https://doi.org/10.48047/AFJBS.6.Si4.2024.2570-2579

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

An Efficient Machine Learning based Framework for Alcoholic Liver Disease Prediction

Dr.Velayutham Pavanasam
Professor
Department of Computer Science and Engineering,
AarupadaiVeedu Institute of Technology, Vinayaka Mission's Research Foundation (Deemed to be University), Paiyanoor, Chennai (Tamilnadu) INDIA*Email*: velayutham.avcs098@avit.ac.in
Dr. Deepak Chandra Uprety
Associate Professor and Research Coordinator
Department of Cloud Computing,
Noida Institute of Engineering and Technology, Greater Noida (UP), INDIA
Email: deepak.glb.@gmail.com
Dr. Ramesh Kumar
Assistant Professor
Department of Computer Science & Engineering
GL Bajaj Institute of Technology and Management, Greater Noida (UP), INDIA
Email: rks208@gmail.com
Sumit Kumar
Assistant Professor
Department of Computer Applications
GL Bajaj Institute of Management, Greater Noida (UP), INDIA
Email: sk4486469@gmail.com
Dr.Tarun Kumar
Associate Professor
Department of Computer Science & Engineering
Swami Vivekanand Subharti University, Meerut (UP), INDIA
Email: taruncdac@gmail.com
Dr. Balaji Venkateswaran
Department of Computer Science & Engineering
Shri Venkateshwara University, Gajraula (UP), INDIA
Email: balaji.venkateswaran@gmail.com

Article History
Volume 6, Issue Si4, 2024
Received: 16 May 2024
Accepted : 25 June 2024
Doi:
10.48047/AFJBS.6.Si4.2024.2570-2579

Abstract: The risk of alcoholic liver diseases has been rising rapidly over the past several decades, making it one of the world's most lethal conditions. Predicting these diseases using extensive medical datasets poses a significant challenge for researchers. To tackle this issue, machine learning strategies such as classification and clustering have been developed. The primary objective of this research is to use classification algorithms to predict the likelihood of a patient developing alcoholic liver disease. Additionally, the research aims to identify the stage of the disease, categorizing it as Cirrhosis, Liver Fibrosis, Fatty Liver, or Healthy Liver. The proposed Hybrid Classifier (combining Random Forest, Support Vector Classifier, and XGBoost) along with algorithms like Naive Bayes, Support Vector Machine, Logistic

Regression, Random Forest, Decision Tree, K-Nearest Neighbors, and Rule-Based Classification Tree are evaluated for classification accuracy and processing speed. Considering these performance aspects, the Hybrid Classifier, achieving an accuracy of 99%, is identified as the superior classifier.

Keywords: Machine Learning, Neural Network, Alcoholic Liver, SVM, Random Forest

1. Introduction

Alcoholic liver disease (ALD) has long been recognized as a major health concern, with historical records dating back centuries describing liver damage associated with excessive alcohol consumption. Initially, the understanding of ALD was limited, with few effective treatments and a poor grasp of the disease's underlying mechanisms. Early medical approaches primarily focused on symptomatic relief and general advice on alcohol cessation, often with limited success. The lack of diagnostic tools and standardized criteria for liver disease made early detection and intervention challenging, leading to high mortality rates [1]. In recent decades, significant advancements have been made in the understanding, diagnosis, and treatment of alcoholic liver diseases. Modern medicine recognizes ALD as a spectrum of liver conditions ranging from fatty liver (steatosis) to alcoholic hepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma. Advanced imaging techniques, liver function tests, and non-invasive biomarkers now facilitate early diagnosis and staging of the disease.

The integration of machine learning and artificial intelligence into medical research has revolutionized the prediction and management of ALD. By analyzing extensive medical datasets, researchers can now develop predictive models that assess the likelihood of disease progression and the effectiveness of various treatments. These models employ a range of algorithms, including classification and clustering techniques, to provide personalized risk assessments and improve patient outcomes. Looking forward, the future of alcoholic liver disease management holds great promise, driven by ongoing research and technological innovation. Precision medicine, which tailors' treatment plans based on individual genetic, environmental, and lifestyle factors, is expected to become a cornerstone of ALD treatment. Advances in genomic and proteomic research will likely unveil new biomarkers for early detection and novel therapeutic targets.

