ORESTA

DEEP LEARNING BASED CIRRHOSIS DETECTION

Anıl Utku *

Department of Computer Engineering, Munzur University, Tunceli, Turkey

Research paper

Abstract: *Cirrhosis is a liver disease caused by long-term liver damage. Scar tissue caused by cirrhosis prevents the liver from working properly. With the hepatitis C virus, 130-150 million people are infected in the world and 350-500 thousand deaths, and 3-4 million new cases are reported every year due to liver disease. In 2030, it is predicted that there will be 40 percent increase in compensated cirrhosis due to the hepatitis C virus, 60 percent increase in decompensated cirrhosis, and 70 percent increase in liver-related deaths. Although it is difficult to diagnose cirrhosis in the early stages, it is very important step for its treatment. Blood tests, imaging tests, and biopsy methods are used to detect cirrhosis. Due to the costs of these tests and the inability to get the test results immediately, the treatment of the patients cannot be started immediately. In this study, a MLP-based deep learning model has been developed for the prediction of cirrhosis. The developed model has been compared with DT, kNN, LR, NB, RF, and SVM. Experimental studies using the accuracy, precision, recall, and F1-score showed that the developed model was more successful than the compared models. Experimental results showed that the developed model had 80.48% accuracy, 85.71% precision, 85.71% recall, and 85.71% F1-score. Experimental results showed that the developed model had a prediction accuracy of over 80% and F1-score of over 85% in cirrhosis detection from blood tests. The developed model can be used in real-world applications to alleviate the workload of healthcare professionals and to develop early diagnosis systems.*

Keywords: *cirrhosis detection, deep learning, machine learning, MLP.*

1. Introduction

Cirrhosis, also called chronic liver disease, causes severe damage to the liver (Arroyo et al., 2016). Different levels of damage to the liver can occur due to various diseases. As a result, various deteriorations occur in the structural functions of the liver and it cannot perform its normal functions (Garcia-Martinez et al., 2013). This is

* Corresponding author.
anilutku@munzur.edu.tr (A. Utku)

the beginning of the cirrhosis process. As a result of the decrease in liver cells that continue to function, the liver begins to harden and shrink. The flow of blood to the hardened tissues becomes difficult and new vascular pathways are formed due to the inability of the blood to reach the tissue. All these events aggravate the cirrhosis table by affecting the liver more negatively (Mozos, 2015). As a result, the liver begins to fail to function and liver failure occurs.

Cirrhosis is a long-lasting and progressive disease (Vranjkovic et al., 2019). In the early stages, the findings may be very mild. As the damage to the liver increases, the symptoms worsen (Younossi and Henry, 2015). The most common symptoms in the early period are; loss of appetite, weight loss, nausea, weakness, and fatigue. These findings get worse in the future. In this process, water accumulation in the body, edema in the legs, swelling in the abdomen, muscle wasting, rapid bruising on the skin, tendency to bleeding, excessive itching, and jaundice are observed (Pinto et al., 2015).

The liver is the body's factory. All the foods taken are used in the liver to make useful and necessary products for the body. One of them, albumin, keeps the fluids in the blood vessels. When liver functions are impaired, albumin synthesis is also affected (Van Zutphen et al., 2016). When the albumin level decreases, the fluids cannot be kept in the vascular bed and leak into the tissues (Guerci et al., 2019). As a result, edema occurs in the legs. Likewise, fluid accumulates in the abdominal cavity. In these patients, bruises may occur on the skin, or the tendency to bleed increases with the slightest impact (Amitrano et al., 2002). The reason for this is that the substances necessary for coagulation cannot be produced as much as they should due to the damage in the liver. Again, as a result of liver failure, some substances accumulate in the blood and severe itching and encephalopathy may occur. Long-term use of alcohol, viral hepatitis B/C, diabetes, obesity, obstruction and inflammation of the biliary tract, chronic heart failure, history of liver disease, and unprotected sex cause cirrhosis (Ginès et al., 2021).

Blood tests, imaging tests such as MRI or ultrasound, and biopsy methods are used to detect cirrhosis (Acharya et al., 2015). The costs of these tests and the inability to get the test results immediately highlight the use of artificial intelligence technologies in the cirrhosis detection. Artificial intelligence methods are used successfully in many applications in the field of medicine. Systems for the diagnosis of diseases can be developed by using artificial intelligence methods. In the literature, there is no study on the detection of cirrhosis using artificial intelligence methods. In this section, studies in the literature in which artificial intelligence methods are used in the field of medical diagnosis are examined.

Goceri (Goceri, 2019) presented a deep learning-based comparative analysis for the detection of skin diseases. U-Net, InceptionV3, InceptionResNetV2, VGGNet and ResNet models were compared. Experimental results showed that ResNet was the most successful model with 80% accuracy, while U-net was the most unsuccessful model with 0.74 accuracy.

Che Azemin et al. (Che Azemin et al., 2020) presented a ResNet-101-based model for detecting COVID-19 cases using chest radiography images. The developed model had 0.82 AUROC, 77.3% sensitivity, 71.8% specificity, and 71.9% accuracy.

Jain et al. (Jain et al., 2021) developed a deep learning-based model for detecting COVID-19 from medical images. In the study, ResNeXt, Inception V3 and Xception models were compared. Experimental results showed that the Xception model was more successful than the compared models with 97.97% accuracy.

Ismael and Şengür (Ismael and Şengür, 2021) proposed a deep learning-based model for detecting COVID-19 from X-ray data. ResNet18, ResNet50, ResNet101, VGG16 and VGG19 were used for feature extraction. Support Vector Machine (SVM) with different kernel functions is used to classify features. A dataset consisting of 380 X-ray images was used. Experimental results showed that ResNet50 and SVM with linear kernel function outperform other models with 94.7% accuracy.

Mostafa et al. (Mostafa et al., 2021) presented comparative analysis of ANN, RF and SVM for Hepatitis C detection from blood test values. The dataset used consists of blood test results of 615 patients. Experimental results showed that RF, ANN and SVM had 98.14%, 88.89% and 96.75% classification accuracy, respectively.

Allugunti (Allugunti, 2022) aims to classify patients as cancer and non-cancerous using mammography images. In the study, the results of the Convolutional Neural Network (CNN), Random Forest (RF) and SVM classification models were compared. Experimental results showed that CNN has 99.65% accuracy, SVM has 89.84% accuracy, and RF has 90.55% accuracy.

