Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi
Afyon Kocatepe University – Journal of Science and Engineering
https://dergipark.org.tr/tr/pub/akufemubid

Araştırma Makalesi / Research Article
e-ISSN: 2149-3367
DOI: https://doi.org/10.35414/akufemubid.1580440
AKÜ FEMÜBİD **25** (2025) 045203 (827-837)
AKU J. Sci. Eng. **25** (2025) 045203 (827-837)

Decision Support Systems Based on Machine Learning for Detection of Liver Diseases

Karaciğer Hastalıklarının Tespiti İçin Makine Öğrenmesine Dayalı Karar Destek Sistemleri

Makale Bilgisi / Article Info
Alındı/Received: 06.11.2024
Kabul/Accepted: 28.02.2025
Yayımlandı/Published: 04.08.2025

Pınar İLTER [1*], Yasin KIRELLİ[2]

[1] *İstinye University, Data Science Department, İstanbul, Türkiye*
[2] *Kutahya Dumlupınar University, Management Information Systems Department, Kütahya, Türkiye*

© 2025 The Authors | Creative Commons Attribution-Noncommercial 4.0 (CC BY-NC) International License

Abstract

The liver, the largest metabolic organ in the human body, plays a crucial role in maintaining essential functions. However, the rise in liver diseases in recent years has become a significant public health issue. There are various causes of these diseases, and the impact of technology in the health field has become necessary to analyze and manage diseases. Although there are various causes of liver diseases, analyzing and diagnosing them is crucial. Technological developments have made it possible to use machine learning methods to analyze large data sets. In this study, machine learning techniques are used in the early diagnosis and classification of liver disease, and the results are evaluated. The 'ILPD (Indian Liver Patient Dataset)' dataset in the UCI Machine Learning Repository has been evaluated using six modern machine learning techniques, including logistic regression, random forest, and decision trees. The study also analyses the effects of data preprocessing, feature selection, and normalization on model performance. Model results show that feature selection and normalization are essential in improving model accuracy. The Logistic Regression model achieved the best performance with an accuracy of 80%, while other algorithms, such as Random Forest and AdaBoost, also achieved high accuracy rates. This research demonstrates the potential of machine learning techniques in the early detection of liver diseases and provides a basis for future work.

Keywords: *Liver Patient Detection; Machine Learning; Health Data; Decision Support Systems.*

Öz

İnsan vücudundaki en büyük metabolik organ olan karaciğer, yaşamsal işlevleri yerine getirmesi açısından kritik bir öneme sahiptir. Ancak son yıllarda artan karaciğer hastalıkları toplum sağlığı üzerinde ciddi bir sorun haline gelmiştir. Bu hastalıkların çeşitli nedenleri bulunmakta olup teknolojinin sağlık alanındaki etkisi ile hastalıkların analizi ve yönetimi önemli hale gelmiştir. Karaciğer hastalıklarının çeşitli etkenleri bulunmakla birlikte, bu hastalıkların analizi ve teşhisi önemlidir. Teknolojik gelişmeler, büyük veri setlerinin analizinde makine öğrenimi yöntemlerini kullanmayı mümkün kılar hale getirmiştir. Bu çalışmada karaciğer hastalığı erken teşhisi ve sınıflandırmada makine öğrenmesi teknikleri kullanılmış ve sonuçları değerlendirilmiştir. UCI Machine Learning Repository'de bulunan "ILPD (Indian Liver Patient Dataset)" veri seti ile Lojistik Regresyon, Rastgele Orman, Karar Ağaçları gibi altı farklı modern makine öğrenmesi teknikleri ile değerlerndirilmiştir. Çalışma ek olarak veri ön işleme, özellik seçimi ve normalizasyonun model performansı üzerindeki etkilerini de analiz etmektedir. Model sonuçları özellik seçimi ve normalizasyonun model doğruluğunu artırmada önemli rol oynadığını göstermektedir. Lojistik Regresyon modeli %80 oranında doğruluk ile en iyi performans elde edilmiştir, Rastgele Orman ve AdaBoost gibi diğer algoritmalar da yüksek doğruluk oranlarına ulaşmıştır. Bu araştırma, karaciğer hastalıklarının erken teşhisinde makine öğrenmesi tekniklerinin potansiyelini ortaya koymakta ve gelecekteki çalışmalar için bir temel sunmaktadır.

Anahtar Kelimeler: *Karaciğer Hastalığı Tespiti; Makine Öğrenmesi; Sağlık Verisi; Karar Destek Sistemleri*

1. Introduction

Liver is a vitally important metabolic organ, with essential tasks such as clearing toxins from the gastrointestinal blood and producing hormones and enzymes. In the case of disease, the body signals many health problems with these reactions. Due to its role in numerous metabolic and detoxification processes, the liver is frequently affected secondarily by disorders originating in other organ systems (Hashem and Mabrouk 2014). Several factors cause liver disease, and if these factors are not detected early, the disease can progress and lead to severe liver failure. In studies, it has been observed that contaminated food consumption, genetic disorders, excessive fat accumulation, hepatocyte damage due to bacterial, viral or bacterial viral or fungal infections, diabetes, and excessive consumption of alcohol or medications are associated with liver disease (Pan et al., 2022). Liver diseases can sometimes have no symptoms

*Sorumlu Yazar/Corresponding Author: Pınar İLTER e-posta/e-mail: pinar.ilter@stu.istinye.edu.tr

in the early stages or can be vague, such as weakness and loss of energy. Symptoms often vary depending on the type and extent of liver disease (Haeng Hur et al., 2024; Wu and Moeckel 2023; Yan et al., 2024).

