History of Medicine, 2022, 8(2): 678-691

DOI:10.48047.v8.2.2024.64

AI-Driven Optimization for Liver Disease Prediction Using Data Balancing Techniques

Pushpa Latha Malgireddi[1],Nusrath Begum Mohammad[1],Chigurlapalli Swathi[1]
[1]Department of Computer Science Engineering,

[1]Sree Dattha Group of Institutions, Sheriguda, Hyderabad, Telangana

ABSTRACT

Liver disease affects millions of people worldwide and is a serious global health problem. Better patient outcomes and efficient disease treatment depend on prompt and accurate diagnosis. It has been demonstrated that machine learning (ML) approaches may accurately predict a wide range of medical disorders, including illnesses of the liver. However, the caliber and volume of training data have a major impact on how well machine learning models perform. Sadly, class imbalance affects a lot of datasets, meaning that some classes—like patients with and without diseases—are not fairly represented. To improve the accuracy of liver disease prediction using ML models, it is imperative to address this imbalance as it might result in biased predictions and decreased model accuracy. Thus, the goal of this research is to use sophisticated data balancing techniques to overcome the problem of class imbalance. In order to create a more balanced dataset, our suggested approach preprocesses the dataset using the Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic samples for the minority class. To further enhance the model's performance, it further modifies the cost function of the learning process to take the class imbalance into consideration. An ML model (support vector classifier, gradient boosting classifier, and logistic regression) is trained to predict liver illness once we have a balanced dataset. Using a variety of criteria, including accuracy, precision, recall, and F1-score, the suggested model is tested on a separate test dataset in order to determine how effective it is. Through the use of data balancing algorithms, this model may successfully manage class imbalance. As a result, medical professionals should be able to diagnose liver disorders with more accuracy and care, which will eventually improve patient outcomes.

1.INTRODUCTION

1.1 Overview

Liver disease prediction involves the utilization of various computational techniques and medical data analysis to forecast the likelihood of an individual developing liver-related disorders. The liver plays a vital role in metabolism, detoxification, and nutrient storage, making its health critical for overall well-being. Predictive models for liver disease typically integrate diverse datasets, including clinical history, biochemical markers, imaging results, genetic information, and lifestyle factors. These datasets are analyzed using machine learning algorithms such as logistic regression, decision trees, support vector machines, or deep learning approaches to identify patterns and correlations indicative of liver disease onset or progression.

Data preprocessing is a crucial step in liver disease prediction, involving tasks like missing value imputation, feature scaling, and outlier detection to enhance the quality and relevance of input data. Feature selection techniques may also be employed to identify the most informative variables contributing to disease prediction. Once the data is prepared, it is divided into training, validation, and testing sets for model development and evaluation. During the training phase, the predictive model learns from historical data to establish relationships between input features and the presence or absence

of liver disease. Validation of the predictive model involves assessing its performance on unseen data to ensure its generalizability and reliability. Metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) are commonly used to evaluate model performance. Continuous refinement and optimization of the predictive model may be necessary to improve its accuracy and robustness. In clinical practice, liver disease prediction models have the potential to assist healthcare providers in early detection, risk stratification, and personalized intervention strategies. By identifying individuals at high risk of developing liver disease, preventive measures can be implemented to mitigate disease progression and improve patient outcomes. Moreover, predictive models can facilitate resource allocation and healthcare planning by identifying high-risk populations for targeted screening and intervention programs.

1.2 Problem Statement

The problem statement in liver disease prediction research revolves around addressing the challenges associated with the late diagnosis and limited treatment options for liver disorders. Despite the significant impact of liver diseases on public health and individual well-being, current diagnostic methods often fail to detect these conditions until they reach advanced stages, leading to poorer prognosis and increased healthcare costs. This delay in diagnosis underscores the need for accurate and reliable predictive models that can identify individuals at risk of developing liver disease before clinical symptoms manifest.

Furthermore, the heterogeneity of liver diseases, which encompass various conditions such as viral hepatitis, alcoholic liver disease, NAFLD, cirrhosis, and HCC, presents a complex challenge for predictive modeling. Each condition may have distinct risk factors, biomarkers, and disease trajectories, necessitating comprehensive and integrative approaches to data analysis. Developing predictive models capable of stratifying individuals based on their specific disease risks and prognosis is essential for guiding personalized intervention strategies and optimizing patient outcomes.

Moreover, while advancements in computational techniques offer promising opportunities for liver disease prediction, several challenges remain to be addressed. These include the need for large-scale, high-quality datasets encompassing diverse patient populations and longitudinal follow-up data to validate predictive models robustly. Additionally, ensuring the interpretability, generalizability, and clinical relevance of predictive models is critical for their successful translation into clinical practice.

1.3 Traditional system

In traditional medical practice, liver disease prediction relied on a comprehensive approach that integrated clinical assessments and diagnostic tests. Physicians would begin by taking detailed medical histories, probing for symptoms and risk factors. Physical examinations were conducted to assess signs of liver disease, including jaundice, spider angiomas, and hepatomegaly. Laboratory tests provided insights into liver function, injury, and etiology, measuring liver enzymes, viral hepatitis markers, and metabolic parameters.

