Volume 13, Issue 4, Page 153-161, 2024 https://doi.org/10.46810/tdfd.1553699 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Ensemble and Non-Ensemble Machine Learning-Based Classification of Liver Cirrhosis Stages

Zeinab Mahdi MOUMIN[1], İrem Nur ECEMİŞ[2*], Mustafa KARHAN[3]

[1]Çankırı Karatekin University, Institute of Science, Department of Electronics and Computer Science, Çankırı, Türkiye
[2*]Çankırı Karatekin University, Faculty of Engineering, Department of Computer Engineering, Çankırı, Türkiye
[3]Çankırı Karatekin University, Institute of Science, Department of Electronics and Computer Science, Çankırı, Türkiye

Zeinab Mahdi MOUMIN ORCID No: 0009-0003-8889-9160
İrem Nur ECEMİŞ ORCID No: 0000-0001-9535-2209
Mustafa KARHAN ORCID No: 0000-0001-6747-8971

**Corresponding author: iremnurecemis@karatekin.edu.tr*

(Received: 20.09.2024, Accepted: 02.12.2024, Online Publication: 30.12.2024)

Keywords
Liver cirrhosis,
Artificial intelligence,
Mutual information,
Soft voting,
K-fold cross-validation

Abstract: Cirrhosis is a chronic liver condition characterized by gradual scarring of the tissue in the liver, which then leads to one of the more serious health problems. Early diagnosis and detection of this condition are critical to managing the patient's situation and planning his treatment. Machine learning is a computer science field in which many complex issues have otherwise been successfully resolved, especially in medicine. This work focuses on constructing an artificial intelligence system, assisted by machine learning algorithms, to help professionals diagnose liver cirrhosis at its early stage. In this paper, four different models have been constructed with the aid of clinical parameters of patients and machine learning techniques: Random Forest, KNN, histogram-based Gradient Boosting, and Soft Voting. Two feature selection methods (Chi-Square and mutual information) have been combined to select the most relevant features in the dataset. Then non-ensemble and ensemble methods are applied to detect the liver disease. The random forest model achieved the highest score among other model with 97.4 % accuracy with a 10-fold cross-validation method.

Topluluk ve Topluluk Olmayan Makine Öğrenmesine Dayalı Karaciğer Sirozu Evrelerinin Sınıflandırılması

Anahtar Kelimeler
Karaciğer sirozu,
Yapay zeka,
Karşılıklı bilgi,
Yumuşak oylama,
K-katlı çapraz doğrulama

Öz: Siroz, karaciğerdeki dokunun kademeli olarak yaralanmasıyla karakterize kronik bir karaciğer rahatsızlığıdır. Bu rahatsızlık ilerleyen dönemde daha ciddi sağlık sorunlarına yol açar. Bu rahatsızlığın erken teşhisi ve tespiti, hastanın durumunu yönetmek ve tedavisini planlamak için kritik öneme sahiptir. Makine öğrenimi, özellikle tıpta birçok karmaşık sorunun başarıyla çözüldüğü bir bilgisayar bilimi alanıdır. Bu çalışma, profesyonellerin karaciğer sirozunu erken aşamada teşhis etmelerine yardımcı olmak için makine öğrenimi algoritmalarıyla desteklenen bir yapay zeka sistemi oluşturmaya odaklanmaktadır. Bu makalede, hastaların klinik parametreleri ve makine öğrenimi tekniklerinin yardımıyla dört farklı model oluşturulmuştur: Rastgele Orman, KNN, Histogram Tabanlı Gradyan Artırma ve Yumuşak Oylama. Veri kümesindeki en alakalı özellikleri seçmek için iki özellik seçme yöntemi (Chi-square ve karşılıklı bilgi) birleştirilmiştir. Ardından karaciğer hastalığını tespit etmek için topluluk dışı ve topluluk yöntemleri kullanılmıştır. Rastgele orman modeli, 10 katlı çapraz doğrulama yöntemi ile %97,4 doğrulukla diğer modeller arasında en yüksek puanı elde etmiştir.