Early diagnosis of liver diseases, prevention of the disease, and early initiation of the treatment process contribute to more promising outcomes for patients. In addition, disease-causing factors may show different trends regionally in the population. In a study on non-alcoholic fatty liver disease (NaFLD), which has recently become a public problem, it has been reported that it is increasing in the young population, especially in adults, in developed countries. The disease begins with an abnormal accumulation of triglycerides in the liver and can cause an inflammatory response in some patients, progressing to cirrhosis and liver cancer. NAFLD is strongly associated with obesity and insulin resistance, but its pathogenesis is still poorly understood, and treatment options are limited (Cohen et al., 2011). Invasive and non-invasive methods are applied to determine the diagnosis of liver diseases.

The widespread use of technology in the healthcare sector enables different methods to diagnose various diseases. One of these methods is machine learning techniques, which can provide significant results. Data stored in digital environments are analyzed using machine learning, statistical and mathematical methods, and numerical learning models are created (Guo et al., 2022; Priya et al., 2018; Yu et al., 2024; Eyupoglu et al., 2024). Machine learning techniques have shown promising results in predicting and classifying liver diseases based on patient data. The use of machine learning techniques in liver disease prediction and classification is a dynamic area of research to reduce healthcare costs (Tokala et al., 2023). These studies are important in developing an effective decision-support system in medical processes. To achieve this goal, extracting and gaining insights from raw data is critical, which is accomplished through knowledge discovery and data mining (KDD) processes. This study conducted to predict liver disease using machine learning methods with the ILDP dataset. The main contributions of this study can be summarized as follows:

- Implementation of multiple machine learning algorithms to assess their effectiveness in liver disease detection.
- Evaluation of the impact of feature selection and normalization on model performance.
- Comparative analysis of different machine learning techniques, including Random Forest and AdaBoost and Ensemble learning.
- Establishing a foundation for future research in liver disease diagnosis using machine learning.

2. Literature Research

The liver performs many vital functions in the body, so if it becomes diseased or damaged, losing these functions can severely impact health. Referred to as hepatic disease, liver disease is a general term that includes all issues impairing the liver's ability to function properly. Typically, a significant reduction in liver function occurs only when more than 75% of the liver tissue is affected.

These are among the functions of the liver:

- Bile production is required for the digestion of food, especially fats

- Storing extra glucose or sugar as glycogen and converting it back into glucose when the body needs energy

- Production of blood clotting factors

- Production of amino acids (building blocks used to make proteins), including those used to control infection

- Urea conversion for excretion of waste products of metabolism in the urine

- Processing medications into their active forms within the body. Cirrhosis refers to lasting scarring in the liver, where healthy liver cells are substituted by scar tissue that cannot carry out normal liver functions.

Acute liver failure can often be reversible, except sometimes when there is a treatable cause, and the liver can return to normal. The first assessment to understand whether the liver injury is acute or chronic, the underlying etiology, and its relationship with systemic diseases is to take the patient's history and perform a physical examination. Based on the findings, the second step to diagnose liver disorders is urine and blood tests, usually performed non-invasively. Liver enzymes are of great importance in the diagnosis of liver diseases. The main enzymes for measuring liver values are Alanine Aminotransferase (SGPT, ALT) and Aspartate Aminotransferase (SGOT, AST). Habits such as smoking and alcohol consumption trigger liver diseases and play a role in increasing these enzymes (Ersoy 2012; Saygın and Baykara 2021). Many factors can lead to the development of liver diseases. These factors can be categorized as follows.

Infection: The liver can become infected with parasites and viruses, leading to inflammation or edema of the liver and compromising liver function. The virus that usually causes liver damage is transmitted through blood or

semen and is usually spread through contaminated food, dirty water or contact with an infected person. Liver infections such as hepatitis A, C, and B can affect humans (Calleja-Panero et al., 2024; Haeng Hur et al., 2024; Maria Jesi and Antony Asir Daniel 2024; Wei et al., 2024; Wu and Moeckel 2023).

Immune System Abnormality: Certain diseases can disrupt the body's immune system, causing it to mistakenly target other organs, including the liver. Examples of such conditions are autoimmune hepatitis, primary biliary cholangitis, and primary sclerosing cholangitis.

Inheritance: An uncommon gene inherited from both parents may lead to an accumulation of certain compounds in the liver, resulting in liver damage. Genetic liver disorders include conditions like Wilson's disease, hemochromatosis, and alpha-1 antitrypsin deficiency.

Cancer and Other Developments: Types of cancer that cause liver diseases include conditions such as liver adenoma, bile duct cancer, and liver cancer.

Other Common causes include long-term alcohol abuse, fatty deposits in the liver (NAFLD), certain medications or over-the-counter treatments, and certain herbal concoctions.

Risk Factors: Factors that may increase the risk of liver disease include excessive alcohol consumption, being overweight, type 2 diabetes, tattooing, body puncture, injection of drugs with used needles, blood transfusion, contact with foreign blood, unprotected sex, exposure to chemicals, and genetic factors (Priya et al., 2018).

The symptoms of liver disease can be varied, and each patient may experience different symptoms. However, in general, some common symptoms of liver disease may include,

Jaundice: Yellowing of the skin and eyes.

Cholestasis: Blockage of bile flow can lead to symptoms such as itching of the skin.

Liver Size: When the liver is more prominent than usual.

Portal Hypertension: High blood pressure caused by blockage of blood vessels in the liver.

Esophageal Varices: Dilated veins on the esophagus (food pipe) due to portal hypertension can cause bleeding.

Ascites: Fluid accumulation in the abdominal cavity, swelling, and discomfort in the abdomen.

Hepatic Encephalopathy: Changes in brain function due to impaired liver function, which can lead to impaired mental status.

Liver Failure: The liver's inability to perform its functions can lead to a range of health issues. To evaluate these symptoms, a physical examination is conducted initially, followed by relevant imaging and diagnostic tests like liver function tests, liver biopsy, ultrasound, CT scan, or MRI, depending on the patient's needs. These assessments help determine the type and extent of liver disease.