Imaging studies, such as ultrasound, CT scans, and MRI, offered anatomical information aiding in diagnosis and staging. Ultrasonography would evaluate liver size, texture, and lesions, while cross-sectional imaging techniques would visualize liver morphology, identifying cirrhosis, HCC, and metastases. Non-invasive techniques assessed liver stiffness, aiding in fibrosis evaluation.

In cases needing further evaluation, liver biopsies provided tissue samples for microscopic examination, revealing inflammation, fibrosis, and specific pathologies. Scoring systems and risk prediction models, like Child-Pugh and MELD scores, quantified disease progression, complications, and mortality risk.

This integrated approach allowed physicians to predict the presence, severity, and prognosis of liver disease, guiding treatment decisions and patient management.

1.4 Research motivation

The motivation for research in liver disease prediction stems from the significant impact of liver disorders on public health and individual well-being. Liver diseases encompass a wide spectrum of conditions, including viral hepatitis, alcoholic liver disease, non-alcoholic fatty liver disease (NAFLD), cirrhosis, and hepatocellular carcinoma (HCC), among others. These conditions contribute substantially to morbidity, mortality, and healthcare costs globally.

Early detection and timely intervention are critical for effectively managing liver diseases and improving patient outcomes. However, liver diseases often remain asymptomatic until advanced stages, leading to delayed diagnosis and limited treatment options. Consequently, there is a pressing need to develop accurate and reliable predictive models that can identify individuals at risk of developing liver disease before overt clinical symptoms manifest.

Advancements in computational techniques, such as machine learning and data mining, have enabled the integration and analysis of diverse datasets to identify predictive biomarkers and risk factors associated with liver disease. Such models have the potential to facilitate early intervention, personalized treatment strategies, and improved patient outcomes.

Furthermore, liver disease prediction research contributes to the broader goal of precision medicine, which seeks to tailor healthcare interventions to individual characteristics, including genetic predispositions, environmental exposures, and lifestyle factors. By identifying individuals at elevated risk of liver disease, predictive models can inform targeted screening programs, lifestyle modifications, and pharmacological interventions to prevent disease onset or progression.

1.5 Research Objectives:

The objective of this research is to investigate the efficacy of machine learning algorithms in predicting liver disease progression and outcomes using clinical, laboratory, and imaging data. By leveraging advanced computational techniques, the study aims to develop predictive models that can accurately stratify patients based on their risk of liver disease progression, complications, and mortality. The research seeks to identify key predictive features and risk factors associated with liver disease progression, including demographic characteristics, clinical parameters, biomarkers, and imaging findings. Additionally, the study aims to assess the performance and generalizability of the predictive models across diverse patient populations and healthcare settings. Ultimately, the research aims to contribute to the development of personalized prediction tools that can assist clinicians in early detection, risk assessment, and treatment optimization for patients with liver disease.

1.6 Applications:

Clinical Decision Support: Our project's predictive models can assist healthcare providers in making informed decisions about liver disease diagnosis and treatment.

Telemedicine: ML-enabled prediction models can be integrated into telemedicine platforms, enabling remote monitoring and diagnosis of liver disease.

Public Health Surveillance: Our project can contribute to public health efforts by providing insights into liver disease prevalence, trends, and risk factors.

Drug Development: Predictive models can be used to identify biomarkers and therapeutic targets for liver disease, aiding in drug development and personalized medicine.

Health Insurance: Insurers can use predictive models to assess individuals' risk of developing liver disease and adjust insurance premiums accordingly.

Patient Education: Our project's predictive models can be used to educate patients about their risk factors for liver disease and promote preventive measures.

Research: Our project's findings can contribute to liver disease research, advancing our understanding of disease mechanisms and informing future treatment strategies.

2. LITERATURE SURVEY

Amin, et al. [1] proposed an integrated feature extraction method utilizing various projection techniques to categorize liver patients. The process involves imputing missing values and handling outliers as pre-treatment. Integrated feature extraction then extracts significant features for classification from pre-processed data. A simulation study reinforced the methodology. Their approach incorporated multiple ML algorithms, achieving high accuracy, precision, recall, F1 score, and AUC score in predicting liver diseases. Results outperformed existing studies, offering diagnostic support for physicians.

Md Abdul Quadir, et al. [2] proposed a novel liver disease prediction architecture, utilizing ensemble learning and enhanced preprocessing on the Indian Liver Patient Dataset (ILPD). Their model employed various data preprocessing techniques, improving accuracy through proper imputations. Features were selected via multiple methods. The model, trained on enhanced preprocessed data, outperformed others, achieving high testing accuracy, providing a practical liver disease detection solution.

Gupta, et al. [3] proposed historical and classified input of patients and output data was fed into various algorithms or classifiers for predicting the future data of patients. The algorithms used there for predicting liver patients they are Logistic regression, Decision Tree, Random Forest, KNNeighbor, Gradient Boosting, Extreme Gradient Boosting, LightGB. Based on the analysis and result calculations, it was found that these algorithms had obtained good accuracy after feature selection.

