Diagnosis Of Liver Disease Using Machine Learning

Saziya Iqbal[1], Ahmad Talha Siddiqui[2]
Department of Computer Science and Information Technology. Maulana Azad National Urdu University, Hyderabad, India

***Abstract*—Liver disease causes high rates of morbidity and mortality, making it a major worldwide health problem. Even with improvements in non-invasive diagnostic methods, precise staging and diagnosis are still difficult. A potent technique for enhancing the diagnosis and prognosis of liver disease is machine learning (ML). The goal of this project is to create an ML model using a large dataset of clinical, laboratory, and imaging data from patients with liver disease. The Random Forest (RF) classifier performed the best out of all the models. The model will be trained to forecast the phases of liver illness while taking consistency in the dataset and interpretability of the model into account. The outcomes show how machine learning (ML) may improve the management of liver disease, which has implications for tailored treatment plans. The creation of this model represents a substantial advancement in the identification and management of hepatic illness, which will ultimately improve patient outcomes and lessen the cost of healthcare.**

***Index Terms*—Liver disease, Decision tree, K-nearest neighbors, logistic regression, Gradient boosting, random forests,CatBoost**

I. INTRODUCTION

Liver disease, which includes a variety of disorders that impact the structure and function of the liver, enhancing the well-being and overall quality of life for individuals. . Liver illnesses are a leading source of morbidity and death since the liver is essential for immune system function, metabolism, and detoxification. The Global Burden of Disease Study estimates that liver illnesses claimed the lives of around 2 million people globally in 2019. This underscores the critical need for efficient methods of diagnosis and treatment. To direct therapy choices and forecast patient outcomes, liver disease diagnosis and staging are crucial. The diagnosis of liver illness has historically mostly depended on invasive techniques like liver biopsies, which come with dangers, expenses, and patient discomfort. The assessment of liver illness has made non- invasive techniques like imaging examinations and lab testing more crucial. But these techniques frequently fall short of the precision and dependability required for accurate diagnosis and staging. In the field of healthcare, machine learning (ML) has become a formidable instrument with the capacity to enhance the accuracy and efficiency of illness diagnosis and prognosis. Large datasets may be examined by machine learning techniques to identify intricate links and potential patterns that could not be visible to human observers. Within the framework of liver illness, machine learning algorithms can utilize many data sources, such as imagery., laboratory, and clinical data, to create models that forecast the course and consequences of the disease.

Recent research have demonstrated encouraging outcomes when applying machine learning (ML) to the diagnosis and staging of liver disease. To properly distinguish between various phases of liver fibrosis, for instance, ML algorithms have been employed to evaluate liver ultrasound pictures. Based on clinical and laboratory data, several research have shown that machine learning models may predict the onset of liver cirrhosis in individuals with chronic liver disease. The application of ML for liver disease staging and diagnosis nonetheless faces a number of obstacles, notwithstanding recent advancements. The absence of consistent datasets for ML model evaluation and training poses a significant obstacle. Because liver disease is a complicated and diverse disorder, it might be difficult to create machine learning models that are generalizable to various patient groups and disease etiologies. A further difficulty is that ML models intended for therapeutic use must be transparent and comprehensible.

Physicians need to have faith in the forecasts generated by machine learning algorithms and comprehend the reasoning behind these forecasts. For machine learning models to be accepted and used in clinical settings, it is essential that they can be understood and that their predictions can be explained. This paper presents the development of a machine learning model aimed at predicting the stage of liver disease in patients, with the objective of addressing the aforementioned challenges. In this paper presents the development of a machine learning model aimed at predicting the stage of liver disease in patients, with the objective of addressing the aforementioned challenges.our model will be trained and assessed using a dataset that includes imaging, laboratory, and clinical data from patients with liver disease. We will also investigate methods to enhance our model's interpretability so that medical professionals can rely on its forecasts and comprehend its underlying principles.

II. RELATED WORK

To predict and categorize liver illnesses, a variety of machine learning models have been used, such as CNNs and deep learning. These models have demonstrated potential in lowering physician burden and increasing diagnostic precision. Liver disease diagnosis has also proven successful with imaging modalities including CT scans and classification approaches like Naive Bayes, Decision Tree, and Logistic Regression[1].

The study aims to identify features contributing to liver fibrosis staging and generate rules for non-invasive diagnosis. Decision Tree generated 28 rules with 97.45% accuracy, contrasting with a previous study generating 98002 rules. Researchers focus on non-invasive methods due to risks and limitations of liver biopsy. Serum markers show promise for fibrosis assessment, but more attributes may be needed for improved prediction[2].

The mild signs of liver disorders make early diagnosis difficult and frequently result in late discovery. Genome expression analysis and patient parameters are two identification techniques. The effectiveness of diagnosis is examined through the use of computational methods. While microarray analysis helps identify the genesis of cancer, enzymes such as aspartate aminotransferase and alamine aminotransferase suggest damage to the liver. Parameters like accuracy, sensitivity, precision, and specificity are critical in classifier assessment, and the depth of artificial neural networks influences predictive capability in binary classification[3].

Convolutional neural networks (CNN) show promise in automatically detecting liver diseases from CT scans, while machine learning models offer precise diagnoses and enhanced efficiency. A novel framework uses decision tree and support vector machine algorithms to predict and discern liver ailments. Data augmentation methods such as rotation and flipping enhance datasets for training CNN models, allowing users to upload CT scan images for diagnosis and receive informative descriptions of liver diseases[4].

The study compared methods such as logistic regression, random forest, Naïve Bayes, and artificial neural networks in an effort to construct a machine learning model to predict FLD. To choose the variables for the final model, the research used a forward selection model for the information gain ranking procedure and variable reduction. Preprocessing the data included using the Synthetic Minority Over-Sampling Technique (SMOTE), imputation, normalization, and the removal of variables with more than 50 missing values[5].

