JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

An Efficient Machine Learning-Based Model for Predicting Acute Liver Failure

Somya Shukla
Computer Science
Shriram College Of Engineering & Management, Banmore
Gwalior, India
shukla.somya28@gmail.com

Jitendra Tyagi
Computer Science
Shriram College Of Engineering & Management, Banmore
Gwalior, India
Jktsrcem2@gmail.com

Abstract: *A clinical phenomenon known as acute-on-chronic liver failure (ACLF) affects people with chronic liver disease. It is characterized by rapid liver cirrhosis and is linked to a high short-term mortality rate. It is characterized by severe organ failure, systemic inflammation, and a bad prognosis. It is possible to classify and predict the course of patients with ACLF using specific prognostic ratings for liver and organ failures. Thus, this research aims to compare the efficacy of numerous "Machine Learning algorithms" to lower the expensive diagnostic cost of chronic liver disease. Several algorithms, including Gradient Boosting and Adaboost, were utilized in this work. The effectiveness of each classification approach was measured using metrics like accuracy, precision, recall, & f1-score in Gradient Boosting. Accuracy is 79.70%, 79%, 78%, and 75%, Adaboost, and Adaboost with Randomized Search CV, respectively. The testing results showed that the highest accuracy was achieved via Gradient Boosting. Our current research also primarily focuses on using clinical data to predict liver disease, and we investigate various data representations during our analysis. And the more accurate model is Ada Boost with RSCV for training in this study.*

Index Terms: *Acute Liver Failure, Disease Prediction, Machine Learning, Gradient Boosting Classifier, Adaboost Classifier.*

I. INTRODUCTION

Collecting samples from patients for diagnostic purposes can be time-consuming and expensive. Numerous tests or a large number of samples from the patient are often required to gather all the information needed for a more accurate diagnosis. Urinalysis "CBC (Complete Blood Count)" & "CMP (Comprehensive Metabolic Panel)" is the most often performed tests. These tests can still yield valuable results that are less expensive than conventional diagnostic procedures.

The liver is responsible for many metabolic processes, including storage, detoxification and glucose synthesis, digestive enzyme generation, protein synthesis, erythrocyte control and much more. Liver fibrosis, cirrhosis, and chronic hepatitis are all forms of chronic liver disease. Both viruses (such as the hepatitis C virus) and the body's immune system can trigger a case of hepatitis. Damage and scarring to the liver's tissue can result from the inflammation brought on by hepatitis infection. The significant differences in liver scarring are fibrosis and cirrhosis. Also, drinking and nonalcoholic fatty liver disease can result in cirrhosis and liver fibrosis. In the early phases of liver disease, among infection & fibrosis, liver failure can be prevented before cirrhosis. With procedures like a CMP and a biopsy, liver disease in all its varieties can be identified. Detection of albumin, alkaline phosphatase, aspartate aminotransferase, gamma-glutamyltransferase, alanine aminotransferase, total protein, creatine, & bilirubin (BIL) is possible with CMP, which also includes a liver function panel.

The CMP test measures various circulating liver-associated chemicals and compares them to reference levels adjusted for age, sex, and body mass index to diagnose and trace the cause of liver disease. Aminotransferases, AST and ALT, act in gluconeogenesis by helping convert ketone bodies to alpha-amino acids. [1]. Although AST is not strictly a liver marker, it may help identify secondary, non-hepatic causes of liver dysfunction because it is present in numerous tissues.

About 70% of all deaths occur because of liver disease.[2]. Developing more precise methods for identifying and diagnosing the liver disease is necessary. The availability and cost of liver function testing for patients should be prioritized. Avoiding expensive and invasive testing could be facilitated using statistical ML (Machine Learning) techniques applied to CMP outcomes for information extraction by doctors.[3]. Exploratory data analysis techniques are crucial in medicine because they can identify patterns in large data sets and speed up and improve identifying risk or disease-related diagnostic indicators. Using these techniques, liver disease may be detected sooner and may not proceed as far as necessitating biopsy or involved treatment in many cases.

Many problems in medical data sets go unnoticed until ML algorithms are applied. This strategy can aid healthcare administration and experts in investigating improved results across a wide range of clinical applications, including language processing, medical image analysis, & tumor or cancer cell detection, by identifying the appropriate features [4]. Various statistical & ML methods (including simulation modeling, categorization, & inference) have been utilized to enhance prediction by researchers and lab professionals [5][6][7]. In terms of clinical results, data drives more than models. Medical diagnostic classification problems are notoriously challenging because of the difficulty of identifying the appropriate target (response variable) and features. Despite its popularity, logistic regression is less effective than other methods, including several that make use of ML (Machine Learning) and DL (Deep Learning) [8][9][10].

The remaining paper is systemized primarily as follows: Section 2 reviews the literature critical to the investigation, while Section 3 presents the study's findings. Section 3 offers the proposed work, including the problem statement, the suggested strategy, the machine learning classifiers, the algorithm, and the flowchart. Then, the dataset details, performance metrics, experimental findings, and discussion are presented in section 4. Section 5 concludes with a discussion of the findings and future studies.

II. LITERATURE OF REVIEW

This section discusses the related work, whereas different researchers presented their work in a similar field.

The study focuses on [11] using the Indian Liver Patient Dataset; several machine learning models have been developed to predict liver illness (ILPD). In this investigation, we used a variety of EL (Ensemble Learning) models, including K-Nearest Neighbor Networks (KNNs), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Random Forests (RFs). The results showed that the highest accuracy, 88%, was achieved by an ensemble of KNN, RF, and SVM models. With RF and KNN ensemble learning, we achieved a TPR (True Positive Rate) of 99% for false positives. PCA (Principal Component Analysis) technique was also found to have a beneficial influence on accuracy in this research.