Studies on datasets related to liver diseases have been analyzed. Within the scope of this study, major literature studies focusing on similar topics are discussed, and the results obtained are analyzed in detail. Saygın and Baykara has been used the ILPD dataset containing the data of Indian Liver Patients in their study. MLP, SVM, LR, DT-CART, and LGBM machine learning methods have been used in the study. It has been reported that the highest diagnostic success is 82.12% with the LGBM machine learning method. Close to the LGBM algorithm, DT-CART, with an 81.13% success rate, and MLP are the second highest-performing methods. According to the sensitivity criterion, the highest rate has been obtained from the LR method, at 95.40%. When the results evaluated according to the precision measure, DT-CART is the most successful method, with 95.40%. According to the F-measure, MLP is the best method, with a rate of 86.81%. When the related studies in this field are analyzed, it is stated that the LGBM machine-learning method has not been applied to the ILPD dataset before (Saygın and Baykara 2021).

Tokala has been used a dataset with liver disease in their study. This dataset has been applied to various classification algorithms such as LR, K-NN, and SVM. According to the findings, both LR and KNN techniques can achieve significant accuracy, but LR has a high sensitivity. It can be concluded that LR is a suitable strategy for disease prediction. A new feature model has been proposed, including Random Forest, SVM, J48, Bayesian Net, and MLP classification methods. When the accuracy of the ML method results is compared, it is stated that the Random Forest method provides better accuracy than LR, SVM, and nearest neighbor methods (Tokala et al., 2023).

Yan et al. have been developed a machine learning model to differentiate non-alcoholic fatty liver disease (NAFLD) from non-alcoholic steatohepatitis (NASH) and compared five different machine learning algorithms using clinical data from 259 patients. The XGBoost algorithm is reported to achieve the highest accuracy rate of 95% (Yan et al., 2024).

Chauhan et al. in their study on liver cancer risk estimation has been compared AdaBoost, XGBoost,

Gradient Boosting and Logistic Regression algorithms, stating that Logistic Regression achieved the highest accuracy and emphasized the significance of specific biomarkers like age, alkaline phosphatase, and aspartate aminotransferase in predicting liver cancer risk (Chauhan et al., 2025).

Carnazzo et al. have been used XGBoost and Explainable AI methods in a study on non-invasive staging of chronic liver diseases (CLD). In their study, they have been reported that laminin (LN), collagen-III N-peptide (PIIIP N-P) and collagen-IV (C-IV) levels, which are ECM markers, are important markers in liver fibrosis stage and the XGBoost model has reached an accuracy ratio 87% (Carnazzo et al., 2025).

Qiu et al. have been developed a machine learning model to predict 28-day prognosis in acute-chronic liver failure (ACLF) patients. They have been identified important variables using LASSO regression and have been demonstrated that achieved the highest performance with the XGBoost algorithm. They also stated that the model is more successful than existing clinical scoring systems (Child-Pugh, MELD, CLIF-SOFA, etc.) and could be used in patient follow-up (Qiu et al., 2024).

Panwar used machine learning algorithms such as Logistic Regression, Decision Tree, Support Vector Machine, Naive Bayes, and Random Forest on the (ILPD) dataset. The performance of different classification techniques on different measurement techniques has evaluated, including accuracy, precision, recall, and specificity. It has been reported that SVM, DT, RF, LR, and NB had 74%, 72%, 72%, 72%, 71%, and 57% accuracy, respectively. The results of the analysis indicated that SVM achieved the highest accuracy. Furthermore, study focuses on using clinical data for liver disease prediction and explores different ways of representing such data through this study (Panvar et al., 2021).

3. Materials and Methods

This study is based on the "ILPD (Indian Liver Patient Dataset)" dataset available at the UCI Machine Learning Repository (UCI Machine Learning Repository). A total of 583 patients have been sampled, and it contains clinical and laboratory data of these patients. In the dataset, 416 people are labeled as liver patients, while 167 people are labeled as healthy individuals. The dataset has been split into an 80% training set and a 20% test set, and this split has been made reproducible by setting the random state parameter to 42. For model training for liver disease diagnosis, machine learning is applied to the original dataset with classification algorithms such as Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN), XGBoost and AdaBoost. Feature selection and normalization procedures have been performed to improve model performance. This step has been carried out by considering attributes including age, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase, total protein, albumin, albumin/globulin ratio, and patient information. After feature selection and normalization, the models have been retrained, and their accuracy performance has been improved. All machine learning operations have been performed in the JupyterLab environment using Python programming language.