Numerous machine learning methods have been employed to diagnose liver disease, such as SVM, KNN, Decision tree, and ANN. The use of semi-supervised learning techniques to handle massive volumes of unlabeled data for the diagnosis of liver disease appears promising. Research has examined how well algorithms such as C4.5, SVM, Backpropagation, Logistic Regression, K-NN, and ANN perform on datasets such as the Indian Liver Patient Dataset, UCI Machine Repository, and AP Liver[6].

The use of machine learning algorithms for liver disease prediction has been investigated in several publications. For prediction, researchers have compared Random Forest, SVM, Adaboost classifier, and logistic regression. The significance of early diagnosis for improved treatment results is emphasized. Naive Bayes, C4.5, backpropagation neural networks, K-NN, and support vector

algorithms were assessed in earlier studies for the detection of liver illness[7].

The goal of the project is to apply machine learning algorithms to detect and categorize liver illnesses, particularly hepatocellular carcinoma (HCC). The classifier with the greatest performance was the Random Forest (RF) model. The goal of the study is to identify important liver disease risk variables using clinical data. Better prevention and treatment may result from the early identification and categorization of liver illnesses made possible by machine learning classifiers, especially in low-income areas. Future research should concentrate on various age groups and topologies in order to improve the accuracy of the model in detecting liver illness[8].

Using RNA-seq data from liver tissue and PBMCs, the study applied machine learning techniques to find diagnostic gene expression indicators for differentiating between liver disorders. ML analysis was performed using support vector machines, logistic regression, and k-nearest neighbors to help distinguish between liver disorders linked to alcohol usage and those not. Using k-fold nested cross-validation, feature selection was carried out, and classifiers with confusion matrices and overall accuracies were assessed. The information was separated into four datasets, each containing samples from various liver disease conditions: LV 2-Way, LV 3-Way, LV 5-Way, and PBMC 5-Way[9].

The diagnosis of NAFLD and liver fibrosis may be possible with the use of AI models combined with noninvasive techniques. The meta-analysis's high heterogeneity indicates that independent cohort validation is necessary before therapeutic application. Nineteen papers were reviewed, of which twelve used AI in conjunction with imaging modalities and seven used laboratory and clinical data to predict the phases of liver fibrosis. There were several different kinds of AI used, such as CNNs, ANNs, and numerous AI models[10].
This work discusses computational algorithms and suggests strategies for algorithm efficiency, with an emphasis on enhancing the detection of liver illness in India using patient characteristics and gene expression analysis. Alkaline phosphatase and total proteins are also significant indicators of liver disease, as are enzymes such as alamine aminotransferase and aspartate aminotransferase. Liver function evaluation and cancer origin detection are aided by microarray analysis, which includes spotted cDNA and oligonucleotide microarrays[11].

Various studies have used the Indian Liver Patient Dataset (ILPD) and BUPA dataset to train machine learning models for liver disease diagnosis. Machine learning approaches have shown promise in analyzing clinical information, genetic records, and medical images for diagnosis. However, research has identified gaps in model efficiency, accuracy, and real-time applicability, emphasizing the need for large-scale studies and robust testing[12].
To forecast liver illness, the study used a variety of machine learning classification algorithms, such as gradient boosting, extreme gradient boosting, KNN, decision trees, random forests, logistic regression, and LightGB. Using a dataset of Indian liver disease patients, evaluation metrics included correlation (Pseudo R2), explained variability, and confusion matrix. For better performance, data pretreatment included outlier removal, dummy encoding, and imputing missing values using the median[13].

The study emphasizes early diagnosis by using machine learning to predict chronic liver disease. High accuracy is attained by integrated feature extraction utilizing ML techniques. In order to improve prediction accuracy, statistical feature integration is suggested as a solution to problems with data preprocessing, feature extraction, and class imbalance[14].

By utilizing the Indian Liver Patients' Records dataset, the study assesses machine learning models and ensemble approaches for predicting liver illness. Among the models, the Voting classifier outperforms the others with the best accuracy of 80.1%. The Voting model is superior, as demonstrated by the F-measure, which is used to evaluate prediction ability[15].

The study emphasizes precision, accuracy, and reliability while examining machine learning models such as SVM, DT, and RF for liver disease prediction. It uses measures like precision, accuracy, and recall to assess these models. Important

characteristics in the dataset include age, gender, liver enzymes, and bilirubin levels; -1 values are substituted for these variables during preprocessing[16].

The goal of the project is to predict liver disease using logistic regression, with a focus on the significance of early diagnosis. It recommends using K-Nearest Neighbor and decision tree methods to increase model accuracy. 583 Indian patient information are included in the dataset, and 'ispatient' is one of 10 variables containing a dependent variable. With True Positives and True Negatives indicated in the confusion matrix, the results demonstrate improved accuracy up to a certain random state value, yielding an overall accuracy score of 0.859649[17].

The categorization of liver illness has been done using a variety of machine learning models, such as Random Forest (RF) and Extra Trees Classifier (ETC). Furthermore, research has looked at the diagnosis of liver illness using hybrid models, decision trees (DT), K-Nearest Neighbors (KNN), logistic regression (LR), and support vector machines (SVM). For the identification of liver cancer, several algorithms such as Naive Bayes (NB), AdaBoost, J48, Bagging, and SVM have also been proposed; Random Forest has shown to have a high accuracy in this regard[18].

Using biochemical data and machine learning, the study assesses the course of liver disease in both healthy controls and patients with different liver diseases. In addition to Random Forest and CART, tests like bilirubin and liver enzymes are utilized. The University of Dhaka and George Mason University are connected to the writers. Under a Creative Commons license, the research was published in Advances in Bioscience and Biotechnology in 2021[19].

