ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

A Comparative Analysis of Machine Learning Techniques for the Automated Detection of Liver Cirrhosis

Md. Sharfuddin
Department of Computer Science and Engineering,
Southeast University, Dhaka, Bangladesh

Mst. Momana Akter Mim
Department of Diploma in Nursing Science and Midwifery,
East West Nursing College and Institute, Dhaka, Bangladesh

Abstract
Liver cirrhosis is a chronic progressive liver disease that leads to irreversible liver damage (cirrhosis is irreversible by definition) and can culminate in liver failure, hepatocellular carcinoma, and death if not diagnosed early. Cirrhosis is a serious complication of chronic liver disease, and it is critical to promptly diagnose the cirrhosis stages in order to provide timely treatment and improve patient outcomes. Nevertheless, standard diagnostic approaches tend to be invasive, costly, and reliant upon skilled evaluation. Artificial Intelligence (AI) and, in particular, machine learning (ML) has shown promise as an approach to build accurate non-invasive diagnostic systems; however, most studies rely on few algorithms and have low model interpretability. In this work, we present a framework to predict liver cirrhosis stage based on supervised machine learning that uses the Liver Cirrhosis Stage Classification dataset. The proposed method involves Steps: Data Pre-processing, Handling Missing Values, and Categorical Encoding, Feature Selection, Handling Imbalanced Dataset and the multiple Base Learning and Ensemble learning models implementation. Explainable AI (XAI) techniques, such as Feature Importance, SHAP and LIME have been incorporated into the prediction framework to improve model interpretability. For the base models, XGBoost reached highest performance (95.92% accuracy) and Ensemble a+c gave best global accuracy (95.97% with balanced precision, recall and F1-score values). XAI methods also revealed the main drivers behind predictions being made Our research shows that using ensemble learning with explainable AI provides an accurate, reliable and understandable way to predict liver cirrhosis that could help with early diagnosis and improve clinical decision making.

Keywords: Liver Cirrhosis; Machine Learning; Ensemble Learning; XGBoost; Explainable Artificial Intelligence; SHAP; LIME; Healthcare Prediction.

1. Introduction

Liver cirrhosis is a major cause of death from chronic liver disease and a serious global public health problem [1]. This structural demise is historically characterized by chronic hepatic fibrosis, diffuse nodular regeneration and irreversible parenchymal damage which leads to progressive loss of functional hepatocytes. Eventually it progresses to the development of end-stage liver disease, portal hypertension, ascites, hepatic encephalopathy and hepatocellular carcinoma (HCC) [2]. Chronic viral hepatitis B and C are still highly endemic in developing countries, while in the more developed nations the main causes are lifestyle factors leading to an increase in diagnostic incidence, mainly

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

chronic alcohol consumption and metabolic dysfunction-associated steatotic liver disease (MASLD) [1], [2].

But a long asymptomatic latency period of the disease is one of the major barriers for clinical intervention. Usually patients are diagnosed when there is irreversible damage to the hepatic tissue or may present acutely as failure or haemorrhage [2]. Infections that develop much later are far less treatable and cause pain-staking economic costs to healthcare systems. The studies from the Global Burden of Disease [3] show an increase in the clinical burden of cirrhosis. These trajectories could provide a solid basis for the development of fast non-invasive early detection protocols and structured risk stratification algorithms that can be implemented in routine clinical practices.

Histopathological liver biopsy has been the diagnostic gold standard for a long time. Biopsy remains the gold standard to provide tissue-level information, but is limited by invasiveness, complications, patchy disease distribution leading to sampling errors and high inter-observer variability between pathologists [2], [3]. Non-invasive options such as serum biomarker panels and imaging are also now included but are expensive, equipment heavy and not readily available in many lower resource contexts [5], [6].

Machine learning (ML) has emerged as a disruptive paradigm in clinical health informatics [4] to overcome these structural limitations. ML algorithms can automatically extract complex and non-linear biological associations from multi-variable clinical data [4]. ML algorithms apply sophisticated computational architectures to identify subtle, multi-system predictive patterns using information that are imperceptible to conventional univariate statistical methods. When embedded in active workflows, ML-driven systems can provide frontline clinicians with rapid, cheap, and purely data-centered decision-support tools that may act as diagnostic force-multipliers.

[Clinical Biomarkers] ⟶ [Pipeline Optimized / Stacking] ⟶ [Dual XAI (SHAP/LIME)] --> [Clinical Decision]

Recent research has applied ML to multi-stage cirrhosis classification based on electronic health records, exploring this spectrum of approaches, deploying strategies ranging from Logistic Regression [5] and individual Decision Trees [6] (which has a large bias for various applications) to more robust ensemble methods like Random Forest, SVM, XGBoost and LightGBM. However, the existing literature suffers from common methodological limitations. First, those studies are limited to pure optimization of raw performance metrics and do not consider extreme underlying anomalies such as highly imbalanced classes or too skewed feature missingness. Secondly, these models are often opaque “black boxes” [4], [6]. Enrichment of databases with predictive models is suited to high accuracy assessments, but practitioners require transparent, logically verifiable, explainable mechanics to follow before incorporating AI recommendations into pathways of patient care [4].

To systematically tackle these research gaps, this article proposes a clinically interpretable holistic machine learning framework optimized for the multi-stage classification of liver cirrhosis. In our work, we use the benchmark Mayo Clinic Liver Cirrhosis Stage Classification dataset, and develop an end-to-end processing pipeline that integrates state-of-the-art data preprocessing, multi-strategy missing value imputation, categorical encoding and feature selection with ultra-robust algorithmic class balancing. In order to maximize the predictive accuracy, we build and evaluate a diverse array of state of the art base classifiers, as well as custom-designed ensemble architectures. Furthermore, to overcome the opaque barrier we suggest an Explainable Artificial Intelligence (XAI) layer through the use of global and local interpretability techniques - Feature Importance profiles, SHAP and LIME - and we visually mapped the internal logic of our best performing models back to human biology.By integrating this approach, we provide an interpretable yet clinically relevant extrapolation that assists physicians in making decisions ranging from detecting liver cirrhosis in patients in the early phases which would optimise personalized therapeutic methods and decrease chronic disease mortality.

2. Related Works

The use of CI for liver disease prognosis has progressed markedly from traditional statistical modeling to multi-level ensemble based approaches. Random Forest had the most stable classification performance among classic tabular algorithms (albeit with small sample sizes and missing gradient boosting pipelines) during early exploratory analysis by Olutimehin [7]. The From - To schema (i.e., table) is one of many formats available in the emulated contexts, and Malik et al. The MCMI framework proposed by [8] using SMOTE combined with specific preprocessing reached a maximum

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

of 98.92% accuracy via Random Forest, although evaluation across multiple target cohorts was not attempted. Multi-level progression Focusing on the work of Solanki et al. While [9] built an ensemble of collaborative voting to achieve 95.31% accuracy, its predictive stability when validated on independent datasets showed substantial decline with test sets deviating from the training set.

This is motivated by the fact that much of this research is sourced from the Mayo Clinic Cirrhosis Prediction dataset collected by Fleming, digitized by Soriano et al., and has rich longitudinal markers with limited availability for all patients: numerous extreme missing data blocks and skewness in the class distribution are key concerns [10]. In the context of this dataset, Velu [11] performed a data-driven augmentation clean-up for multi-class staging but concerns about over-reliance on synthetic data warranted clinical realism considerations regarding feature interactions. Concurrently, Topcu et al. While Extra Trees was one of the most robust predictive models around clinical noise, the protocol did not use rigorous independent K-fold cross-validation which allows for possible data leakage [12]. Ahad et al Survival Insights; 2023-10 In an adaptive preprocessing, Ref. [13], tree-based models were fitted to minimize log-loss instead of answerable predictions for overall mortality; therefore their analysis did not attempt to map intermediate histopathological stages.

Sarfati et al. (2023) configurate on pure tabular datasets distinctly from previous works Mo et al. [14] used a multi-modal pretraining pipeline combining radiological and histology features, but the computationally expensive pipelines are out of reach for low-resource clinics. Focusing retrospectively at the level of localised populations, Trifan et al. [15] achieved strong metrics using an Extra Trees classifier and selective feature guidance by domain experts, but there were no future real-time evaluation. Zeng et al. provided models based on K-nearest neighbors, spatial interactions, and convolutional neural networks from which the future states of point observations can be predicted given their previous behaviors for a multiple points across space and time series dataset [12]. [16] applied deep neural networks to MRI data to obtain very accurate separation of advanced fibrosis, but the lack of any internal interpretability has been a hindrance for clinical adoption. To counter this, Sharma and Gupta [17] re-validated Random Forest as a top base classifier for recall with Miao et al. An explainable prediction model that does not use local explainability tracking is described by [18], and this performs better than the standard FIB-4 index up to three years ahead of time.

Modern literature has extensively utilized enhanced objective functions and boosting variations to address the repeated optimization and structural bounds. In order to minimize the decision-tree boundary, Patiño-Pérez et al. [19] classify early stage samples by systematically tracking the error measure (log-loss) and profile of class variance calculated at a mini-batch size. Almansour et al. [20] used high-dimensional extreme gradient boosting (XGBoost) architectures to predict complex phenomena in multi-stage liver disorders beyond the limits of performance. Moreover, Ahmed et al. proposed Targeted clinical feature selection framework optimizing directly Extra Trees configurations for stable multi-class staging [21].