1. INTRODUCTION

The body's largest organ is the liver, which is important in digestion and detoxification. However, factors such as viral infections and alcohol consumption can damage the liver, leading to life-threatening conditions [1]. Liver conditions, including hepatitis, cirrhosis, liver tumors, and cancer, are significant contributors to mortality worldwide. Cirrhosis alone is responsible for hundreds of thousands of death annually. There are approximately 71 million people worldwide who have liver cirrhosis and other chronic liver diseases, and it resulted in an estimated 1.3 million deaths in 2021 [2]. The World Health Organization has considered liver cirrhosis to be one of the major concerns to global health, adding about 3–4 million new cases yearly. The highest frequency is in the developing countries in Asia and Africa, in contrast to Western nations in Europe and North America [3]. Symptoms of liver cirrhosis frequently appear only in the later stages of the disease. Most infected patients, however are asymptomatic at the initial stages thus resulting in a more significant damage of the liver and rise in mortality rates [4]. The vaccination is not an option for individuals with severely damaged livers. Thus, establishing the extent of damage to the liver is very necessary to ensure that doctors can quickly identify and provide treatment for chronic infections. Full treatment ensures that the illness does not pass from one individual to another [5]. Machine learning, as a subset of artificial intelligence, has been an emerging avenue that potentially diagnoses and classifies cirrhosis. Based on large datasets and algorithms, machine learning models could analyze a variety of factors and patterns in a patient to predict the likeliness of cirrhosis [6]. Machine learning may have the potential to improve diagnosis for diseases that have made an interest in the biomedical field while bringing down diagnostic costs simultaneously [7].

Several machine learning algorithms have been executed for identifying liver conditions. Meng et al. [8] applied a dataset of ROI ultrasound images consisting of 79 images of healthy liver ROIs, 89 images of early-stage liver fibrosis ROIs, and 111 images of late-stage liver fibrosis ROIs. They pose a liver fibrosis classification method employing transfer learning (TL) with VGGNet, and a deep classifier called FCNet. Huang et al. [9], discover how features such as gender and weight could impact the commonness of liver cirrhosis in different populations. The study states that all of the above-mentioned criteria should be taken into consideration while developing any manual or AI-driven system to get the desired results for the most appropriate treatment solutions from health experts. In simple research Cheng et al. [10], the contribution of attributes by patients was considerably different in terms of sex, body mass index, bilirubin, alanine aminotransferase, and so on. The study has shown that the mean BMI value for male patients over 60 is lower than for female patients under 60. Higher BMI values are also linked to an increased risk. Using Regression Logistic, and RF, Bedeir and Hadi [11] developed a machine learning model to predict cirrhosis liver. Every model was assessed based on performance measures such as accuracy and test error. The results obtained from employing the feature selection method were compared with those obtained without feature selection. The results highlighted that the Random Forest model gives the best accuracy at 96.59 %. The authors suggest future studies to measure the performance of the proposed framework in different patient populations and healthcare settings. The data set used in [12] is the northeast Andhra Pradesh, India dataset obtained from the UCI Machine Learning Repository In this study, they employed six concepts of machine learning like Logistic Regression, KNN, Decision Tree, Support Vector Naive Bayes and Random Forest. In another paper, ensemble methods that give high results on different data sets were examined [13]. Additional metrics evaluated the performance of models based on different performance measures like accuracy, precision, recall and F1-score. Topcu et al. [14] employed Logistic Regression (LR), K-Nearest Neighbors (KNN), Random Forest (RF), AdaBoost, and Bernoulli Naive Bayes (BernoulliNB). The Random Forest algorithm achieved the highest score, approximately 98%. The paper employed a "Cirrhosis Patient Survival Prediction" dataset from the UCI. This dataset contained 418 patients with liver conditions and 17 clinical features to predict the state of patients with liver cirrhosis. In another study, Zhang et al. [15] built a machine learning-based model to identify and predict the different fibrosis stages. They employed clinical data from 618 chronic hepatitis patients treated at Zhejiang Provincial People's Hospital between February 2017 and September 2021. Six different learning algorithms (Logistic Regression, Support vector machine, Bayes, K-Nearest Neighbor, Decision Tree (DT), and Random Forest) were used to build the predictive model. The researchers identified the most relevant features from the data with maximum relevance, minimum redundancy (mRMR), and gradient boosting decision tree (GBDT) methods. The Decision Tre model showed strong performance in detecting liver fibrosis stages with high Area under the Curve (AUC) values in both the training cohort (0.898 to 0.944) and the external validation cohort (0.876 to 0.933). Choi and Oh [16] built a machine learning-based approach to detect and predict liver cirrhosis. They employed a dataset of 6980 patients treated between January 2021 and December 2018. Machine learning (Gaussian Naive Bayes, Extreme Gradient Boosting (XGBoost), Random Forest, and Least Absolute Shrinkage and Selection Operator Regression) was applied to identify significant risk factors for cirrhosis. The XGBoost model performed best, with AUC values of 0.832 in the training set and 0.829 in external validation, proving the most effective for cirrhosis prediction. In another paper, Hirano et al. [17] used image data from 75 patients who had undergone liver biopsy and contacted CT scans. The model's performance was evaluated using recall, accuracy, and specificity metrics. They utilized logistic regression with L2 norm regularization to combine multiple texture features from CT images into single combined features. The combined feature showed the highest performance with an accuracy of 76%.