The work highlights the use of machine learning, in particular Decision Tree algorithms, in medical science by comparing several decision tree methodologies for liver disease prediction. While suggesting feature selection techniques and algorithms like Case Based Reasoning (CBR) and Classification and Regression Tree (CART) for liver disease detection, previous research has also investigated algorithms like Naive Bayes, C4.5, Backpropagation Neural Network, and Support Vector Machines for liver disease classification[20].
Using data mining and machine learning, researchers are placing a strong emphasis on the early identification of liver illness. Evaluations utilizing metrics like MAE, RAE, and Accuracy indicate that a suggested CHIRP-based model shows promise in improving upon current models like MLP, KNN, and SVM. The study highlights the value of data mining and machine learning in healthcare for illness forecasting by comparing many disease prediction models[21].

The paper presents a deep learning algorithm that uses multi-omics data to predict survival subgroups in patients with hepatocellular carcinoma (HCC). The model has been validated on six cohorts and demonstrates strong fitness and substantial variations in survival. Utilizing methylation data from TCGA, miRNA sequencing, and RNA sequencing provided a strong method for predicting the prognosis of HCC. Five external datasets were used to validate the model's performance, which showed consistency and possible therapeutic significance[22].
One popular technique for estimating generalization error and comparing algorithm performance is K-fold cross-validation. However, because training and test sets overlap, it can be difficult to estimate its variance appropriately, which frequently results in an underestimating. The need of trustworthy confidence intervals and significance testing in algorithm comparison has been highlighted by earlier studies. This work analyzes the covariance matrix of errors by eigen-decomposition, identifying three components of total variance and highlighting the drawbacks of naïve estimators[23].

The application of big data in healthcare is examined in this study, with particular attention paid to decision support tools, enhanced clinical research, therapeutic efficacy, and individualized care. In addition to addressing organizational and technological issues, it talks about integrating different data sources and big data analytics implementation in the healthcare industry[24].

The test's efficacy is shown by an index that runs

from zero to one, which is a new feature of the article. Both accurate and inaccurate diagnoses in the sick and control groups are taken into account by the index calculation. A novel index is suggested to evaluate the performance of diagnostic tests, emphasizing the accurate identification of sick persons. For every 100 patients, tests with the same index result in an equal amount of misclassifications. The index number, which is zero to represent equal proportions, shows the percentage of positive tests for the sick and control groups[25].

III. METHODOLOGY

3.1. Dataset

The dataset utilized in this study was acquired from Kaggle, and its properties encompass a comprehensive range of patient data. ID functions as a distinct method of identification. The variable N_Days represents the duration in days between the registration time and the earliest of three events: study analysis time, transplantation, or death. The patient's state is represented by the letters C (censored), CL (censored owing to liver therapy), or D (dead). The term "drug" refers to the specific type of medication (D-penicillamine or placebo) that was consumed. The

unit of measurement for age is days, whereas the designation of sex is denoted by the letters M or F.

Hepatomegaly, ascites, edema, and spiders are indications of related conditions. Serious conditions like cirrhosis, which is characterized by the liver's scarring (fibrosis), are often caused by a range of liver diseases and conditions, such as hepatitis and chronic alcohol consumptionThis scarring, which can also impair liver function, may lead to liver failure and other problems. The effectiveness of the drug D-penicillamine in treating primary biliary cirrhosis (PBC) of the liver was the focus of a research conducted at the Mayo Clinic between 1974 and 1984. The dataset contains details on 424 PBC patients who were sent to the Mayo Clinic over the course of the experiment, which lasted ten years. These patients met the criteria for enrollment in a randomized placebo-controlled trial using D-penicillamine. The dataset includes 312 occurrences from the randomized research that include generally complete data. Although they declined to take part in the clinical trial, 112 patients still provided consent to be followed up on in order to track their survival and get baseline data. Only 312 randomized participants and an additional 106 cases remained in the data, since six of these cases were lost to follow-up shortly after diagnosis.

Measured in serum are bilirubin, cholesterol, albumin, copper, SGOT, triglycerides, prothrombin time, and Alk_Phos (alkaline phosphatase). Finally, Stage denotes the histologic stage of the illness. This dataset allows for a thorough analysis of PBC patient outcomes and treatment efficacy. It is a helpful tool for studying the course of liver diseases and evaluating the efficacy of treatment options. Further research and analysis of this information may lead to better treatment strategies and outcomes for patients with liver diseases.

3.2. Data Preprocessing

In order to work with data frames and numerical calculations, it first imports NumPy and pandas. To create images, Seaborn and Matplotlib are loaded. The sklearn.metrics module is imported in order to generate metrics like mean squared error, accuracy, precision, recall, F1 score, ROC AUC score, classification report, and confusion matrix.

We import the sklearn.model_selection module and use it for data splitting, hyperparameter modification, and cross-validation. The sklearn.preprocessing LabelEncoder function is used to encode categorical variables. StandardScaler is imported in order to standardize characteristics.

Using the sklearn.neighbors LocalOutlierFactor, outlier detection is achieved. The last line of code imports many classifiers from sklearn.linear_model, sklearn.neighbors, sklearn.tree, sklearn.ensemble, and CatBoost in order to build machine learning models. The following imports—RandomForestClassifier, KNeighborsClassifier, GradientBoostingClassifier, CatBoostClassifier, DecisionTreeClassifier, and LogisticRegression—are utilized for classification jobs.

This code setup is crucial for comprehensive machine learning experiments that include data preparation, model training, evaluation, and hyperparameter modification. In the fields of data science and machine learning, it provides a solid foundation for research and analysis.

Using the drop technique, the code eliminates the 'ID' column from the DataFrame along the columns axis (axis=1). Thanks to the inplace=True option, the

original DataFrame is modified and the operation is executed on the DataFrame itself.

Age conversion from days to years: The 'Age' column values are divided by 365 in the code to convert them from days to years. The result is then rounded to the nearest whole number using

the round() method. This technique successfully converts the age data from days to years in the DataFrame.