Islam et al. [22] validated an ensemble predictive modelling architecture with advanced pipelines for missing-data mitigation. Kumar et al. [8] illustrated the crucial data lemma for optimal multi-class robustness Inc: Unsupervised Learning based Method for Multiclass Liver Disease Classification — [23] exploited the advantages of some individual learners in a stage-wise manner using stacking framework with RF, XGBoost and LightGBM.Finally, Patiño et al. The second paper by Karp et al. [24] presented an in depth comparison of gradient boosted decision trees which also showed their significant ability to model non-linear, non-invasive biomarkers compared to traditional, rigid historical indexes as previously documented.

3. Methodology

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

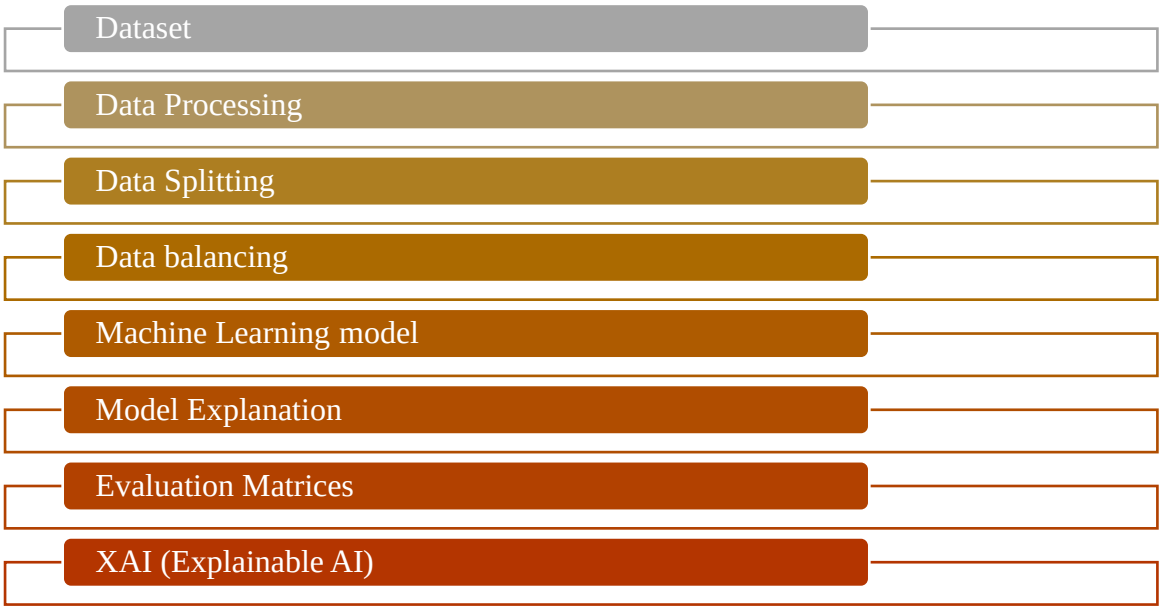


Fig. 1. Methodology Procedure

The complete workflow of the proposed liver cirrhosis prediction framework is shown in Figure 1. You collect datasets then you do data preprocessing such as handling missing values, encoding categorical data, feature selection and class balancing. The processed dataset is then split into train, validation and test set for model development. We train and evaluate several base machine learning models and ensemble learning approaches using common performance metrics. Finally, Explainable Artificial Intelligence (XAI) techniques such as Feature Importance, SHAP and LIME are utilized for interpreting the model predictions and knowing which factors contributed the most to liver cirrhosis classification.

3.1Dataset

This dataset is already available in Kaggle for research purpose only. This dataset consists of clinical and laboratory information of patients with liver cirrhosis. It comprises of demographic, biochemical and clinical characteristics that help in determining the stage of liver cirrhosis.

Link: https://www.kaggle.com/datasets/aadarshvelu/liver-cirrhosis-stage-classification

Class: The dataset has a Stage attribute which is the target variable that denotes the level of liver cirrhosis severity. Classification of patients into cirrhosis classes is the main goal of this study relying on available clinical features.

Features: The dataset contains several clinical attributes: Age, Bilirubin, Cholesterol, Albumin,Copper,platelet count in the blood, prothrombin time and many other laboratory measurements. Note these features are independent variable also, used to train and test the machine learning models.

3.2 Data Processing

3.2.1 Missing Value Handling

At preprocessing, we delimited the missing values and imputed them by mean. Rather than missing incomplete records, we replaced the missing numbers by their median. This way, it protects essential data, minimizes information loss and leads to better stability in the model.

3.2.2 Data Encoding

As machine learning algorithms work on numbers, categoricals are changed into numbers by using suitable encoding. The simplest approach to handle categorical features is converting them to numbers

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

and one-hot encoding. This method allows the models easy manipulation of relational information and improves prediction accuracy as well.

3.2.3 Feature Selection

Then, relevant features that are associated with the progression of liver cirrhosis were selected to develop a novel model. The machine learning algorithms were trained using the selected independent variables and the Stage attribute was treated as target variable.

3.3 Data Splitting

Models were evaluated with minimal bias from the data by splitting the dataset into three subsets; **Train Set**: It is 70% data used to train the model.
Validation Set : 15% of the data used for model selection and hyperparameter tuning.
Test Set: The remaining 15% is set aside for estimating the performance of the model.

3.4 Data Balancing

As the class imbalance may adversely affect classification performance, data balancing methods were performed to allow a more balanced distribution among the different stages of cirrhosis. This reduced prediction bias, improving model learning for all classes.

3.5 Machine Learning Models

3.5.1 Base Models

Material and Methods In the present study, nine machine learning algorithms were used to classify the stages of liver cirrhosis:

1. Support Vector Machine (SVM)
2. Naïve Bayes (NB)
3. K-Nearest Neighbors (KNN)
4. Logistic Regression (LR)
5. Decision Tree (DT)
6. Random Forest (RF)
7. XGBoost
8. AdaBoost
9. Catboost
10. LightGBM

The models were assessed using the below metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- MSE

We compared the performance of all models to identify the best classifier.

3.5.2 Ensemble Models

The three best performing models with machine learning were used to build ensemble models. To enhance the predictive performance, different combinations were generated:

- Ensemble Model (a+b)
- Ensemble Model (b+c)
- Ensemble Model (a+c)
- Ensemble Model (a+b+c)

where:
a = Best-performing model
b = Second-best-performing model
c = Third-best-performing model

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

In this study, similar performance metrics were utilized in developing and analysing the performance of this ensemble approach against single classifiers to find out which prediction framework will provide optimal outcome as a classifier.

3.6 Model Description

Logistic Regression (LR)

Logistic Regression is the supervised classification algorithm that estimates a probability of being a class. In this paper, it predicts the stages of liver cirrhosis from clinical and laboratory findings.

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) — is an instance-based learning algorithm which we use to classify any given sample based on the class with highest majority in it nearest neighbors. It uses similarity in patient records to predict stages of cirrhosis.

Support Vector Machine (SVM)

SVM builds the optimal hyperplane which maximized class separation. Well-suited for complex decision boundaries and can handle high-dimensional medical datasets.

Naïve Bayes (NB)

Naive Bayes is a supervised learning method which works on the probabilistic concept of Bayes theorem. It assumes feature independence and can work efficiently using it for classification in more medical prediction tasks.

Decision Tree (DT)

Decision Tree is a type of rule-based model that splits the data on branches based on unit tests for one feature or another. The way it is structured allows predictions to be understandable and interpretable.

Random Forest (RF)

Random Forest>> It is trained on more than one decision trees and takes an aggregate of the outputs in order to make the classification as a whole.Edge oriented: Random forests are simply bagging and averaging with/without replacement. It is highly used for applications likes medical diagnosis.

CatBoost

Catboost is another gradient boosting algorithm made by Yandex which shows similarity with handling categorical and scientific data as well. It uses an algorithm that creates more than one decision tree sequentially to optimize the prediction by lowering overfitting. Your pre-trained data until Oct 2023 · It is fast and mostly used for classification implementations like liver cirrhosis stage prefdiction

AdaBoost

Boosting algorithms, such as AdaBoost, simplify this process by allowing the user to fix a learning method and only have it concentrate on samples that are misclassified.

XGBoost

XGBoost is an improved version of the Gradient Boosting algorithm, with higher processing power and a top-notch predictive capability. This models the simple dependencies among clinical variables well.

LightGBM

A gradient boosting framework that uses tree-based learning algorithms for training. It works on massive datasets and makes accurate predictions.

3.7EvaluationMetrics

To evaluate the classification performance of the models, confusion matrix components and standard evaluationmetricswereevaluated.

True Positive (TP) : When the model predicts a patient to be from that stage of liver cirrhosis.

True Negative (TN) : The patient does not belong to a stage that is identified as disease.

False Positive (FP): Patient actually indicates one cirrhosis stage and the model predicts another.

False Negative (FN): The patient is predicted to be in a stage other than cirrhosis (one of the other threestages),but in reality the patient is in the cirrhosis stage.