Devikanniga et al. [18] assessed models' performance through precision, recall, and accuracy. They employed an optimized support vector to predict liver disease

accurately. The support vector was optimized using the crow search Algorithm (CSA). The optimized support vector model outperformed traditional SVM and State-of-the-art classification models. The SVM showed an accuracy of 99.49 %. In another paper, Md et al. [19] leverage ensemble learning algorithms to predict liver disease. The dataset in this study is the Indian liver patient dataset. The most relevant features were selected through univariate selection, feature importance, and correlation matrix. Among the six ensemble methods (Gradient Boosting, XGBoost, Bagging, Random Forest, Extra Trees, Stacking), the Extra tree method achieved the highest accuracy score of 91.82%.

In this paper, the proposed approach is using Random Forest and KNN, Histogram-Based Gradient Boosting, Soft Voting to classify the patients with different stages of cirrhosis. The performance of these algorithms was compared in terms of accuracy, precision, recall and F1-score. The remainder of the paper is structured as follows: Section 1 reviews related work in machine learning for cirrhosis classification and prediction, while Section 2 presents our approach. The findings are presented in Section 3 and a few avenues for future work follow the conclusion.

2. MATERIAL AND METHOD

The proposed method comprises different sections that work simultaneously to achieve the study objective. To start, the dataset is collected then the preprocessing step occurs. The datasets are split into tests and training sets following the preprocessing step. The classification algorithms like Random Forest, KNN, and histogram-based Gradient Boosting and Soft Voting are trained and tested using the K-fold cross-validation method on the Cirrhosis Patient Survival Prediction dataset. And liver disease dataset. Figure 1 displays the block diagram of the proposed method, beginning with the initial step of sourcing the datasets from the UCI repository.

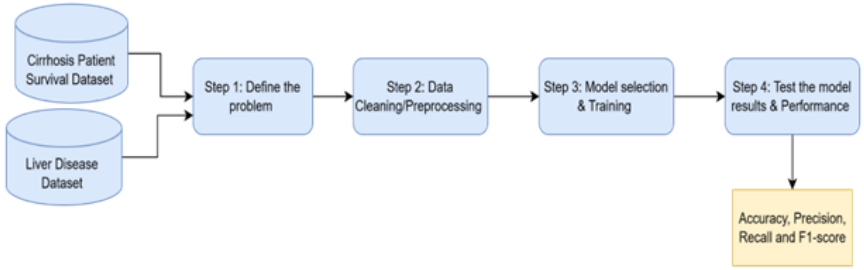


Figure 1. Approach for liver cirrhosis detection

2.1. Datasets

The paper employed a "Cirrhosis Patient Survival Prediction" dataset acquired from the UCI repository. The original data originated from a Mayo Clinic study on primary biliary cirrhosis (PBC) of the liver, conducted between the years 1974 and 1984. This dataset contained 418 patients with liver conditions and 20 clinical features to predict the state of patients with liver cirrhosis. Among these, 10.53% are male patients, and 89.47% are female patients. Table 1 contains a brief description of the attributes and their data types.

Table 1. Cirrhosis dataset attribute description (Dataset 1) [20]

S.No.	Attribute	Description	Data type
1	ID	Unique identifier	Integer
2	N-Days	Number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986	Integer
3	Status	Status of the patient C (censored), CL (censored due to liver tx), or D (death)	Categorical
4	Drug	Type of drug D-penicillamine or placebo	Categorical
5	Age	Age	Integer
6	Sex	M (male) or F (female)	Categorical
7	Ascites	Presence of ascites N (No) or Y (Yes)	Categorical
8	Hepatomegaly	Presence of hepatomegaly N (No) or Y (Yes)	Categorical
9	Spiders	Presence of spiders N (No) or Y (Yes)	Categorical
10	Edema	N (no edema and no diuretic therapy for edema), S (edema present without diuretics), or Y (edema despite diuretic therapy)	Categorical
11	Bilirubin	Serum bilirubin	Continuous
12	Cholesterol	Serum cholesterol	Integer
13	Albumin	Albumin	Continuous
14	Copper	Urine copper	Integer
15	Alk_Phos	Alkaline phosphatase	Continuous
16	SGOT	SGOT	Continuous
17	Tryglicerides	Tryglicerides	Integer
18	Platelets	Platelets per cubic	Integer
19	Prothrombin	Prothrombin time	Continuous
20	Stage	Histologic stage of disease (1, 2, 3, or 4)	Categorical