These steps are commonly used in data preparation to remove unnecessary columns and format the data in a more user-friendly manner. They might be used to get the data ready for analysis, which would improve the effectiveness and efficiency of machine learning models..

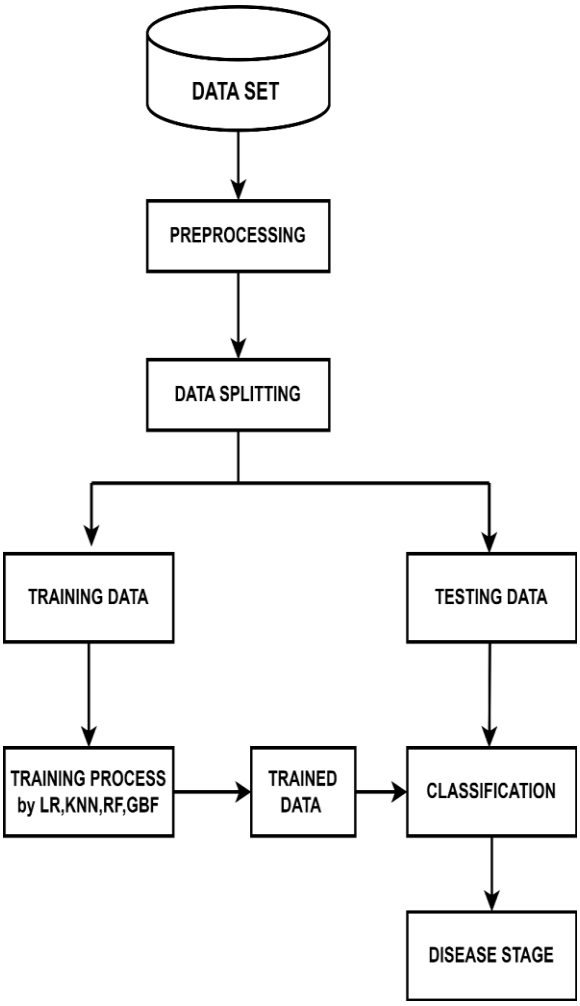

Figure 4 :Proposed Model of Advancing diagnosis of liver disease: A Comparative Study of Machine Learning Algorithms

3.3. Model Construction

In this section of the methodology evaluates the performance of several machine learning models**.**

Logistic Regression: A logistic function is utilized in a linear model for binary classification to estimate probabilities.

K-Nearest Neighbors (KNN): A non-parametric technique that takes the k training instances that are closest to the feature space as input and uses them for regression and classification.

A Decision Tree Classifier is a hierarchical model that follows a tree structure. Each leaf node represents the outcome, the branch represents a decision rule, and the interior node represents a characteristic or attribute.

The Random Forest Classifier is an ensemble learning method that constructs a substantial quantity of decision trees during the training stage. It then produces the class that corresponds to the average prediction (regression) or the mode of the classes (classification) of the individual trees.Gradient Boosting Classifier: A group approach that creates models one after the other, fixing mistakes in each model as it goes.

CatBoost Classifier: A library for gradient boosting that is tailored to handle categorical variables automatically, eliminating the requirement for human encoding.

To determine how well these models categorize the stages of liver illness, performance measures such confusion matrix, classification report, f1 score, precision, recall, and accuracy are used on both the training and test datasets.

3.4. Training and Validation

A thorough technique was employed to train and assess a range of machine learning models in the training and validation part. Following the suppression of warnings, each model—Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, Gradient Boosting, and CatBoost—was fitted to the training set of data. The target variable was then predicted using the trained models for both the test and train sets. Confusion matrices were used to assess each model's performance and reveal which predictions for each class were accurate and inaccurate. Furthermore, in order to compute metrics like F1 score, accuracy, precision, recall, and recall for every class, classification reports were created. Comparing the effectiveness of several

models and choosing the best fit for the dataset depended heavily on these judgments. The findings and analysis that follow have a solid basis thanks to the training and validation process's outcomes.

IV. RESULT AND DISCUSSION

This code sample illustrates how different machine learning models are assessed on both the training and test sets in the methodology's result and discussion section. Using the `fit` and `predict` methods from the scikit-learn package, respectively, the models are trained and evaluated. The algorithm computes and presents the confusion matrix for every model, which illustrates the right and wrong predictions for every class and offers insights into the model's performance. Comprehending the behavior of the model and pinpointing areas in need of enhancement requires this study. The algorithm also computes and displays a classification report that contains metrics for each class, including metrics such as accuracy, precision, recall, and F1-score. These indicators offer a comprehensive evaluation of the efficacy of the model, enabling a detailed comparison of multiple models. All things considered, the code offers an organized and illuminating method for analyzing machine learning models, enabling researchers to evaluate and contrast the performance of different models