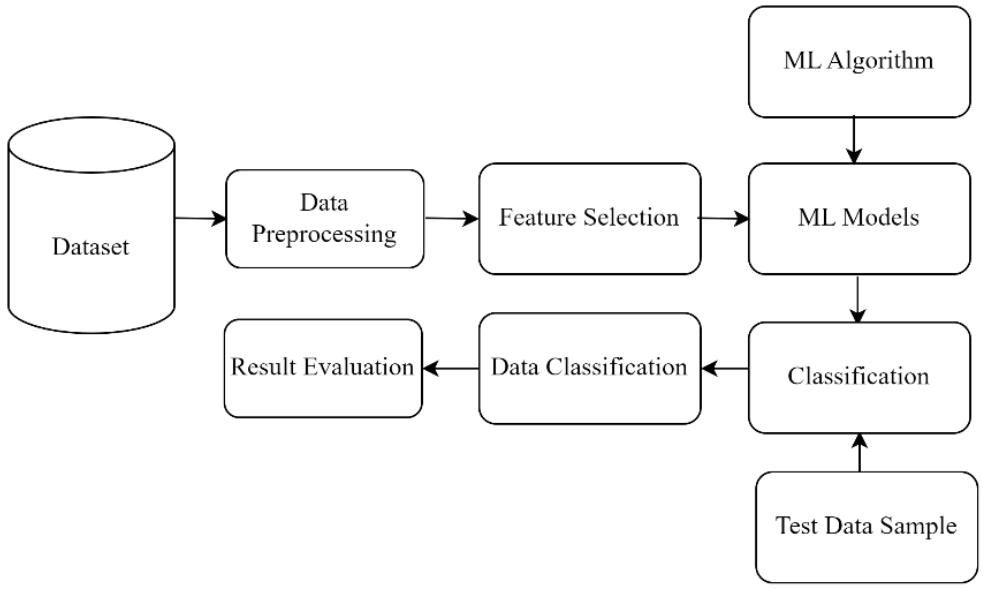


Figure 1. Process Flow Chart

The dataset used has been used in studies on the diagnosis of liver diseases in Indian patients. It is available in the UCI machine learning repository under the name ILPD. A total of 583 patients are sampled in the dataset, 416 with liver disease and 167 healthy patients. These sampled patients consisted of 441 males and 142 females. There are 11 features in total in the dataset. There are 10 of them as features, and 1 of them indicates whether the disease is present or not. Additional information about the dataset is given in Table 1.

It has been determined that there is missing data in the "A/G Ratio" variable because of the activities carried out to understand the data set. In order not to lose the data or disrupt its integrity, the missing data have been filled by averaging the other records. It is presented in Table 2.

The basic statistical characteristics of the data set, its general distribution, central tendency and a prediction for the spread of the distribution are presented in Table 3.

Table 1. Data Set Properties

Features	Data Type
Age	Numeric
Gender	Categorical
TB	Numeric
DB	Numeric
AlkPhos	Numeric
Sgpt	Numeric
Sgot	Numeric
TP	Numeric
Albumin	Numeric
A/G Ratio	Numeric
Is_Patient	Numeric

The distribution of women and men in the dataset is presented in Figure 2, with a male-dominated distribution. According to gender and age distribution, it is observed that the prevalence of liver disease is higher in the 35-60 age range and in men and is presented in Figure 3.

Table 2. Data Set Sample

	age	gender	tb	db	Alkphos	sgpt	Sgot	tp	albumin	a/g_ratio	is_patient
1	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.90	1
2	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
3	62	Male	7.3	4.1	490	60	68	7.0	3.3	0.89	1
4	58	Male	1.0	0.4	182	14	20	6.8	3.4	1.00	1
5	72	Male	3.9	2	195	27	59	7.3	2.4	0.40	1

Table 3. Statistical Distribution

	age	tb	db	Alkphos	sgpt	sgot	tp	albumin	a/g_ratio	is_patient
count	583	583	583	583	583	583	583	583	583	583
mean	44.746	3.298	1.486	290.576	80.713	109.910	6.483	3.141	0.947	0.713
std	16.189	6.209	2.808	242.937	182.620	288.918	1.085	0.795	0.318	0.452
min	4.000	0.400	0.100	63.000	10.000	10.000	2.700	0.900	0.300	0.000
25%	33.000	0.800	0.200	175.500	23.000	25.000	5.800	2.600	0.700	0.000
50%	45.000	1.000	0.300	208.000	35.000	42.000	6.600	3.100	0.947	1.000
75%	58.000	2.600	1.300	298.000	60.500	87.000	7.200	3.800	1.100	1.000
max	90.000	75.00	19.700	2110.000	2000.000	4929.000	9.600	5.500	2.800	1.000