Terlapu et al. (Terlapu et al., 2022) presented comparative analysis of the Probabilistic Neural Network (PNN), SVM, k Nearest Neighbours (kNN), RF, Decision Tree (DT), and Naïve Bayes (NB) for hepatitis C detection. Experimental results showed that PNN outperformed machine learning models with 99.6% accuracy. RF was the most successful model among machine learning models with 97.5% accuracy.

Almadhoun and Abu-Naser (Almadhoun and Abu-Naser, 2022) presented a deep learning-based comparative analysis for brain tumor detection. Using a dataset of 10,000 images, the proposed model, Inception, VGG16, MobileNet, and ResNet models were compared. Experimental results showed that the VGG16 outperformed the compared models with 99.86% accuracy.

Luetkens et al. (Luetkens et al., 2022) presented deep learning-based comparative analysis for cirrhosis detection from liver MR images. The dataset used consists of 465 patient data. Experimental studies using the ResNet50 and DenseNet121 models showed that ResNet50 had 0.823 precision.

Pasyar et al. (Pasyar et al., 2022) developed a hybrid deep learning model for detecting liver diseases such as hepatitis and cirrhosis from ultrasound images. Transfer learning has been applied to the ResNet and AlexNet architectures. The voting method was used to weight the results obtained by each model. Experimental results showed that the hybrid ResNet50 model had classification accuracy of over 86%.

Cirrhosis does not usually cause symptoms in the early stages. However, as the degree of the disease progresses and the level of damage to the liver increases, the symptoms and the severity of these symptoms increase. Liver diseases cause other diseases in the body and can pose great dangers to the body. For this reason, diagnosis and treatment of liver diseases such as cirrhosis at an early stage are vital

for the human body. Blood tests are mainly used to diagnose liver diseases. However, cirrhosis diagnosis made with traditional methods. These methods have limitations in terms of cost, time, and accuracy compared to artificial intelligence methods. Especially in hospitals where these tests are carried out intensively, limitations in terms of the decision-making phase and time come to the fore in terms of the human factor. The motivation of this study is to offer a solution to the limitations of traditional methods used in cirrhosis detection. In this study, a deep learning model was developed to detect cirrhosis patients from blood test values. It was aimed to provide a pre-diagnosis system for the detection of cirrhosis by easing the workload of healthcare professionals.

The main contributions of this study to the literature can be summarized as follows:

- This is the first study in the literature on cirrhosis detection using artificial intelligence methods.

- In this study, Multilayer Perceptron (MLP) based deep learning model were developed for cirrhosis detection.

- The developed model were compared with DT, kNN, Logistic Regression (LR), NB, RF and SVM using accuracy, precision, recall and F1-score.

- Experimental results showed that the developed model was more successful than the compared models with 80.48% classification accuracy.

2. Deep Learning Based Cirrhosis Detection

Many diseases cause people to die as a result of late diagnosis. For this reason, experts recommend that it is appropriate to perform screening tests at regular intervals. Despite this recommendation of experts, many people do not care about health screenings and do not go to the doctor without any symptoms. Artificial intelligence is used for the effective use of human resources, rapid diagnosis, and treatment, supporting healthcare professionals in many ways. Artificial intelligence is used in the diagnosis and treatment of diseases, in determining the appropriate tools for treatment and in medical decision support systems. In this study, it was aimed to determine whether people had cirrhosis according to their demographic characteristics and blood test results. For this purpose, MLP-based deep learning model was developed. The developed model was compared with DT, kNN, LR, NB, RF and SVM using accuracy, precision, recall, and F1-score.

2.1. Dataset

The dataset used in this study consists of patients' demographic information and blood test values. The dataset is publicly available via https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset. The dataset consists of blood test data from a total of 418 PBC patients. The dataset consists of 19 attributes. 'ID' is a unique sequence number of patients.

'N_Days' represents the number of days from the patient's registration date to the date of the transplant, the date of the patient's death, or July 1986. 'Status' refers to

the patient's status as C (censored), CL (censored due to liver transplant), or D (death). 'Drug' refers to the type of drug, either D-penicillamine or placebo.

'Age' refers to the patient's age in days. Sex denotes gender as M for males and F for females. 'Acid' refers to the presence of acid as N for no and Y for yes. 'Hepatomegaly' refers to the presence of hepatomegaly as N for no and Y for yes. 'Spiders' refers to the presence of spiders as N for no and Y for yes.

'Edema' refers to the presence of edema as N (no edema and no diuretic therapy for edema), S (existing edema without diuretics or edema resolving with diuretics), or Y (edema despite diuretic therapy). 'Bilirubin' refers to the amount of serum bilirubin in mg/dl. 'Cholesterol' means the amount of cholesterol in mg/dl. 'Albumin' refers to the amount of albumin in mg/dl. 'Copper' refers to the amount of copper in the urine in µg/day. 'Alk_Phos' refers to the amount of alkaline phosphatase in U/liter.

'SGOT' refers to the SGOT value in U/ml. 'Triglycerides' refers to the triglycerides value in mg/dl. 'Platelets' refers to the platelet's value per cubic ml/1000. 'Prothrombin' refers to the prothrombin time in seconds.

'Stage' refers to the stage of the disease as 1, 2, 3, or 4. In the dataset, missing data (NA) of the features are filled with the mean values of the relevant column. The categorical features in the dataset were replaced with a numeric value.

The categorical values of the 'Sex' attribute were changed to 0 for 'M' and 1 for 'F'. The categorical values of the 'ascites' attribute were changed to 0 for 'N' and 1 for 'Y'.

The categorical values of the 'Drug' attribute were changed to 0 for 'D-penicillamine' and 1 for 'Placebo'. The categorical values of the 'hepatomegaly' attribute were changed to 0 for 'N' and 1 for 'Y'.

The categorical values of the 'Spiders' attribute were changed to 0 for 'N' and 1 for 'Y'. The categorical values of the 'Edema' attribute were changed to 0 for 'N', 1 for 'Y', and -1 for 'S'. The categorical values of the 'status' attribute were replaced with 0 for 'C', 1 for 'CL', and -1 for 'D'.