[12] choose the most effective machine learning technique. When evaluating the Dataset, we only use the most accurate settings. Model accuracy for the heart attack ML model was most excellent in Logistic Regression (LR) setting, at 93.41%. Meanwhile, the linear regression model had the lowest accuracy, at only 60.10% [12].

[13] aims to identify the characteristics that significantly aid in identifying liver fibrosis and to develop guidelines to help doctors treat patients at all stages using a clinically non-invasive method. Additionally, estimates and comparisons of the effectiveness of various Classifiers like MLP (Multi-layered Perceptron), RFs, & LR can handle both extensive and minimal datasets. Decision Tree only produced 28 rules from the same data, while previous studies generated 98002 rules with an accuracy rate of over 99.97%. Regulations based on the study's findings improved histological staging of liver fibrosis with a prediction accuracy of 97.45 percent.

In this work, [14] proposed a method for optimizing performance that is implemented using both the training data & elements that influence the model. AUC (Area Under the Curve) was calculated to be 72.5% for the naive Bayes algorithm and 63.19% for the k-nearest neighbor algorithm (KNN).

In this paper, [15] have compared four alternative ML algorithms for the task of categorizing the Indian Liver Patient Dataset (ILPD): Decision Tree (DT), Extra Trees (ET), Logistic Regression (LR), and Random Forest (RF). Pearson Correlation Coefficient based feature selection (PCC-FS) is applied to exclude extraneous features from the collection. Also, a boosting algorithm (AdaBoost) is utilized to enhance the predictive performance of those algorithms. The comparative analysis evaluated accuracy, ROC, F-1 score, precision, and recall. After comparing experimental results, we have found that boosting ET provides the highest accuracy of 92.19%.

In [16] focuses on a health care data set that deals with liver disease and compares the effectiveness of the mentioned three strategies using the Silhouette coefficient. Prediction accuracy is determined by the Silhouette coefficient, which determines K-Means to be the most effective method. Following this, the best approach for predicting liver disorders using unsupervised machine learning will be identified by contrasting all of the findings based on the accuracy of the predictions as well as the amount of computational effort required.

This study [17] classifies patients based on serum biomarkers & clinical data to assess the efficacy of several machine learning algorithms for predicting fibrosis progression. Age, AST, platelet count, & albumin were found to have statistically significant relationships with advanced fibrosis. Investigational ML algorithms successfully predicted advanced fibrosis in HCC patients with AUROC of 0.73 to 0.76 & accuracy of 66.3 to 84.4 percent. Conclusions: Machine-learning methods are an alternate strategy for estimating the chance of severe liver fibrosis because of chronic hepatitis C.

[18] proposed that a one-of-a-kind solution based on stochastic gradient descent be developed for learning with an abstention paradigm. This solver was then utilized to develop an efficient and innovative method for identifying liver illnesses. Our findings, which were derived from an analysis of data collected from around one hundred patients at MINAR in Multan, Pakistan, indicate that the performance of the suggested method is comparable to that of trained medical experts.

III. PROPOSED WORK

A. Problem Statement

In this world, people with liver cirrhosis suffer from long waiting times to be diagnosed due to limited medical resources and long diagnosis processes. Liver cirrhosis is now regarded as the leading common cause of death among people. Liver cirrhosis progresses slowly, and if diagnosed initially, there is a possibility of prolonged survival, which is considerably increased. Reducing diagnostic delays improves early detection and makes treatment outcomes in liver cirrhosis. Thus, reducing these delays improves early detection and makes treatment cost-effective. Doctors or medical practitioners can take prompt action. Again, liver cirrhosis is sometimes misdiagnosed due to a lack of proper tools and longer diagnosis processes. A long waiting time to diagnose liver cirrhosis may increase the possibility of the disease spreading. The motivation behind this study is that liver cirrhosis has become a common disease worldwide. The death rate due to the disease is becoming alarming. Early detection of the disease may reduce the complication of the disease misfortune on patients. The ease of use of innovative technologies such as the one anticipated in this research may help alleviate the troubles of holdup in uncovering and treating liver cirrhosis. Also, data mining tools can assist physicians in predicting and diagnosing the disease to enhance necessary treatment. One more significant drive behind this study is to advance on the works of previous researchers who contribute to this particular field of study.

B. Proposed Methodology

A machine learning-based approach is suggested for predicting acute liver failure to address the issues above. The primary goal of the suggested method is to forecast the effectiveness of machine learning-based categorization strategies for patients with liver disease. This section outlines the research methodology used for the study. The suggested system comprises some modules, one for each stage of the procedure. The first phase of this proposed part is to collect and load the dataset. The liver failure dataset[1] is used in this work which is collected from the kaggle. Next, apply the data pre-processing technique to preprocess the dataset. In the pre-processing technique, eliminate the missing and noisy data. After performing the data preprocessing, select the features randomly. Then, split the data into two phases, i.e., the testing set (20%) and the training set (80%). Then, apply the machine learning algorithms to perform the classification. And the machine learning classifiers are Gradient Boosting (GB) Classifier and Ada Boost (AB) Classifier; apart from this, both classifiers are also used with Randomized Search CV. Then finally, the proposed method is graded on how well it performs in terms of accuracy, precision, recall &, and F1-Score. Experimental findings demonstrate that the suggested system outperforms conventional methods. This overall procedure is described in a flow diagram which is presented below.