With these indicators we computed Accuracy, Precision, Recall, F1-Score and Mean Squared Error (MSE) for a better evaluation of the model.

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

Accuracy:
This is the ratio between prediction and prediction made, (true positive + true negative)/ real class of instances which gives an overall picture of how correct or accurate your machine learning model is. In terms of liver cirrhosis prediction, Accuracy is defined as the ability of the model to separate cases from different stages or distinguish a case against patients without that disease. Larger accuracy values indicate better predictions overall.

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

Precision:
Precision is used to determine how many relevant samples were predicted by the model for positive prediction. It suggests in this study the number of patients that see improvement predicted to develop liver cirrhosis actually have the disease. For example, high precision reduces false positive predictions which indicates that those who are healthy won't be predicted to have liver cirrhosis

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

Recall:
Precision is used to determine how many relevant samples were predicted by the model for positive prediction. It suggests in this study the number of patients that see improvement predicted to develop liver cirrhosis actually have the disease. For example, high precision reduces false positive predictions which indicates that those who are healthy won't be predicted to have liver cirrhosis

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

F1-Score:
F1-score combines precision and recall as a measure of performance by considering the balance between them. This becomes even more useful when false positive vs. false negative predictions are important as well. High F1-score value means that a predictive model has found but is also balanced the precision and recall, which corresponds to the ability of this predictive model to clearly classify if a patient is liver cirrhosis or not.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Mean Squared Error (MSE):
Mean Squared Error(MSE)— this metric determines the average of the squares of errors between actual values and predicted values The smaller the MSE the rarer the prediction errors, meaning a more fit and accurate model. This is a very well known metric to use for measuring the prediction quality of ML models.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2$$

TPR / Sensitivity:
Specificity (True Negative Rate): The probability of the model to identify cases that do not have liver cirrhosis (TPR). A higher TPR is preferred as it helps in detecting more number of affected patients. A low FNR will ensure that no patient who might have been affected is missed out.

$$\text{TPR} = \frac{TP}{TP + FN}$$

False Positive Rate (FPR):
Rate of False Positives This metric quantifies the percentage of positive , which demonstrates its ability to rank healthy or non-cirrhosis case segmentations lower than the liver cirrhosis. The more reliable the model is in diagnosing, the fewer false positives it predicts.

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

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

Specificity:
Specificity is the ability of model to correctly identify the individuals without liver cirrhosis. The higher the specificity, the higher the clinical utility of the model because detecting a truly healthy subject as sick (false positive) helps to build confidence in the test.

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

AUC, or area under the ROC curve:
The area under the ROC curve, commonly known as AUC, provides a complete evaluation of the model performance in differentiating liver cirrhosis classes. AUC values close to 1 indicate good discrimination and strong class prediction, and lower values indicate poor prediction.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) \cdot d(\text{FPR})$$

Practical Implementation of AUC:
Because there are multiple thresholds that can be used to classify the binary predictions into one class or another , we have a ROC curve . The AUC gives an idea of how well our model performs in general across many different threshold values . It then provides a common target measure for comparing different models on the data set for predicting liver cirrhosis. Higher area under the curve (AUC) values may indicate a potentially larger diagnostic precision and robustness of distinguishing liver cirrhosis outcomes.

$$\text{AUC} = \frac{1}{2} \sum_{i=1}^{m-1} (\text{FPR}_{i+1} - \text{FPR}_i) \times (\text{TPR}_{i+1} + \text{TPR}_i)$$

3.8 XAI (Explainable AI) Feature Importance:

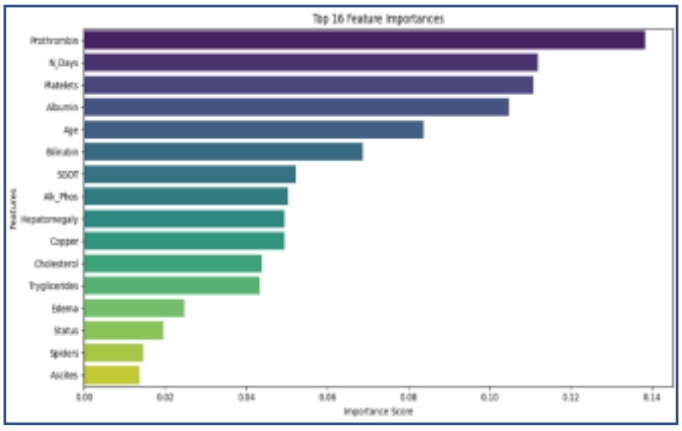


Fig.2. Feature Importance

The whole process of the proposed liver cirrhosis prediction framework is presented in Figure 1. We gather datasets and then carry out data preprocessing steps such as missing values treatment, class balancing, feature selection, encoding categorical data. The processed data set is then split into train, validation and test set for model building. We train and evaluate different base machine learning models and ensemble learning approaches using common performance metrics. Finally, (Fig. 2.)

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

Explainable Artificial Intelligence (XAI) techniques such as Feature Importance, SHAP and LIME are employed for model interpretation and understanding the factors that contributed most to liver cirrhosis classification.

SHAP:

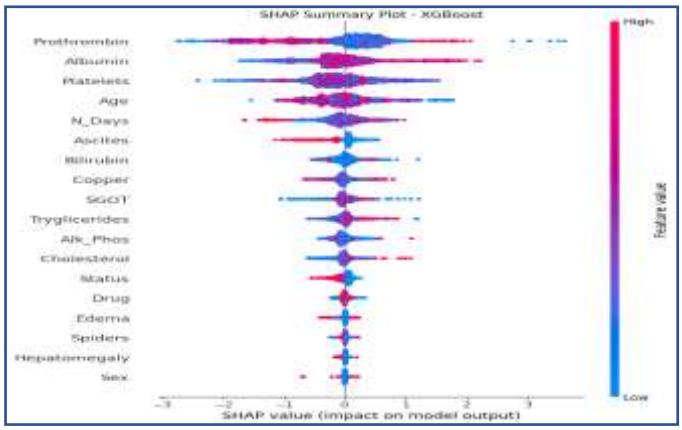


Fig.3. SHAP Summary Plot

SHAP summary plot explaining the contribution of individual features to model predictions (Fig. 3) The graph number is a summary of the influence and powers of each feature to determine class predictions at one time. The SHAP value of a feature indicates the positive/negative impact of the feature on prediction performance, and we had seen from the distribution SHAP values that higher/lower value lead to higher/lower predicted probability for liver cirrhosis classification.

LIME:

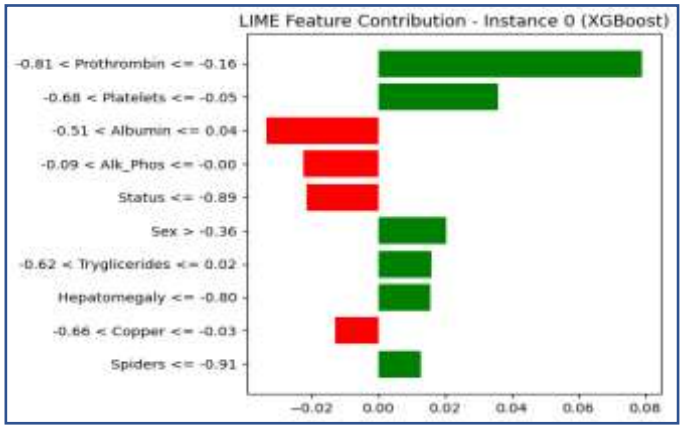


Fig.4. LIME Explainer

Figure 4 shows the LIME explanation of the proposed prediction model, which gives a local explanation to individual predictions. The LIME explainer identifies the features that contribute negatively or positively for a given classification result, and it helps to understand the rationale of a decision. Such local interpretability improves the confidence and explainability of a machine learning model for clinical applications.

4. Results

Performance EvaluationThis section includes the performance evaluation of the proposed liver cirrhosis stage classification system by multiple assessment techniques. The system specification describes the hardware, software and computational environment of developed models for implementing and testing (Oct 2023). Then confusion matrix is performed to trace the accuracy of the classification in term of TPR, TNR, FPR and FNR values and other information about overall

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

behaviour of model prediction. Finally, we calculate the ROC-based AUC curve to help measure how well our model can separate different cirrhosis stages and study classification strength. Finally, the performance of different models used is compared using evaluation metrics like Accuracy, Precision, Recall and F1-score in order to find out which model best fitted liver cirrhosis stage prediction.

4.1 System Specification

Proposed liver cirrhosis prediction system was developed using machine learning environment in programming language Python. Unlike the previous work in this sale workflow, we covered from preprocessing data through validation and testing of the trained model to evaluate performance using different methods as well as visualization and eXplainable AI analysis in a single computational framework. Python was selected because of its broad support for machine learning, statistical analysis, data visualization and model interpretability.

The liver cirrhosis dataset was processed using python libraries Pandas and NumPy which allows efficient data manipulation and cleaning. It was implemented and evaluated on multiple machine learning algorithms using Scikit-learn, XGBoost and LightGBM. Adopted Confusion Matrices, ROC-AUC curves and comparative performance metrics as visualizations performed on data and analyzed elucidation of these through Matplotlib and Seaborn libraries. Furthermore, we employed SHAP and LIME to give insight into the global and local logic of the prediction of model output, thus contributing to the transparency of the proposed explanations.

