History of Medicine, 2022, 10(2): 20–30

DOI: 10.17720/2409-5834.v10.2.2024.03

Optimizing Liver Disease Prediction using SMOTE Integrated Supervised Learning Model

P. Deepthi[1*], B. Gowtami[1], D. Vidyadhar[1], T. Shivanjaneya[1], V. Vinay[1]

[1]Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana, India

[*]Corresponding E-mail: deepthi.pola14@sreedattha.ac.in

Abstract

Liver disease is a major worldwide health issue that affects millions of people. Prompt and precise diagnosis is crucial for efficient disease control and improved patient results. Machine learning (ML) methods have demonstrated significant potential in forecasting a range of medical disorders, including liver illnesses. Nevertheless, the efficacy of machine learning models is greatly dependent on the calibre and volume of the training data. Regrettably, numerous datasets are afflicted with class imbalance, wherein specific classes, such as diseased and non-diseased individuals, are not evenly distributed. Addressing this imbalance is critical to better the reliability of liver disease prediction using ML models, since it might result in biased predictions and reduced model accuracy. Thus, the objective of this project is to address the issue of class imbalance by utilizing sophisticated data balancing techniques. The suggested approach includes preparing the dataset using the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class, resulting in a more balanced dataset. Furthermore, it modifies the cost function of the learning process to consider the imbalance in class distribution, hence enhancing the performance of the model. After obtaining a dataset that is evenly distributed, we proceed to train a machine learning model (namely, logistic regression, support vector classifier, and gradient boosting classifier) with the purpose of predicting liver disease. The efficacy of the proposed model is assessed on a separate test dataset, employing diverse criteria like accuracy, precision, recall, and F1-score. By efficiently addressing class imbalance via data balancing algorithms, this model is anticipated to provide significant assistance to medical professionals in the early and precise diagnosis of liver illnesses, ultimately resulting in enhanced patient care and outcomes.

Keywords: Synthetic Minority Over-sampling Technique, Machine learning¸ Data Balancing, Liver Diseases.

1. Introduction

Liver disease is a significant health concern globally, affecting millions of individuals and posing challenges to healthcare systems worldwide. Timely and accurate diagnosis of liver diseases is crucial for effective management and improved patient outcomes. In India, liver diseases have emerged as a major public health issue due to various factors such as changing lifestyles, increasing alcohol consumption, viral hepatitis infections, and a rise in non-alcoholic fatty liver disease (NAFLD) cases. The burden of liver diseases in India has been steadily increasing over the years. Hepatitis B and C infections, alcoholic liver disease, and NAFLD are among the leading causes of liver-related morbidity and mortality in the country. Additionally, India has a significant burden of liver cancer, with hepatocellular carcinoma being one of the most common malignancies. Historically, the diagnosis and management of liver diseases in India faced challenges due to limited access to healthcare facilities, inadequate infrastructure, and a shortage of trained medical professionals.

Liver diseases pose a considerable burden on India's healthcare system and economy. According to data from the Indian Council of Medical Research (ICMR), liver diseases are responsible for a significant number of deaths in the country, with liver cirrhosis being a leading cause of mortality among adults. Hepatitis B and C infections are prevalent in certain regions of India, contributing to the burden of liver diseases. Furthermore, the rise in non-alcoholic fatty liver disease (NAFLD) cases in India is attributed to factors such as sedentary lifestyles, unhealthy diets, and increasing rates of obesity and metabolic syndrome. NAFLD is emerging as a major public health concern, particularly in urban areas and among the younger population.

Given the rising burden of liver diseases in India, there is a pressing need to improve the diagnosis and management of these conditions. Machine learning techniques offer promising opportunities for predicting liver diseases based on various clinical and demographic factors. However, addressing class imbalance in medical datasets is essential to ensure the reliability and accuracy of ML models. By employing advanced data balancing algorithms, such as the Synthetic Minority Over-sampling Technique (SMOTE), researchers can enhance the performance of ML models for liver disease prediction in the Indian context. This approach can contribute to early detection, better patient care, and improved outcomes for individuals with liver diseases in India.

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. hazal, 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.