The distribution of the patients according to the disease stages is shown in Fig 1.
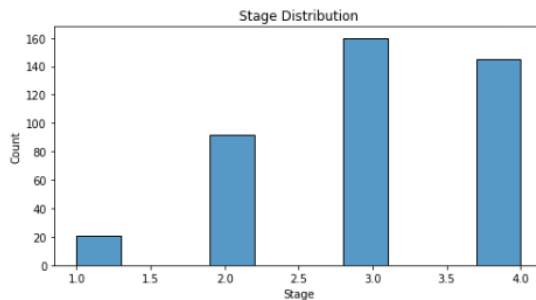

Figure 1. The distribution of the patients according to the disease stages

The heatmap showing the relationships of the features in the dataset is shown in Fig. 2. The heatmap was used for feature selection.
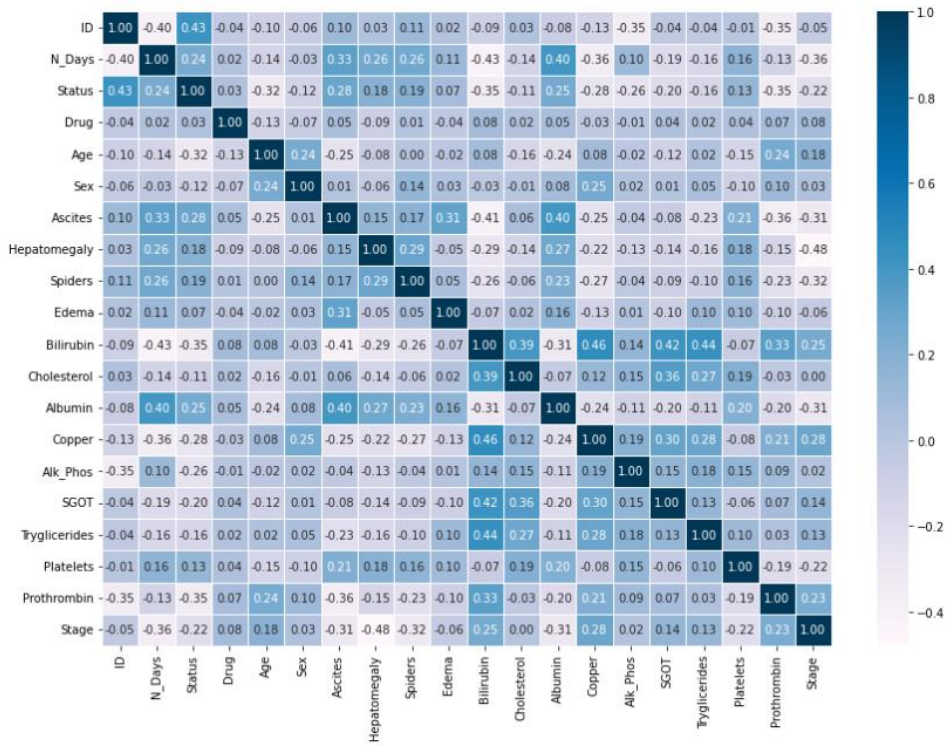
Figure 2. The heatmap used for feature selection

The distributions of the features are shown in Fig. 3.