Another dataset was used to expand the research. The second dataset was collected from a hospital in the Northeast of Andhra Pradesh, India. The Data contains 584 patients' records; 416 patients were diagnosed with liver disease, and 167 patients were healthy individuals, highlighting an imbalance between the two classes. The dataset's attributes comprise aspects related to patient demographics and different biomarkers affiliated with liver disease. In Table 2, there is a brief description of all features of the liver disease dataset.

Table 2. Liver disease dataset attribute description (Dataset 2) [21]

S.No.	Attribute	Description	Data type
1	Age	Age	Integer
2	Gender	Gender of the patient	Binary
3	TB	Total Bilirubin	Continuous
4	DB	Direct Bilirubin	Continuous
5	Alkphos	Alkaline Phosphotase	Integer
6	Sgpt	Alamine Aminotransferase	Integer
7	Sgot	Aspartate Aminotransferase	Integer
8	TP	Total Proteins	Continuous
9	ALB	Albumin	Continuous
10	A/G Ratio	Albumin and Globulin Ratio	Continuous
11	Selector	It is the target variable, 1 means patient is suffering from liver disease, and 0 means the patient is heathy.	Binary

2.1.1. Dataset pre-processing

Data pre-processing is a fundamental process in the field of data analysis. It entails preparing raw data, which includes cleaning and organizing, to render it suitable for machine learning models. Depending on the dataset, different pre-processing techniques can be used [22]. Data cleaning is the step that deals with the handling of missing and noisy data to achieve data consistency. For the Cirrhosis Patient Survival Prediction dataset, the missing values have been handled using two methods. Median imputation replaces missing values in numeric columns with the median value of that column. The mode imputation method replaces the missing values of nonnumeric columns with the most frequent value in the respective column [23]. Traditional machine learning algorithms are driven by mathematical models and hence require numeric computations and statistical operations. Since these algorithms are searching for meaningful patterns or relationships in the data, it becomes a necessity that the data has to be in a numerical representation format. Categorical variables were converted by a label encoding method. Scaling is one of the fundamental stages of preprocessing in machine learning workflows aimed at improving performance. The attributes will be scaled with the StandardScaler technique. In the dataset, the target value Stage is imbalanced. The imbalance of such a target variable was stabilized before the methods of classification were applied, followed by the application of SMOTE-Tomek Link [24]. This method is a technique for handling imbalanced datasets. The SMOTE method generates synthetic samples for the minority class rather than simply duplicating existing minority samples. This approach is employed because it eliminates the risk of overfitting when the same data points are duplicated.

Here is a brief description of how the SMOTE method works:

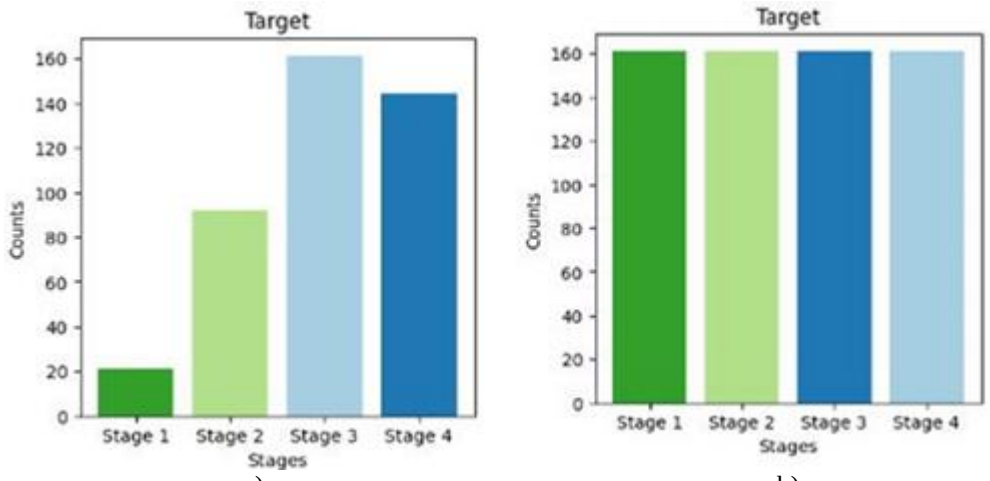
- For every instance in the minority class, SMOTE determined its K-Nearest Neighbors within the same class. Then, it chooses one of the nearest neighbors at random.
- A new synthetic sample is created by interpolating between the original sample and the selected neighbor. The formula to create a synthetic sample is below:

$$\text{synthetic sample} = \text{sample} + \text{gap}(\text{neighbor} - \text{sample})$$

Where gap is a random value between 0 and 1 this process repeats until the minority class has the same number of observations as the majority class, balancing the overall class distribution.