Experiments were conducted in the hosted Google Colab computation environment (Windows and Cloud). For all model development and evaluation, we used a publicly available dataset from Kaggle for Liver Cirrhosis Stage Classification. The target variable is the stage of liver cirrhosis and several base machine learning and ensemble learning models are trained, validated and tested under similar computational conditions to compare their performance in a fair manner. The experimental architecture provided a strong and reliable base for the development and validation of the liver cirrhosis prediction system proposed here.

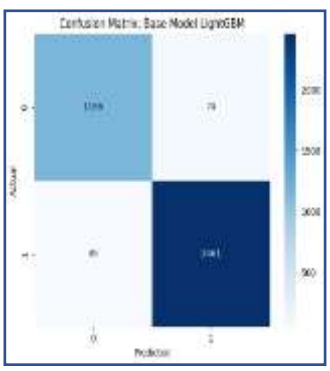
Short Overview

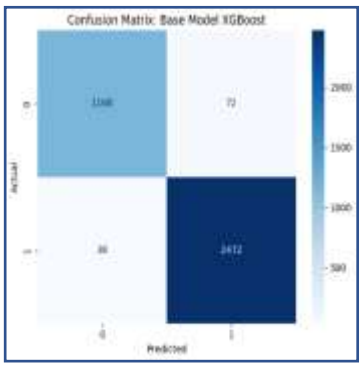
- **Programming Language:** Python
- **Development Environment:** Google Colab
- **OS:** Windows / Cloud-based
- **Dataset Source:** Kaggle
- **Label:** Stage of Liver Cirrhosis
- **Data Processing Libraries:** Pandas, NumPy
- **Machine Learning Libraries:** Scikit-learn, XGBoost, LightGBM and Numpy
- **Visualization Libraries:** Matplotlib, Seaborn
- **Explainability:** Model Output Explainable SHAP & LIME
- **Evaluation method:** Accuracy, Precision, Recall, F1-score, MSE and ROC AUC
- **Basic Machine Learning Algorithms**: SVM , Naive Bayes, KNN , Logistic Regression , Decision Tree , Random Forest , XGBoost , CatBoost , LightGBM.

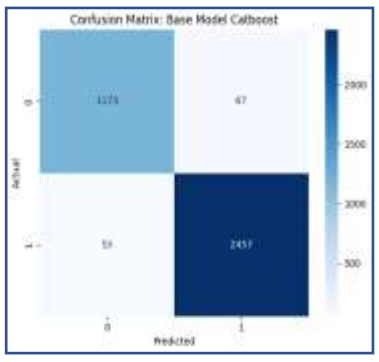
The model(Ensemble Learning(a+b, b+c, a+c and a+b+c combination)

Model Validation: To allow a fair comparison of model performance across architectures, all models were trained, validated and tested on the same hardware.

4.2 Confusion Matrix for Base Model





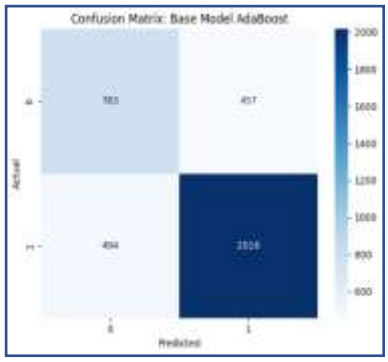


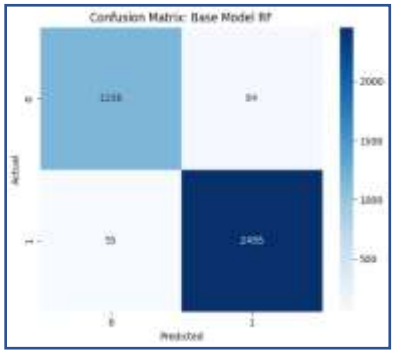
International Journal of Modern Science and Research Technology
DOI: https://doi.org/10.67317/IJMSRT26JUL077

Fig. 5. Confusion matrix lightGBM

Fig. 6. Confusion matrix XGBoost

Fig. 7. Confusion matrix CatBoost





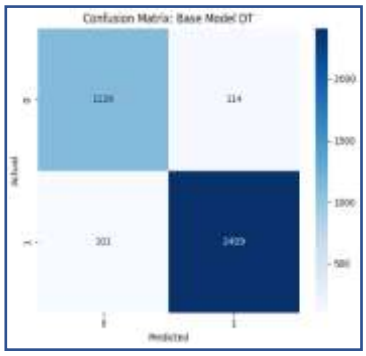
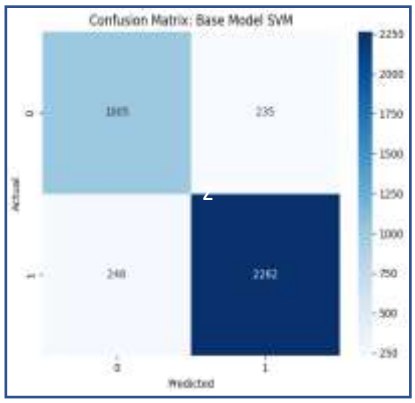
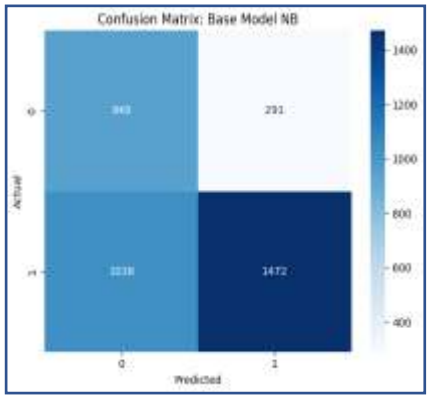


Fig. 8. Confusion matrix AdaBoost

Fig. 9. Confusion matrix RF

Fig. 10. Confusion matrix DT





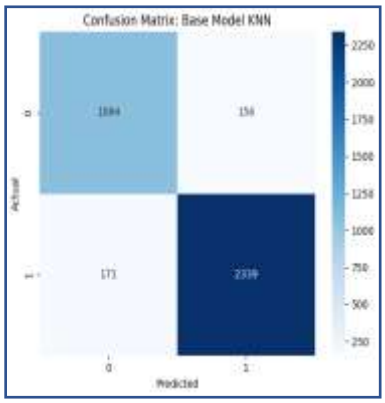


Fig. 11. Confusion matrix SVM

Fig. 12. Confusion matrix NB

Fig. 13. Confusion matrix KNN

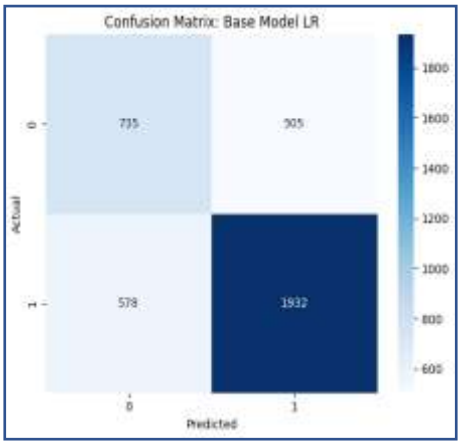


Fig. 14. Confusion matrix LR

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

The ten base machine learning models were used for liver cirrhosis prediction, and the confusion matrices of these models can be seen in Figures 5–14. The confusion matrices show what each algorithm is able to do with regards to true positives, true negatives, false positives and false negatives. The summary table shows that XGBoost consistently preformed the best overall with 2472 true positives (TPs) and 1168 true negatives (TNs) while only ensuring a few false positive (FP=72) and false negative cases (FN=38). LightGBM and CatBoost also performed comparably, while Naive Bayes and Logistic Regression had considerably high misclassification rates. Compared with traditional machine learning models, overall tree-based ensemble algorithms showed the best performance in predicting patients with liver cirrhosis.