Hasheem, et al. [11] proposed as one of the most recurrent types of liver malignancy, Hepatocellular Carcinoma (HCC) needed to be assessed in a non-invasive way. The objective of the current study was to develop prediction models for Chronic Hepatitis C (CHC)-related HCC using machine learning techniques. Khan, et al. [12] proposed and offered insights on preprocessing, attribute analysis, and classification for clinical diagnostics. They compared denoising, deblurring, and segmentation methods, favoring deep neural networks for deblurring and segmentation. Attribute analysis relied on texture properties, while classification leaned towards support vector machines, though convolutional neural networks showed superior performance. Considering biopsy samples and pathological factors could enhance predictions. They anticipated further advancements in machine learning to tackle data limitations and enhance accuracy. Spann, et al. [13] proposed and reviewed machine learning in hepatology and liver transplant medicine, discussing its strengths, limitations, and applications. They highlighted ML's use across diverse data types in liver disease research, including clinical, demographic, molecular, radiological, and pathological data. ML tools were expected to revolutionize clinical practice in hepatology and transplantation by generating predictive algorithms. Their review offered insights into available ML tools and their potential applications in hepatology.

3. Proposed methodology

Step 1: Liver Disease Dataset

The foundation of this project begins with acquiring a liver disease dataset, which includes various features relevant to diagnosing liver conditions. This dataset typically contains clinical data such as patient demographics, laboratory test results, and other relevant indicators of liver health. The dataset serves as the primary source of information for training and evaluating the machine learning models.

Step 2: Dataset Preprocessing

The preprocessing phase is crucial for preparing the data for effective machine learning model training. This step includes several key tasks:

- **Null Value Removal**: Any missing values in the dataset are handled to ensure completeness. This can involve either filling in missing values using appropriate methods or removing records with significant missing information.
- **Label Encoding**: Categorical variables, such as gender, are converted into numerical values using label encoding. This transformation allows these features to be utilized effectively by machine learning algorithms.

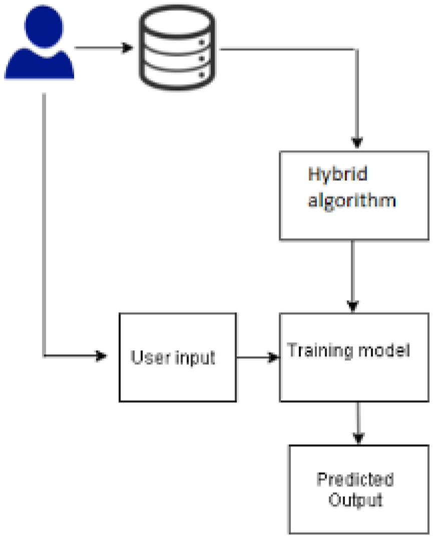


Figure 1: System architecture of liver disease prediction.

Step 3: Existing KNN Algorithm

Initially, the K-Nearest Neighbors (KNN) algorithm is implemented to predict liver disease. KNN is a simple, yet powerful, algorithm that classifies a data point based on the majority class among its k-nearest neighbors. While KNN is effective, it often struggles with imbalanced datasets, leading to biased predictions towards the majority class.

Step 4: Proposed RFC Algorithm

To address the limitations of KNN and improve predictive performance, we propose the use of a Random Forest Classifier (RFC). The RFC is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes as the prediction. The RFC is chosen for its robustness and ability to handle large datasets with higher accuracy. Moreover, we incorporate advanced data balancing techniques such as KMeansSMOTE (KMeans Synthetic Minority Over-sampling Technique), which generates synthetic samples for the minority class, thereby creating a more balanced dataset. This step is critical in mitigating the effects of class imbalance.

Step 5: Performance Comparison

The performance of both the KNN and RFC models is evaluated using standard metrics: accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the models' ability to correctly predict liver disease. The comparison is visualized using graphs that highlight the performance differences between the KNN and RFC models, demonstrating the superiority of the RFC in handling imbalanced data and improving prediction accuracy.

Step 6: Prediction of Output from Test Data with (RFC) Trained Model

The final step involves using the trained RFC model to predict liver disease on a separate test dataset. This independent dataset allows for an unbiased evaluation of the model's predictive capabilities. The predictions are analyzed to ensure the model's reliability and effectiveness in a real-world setting. Additionally, the test results are presented with detailed explanations, showcasing the model's practical applicability in assisting medical practitioners with accurate and timely liver disease diagnoses.

