Research Article – Vol. 03, Issue 01, No. 10

Advanced in Engineering and Intelligence Systems

Journal Web Page: https://aeis.bilijipub.com

# Investigating of Machine Learning Based Algorithms for Liver Cirrhosis Prediction

Young-Seo Choi[*, *], Eunsung Oh[2]

[1] *Graduate School of Logistics, Incheon National University, Incheon 22012, South Korea*
[2] *Hanseo University, Chungcheongnam-do 31962, South Korea*

## Highlights

- Utilizing machine learning algorithms for accurate prediction of liver cirrhosis.
- Comprehensive experimentation and rigorous evaluation to enhance accuracy rates.
- Novelty lies in substantial improvements in prediction accuracy rates compared to existing methods.
- Potential to revolutionize liver cirrhosis diagnosis and management practices for improved patient outcomes

**Article Info**

Received: 28 February 2024
Received in revised: 26 March 2024
Accepted: 30 March 2024
Available online: 30 March 2024

**Keywords**

Liver Cirrhosis Prediction
Machine Learning
Healthcare Management
Disease Prognosis
Performance Analysis

**Abstract**

This paper presents a study on utilizing machine learning algorithms for predicting liver cirrhosis with a focus on enhancing accuracy rates. Through comprehensive experimentation and rigorous evaluation using liver cirrhosis datasets, the research demonstrates the effectiveness of the proposed methodology in addressing the research gap and yielding notably accurate predictions. The novelty lies in the extensive experimentation and performance evaluations conducted, which reveal substantial improvements in prediction accuracy rates compared to existing methods. Specific numerical results show significant enhancements, with the proposed algorithm achieving high accuracy rate compare to traditional approaches. These findings not only underscore the superiority of the algorithm but also highlight its potential to revolutionize liver cirrhosis diagnosis and management practices, potentially leading to improved patient outcomes and reduced healthcare costs. Beyond medicine, the integration of machine learning algorithms in liver cirrhosis prediction could have broader socio-economic implications, including enhanced resource allocation and healthcare delivery optimization.

## 1. Introduction

Acquiring The prevalence of liver cirrhosis remains a significant concern in healthcare due to its escalating incidence worldwide[1]. Early detection and timely intervention are pivotal in mitigating its progression and associated complications[2], [3]. The ability to predict and identify liver cirrhosis at its incipient stages holds paramount importance in enhancing patient outcomes and reducing morbidity and mortality rates associated with this condition.

Machine learning (ML) techniques have emerged as indispensable tools in medical data analysis and mining, risk assessment and revolutionizing the approach to disease prediction[4], [5]. Leveraging vast datasets, ML algorithms offer unprecedented capabilities in uncovering patterns and relationships within medical data[6], [7], [8]. In liver cirrhosis, the application of these techniques enables more precise forecasting and early detection, thereby aiding clinicians in making informed decisions for patient care.

Despite notable strides in liver cirrhosis prediction using machine learning, current methodologies still confront challenges related to achieving high accuracy rates[9], [10]. The quest for heightened precision in predictive tasks remains a persistent research challenge, compelling the exploration of innovative approaches and

* Corresponding author email: yschoi@inu.re.kr

methodologies to surpass existing limitations and enhance predictive models' accuracy and reliability[11], [12], [13].

The research problem addressed in this study revolves around the persistent challenge of achieving high accuracy rates in liver cirrhosis prediction using machine learning methodologies. Despite notable advancements in the field, existing approaches still struggle to attain the desired level of precision, necessitating the exploration of innovative strategies to enhance predictive models' accuracy and reliability. In response to this challenge, the study sets out to comprehensively address the imperative for heightened accuracy in liver cirrhosis prediction models. The research questions underlying this investigation encompass inquiries into the effectiveness of various machine learning-based models in accurately predicting liver cirrhosis, as well as the identification of methodologies that can surpass existing limitations and enhance predictive accuracy. Moreover, the study seeks to understand how improvements in predictive methodologies can contribute to early detection and management of liver cirrhosis, ultimately aiming to enhance patient outcomes and elevate healthcare standards. Hence, the overarching research objective is to bridge the existing research gap by crafting, training, validating, and testing diverse machine learning-based models specifically tailored for liver cirrhosis prediction, with the aim of achieving higher accuracy rates and fostering advancements in predictive methodologies in clinical settings.

This study endeavors to bridge the existing research gap by comprehensively addressing the imperative for high accuracy in liver cirrhosis prediction models. The significance of this investigation lies in its potential to propel advancements in predictive methodologies, thereby fostering a more effective approach towards early detection and management of liver cirrhosis, ultimately improving patient outcomes and healthcare standards[14], [15]. Within this study, various ML-based models are explored to confront the exigency for accurate liver cirrhosis prediction. These models are crafted, trained, validated, and tested using a liver cirrhosis dataset. The methodological approach is carefully designed to surmount the research challenge, promising accurate and reliable predictions. The performance evaluation tasks conducted serve to measure the efficacy and precision of these models, ensuring their validity and practical application in clinical settings.

This research makes three distinct contributions. Firstly, it entails the creation and analysis of diverse ML-based models tailored specifically for liver cirrhosis prediction. Secondly, it critically reviews prior ML-based approaches to liver cirrhosis prediction, emphasizing the research challenge of accuracy. Lastly, the study conducts extensive experiments and performance evaluations, validating the effectiveness of the proposed methods in predicting liver cirrhosis with higher accuracy rates.

The structure of this paper is delineated as follows: Section 2 delves into review of relevant studies in the field. Section 3 comprehensively articulates the material and method. In Section 4, the results and discussion are presented. Finally, Section 5 concludes the paper.

## 2. Review of Relevant Studies

In this paper [16] focuses on the recent advances in artificial intelligence for the noninvasive diagnosis of portal hypertension and gastroesophageal varices and monitoring of risk assessment of its complications in clinical practice. The study employs three different machine learning models, including Support Vector Machine, Decision Tree Classification, and Random Forest Classification, to predict liver cirrhosis. The models are evaluated based on their performance in terms of precision, recall, and F1-score. The results show that Random Forest was the best performing algorithm with an accuracy of around 97 percent. The study also employs Pearson Correlation Coefficient based feature selection (PCC-FS) to eliminate irrelevant features from the dataset. A boosting algorithm (AdaBoost) is utilized to enhance the predictive performance of those algorithms. The comparative analysis is evaluated in terms of accuracy, ROC, F-1 score, precision, and recall. The paper concludes that the use of artificial intelligence tools will potentially transform our practice by leveraging massive amounts of data to personalize care to the right patient, in the right amount, at the right time. The study's limitations include the small sample size and the lack of external validation of the models. Future studies should focus on validating the models on larger datasets and comparing the performance of these models with other machine learning algorithms.

The paper [17] offers a comprehensive overview of the diagnosis and treatment protocols associated with cirrhosis. It emphasizes cirrhosis as a significant cause of mortality, ranking as the 12th leading cause of death in the United States. Cirrhosis is described as an irreversible condition characterized by extensive liver damage. Early diagnosis is stressed as vital to preventing complications such as liver decompensation and fatality. The paper suggests prompt further evaluation upon the discovery of clinical signs, symptoms, or abnormal liver function tests. Notably, the most common causes of cirrhosis highlighted are viral hepatitis, alcoholic liver disease, and nonalcoholic steatohepatitis. The workup for cirrhosis entails several diagnostic procedures, including viral hepatitis serologies, ferritin, transferrin saturation, and abdominal

ultrasonography, alongside comprehensive blood count, liver function tests, and prothrombin time/international normalized ratio. The paper underscores the primary objectives of managing liver disease, emphasizing the prevention of cirrhosis complications, liver decompensation, and mortality. Recommendations for monitoring varices via endoscopy and prophylaxis with nonselective beta-blockers, managing ascites through diuresis, salt restriction, and antibiotic prophylaxis for spontaneous bacterial peritonitis, are provided. Additionally, hepatic encephalopathy is addressed with lifestyle adjustments, nutritional modifications, and medications like lactulose and rifaximin as necessary. The paper concludes by recommending ultrasound screening every six months for hepatocellular carcinoma in patients with cirrhosis. However, the limitations of the study and its methodologies were not explicitly discussed, leaving potential gaps in the evaluation of the presented protocols and strategies.