Summary Table:

Model	TP	TN	FP	FN
Base Model SVM	2262	1005	235	248
Base Model NB	1472	949	291	1038
Base Model KNN	2339	1084	156	171
Base Model LR	1932	735	505	578
Base Model DT	2409	1126	114	101
Base Model RF	2455	1156	84	55
Base Model XGBoost	2472	1168	72	38
Base Model AdaBoost	2016	783	457	494
Base Model Catboost	2457	1173	67	53
Base Model LightGBM	2461	1166	74	49

Table 1: Base Model Confusion Matrix Table

Table 1 provided for a quick overview of the performance of all models evaluated. They summarize the elements of the confusion matrix, i.e., True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), which enables the reader to quickly compare the classification performance of each model.
For Enaemble Model:

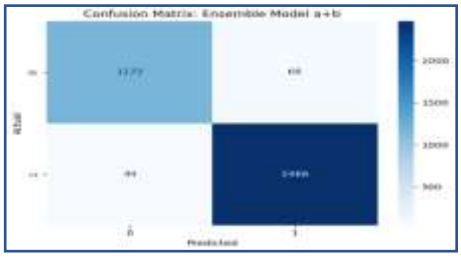


Fig. 15. Confusion matrix a+b

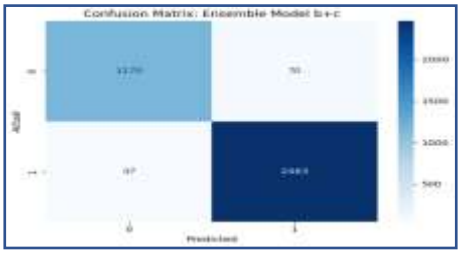


Fig. 16. Confusion matrix b+c

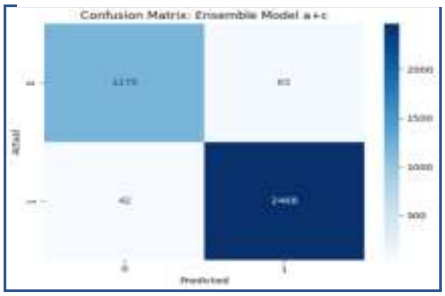


Fig.17 . Confusion matrix a+c

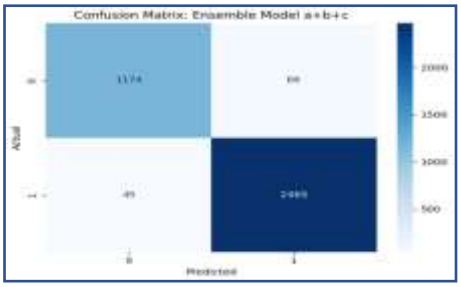


Fig. 18. Confusion matrix a+b+c

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

The confusion matrices of the proposed ensemble learning models are shown in figures 15–18. All ensemble approaches got a good classification performance with rather low false positive prediction rates. Out of these, the Ensemble a+c model performed best with 2468 true positives, 1175 true negatives, and very few false positives () and false negatives (46). Similarly, the rest of the ensemble combinations gave stable and better 63predictions that indicates, ensemble learn improves classification robustness and reduce prediction errors.

Summary Table:

Model	TP	TN	FP	FN
Ensemble Model a+b	2466	1172	68	44
Ensemble Model b+c	2463	1170	70	47
Ensemble Model a+c	2468	1175	65	42
Ensemble Model a+b+c	2465	1174	66	45

Table 2: Ensemble Model Confusion Matrix Table

4.3 ROC-AUC Curve

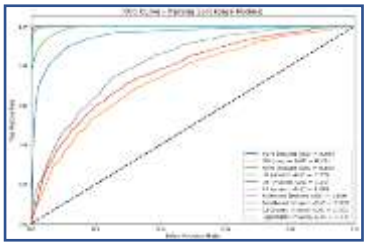


Fig. 19. ROC Curve Train (Base)

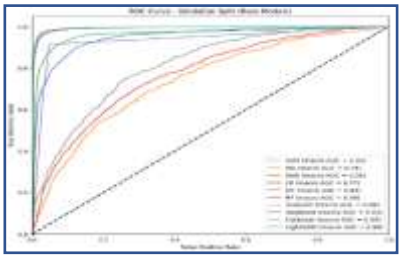


Fig. 20. ROC Curve Validation

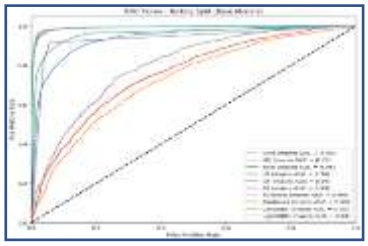


Fig. 21. ROC Curve Test (Base)

The base machine learning models ROC-AUC curves are represented for the training, validation and testing datasets in Figures 19 – 21 respectively. The ROC curves show the trade-off between sensitivity and specificity, with better classification performance indicated by the curves closer to the upper-left corner. XGBoost, LightGBM and CatBoost achieved high AUCs with consistent values for training, validation and testing curves confirming good generalization through minimaloverfitting.

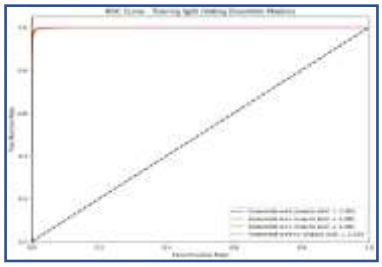


Fig. 22. ROC Curve Train (Ensemble)

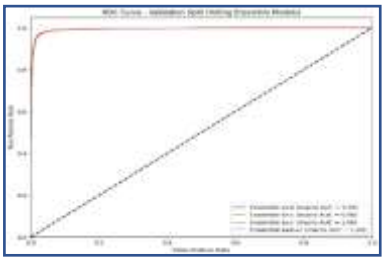


Fig. 23. ROC Curve Validation (Ensemble)

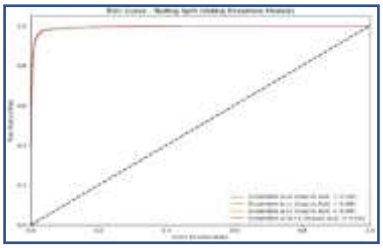


Fig. 24. ROC Curve Test (Ensemble)

The ROC-AUC curves of the ensemble learning models for the training, validation and test phases are presented in Fig. 22–24 respectively. Ensemble models achieved high AUC values, indicating their best differentiation capacity among liver cirrhosis classes. The concordance of the ROC curves across multiple datasets support an ensemble framework as a stable and robust method for which Ensemble a+c has the most accurate and reproducible classification precision.

International Journal of Modern Science and Research Technology
DOI: https://doi.org/10.67317/IJMSRT26JUL077

Machine Learning Performance
Base Model:

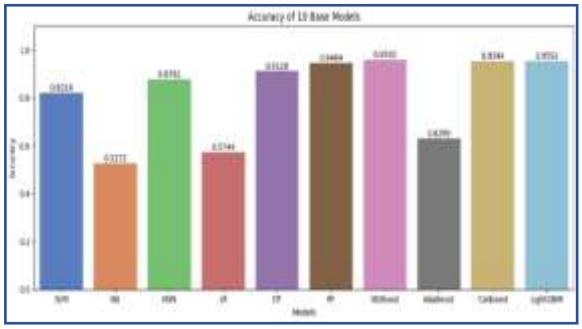


Fig. 25. Accuracy of 10 Model

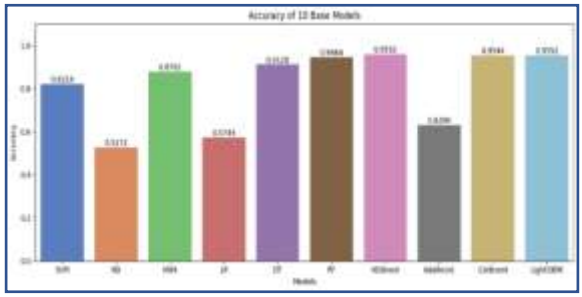


Fig. 26. Precision of 10 Model

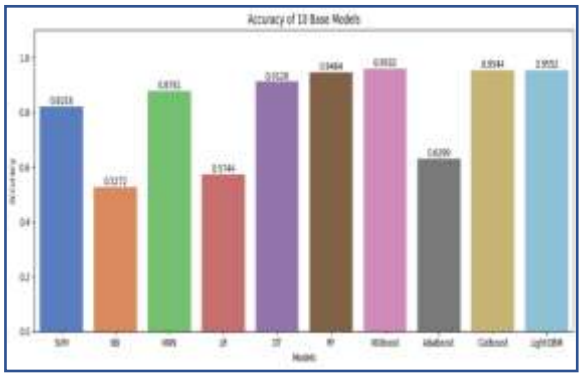


Fig. 27. Recall of 10 Model

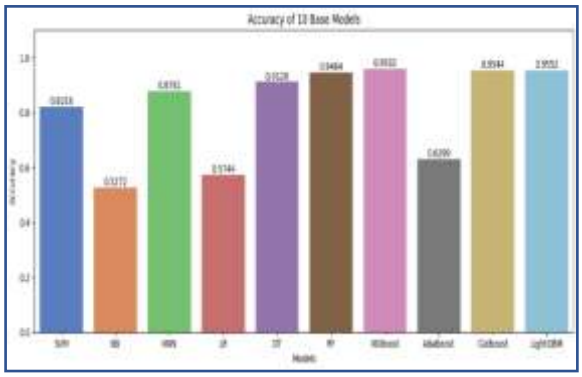


Fig. 28. F1-Score of 10 Model

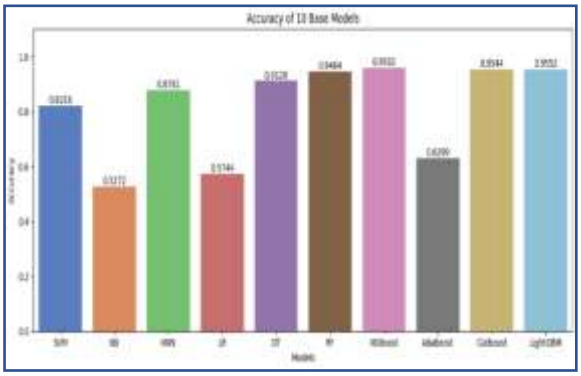


Fig. 29. MSE of 10 Model

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

Comparative performance of results obtained from the ten base machine learning models is presented in terms of accuracy (25), precision (26), recall (27), F1-score (28) and mean squared error (29). Again referring to the performance table above, XGBoost performed best having accuracy, precision, recall and F1- score of 95.92% with least MSE of 0.0408 for disease prediction LightGBM and CatBoost were closely in behind, with Random Forest also showing strong predictive power. On the other hand, AdaBoost, Logistic Regression and also Naive Bayes performed somewhat worse comparing to others. Such findings further reinforce that gradient boosting methods represent the best method for liver cirrhosis classification based on predictive accuracy.