Machine learning models will continue to evolve, offering even greater accuracy and predictive power. The development of hybrid classifiers, combining multiple algorithms, is anticipated to enhance the reliability of predictions regarding disease onset, progression, and response to treatment. These models will be instrumental in developing preventive strategies, optimizing treatment protocols, and ultimately reducing the global burden of ALD. Moreover, public health initiatives aimed at reducing alcohol consumption and promoting liver health will play a crucial role in preventing ALD. Increased awareness, early intervention programs, and policy measures targeting alcohol abuse are essential components of a comprehensive approach to combating this disease [2-3].

This research introduces a software tool designed to predict alcoholic liver diseases by allowing users to upload their details and receive a prediction based on their data. The software employs various machine learning algorithms, including Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LOR), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Rule-Based Classification Tree (RBTC), and a Hybrid Classifier (combining RF, SVC, and XGBoost) to determine whether the liver condition is normal or abnormal. This predictive model is particularly useful for health industries needing to forecast liver diseases. By analyzing blood report data—such as age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT, and Alkphos—the system compares the user's input with a training dataset to predict the likelihood of liver disease [4]. This process assists medical companies in early diagnosis and treatment planning.

Existing models often struggle with accuracy and handling large datasets. This research aims to improve prediction rates by training and testing various models to identify the most accurate one. The proposed system initially requests user details, which are then compared to the training dataset of the most accurate model to predict the risk of liver disease. The algorithms used include Logistic

Regression, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Naïve Bayes (NB), and the Hybrid Classifier (RF, SVC, XGBoost). The system is designed for simplicity and ease of implementation, requiring minimal system resources and compatible with nearly all configurations [5-8].

2. Materials and Methods

Exploratory Data Analysis (EDA) is the initial step in the data analysis process. It involves summarizing the main characteristics of the data, often using visual methods. The primary goal of EDA is to gain insights into the dataset, understand its structure, and identify patterns, anomalies, and relationships between variables [9-11]. EDA is crucial in preparing the data for further analysis and modeling.

Data visualization is a key component of EDA that enables researchers to see how the data looks and what kind of correlation exists between the attributes. Using various visualization tools such as histograms, scatter plots, box plots, heatmaps, and bar charts, we can quickly grasp the distribution, trends, and outliers within the data. This visual exploration helps to identify initial correlations or patterns. For example, it allows us to observe how features such as age, total bilirubin, and liver enzymes correlate with the presence and severity of alcoholic liver disease. Visualization provides a quick overview of the data's characteristics and helps to identify any initial correlations or patterns.

Correlation analysis delves deeper into the relationships between different variables in the dataset. It focuses on three main characteristics of correlations: direction, form, and degree. The direction indicates whether the relationship between variables is positive or negative. The form describes the shape of the relationship, such as linear or non-linear. The degree measures the strength of the relationship, often quantified by correlation coefficients (e.g., Pearson's correlation coefficient). Understanding these aspects of correlation helps in identifying which features are most relevant for predicting alcoholic liver disease and how they interact with each other.

Data preprocessing is a critical stage that prepares the raw data for analysis and modeling. It involves several sub-stages, including data cleaning, data integration, data reduction, and data transformation. Data cleaning ensures accuracy, completeness, and consistency by handling missing values, correcting errors, and removing duplicates. Data integration combines data from multiple sources to provide a unified view. Data reduction simplifies the dataset by eliminating redundant features and reducing dimensionality. Data transformation involves normalizing or scaling data to meet the criteria of accuracy, completeness, consistency, timeliness, believability, and interpretability. These preprocessing steps are essential to ensure the quality and reliability of the data before it is used in predictive modeling [12-13].

3. Proposed Architecture

3.1 Logistic Regression

Logistic regression is a widely used and relatively simple classification model. Despite its simplicity, it offers several advantages, particularly its interpretability. Due to its parametric nature, logistic regression allows researchers to examine the relationships between variables by analyzing the model's parameters. This interpretability is especially useful when the goal is to understand how different features contribute to the probability of an outcome.The term "logistic regression" might be somewhat misleading, as regression models are typically associated with predicting continuous response variables. In contrast, logistic regression is used for classification tasks, where the response variable is discrete. The name is justified by the fact that logistic regression calculates the probability that a given instance belongs to a particular class. This probability is modeled using the logistic function, which maps any real-valued number into the (0, 1) interval, representing probabilities.