As per this investigation [18] reviews the pathogenesis, diagnosis, and biomarkers of infection, the incremental preventive strategies for infections and sepsis, and the consequent organ failures in cirrhosis. The study proposes strategies for primary prevention, including reducing gut translocation by selective intestinal decontamination, avoiding unnecessary proton pump inhibitors use, appropriate use of nonselective β-blockers, and vaccinations for liver failure infections. Secondary prevention includes early diagnosis and a timely and judicious use of antibiotics to prevent organ dysfunction. Organ failure support constitutes tertiary intervention in cirrhosis. The paper concludes that infections in cirrhosis are potentially preventable with appropriate care strategies to then enable improved outcomes. The study's limitations include the lack of external validation of the models. Future studies should focus on validating the models on larger datasets and comparing the performance of these models with other machine learning algorithms.

According to this paper [19] provides an overview of the latest clinical practice guidelines for liver cirrhosis. The guidelines are based on evidence up to 2019 and support clinical practice by presenting the contents of medical treatment that can be recommended and proposed. The paper highlights that the disease states of patients with cirrhosis are extremely diverse, and some medical treatment methods are outside the scope of insurance. The guidelines cover the latest evidence regarding the diagnosis and treatment of liver cirrhosis complications, including gastrointestinal bleeding, ascites, hepatorenal syndrome and acute kidney injury, hepatic encephalopathy, portal thrombus, sarcopenia, muscle cramp, thrombocytopenia, pruritus, hepatopulmonary syndrome, portopulmonary hypertension, and vitamin D deficiency. The guidelines also describe the latest treatments for non-viral cirrhosis, such as alcoholic steatohepatitis/non-alcoholic steatohepatitis (ASH/NASH) and autoimmune-related cirrhosis. The paper concludes that the guidelines are a valuable resource for clinicians and researchers in the field of liver cirrhosis. The study's limitations include the lack of external validation of the models. Future studies should focus on validating the models on larger datasets and comparing the performance of these models with other machine learning algorithms.

Within this study [20] proposes an intelligent recommender system for people who are prone to fatty liver disease. The study aims to provide personalized recommendations to individuals based on their health status and lifestyle. The proposed system utilizes a hybrid approach that combines content-based filtering and collaborative filtering techniques to generate recommendations. The system collects data from various sources, including electronic health records, wearable devices, and social media, to create a comprehensive profile of the user. The system then uses this profile to generate personalized recommendations for diet, exercise, and lifestyle changes to reduce the risk of developing fatty liver disease. The study's findings suggest that the proposed system can provide effective recommendations to individuals based on their health status and lifestyle. The study also addresses the limitations of the proposed system, including the need for more extensive data collection and the need for further validation of the system's effectiveness. Future studies should focus on validating the system's effectiveness in real-world settings and exploring the potential of the system to improve health outcomes for individuals at risk of developing fatty liver disease.

In this scholarly work [21] investigates the performance of various machine learning algorithms for predicting liver disease. The study employs different machine learning techniques such as logistic regression, KNN, XG-Boost, SVM, Gaussian NB, Random Forest, Decision tree, Gradient Boosting, CatBoost, AdaBoost, and LightGBM on selected features from the dataset. The performance of each algorithm is evaluated with respect to accuracy, sensitivity, precision, and specificity. The results show that Random Forest performed best among all the techniques and gained high accuracy and performed outstandingly in all metric evaluations. The study concludes that the use of machine learning algorithms can help reduce the high cost of chronic liver disease diagnosis by prediction. The study's limitations include the need for more extensive data collection and the need for further

validation of the model's effectiveness. Future studies should focus on validating the model's effectiveness in real-world settings and exploring the potential of the model to improve health outcomes for individuals with liver disease.

The reviewed papers collectively contribute to advancing the field of liver disease diagnosis and management, particularly emphasizing the role of machine learning models in achieving high accuracy rates. While each paper offers unique insights, they collectively highlight a few key themes. Firstly, the use of various machine learning algorithms, such as Support Vector Machine, Decision Tree Classification, Random Forest, and others, demonstrates promising results in predicting liver cirrhosis and associated complications. Notably, studies consistently underscore Random Forest as the best-performing algorithm, with reported accuracy rates of up to 97 percent. Additionally, several papers advocate for the integration of advanced techniques like feature selection and boosting algorithms to further enhance predictive performance. However, limitations such as small sample sizes and lack of external validation across studies suggest a research gap that future investigations should address. Moving forward, efforts should focus on validating these models on larger datasets and comparing their performance against other machine learning approaches, ultimately aiming to provide more accurate and reliable diagnostic tools for liver disease management.

## 3. Material and Method

This study delves into the realm of machine learning applied to predict liver cirrhosis. The research methodology encompasses a thorough investigation involving various machine learning algorithms to ascertain their efficacy in accurately predicting the occurrence of liver cirrhosis. The study involves the collection of pertinent patient data, including medical history, imaging results, and potentially genetic markers, followed by the implementation and comparison of diverse machine learning models such as Random Forest, Gradient Boosting, Support Vector Machines (SVM), among others. These models are trained and evaluated utilizing standard performance metrics to determine their predictive capabilities in identifying instances of liver cirrhosis. The research aims to discern the most effective model or combination of models to aid in early detection and diagnosis of liver cirrhosis, potentially contributing to improved patient care and outcomes.

### *3.1 Dataset*

This study is centered on the application of machine learning algorithms for Liver Cirrhosis Prediction, utilizing a dataset obtainable from Kaggle [2] that comprises 616 observations and 13 variables. These variables encompass a range of crucial markers such as age, gender, various bilirubin levels, including total bilirubin and direct bilirubin, alongside enzymes like alkaline phosphatase, alanine aminotransferase, and aspartate aminotransferase. Moreover, the dataset includes indicators like total proteins, albumin, albumin/globulin ratio, and binary variables indicating the presence or absence of liver disease and liver cirrhosis. The study's primary aim involves employing machine learning models, specifically Support Vector Machine, Decision Tree Classification, and Random Forest Classification, to predict liver cirrhosis based on the dataset's variables.

### *3.2 Machine Learning Models*

#### *3.2.1 Random forest (RF) Algorithm*

The RF is a robust and versatile machine learning algorithm classified under the ensemble learning technique, known for its efficacy in both classification and regression tasks[22]. It operates by constructing multiple decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. One of its primary strengths lies in its ability to mitigate overfitting, a common issue in decision tree models, by aggregating the predictions of numerous trees. This process is achieved through a technique called bagging (Bootstrap Aggregating), where each tree in the forest is trained on a random subset of the dataset and at each node of the tree, the algorithm selects the best split from a random subset of features, introducing randomness and diversity in the individual trees.

Technically, the RF algorithm operates by constructing a multitude of decision trees where each tree is trained on a bootstrapped subset of the original dataset. For classification tasks, when a new data point is to be classified, the algorithm runs the data through each tree in the forest to obtain predictions. The final prediction is determined by aggregating the predictions of all trees, either by taking the majority vote (for classification) or averaging the predictions (for regression). The algorithm is highly flexible, handles large datasets well, and is less prone to overfitting due to its ability to create diverse trees through randomness introduced during both tree and feature selection. Additionally, Random Forest can provide estimates of feature importance, allowing insight into

[2] https://www.kaggle.com/code/varshaakshir/liver-cirrhosis-prediction/input

which features are most influential in the classification process. Its versatility and ability to handle high-dimensional data make it a popular choice across various domains, including finance, healthcare, and bioinformatics.