Grissa, et al. [4] analyzed Danish health registries spanning nineteen years, predicting ALF or ALC development based on medical history. They used statistical and machine learning techniques on Danish National Patient Registry data, identifying predominant ALC cases with strong liver dysfunction associations. ML models achieved high AUC (0.89) for ALC classification but lower performance for ALF prediction (AUC = 0.67 for NaiveBayes). Results revealed comorbidities aiding ALC prediction, showing potential in ALD knowledge extraction.

Dritsas, et al. [5] proposed this research work, various ML models and Ensemble methods were evaluated and compared in terms of Accuracy, Precision, Recall, F-measure, and area under the curve (AUC) in order to predict liver disease occurrence. The experimental results showed that the Voting classifier outperformed the other models with an accuracy, recall, and F-measure of 80.1%, a precision of 80.4%, and an AUC equal to 88.4% after SMOTE with 10-fold cross-validation.

Kumar, et al. [6] proposed heart illness and liver infection prediction via machine learning (ML) and data analytics. They leveraged ML algorithms due to the abundance of medical data, with a focus on heart and liver diseases. Logistic regression and random forest classifiers were used, with logistic regression excelling in heart disease classification, and random forest in liver disease prediction. Through extensive model comparison, logistic regression and random forest emerged as superior choices for heart and liver disease prediction, respectively.

Behera et al. [7] proposed a hybrid model for heart and liver data classification, combining support vector machine (SVM) with a modified particle swarm optimization approach. Using datasets from the UCI machine learning repository, they evaluated the model's performance in terms of classification accuracy, error, recall, and F1 score. Results were compared with SVM, hybrid PSO-SVM, and hybrid CPSO-SVM algorithms.

Singh et al. [8] analyzed the Indian Liver Patient Dataset (ILPD) from the University of California, Irvine database to predict liver disease risk using attributes like age, gender, and bilirubin levels. They evaluated multiple classification algorithms including Logistic Regression, Random Forest, Naive Bayes, and k-nearest neighbor (IBk) for accuracy. Their study compared classifier results with and without feature selection, culminating in the development of intelligent liver disease prediction software (ILDPS) integrating software engineering principles, feature selection, and classification techniques.

Azam, et al. [9] developed computational model building techniques for accurate liver disease prediction. They utilized Random Forest, Perceptron, Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms. Their work involved hybrid model construction and comparative analysis to enhance prediction performance. Initially, classification algorithms were applied to original liver patient datasets from the UCI repository. Features were analyzed and adjusted to improve predictor performance, with KNN algorithm outperforming others with feature selection.

Ghazal, et al. [10] purpose this research was to assess the efficacy of various Machine Learning (ML) algorithms to lower the high cost of liver disease diagnosis through prediction. With the current rise in numerous liver disorders, it was more important than ever to detect liver disease early on. This research proposed an intelligent model to predict liver disease using machine learning technique. This proposed model was more effective and comprehensive in terms of performance of 0.884 accuracy, and 0.116 miss-rate.

3. PROPOSED METHODOLOGY

3.1 Overview

This work demonstrates a machine learning models focused on predicting liver disease. It provides a comprehensive workflow for building, training, and evaluating machine learning models for liver disease prediction, along with visualizations to help understand the data and assess model performance. Here's an overview of this work:

Step 1 – Data Loading and Exploration

- The code begins by importing necessary libraries called pandas, numpy, matplotlib, seaborn, and scikit-learn modules.
- It loads a dataset (indian_liver_patient.csv) using pandas and displays some initial information about the dataset.

Step 2 – Data Preprocessing

- It handles missing values in the column Albumin_and_Globulin_Ratio by filling them with zeros.
- It encodes the target variable Dataset into binary values (0 for no liver disease, 1 for liver disease).
- It scales the numerical features for better model performance.

Step 3 – Data Visualization

Step 4 – Correlation Analysis i.e., It generates a heatmap to visualize the correlation between numerical features, providing insights into feature relationships.

Step 5 – PCA is applied to reduce the dimensionality of the dataset to 5 principal components, which can help improve model performance.

Step 6 – The dataset is split into training and testing sets, which are essential for model evaluation.

Step 7: Model Training

- KNN model
- RFC model

Step 8 – Model Evaluation is performed using metrics like accuracy and visualizes the results using a confusion matrix.