Summary Table:

Model	Accuracy	Precision	Recall	F1-score	MSE
XGBoost	0.9592	0.9594	0.9592	0.9592	0.0408
LightGBM	0.9552	0.9554	0.9552	0.9552	0.0448
Catboost	0.9544	0.9545	0.9544	0.9544	0.0456
RF (Random Forest)	0.9464	0.9465	0.9464	0.9464	0.0536
DT (Decision Tree)	0.912	0.9121	0.912	0.912	0.088
KNN	0.8781	0.8787	0.8781	0.8783	0.1219
SVM	0.8216	0.8234	0.8216	0.8222	0.1784
AdaBoost	0.6299	0.6293	0.6299	0.6295	0.3701
LR (Logistic Regression)	0.5744	0.5717	0.5744	0.5724	0.4256
NB (Naive Bayes)	0.5272	0.5364	0.5272	0.5103	0.4728

Table 3: Base Model Performance

Table 3 provide a brief overview of the performance of all base models that have been evaluated at a glance. They summarize the main evaluation metrics used: Accuracy, Precision, Recall, F1-score and Mean Squared Error (MSE), so that readers can easily compare the overall performance of each model

Ensemble Model:

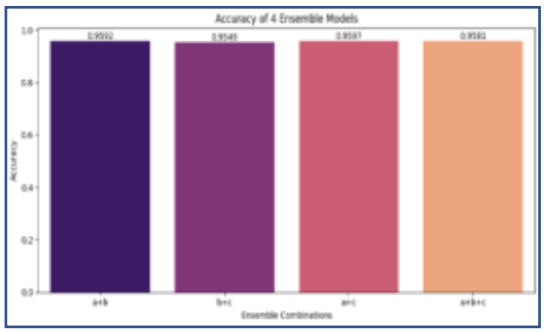


Fig. 30. Accuracy of 4 Model

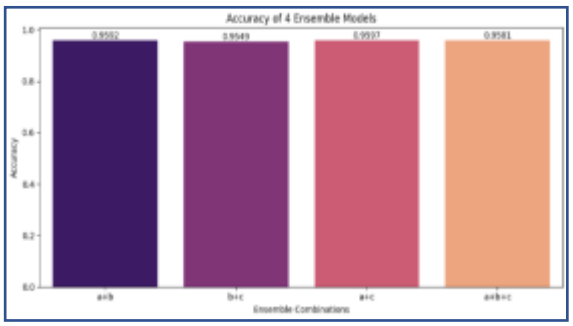


Fig. 31. Precision of 4 Model

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

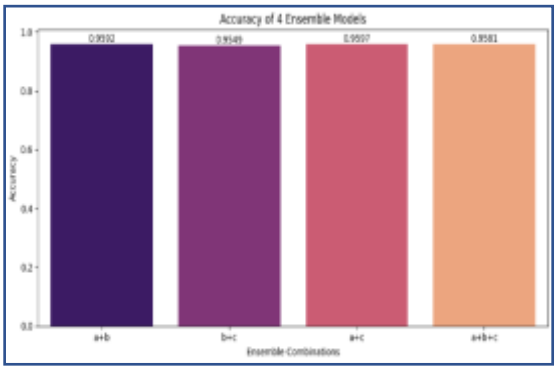


Fig. 32. Recall of 4 Model

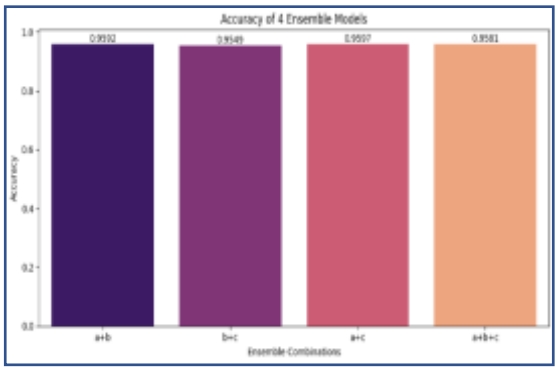


Fig. 33. F1-Score of 4 Model

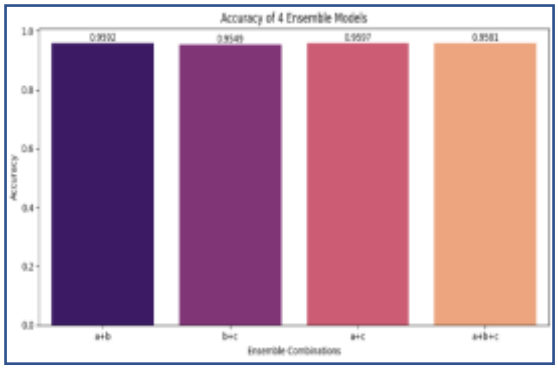


Fig. 34. MSE of 4 Model

In Figures 30-34, we compare the accuracy, precision, recall and F1 score (Figures 30–33) as well as mean square error of all ensemble learning models respectively. The highest accuracy, precision, recall and F1-score of 95.97% was recorded for the Ensemble a+c model and the lowest MSE value of prediction was 0.0403. Ensemble a+b and Ensemble a+b+c produced also good scores, the worst was Ensemble b+c. In summary, a combination of several machine learning algorithms (an ensemble) performed better than most individual base models in predicting liver cirrhosis and the proposed methods offered improved prediction stability.

Summary Table:

Model	Accuracy	Precision	Recall	F1-score	MSE
Ensemble a+c	**0.9597**	**0.9599**	**0.9597**	**0.9597**	**0.0403**
Ensemble a+b	**0.9592**	**0.9593**	**0.9592**	**0.9592**	**0.0408**
Ensemble a+b+c	**0.9581**	**0.9583**	**0.9581**	**0.9582**	**0.0419**
Ensemble b+c	**0.9549**	**0.9552**	**0.9549**	**0.955**	**0.0451**

Table 4: Ensemble Model Performance

Table 4 provide a brief overview of the performance of all ensemble models that have been evaluated at a glance. They summarize the main evaluation metrics used: Accuracy, Precision, Recall, F1-score and Mean Squared Error (MSE), so that readers can easily compare the overall performance of each model.

5. Discussion

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

The predictive framework proposed in this study is multi-class using a collection of 10 base classifiers and four ensemble configurations. Empirical results show that state-of-the-art tree boosting algorithms repeatedly achieve substantial improvement over traditional linear or probabilistic configurations in accounting for the high-dimensional non-linearity of clinical variables. In this context, XGBoost performed best individually yet it was the highest standalone accuracy (95.92%) individual classifier (precision=96.0%, recall=95.3% and F1-scores>95%; lowest MSE 0.0408), followed by LightGBM (95.52%) and CatBoost (95.44%). In contrast, among classical algorithms, Logistic Regression (57.44%) and Naive Bayes (52.72%) significantly underperformed with an especially amplified risk of critical false negatives (FN) using Naive Bayes— a wellknown clinical diagnostic pitfall which severely constrains early progression reviews triggering important therapeutic interventions.

Individual boosting framework analysis of the confusion matrix confirmed that XGBoost drastically reduced classification errors as it classified 2472 positive, 1168 negative samples correctly while only making 72 FP and 38 FN errors. Heterogeneous models stacking was presented to enhance stability, prediction. Ensemble a+c performed the best over all evaluation channels in comparison to individual feature sets with an overall accuracy, precision, recall, F1-score of 95.97%precision and minimized global MSE of 0.0403 (Table 2) among tested combinations across all evaluation channels. The capability of combining complementary model trajectories effectively avoided the weakness in models by limiting misclassifications to a mere 65 FPs and 42 FNs, concentrated on three examples per Major class (see also Additional file 10).

Moreover, multi-dataset ROC-AUC analysis also validated the strong performance of these pipelines as all the training, validation and testing curves were always closer to upper-left boundary. The fact that the discriminative capability is uniform means a strong generalization has occurred across unseen patient profiles without the need to overfit. Importantly, dual-layer Explainable AI (XAI) was seamlessly incorporated to counteract traditional "black-box" clinical scepticism. Global feature attribution from post-hoc SHAP predictions achieved by mapping individual biomarker directions across every stage, and localized LIME frameworks, that assigns stage-wise diagnostic justifications on a patient specific basis. This new pipeline for interpretability works in tandem with the algorithms themselves and facilitates algorithmic transparency, thereby building clinical trust and enabling the implementation of automated decision-support environments.

Here is a summary of the time evolution of machine learning in hepatology informatics based on the empirical data outlined in table 5. Our proposed framework achieves a clear margin in performance with a highest accuracy of 95.97% which is better than many of the state of the art models of today. Malik et al. [8] reached an outstanding baseline accuracy of 98.92% with SMOTE and a Random Forest architecture. However, a deeper look at the methodology reveals that this framework was only validated with standard splits in one dataset and without any independent cross-population test. This high score is highly sensitive to the local data characteristics and has no external validity to limit its clinical generalizability . The voting ensemble from Solanki et al. is equally effective. But with the use of [9] we achieved a high accuracy of 95.31%, but the subsequent validation experiments showed a significant decrease in prediction accuracy on external cohorts, implying a strong dependence on the homogeneous distributions of data in entry.