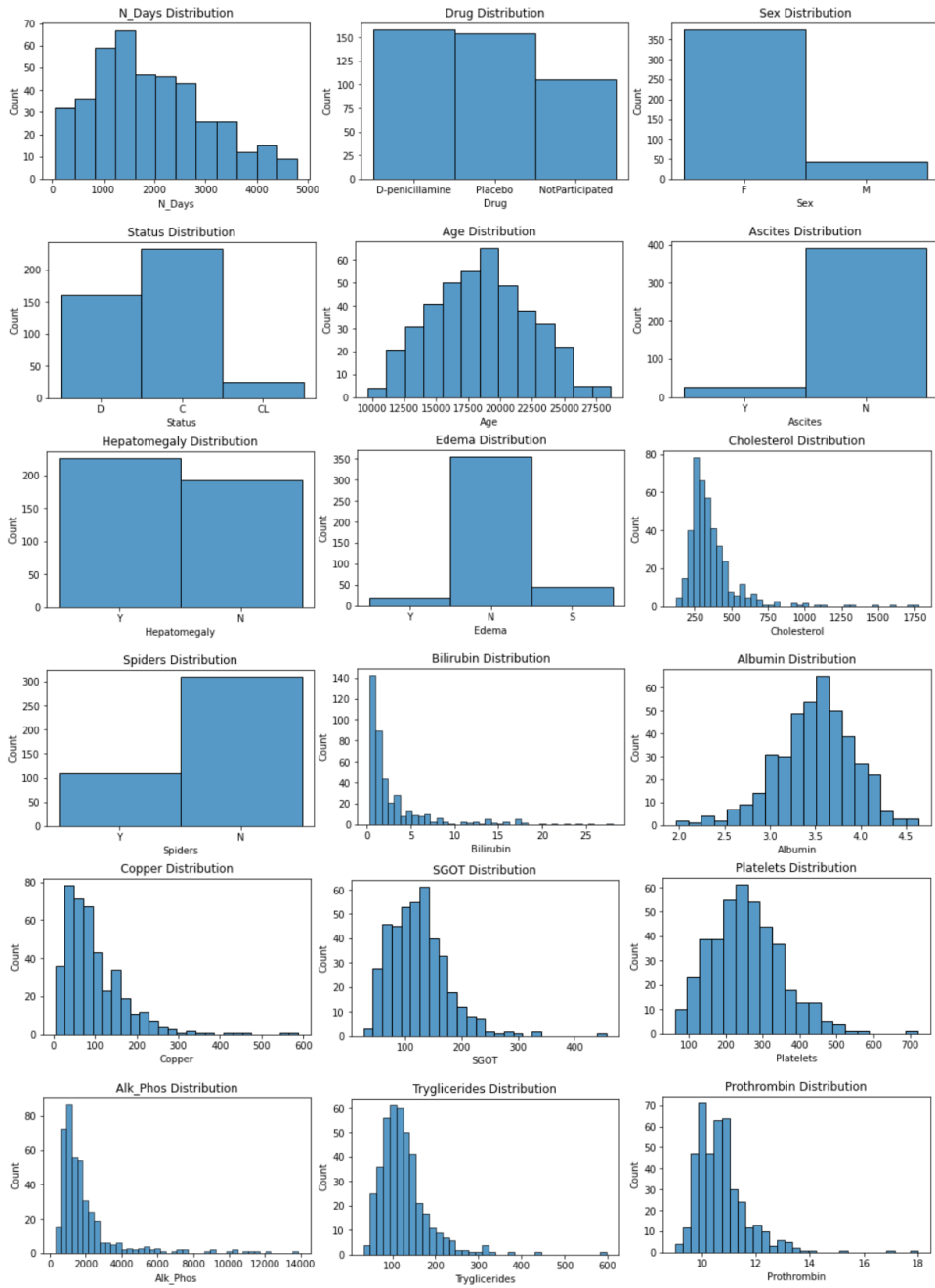Deep Learning Based Cirrhosis Detection

Figure 3. The distributions of the features

The attributes with categorical values in the dataset are shown in Figure 4.
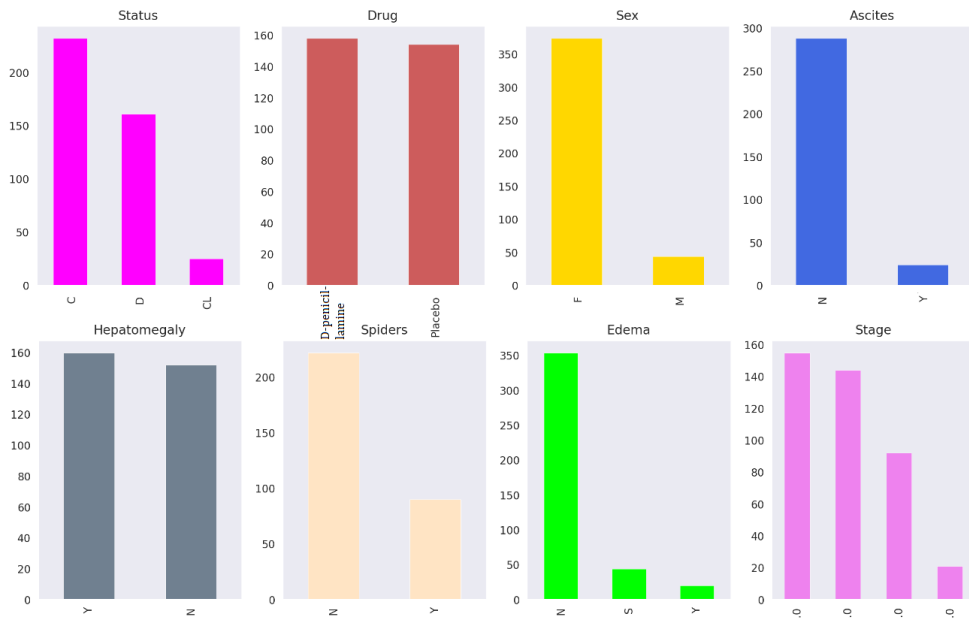
Figure 4. The attributes with categorical values in the dataset

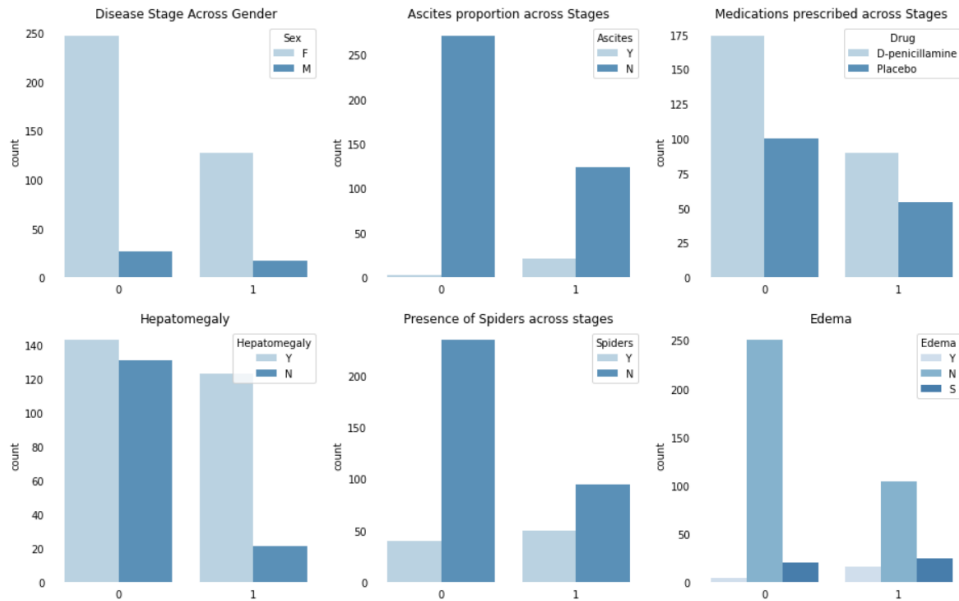Relationships between features other than blood values are shown in Fig. 5.

Figure 5. Relationships between features other than blood values

The density graphs of the features obtained from the blood tests and the age feature according to their classes are shown in Fig. 6.
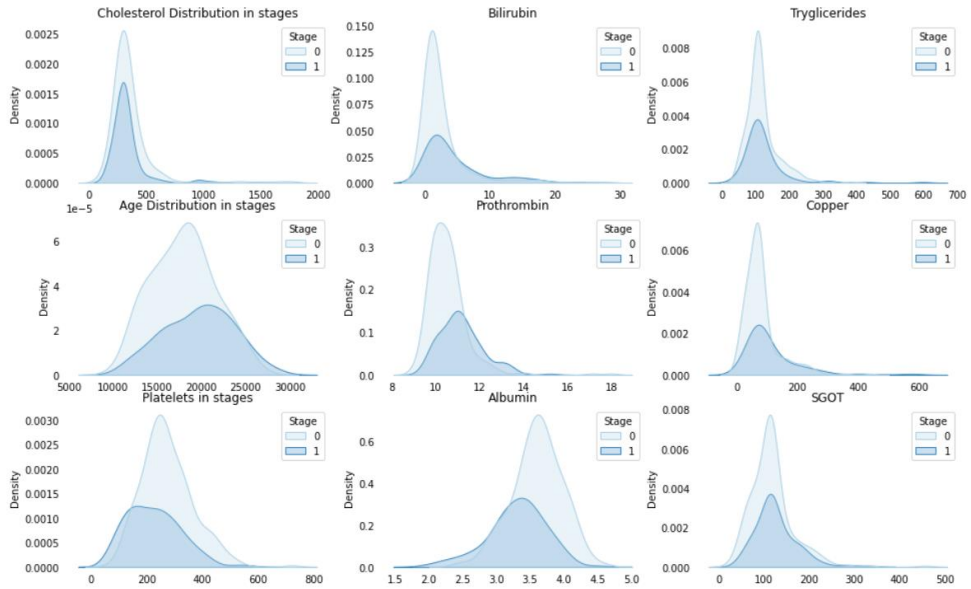
Figure 6. The density graphs of the features obtained from the blood tests and the age feature

2.2. Developed Model

MLP has a structure in which many neurons with non-linear activation functions are hierarchically connected. MLP consists of an input, one or more intermediate, and an output layer. The input layer receives the input strings to be processed. Inputs are forwarded to the network using the weights between the input layer and the hidden layers. In hidden layers, activation functions such as ReLU, sigmoid, and tanh are used. Input sequences processed in hidden layers are transmitted over the network with the help of these activation functions. These processes are repeated as many times as the number of hidden layers in the network structure. The output layer performs tasks such as regression or classification. In the output layer, activation functions are used according to the type of problem. For example, sigmoid is used for binary classification and softmax activation functions are used for multi-class classification problems. The neurons in the MLP are trained using the backpropagation algorithm.