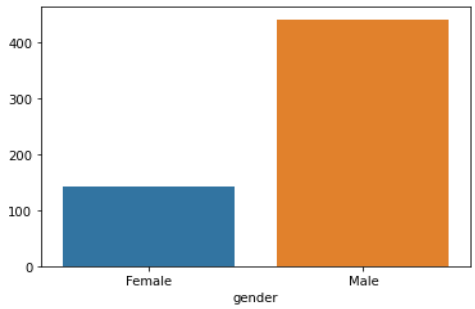


Figure 2. Female Male Ratio

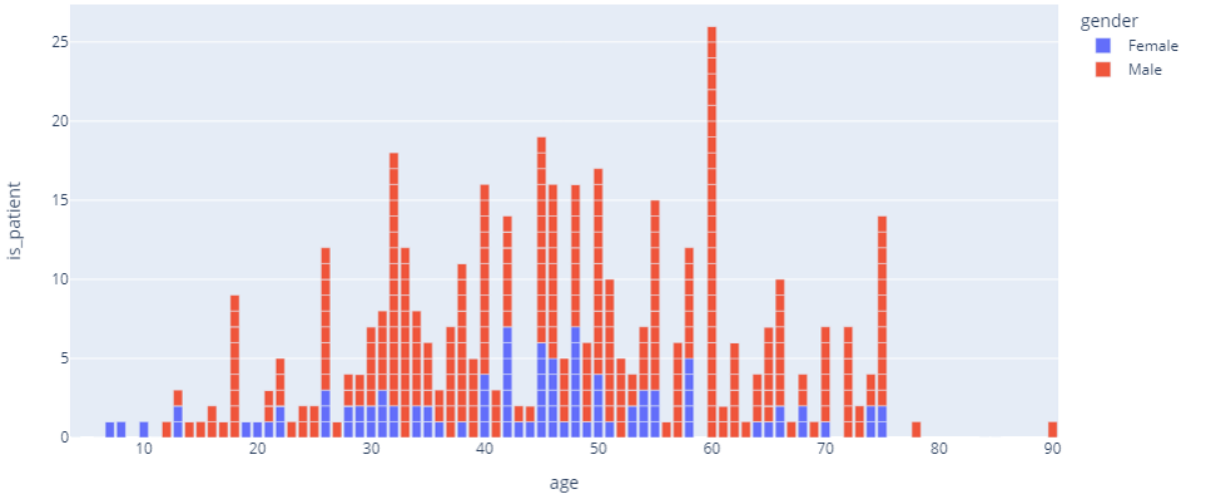


Figure 3. Liver Disease by Gender and Age

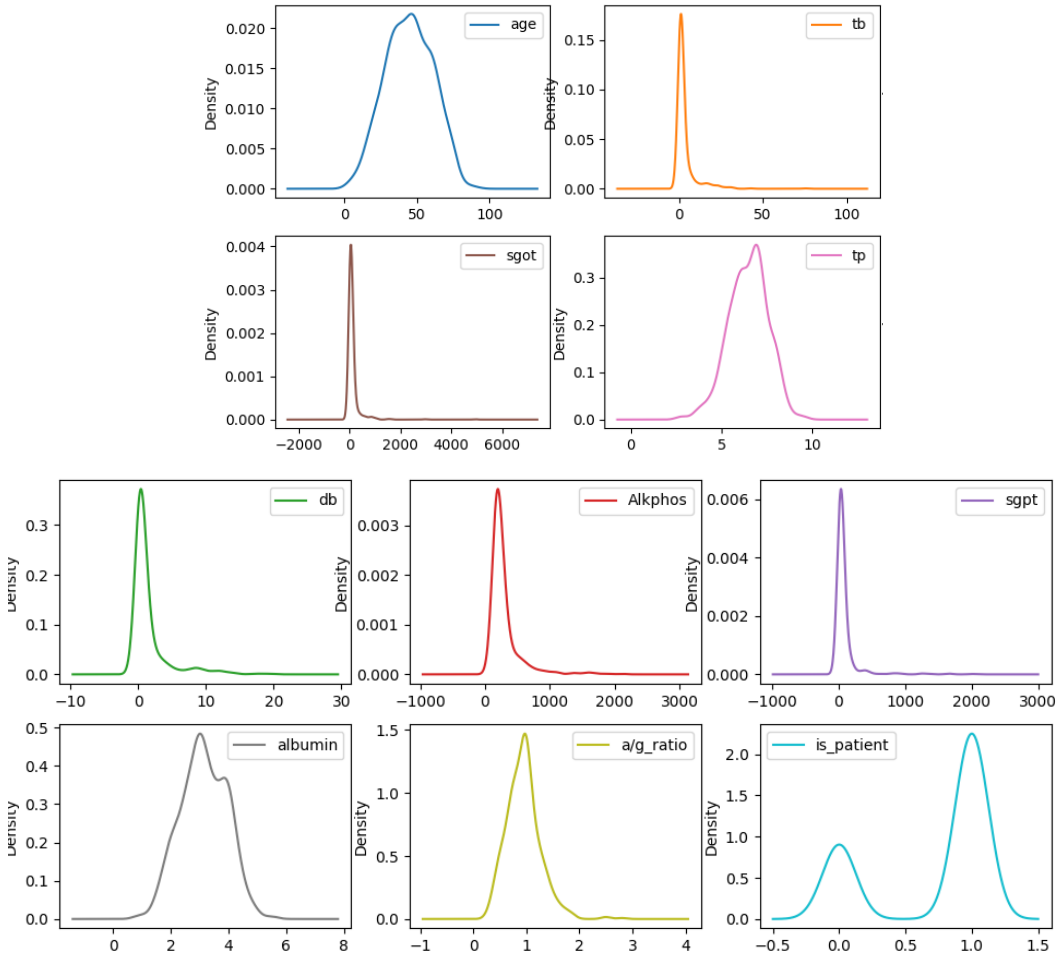


Figure 4. Attributes Distribution

The frequency distribution analysis performed on the obtained data set explains the differences between the attributes of the patients. As a result of this analysis, it has been observed that age (age), total protein (tp), albumin, and albumin/globulin ratio (a/g_ratio) variables have a normal distribution. In contrast, total bilirubin (tb), direct bilirubin (db), alkaline phosphatase (alkphos), SGOT (sgot), and SGPT (sgpt) variables have a right skewed distribution. These findings show that patients exhibited a more homogeneous distribution in terms of age, total protein level, albumin amount, and albumin/globulin ratio. However, there is a significant skewness in total bilirubin, direct bilirubin, alkaline phosphatase, SGOT, and SGPT values related to liver function, which are presented in Figure 4. In the relationship between the correlation matrix and the attributes, there is a strong positive correlation between total bilirubin and direct bilirubin (0.87). This indicates that these two values are closely

linked. The correlation between age and other variables is generally low. The highest negative correlation is seen between age and albumin at -0.27. This may indicate that albumin levels generally decrease with increasing age. The correlation between alkaline phosphatase and SGPT is low. This indicates that these two values are independent of each other. There is a strong positive correlation between SGPT and SGOT (0.79). This indicates that these two values are closely linked. There is a strong positive correlation between total protein and albumin (0.78). This indicates that the total protein level usually parallels the albumin level. There is a strong positive correlation between albumin and a/g_ratio (0.69). This indicates that albumin level is closely related to a/g ratio. It is presented in Figure 5. The significance levels between some attributes have been measured, and the results are presented in Table 3.