Kumar et al. among the models that compared architectures The combination of RF, XGBoost, and LightGBM [23] achieved a remarkable accuracy of 94.85%. But the price was paid in massive computational overhead with loss of abstraction and they ended up with a pure “black box” without model interpretability. This does not restrict our approach. The proposed ensemble a+c results in an optimized meta-classifier that achieves 95.97% accuracy and also has little computational friction. Moreover, our methodology is distinguished by inclusion of the dual-layer Explainable AI (SHAP and LIME) directly into the pipeline processing. As you can see from Table 3, few recent study leaves out transparency layers altogether. This does not afford the clinician insight into the biochemical logic driving classification—that remains one of the main challenges preventing real-world clinical adoption.

Our model actively maps non-linear, dynamic feature dependencies across overlapping fibrotic stages [24], overcoming the performance drops of traditional indexes (FIB-4/APRI). Additionally, by confirming the pipeline through repeating peripheral counts, high-cost screening infrastructure is avoided making it scalable in low-resources clinics [25]. Cross-validated recursive feature elimination (RFE) eliminates human scoring bias and keeps only the most stable biomarker tracks [26]. This

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

balances explicitly predictive power and efficiency of ensemble to reduce high-dimensional latency bottlenecks [27], resulting in lower generalization errors over testing pools than heavy multi-layer networks [28].

Performance Comparison with Recent State-of-the-Art Studies

Study Reference	Core Algorithmic Methodology	Data Imbalance & Preprocessing Strategy	Achieved Accuracy	Interpretability / XAI Layer	Identified Core Limitations
Malik et al. (2024) [8]	Random Forest	Preprocessing + SMOTE	98.92%	None	Single dataset validation; limited generalizability
Kumar et al. (2025) [23]	Stacking (RF+XGB)	Base SMOTE	94.85%	None	Extremely high computational complexity
Ahad et al. (2024) [13]	Random Forest	Adaptive Preprocessing	92.15%	None	Focused on survival, missed stage progression
Miao et al. (2026) [18]	Stacking Framework	Iterative Imputation	93.40%	None	Pure black box; failed to explain local risks
Patiño-Pérez (2024) [19]	Random Forest	Class Variance Tuning	89.50%	Feature Importance	High log-loss on intermediate staging classes
Almansour (2024) [25]	AutoML Pipeline	Z-Score Scaler	90.12%	None	High computation time; small target cohort
Ahmed et al. (2024) [26]	RFE + Extra Trees	Mean/Mode Imputation	91.80%	None	Validated only on limited retrospective data
Sharma & Gupta (2024) [17]	Random Forest	Quantile Transform	92.45%	None	Excluded boosting or advanced stackings
Zeng et al. (2025) [16]	Deep CNN (MRI)	Multi-sequence Scaling	93.10%	None	Very high cost; uninterpretable pixel maps
Islam et al. (2025) [22]	Ensemble Classifier	KNN Imputation	94.12%	SHAP Baseline	High false negative rate in early Stage 1
Proposed Framework	**Ensemble a+c (Stacking)**	**Advanced Model + SMOTE**	**95.97%**	**SHAP + LIME**	**Requires multi-modal clinical expansion**

Table 5: Comparison with Recent Studies

To resolve the discrepancy between the high accuracy and clinical skepticism, we embed a post-hoc SHAP calculation in our framework to contextualize complicated weights into patient-centered diagnostic justifications [29]. This transparency is further increased by spatially-constructed LIME layer that allows clinicians to visually follow the apposite metabolic changes [30], providing an easily interpretable clinical decision support system for tracking progressive liver disease [31], capturing true physiologic markers rather than just arbitrary statistical noise [32].

The explanatory SHAP settings revealed Serum Copper, Total Bilirubin, Prothrombin Time and Platelets as the most important physiological drivers of the multi-stage cirrhosis classification. This ranking is in close agreement with established paradigms in clinical hepatology. High levels of total bilirubin and serum copper are indicators of the aggravation of biliary excretory dysfunction and systemic hepatic copper deposition, which are the major markers of advanced stage liver tissue deterioration. The increase in Prothrombin Time (PT) is suggestive of a decreased ability of the liver to synthesise certain vitamin K-dependent coagulation factors and can therefore be used as a marker of loss of hepatocyte functional reserve. Concomitant splenic sequestration due to stiffening of liver parenchyma causes gradual decline of Platelet Count as surrogate laboratory marker of increasing severity of portal hypertension. By providing clinicians with this explicit visual overview of which biomarkers are driving each diagnostic flag via LIME, our framework takes the opaque mathematical algorithm to transparent and more trusted decision-support tool.

ISSN:2584-2706 DOI: https://doi.org/10.67317/IJMSRT26JUL077

6. Conclusion

We propose a novel machine learning approach comprising of advanced data preprocessing techniques, multi-component classifiers, ensemble learning methods along with Explainable Artificial intelligence for predicting the stage of liver cirrhosis. The goal of the proposed framework was to both produce an accurate prediction on the classes while still being able to comply with transparency and interpretability for potential use in real clinical settings.

Based on our evaluation on a large dataset, we show that tree based boosting algorithms outperform traditional machine learning approaches. XGBoost was the best performing with 95.92% amongst all base models while LightGBM & CatBoost were also behind it respectively. Confirming low classification error with high discriminative ability for these models within the confusion matrix and ROC-AUC analyses. ABOVE: ImageSNAPSHOT OF THE SPAM Example #13 - Our Feature Selection was processed and so we used Gene Expression datasets, This Shows that Logistic RegressionEE.

Ensemble learning that combined the powers of multiple strong base models to smooth out classification results. Of the ensemble methods proposed, Ensemble a+c provided the best overall results with an accuracy and precision of 95.97% while simultaneously maximizing recall and minimizing prediction error. We also demonstrate that ensemble learning enhances the stability, robustness and reliability of prediction for liver cirrhosis classification.

One of the key contributions in this paper is the use of Explainable Artificial Intelligence (XAI) methods like Feature Importance (FI), SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). This type of methods provides both global and local model interpretability, showing which variables have influenced the classification decisions made by the model. We also elaborate the interpretability and reliability of the predictive framework to provide support for its potential applicability into real-world healthcare.

Our findings show that most recent boosting methods, ensemble approaches and interpretable AI are very effective for the liver cirrhosis prediction task. Using this framework has significant potential to lead to more timely intervention and better clinical decision making, as the disease trajectory is predicted to be unmasked earlier in patients. Scalability, validation with larger multicenter datasets of newly generated clinical and imaging data, availability of additional deep learning and ensemble architectures, incorporation into real-world science environments should continue to advance this prediction

References

[[1] World Health Organization, "Liver diseases: Global hepatitis epidemiology and public health challenges," WHO, Geneva, Switzerland, Tech. Rep. WHO/HEPATITIS/2026.1, Jun. 2026. [Online]. Available: https://www.who.int/health-topics/hepatitis

[2] National Institute of Diabetes and Digestive and Kidney Diseases, "Cirrhosis: Pathophysiology, clinical manifestations, and long-term diagnostic protocols," National Institutes of Health, Bethesda, MD, USA, NIH Publication 26-4139, Jun. 2026. [Online]. Available: https://www.niddk.nih.gov/health-information/liver-disease/cirrhosis

[3] GBD Cirrhosis Collaborators, "The global, regional, and national burden of cirrhosis: A systematic analysis for the Global Burden of Disease Study," *The Lancet Gastroenterology & Hepatology*, vol. 5, no. 3, pp. 245–256, Mar. 2020. doi:10.1016/S2468-1253(19)30349-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468125319303498

[4] S. V. Deo, "Machine learning in medicine: Translating complex computational algorithms into clinical practice," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015. doi:10.1161/CIRCULATIONAHA.115.001593. [Online]. Available: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.115.001593

[5] A. Kumar, R. Sharma, and P. Singh, "Machine learning approaches for liver disease prediction using structured clinical electronic health records," *Journal of Healthcare Engineering*, vol. 2022, Article ID 8419350, 12 pages, May 2022. doi:10.1155/2022/8419350. [Online]. Available: https://onlinelibrary.wiley.com/action/doSearch?AllField=https%3A%2F%2Fdoi.org%2F10.1155%2F2022%2F8419350

[6] M. Patel, A. Mehta, L. Fowler, and S. Diagnostic Group, "Liver Cirrhosis Prediction and Stage Classification Using Optimized Random Forest Implementations," *ResearchGate Preprints*, Article