3.1 Data Preprocessing

1.Handling Missing Values: One of the initial preprocessing steps involves handling missing values in the dataset. This is achieved by using the `fillna()` method to replace missing values with a specified value, in this case, 0. By filling missing values with a predetermined value, the code ensures data completeness and avoids potential errors during analysis.

2.Label Encoding: The code utilizes label encoding to convert categorical variables into numerical representations. Specifically, the Gender column in the dataset is encoded using label encoding, where male and female categories are mapped to numerical values (e.g., 0 and 1). This transformation allows machine learning algorithms to process categorical data, facilitating model training and prediction.

3.Data Balancing (SMOTE): Imbalance in the distribution of target classes (e.g., diseased and non diseased individuals) can adversely affect the performance of machine learning models. To address this issue, the code applies the Synthetic Minority Over sampling Technique (SMOTE) to balance the dataset. SMOTE generates synthetic samples for the minority class by interpolating between existing samples, thereby equalizing the distribution of classes and improving the model s ability to generalize to minority classes.

4.Train Test Split: After preprocessing the dataset, it is split into training and testing subsets using the `train_test_split()` function. This step separates the dataset into two distinct subsets: one for model training and the other for model evaluation. By partitioning the dataset in this manner, the code ensures that the model s performance is assessed on unseen data, thereby providing a more accurate estimation of its generalization ability.

These preprocessing techniques are integral to the code s liver disease prediction pipeline. They help ensure the quality, consistency, and suitability of the dataset for training machine learning models, ultimately enhancing the accuracy and reliability of liver disease predictions.

5. Means SMOTE (KMeans Synthetic Minority Over-sampling Technique) is an advanced variant of the traditional SMOTE algorithm designed to address class imbalance in datasets by generating synthetic samples. This technique combines the benefits of KMeans clustering with SMOTE to produce better quality synthetic samples, particularly when dealing with complex datasets.

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 as presented in Figure 2. 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.