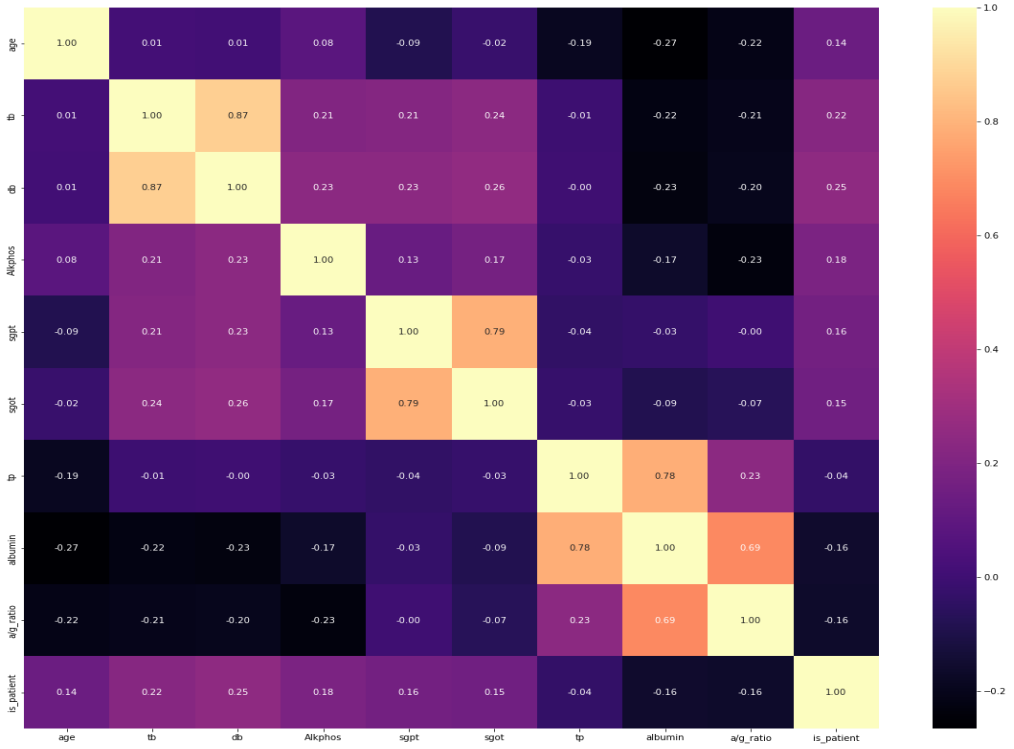


Figure 5. Correlation matrix

Table 3. Significance Levels Between Attributes

	Attribute	Attribute	Correlation	P-Value	Result
0	sgot	sgpt	0.791966	1,35E-120	Significant
1	sgot	tb	0.237831	6,11E-03	Significant
2	sgot	db	0.257544	2,76E-04	Significant
3	sgot	albumin	-0.085290	3,95E+04	Significant
4	sgot	tp	-0.025645	5,37E+05	Not Significant
5	sgot	a/g_ratio	-0.070024	9,12E+04	Not Significant
6	sgpt	tb	0.214065	1,81E-01	Significant
7	sgpt	db	0.233894	1,10E-02	Significant
8	sgpt	albumin	-0.029742	4,74E+05	Not Significant
9	sgpt	tp	-0.042518	3,05E+05	Not Significant
10	sgpt	a/g_ratio	-0.002374	9,54E+05	NotSignificant
11	tb	db	0.874618	7,87E-179	Significant
12	tb	albumin	-0.222250	5,87E-02	Significant
13	tb	tp	-0.008099	8,45E+05	Not Significantt
14	tb	a/g_ratio	-0.206159	5,14E-01	Significant
15	db	albumin	-0.228531	2,40E-02	Significant
16	db	tp	-0.000139	9,97E+05	Not Significantt
17	db	a/g_ratio	-0.200004	1,13E+00	Significant
18	albumin	tp	0.784053	1,93E-116	Significant
19	albumin	a/g_ratio	0.686322	2,17E-76	Significant
20	tp	a/g_ratio	0.233904	1,10E-02	Significant

The combination of Feature Selection and Normalization Filter preprocessing steps is a strategy to optimize the training of the model as Table 4. Correlation-based feature selection helps the model focus on essential features, while normalization corrects for scale differences in the dataset, allowing the model to produce more balanced and consistent results. In this way, the learning accuracy of classification methods can be increased, and more reliable predictions can be made.

Table 4. Feature Selection

Feature Abbreviation	Description
Age	Age
TB	Total Bilirubin
Sgpt	Alanine Aminotransferase
Sgot	Sgot
TP	Total Protein

4. Model Evaluation and Discussion

The presented machine learning models are evaluated with various success metrics. These metrics allow us to analyze model performance in various aspects. In order to evaluate the performance of each model, a confusion matrix has been calculated. This matrix indicates the accuracy of the model in a 2x2 matrix format by comparing the predictions made with the actual values. Each cell indicates whether the model correctly or incorrectly predicted the positive and negative classes. In this way, essential performance measures such as precision, sensitivity, and F1-Score were obtained (Eyüpoğlu, 2020). The confusion matrix helps to analyze the classification ability of each model in detail and to reach reliable results (Chen et al., 2024; Kistenev et al., 2022; Lampignano et al., 2023; Otesteanu et al., 2024, Yavuz and Eyupoglu, 2020).

TP (True Positive): The number of values classified as 1 that are indeed 1.

FP (False Positive): The number of values that are classified as 1 but are actually 0.

TN (True Negative): The number of values classified as 0 that are actually 0.

FN (False Negative): The number of values that are classified as 0 but are actually 1.

Precision: Returns the proportion of correctly classified data.

$$Precision = \frac{TP}{TP + FP}$$

F-measure: Returns the harmonic mean of Precision and Recall values.

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Sensitivity: Returns the ratio of data classified as positive to the actual positive data.

$$Sensitivity = \frac{TP}{TP + FP}$$

Specificity: Returns the ratio of data classified as negative to actual negative data.

$$\text{Specifity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Accuracy: Returns the accuracy value itself and is calculated as follows.

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

The effects of feature extraction and normalization on performance before and after model training have been evaluated. In Table 5, the success rates before feature extraction and normalization are compared with the rates after these operations on the dataset.

Table 5. Effect of Normalization Process on Model Results

ML Models	Accuracy Rates	
	Before Normalization	After Normalization
Logistic Regression-LR	75%	80%
Random Forest - RF	73%	76%
Decision Tree- DT	71%	68%
KNeighbors Classifier - KNN	66%	56%
XGBClassifier - XGB	71%	72%
AdaBoost Classifier-ADABC	70%	75%
Ensemble-BaggingRegressor	71%	72%

Feature extraction and normalization played an essential role in improving the overall accuracy and performance of the model. The success rates obtained in the first case are lower than those observed after feature extraction and normalization. This is because the features in the dataset have not been scaled and significantly modified. Feature extraction allowed the model to focus on more important and salient features, while normalization helped to represent different attributes at similar scales. As a result, these processes allowed the model to generalize better and provide more reliable results in real-world applications. Table 6 shows the comparative model results.