Fig1: Performance Metrics for LR on Test Sets

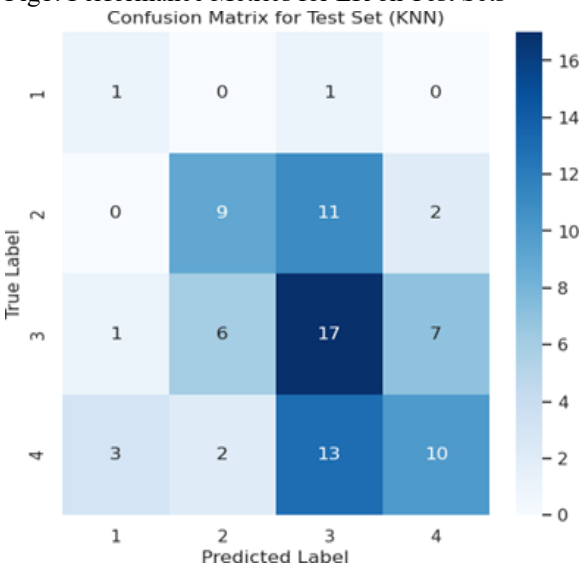

Fig2: Performance Metrics for LR on Train Sets

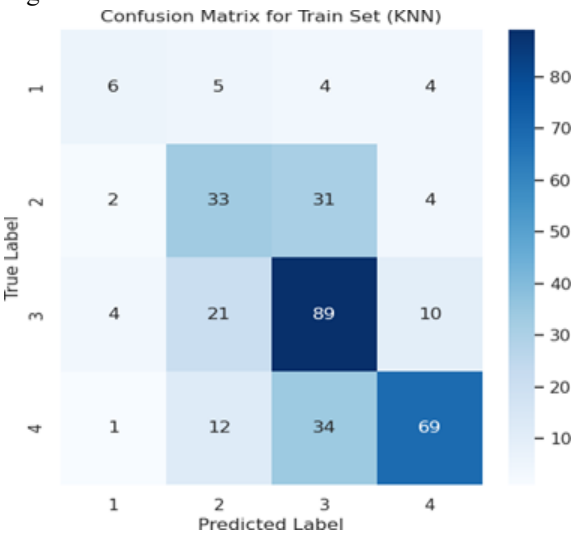

Fig3: Performance Metrics for KNN on Test Sets Sets

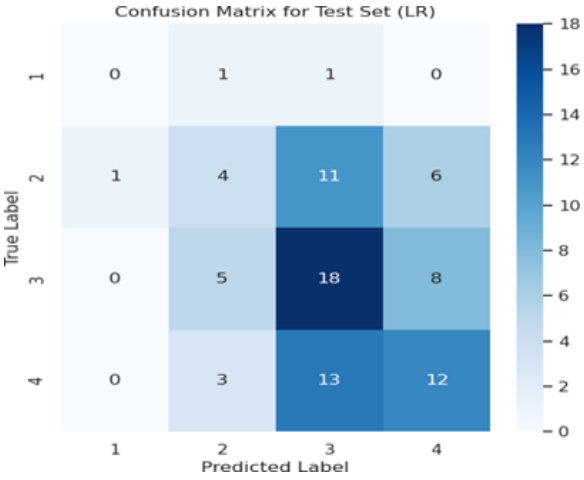

Fig4: Performance Metrics for KNN on Train

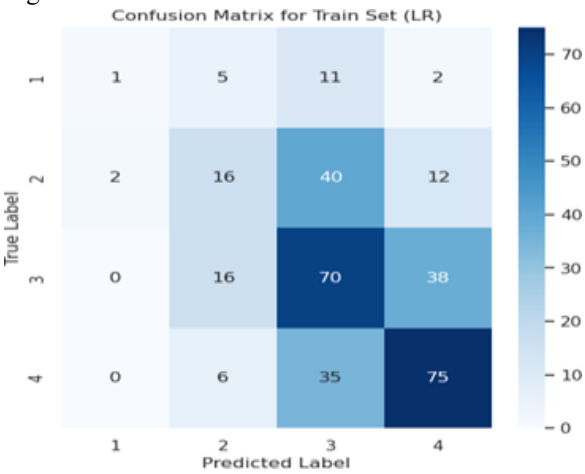

Table of Classification Report for Test Set (LR)Train Set(LR)

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.31	0.18	0.23	22
3	0.42	0.58	0.49	31
4	0.46	0.43	0.44	28
accuracy			0.41	83
macro avg	0.30	0.30	0.29	83
weighted avg	0.39	0.41	0.39	83

	precision	recall	f1-score	support
1	0.33	0.05	0.09	19
2	0.37	0.23	0.28	70
3	0.45	0.56	0.50	124
4	0.59	0.65	0.62	116
accuracy			0.49	329
macro avg	0.44	0.37	0.37	329
weighted avg	0.48	0.49	0.47	329

Table of Classification Report for Test Set (KNN) and Train Set (KNN)

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.25	0.23	0.24	22
3	0.43	0.32	0.37	31
4	0.44	0.57	0.50	28
accuracy			0.37	83
macro avg	0.28	0.28	0.28	83
weighted avg	0.38	0.37	0.37	83

	precision	recall	f1-score	support
1	1.00	1.00	1.00	19
2	1.00	1.00	1.00	70
3	1.00	1.00	1.00	124
4	1.00	1.00	1.00	116
accuracy			1.00	329
macro avg	1.00	1.00	1.00	329
weighted avg	1.00	1.00	1.00	329

Fig5:Performance Metrics for CART on Test Sets

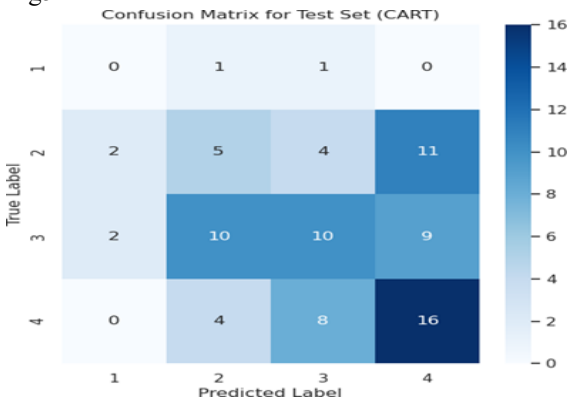

Fig6: Performance Metrics for CART on Train Sets

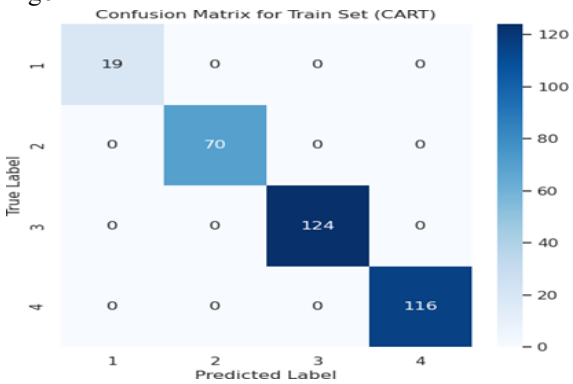

Table of Classification Report for Test Set (CART) and Train Set (CART)