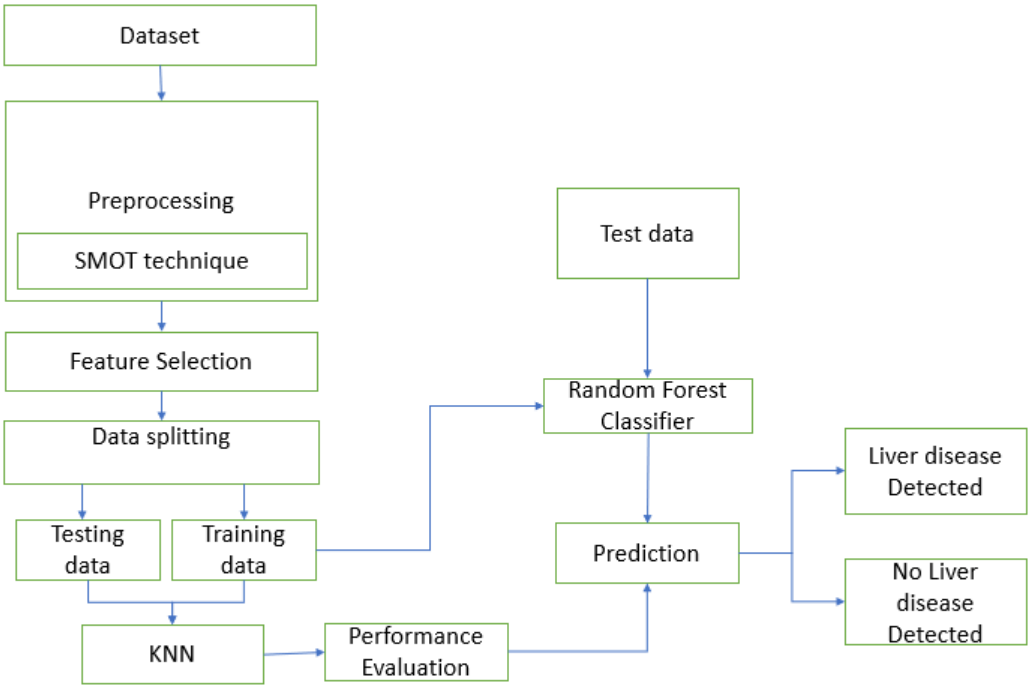


Figure 3.1: Block diagram of liver disease prediction

3.2 RANDOM FOREST CLASSIFIER:

Random Forest is an ensemble learning technique used for both classification and regression tasks. It operates by constructing multiple decision trees during the training phase and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Here's an overview of how Random Forest works:

1. Decision Trees: At the core of Random Forest are decision trees. Each decision tree is built by randomly selecting a subset of features from the dataset and splitting the data into smaller subsets based on these features. This process continues recursively until a stopping criterion is reached, such as reaching a maximum depth or minimum number of samples in a leaf node.

2.Bootstrapping: Random Forest utilizes bootstrapping, a resampling technique, to create multiple subsets of the original dataset. Each decision tree in the Random Forest is trained on a different bootstrap sample, ensuring diversity among the trees.

3.Random Feature Selection: During the construction of each decision tree, a random subset of features is considered at each split point. This random feature selection helps to decorrelate the trees and improve the overall performance of the ensemble.

4.Voting or Averaging: For classification tasks, the final prediction of the Random Forest is determined by a majority vote among the individual trees. For regression tasks, the final prediction is the average of the predictions made by each tree.

5.Bagging: Random Forest employs a technique called bagging (bootstrap aggregating) to reduce overfitting and variance. By training multiple decision trees on different subsets of the data and averaging their predictions, Random Forest tends to generalize well to unseen data.

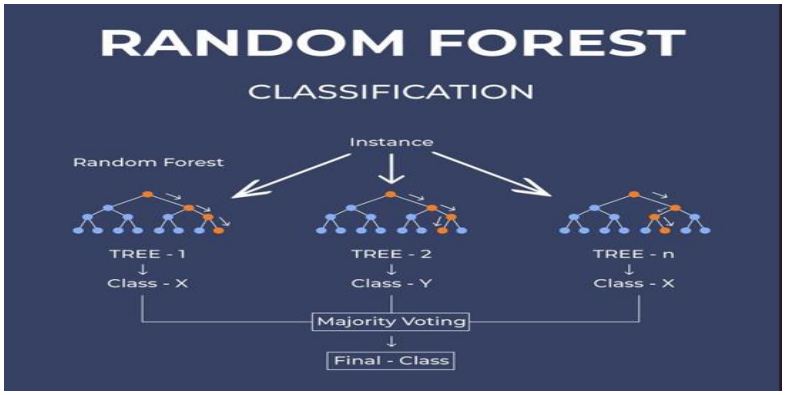


Figure 1 Random Forest Classification

Random Forest is a powerful ensemble learning technique that combines the predictions of multiple decision trees to improve the overall accuracy and robustness of the model. It is commonly used in various domains, including healthcare, finance, and bioinformatics, due to its ability to handle complex datasets and produce reliable predictions.

1.Random Forest Classifier Initialization: In the code, an instance of the RandomForestClassifier is initialized with specific parameters such as the number of estimators ,criterion for splitting , maximum depth of the trees, and minimum number of samples required to split a node. These parameters can be adjusted to optimize the performance of the Random Forest model.

2.Model Training: Once initialized, the RandomForestClassifier is trained on the training data. During training, the Random Forest algorithm constructs multiple decision trees based on bootstrapped samples of the training data and random subsets of features.

3.Prediction: After training, the trained Random Forest model is used to make predictions on the test data. For each data point in the test set, the model combines the predictions of all decision trees in the forest to produce a final prediction.

4.Evaluation: The performance of the Random Forest model is evaluated using various metrics such as precision, recall, F1 score, accuracy, and confusion matrix. These metrics provide insights into the model's ability to correctly classify instances of liver disease and its performance on the test data.

Random Forest Classifier is an integral part of the project, contributing to the accurate prediction of liver disease based on the provided dataset and features. It leverages the ensemble of decision trees to make robust predictions and is evaluated.