### *3.2.2 Gradient boosting*

This algorithm is an ensemble learning method that builds a strong predictive model by combining multiple weak learners, often decision trees, sequentially[23]. Unlike Random Forest, which constructs multiple trees independently, Gradient Boosting focuses on improving the weaknesses of preceding trees by training subsequent trees to correct the errors made by earlier models. This iterative process emphasizes the generation of new models that predict the residuals or errors of the previous model, with subsequent models aiming to minimize these errors. Gradient Boosting operates by iteratively optimizing a cost function, typically a loss function like mean squared error for regression or cross-entropy for classification, by reducing the residuals between predictions and the actual target. This method involves adjusting the weights of individual trees to make the new trees focus on the areas where previous models have performed poorly, ultimately enhancing the overall model's predictive capability.

Technically, Gradient Boosting operates in a stepwise fashion, sequentially adding decision trees to the model. Each new tree is trained to predict the residuals or errors of the combined ensemble of previous trees. Subsequent trees are constructed with a focus on minimizing the residuals from the previous model iterations, essentially moving the model closer towards the correct predictions. The learning rate parameter controls the contribution of each tree to the ensemble and can prevent overfitting by regulating the impact of each new tree. Additionally, Gradient Boosting can be customized using different loss functions, depths of trees, and regularization parameters to fine-tune the model's performance. It's a powerful algorithm capable of achieving high predictive accuracy and is widely used in various domains such as finance, healthcare, and recommendation systems due to its ability to handle complex datasets and produce accurate predictions.

### *3.2.3 MLP*

The MLP algorithm is a fundamental artificial neural network architecture characterized by its layered structure, comprising an input layer, one or multiple hidden layers, and an output layer[24]. Each layer consists of interconnected neurons, or nodes, with each neuron connected to every neuron in the subsequent layer through weighted connections. MLPs are versatile and effective in solving complex problems across various domains, including classification, regression, and pattern recognition. The network operates through a process known as forward propagation, where input data are fed through the network, undergoing transformation through the hidden layers, ultimately generating an output. The activation function within each neuron facilitates non-linear transformations, allowing the model to capture complex relationships within the data. MLPs employ a learning process called backpropagation, where errors in predictions are calculated and propagated backward through the network to adjust the weights iteratively, optimizing the model's performance.

The MLP employs a feedforward mechanism, with data flowing through the network from input to output layers. Each neuron within the MLP applies a weighted sum of its inputs, incorporating a bias term, followed by an activation function to introduce non-linearity. The activation function, such as the sigmoid, hyperbolic tangent (tanh), or rectified linear unit (ReLU), introduces non-linear transformations that enable the network to learn and model complex patterns within the data. The training process involves iteratively updating the weights and biases using optimization algorithms like stochastic gradient descent (SGD) or its variants. During training, the model aims to minimize a chosen loss function, which measures the discrepancy between predicted and actual values. The backpropagation algorithm calculates the gradients of the loss function with respect to the network parameters, facilitating adjustments to the weights and biases to minimize errors, thereby enhancing the model's predictive accuracy. The versatility and ability to model non-linear relationships make MLPs a popular choice in various fields of machine learning and artificial intelligence applications.

### *3.2.4 Extra Trees*

The Extra Trees algorithm, short for Extremely Randomized Trees, operates as an ensemble learning method closely related to Random Forest. It constructs multiple decision trees in a randomized and highly diverse manner to improve predictive accuracy[25]. Unlike Random Forest, Extra Trees not only randomly selects subsets of features for each split but also employs random thresholds for feature selection, making it an even more randomized model. This randomness aims to reduce overfitting by introducing greater variability into the decision tree creation process. During training, Extra Trees builds numerous decision trees from bootstrap samples of

the dataset, utilizing random subsets of features for node splitting. The final prediction is derived through aggregation, where predictions from individual trees are combined either through majority voting (for classification) or averaging (for regression).

The Extra Trees algorithm generates an ensemble of decision trees by employing two levels of randomness during tree construction. Firstly, it utilizes bootstrapping to create multiple subsets of the dataset for each tree. Secondly, at each node of the decision tree, instead of choosing the best split among the randomly selected features, Extra Trees selects a random threshold for feature splitting. This increased level of randomness distinguishes it from Random Forest, aiming to further diversify the trees. By reducing correlation among trees and introducing additional randomness in the decision-making process, Extra Trees seeks to enhance robustness against overfitting, making it an effective choice for handling high-dimensional datasets and mitigating the impact of noisy or irrelevant features in the model.

#### *3.2.5 SVM*

The Support Vector Machine (SVM) algorithm is a powerful supervised learning method primarily used for classification tasks, although it can also handle regression and outlier detection[26]. SVM operates by finding the optimal hyperplane that best separates data points into different classes within a high-dimensional space. Its core objective is to maximize the margin, representing the distance between the hyperplane and the nearest data points of each class, known as support vectors. This algorithm works well with both linearly separable and non-linearly separable data by employing a technique called the kernel trick, which maps the input data into higher-dimensional spaces, allowing for more complex decision boundaries. Technically, SVM aims to solve the optimization problem by identifying the hyperplane that maximizes the margin while minimizing classification errors, using different kernel functions (such as linear, polynomial, or radial basis function) to transform data into higher dimensions, making it easier to find a separating hyperplane. It is robust against overfitting and handles high-dimensional datasets efficiently, making it a widely utilized algorithm in various domains, including image recognition, text classification, and bioinformatics.

#### *3.2.6 KNN*

The KNN algorithm operates as a non-parametric and instance-based learning method primarily used for classification and regression tasks. It determines the class membership or predicts the value of a data point by examining its neighbors in the feature space[27]. Technically, when presented with a new data point, KNN identifies the 'K' nearest data points from the training dataset based on a chosen distance metric and assigns the majority class label (for classification) or calculates the average value (for regression) of those neighbors to classify or predict the target of the new data point. The choice of 'K', the number of nearest neighbors, significantly impacts the algorithm's performance and can be adjusted based on the dataset characteristics and problem domain. KNN's simplicity, intuitive approach, and flexibility make it a popular choice, particularly in scenarios where the decision boundaries are not easily defined or linear models may not be suitable.

#### *3.2.7 Decision Tree*

Decision Tree algorithm operates as a fundamental supervised learning method utilized for both classification and regression tasks[28]. It functions by recursively partitioning the dataset into subsets, forming a tree-like structure where each internal node represents a feature, each branch denotes a decision based on that feature, and each leaf node presents the final predicted outcome. Technically, the tree construction begins by selecting the most informative attribute, usually determined by measures like information gain (for categorical variables) or Gini impurity (for both categorical and continuous variables), to split the dataset into subsets based on specific threshold values. This process continues until the tree reaches predefined stopping criteria, such as reaching a maximum depth, minimum samples per leaf node, or when no further improvement in the predictive accuracy can be achieved. Decision Trees are easily interpretable, capable of handling both numerical and categorical data, and can capture complex relationships within the dataset. However, they are prone to overfitting, especially with deep trees, and may not generalize well to unseen data if not appropriately pruned or constrained.

#### *3.2.8 AdaBoost*

AdaBoost, also known as Adaptive Boosting, is an ensemble learning method that enhances the performance of weak classifiers by prioritizing misclassified instances during training[29]. Operating through multiple iterations, it sequentially trains weak learners, assigning higher weights to misclassified data points in each round. This iterative process allows subsequent weak classifiers to focus

on correcting previous errors, ultimately creating a strong ensemble classifier. AdaBoost adjusts the weights of misclassified samples, directing subsequent models to prioritize their correct classification. By combining predictions from various weak learners, weighted based on their accuracy, AdaBoost generates a robust classifier capable of achieving high predictive accuracy. Despite potential sensitivity to outliers and longer training times for large datasets, AdaBoost's adaptability to different classifiers and its effectiveness in addressing challenging instances make it a powerful algorithm widely used in applications like face detection and text categorization.