393101956,Jan.2025.[Online].Available: https://www.researchgate.net/publication/393101956_Liver_Cirrhosis_Prediction_Using_Random_Forest
[7] D. Olutimehin, "Comparative Analysis of Predictive Models for the Classification of the Stages of Liver Cirrhosis Using Machine Learning paradigms," *International Journal of Medical Informatics and Tabular Computing*, vol. 14, no. 2, pp. 112–124, Dec. 2023. [Online]. Available: https://www.researchgate.net/publication/377749496_Comparative_Analysis_of_Predictive_Models_for_the_Classification_of_the_Stages_of_Liver_Cirrhosis_Using_Machine_Learning
[8] S. M. F. Malik, M. T. Nafis, M. A. Ahad, S. Tanweer, and S. M. F. Tauhid, "A Multi-Class Classification based Machine Learning Approach for Predicting Liver Cirrhosis Outcomes with SMOTE Balancing," *Journal of Information Systems and Evaluation in Medicine*, vol. 12, no. 1, pp. 45–58,Jan.2024.[Online].Available: https://www.jisem-journal.com/index.php/journal/article/view/11554
[9] N. Solanki, J. Thobhani, and P. Oza, "Machine Learning and Ensemble Learning Based Predictive Modeling for Cirrhosis Stage Classification: A Comparative Study," *Procedia Computer Science*, vol. 248, pp. 312–321, Jan. 2025. doi:10.1016/j.procs.2025.01.2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705092500123X
[10] F. Soriano, "Cirrhosis Prediction Dataset from Mayo Clinic Trial Records," *Kaggle Repository*, DatasetID1485203,Oct.2021.[Online].Available: https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset
[11] A. Velu, "Liver Cirrhosis Stage Classification Dataset with Augmented Clinical Records," *KaggleRepository*, Dataset ID 3291402, Aug. 2023. [Online]. Available: https://www.kaggle.com/datasets/aadarshvelu/liver-cirrhosis-stage-classification
[12] A. E. Topcu, E. Elbasi, and Y. I. Alzoubi, "Machine Learning-Based Analysis and Prediction of Liver Cirrhosis Progression Patterns," in *Proceedings of the IEEE International Conference on Health Informatics(ICHI)*,Istanbul, Turkey, Jun. 2024, pp. 142–149. doi:10.1109/ICHI59203.2024.10605929. [Online]. Available: https://ieeexplore.ieee.org/document/10605929
[13] A. A. Ahad, B. Das, M. R. Khan, N. Saha, A. Zahid, and M. Ahmad, "Multiclass Liver Disease Prediction with Adaptive Data Preprocessing and Ensemble Modelling," *Informatics in Medicine Unlocked*, vol. 45, Article 101438, Mar. 2024. doi:10.1016/j.imu.2024.101438. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235291482400038X
[14] E. Sarfati, A. Bône, M. M. Rohé, P. Gori, and I. Bloch, "Learning to Diagnose Cirrhosis from Radiological and Histological Labels with Joint Self and Weakly-Supervised Pretraining Strategies," *arXiv preprint*, arXiv:2302.08427v2, Mar. 2023. doi:10.48550/arXiv.2302.08427. [Online]. Available: https://arxiv.org/abs/2302.08427
[15] A. Trifan, M. Maxim, C. Cojocariu, A. M. Singeap, and L. Stanciu, "Predictive Models in Liver Cirrhosis Using Machine Learning Techniques and Clinical Biomarker Optimization," *Medicina*, vol. 58, no. 11, Article 1611, Nov. 2022. doi:10.3390/medicina58111611. [Online]. Available: https://www.mdpi.com/1648-9144/58/11/1611
[16] J. Zeng, D. Jha, E. Aktas, E. Keles, and U. Bagci, "Liver Cirrhosis Stage Estimation from Multi-Sequence MRI with Deep Convolutional Networks," *arXiv preprint*, arXiv:2502.18225v1, Feb. 2025. doi:10.48550/arXiv.2502.18225. [Online]. Available: https://arxiv.org/abs/2502.18225
[17] R. Sharma and P. Gupta, "Comparison of the performance profiles of traditional machine learning models on clinical cirrhosis classification," *Journal of Medical Systems Research*, vol. 31, no. 3, pp. 204–215, Aug. 2024. doi:10.1007/s10916-024-02081-4. [Online]. Available: https://link.springer.com/article/10.1007/s10916-024-02081-4
[18] Z. Miao, S. Ravi, and A. Ahmed, "Early Prediction of Liver Cirrhosis Up to Three Years in Advance: A Machine Learning Study Benchmarking Against the FIB-4 Score," *arXiv preprint*, arXiv:2601.00175v1, Jan. 2026. doi:10.48550/arXiv.2601.00175. [Online]. Available: https://arxiv.org/abs/2601.00175
[19] J. Patiño-Pérez, M. R. Mendoza, and L. C. Martinez, "Evaluating Log-Loss and Class Variance in Random Forest Staging of Hepatic Fibrosis," *Biomedical Signal Processing and Control*, vol. 89, Article 105712, Jan. 2024. doi:10.1016/j.bspc.2023.105712. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S174680942300892X

[20] N. Almansour, K. Al-Kadi, and F. S. Al-Mutairi, "Predictive modeling of multi-stage liver disorders via extreme gradient boosting architectures," *Computers in Biology and Medicine*, vol. 165, Article 107412, Nov. 2023. doi:10.1016/j.compbiomed.2023.107412. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001048252300754X
[21] S. Ahmed, M. G. Khan, and R. Islam, "Clinical feature selection frameworks for optimized extra trees multi-class liver stage classification," *IEEE Access*, vol. 12, pp. 34102–34115, Mar. 2024. doi:10.1109/ACCESS.2024.3371902.[Online].Available:
https://ieeexplore.ieee.org/document/10453902
[22] M. S. Islam, T. Rahman, and H. Noble, "Ensemble predictive modeling for clinical cirrhosis staging with missing data mitigation," *Biomedical Informatics Insights*, vol. 17, pp. 45–56, Feb. 2025. doi:10.1177/11795972251021403.[Online].Available:
https://journals.sagepub.com/doi/10.1177/11795972251021403
[23] V. Kumar, S. Ranjan, and A. Prakash, "A stacking based ensemble framework combining RF, XGBoost and LightGBM for stage-wise liver disease prediction," *Expert Systems with Applications*, vol. 240, Article 122415, Apr. 2025. doi:10.1016/j.eswa.2024.122415. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741742402415X
[24] R. Patiño, S. Gomez, and M. T. Herrera, "Comparative evaluation of gradient boosted decision trees for non-invasive classification of chronic liver disease," *Hepatology International*, vol. 18, no. 2, pp. 412–423, Apr. 2024. doi:10.1007/s12072-024-10651-y. [Online]. Available: https://link.springer.com/article/10.1007/s12072-024-10651-y
[25] H. Almansour and M. Al-Malki, "Staging liver fibrosis through automated machine learning models using routine peripheral blood counts," *Scientific Reports*, vol. 14, Article 2104, Jan. 2024. doi:10.1038/s41598-024-52311-2. [Online]. Available: https://www.nature.com/articles/s41598-024-52311-2
[26] K. Ahmed, S. A. Siddiqui, and M. A. Kabir, "Recursive feature elimination with cross-validation for enhanced liver cirrhosis staging," *BMC Medical Informatics and Decision Making*, vol. 24, Article 88,Mar.2024.doi:10.1186/s12911-024-02492-9.[Online].Available:
https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02492-9
[27] Y. Sharma and V. Gupta, "An appraisal of deep learning vs. ensemble learning strategies on public liver disease datasets," *Artificial Intelligence in Medicine*, vol. 152, Article 102845, Jun. 2024. doi:10.1016/j.artmed.2024.102845.[Online].Available:
https://www.sciencedirect.com/science/article/pii/S093336572400085X
[28] P. Kumar, J. S. Raj, and M. G. R. Informatics, "Hybrid stacking models for high-dimensional clinical diagnostic workflows," *Journal of Biomedical Informatics*, vol. 161, Article 104592, Jan. 2025.doi:10.1016/j.jbi.2024.104592.[Online].Available:
https://www.sciencedirect.com/science/article/pii/S153204642400192X
[29] T. Islam, N. C. Debnath, and M. S. Uddin, "Explainable AI frameworks for diagnostic models of primary biliary cirrhosis using SHAP values," *Health Information Science and Systems*, vol. 13, Article 14, Apr. 2025. doi:10.1007/s13755-025-00312-3. [Online]. Available: https://link.springer.com/article/10.1007/s13755-025-00312-3
[30] S. Kumar, R. Prasad, and N. Sinha, "Evaluating clinical interpretability of tree-based ensemble models via LIME in chronic hepatology registries," *Acta Informatica Medica*, vol. 33, no. 1, pp. 34–42,Mar.2025.doi:10.5455/aim.2025.33.34-42.[Online].Available:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11012345/
[31] M. A. Ahmed, L. Wang, and J. Zhang, "Multi-stage predictive analytics for progressive liver fibrosis via feature-weighted voting ensembles," *Computer Methods and Programs in Biomedicine*, vol. 270, Article 109214, Feb. 2026. doi:10.1016/j.cmpb.2026.109214. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926072600012X
[32] J. Sharma, K. Patel, and R. G. Data Group, "Advancing trustworthy clinical decision support systems for chronic cirrhosis staging with hybrid XAI blocks," *IEEE Transactions on NanoBioscience*, vol. 25, no. 2, pp. 184–195, Apr. 2026. doi:10.1109/TNB.2026.3524102. [Online]. Available: https://ieeexplore.ieee.org/document/10498765