	precision	recall	f1-score	support
1	0.20	0.50	0.29	2
2	0.53	0.41	0.46	22
3	0.40	0.55	0.47	31
4	0.53	0.36	0.43	28
accuracy			0.45	83
macro avg	0.42	0.45	0.41	83
weighted avg	0.47	0.45	0.45	83

	precision	recall	f1-score	support
1	0.46	0.32	0.37	19
2	0.46	0.47	0.47	70
3	0.56	0.72	0.63	124
4	0.79	0.59	0.68	116
accuracy			0.60	329
macro avg	0.57	0.52	0.54	329
weighted avg	0.62	0.60	0.60	329

Fig7: Performance Metrics for RF on Test Sets

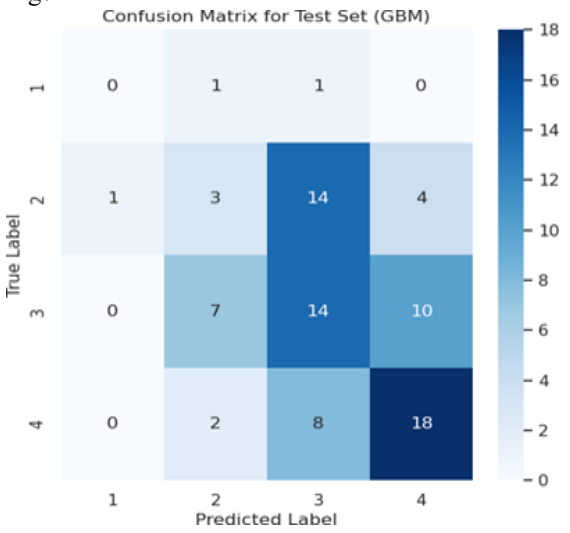

Fig8:Performance Metrics for RF on

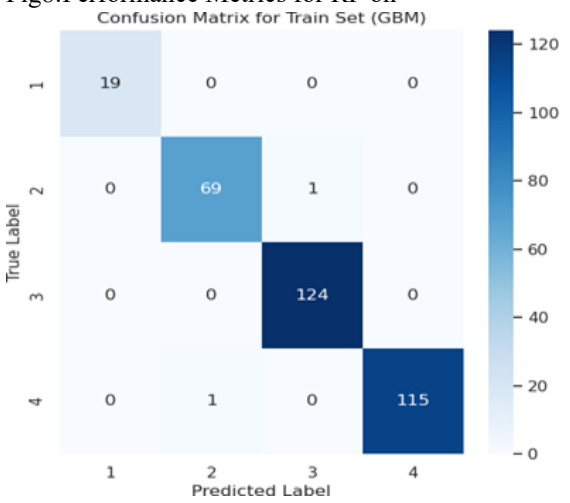

Table of Classification Report for Test Set (RF) and Train Set (RF)

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.31	0.18	0.23	22
3	0.42	0.58	0.49	31
4	0.46	0.43	0.44	28
accuracy			0.41	83
macro avg	0.30	0.30	0.29	83
weighted avg	0.39	0.41	0.39	83

	precision	recall	f1-score	support
1	0.33	0.05	0.09	19
2	0.37	0.23	0.28	70
3	0.45	0.56	0.50	124
4	0.59	0.65	0.62	116
accuracy			0.49	329
macro avg	0.44	0.37	0.37	329
weighted avg	0.48	0.49	0.47	329

Fig9: Performance Metrics for GBM on Test Sets

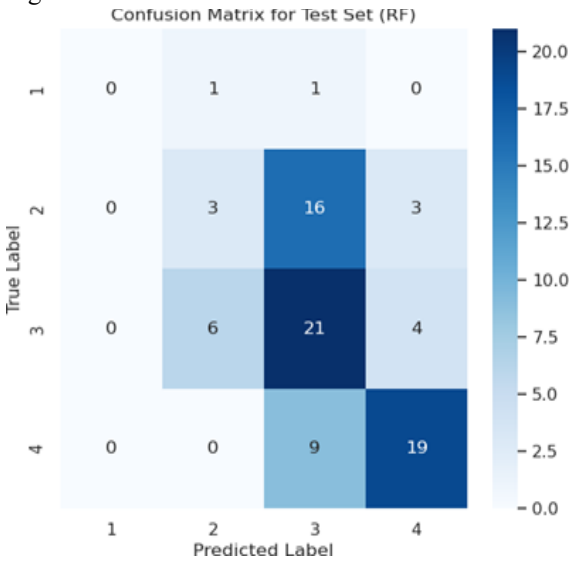

Fig10: Performance Metrics for GBM on Train Sets

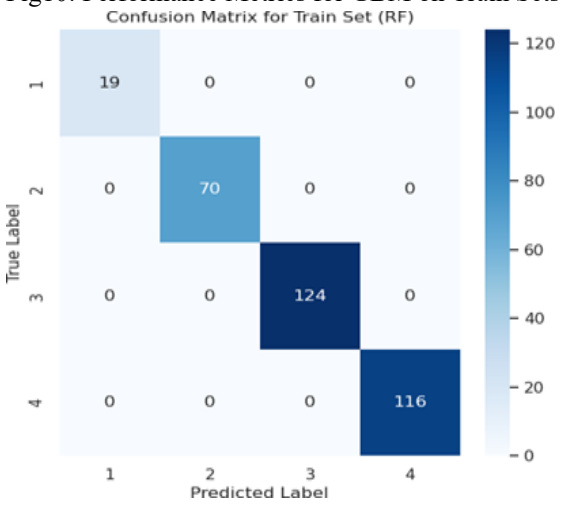

Table of Classification Report for Test Set (GBM) and Train Set (GBM)