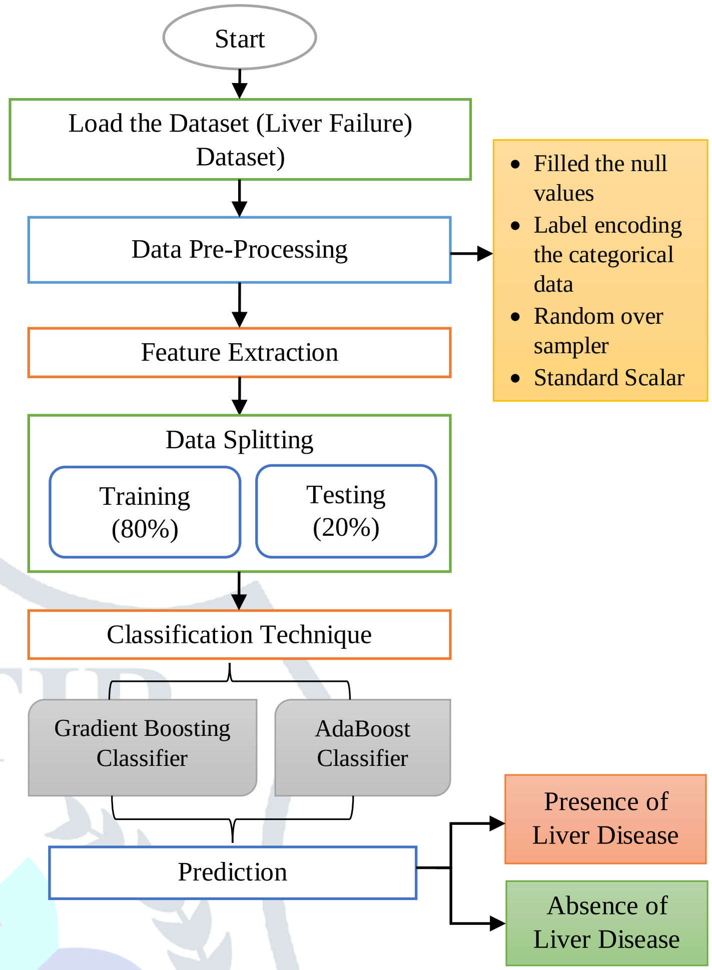


Figure 1: Flow Diagram of the Proposed Methodology

"The following modules, which will be covered further in this article, are used to implement this system. 1) Dataset Collection, 2) Data Preprocessing, 3) Feature Selection/Extraction, 4) Data Splitting, 5) Machine Learning Classifiers are the steps in this procedure".

a) Data Collection

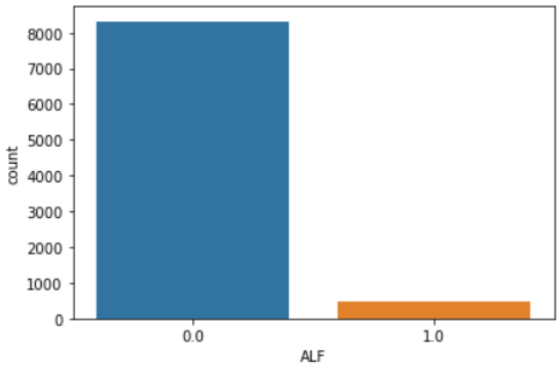
For this work, I will utilize the Liver Failure dataset collected from Kaggle. Liver cirrhosis occurs when scar tissue replaces normal liver cells. The liver's regular function is disrupted as a result. Damage to the liver from cirrhosis takes a (chronic liver disease or acute liver failure). Constant exposure to toxins causes progressive liver damage. The liver is the largest organ found within the body. It's the right side of your belly, just under your ribs.

b) Data Pre-Processing

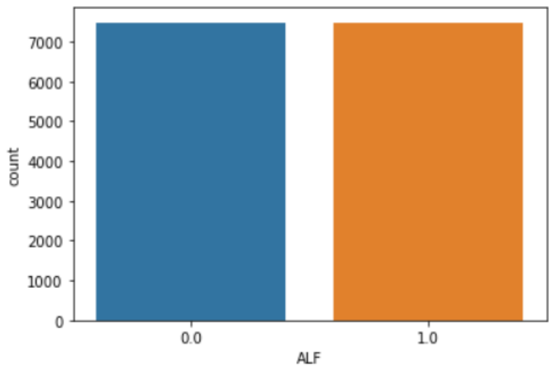
It is an essential step in the ML process since it prepares raw data for use in model development and training. The training and assessment of data preparation methods are crucial for the effective implementation of machine learning. Several preprocessing techniques can help ensure that classifiers are as accurate as possible. In the actual world, there is missing or noisy data. Thus, the data is preprocessed to reduce flaws and provide accurate forecasts. The following preprocessing techniques are missing value removal and standard scalar, which have been successfully used in classifiers. Several missing and noisy data are present to reduce these inaccuracies and make accurate forecasts in the real world. Each feature is guaranteed to have the

[1] *https://www.kaggle.com/rahul121/acute-liver-failure*

same mean and variance because of the standard scalar, which also ensures that each feature's coefficient is the same. Missing value rows are eliminated from the dataset.



a) Before Balancing



b) After Balancing

Figure 2: Count Plot of Data Balancing

Figure 2 represents a count plot visualization for data balancing. There are two plots given in this figure. The first graph represents the before data balancing graph, and the second represents the after-data balancing graph. For these graphs, the x-axis shows the total no. of classes, while the y-axis shows the total no. of observations.