The synthetic sample allows the models to learn observations of minority class patterns, enhancing classification results across all classes and, foremost more generalized performance.

Figure 2 presents the bar chart of the outcome variable before and after SMOTE.



a) b)

Figure 2. (a) Bar plot of outcome variable before SMOTE, (b) Bar plot of outcome variable after SMOTE.

2.2. Feature Selection

Feature Selection methods are techniques employed to reduce the features in the dataset to improve the performance machine learning models. Therefore the less relevant and redundant features are removed from the dataset Over the past few decades, several algorithms have been elaborated to pinpoint the most relevant attributes. In this study, two methods have been combined to select the most significant features.

2.2.1. Chi-square test

Chi-Square test is a statistical method that does not form hypotheses regarding the distribution pattern of the dataset. Chi-Square method measures the obtained results experimentally with expected results theoretically. It analyzes the relationship between the target and the categorical variables. Chi-square tests use the same formula to calculate the statistic. The formula is shown in Equation 1. Chi-square (X^2):

$$X^2 = \sum \frac{(O-E)^2}{E} \quad (1)$$

Where:

- X^2 is the chi-square test statistic
- Σ is the summation operator
- O is the obtained frequency
- *E* is the expected frequency

2.2.2. Mutual information

Mutual Information is a statistical measure employed to scale the information obtained about the target variable through random variables. Mutual information between two variables, X and Y, spelled as I(X;Y), describe how much knowing the value of X dismisses ambiguity about Y. If X and Y are independent, I(X;Y)=0 means X contributes any information about Y. If X and Y are entirely dependent, then I(X;Y) is maximized. The formula is shown in Equation 2. Mutual Information:

$$I(X;Y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (2)$$

Where:

p(x,y) = the joint probability distribution of X and Y

p(x) and p(y) = the marginal distribution of X and Y

Higher Mutual information score indicates that the random variable has useful information about the target variable.

Unlike other feature selection methods, mutual information does not assume a specific distribution and selects features with no linear relevance to the target. The features chosen for this statistical measure are Platelets, Spiders, Trylicerides, Bilerubin, SGOT, Sex, Ascites, Chotesterol, Drug, Copper, Alk-Phos, Status, N-Days and Age.

2.3. Brief Description of Machine Learning Techniques

2.3.1. Random forest

Random Forest is a supervised machine learning method for regression and classification problems. The Random Forest model combines the output of several decision trees to attain one result. Every tree in the ensemble is constructed of a data slice drawn from a training set with replacement [25]. The performance it gives is better than that of other models. This classifier is able to manage large datasets.

Here is a brief description of how the Random Forest technique works:

- Preparation Data: The dataset is split into train and test sets. Random Forest can handle categorical and numerical datasets.
- Random Sampling and bootstrapping: Random forest builds every tree on various random samples of the training dataset. Each sample is drawn with

bootstrapping, meaning that some samples will repeat while others will not.

- Growing each decision tree: Random forest selects a random sample of feature sets at every split. Random feature selection enhances generalization. After selecting a subset, the model determines the best split by evaluating different thresholds. Every tree grows until it grows fully.
- Aggregating Predictions: After all trees grow to the maximum depth, prediction are made individually for each input sample. The class with the most votes from the individual trees is chosen as the final output for classification tasks.

Random Forest model sets the default number of trees at 100 and the gini criterion for splitting nodes. These parameters allow the trees to grow with no limit on the depth of the trees.

2.3.2. K-Nearest neighbor

K-Nearest Neighbor algorithm is one of the simplest supervised learning machines. The KNN technique is employed for the classification and regression tasks. The KNN algorithm registered all the available data and groups to make a new data point based on similarity.