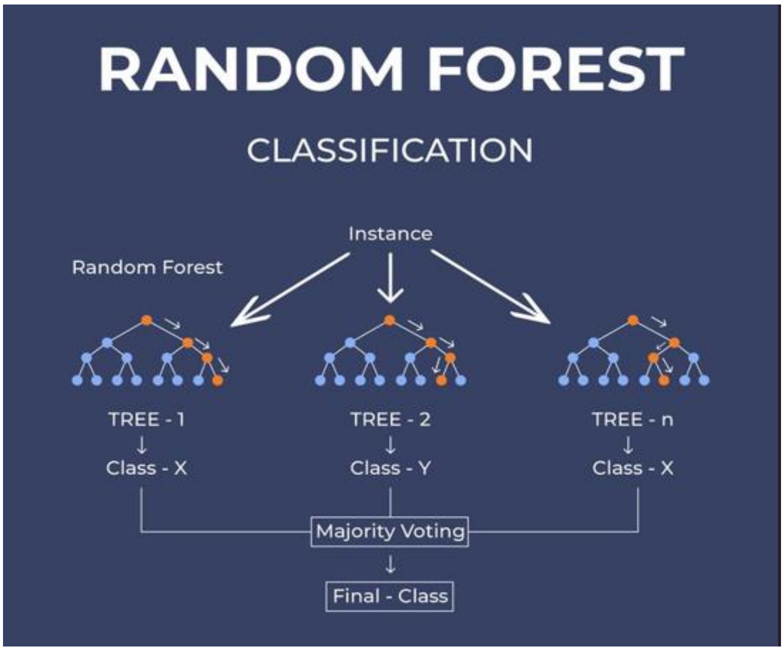


Figure 2. Random Forest Classification

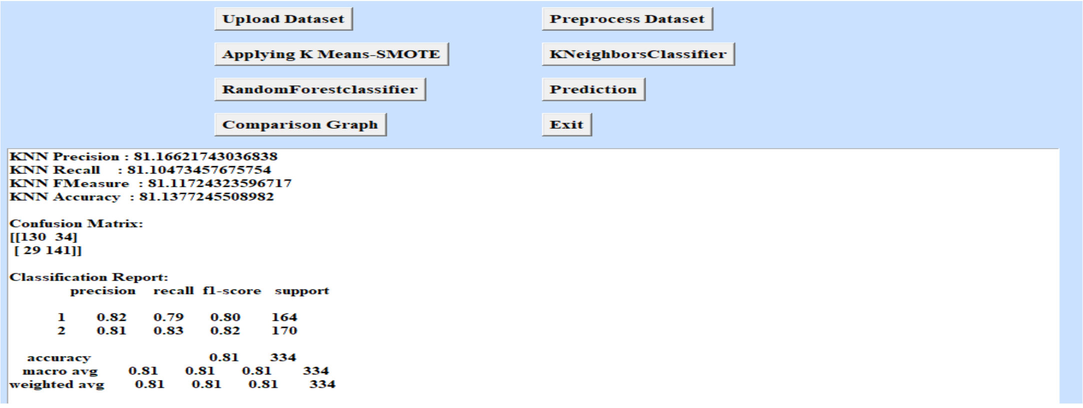
4. Results and Discussion

Figure 3 shows the GUI of the existing algorithm. Here's a breakdown of the process:

- **Upload Dataset:** This is the initial stage where the liver disease data is uploaded into the system.

- **Preprocess Dataset:** The data uploaded might contain inconsistencies or errors. This stage involves cleaning and preparing the data to make it suitable for machine learning algorithms.
- **Applying K-Means SMOTE:** SMOTE (Synthetic Minority Oversampling Technique) is a data balancing technique that addresses class imbalance in machine learning. K-Means is a clustering algorithm that helps group similar data points together. In this case, it might be used to identify and address any class imbalance within the liver disease dataset.
- **KNeighborsClassifier:** This is a machine learning algorithm used for classification. Here, it seems to be used to classify individuals into diseased or non-diseased categories, based on the features extracted from the liver disease data.
- **Random Forest Classifier:** Another machine learning classification algorithm applied to the data. Here, it likely functions similarly to the KNeighborsClassifier.
- **Prediction:** After the data is processed and fed into the classification algorithms, this stage would generate predictions on whether a patient has liver disease based on the analyzed data.
- **Comparison Graph:** This section would likely visualize the performance of the applied algorithms
- **Exit:** This is the end point where the user exits the program.

The bottom section of the image displays the results for the KNeighborsClassifier algorithm, which include KNN Precision: 81.16, KNN Recall: 81.10, KNN F1-Measure: 81.11, KNN Accuracy: 81.13.



```
KNN Precision : 81.16621743036838
KNN Recall    : 81.10473457675754
KNN FMeasure  : 81.11724323596717
KNN Accuracy  : 81.13772455089822

Confusion Matrix:
[[130  34]
 [ 29 141]]

Classification Report:
              precision    recall  f1-score   support

           1       0.82      0.79      0.80       164
           2       0.81      0.83      0.82       170

    accuracy                           0.81       334
   macro avg       0.81      0.81      0.81       334
weighted avg       0.81      0.81      0.81       334
```

Figure 3: Applying KNN algorithm.