1) Label Encoding

It is often used as an encoding method for categorical data. Here, an individual number is allocated to each label based on its position in the alphabet. Most tabular data sets have both numeric and nominal categories in separate columns. A computer can only process numerical data. Machine Learning algorithms are substantially the same in this regard. That's why we have to transform our categorical columns into numeric ones: so that a machine learning algorithm can read and make sense of them. The term "categorical encoding" describes this procedure. Translating concepts into numeric values is known as "categorical encoding."

2) Random Over Sampler (Oversampling technique)

In resampling, the training dataset is modified so that the chosen samples represent a different class distribution. This is a straightforward method for solving problems of unequal classification. The simplest method is random resampling, consisting of randomly selecting instances for the changed dataset. For imbalanced classifications, oversampling and undersampling are the two most common random resampling strategies.

- **Random Undersampling**: Randomly delete examples in the majority class.
- **Random Oversampling**: In the minority class, randomly duplicate some samples.

We can achieve random oversampling by selecting samples from the underrepresented group and replacing them with new ones. We achieve random undersampling by selecting samples randomly from the majority class & removing them from the training dataset. Both approaches may be used repeatedly until a desirable class distribution is reached in the training dataset—for example, even distribution across all classes.

3) Standard Scalar

StandardScaler is a crucial step in the preprocessing of most ML models, performed to normalize the scope of the functional capabilities of the input dataset. It removes mean data and adjusts the data normalization to the variance of the unit. However, outliers have an effect when calculating the empirical mean and standard deviation, reducing the range of characteristic values.

The steps listed below are necessary for standardizing a value:

$$y = (x - \text{mean}) / S.D \quad (1)$$

"Here, mean is determined as follows":

$$\text{mean} = \text{sum}(x) / \text{count}(x) \quad (2)$$

"S.D is calculated as":

$$S.D = \text{sqrt}\ (\text{sum}\ ((x - \text{mean})\ \hat{}\ 2\ ) / \text{count}(x)) \quad (3)$$

c) Feature Extraction

The purpose of a method known as "Feature Extraction" is to decrease the size of the dataset by generating new features from existing ones. (also then discarding original features). Afterward, this new, pared-down collection of features must be able to review the original set of features effectively. A condensed version can be derived from the original set of features by combining them in this fashion. No particular feature selection approaches are utilized in this experiment; however, feature selection differs from Feature Extraction in that it seeks to prioritize the value of the current characteristics in the dataset and reject less important ones.

d) Dataset Splitting

In machine learning, data splitting is widely implemented to prevent overfitting. Typically, machine learning models divide the initial data into two sets. Testing set & training set are two sets that are used most often. The testing set is made up of 20 percent of the whole dataset, and the training set is made up of 80 percent of the dataset.

e) Machine Learning Classifiers

1) Gradient Boosting Classifier

Gradient boosting machines, and maybe GBMs, use a learning process in which new models are sequentially fitted to new data to give a more exact estimate of the response variable. The basic idea behind this strategy is to train the new base learners to have the most significant correlation with the ensemble's overall negative gradient of the loss function. If the loss function is a traditional squared-error loss, the learning process will result in successive error-fitting. However, the loss functions used can be random to provide more intuition. In general, the researcher must choose the loss function to employ. A wide range of loss functions have been derived thus far, and one can construct task-specific loss. GBMs can be easily adjusted to any given data-driven activity because of their versatility. It gives lots of flexibility to model design, making choosing an appropriate loss function matter of trial & error. However, because boosting approaches are so simple to employ, it is straightforward to experiment with different model topologies. Furthermore, GBMs have shown considerable effectiveness in several machine learning and data mining issues and practical applications [19].

It builds a series of weak models to reduce the loss function and then uses it as the basis for building more robust models. This loss function is quantified using the gradient descent technique. When new models are constructed that better fit the observations, the total accuracy is improved by employing a loss function. However, boosting must be turned off sometimes, or the model would overfit. To determine when to end, you can use either a predetermined prediction accuracy cutoff or a fixed maximum number of models.[20]

2) Adaboost Classifier

A well-known algorithm in ML is boosting. Adaboost Algorithm is the most utilized Boosting Algorithm. An algorithm known as AdaBoost (AB) is an ensemble boosting technique developed by [21]. AdaBoost's core principle is combining many sequentially taught base classifiers to produce a more efficient model with improved prediction performance compared to each independently trained classifier. While training, erroneously predicted samples are given a boost in their weights.

In addition, to comprehend the idea behind the AdaBoost method, we'll pretend that we have n observations in our training dataset and that xi and Yi are the variables of interest. Consider the sum of all classifiers to be the "base." Equal weights distributions are used to begin training the primary classifier. Re-weighting each wrongly predicted sample in future training rounds will raise the likelihood of adequately classifying it. This training procedure is repeated until the stopping criteria, or all training samples have been properly categorized. A linear combination of the initial classifiers CJ results in the final model C(x). The following equation explains how to use it. [22].

$$C(x) = \sum_{j=1}^{n_{base}} w_j c_j(x), \tag{4}$$

The variable WJ represents weights for a given classifier.

C. Proposed Algorithm

To combine these classifiers' strengths, we propose a new consolidation algorithm. This section provides the proposed algorithm in which every step is defined briefly.

Algorithm 1: Proposed Machine Learning-based Algorithm

Input: Liver Failure Dataset
Output: Disease Prediction

Strategy:

Step 1. Initialize the system
Step 2. Importing all essential libraries of python.
Step 3. Gather and load the used Liver Failure Dataset from the Kaggle repository.
Step 4. Apply different data preprocessing techniques to preprocess the dataset.