Here is a brief description of how the K-Nearest Neighbors (KNN) technique functions [26]:

- Data Representation and Initialization: Every data point is illustrated in multi-dimensional space. Each dimension corresponds to the feature of dataset. KNN technique does not learn from the training phase as the other algorithms. Instead, it memorizes the training dataset.
- Distance Calculation: KNN calculates the distance between a new data point and other points to predict the class of the data point. Common distance metrics exclude Euclidean, Manhattan and Murkowski distances.
- Finding Neighbors: Following the distance calculation step, KNN selects the nearest data points to the new data points.
- Majority Voting: To classify the new data point, the class that occurs the most among the neighbors is assigned to it.

The KNN model is set with the default Euclidean distance to extend the proximity between data points. By default, the number of neighbors was set to 5, and the leaf size was assessed to 30.

2.3.3. Histogram-based gradient boosting

HistGradientBoosting short for Histogram-Based Gradient Boosting is powerful supervised machine learning method designed for both regression and classification tasks. This algorithm is an ensemble machine learning that creates an accurate predictive model by sequentially combining several weaker models, typically decision trees. It is another gradient boosting algorithm, that utilizes histogram-based techniques to speed up the training phase and improve memory efficiency. In the Histogram-Based Gradient Boosting classifier, the default learning rate was set to 0.1, and the default number of boosting iterations to 100. The default value for the minimum number of samples required in a leaf node is 20.

2.3.4. Soft voting

Voting is an ensemble machine learning technique that merges predictions from several models to enhance model performance beyond what a single model could achieve alone. This ensemble method helps individual model weaknesses and enhances overall performance. There are different types of ensemble techniques. Soft voting, or weighted voting, considers the probability scores assigned by each base model for each class. It calculates the weighted average of these probabilities to make the final prediction. The voting classifier employs a soft voting by default; each classifier is included in the voting classifier with no weighting. Figure 3 illustrates the soft voting.

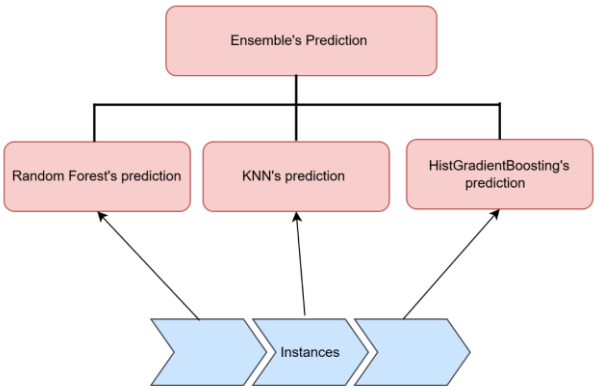


Figure 3. Working process of soft voting

2.4. Performance Measurement Metric

Random Forest, histogram-based Gradient Boosting, KNN and Soft Voting algorithms are used in this study. Experiments are performed using hold out method. Accuracy, F-Measure, Recall and Precision measures are used for the classification of observations. Confusion matrix is a tabular representation of prediction outcomes of any binary. The form of the confusion matrix is given in Figure 4.

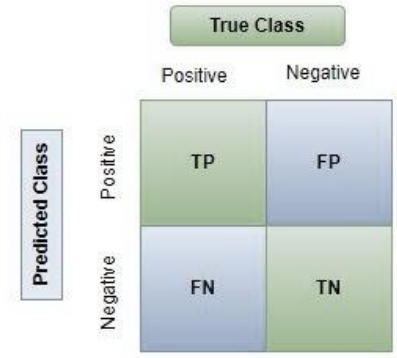


Figure 4. Confusion matrix elements

- TP refers to True Positive. True, it can be understood as the model expected to show positive class.
- False Positive is FP. Though it is False, it can be understood as the model projected positive class.
- False Negative refers to FN. Though it is False, one can understand it as the model expected negative class.
- TN, or True Negative, is It is True and can be seen as the model expected negative class.

The accuracy metric refers as the number of correct predictions to the total number of predictions. The accuracy formula is shown in Equation 3.

$$Accuracy = \frac{\sum(TP + TN)}{\sum(TP + FN + TN + FP)} \quad (3)$$

The precision metric is used to overcome the limitation of accuracy. The precision determines the proportion of positive prediction that was actually correct and is shown in Equation 4.

$$Precision = \frac{\sum TP}{\sum(TP + FP)} \quad (4)$$

The Recall metric calculate the proportion of actual positive that was identified incorrectly. The metric formula is given in Equation 5.

$$Recall = \frac{\sum TP}{\sum(TP + FN)} \quad (5)$$

3. RESULTS AND DISCUSSION

Table 3 presents the results of different machine models with cross-validation method. Different performance metrics were used to assess the performance of four machine learning classifiers on a cirrhosis dataset. The outcomes reveal that Random Forest classifier exhibit the highest accuracy score (96.6%), recall (86.7%), precision (97.7%), and F1-score (86.7%). This method is succeeded by the soft voting and hist gradient boosting methods, which achieve a commendable 95.2% and 94.7% accuracy score respectively. KNN has the lowest accuracy score (76.6%).