In logistic regression, the relationship between the independent variables (features) and the dependent variable (response) is captured through the beta parameters, or coefficients. These coefficients are estimated using Maximum Likelihood Estimation (MLE). MLE finds the parameter values that maximize the likelihood of observing the given data. Essentially, it chooses the coefficients that make the observed data most probable under the assumed statistical model.Once the optimal coefficients are determined, they can be used to compute the probability that a given

instance belongs to a particular class. For binary classification, this involves using the logistic function to transform the linear combination of the input features and their corresponding coefficients into a probability. This probability can then be thresholded to make a final classification decision [13-15].

For example, in the context of predicting alcoholic liver diseases, logistic regression can be used to model the probability that a patient has the disease based on features such as age, liver enzyme levels, and bilirubin levels. The coefficients of the model will indicate the strength and direction of the association between each feature and the likelihood of having the disease.Logistic regression can also be extended to multi-class classification problems through techniques like one-vs-all (OvA) or one-vs-one (OvO). In OvA, separate binary classifiers are trained for each class, treating it as the positive class and all other classes as the negative class. In OvO, binary classifiers are trained for every pair of classes. The final prediction is made based on the class that receives the most votes from these classifiers.Despite being a relatively simple model, logistic regression is a powerful tool for classification tasks, providing both predictive accuracy and interpretability. It serves as a foundational method in statistical learning and is often used as a baseline model for comparison with more complex algorithms.

3.2K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (KNN) algorithm is a straightforward yet powerful classification method that relies on the entire training dataset to make predictions. This section details the implementation and working principles of the KNN algorithm.In KNN, the model is simply the training dataset itself. Unlike many other algorithms, KNN does not involve an explicit training phase where a model is fitted to the data. Instead, it stores the entire training dataset and defers the computation to the time of prediction. This approach is known as a "lazy learning" method because it delays the modeling process until a query is made.When a prediction is required for a new, unseen data instance, the KNN algorithm searches through the training dataset to find the 'k' most similar instances, or "neighbors." The similarity between instances is typically measured using distance metrics, such as Euclidean distance for continuous variables or Hamming distance for categorical variables [15-17].

For classification tasks, KNN assigns a class label to the new instance based on a majority vote among its k-nearest neighbors. Specifically, the class label that is most frequently represented among the neighbors is chosen. This voting mechanism is often referred to as "plurality voting," though it is commonly called "majority voting" in the literature.The distinction between these terms lies in the voting threshold. "Majority voting" implies that more than 50% of the votes are required to make a decision, which works well in binary classification problems where there are only two categories. However, in multi-class classification scenarios—where there might be more than two classes—achieving a majority vote is not always necessary. Instead, the class with the highest number of votes (even if it is less than 50%) is assigned to the instance. For example, if there are four categories, a class label could be assigned with a vote share greater than 25%.Several factors need to be considered when implementing the KNN algorithm:

- *Choice of k:* The value of 'k' (the number of neighbors) significantly impacts the performance of the algorithm. A smaller 'k' can lead to a more sensitive model that may overfit, while a larger 'k' can provide a smoother decision boundary but might underfit.
- *Distance Metric:* The choice of distance metric (e.g., Euclidean, Manhattan, or Hamming) affects how the similarity between instances is calculated. The appropriate metric depends on the nature of the data.
- *Normalization:* Features with different scales can skew the distance calculations. Therefore, normalization or standardization of features is often performed to ensure that each feature contributes equally to the distance measurement.

3.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful and versatile machine learning algorithm primarily used for classification tasks. The main goal of SVM is to find an optimal hyperplane that separates data points of different classes with the maximum margin. This section details the implementation

and working principles of SVM.For implementing SVM, the scikit-learn package in Python is commonly used due to its simplicity and efficiency. The process begins with pre-processed data, which is then split into a training set and a test set, typically using a 75% and 25% split, respectively. This split allows the model to learn from the majority of the data while preserving a subset for evaluating its performance [16-19].