The developed MPL-based model takes the blood test data of the patients as input and predicts whether the patient has cirrhosis. The architecture of the developed model is shown in Fig. 7.
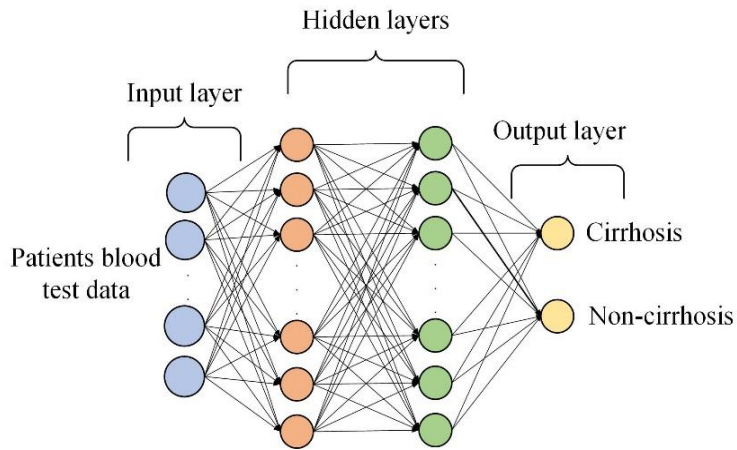
Figure 7. The architecture of the developed MLP-based model

The developed model consists of an input layer in which demographic data of the patients and blood test data are presented as input. There are 2 hidden layers for the model to calculate. Hyperparameter analysis studies were carried out using GridSearchCV to determine the number of neurons and epochs in the hidden layers. ReLU activation function is used in the input layer. In hidden layers, ReLu activation function is used to sort the layers and make nonlinear calculations. Since the binary classification is done, the sigmoid activation function is used in the output layer.

2.3. Evaluation Metrics

Classification algorithms aim to predict categorical values with two or more classes. Accuracy, precision, recall, and F1-score metrics are used to measure the performance of classification algorithms. These metrics are calculated using the confusion matrix. The confusion matrix is used to interpret the results of classification models and evaluate the relationship between actual values and predicted values. The confusion matrix is shown in Table 1.

Table 1. Confusion matrix

		Actual values	
		Positive (1)	Negative (0)
Predicted values	Positive (1)	TP	FP
	Negative (0)	FN	TN

TP is the number of patients that are actually cirrhosis and the classifier also predicts cirrhosis. TN is the number of cases that are actually non-cirrhosis and the classifier also predicts non-cirrhosis. FP is the number of patients that the classifier predicts as cirrhosis, but non-cirrhosis. FN is the number of patients that the classifier predicts as non-cirrhosis, but cirrhosis.

Accuracy is calculated as the ratio of the number of samples classified correctly by the model to the total number of samples, as seen in Eq. (1).

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

Precision refers to how many of the positively predicted values are actually positive. Precision is calculated using Eq. (2).

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

Recall is a metric that shows how many of the samples that should be predicted positively are correctly predicted. Recall is calculated using Eq. (3).

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

The F1-score is calculated using precision and recall values. F1-score is calculated as seen in Eq. (4).

$$F1-\text{score} = \frac{2.\text{Precision}.\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

3. Experimental Results

In this study, MLP-based deep learning model was developed to detect cirrhosis patients. The experimental results of the developed model were extensively compared with DT, kNN, LR, NB, RF and SVM. The accuracy, precision, recall, and F1-score obtained for each model were compared. Parameter analysis studies were conducted using GridSearchCV to determine the parameters of models.

Cross validation has been used to eliminate the overfitting problem and to increase the quality of the models. Cross-validation has been made by choosing the k value as 10. All models were run on 10 randomly generated datasets using cross-validation, and the results obtained were averaged.

The confusion matrix and experimental results for DT are shown in Table 2 and Table 3.

Table 2. Confusion matrix for DT

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	20	6
	Non-cirrhosis (0)	8	7

As seen in Table 2, TP is 20, FP is 6, FN is 8, and TN is 7. Experimental results showed that DT correctly detected 20 of 28 cirrhosis patients and 7 of 13 non-

cirrhosis patients. Of the 41 patients in the dataset, 27 patients were correctly classified and 14 patients were misclassified.

Table 3. Accuracy, precision, recall and F1-score values for DT

Accuracy	Precision	Recall	F1-score
0.6585	0.7692	0.7142	0.7406

As seen in Table 3, the accuracy of DT is 0.6585, the precision is 0.7692, the recall is 0.7142, and the F1-score is 0.7406. The confusion matrix and experimental results for kNN are shown in Table 4 and Table 5.

Table 4. Confusion matrix for kNN

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	18	9
	Non-cirrhosis (0)	10	4

As seen in Table 4, TP is 18, the FP is 9, FN is 10, and TN is 4. Experimental results showed that kNN correctly detected 18 of 28 cirrhosis patients and 4 of 13 non-cirrhosis patients. Of the 41 patients in the dataset, 22 patients were correctly classified and 19 patients were misclassified.

Table 5. Accuracy, precision, recall and F1-score values for kNN

Accuracy	Precision	Recall	F1-score
0.5365	0.6666	0.6428	0.6544

As seen in Table 5, the accuracy of kNN is 0.5365, the precision is 0.6666, the recall is 0.6428, and the F1-score is 0.6544.

The confusion matrix and experimental results for LR are shown in Table 6 and Table 7.