- Filled the null values
- Label encoding the categorical data
- Random Over Sampler (Oversampling technique)
- Standard Scaler

Step 5. Following the feature extraction/ selection Technique for selecting and extracting the features, there is no specific feature selection technique.
Step 6. Splitting the dataset into the training set and testing set

- Training set (80%)
- Testing Set (20%)

Step 7. Apply different machine learning classifiers.

- Gradient Boosting Classifier
- Adaboost Classifier
- Gradient Boosting Classifier with Randomized Search CV
- Adaboost Classifier with Randomized Search CV.

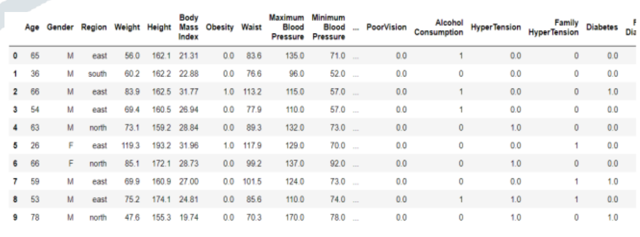
Step 8. Create the suggested machine-learning classifier models from the training dataset.
Step 9. Testing the ML classifiers' dataset based on the classifier model.
Step 10. Compare experimental classifier performance results.
Step 11. Using multiple metrics, determine the optimum algorithm. Then, predict the Liver disease
Step 12. Terminate the Model.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The results and their interpretation are discussed here. Begin by detailing the dataset's structure and contents. The next step is to create a visual representation of the experiment's results and an explanation of those results. This experiment is done on the python simulation tool, with the jupyter notebook simulation platform.

A. Dataset Description

The JPAC Center for Health Diagnosis and Control has been surveying Indian adults since 1990. The center conducted in-depth interviews and physical examinations and took blood samples employing trained employees to compile a wealth of demographic and health data. Information from the 2008-2009 and 2014-2015 surveys was used to compile this data collection, which includes responses from 8,785 persons aged 20 and older.



	Age	Gender	Region	Weight	Height	Body Mass Index	Obesity	Waist	Maximum Blood Pressure	Minimum Blood Pressure	...	PoorVision	Alcohol Consumption	HyperTension	Family HyperTension	Diabetes
0	65	M	east	56.0	162.1	21.31	0.0	83.6	135.0	71.0	...	0.0	1	0.0	0	0.0
1	36	M	south	60.2	162.2	22.88	0.0	76.6	96.0	52.0	...	0.0	0	0.0	0	0.0
2	66	M	east	83.9	162.5	31.77	1.0	113.2	115.0	57.0	...	0.0	1	0.0	0	1.0
3	54	M	east	69.4	160.5	26.94	0.0	77.9	110.0	57.0	...	0.0	1	0.0	0	0.0
4	63	M	north	73.1	159.2	28.84	0.0	89.3	132.0	73.0	...	0.0	0	1.0	0	0.0
5	26	F	east	119.3	193.2	31.96	1.0	117.9	129.0	70.0	...	0.0	0	0.0	1	0.0
6	66	F	north	85.1	172.1	28.73	0.0	99.2	137.0	92.0	...	0.0	0	1.0	0	0.0
7	59	M	east	69.9	160.9	27.00	0.0	101.5	124.0	73.0	...	0.0	0	0.0	1	1.0
8	53	M	east	75.2	174.1	24.81	0.0	85.6	110.0	74.0	...	0.0	1	1.0	1	0.0
9	78	M	north	47.6	155.3	19.74	0.0	70.3	170.0	78.0	...	0.0	0	1.0	0	1.0

10 rows × 30 columns

Figure 3: Dataset Visualization

The following figure 3 represents the visualization of the dataset. There are 30 columns and ten rows defined in this diagram. There are columns for "age, gender, region, weight, height, BMI, obesity, waist, maximum blood pressure, minimum blood pressure, bad eyesight, alcohol usage, hypertension, and family history of hypertension and diabetes." In contrast, data values for people aged 65 to 78 are listed in columns.

B. Performance Measures

- ***Accuracy:*** It is defined as the simple ratio that compares the number of points that were categorized correctly to the total number of points.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

- ***Precision:*** It is the percentage of cases that were classified correctly out of all classified cases.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

- ***Recall:*** The proportion of cases accurately classified out of the total number of instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (7)$$

- ***F1-Score:*** It represents the harmonic mean of precision & recall.

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (8)$$

C. Screenshots of Experimental Results

This section visualizes the results and describes their interpretation.

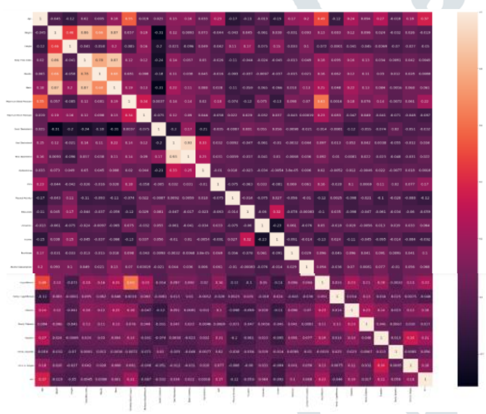


Figure 4: Correlation Matrix

Figure 4 represents the Correlation matrix of a dataset. After performing fundamental statistical analysis on the dataset, the correlation among each column was examined.