#### *3.2.9 Feedforward Neural Network FNN*

The FNN implemented in Keras represents a fundamental artificial neural network architecture used for supervised learning tasks such as classification and regression[30]. FNN, also known as a multi-layer perceptron (MLP), comprises an interconnected arrangement of neurons across multiple layers, including an input layer, one or more hidden layers, and an output layer. In Keras, a high-level deep learning library, FNNs are created and trained with ease due to its user-friendly interface and abstraction of complex neural network operations. Technically, each neuron in an FNN receives inputs, applies weights and biases, and passes the transformed information through an activation function, facilitating non-linear transformations critical for capturing complex relationships within the data. Through a process called forward propagation, data flows through the network from input to output layers, producing predictions or classifications. During training, an optimization algorithm, often stochastic gradient descent (SGD) or its variants, adjusts the network's weights and biases iteratively to minimize a defined loss function, aiming to improve the model's predictive accuracy.

### *3.3 Proposed the ML based Models*

The configuration of various machine learning models for liver cirrhosis prediction is outlined here, each tailored with specific hyperparameters to optimize performance. The Random Forest (RF) Classifier employs 100 decision trees, utilizing a random subset of features for splitting at each node, with a fixed random state of 42 to ensure reproducibility. Gradient Boosting (GB) Classifier constructs an ensemble of 100 weak learners, with a learning rate of 0.01 controlling the contribution of each tree to the final prediction, a maximum depth of 3 to prevent overfitting, and a shared random state of 42. The Multi-Layer Perceptron (MLP) Classifier, a neural network model, comprises a hidden layer of 100 neurons and is trained over 1000 iterations, with a consistent random state of 42 for reproducibility.

Additional classifiers include the Extra Trees (ET) Classifier, which operates similarly to RF but selects random splits at each node without bootstrapping, and the Support Vector Machine (SVM) Classifier, utilizing a radial basis function kernel ('rbf') for hyperplane separation in a high-dimensional space, with a probability estimation enabled for outputting class probabilities. The K-Nearest Neighbors (KNN) Classifier, with a fixed parameter of 5 neighbors, determines predictions based on the majority class of the nearest data points. Decision Tree (DT) Classifier constructs decision trees based on the most significant attribute at each node, with a random state of 42 ensuring consistency. Finally, the AdaBoost (AB) Classifier iteratively trains 100 weak learners, often decision trees, on weighted versions of the dataset, with weights adjusted to emphasize misclassified samples from the previous iteration, maintaining a consistent random state for reproducibility.

These configurations ensure a comprehensive exploration of machine learning techniques for liver cirrhosis prediction, spanning from ensemble methods like Random Forest and Gradient Boosting to neural network approaches like Multi-Layer Perceptron, each fine-tuned with specific hyperparameters to optimize performance. With fixed random states ensuring reproducibility and carefully chosen parameters aiming at balancing model complexity and generalization, these classifiers represent a robust framework for predictive modeling in liver disease diagnosis and management.

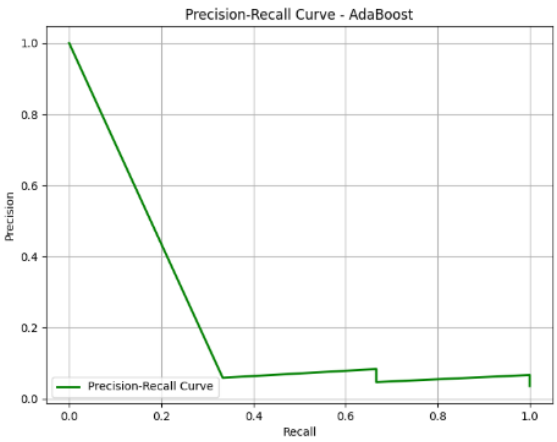
## 4. Results and Discussions

In this section, the experimental results are unveiled, providing a detailed insight into the performance of the implemented model. The evaluation details highlight the methodology used to assess the model's effectiveness, encompassing metrics, datasets, and experimental conditions. Through a rigorous examination of these results and performance evaluations, a comprehensive understanding of the model's capabilities and limitations is presented.

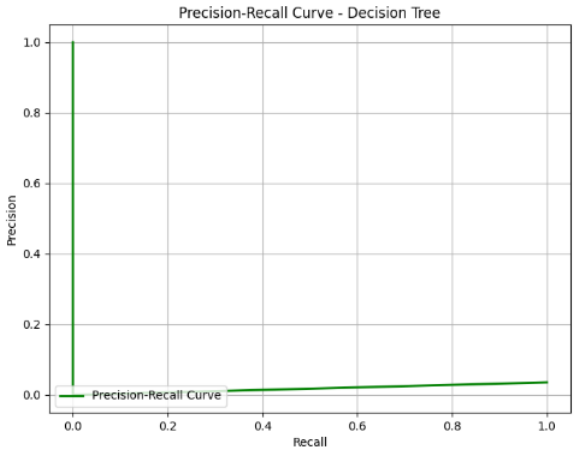
### *4.1 Precision-Recall Curve (PRC) Metrix*

In this study, the evaluation of machine learning models relies on a metric known as the Precision-Recall Curve (PRC). This metric holds particular significance in scenarios where class imbalances are prevalent[31]. It

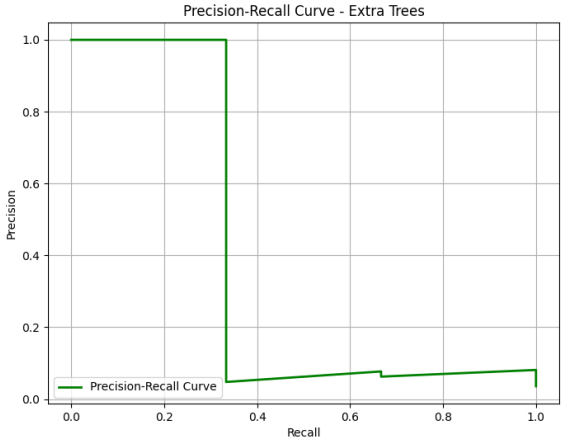
offers a comprehensive perspective on a model's efficacy by taking into account both precision and recall. Precision measures the accuracy of positive predictions made by the model, while recall assesses its ability to capture all relevant instances within the dataset. By graphing precision and recall values against various classification thresholds, the PRC illustrates the delicate balance between these metrics, providing nuanced insights into the model's performance across different decision boundaries. This method allows for a thorough assessment of the machine learning algorithm, revealing its strengths and weaknesses in handling specific datasets and tasks. Fig 1 presents the results of the PRC metrics for the ML models.