In this study the KNN, DT and RF algorithms have been shown lower performance compared to other models. The accuracy rate of the KNN algorithm is 56% and the main reasons for this low performance are the insufficient

scaling process and data imbalance. The selection of the number of neighbors as 2 is estimated to increase the sensitivity of the model to noise and reduce its generalization ability. The DT algorithm has an accuracy rate of 68%. Possible reasons are that the model overfits the training data and underperforms on the test data due to its tendency to overfit. One of the reasons that the Random Forest algorithm could not reach the expected level with an accuracy of 76% is that the number of trees that are used is considered to be the default values (100) and the feature selection is not optimized. There are recommendations for improvement strategies for future work to increase the performance of these models. Data scaling methods should be applied and class imbalance should be addressed by methods using SMOTE. In order to prevent the DT model from overfitting the maximum depth should be limited and pruning techniques should be used. In the RF model's generalization ability of the model can be improved by increasing the number of trees and more significant variables can be included in the model through feature selection.

Table 6. Model Performance Metrics

ML Models	Performance Measures				
	Accuracy	Precision	F-Score	Sensitivity	Specificity
Logistic Regression-LR	80%	0.796	0.882	0.988	0.266
Random Forest - RF	76%	0.817	0.844	0.873	0.433
Decision Tree- DT	68%	0.777	0.790	0.804	0.333
KNeighborsClassifier- KNN	56%	0.790	0.657	0.563	0.566
XGBClassifier-XGB	72%	0.831	0.811	0.793	0.533
AdaBoost Classifier-ADABC	75%	0.862	0.826	0.793	0.633
Ensemble-BaggingRegressor	72%	0.795	0.822	0.850	0.366

Table 5 shows that the LR model performs the best with an accuracy of 80%, while the precision is 79.6% and the sensitivity is 98.8%. This demonstrates the model's high accuracy in positive predictions and its effectiveness in detecting true positive cases. The RF model stands out with an accuracy of 76% and a precision of 81.7%. This highlights the model's accuracy in positive predictions as well as its ability to recognize true positive cases successfully. The DT model, which has an accuracy rate of 68%, shows a more uneven performance in terms of precision and sensitivity compared to the other models. With an accuracy of 56%, KNN performs less well compared to the others. However, the sensitivity value of 56.3% represents the model's success in detecting positive cases. With 72% accuracy, the XGB model stands out for its high precision (83.1%) and specificity (79.3%). This highlights the model's accuracy in positive predictions and its ability to classify negative cases accurately. With an accuracy of 75%, ADABC also performs strongly on other metrics. High values of precision (86.2%) and specificity (79.3%) strongly support the model's classification ability. In addition, the Ensemble-Bagging Regressor method combines the outputs of multiple models to improve overall performance, achieving 72% accuracy, 79.6% precision and 85% sensitivity. It demonstrates a balanced and robust classification capability that leverages the strengths of ensemble learning to improve model stability.

5. Conclusions

This study comprehensively evaluates the performance of different models through classification experiments on the ILDP dataset. The obtained results contribute to the understanding of the important factors affecting model performance. Future work could include an in-depth investigation of different metrics and parameters to improve model performance further. In this context, research on larger datasets, feature engineering, and more complex model architectures can increase the potential to optimize classification performance. The results of this study inspire future research in the early diagnosis and treatment of ILDP or similar diseases.

Declaration of Ethical Standards

The authors declare that they comply with all ethical standards.

Credit Authorship Contribution Statement

Author-1: Conceptualization, investigation, methodology and software, visualization and writing – original draft.

Author-2: Conceptualization, investigation, methodology and software, supervision and writing – review and editing.

Declaration of Competing Interest

The authors have no conflicts of interest to declare regarding the content of this article.

Data Availability Statement

All data generated or analyzed during this study are included in this published article.

6. References

Calleja-Panero, J. L., Esteban Mur, R., Jarque, I., Romero-Gómez, M., Group, S. R., García Labrador, L., & González Calvo, J. (2024). Chronic liver disease-associated severe thrombocytopenia in Spain: Results from a retrospective study using machine learning and natural language processing. Gastroenterologia y Hepatologia, 47(3), 236–245. https://doi.org/10.1016/j.gastrohep.2023.05.010

Carnazzo, V., Pignalosa, S., Tagliaferro, M., Gragnani, L., Zignego, A. L., Racco, C., ... & Basile, U. (2025). Exploratory study of extracellular matrix biomarkers for non-invasive liver fibrosis staging: A machine learning approach with XGBoost and explainable AI. Clinical Biochemistry, 135, 110861.

Chauhan, R., Sahi, A., Kaur, H., & Avasthi, S. (2025). Machine learning algorithm approach in risk prediction of liver cancer. In Machine Learning Models and Architectures for Biomedical Signal Processing, Academic Press, 547-570.

Chen, W., Zhang, Y. J., Wu, W., Yang, H., & Huang, W. (2024). Machine learning-based predictive model for abdominal diseases using physical examination datasets. Computers in Biology and Medicine, 173, 108249. https://doi.org/10.1016/j.compbiomed.2024.108249

Cohen, J. C., Horton, J. D., & Hobbs, H. H. (2011). Human fatty liver disease: old questions and new insights. Science, 332(6037), 1519–1523.

Ersoy, O. (2012). Karaciğer enzim yüksekliğinin değerlendirilmesi. Ankara Medical Journal, 12(3), 129–135

Eyupoglu, C., & Karakuş, O. (2024). Novel CAD diagnosis method based on search, PCA, and AdaBoostM1 techniques. Journal of Clinical Medicine, 13(10), 2868.