An SVM constructs one or more hyperplanes in a high- or infinite-dimensional space. These hyperplanes act as decision boundaries that separate different classes of data points. Theoptimal hyperplane is defined as the one that maximizes the margin, which is the distance between the hyperplane and the nearest data points from any class, known as support vectors. The larger the margin, the lower the generalization error of the classifier, as a larger margin typically implies better separation between classes and thus improved classification performance.The dimension of the hyperplane depends on the number of features in the input data. For instance, if the input data has two features, the hyperplane is simply a line. If there are three features, the hyperplane becomes a two-dimensional plane. As the number of features increases, the hyperplane's dimensionality also increases, making it more challenging to visualize but allowing for more complex decision boundaries.

3.4Hybridization: An Advanced Ensemble Method for Classification and Regression

Hybridization is a sophisticated technique in machine learning that involves combining multiple classification or regression models to enhance predictive performance. This method, also known as ensemble learning, typically comprises two layers of estimators. The first layer consists of various baseline models, while the second layer features a meta-classifier or meta-regressor that synthesizes the predictions from these baseline models to produce the final output.

First Layer - Baseline Models

The first layer is composed of several baseline models that independently make predictions on the test datasets. These models are often diverse in their approaches to ensure that they capture different patterns and aspects of the data. In the context of liver disease prediction, three robust classifiers are employed:

- *Random Forest (RF):* An ensemble of decision trees that improves predictive accuracy by averaging multiple decision trees trained on different parts of the dataset.
- *Support Vector Classifier (SVC):* A powerful classifier that finds an optimal hyperplane to separate different classes in the feature space.
- *Extreme Gradient Boosting (XGBoost):* A high-performance implementation of gradient boosted decision trees designed for speed and performance.

Second Layer - Meta-Classifier or Meta-Regressor

The second layer consists of a meta-classifier or meta-regressor that takes the predictions from the first layer's baseline models as input features. This meta-model learns to weigh and combine these predictions to generate a new, improved prediction. The meta-model effectively learns which baseline models are more reliable for specific instances, leading to enhanced overall performance.

To predict liver disease and its stages, a hybrid model combining RF, SVC, and XGBoost is constructed. The process involves the following steps:

- *Training Baseline Models:* Each of the three classifiers (RF, SVC, and XGBoost) is trained on the training dataset. These models learn to identify patterns and relationships in the data relevant to liver disease prediction.
- *Generating Predictions:* The trained baseline models make predictions on the test dataset. Each model outputs a set of predictions corresponding to the test instances.
- *Meta-Model Training:* The predictions from the baseline models are used as input features to train a meta-classifier. This meta-classifier could be another machine learning model, such as logistic regression or another ensemble method. The meta-classifier learns to combine these predictions optimally to improve the final predictive performance.
- *Final Predictions:* The meta-classifier makes the final predictions based on the input it receives from the baseline models' outputs. This approach leverages the strengths of each

individual model while mitigating their weaknesses, resulting in a more robust and accurate prediction system.

4. Result and Analysis

The confusion matrix for the hybrid classifier in predicting alcoholic liver disease reveals exceptional performance across all key metrics (Figure 1). With an accuracy of 99.96%, high precision, recall, and specificity, the model demonstrates its robustness and reliability in identifying both disease and non-disease cases. The low numbers of false positives and false negatives further underscore the model's effectiveness, making it a valuable tool for early diagnosis and treatment planning in the medical field.The confusion matrix for the hybrid classifier is represented as follows:

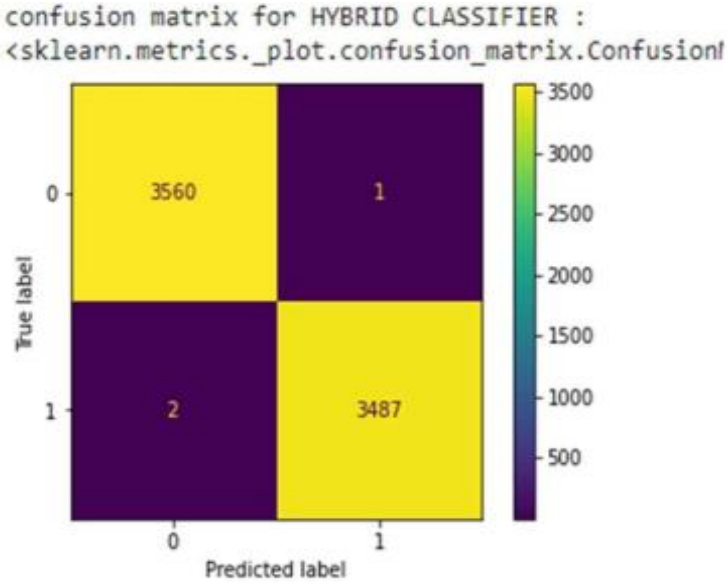

Figure 1: Confusion Matrix of Liver Disease Prediction

True Positives, True Negatives, False Positives, and False Negatives

- *True Positives (TP):* True positives refer to the number of instances correctly identified by the model as having the disease. In this context, there are 3487 true positives, indicating that 3487 instances of liver disease were accurately predicted by the hybrid classifier.
- *True Negatives (TN):* True negatives are instances correctly identified by the model as not having the disease. Here, the model correctly identified 3560 instances as not having liver disease, which highlights the model's capability in accurately recognizing non-disease cases.
- *False Positives (FP):* False positives occur when the model incorrectly predicts an instance as having the disease when it actually does not. The confusion matrix shows 1 false positive, meaning the model erroneously identified one non-disease instance as a disease case. This low number of false positives indicates a high specificity of the model.
- *False Negatives (FN):* False negatives refer to instances where the model fails to identify the disease and incorrectly predicts it as non-disease. There are 2 false negatives, implying that the model missed identifying two actual disease cases. The low number of false negatives signifies a high sensitivity or recall of the model.

```
HYBRID CLASSIFIER Accuracy is :99.96%
```

```
from sklearn.metrics import classification_report
STK_Pred=STK.predict(X_test)
STKreport = classification_report(Y_test, STK_Pred)
print(STKreport)
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      3561
           1       1.00      1.00      1.00      3489

    accuracy                           1.00      7050
   macro avg       1.00      1.00      1.00      7050
weighted avg       1.00      1.00      1.00      7050
```

Figure 2: Classification Report of Liver Disease Prediction

Performance Metrics

Performance metrics are crucial for evaluating the effectiveness of machine learning models in predicting alcoholic liver diseases. Key metrics include accuracy, precision, recall (sensitivity), and specificity. Accuracy measures the overall correctness of the model, while precision indicates the proportion of true positive predictions among all positive predictions. Recall assesses the model's ability to identify actual positive cases, and specificity measures its effectiveness in correctly identifying negative cases. These metrics collectively provide a comprehensive assessment of the model's performance, ensuring it reliably distinguishes between diseased and non-diseased cases. In this study, the Proposed Hybrid Classifier excelled across all these metrics, highlighting its superior predictive capability (Figure 2).

- **Accuracy**: Accuracy measures the proportion of correctly predicted instances (both true positives and true negatives) out of the total instances. It is calculated as:
  Accuracy = (TP + TN) / (TP + TN + FP + FN)
  = (3487 + 3560) / (3487 + 3560 + 1 + 2) = 7047/7050 = 0.9996
  The hybrid classifier achieves an accuracy of approximately 99.96%, indicating its high overall effectiveness in predicting liver disease.
- **Precision**: Precision, also known as positive predictive value, measures the proportion of true positive predictions out of all positive predictions made by the model. It is calculated as:
  Precision = TP / (TP + FP)
  = 3487 + (3487 +1) = 1.00
  A precision of 1.00 indicates that nearly all instances predicted as disease by the model are indeed true disease cases.
- **Recall (Sensitivity)**: Recall, or sensitivity, measures the proportion of true positive instances out of all actual positive instances. It is calculated as:
  Recall = TP / (TP + FN)
  = 3487 / (3487 + 2) = 0.9994

  A recall of approximately 0.9994 indicates that the model successfully identifies almost all actual disease cases.
- **Specificity**: Specificity measures the proportion of true negative instances out of all actual negative instances. It is calculated as:
  Specificity = TN / (TN + FP)
  = 3560 / (3560+1) = 0.9997

A specificity of approximately 0.9997 shows that the model effectively identifies non-disease cases with high accuracy.