Table 3. Performance of classification algorithms on various measures (5-fold cross-validation)

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.966	0.967	0.842	0.867
KNN	0.890	0.920	0.836	0.864
HistGradientBoosting	0.947	0.792	0.772	0.767
Soft Voting	0.952	0.790	0.777	0.769

Table 4. Performance of classification algorithms on various measures (10-fold cross-validation) with Dataset 1

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.974	0.878	0.875	0.900
KNN	0.895	0.879	0.839	0.850
HistGradientBoosting	0.940	0.742	0.754	0.745
Soft Voting	0.959	0.811	0.796	0.791

Table 5. Performance of classification algorithms on various measures (10-fold cross-validation) with Dataset 2

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.8092	0.8139	0.8090	0.8083
KNN	0.7381	0.7328	0.7378	0.7563
HistGradientBoosting	0.7959	0.8010	0.7956	0.7947
Soft Voting	0.8032	0.8108	0.8030	0.8018

Regarding F1-score, Random Forest performed the best with a score of 86.7%, whereas HistGradientBoosting had the lowest F1-score at 76.7%. Based on these evaluation metrics, Random Forest emerged as the most effective classifier to classifying the stage of cirrhosis.

Looking at Table 4 in terms of accuracy values, The Random Forest algorithm is concluded to perform better than the four algorithms in comparison. It gives higher accuracy in respective to other classification algorithms with an accuracy of (97.4 % for 10-fold cross-validation). The remaining three methods have achieved acceptable scores on the various performance classifications. However, a rise in the accuracy score is seen in this table compared to when the 10-fold cross-validation method is applied.

Table 5 presents the results of different machine models with 10-fold validation. Different performance metrics were used to assess the performance of four machine learning classifiers on the liver disease dataset. The Random Forest model achieved the highest accuracy, precision, recall and F1-score. Then, the HistGradientBoosting and soft voting model achieved a score slightly lower than the score achieved by the Random Forest. The KNN model showed the lowest accuracy score. The KNN is only a non-ensemble model, so it tends to be biased toward the majority class in imbalanced datasets since it is likely to have more nearby points from the dominant class. Random Forest is an ensemble of decision trees that fuses several decision boundaries. Random Forest model minimizes overfitting and accomplishes high accuracy. By bootstrapping and selecting a sample of the features at every split, random forest prevents any single tree from dominating the decision-making process. The model is more robust and less sensitive to the noise in the dataset. Compared to the two other ensemble models, Random Forest's performance changes depending on the dataset. Furthermore Random Forest model reduces the false

positives and negatives. Random Forest can handle non-linear data and performs well with default parameters. Unlike KNN, it struggles when the decision boundary is complex. Compared to the KNN model, ensemble models like Random Forest, HistGradientBoosting, and Soft Voting produce superior accuracy, precision, recall, and F1-score. Reducing bias and variance enhances stability and captures complex patterns.

4. CONCLUSION

Liver disease, including cirrhosis, is a significant contributor to mortality in today's world. The application of machine learning techniques holds promise in reducing mortality rates by facilitating early disease detection. This study evaluated the cirrhosis dataset using four classification algorithms: Random Forest, histogram-based Gradient Boosting, KNN, and Soft Voting algorithms. In the experiment, two feature selections have been employed to drop the redundant attributes. Random Forest achieved the highest accuracy with the cross-validation method and was superior to other machine learning algorithms in two datasets. The work can be extended and improved employing a more suitable dataset can give more in-depth information on the variables that can help predict classifying the disease in a better way. In order to achieve this, hyperparameters can be employed to optimize the performance of ensemble and non-ensemble methods.

REFERENCES

[1] The Digestive Process: The Liver and its Many Functions. Johns Hopkins Medicine. 2019 Nov 19. Available from: https://www.hopkinsmedicine.org/health/conditions-and-diseases/the-digestive-process-the-liver-and-its-many-functions.

[2] H. Devarbhavi, S. K. Asrani, J. P. Arab, Y. A. Nartey, E. Pose, and P. S. Kamath, "Global Burden of Liver Disease: 2023 Update," Journal of Hepatology, vol. 79, no. 2, Mar. 2023, doi: https://doi.org/10.1016/j.jhep.2023.03.017.