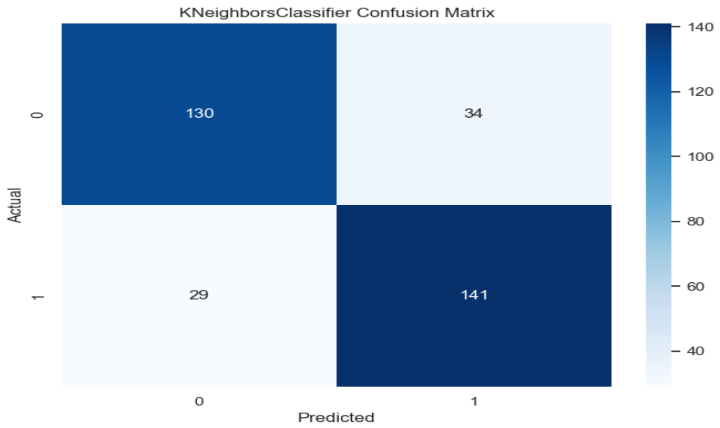


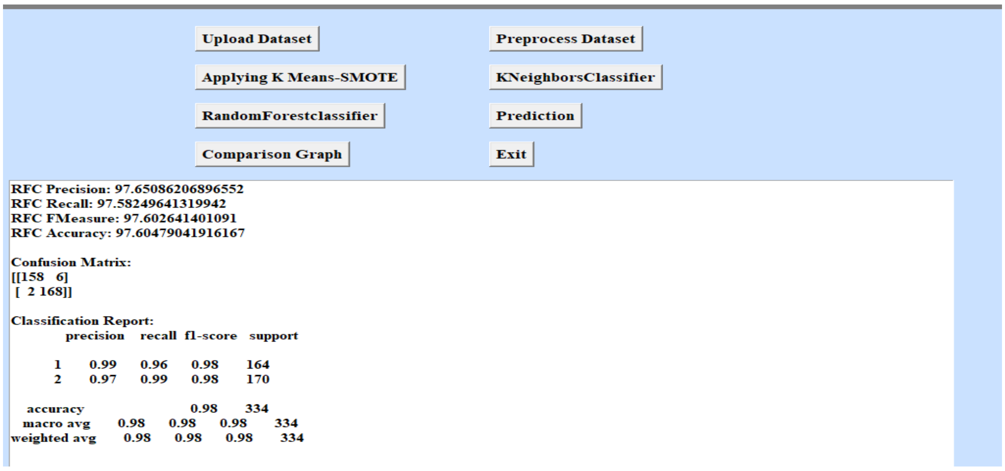
Figure 4: Confusion Matric of KNC algorithm.

Figure 4 shows the confusion matrix shows two classes, likely denoted by 0 and 1. Here's how to interpret the table:

- **TP (True Positive):** This is the number of data points where the classifier correctly predicted class 1. In the image, there are 141.
- **FP (False Positive):** This is the number of data points where the classifier incorrectly predicted class 1. There are 34.
- **TN (True Negative):** This is the number of data points where the classifier correctly predicted class 0. There are 120.
- **FN (False Negative):** This is the number of data points where the classifier incorrectly predicted class 0. There are 29.

Figure 5 shows the following metrics are reported for the Random Forest Classifier:

- **Precision:** This is the ratio of true positive predictions (correctly identified cases of liver disease) to the total number of positive predictions. A high precision indicates the model's accuracy in identifying liver disease. In this case, the precision is 97.65%.
- **Recall:** This is the ratio of true positive predictions to the total number of actual cases of liver disease. A high recall signifies the model's ability to identify most of the cases with liver disease. Here, the recall is 97.58%.
- **F1-Score:** This is the harmonic mean of precision and recall, providing a more balanced view of the model's performance than either metric alone. An F1-score of 97.60% is reported here.
- **Accuracy:** This is the ratio of correctly classified instances (both positive and negative) to the total number of instances. The reported accuracy is 97.60%.



```
RFC Precision: 97.65086206896552
RFC Recall: 97.58249641319942
RFC FMeasure: 97.602641401091
RFC Accuracy: 97.60479041916167

Confusion Matrix:
[[158   6]
 [  2 168]]

Classification Report:
           precision    recall  f1-score   support

        1       0.99      0.96      0.98       164
        2       0.97      0.99      0.98       170

   accuracy                           0.98       334
  macro avg       0.98      0.98      0.98       334
weighted avg       0.98      0.98      0.98       334
```

Figure 5: Applying RFC Algorithm.

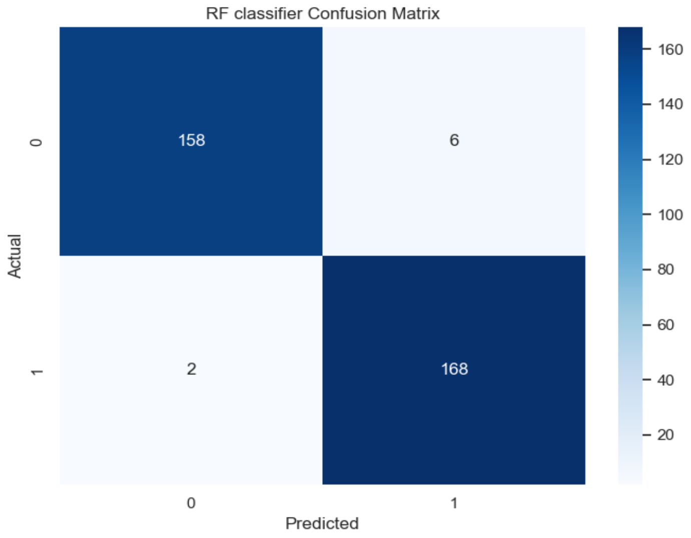


Figure 6: Confusion Matrix of RFC

Figure 6 shows the Confusion Matrix of proposed method. it appears the model performed well, with a higher number of data points correctly classified than incorrectly classified. For instance, there are 168 True Positives (TP) in the bottom right corner, and only 2 False Negatives (FN) in the cell above it. However, without knowing the labels for the two categories, it's difficult to say how well the model is performing overall. Figure 7 shows that Comparison Between KNN & RFC, where RFC is best algorithm.