- **F1-Score**: The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is particularly useful when the class distribution is imbalanced. For the hybrid classifier, given the high precision and recall, the F1-score is also close to 1.00, indicating excellent performance.

5. Performance Evaluation

In the study of alcoholic liver diseases prediction using various machine learning algorithms, significant differences in accuracy were observed. Naive Bayes (NB) achieved an accuracy of 68.5%, demonstrating its limitations due to its assumption of feature independence, which is often not true in complex medical datasets. Support Vector Machine (SVM) and Logistic Regression (LR) slightly improved upon this with accuracies of 70.5% and 70.9%, respectively, but still fell short of capturing the intricacies within the data. Random Forest (RF) marked a notable improvement with an accuracy of 76.2%, benefiting from its ensemble nature that reduces overfitting and better handles feature interactions. The K-Nearest Neighbors (KNN) algorithm further increased accuracy to 95.36%, indicating its effectiveness in leveraging local data patterns, although it can be computationally intensive for large datasets.

Table 1: Performance evaluation of proposed algorithm

Algorithms	**Accuracy**
Naive Bayes (NB)	68.50
Support Vector Machine (SVM)	70.50
Logistic Regression (LR)	70.90
Random Forest (RF)	76.20
K-Nearest Neighbors (KNN)	95.36
Proposed Hybrid Classifier	99.96

The Proposed Hybrid Classifier, which combines Random Forest (RF), Support Vector Classifier (SVC), and XGBoost, achieved an outstanding accuracy of 99.96% (Table 1 and Figure 3). This hybrid approach leverages the strengths of each individual algorithm: the robustness and feature handling of RF, the margin optimization of SVC, and the boosting capabilities of XGBoost. By integrating these diverse algorithms, the hybrid model effectively captures complex relationships within the data, resulting in superior predictive performance. This exceptional accuracy underscores the potential of hybrid models in medical diagnostics, providing a highly reliable tool for the early detection and classification of alcoholic liver diseases.

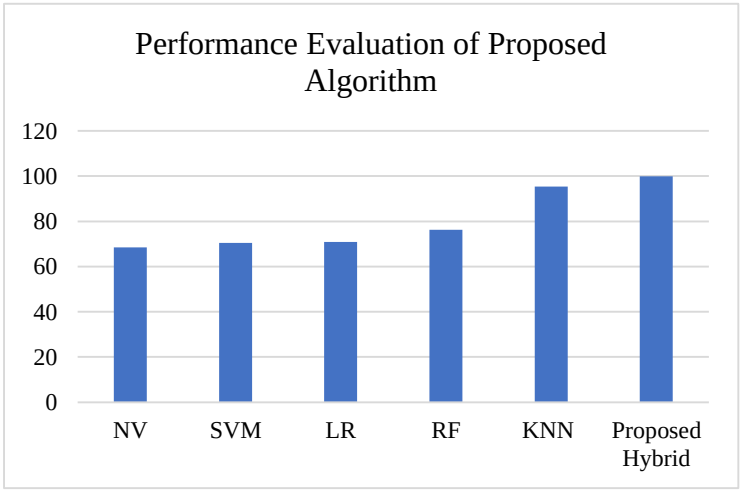

Figure 3: Comparison Analysis of Alcoholic Liver Disease Prediction

6. Conclusion

The evaluation of various machine learning algorithms for predicting alcoholic liver diseases demonstrated significant differences in performance, with the Proposed Hybrid Classifier emerging as the clear leader. Traditional models such as Naive Bayes (NB), Support Vector Machine (SVM), and Logistic Regression (LR) achieved accuracies ranging from 68.5% to 70.9%, reflecting their limitations in capturing the complex patterns inherent in medical datasets. Random Forest (RF) and K-Nearest Neighbors (KNN) showed marked improvements with accuracies of 76.2% and 95.36%, respectively. However, while these models performed better, they still fell short of the highest standards of accuracy required for reliable medical diagnostics.The Proposed Hybrid Classifier, which combines Random Forest (RF), Support Vector Classifier (SVC), and XGBoost, achieved an exceptional accuracy of 99.96%. This hybrid approach leverages the complementary strengths of its constituent algorithms, effectively capturing intricate relationships within the data. The remarkable accuracy of the hybrid model underscores its potential as a highly reliable tool for the early detection and classification of alcoholic liver diseases. This superior performance highlights the advantages of using ensemble methods in medical diagnostics, paving the way for more accurate and dependable disease prediction models in clinical practice.