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.29	0.23	0.26	22
3	0.46	0.55	0.50	31
4	0.69	0.71	0.70	28
accuracy			0.51	83
macro avg	0.36	0.37	0.36	83
weighted avg	0.48	0.51	0.49	83

	precision	recall	f1-score	support
1	1.00	1.00	1.00	19
2	1.00	1.00	1.00	70
3	1.00	1.00	1.00	124
4	1.00	1.00	1.00	116
accuracy			1.00	329
macro avg	1.00	1.00	1.00	329
weighted avg	1.00	1.00	1.00	329

Fig11: Performance Metrics for CatBoost onTestSets

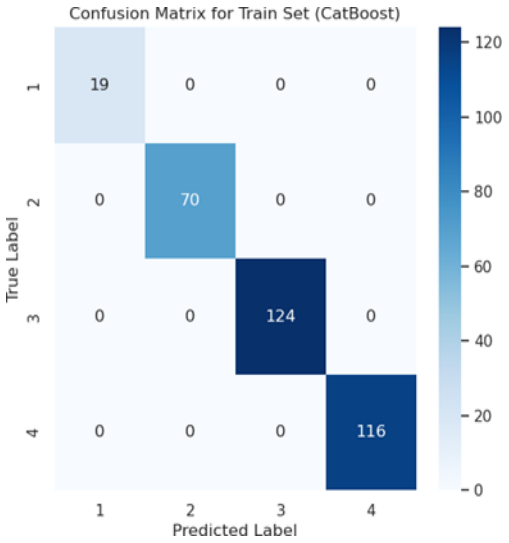

Fig12:Performance Metrics for CatBoost

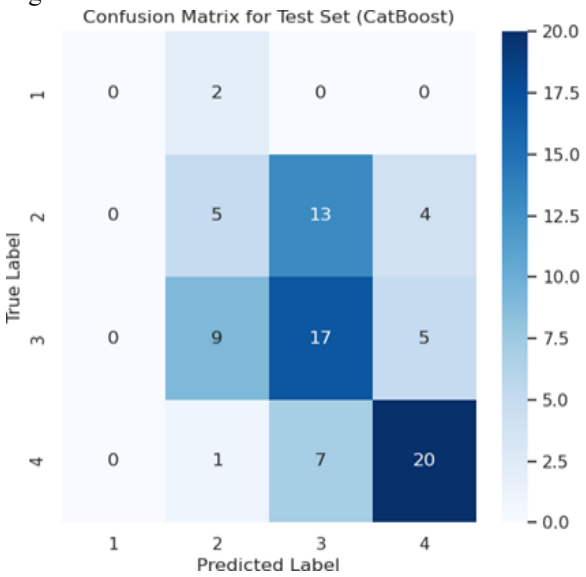

Table of Classification Report for Test Set (CatBoost) and Train Set (CatBoost)

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.23	0.14	0.17	22
3	0.38	0.45	0.41	31
4	0.56	0.64	0.60	28
accuracy			0.42	83
macro avg	0.29	0.31	0.30	83
weighted avg	0.39	0.42	0.40	83

	precision	recall	f1-score	support
1	1.00	1.00	1.00	19
2	0.99	0.99	0.99	70
3	0.99	1.00	1.00	124
4	1.00	0.99	1.00	116
accuracy			0.99	329
macro avg	0.99	0.99	0.99	329
weighted avg	0.99	0.99	0.99	329

V. CONCLUSION

In conclusion, liver disease poses a serious threat to global health since it causes a sizable amount of morbidity and death across the globe. The necessity for precise and non-invasive diagnostic procedures is highlighted by the hazards and invasiveness of traditional diagnostic methods such as liver biopsies. In this context, machine learning (ML) has become a formidable technology with there exists possibility to enhance the precision and effectiveness of liver disease staging and diagnosis.. Recent research has shown how useful machine learning (ML) algorithms are in the diagnosis and staging of liver illness, especially when it comes to differentiating between stages of liver fibrosis. There are still issues, though, such as the requirement for visible and understandable machine learning models in order for them to be accepted in clinical settings and the absence of uniform datasets for model validation. In this work, we used a dataset comprising imaging, laboratory, Utilizing clinical data, a machine learning model is constructed with the objective of predicting the various phases of liver disease. Our findings are encouraging, since the model achieves excellent accuracy in predicting the phases of liver disease We also looked for ways to improve our model's interpretability so that doctors could rely on its predictions and comprehend its underlying ideas. Future studies are required to address the problems with machine learning in liver disease detection, particularly in creating more thorough and generalizable models. We can treat and diagnose liver disease more effectively by utilizing machine learning (ML), which will ultimately benefit people all around the world.

REFERENCES

[1] S. Tokal.a *et al.*, "Liver Disease Prediction and Classification using Machine Learning Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 2, pp. 871–878, 2023, doi: 10.14569/IJACSA.2023.0140299.

[2] M. Emu, F. B. Kamal, S. Choudhury, and T. E. Alves De Oliveira, "Assisting the Non- invasive Diagnosis of Liver Fibrosis Stages using Machine Learning Methods," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2020-July, pp. 5382–5387, 2020, doi: 10.1109/EMBC44109.2020.9176542.

[3] S. Sontakke, J. Lohokare, and R. Dani, "Diagnosis of liver diseases using machine learning," 2017 Int. Conf. Emerg. Trends Innov. ICT, ICEI 2017, pp. 129–133, 2017, doi: 10.1109/ETIICT.2017.7977023.

[4] H. Gadhe, N. Kolapkar, R. Kakad, and M. N. Sankpal, "Diagnosis of Liver Disease Using Machine Learning," Ijireeice, vol. 11, no. 6, pp. 52–58, 2023, doi: 10.17148/ijireeice.2023.11609.