Table 6. Confusion matrix for LR

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	20	7
	Non-cirrhosis (0)	8	6

As seen in Table 6, TP is 20, FP is 7, FN is 8, and TN is 6. Experimental results showed that LR correctly detected 20 of 28 cirrhosis patients and 6 of 13 non-cirrhosis patients. Of the 41 patients in the dataset, 26 patients were correctly classified and 15 patients were misclassified.

Table 7. Accuracy, precision, recall and F1-score values for LR

Accuracy	Precision	Recall	F1-score
0.6341	0.7407	0.7142	0.7272

As seen in Table 7, the accuracy of LR is 0.6341, the precision is 0.7407, the recall is 0.7142, and the F1-score is 0.7272.

The confusion matrix and experimental results for NB are shown in Table 8 and Table 9.

Table 8. Confusion matrix for NB

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	15	11
	Non-cirrhosis (0)	13	2

As seen in Table 8, TP is 15, FP is 11, FN is 13, and TN is 2. Experimental results showed that NB correctly detected 15 of 28 cirrhosis patients and 2 of 13 non-cirrhosis patients. Of the 41 patients in the dataset, 17 patients were correctly classified and 24 patients were misclassified.

Table 9. Accuracy, precision, recall and F1-score values for NB

Accuracy	Precision	Recall	F1-score
0.4146	0.5769	0.5357	0.5555

As seen in Table 9, the accuracy of NB is 0.4146, the precision is 0.5769, the recall is 0.5357, and the F1-score is 0.5555.

The confusion matrix and experimental results for RF are shown in Table 10 and Table 11.

Table 10. Confusion matrix for RF

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	22	5
	Non-cirrhosis (0)	6	8

As seen in Table 10, TP is 22, FP is 5, FN is 6, and TN is 8. Experimental results showed that RF correctly detected 22 of 28 cirrhosis patients and 8 of 13 non-cirrhosis patients. Of the 41 patients in the dataset, 30 patients were correctly classified and 11 patients were misclassified.

Table 11. Accuracy, precision, recall and F1-score values for RF

Accuracy	Precision	Recall	F1-score
0.7317	0.8148	0.7857	0.7999

As seen in Table 11, the accuracy of RF is 0.7317, the precision is 0.8148, the recall is 0.7857, and the F1-score is 0.7999.

The confusion matrix and experimental results for SVM are shown in Table 12 and Table 13.

Table 12. Confusion matrix for SVM

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	22	5
	Non-cirrhosis (0)	6	8

As seen in Table 12, TP is 22, FP is 5, FN is 6, and TN is 8. Experimental results showed that SVM correctly detected 22 of 28 cirrhosis patients and 8 of 13 non-cirrhosis patients. Of the 41 patients in the dataset, 30 patients were correctly classified and 11 patients were misclassified.

Table 13. Accuracy, precision, recall and F1-score values for RF

Accuracy	Precision	Recall	F1-score
0.7317	0.8148	0.7857	0.7999

As seen in Table 13, the accuracy of SVM is 0.7317, the precision is 0.8148, the recall is 0.7857, and the F1-score is 0.7999.

The confusion matrix and experimental results for developed model are shown in Table 14 and Table 15.

Table 14. Confusion matrix for developed model

		Actual values	
		Cirrhosis (1)	Non-cirrhosis (0)
Predicted values	Cirrhosis (1)	24	4
	Non-cirrhosis (0)	4	9

As seen in Table 14, TP is 24, FP is 4, FN is 4, and TN is 9. Experimental results showed that the developed model correctly detected 24 of 28 cirrhosis patients and 9 of 13 non- cirrhosis patients. Of the 41 patients in the dataset, 33 patients were correctly classified and 8 patients were misclassified.

Table 15. Accuracy, precision, recall and F1-score values for developed model

Accuracy	Precision	Recall	F1-score
0.8048	0.8571	0.8571	0.8571

As seen in Table 15, the accuracy of proposed model is 0.8048, the precision is 0.8571, the recall is 0.8571, and the F1-score is 0.8571.

Comparative experimental results according to accuracy, precision, recall and F1-score values for DT, kNN, LR, NB, RF, SVM and developed model are shown in Table 16 and Fig. 8.

Table 16. Comparative experimental results

Model	Accuracy	Precision	Recall	F1-score
DT	0.6585	0.7692	0.7142	0.7406
kNN	0.5365	0.6666	0.6428	0.6544
LR	0.6341	0.7407	0.7142	0.7272
NB	0.4146	0.5769	0.5357	0.5555
RF	0.7317	0.8148	0.7857	0.7999
SVM	0.7317	0.8148	0.7857	0.7999
Developed model	0.8048	0.8571	0.8571	0.8571

As seen in Table 16 and Fig. 8, the developed model has more successful results than the other models compared.
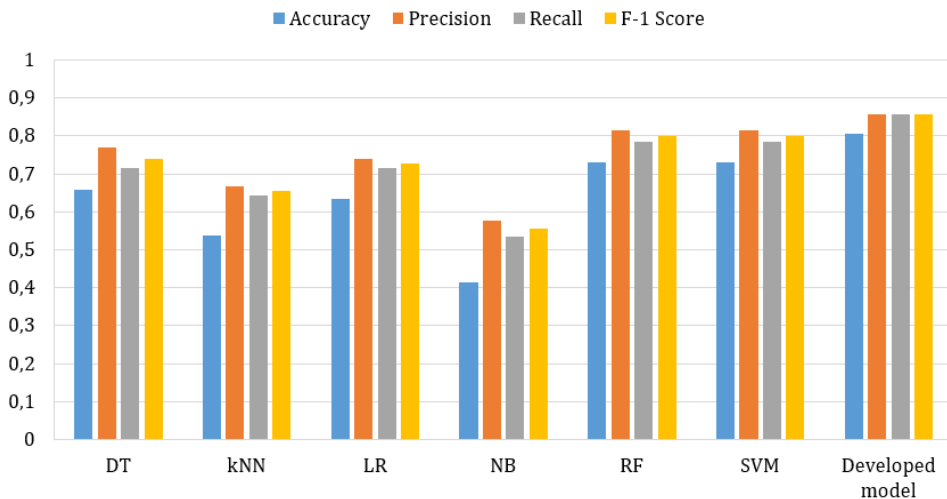

Figure 8. Comparative experimental results

As can be seen in Table 16 and Fig. 8, the developed model showed better classification performance in detecting cirrhosis patients compared to other models. After the developed model, RF, SVM, DT, LR, kNN and NB are the models with the successful results, respectively.