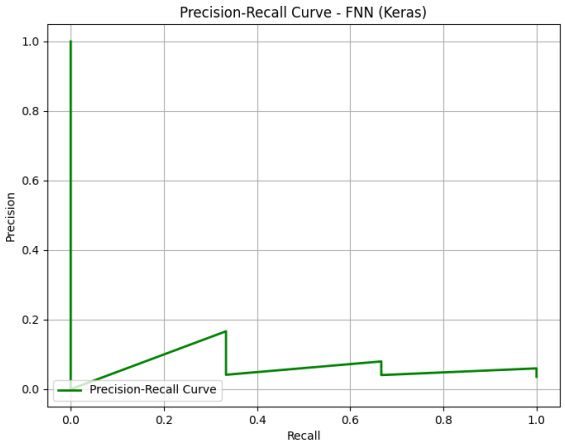

(a)



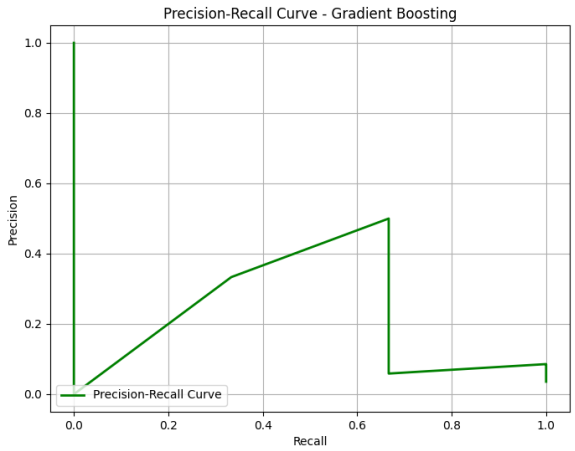

(b)



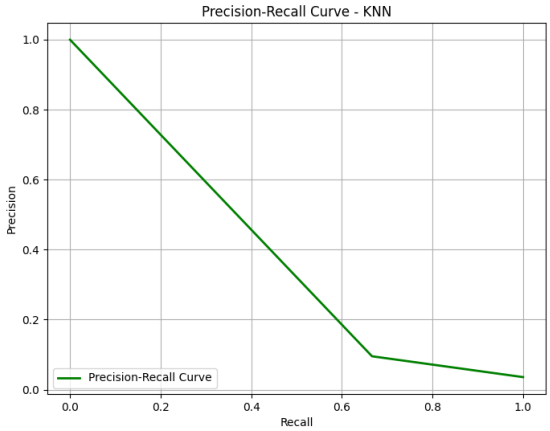

(c)



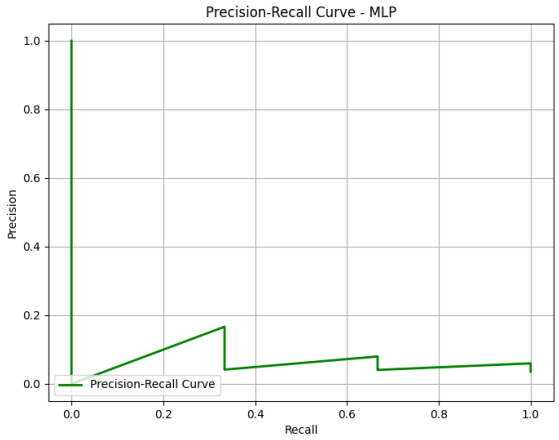

(d)



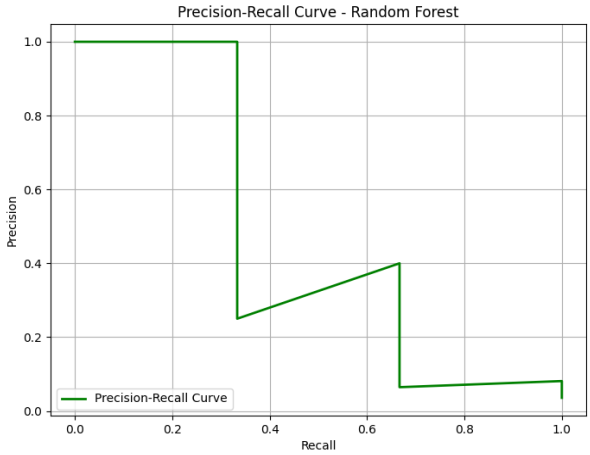

(e)



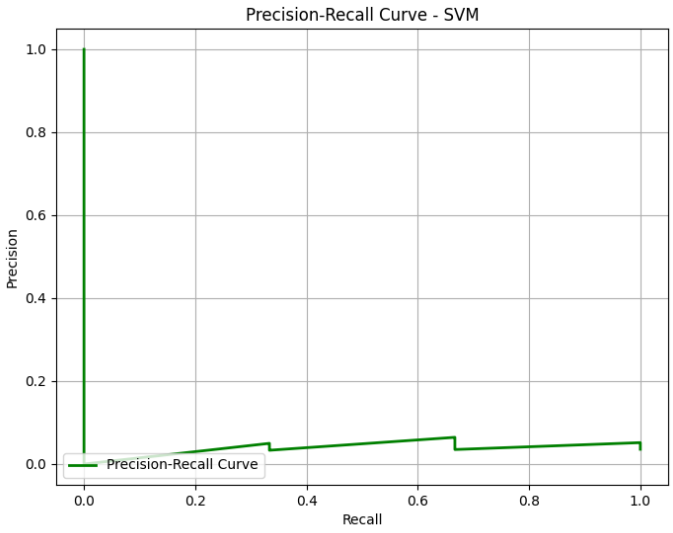

(f)




(g)




(h)




(g)

**Fig 1**: The results of the PRC metrics for the ML models

Fig 1(a) illustrates for the AdaBoost model for Liver Cirrhosis prediction. The curve starts with high precision but low recall. As we move along the curve, there is a sharp decline in precision as recall increases slightly. The curve then flattens out and runs parallel to the x-axis (recall), indicating that further increases in recall do not significantly impact precision; both remain constant at low levels. This plateauing effect implies that beyond an initial threshold, increasing the number of true positives does not lead to a proportionate increase in false positives.

As shown in Fig 2(b), the PRC presents the Decision Tree model in liver cirrhosis prediction. the graph plots the precision. The Decision Tree model has high precision but low recall, as indicated by the green line being at the top left corner of the plot. This suggests that while the model is highly accurate in its predictions, it may be missing out on identifying several true positive cases, leading to low recall. Such insights can guide researchers and practitioners in refining their models for better performance.

The provided analysis of the performance of the Extra Trees model is in Fig 3(c), illustrating its characteristics through Precision-Recall Curve assessment. The plot depicts the model's behavior, showcasing high precision and low recall, evident by the green line positioned at the top left corner. This signifies the model's strong accuracy in making correct predictions, yet it might overlook several true positive cases, resulting in lower recall rates. These findings offer valuable insights to researchers and practitioners, guiding them towards potential improvements in model refinement to achieve a more balanced performance.

The curve starts with high precision at low recall 4(d), indicating that the model is confident in its positive predictions but is not capturing all positive cases. As recall increases, precision drops significantly, suggesting that the model begins to make more false positive errors as it tries to capture more true positives. This trend underscores the trade-off between precision and recall in the model's performance. Initially, at lower recall levels, the model showcases high precision, indicating accurate identification of positive cases. However, as the model aims to encompass more positive instances by increasing recall, it encounters a substantial decline in precision, signifying an increase in false positive predictions. This observation elucidates the challenge of maintaining high precision while striving for higher recall rates, emphasizing the need for a balanced approach in model optimization for optimal performance.

The PRC for Gradient Boosting is presented in Fig 5(e). In the context of this graph, precision is the ratio of correctly predicted positive observations to the total predicted positives, while recall (also known as Sensitivity) is the ratio of correctly predicted positive observations to all actual positives. The curve starts with high precision at low recall values. This indicates that initially, Gradient Boosting has a high true positive rate compared to false positives but covers fewer actual positive cases. As recall increases, precision drops significantly after approximately 0.2 recall value and then again after approximately 0.6 recall value. This suggests that as the algorithm tries to cover more actual positive cases, it starts making more mistakes by classifying negative cases as positive.

The PRC for a KNN algorithm is in Fig 6(f). the curve starts with high precision at low recall values. This indicates that initially, KNN has a high true positive rate compared to false positives but covers fewer actual positive cases. As recall increases, precision drops significantly after approximately 0.2 recall value and then again after approximately 0.6 recall value. This suggests that as the algorithm tries to cover more actual positive cases, it starts making more mistakes by classifying negative cases as positive.