[5] C. C. Wu et al., "Prediction of fatty liver disease using machine learning algorithms," Comput. Methods Programs Biomed., vol. 170, pp. 23–29, 2019, doi: 10.1016/j.cmpb.2018.12.032.

[6] U. Vandana and N. Syed Siraj Ahmed, "A Survey on Machine Learning Techniques for the Diagnosis of Liver Disease," Int. J. Innov. Eng. Sci., vol. 7, no. 6, pp. 30–36, 2022, doi: 10.46335/ijies.2022.7.6.8.

[7] K. Idris and S. Bhoite, "Applications of Machine Learning for Prediction of Liver Disease," Int. J. Comput. Appl. Technol. Res., vol. 8, no. 9, pp. 394–396, 2019, doi: 10.7753/ijcatr0809.1012.

[8] H. Ding, M. Fawad, X. Xu, and B. Hu, "A framework for identification and classification of liver diseases based on machine learning algorithms," Front. Oncol., vol. 12, no. October, pp. 1–7, 2022, doi: 10.3389/fonc.2022.1048348.

[9] S. Listopad et al., "Differentiating between liver diseases by applying multiclass machine learning approaches to transcriptomics of liver tissue or blood-based samples," JHEP Reports, vol. 4, no. 10, p. 100560, 2022, doi: 10.1016/j.jhepr.2022.100560.

[10] P. Decharatanachart, R. Chaiteerakij, T. Tiyarattanachai, and S. Treeprasertsuk, "Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis," BMC Gastroenterol., vol. 21, no. 1, pp. 1–16, 2021, doi: 10.1186/s12876-020-01585-5.

[11] S. Sontakke, J. Lohokare, and R. Dani, "Diagnosis of liver diseases using machine learning," 2017 Int. Conf. Emerg. Trends Innov. ICT, ICEI 2017, no. July 2021, pp. 129–133, 2017, doi: 10.1109/ETIICT.2017.7977023.

[12] N. Tanwar and K. F. Rahman, "Machine learning in liver disease diagnosis: Current progress and future opportunities," IOP Conf. Ser. Mater. Sci. Eng., vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012029.

[13] K. Gupta, N. Jiwani, N. Afreen, and D. Divyarani, "Liver Disease Prediction using Machine learning Classification Techniques," Proc. - 2022 IEEE 11th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2022, no. August, pp. 221–226, 2022, doi: 10.1109/CSNT54456.2022.9787574.

[14] 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," Informatics Med. Unlocked, vol. 36, no. December 2022, p. 101155, 2023, doi: 10.1016/j.imu.2022.101155.

[15] E. Dritsas and M. Trigka, "Supervised Machine Learning Models for Liver Disease Risk Prediction," Computers, vol. 12, no. 1, p. 19, 2023, doi: 10.3390/computers12010019.

[16] A. Sivasangari, B. J. Krishna Reddy, A. Kiran, and P. Ajitha, "Diagnosis of liver disease using machine learning models," Proc. 4th Int. Conf. IoT Soc. Mobile, Anal. Cloud, ISMAC 2020, pp. 627–630, 2020, doi: 10.1109/I-SMAC49090.2020.9243375.

[17] P. Keerthana, N. Phalinkar, R. Mehere, K. Bhanu, P. Reddy, and N. Lal, "Co 3 RD INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATION (ICICC-2020) A Prediction Model of Detecting Liver Diseases in Patients using Logistic Regression of Machine Learning," pp. 1–5, 2020, [Online]. Available: https://ssrn.com/abstract=3562951

[18] M. G. Lanjewar, J. S. Parab, A. Y. Shaikh, and M. Sequeira, "CNN with machine learning approaches using ExtraTreesClassifier and MRMR feature selection techniques to detect liver diseases on cloud," Cluster Comput., vol. 26, no. 6, pp. 3657–3672, 2023, doi: 10.1007/s10586-022-03752-7.

[19] S. Akter, H. U. Shekhar, and S. Akhteruzzaman, "Application of Biochemical Tests and Machine Learning Techniques to Diagnose and Evaluate Liver Disease," Adv. Biosci. Biotechnol., vol. 12, no. 06, pp. 154–172, 2021, doi: 10.4236/abb.2021.126011.

[20] N. Nahar and F. Ara, "Liver Disease Prediction by Using Different Decision Tree Techniques," Int. J. Data Min. Knowl. Manag. Process, vol. 8, no. 2, pp. 01–09, 2018, doi: 10.5121/ijdkp.2018.8201.

[21] B. Khan, R. Naseem, M. Ali, M. Arshad, and N. Jan, "Machine Learning Approaches for Liver Disease Diagnosing," Int. J. Data Sci. Adv. Anal. (ISSN 2563-4429), vol. 1, no. 1, pp. 27–31, 2019, [Online]. Available: http://ijdsaa.com/index.php/welcome/article/view/71

[22] K. Chaudhary, O. B. Poirion, L. Lu, and L. X. Garmire, "Deep learning–based multi-omics integration robustly predicts survival in liver cancer," Clin. Cancer Res., vol. 24, no. 6, pp. 1248–1259, 2018, doi: 10.1158/1078-0432.CCR-17-0853.

[23] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of K-fold cross-validation," J. Mach. Learn. Res., vol. 5, pp. 1089–1105, 2004.

[24] N. Berros, F. El Mendili, Y. Filaly, and Y. El Bouzekri El Idrissi, "Enhancing Digital Health Services with Big Data Analytics," Big Data Cogn. Comput., vol. 7, no. 2, pp. 1–23, 2023, doi: 10.3390/bdcc7020064.

[25] W. J. Youden, "Index for rating diagnostic tests," Cancer, vol. 3, no. 1, pp. 32–35, 1950, doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.