Experimental results showed that the developed model had a classification accuracy of over 80% and an F1-score close to 86%. The obtained results showed

that the developed model can be successfully applied in the cirrhosis detection and can be used in early diagnosis systems.

The accuracy/loss graphs of the developed model during the training and validation are shown in Fig. 9.
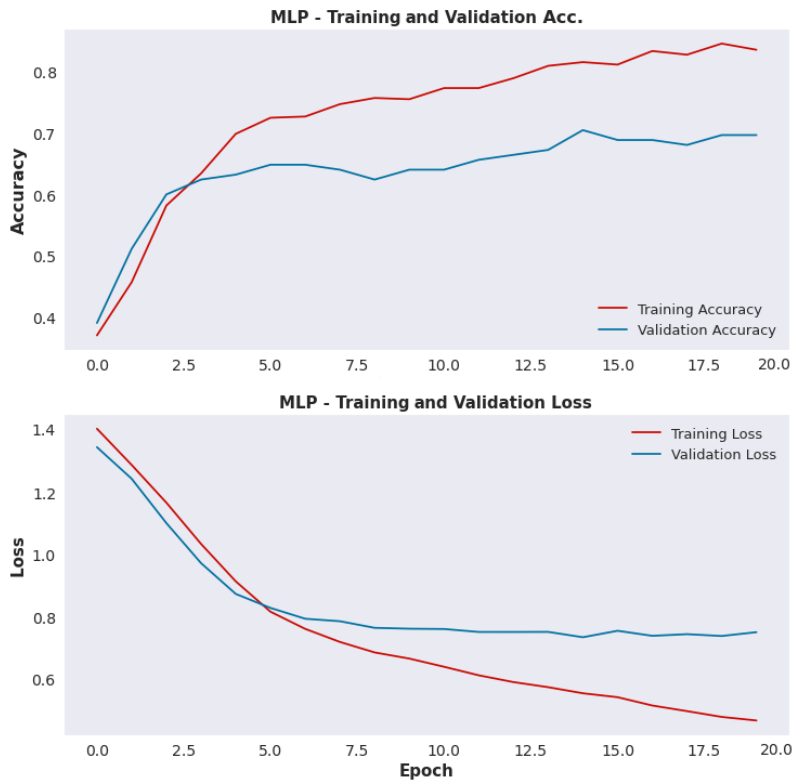

Figure 9. The accuracy/loss graphs of the developed model during the training and validation

Experimental results showed that the proposed model was more successful than other models compared according to accuracy, recall and F1-score metrics. The accuracy value is important because it shows the number of correctly classified patients. Precision value is important because it shows how many patients with predicted cirrhosis have cirrhosis. The Recall value is important because it shows how many of the patients who should have been predicted as cirrhosis were correctly predicted. F1-score is a harmonic mean value calculated according to precision and recall values. In other words, successful models are expected to have a higher F1-score value.

ROC and precision-recall curve graphs of the developed model are shown in Fig. 10.a and Fig. 10.b.
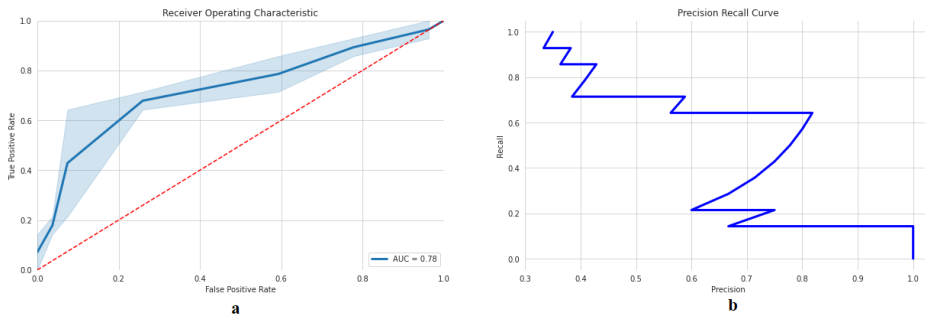
Figure 10. ROC and precision-recall curve graphs of the developed model

The ROC curve is a very important performance measure for classification problems. ROC is a probability curve. AUC is the area under the curve and represents the degree or measure of discrepancy. In the ROC curve, there is False Positive Rate (FPR) on the X axis and True Positive Rate (TPR) on the Y axis. As the area under the curve increases, the discrimination performance between classes increases. TPR is the recall value. In other words, it is the detection rate of cirrhosis patients. FPR is the rate of erroneous prediction for non-cirrhosis patients.

4. Conclusions

Cirrhosis is an advanced chronic liver disease. Different levels of damage occur in the liver due to different diseases. Depending on these reasons, the cirrhosis process begins with the development of structural changes in the liver. As a result, the number of functional liver cells decreases, and the liver hardens and shrinks. The resistance to the blood that has to pass through increases. When the blood cannot flow from here, the intravascular pressure increases in the areas where the blood comes from. The blood, which cannot reach the liver due to the increased pressure, looks for other ways to reach the liver and creates new vascular pathways. As a result, liver functions gradually deteriorate and signs of liver failure occur.

Artificial intelligence technologies are used in medical application areas such as disease diagnosis, surgery, drug development, analysis of radiological images and lesions, and personalized therapy. In this study, MLP-based deep learning model was developed for cirrhosis detection. The developed model was compared with DT, kNN, LR, NB, RF, and SVM. Experimental studies using the accuracy, precision, recall, and F1-score showed that the developed model was more successful than the compared models. Experimental results showed that the developed model has 80.48% accuracy, 85.71% precision, 85.71% recall and 85.71% F1-score.

The fact that RF is more successful than DT can be interpreted by the bagging technique of RF. RF creates multiple decision trees. It evaluates the results of these trees using the voting method.

The fact that RF is more successful than kNN can be explained as random samples selected from the dataset. kNN only tries to include the closest instances in the same class. However, in this case, samples outside of local similarity will not be classified correctly by kNN.

The fact that RF is more successful than NB can be interpreted as NB's inability to represent the complex behavior of models due to the small model size. On the other hand, the size of RF is very large. RF can adapt to the dynamic structure and change of data.

The fact that SVM is more successful than RF can be interpreted as the presence of numerical and categorical features in the dataset. RF works with a mixture of numerical and categorical features. RF is advantageous when features are of various scales. SVM maximizes the margin between different points and calculates the distance between points. In the classification problem, RF gives the probability of belonging to the class, while SVM gives the points closest to the boundary between classes. Due to the categorical and numerical coexistence of the features in the data, RF performed better than SVM.