References:

[1] M.Ghoshetal.,“Acomparativeanalysisofmachinelearningalgorithmstopredictliverdisease, ” Intell. Autom. Soft Comput., Vol. 30, No. 3, pp. 917–928, 2021.

[2] Jagdeep Singha et al. “Software Based Prediction of Liver Disease with Feature Selection and Classification Techniques” ELSEVIER 2020.

[3] R. H. Lin, “An intelligent model for liver disease diagnosis,” Artif. Intell. Med., Vol. 47,No.1, pp. 53–62, 2009.

[4] GreeseGupta et al. “A Web Based Framework for Liver Disease Diagnosis using Combined Machine Learning Models” © IEEE 2020.

[5] A.K.M.S.Rahman et al. “Acomparativestudyonliverdiseasepredictionusingsupervisedmachinelearning algorithms,”Int. J. Sci.Technol. Res., vol.8, no. 11, pp.419–422, 2019.

[6] P. Sug, “On the optimality of the simple Bayesian classifier under zero-one loss”, Machine Learning 29 (2–3) (1997)103–130.

[7] Michael J Sorich, “Anintelligent model for liver disease diagnosis”, Artificial Intelligence in M edicine2009;47:53-62.

[8] C.J.A.Kannan,“ComparativeAnalysisofMachineLearningTechniques forIndianLiverDiseasePatients,”6th Int. Conf.Adv. Comput. Commun.Syst., 2020.

[9] Maria Alex Kuzhippallil, Carolyn Joseph,Kannan A, “Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients” © IEEE 2020.

[10] A.Singh,“AComprehensiveGuidetoEnsembleLearning(withPythoncodes),”2023.

[11] A.Quadir et al.“EnhancedPreprocessing Approach Using Ensemble Machine Learning Algorithms for DetectingLiverDisease,”2023.

[12] R.Ashtagi et al. “InternationalJournalof IntelligentSystemsandApplicationsin EngineeringRevolutionizingEarlyLiverDiseaseDetection:ExploringMachineLearning and Ensemble Models,” Orig. Res. Pap. Int. J. Intell. Syst. Appl. Eng. IJISAE,vol.2024,no.13s, pp.528–541,2024.

[13] M.O.Edehetal.,“ArtificialIntelligence-BasedEnsembleLearningModelforPredictionof Hepatitis C Disease,” Front. Public Heal., vol. 10, no. April, 2022.

[14] E.DritsasandM.Trigka, “SupervisedMachineLearningModelsforLiver DiseaseRiskPrediction,”Computers, vol.12, no. 1,p. 19, 2023.

[15] L. Meng, W. Treem, G. A. Heap, and J. Chen, “A stacking ensemble machine learningmodel to predict alpha-1 antitrypsin deficiency-associated liver disease clinical outcomesbased on UK Biobank data,” Sci. Rep., vol. 12, no. 1, pp. 1–18, 2022.

[16] Z.Jin et al.“RFRSF:EmployeeTurnoverPrediction Based on Random Forests and Survival Analysis,” in Lecture Notes inComputer Science, vol. 12343LNCS, 2020, pp.503–515.
[17] B.V.bAshfaqur and Rahmana,“Ensembleclassifiergenerationusingnon-uniformlayeredclusteringand GeneticAlgorithm,”Knowledge-BasedSyst., 2013.
[18] Y. Freund and R. E. Schapire, “Experiments with a New Boosting Algorithm,” Proc. 13thInt.Conf. Mach. Learn., pp. 148–156, 1996.
[19] S.DžeroskiandB.Ženko,“Iscombiningclassifierswithstackingbetterthanselectingthebestone ?,”Mach. Learn., vol. 54, no. 3, pp. 255–273, 2004