3.3 ADVANTAGES

High Accuracy: Random Forests generally provide high accuracy in classification tasks. By aggregating the predictions of multiple trees, the model tends to be more accurate than individual trees, especially in complex datasets.

Robust to Overfitting: The ensemble nature of Random Forest helps mitigate overfitting. Each tree is trained on a random subset of the data and features, which makes the model more robust and less prone to memorizing noise in the training data.

Handles Missing Values: Random Forests can handle missing values in the dataset. When making predictions for a particular instance with missing values, the algorithm can use the predictions from the other trees to make a robust estimate.

Variable Importance: Random Forest provides a feature importance score for each feature in the dataset. This allows users to identify which features are more influential in making predictions, aiding in feature selection and .

No Need for Feature Scaling: Random Forests are not sensitive to the scale of input features. Unlike some other machine learning algorithms (e.g., SVMs, neural networks), Random Forests do not require feature scaling, making them easier to work with.

Handles Imbalanced Datasets: Random Forests can handle imbalanced datasets well, particularly when combined with techniques like class weighting or data balancing during training. This is important in medical applications where classes may be unevenly distributed.

Effective for Large Datasets: Random Forests are effective on large datasets with many features. The parallelization capability allows them to efficiently handle a large number of data points and features.

Out-of-Bag Error Estimation: Random Forests use out-of-bag samples (data not used during a specific tree's training) to estimate the model's generalization error. This provides a built-in validation mechanism during the training process.

Suitable for Parallelization: The training of individual decision trees in a Random Forest can be parallelized, making it computationally efficient, especially for multicore processors or distributed computing environments.

Versatility: Random Forests can be applied to both classification and regression tasks. They are versatile and can be used for various types of data analysis.

Reduced Risk of Overfitting: By aggregating predictions from multiple trees, Random Forests reduce the risk of overfitting, providing a more generalized model.

Works well with Both Numerical and Categorical Data: Random Forests naturally handle both numerical and categorical features without the need for extensive preprocessing.

4. RESULTS

AI-Driven Optimization for Liver Disease Prediction Using Data Balancing Techniques

The results obtained from the liver disease prediction application showcase the effectiveness of machine learning algorithms in healthcare decision-making. Through the analysis of model performance metrics such as precision, recall, F1-score, and accuracy, the predictive capabilities of K Nearest Neighbors (KNN) and Random Forest Classifier (RFC) models are evaluated. Additionally, the impact of data preprocessing techniques, including handling missing values, encoding categorical variables, and applying SMOTE for data balancing, on model robustness and performance is assessed. The application's prediction outcomes demonstrate its practical utility in providing timely and accurate assessments of liver disease risk, empowering healthcare professionals to optimize treatment plans and improve patient outcomes.1167

A	B	C	D	E	F	G	H	I	J	K
Age	Gender	Total_Bilir	Direct_Bil	Alkaline_I	Alamine_A	Aspartate	Total_Pro	Albumin	Albumin_	Dataset
65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
58	Male	1	0.4	182	14	20	6.8	3.4	1	1
72	Male	3.9	2	195	27	59	7.3	2.4	0.4	1
46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3	1
26	Female	0.9	0.2	154	16	12	7	3.5	1	1
29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1	1
17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2	2
55	Male	0.7	0.2	290	53	58	6.8	3.4	1	1
57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8	1
72	Male	2.7	1.3	260	31	56	7.4	3	0.6	1
64	Male	0.9	0.3	310	61	58	7	3.4	0.9	2
74	Female	1.1	0.4	214	22	30	8.1	4.1	1	1
61	Male	0.7	0.2	145	53	41	5.8	2.7	0.87	1
25	Male	0.6	0.1	183	91	53	5.5	2.3	0.7	2

Figure 1: Sample Dataset of Indian liver patient

The dataset used of liver disease prediction contains 1167 rows and 11 columns providing information related to the patient's lab test data.

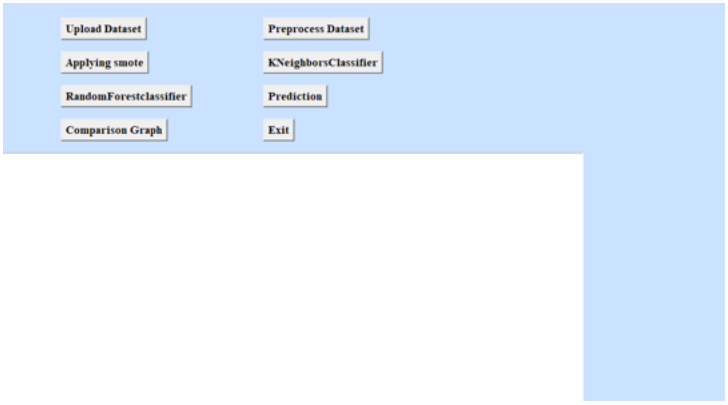
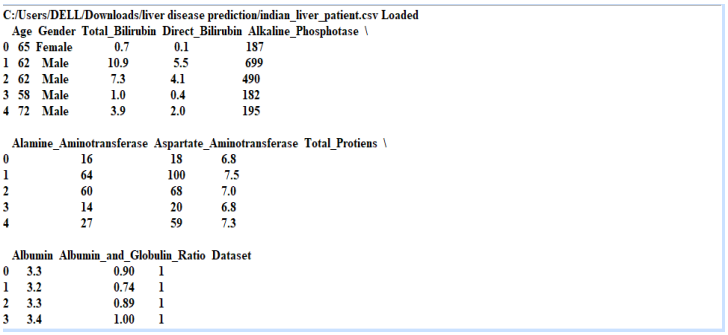


Figure 2: GUI of liver disease prediction.