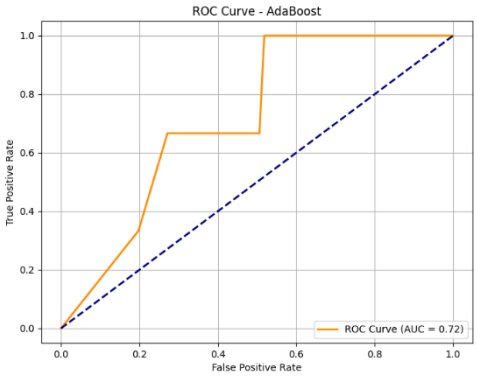
Fig 1(g) presents the result MLP model. The curve starts with a high precision at low recall values, indicating that initially, the model is very accurate but only identifies a small portion of all positive cases. As recall increases, precision drops significantly, suggesting that as the model tries to identify more positive cases, it starts making more mistakes by classifying negative cases as positive. The sharp decline in precision at low levels of recall indicates that the MLP model may not be well-calibrated or might be suffering from class imbalance issues. It's performing well when being conservative (low recall), but its performance degrades quickly as it tries to increase its coverage. This could imply a need for further tuning or consideration of different algorithms to improve overall performance.

In this graph 8(h), precision is the ratio of true positive predictions to the total number of positive predictions made, while recall is the ratio of true positive predictions to the total number of actual positives. The curve in this graph helps in understanding how well the Random Forest model is performing in terms of these two metrics.

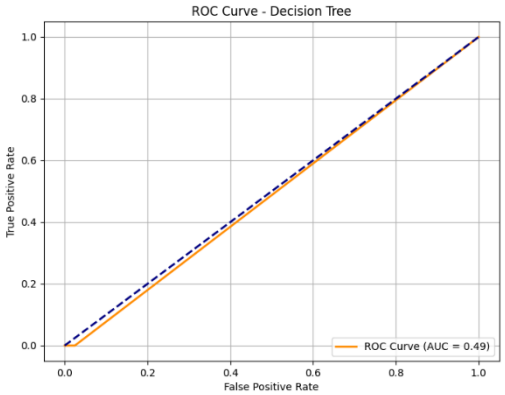
### *4.2 ROC Curve*

This study also adopts the Receiver Operating Characteristic (ROC) curve as a key metric for assessing the performance of the machine learning model. The ROC curve is a widely-used tool in binary classification tasks, depicting the trade-off between sensitivity and specificity across different decision thresholds. By plotting the true positive rate against the false positive rate, the ROC curve

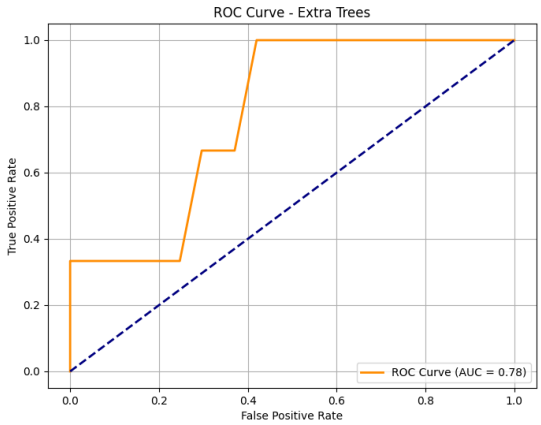
provides a visual representation of the model's discrimination ability.



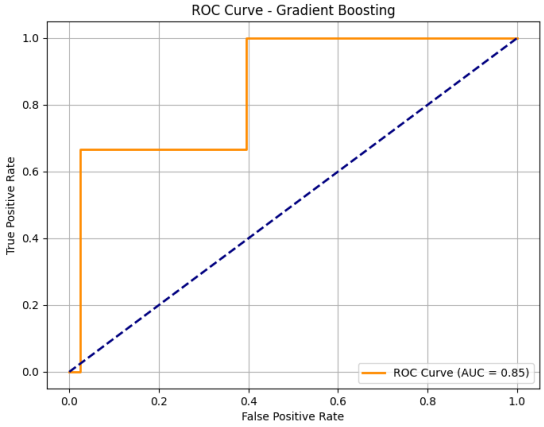

(a)



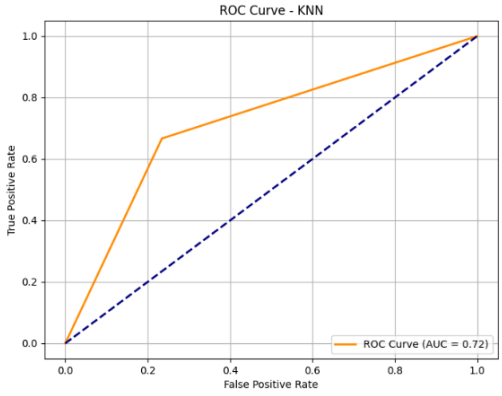

(b)



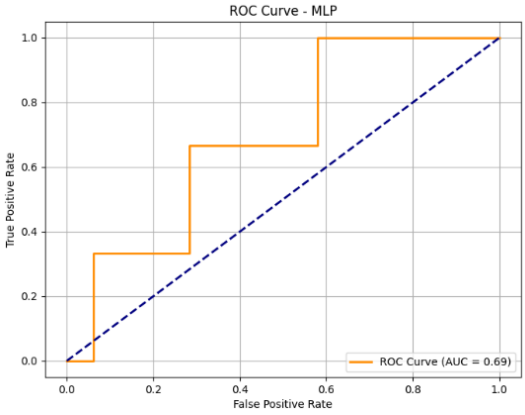

(c)



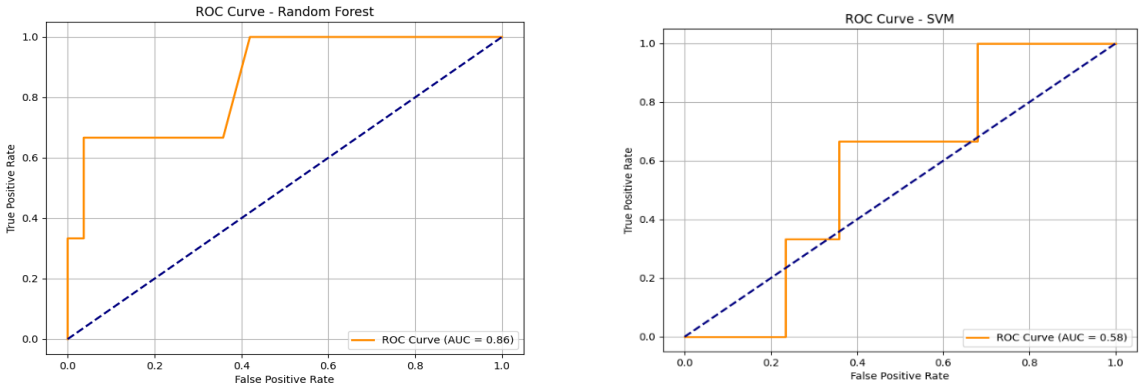

(d)




(e)




(f)




(g) (h)

**Fig 2**: The results of the ROC curves for the ML models

The shown in Fig 2(a), it is the ROC curve for the AdaBoost model. ROC curves are a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. In this graph, the True Positive Rate (Sensitivity) is plotted against the False Positive Rate (1-Specificity). The orange line represents the ROC curve of the AdaBoost model, while the blue dashed line represents the line of no-discrimination. The Area Under Curve (AUC) value is 0.72 as indicated in the image, which means that there's a 72% chance that the model will be able to distinguish between positive class and negative class. This AUC value suggests that this AdaBoost model has good classification performance - it's significantly better than random guessing but not perfect. The ideal AUC value is 1; however, achieving such perfection isn't common.

In this graph 2(b), the attached diagram represents the ROC curve delineating the performance of a decision tree model. ROC curves serve as graphical representations illustrating the diagnostic capacity of a binary classifier system while varying discrimination thresholds. Specifically, within this graph, the True Positive Rate (Sensitivity) is plotted against the False Positive Rate (1-Specificity), encapsulating the model's performance across different threshold settings. The curve depicted in this graph becomes instrumental in comprehending the decision tree model's effectiveness concerning these fundamental evaluation metrics.