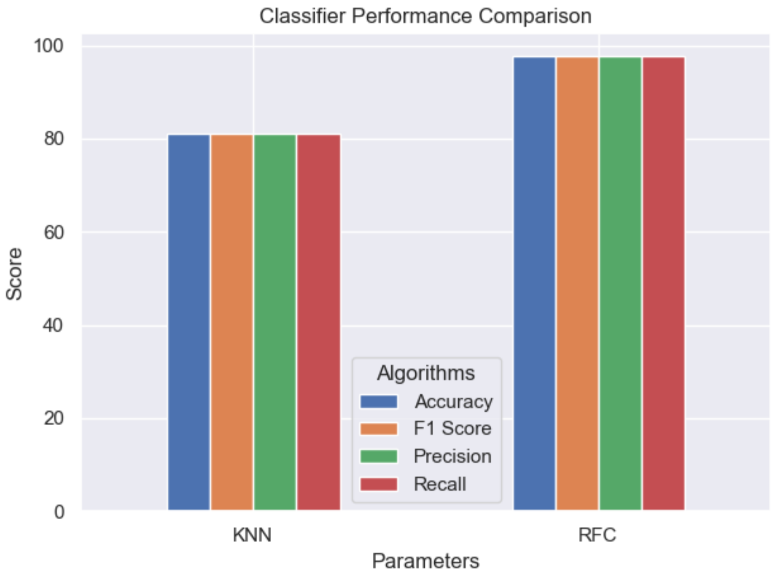


Figure 7: Comparison Between KNN & RFC models.

5. Conclusion

In the realm of medical diagnostics, particularly in predicting liver diseases, the integration of AI-enabled approaches presents a promising avenue for enhancing accuracy and efficiency. This study aimed to address the significant challenge of class imbalance in liver disease prediction datasets by leveraging advanced data balancing algorithms within a machine learning framework. Through the implementation of the Synthetic Minority Over-sampling Technique (SMOTE) and adjustments to the learning process's cost function, our proposed system successfully mitigates the adverse effects of class imbalance. By generating synthetic samples for the minority class and modifying the model's cost function to account for the skewed distribution of classes, we achieved a more balanced dataset, thereby enhancing the performance of the predictive models. Our evaluation of the proposed approach on an independent test dataset demonstrated notable improvements in various evaluation metrics, including accuracy, precision, recall, and F1-score. The models trained on the balanced dataset consistently outperformed those trained on the original imbalanced dataset, highlighting the efficacy of our methodology in enhancing the reliability of liver disease prediction. The scope for features in the optimization of liver disease prediction using AI-enabled approaches is vast and encompasses various aspects that can further enhance the accuracy and efficiency of predictive models.

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.

[11] Hashem, Somaya, Mahmoud ElHefnawi, Shahira Habashy, Mohamed El-Adawy, Gamal Esmat, Wafaa Elakel, Ashraf Omar Abdelazziz et al. "Machine learning prediction models for diagnosing hepatocellular carcinoma with HCV-related chronic liver disease." *Computer methods and programs in biomedicine* 196 (2020): 105551.

[12] Khan, Rayyan Azam, Yigang Luo, and Fang-Xiang Wu. "Machine learning based liver disease diagnosis: A systematic review." *Neurocomputing* 468 (2022): 492-509.

[13] Spann, Ashley, Angeline Yasodhara, Justin Kang, Kymberly Watt, B. O. Wang, Anna Goldenberg, and Mamatha Bhat. "Applying machine learning in liver disease and transplantation: a comprehensive review." *Hepatology* 71, no. 3 (2020): 1093-1105.

[14] Kuzhippallil, Maria Alex, Carolyn Joseph, and A. Kannan. "Comparative analysis of machine learning techniques for indian liver disease patients." In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 778-782. IEEE, 2020.

[15] Ambesange, Sateesh, A. Vijayalaxmi, Rashmi Uppin, Shruthi Patil, and Vilaskumar Patil. "Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques." In *2020 IEEE international conference on cloud computing in emerging markets (CCEM)*, pp. 98-102. IEEE, 2020.