a) Testing Results

1) Results of Gradient Boosting Classifier

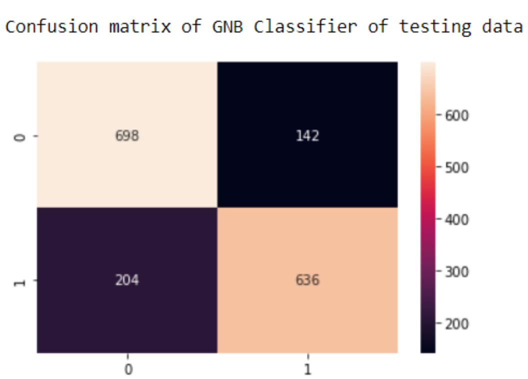


Figure 5: Confusion Matrix for GNB Classifier of Testing Data

Figure 5 illustrates the confusion matrix used by the Gradient Boosting classifier. In this fig., the x-axis depicts the predicted label, while the y-axis depicts the true labels in this matrix. This matrix shows the binary classification. In this matrix, the values are 698 true negative and 204 true positive, while 142 false negative, and 636 false positive, respectively.

```
Classification report GradientBoostingClassifier of testing data
              precision    recall  f1-score   support

           0       0.77      0.83      0.80       840
           1       0.82      0.76      0.79       840

    accuracy                           0.79      1680
   macro avg       0.80      0.79      0.79      1680
weighted avg       0.80      0.79      0.79      1680
```

Figure 6: Classification Report of GB Classifier

Figure 6 shows the classification report for the performance of the GB Classifier. For the 0 label data class, the Gradient Boosting Classifier precision, recall, and f1 score are 77%, 83%, and 80%, and for the one label data class, precision is 82%, recall is 76%, and f1 score is 79%. Also, different values of the weighted avg and the macro average are given in this figure, and the model accuracy is 79%, respectively.

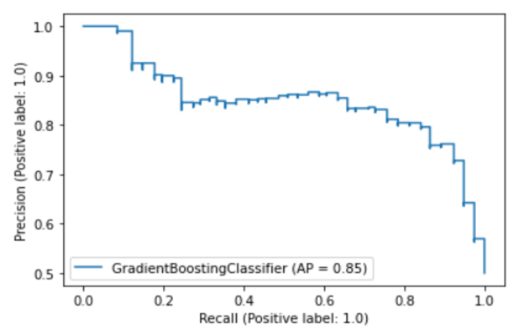


Figure 7: Precision-Recall Curve for GB Classifier

Figure 7 represents a Precision-Recall curve GB classifier. In this figure, the x-axis represents a recall (positive label:1.0) with a 0.0 to 1.0 number, and the y-axis represents a Precision (positive label:1) with a 05 to 10 number. The area under the curve is "x-axis 1.0" to "y-axis 1.0". Blue Line mentions the Random Search CV(AP=0.85) and this zig-zag line graph.

2) Results of Adaboost Classifier

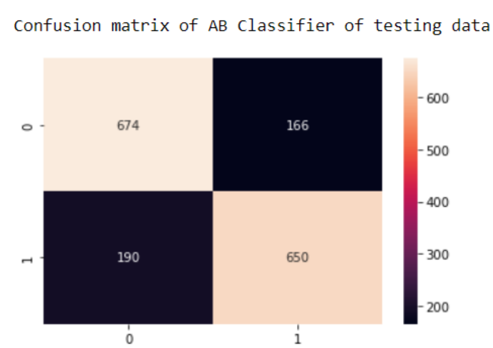


Figure 8: Confusion matrix of Adaboost Classifier

A confusion matrix for the Adaboost classifier is shown in fig. 8. In this matrix, values are given as 674 true negative and 190 true positive while 166 false negative, and 650 false positive, respectively.

Classification report AdaBoostingClassifier of testing data

	precision	recall	f1-score	support
0	0.78	0.80	0.79	840
1	0.80	0.77	0.79	840
accuracy			0.79	1680
macro avg	0.79	0.79	0.79	1680
weighted avg	0.79	0.79	0.79	1680

Figure 9: Classification Report of Adaboost Classifier

Figure 9 shows the classification report for the performance of the Ada Boost Classifier: For a class of data with 0 labels, the Ada Boost Classifier's precision, recall, and f1 score are 78%, 80%, and 79%, respectively, while for a class of data with one label, these values are 80%, 77%, and 79%. Also, different values of weighted avg and macro average are given in this figure, and the model accuracy is 79%, respectively.

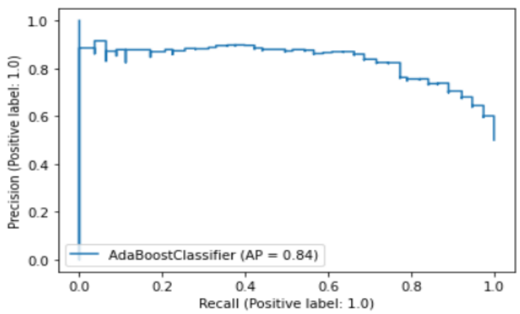


Figure 10: Precision-Recall Curve for Adaboost classifier

Figure 10 shows the accuracy and recall of the Adaboost classifier. This X-axis represents a recall (positive label:1.0) with a 00 to 10 number given, and the Y-axis represents a Precision (positive label:1.0withas a 05 to 10 number given. The area under the curve is x-axis 10 to y-axis 10. Blue Line mentions the Random Search CV(AP=0.84) and this zig-zag line graph.