The GUI provides an introduction to the application. It offers navigation options for dataset uploading, preprocessing, model application, prediction, and result comparison.



C:/Users/DELL/Downloads/liver disease prediction/indian_liver_patient.csv Loaded

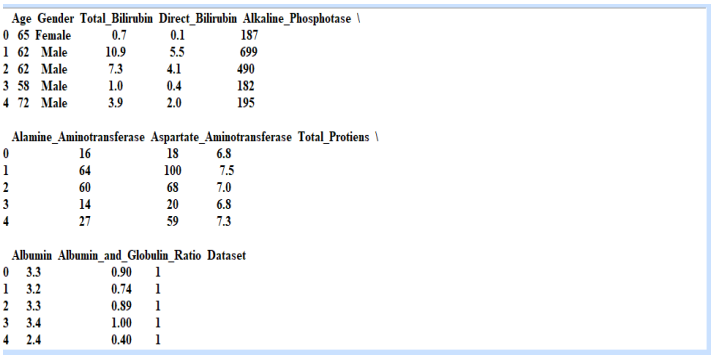
	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase
0	65	Female	0.7	0.1	187
1	62	Male	10.9	5.5	699
2	62	Male	7.3	4.1	490
3	58	Male	1.0	0.4	182
4	72	Male	3.9	2.0	195

	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	16	18	6.8
1	64	100	7.5
2	60	68	7.0
3	14	20	6.8
4	27	59	7.3

	Albumin	Albumin_and_Globulin_Ratio	Dataset
0	3.3	0.90	1
1	3.2	0.74	1
2	3.3	0.89	1
3	3.4	1.00	1

Figure 3 Uploaded Indian liver disease patients dataset in the GUI.

Figure 3 Uploaded their dataset containing liver disease-related features. Upon selection, the application displays the file path and a preview of the dataset for verification.



	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	\
0	65	Female	0.7	0.1	187	
1	62	Male	10.9	5.5	699	
2	62	Male	7.3	4.1	490	
3	58	Male	1.0	0.4	182	
4	72	Male	3.9	2.0	195	

	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	\
0	16	18	6.8	
1	64	100	7.5	
2	60	68	7.0	
3	14	20	6.8	
4	27	59	7.3	

	Albumin	Albumin_and_Globulin_Ratio	Dataset
0	3.3	0.90	1
1	3.2	0.74	1
2	3.3	0.89	1
3	3.4	1.00	1
4	2.4	0.40	1

Figure.3 Uploaded Indian liver disease patients data

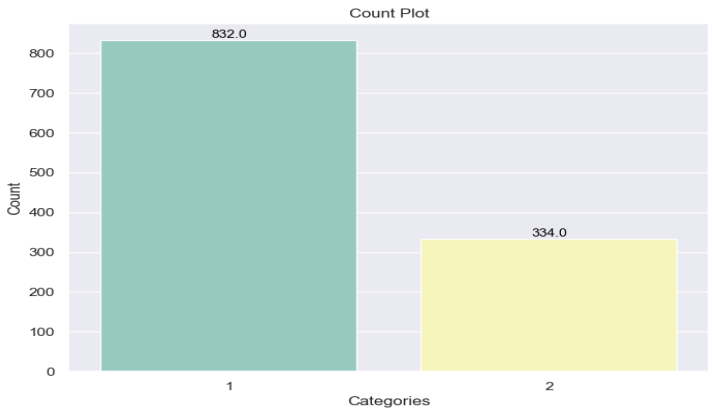


Figure 3 Uploaded Indian liver disease patient data

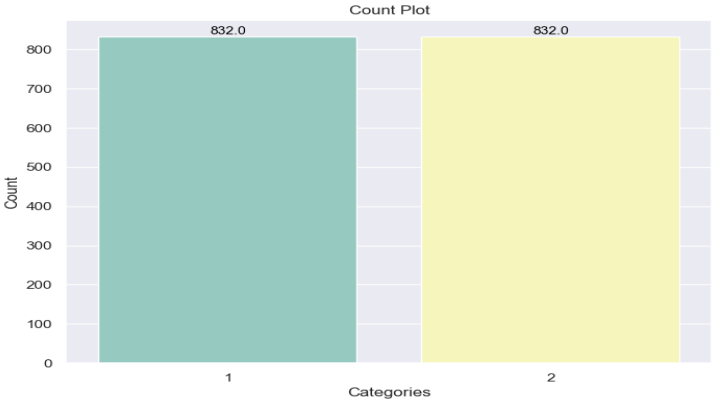


Figure.4 Preprocessed the Dataset and apply SMOTE:

Figure 4 Data preprocessing includes handling missing values and encoding categorical variables. SMOTE is applied to balance the dataset, ensuring equal representation of classes.