The fact that the developed model is more successful than other models can be interpreted as a large number of input data presented to the network. The more inputs presented to the network, the network learns better and makes better predictions. But basically machine learning methods require fewer input data.

Studies in the literature have generally focused on hepatitis C and liver diseases detection. Disease detection is made from blood test values with ultrasound or MR images. To the best of our knowledge, there is no study in the literature for cirrhosis detection using this dataset. Therefore, the experimental results could not be compared with the studies in the literature.

The size of the dataset used is one of the limitations of this study. The dataset consists of 418 PVC patient data and 19 attributes. Making more different measurements in blood tests and increasing the number of patients can enable the developed model to be trained more successfully. In addition, classification accuracy can be increased by using advanced deep learning models such as more hybrid models.

References

Acharya U. R., Faust O., Molinari F., Sree S. V., Junnarkar S. P., Sudarshan V. (2015). Ultrasound-based tissue characterization and classification of fatty liver disease: A screening and diagnostic paradigm. Knowledge-Based Systems, 75, 66-77. https://doi.org/10.1016/j.knosys.2014.11.021

Aksu D., Üstebay S., Aydin M. A., Atmaca T. (2018). Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm. In International symposium on computer and information sciences, 141-149. https://doi.org/10.1007/978-3-030-00840-6_16

Allugunti V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. International Journal of Engineering in Computer Science, 4(1), 49-56.

Almadhoun H. R., Abu-Naser S. S. (2022). Detection of Brain Tumor Using Deep Learning. International Journal of Academic Engineering Research (IJAER), 6(3).

Amitrano L., Guardascione M. A., Brancaccio V., Balzano A. (2002). Coagulation disorders in liver disease. In Seminars in liver disease, 22 (1), 83-96. https://doi.org/10.1055/s-2002-23205

Arroyo V., Moreau R., Kamath P. S., Jalan R., Ginès P., Nevens F., Schnabl B. (2016). Acute-on-chronic liver failure in cirrhosis. Nature reviews Disease primers, 2(1), 1-18. https://doi.org/10.3390/jcm10194406

Che Azemin M. Z., Hassan R., Mohd Tamrin M. I., Md Ali M. A. (2020). COVID-19 deep learning prediction model using publicly available radiologist-adjudicated chest X-ray images as training data: preliminary findings. International Journal of Biomedical Imaging. https://doi.org/10.1155/2020/8828855

Garcia-Martinez R., Caraceni P., Bernardi M., Gines P., Arroyo V., Jalan R. (2013). Albumin: pathophysiologic basis of its role in the treatment of cirrhosis and its complications. Hepatology, 58(5), 1836-1846. https://doi.org/10.1002/hep.26338

Ginès P., Krag A., Abraldes J. G., Solà E., Fabrellas N., Kamath P. S. (2021). Liver cirrhosis. The Lancet, 398(10308), 1359-1376. https://doi.org/10.1016/S0140-6736(21)01374-X

Goceri E. (2019). Skin disease diagnosis from photographs using deep learning. In ECCOMAS thematic conference on computational vision and medical image processing, 239-246. https://doi.org/10.1007/978-3-030-32040-9_25

Guerci P., Ergin B., Uz Z., Ince Y., Westphal M., Heger M., Ince C. (2019). Glycocalyx degradation is independent of vascular barrier permeability increase in nontraumatic hemorrhagic shock in rats. Anesthesia & Analgesia, 129(2), 598-607. https://doi.org/10.1213/ane.0000000000003918

Ismael A. M., Şengür A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Systems with Applications, 164, 114054. https://doi.org/10.1016/j.eswa.2020.114054

Jadhav S. D., Channe H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. International Journal of Science and Research (IJSR), 5(1), 1842-1845. https://doi.org/10.21275/v5i1.nov153131

Jain R., Gupta M., Taneja S., Hemanth D. J. (2021). Deep learning based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence, 51(3), 1690-1700.

Luetkens J. A., Nowak S., Mesropyan N., Block W., Praktiknjo M., Chang J., Attenberger U. (2022). Deep learning supports the differentiation of alcoholic and other-than-alcoholic cirrhosis based on MRI. Scientific reports, 12(1), 1-8. https://doi.org/10.1038/s41598-022-12410-2

Mostafa F., Hasan E., Williamson M., Khan H. (2021). Statistical Machine Learning Approaches to Liver Disease Prediction. Livers, 1(4), 294-312. https://doi.org/10.3390/livers1040023

Mozos I. (2015). Arrhythmia risk in liver cirrhosis. World journal of hepatology, 7(4), 662. https://doi.org/10.1007/s10489-020-01902-1

Pasyar P., Mahmoudi T., Kouzehkanan S. Z. M., Ahmadian A., Arabalibeik H., Soltanian N., Radmard A. R. (2021). Hybrid classification of diffuse liver diseases in ultrasound images using deep convolutional neural networks. Informatics in Medicine Unlocked, 22. https://doi.org/10.1016/j.imu.2020.100496

Pinto R. B., Schneider A. C. R., da Silveira, T. R. (2015). Cirrhosis in children and adolescents: An overview. World journal of hepatology, 7(3), 392. https://doi.org/10.4254/wjh.v7.i3.392

Terlapu P. V., Gedela S. B., Gangu V. K., Pemula R. (2022). Intelligent diagnosis system of hepatitis C virus: A probabilistic neural network based approach. International Journal of Imaging Systems and Technology. https://doi.org/10.1002/ima.22746

Van Zutphen T., Ciapaite J., Bloks V. W., Ackereley C., Gerding A., Jurdzinski A., Bandsma R. H. (2016). Malnutrition-associated liver steatosis and ATP depletion is caused by peroxisomal and mitochondrial dysfunction. Journal of hepatology, 65(6), 1198-1208. https://doi.org/10.1016/j.jhep.2016.05.046

Vranjkovic A., Deonarine F., Kaka S., Angel J. B., Cooper C. L., Crawley A. M. (2019). Direct-acting antiviral treatment of HCV infection does not resolve the dysfunction of circulating CD8+ T-cells in advanced liver disease. Frontiers in immunology, 10, 1926. https://doi.org/10.3389/fimmu.2019.01926

Younossi Z., Henry L. (2015). Systematic review: patient-reported outcomes in chronic hepatitis C-the impact of liver disease and new treatment regimens. Alimentary pharmacology & therapeutics, 41(6), 497-520. https://doi.org/10.1111/apt.13090