3) *Results of Gradient Boosting with Randomized Search CV*

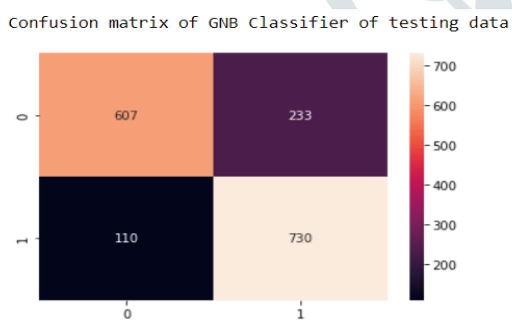


Figure 11: Confusion matrix of GNB classifier

Figure 11 represents a confusion matrix for the GNB classifier of testing data. In this matrix, the values are 607 true negative and 110 true positive while 233 false negative and 730 false positive, respectively.

Classification report GradientBoostingClassifier of testing data

	precision	recall	f1-score	support
0	0.85	0.72	0.78	840
1	0.76	0.87	0.81	840
accuracy			0.80	1680
macro avg	0.80	0.80	0.79	1680
weighted avg	0.80	0.80	0.79	1680

Figure 12: Classification report of GB classifier

Figure 12 shows the classification report for the performance of the Gradient Boosting with Randomized Search CV. The Gradient Boosting with Randomized Search CV precision, for 0 label data "f1 score" are 85%, and "recall", 80%, and 79% and "precision" 80%, "recall" 77% and 79% of "f1 score" for one label data. Also, different values of the weighted avg and the macro average are given in this figure, and the model accuracy is 79%, respectively.

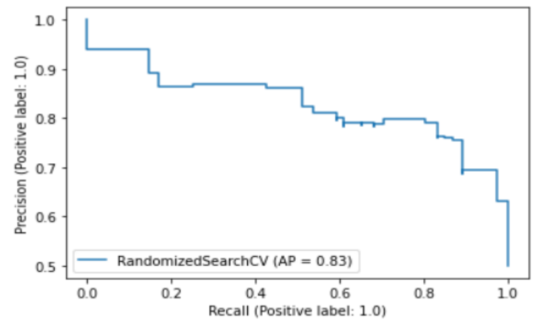


Figure 13: Precision-Recall Curve

Figure 13 represents a Precision-Recall curve. This X-axis represents a recall (positive label:1.0) with 00 to 10 numbers given, and Y-axis represents a Precision(positive label:1.0) with a 05 to 10 number given. Area under the curve is x-axis 10 to y-axis 10. Blue Line mentions the Random Search CV(AP=0.83) and this zig-zag line graph.

4) *Results of Ada Boosting Classifier with Randomized Search CV*

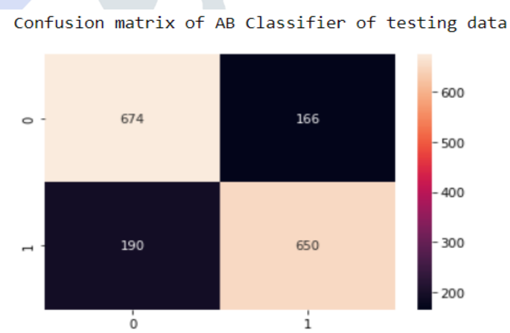


Figure 14: Confusion Matrix of AB classifier

Figure 14 represents a confusion matrix for the AB classifier. In this matrix, the values are 674 true negative and 190 true positive while 166 false negative, and 650 false positive, respectively.

Classification report AdaBoostingClassifier of testing data

	precision	recall	f1-score	support
0	0.78	0.80	0.79	840
1	0.80	0.77	0.79	840
accuracy			0.79	1680
macro avg	0.79	0.79	0.79	1680
weighted avg	0.79	0.79	0.79	1680

Figure 15: Classification Report of Ada boost classifier

Figure 15 shows the classification report for the performance of the Ada Boosting Classifier with Randomized Search CV. The different values of the weighted avg and the macro average are given in this figure, and model accuracy is 79%, respectively.

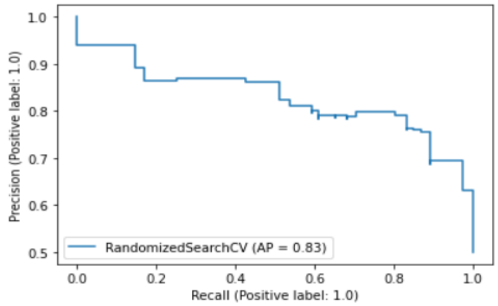


Figure 16: Precision-Recall Curve

Figure 16 represents a Precision-Recall curve Ada boosting classifier with randomized search CV. This X-axis represents a recall (positive label:1.0) with a 00 to 10 number given, and Y-axis represents a Precision(positive label:1.0) with a 05 to 10 number given. The area under the curve is x-axis 10 to y-axis 10. Blue Line mentions the Random Search CV(AP=0.83) and this zig-zag line graph.

b) Training Results

1) Gradient Boost Classifier:

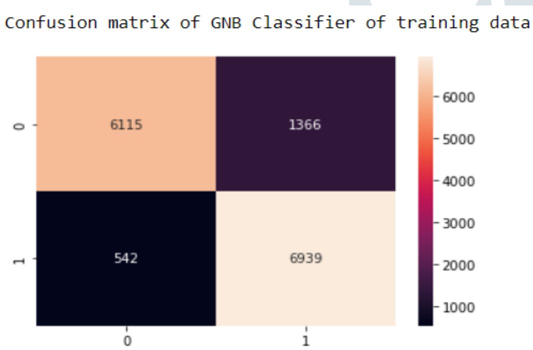


Figure 17: Confusion Matrix of GNB classifier

Figure 17 represents a confusion matrix for the GNB classifier. In this matrix, the values are 6115 true negative and 542 true positive, while 1366 false negative and 6939 false positive, respectively.