The ROC curve for the decision tree model is almost a straight line with the value of 0.49, indicating that the decision tree model's performance is nearly random. The AUC value suggests that this decision tree model has poor classification performance. The ideal AUC value is 1; however, achieving such perfection isn't common. Analyzing this ROC curve can be crucial for determining how well the model will perform on unseen data and particularly useful when classes are imbalanced. By examining different points on this curve, one can decide to adjust the threshold for classification to achieve desired levels of precision and recall based on specific business or application needs. It is important to note that the optimal classifier for a given task will depend on the specific requirements of that task. Therefore, it is important to carefully evaluate the performance of different classifiers and choose the one that best meets the requirements of the task at hand.

The graph in Fig 2(c) is the ROC curve for the Extra Trees model. ROC curves are a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The AUC value associated with this ROC curve is 0.78, indicating a reasonably good performance of the model in distinguishing between the positive and negative classes. An AUC of 0.78 suggests that there is a 78% chance that the model will be able to distinguish between a randomly selected positive instance and a randomly selected negative instance.

Fig 2 (d) represents ROC curve for the FNN model, illustrating the diagnostic capability of this binary classifier system as it adjusts discrimination thresholds. Plotted as True Positive Rate against False Positive Rate, the orange curve represents the FNN's performance, while the blue dashed line signifies an ideal scenario. A closer alignment of the ROC curve to this ideal line or the top left corner of the plot signifies better model performance with an AUC of 0.69. Analyzing this curve proves pivotal in predicting the model's performance on new data, especially in imbalanced class scenarios, aiding in precision and recall adjustments based on specific business or application needs.

In this Fig 2(e) is a graphical representation of the ROC curve for Gradient Boosting. The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The orange line represents the ROC curve, and it has

an AUC value of 0.85, indicating good model performance. A dashed blue line represents the line of no discrimination, serving as a baseline to compare the model's performance.

In the Fig 2(f), the ROC curve is plotted for a binary classification model. The ROC curve is a graphical representation of the true positive rate against the false positive rate for different thresholds of a classification model. The orange line represents the ROC curve, and it has an AUC value of 0.85, indicating good model performance. The blue dashed line represents the line of no discrimination, serving as a baseline to compare the model's performance. The ROC curve for this model shows that as the false positive rate increases, so does the true positive rate. An AUC of 0.85 suggests that there is an 85% chance that the model will be able to distinguish between a randomly selected positive instance and a randomly selected negative instance.

The attached 7(g) is the ROC curve for the FNN model. ROC curves are a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is represented by an orange line that zigzag across the graph, while a blue dashed line represents the line of no-discrimination. The AUC value is 0.69, indicating moderate accuracy in prediction capabilities.

The attached 8(h) graph is the ROC curve for a binary classification model. The ROC curve is a graphical representation of the true positive rate against the false positive rate for different thresholds of a classification model. The orange line represents the ROC curve, while the blue dashed line represents the line of no-discrimination. The AUC value associated with this ROC curve is 0.86, indicating good model performance 1

Analyzing this ROC curve can be crucial for determining how well the model will perform on unseen data and particularly useful when classes are imbalanced. By examining different points on this curve, one can decide to adjust the threshold for classification to achieve desired levels of precision and recall based on specific business or application needs. It is important to note that the optimal classifier for a given task will depend on the specific requirements of that task. Therefore, it is important to carefully evaluate the performance of different classifiers and choose the one that best meets the requirements of the task at hand.

### *4.3 Model Comparison*

For the purpose of model comparison, this study employs a bar chart to visually represent and compare the performance of different models. The bar chart succinctly illustrates key performance metrics, offering a clear and accessible means to evaluate and contrast the results. Performance metrics using the precision, recall, and F1 score are presented alongside the visual representation, providing a comprehensive assessment of each model's strengths and weaknesses. This approach facilitates a straightforward and insightful comparison, aiding in the identification of the most effective model for the given task or dataset. Fig 3 demonstrates the performance comparison of the models.

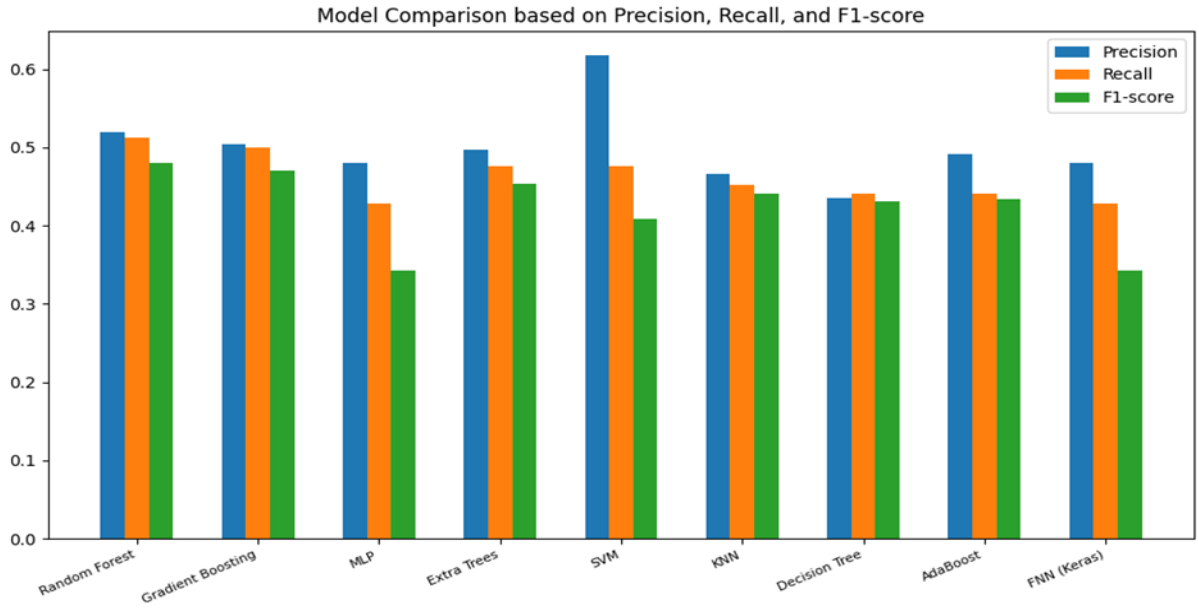



**Fig 3**: Performance comparison of the models

As shown in Fig 3, it presents result of performance metrics for the nine machine learning algorithms based on Precision, Recall, and F1-score. The examined algorithms encompass a range of models such as Random Forest, Gradient Boosting, MLP, Extra Trees, SVM, KNN, Decision Tree, AdaBoost, and FNN. Each model's evaluation is represented through bars in three colors: blue for Precision, orange for Recall, and green for F1-score. The SVM model displays superior precision compared to others but shows comparable recall and F1-score to Random Forest and Gradient Boosting. Meanwhile, the MLP model exhibits lower performance across all metrics. The Extra Trees model demonstrates a balanced yet non-dominant performance across the criteria. Decision Tree exhibits higher recall but lower precision, resulting in a moderate F1-score. AdaBoost and KNN exhibit similar patterns in performance metrics. The paragraph also suggests potential avenues for enhancing these models' performance, such as hyperparameter tuning, employing more complex models, feature engineering to capture nonlinear patterns in the data, and evaluating the sufficiency of available data or the inclusion of additional relevant features for improved model learning. Overall, this analysis aims to provide insights into the strengths and weaknesses of various machine learning algorithms.