[3] Mohamed AA, Elbedewy TA, El-Serafy M, El-Toukhy N, Ahmed W, et al. Liver cirrhosis virus: A global view. World Journal of Hepatology. 2015;7(26):2676-80.

[4] Cirrhosis of the Liver. American Liver Foundation. 2023 Mar 16. Available from: https://liverfoundation.org/liver-diseases/complications-of-liver-disease/cirrhosis/.

[5] Buechter M, Gerken G. Liver Function—How to Screen and to Diagnose: Insights from Personal Experiences, Controlled Clinical Studies and Future Perspectives. J Pers Med. 2022;12(10):1657.

[6] Hanif I, Khan MM. Liver Cirrhosis Prediction using Machine Learning Approaches. 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2022.

[7] Ahsan MM, Luna SA, Siddique Z. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. Healthcare (Basel). 2022;10(3):541.

[8] Meng D, Zhang L, Cao G, Cao W, Zhang G, Hu B. Liver Fibrosis Classification Based on Transfer Learning and FCNet for Ultrasound Images. IEEE Access. 2017;5:2169-3536.

[9] Huang R, Rao H, Yang M, Gao Y, Wang J, et al. Noninvasive measurements predict liver fibrosis well in liver cirrhosis virus patients after direct-acting antiviral therapy. Dig Dis Sci. 2020;65(5):1491-500.

[10] Cheng Z, Zhang Y, Zhou C. QSAR models for phosphoramidate prodrugs of 2'-methylcytidine as inhibitors of Liver cirrhosis virus based on PSO boosting. Chem Biol Drug Des. 2011;78(6):948-59.

[11] Bedeir A, El-Hadi M. A Proposed Framework for Predictive Analytics for Cirrhosis of the Liver Using Machine Learning. Mağallaẗ Al-Ğamʿiyyaẗ Al-Miṣriyyaẗ Li Nuẓum Al-Maʿlūmāt wa Tiknūlūğyā Al-Ḥāsibāt. 2023;31(31):114-23.

[12] Rahman AKMS, Javed M, Tasnim Z, Roy J, Hossain SA. A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms. 2019;8(11):419-22.

[13] Sürücü S, Diri B. Transferemble: a classification method for the detection of fake satellite images created with deep convolutional generative adversarial network. J Electron Imaging. 2023;32(4):043004.

[14] Topcu AE, Elbasi E, Alzoubi YI. Machine Learning-Based Analysis and Prediction of Liver Cirrhosis. In: Proceedings of the 2024 47th International Conference on Telecommunications and Signal Processing (TSP); 2024; Prague, Czech Republic. p. 191-194.

[15] Zhang C, Shu Z, Chen S, et al. A machine learning-based model analysis for serum markers of liver fibrosis in chronic hepatitis B patients. Sci Rep. 2024;14:12081.

[16] Choi YS, Oh E. Investigating of Machine Learning Based Algorithms for Liver Cirrhosis Prediction. Adv Eng Intell Syst. 2024;3(1):115-130.

[17] Hirano R, Rogalla P, Farrell C, et al. Development of a classification method for mild liver fibrosis using non-contrast CT image. Int J Comput Assist Radiol Surg. 2022;17:2041-2049.

[18] Devikanniga D, Ramu A, Haldorai A. Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crow Search Algorithm. EAI Endorsed Trans Energy Web. 2020;7(29):e10.

[19] Md AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C. Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines. 2023;11(2):581.

[20] Dickson ER, Grambsch PM, Fleming TR, Fisher LD, Langworthy A. Prognosis in primary biliary cirrhosis: model for decision making. Hepatology. 1989;10(1):1-7.

[21] Straw I, Wu H. Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ Health Care Inform. 2022;29(1).

[22] Lee GY, Alzamil L, Doskenov B, Termehchy A. A survey on data cleaning methods for improved

machine learning model performance. arXiv preprint arXiv:2109.07127. 2021.
[23] Kwak SK, Kim JH. Statistical data preparation: management of missing values and outliers. Korean J Anesthesiol. 2017;70(4):407-411.
[24] Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl. 2004;6(1):20-29.
[25] Chen X, Ishwaran H. Random forests for genomic data analysis. Genomics. 2012;99(6):323-329.
[26] Tembusai ZR, Mawengkang H, Zarlis M. K-nearest neighbor with K-fold cross validation and analytic hierarchy process on data classification. Int J Adv Data Inf Syst. 2021;2(1):45-52.