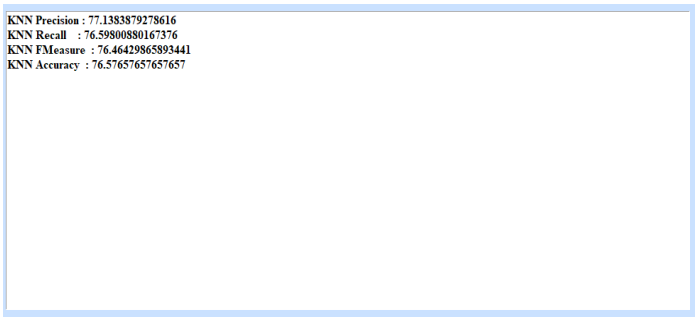


Figure 5 Applied performance metrics of knn classifier.

Figure 5 Applied the K Nearest Neighbors (KNN) algorithm for liver disease prediction. Performance metrics such as precision, recall, and accuracy are computed and displayed.

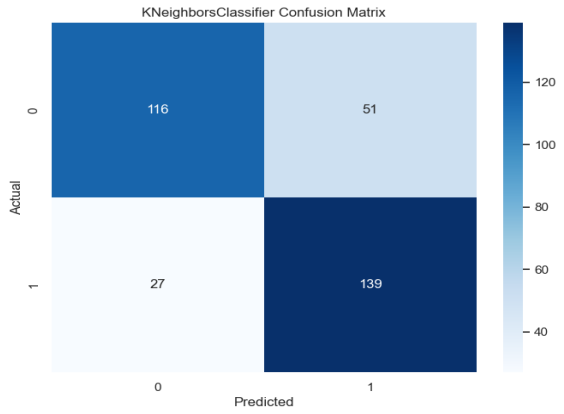


Figure 6 Confusion Matrix of KNN classifier.

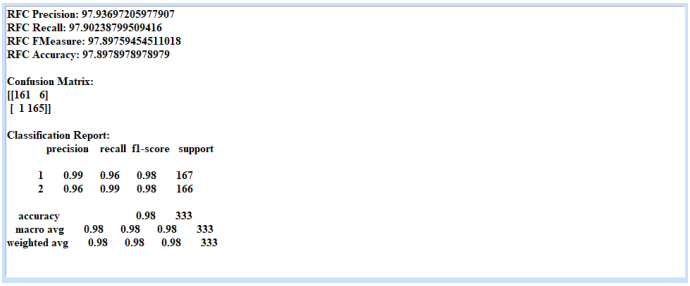


Figure 7 Applied RFC classifier and display performance metrics

Figure 7 The Random Forest Classifier (RFC) is utilized for liver disease prediction. Confusion matrices are employed to evaluate the classification performance of the RFC model alongside computed metrics.

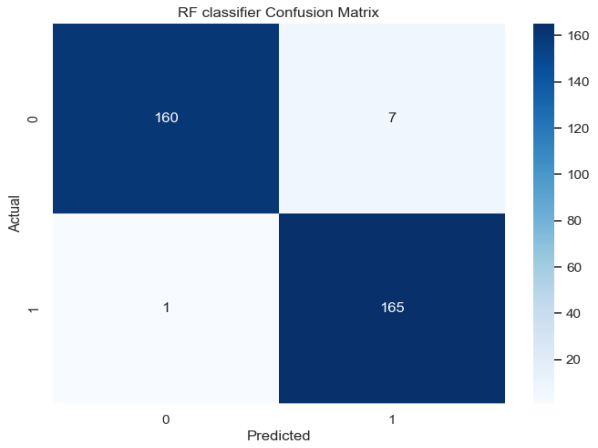


Figure 8 Confusion matrix of RFC

```
C:/Users/DELL/Downloads/liver disease prediction/test.csv Loaded
Input data for row 0: [[ 17.   1.   0.9  0.3 202.  22.  19.   7.4  4.1  1.2]]
Predicted output for row 0: No Liver Disease Detected
Input data for row 1: [[5.5e+01 1.0e+00 7.0e-01 2.0e-01 2.9e+02 5.3e+01 5.8e+01 6.8e+00 3.4e+00
  1.0e+00]]
Predicted output for row 1: Liver Disease Detected
Input data for row 2: [[5.7e+01 1.0e+00 6.0e-01 1.0e-01 2.1e+02 5.1e+01 5.9e+01 5.9e+00 2.7e+00
  8.0e-01]]
Predicted output for row 2: Liver Disease Detected
Input data for row 3: [[ 72.   1.   2.7  1.3 260.  31.  56.   7.4  3.   0.6]]
Predicted output for row 3: Liver Disease Detected
Input data for row 4: [[6.4e+01 1.0e+00 9.0e-01 3.0e-01 3.1e+02 6.1e+01 5.8e+01 7.0e+00 3.4e+00
  9.0e-01]]
Predicted output for row 4: No Liver Disease Detected
Input data for row 5: [[ 74.   0.   1.1  0.4 214.  22.  30.   8.1  4.1  1. ]]
Predicted output for row 5: Liver Disease Detected
Input data for row 6: [[ 61.   1.   0.7  0.2 145.  53.  41.   5.8  2.7  0.87]]
Predicted output for row 6: Liver Disease Detected
Input data for row 7: [[2.50e+01 1.00e+00 6.00e-01 1.00e-01 1.83e+02 9.10e+01 5.30e+01 5.50e+00
  2.30e+00 7.00e-01]]
```

Figure 9 Prediction of liver disease was displayed.