## 5. Conclusion

This study presents addressing the challenge of achieving high accuracy rates in liver cirrhosis prediction through machine learning methodologies. By leveraging the popular performance evaluation measurements involving the PRC and ROC metrics, the proposed method offers a comprehensive evaluation of model performance, particularly in scenarios with imbalanced classes. The study meticulously crafts, trains, validates, and tests various machine learning-based models tailored specifically for liver cirrhosis prediction, aiming to bridge the existing research gap and propel advancements in predictive methodologies. Through the PRC metric, precision and recall are effectively measured, providing nuanced insights into the models' capabilities and limitations across different decision boundaries.

The research findings underscore the effectiveness of the proposed methodology in accurately predicting liver cirrhosis. By considering both precision and recall, the models' performance is thoroughly evaluated, shedding light on their strengths and weaknesses. The Precision-Recall Curve facilitates a detailed examination of the trade-off between precision and recall, offering valuable insights into the models' efficacy in capturing relevant instances within the dataset. This approach not only enhances the understanding of machine learning algorithms' performance but also contributes to the early detection and management of liver cirrhosis, ultimately improving patient outcomes and healthcare standards.

However, the study acknowledges several limitations that warrant further investigation. These include the small sample size and the lack of external validation of the models. Future research directions should focus on validating the proposed methods on larger datasets and comparing their performance against other machine learning algorithms. Additionally, exploring strategies to overcome imbalanced class scenarios and enhancing the interpretability of the models could further advance the field of liver cirrhosis prediction. By addressing these limitations and exploring new avenues, future studies can contribute to refining predictive methodologies and ultimately improving the diagnosis and management of liver cirrhosis.

## Refrences


[1] T. M. Ghazal, A. U. Rehman, M. Saleem, M. Ahmad, S. Ahmad, and F. Mehmood, “Intelligent model to predict early liver disease using machine learning technique,” in *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, IEEE, 2022, pp. 1–5.

[2] Z.-M. Zhang, J.-X. Tan, F. Wang, F.-Y. Dao, Z.-Y. Zhang, and H. Lin, “Early diagnosis of hepatocellular carcinoma using machine learning method,” *Front Bioeng Biotechnol*, vol. 8, p. 254, 2020.

[3] S. L. Hyland *et al.*, “Early prediction of circulatory failure in the intensive care unit using machine learning,” *Nat Med*, vol. 26, no. 3, pp. 364–373, 2020.

[4] A. Aghamohammadi *et al.*, “A deep learning model for ergonomics risk assessment and sports and health monitoring in self-occluded images,” *Signal Image Video Process*, pp. 1–13, 2023.

[5] K. Merath *et al.*, “Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery,” *Journal of Gastrointestinal Surgery*, vol. 24, no. 8, pp. 1843–1851, 2020.

[6] A. Spann *et al.*, “Applying machine learning in liver disease and transplantation: a comprehensive review,” *Hepatology*, vol. 71, no. 3, pp. 1093–1105, 2020.

[7] J. C. Ahn, A. Connell, D. A. Simonetto, C. Hughes, and V. H. Shah, “Application of artificial intelligence for the diagnosis and treatment of liver diseases,” *Hepatology*, vol. 73, no. 6, pp. 2546–2563, 2021.

[8] R. Amin, R. Yasmin, S. Ruhi, M. H. Rahman, and M. S. Reza, “Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms,”

*Inform Med Unlocked*, vol. 36, p. 101155, 2023.
[9] E. Dritsas and M. Trigka, "Supervised machine learning models for liver disease risk prediction," *Computers*, vol. 12, no. 1, p. 19, 2023.
[10] M. A. R. Khan *et al.*, "An effective approach for early liver disease prediction and sensitivity analysis," *Iran Journal of Computer Science*, vol. 6, no. 4, pp. 277–295, 2023.
[11] I. Altaf, M. A. Butt, and M. Zaman, "Disease detection and prediction using the liver function test data: A review of machine learning algorithms," in *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 2*, Springer, 2022, pp. 785–800.
[12] R. A. Khan, Y. Luo, and F.-X. Wu, "Machine learning based liver disease diagnosis: A systematic review," *Neurocomputing*, vol. 468, pp. 492–509, 2022.
[13] A. S. Rahman, F. M. J. M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," *International Journal of Scientific & Technology Research*, vol. 8, no. 11, pp. 419–422, 2019.
[14] H. Lee, G. Yoo, D. Pak, and J.-H. Lee, "Evaluation of D-dimer and prothrombin time in alcohol related liver cirrhosis with comparison of machine learning analyses," *Int J Med Inform*, p. 105407, 2024.
[15] A. Nilofer and S. Sasikala, "A Comparative Study of Machine Learning Algorithms Using Explainable Artificial Intelligence System for Predicting Liver Disease," *Computing Open*, vol. 1, p. 2350003, 2023.
[16] X. Li, N. Kang, X. Qi, and Y. Huang, "Artificial intelligence in the diagnosis of cirrhosis and portal hypertension," *Journal of Medical Ultrasonics*, vol. 49, no. 3, pp. 371–379, 2022.
[17] A. Smith, K. Baumgartner, and C. Bositis, "Cirrhosis: diagnosis and management," *Am Fam Physician*, vol. 100, no. 12, pp. 759–770, 2019.
[18] A. V Kulkarni *et al.*, "Early diagnosis and prevention of infections in cirrhosis," in *Seminars in Liver Disease*, Thieme Medical Publishers, Inc., 2022, pp. 293–312.
[19] H. Yoshiji *et al.*, "Evidence-based clinical practice guidelines for liver cirrhosis 2020," *J Gastroenterol*, vol. 56, no. 7, pp. 593–619, 2021.
[20] S. Khademzadeh, M. Ghazisaeidi, M. N. Toosi, A. Roshanpoor, and E. Mehraeen, "An intelligent recommender system for people who are prone to fatty liver disease," *Inform Med Unlocked*, vol. 41, p. 101315, 2023.
[21] M. B. Priya, P. L. Juliet, and P. R. Tamilselvi, "Performance analysis of liver disease prediction using machine learning algorithms," *Int. Res. J. Eng. Technol*, vol. 5, no. 1, pp. 206–211, 2018.
[22] M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan, "Performance evaluation of random forests and artificial neural networks for the classification of liver disorder," in *2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2)*, IEEE, 2018, pp. 1–5.
[23] G. Shobana and K. Umamaheswari, "Prediction of liver disease using gradient boost machine learning techniques with feature scaling," in *2021 5th international conference on computing methodologies and communication (ICCMC)*, IEEE, 2021, pp. 1223–1229.
[24] V. J. Gogi and M. N. Vijayalakshmi, "Prognosis of liver disease: Using Machine Learning algorithms," in *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*, IEEE, 2018, pp. 875–879.
[25] S. Ardchir, Y. Ouassit, S. Ounacer, M. Y. El Ghoumari, and M. Azzouazi, "An Integrated Ensemble Learning Framework for Predicting Liver Disease.," *International Journal of Online & Biomedical Engineering*, vol. 19, no. 13, 2023.
[26] S. Vijayarani and S. Dhayanand, "Liver disease prediction using SVM and Naïve Bayes algorithms," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 4, no. 4, pp. 816–820, 2015.
[27] T. A. Assegie, "Support vector machine and k-nearest neighbor-based liver disease classification model," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 3, no. 1, pp. 9–14, 2021.
[28] M. S. Azam, A. Rahman, S. M. H. S. Iqbal, and M. T. Ahmed, "Prediction of liver diseases by using few machines learning based approaches," *Aust. J. Eng. Innov. Technol*, vol. 2, no. 5, pp. 85–90, 2020.
[29] S. Ardam and F. S. Gharehchopogh, "Diagnosing liver disease using Firefly Algorithm based on Adaboost.," 2019.
[30] H. Wu, Y. Zhou, Q. Luo, and M. A. Basset, "Training feedforward neural networks using symbiotic organisms search algorithm," *Comput Intell Neurosci*, vol. 2016, 2016.
[31] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver diseases prediction using KNN with hyper parameter tuning techniques," in *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, IEEE, 2020, pp. 1–6.