Eyüpoğlu, C. (2020). Kronik böbrek hastalığının erken tanısı için yeni bir klinik karar destek sistemi. Avrupa Bilim ve Teknoloji Dergisi, (20), 448-455.

Guo, W., Shi, Z., Zeng, T., He, Y., Cai, Z., & Zhang, J. (2022). Metabolic study of aristolochic acid I-exposed mice liver by atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry imaging and machine learning. Talanta, 241, 123261. https://doi.org/10.1016/j.talanta.2022.123261

Haeng Hur, M., Cheuk-Fung Yip, T., Up Kim, S., Woong Lee, H., Ah Lee, H., Lee, H.-C., Lai-Hung Wong, G., Wai-Sun Wong, V., Yong Park, J., Hoon Ahn, S., Kyung Kim, B., Young Kim, H., Seok Seo, Y., Shin, H., Park, J., Ko, Y., Park, Y., Bin Lee, Y., Jong Yu, S., … Lee, J.-H. (2025). Journal Pre-proof A machine learning model to predict liver-related outcomes after the functional cure of chronic hepatitis B including those for text and data mining, AI training, and similar technologies. Journal of Hepatology. 82, 2, 235-244 https://doi.org/10.1016/j.jhep.2024.08.016

Hashem, E. M., & Mabrouk, M. S. (2014). A study of support vector machine algorithm for liver disease diagnosis. American Journal of Intelligent Systems, 4(1), 9–14.

Kistenev, Y. V., Vrazhnov, D. A., Shnaider, E. E., & Zuhayri, H. (2022). Predictive models for COVID-19 detection using routine blood tests and machine learning. Heliyon, 8, 10, e11185 https://doi.org/10.1016/j.heliyon.2022.e11185

Lampignano, L., Tatoli, R., Donghia, R., Bortone, I., Castellana, F., Zupo, R., Lozupone, M., Panza, F., Conte, C., & Sardone, R. (2023). Nutritional patterns as machine learning predictors of liver health in a population of elderly subjects. Nutrition, Metabolism and Cardiovascular Diseases, 33(11), 2233–2241. https://doi.org/10.1016/j.numecd.2023.07.009

Maria Jesi, P., & Antony Asir Daniel, V. (2024). Differential CNN and KELM integration for accurate liver cancer detection. Biomedical Signal Processing and Control, 95, 106419. https://doi.org/10.1016/j.bspc.2024.106419

Otesteanu, C. F., Caldelari, R., Heussler, V., & Sznitman, R. (2024). Machine learning for predicting Plasmodium liver stage development in vitro using microscopy imaging. Computational and Structural Biotechnology Journal, 24, 334–342. https://doi.org/10.1016/j.csbj.2024.04.029

Pan, A., Mukhopadhyay, S., & Samanta, S. (2022). Liver disease detection: evaluation of machine learning algorithms performances with optimal thresholds. International Journal of Healthcare Information Systems and Informatics (IJHISI), 17(2), 1–19.

Priya, M. B., Juliet, P. L., & Tamilselvi, P. R. (2018). Performance analysis of liver disease prediction using machine learning algorithms. Int. Res. J. Eng. Technol, 5(1), 206–211.

Qiu, S., Zhao, Y., Hu, J., Zhang, Q., Wang, L., Chen, R., ... & Han, T. (2024). Predicting the 28-day prognosis of acute-on-chronic liver failure patients based on machine learning. Digestive and Liver Disease, 56(12), 2095-2102.

Saygın, E., & Baykara, M. (2021). Karaciğer yetmezliği teşhisinde özellik seçimi kullanarak makine öğrenmesi yöntemlerinin başarılarının ölçülmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 33(2), 367–377.

Tokala, S., Hajarathaiah, K., Gunda, S. R. P., Botla, S., Nalluri, L., Nagamanohar, P., Anamalamudi, S., & Enduri, M. K. (2023). Liver disease prediction and classification using machine learning techniques. International Journal of Advanced Computer Science and Applications, 14, 2. https://dx.doi.org/10.14569/IJACSA.2023.0140299

Wei, H., Luo, S., Bi, Y., Liao, C., Lian, Y., Zhang, J., & Huang, Y. (2024). Plasma microRNA-15a/16-1-based machine learning for early detection of hepatitis B virus-related hepatocellular carcinoma. Liver Research, 8(2), 105–117. https://doi.org/10.1016/j.livres.2024.05.003

Wu, B., & Moeckel, G. (2023). Application of digital pathology and machine learning in the liver, kidney and lung diseases. Journal of Pathology Informatics, 14, 100184. https://doi.org/10.1016/j.jpi.2022.100184

Yan, Y., Chen, Q., Xiang, Z., Wang, Q., Long, Z., Liang, H., Ameer, S., Zou, J., Dai, X., & Zhu, Z. (2024). Amino acid metabolomics and machine learning-driven assessment of future liver remnant growth after hepatectomy in livers of various backgrounds. Journal of Pharmaceutical and Biomedical Analysis, 249, 116369. https://doi.org/10.1016/j.jpba.2024.116369

Yan, Y., Gan, D., Zhang, P., Zou, H., & Li, M. (2024). A machine learning-based predictive model discriminates nonalcoholic steatohepatitis from nonalcoholic fatty liver disease. Heliyon, 10(21), e38848. https://doi.org/10.1016/j.heliyon.2024.e38848

Yavuz, E., & Eyupoglu, C. (2020). An effective approach for breast cancer diagnosis based on routine blood analysis features. Medical & Biological Engineering & Computing, 58(7), 1583-1601.

Yu, R., Huang, Y., Hu, X., & Chen, J. (2024). Analysis of machine learning based integration to identify the crosslink between inflammation and immune response in non-alcoholic fatty liver disease through bioinformatic analysis. Heliyon, 10(14), e32783. https://doi.org/10.1016/j.heliyon.2024.e32783

Internet References

1-UCI Machine Learning Repository. Retrieved September 18, 2024, from https://archive.ics.uci.edu/