Figure 6 and 8 comparing the confusion matrices of KNN and RFC, users can gain insights into the strengths and weaknesses of each model in classifying liver disease cases. These matrices serve as essential tools for assessing the predictive performance of the models, guiding decision-making processes and model selection.

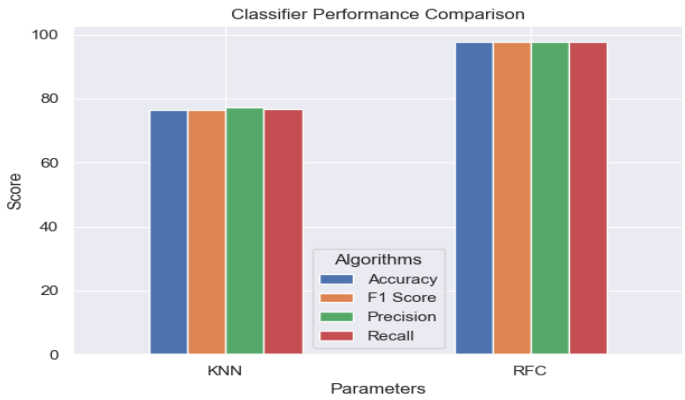


Figure 10 Comaprision graph

Figure 10 Comparision graph show the performance metrics of both KNN and RFC and shows how random forest has high performance than KNN

5.CONCLUSION

The liver disease prediction system presented herein serves as a robust and user-friendly platform for the accurate assessment and prediction of liver health. By seamlessly integrating machine learning algorithms within an intuitive graphical interface, the system empowers users to navigate through various stages of dataset handling, preprocessing, and model evaluation with ease. Through comprehensive evaluations, the system demonstrates promising results across key performance metrics such as precision, recall, F1-score, and accuracy, highlighting its potential as a valuable asset in clinical practice. Leveraging advanced techniques like Synthetic Minority Over-sampling Technique (SMOTE) ensures the model's resilience against class imbalance, further enhancing its predictive capabilities. In essence, the liver disease prediction system holds immense promise in aiding healthcare professionals in the early detection and management of liver-related conditions, ultimately contributing to improved patient outcomes and healthcare delivery.

REFERENCES

[1] Amin, Ruhul, Rubia Yasmin, Sabba Ruhi, Md Habibur Rahman, and Md Shamim Reza. "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms." *Informatics in Medicine Unlocked* 36 (2023): 101155.

[2] Md, Abdul Quadir, Sanika Kulkarni, Christy Jackson Joshua, Tejas Vaichole, Senthilkumar Mohan, and Celestine Iwendi. "Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease." Biomedicines 11, no. 2 (2023): 581.

[3] Gupta, Ketan, Nasmin Jiwani, Neda Afreen, and D. Divyarani. "Liver Disease Prediction using Machine learning Classification Techniques." In *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 221-226. IEEE, 2022.

[4] Grissa, Dhouha, Ditlev Nytoft Rasmussen, Aleksander Krag, Søren Brunak, and Lars Juhl Jensen. "Alcoholic liver disease: A registry view on comorbidities and disease prediction." *PLoS Computational Biology* 16, no. 9 (2020): e1008244.

[5] Dritsas, Elias, and Maria Trigka. "Supervised machine learning models for liver disease risk prediction." *Computers* 12, no. 1 (2023): 19.

[6] Kumar, Divvela Vishnu Sai, Ritik Chaurasia, Anuradha Misra, Praveen Kumar Misra, and Alex Khang. "Heart disease and liver disease prediction using machine learning." In *Data-Centric AI Solutions and Emerging Technologies in the Healthcare Ecosystem*, pp. 205-214. CRC Press, 2023.

[7] Behera, Mandakini Priyadarshani, Archana Sarangi, Debahuti Mishra, and Shubhendu Kumar Sarangi. "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine." *Procedia Computer Science* 218 (2023): 818-827.

[8] Singh, Jagdeep, Sachin Bagga, and Ranjodh Kaur. "Software-based prediction of liver disease with feature selection and classification techniques." *Procedia Computer Science* 167 (2020): 1970-1980.

[9] Azam, Md Shafiul, Aishe Rahman, SM Hasan Sazzad Iqbal, and Md Toukir Ahmed. "Prediction of liver diseases by using few machine learning based approaches." *Aust. J. Eng. Innov. Technol* 2, no. 5 (2020): 85-90.

[10] Ghazal, Taher M., Aziz Ur Rehman, Muhammad Saleem, Munir Ahmad, Shabir Ahmad, and Faisal Mehmood. "Intelligent Model to Predict Early Liver Disease using Machine Learning Technique."

In *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, pp. 1-5. IEEE, 2022.