Classification report GradientBoostingClassifier of training data

	precision	recall	f1-score	support
0	0.92	0.82	0.87	7481
1	0.84	0.93	0.88	7481
accuracy			0.87	14962
macro avg	0.88	0.87	0.87	14962
weighted avg	0.88	0.87	0.87	14962

Figure 18: Classification Report of GB classifier of Training Data

The performance of the Gradient Boost Classifier is described in Figure 18. In the absence of labels, the GB classifier achieves a 92% precision, 82% recall, and 87% f1 score, but these values drop to 84%, 93%, and 88%, respectively. Also, different values of the weighted avg and the macro average are given in this figure, and the model accuracy is 87%, respectively.

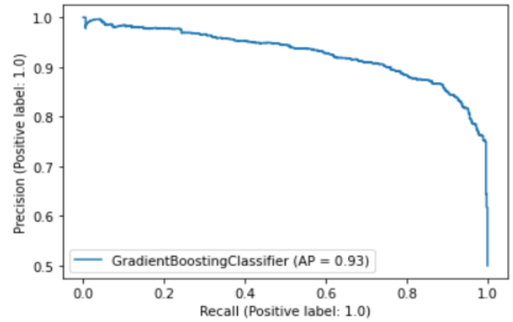


Figure 19: Precision-Recall Curve of GB classifier

Figure 19 represents a Precision-Recall curve of the GB classifier for testing. In this figure, the AP value of the gradient boosting classifier is 0.93.

2) Ada Boost Classifier

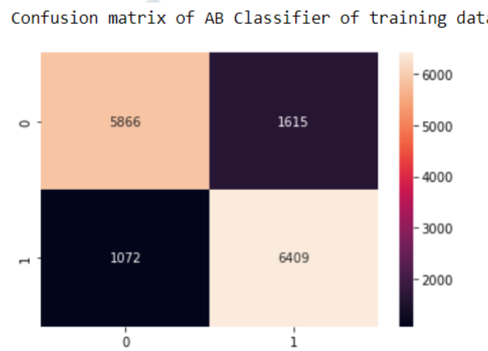


Figure 20: Confusion Matrix of AB classifier

Figure 20 represents a confusion matrix for the AB classifier. In this matrix, the values are given as 5866 true negative and 1072 true positive while 1615 false negative, and 6409 false positive, respectively.

Classification report GradientBoostingClassifier of training data

	precision	recall	f1-score	support
0	0.92	0.82	0.87	7481
1	0.84	0.93	0.88	7481
accuracy			0.87	14962
macro avg	0.88	0.87	0.87	14962
weighted avg	0.88	0.87	0.87	14962

Figure 21: Classification report of GB classifier

Figure 21 shows the classification report for the performance of the Ada Boost Classifier. The Ada Boost For a class of data with 0 labels, the classifier's “precision”, “recall”, & “f1 score” is 92%, 82%, & 87%, respectively, and for data with one label, they are 84%, 93%, and 88%. Additionally, different weighted average and macro average values are shown in this image. The model's accuracy is 87% and 87%, respectively.

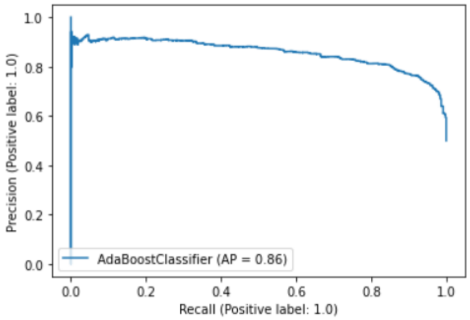


Figure 22: Precision-Recall Curve

Figure 22 represents a Precision-Recall curve of the Adaboost classifier. In this figure, the AP value of the AdaBoost classifier is 0.86.

3) Gradient Boosting with Randomized Search CV

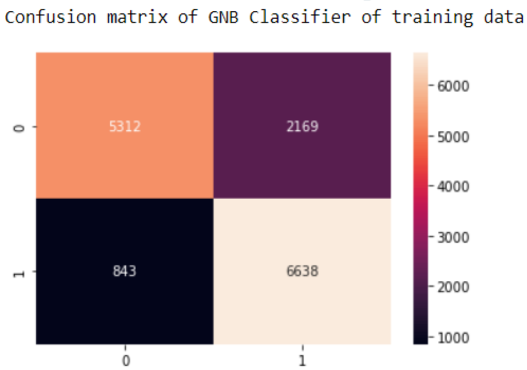


Figure 23: Confusion Matrix of GNB classifier

Figure 23 represents a confusion matrix for the GNB classifier. In this matrix, the values are given as 5312 true negative and 843 true positive while 2169 false negative, and 6638 false positive, respectively.

Classification report GradientBoostingClassifier of training data

	precision	recall	f1-score	support
0	0.86	0.71	0.78	7481
1	0.75	0.89	0.82	7481
accuracy			0.80	14962
macro avg	0.81	0.80	0.80	14962
weighted avg	0.81	0.80	0.80	14962

Figure 24: Classification report of GB classifier

Figure 24 shows the classification report for the performance of the Gradient Boosting with Randomized Search CV. The Gradient Boosting with Randomized Search CV For a class of data with 0 labels, precision, recall, and f1 score are 86%, 71%, and 78%, respectively; for a class of data with one label, these values are 75%, 89%, and 82%. Also, different values of the weighted avg and the macro average are given in this figure, and the model accuracy is 80%, respectively.

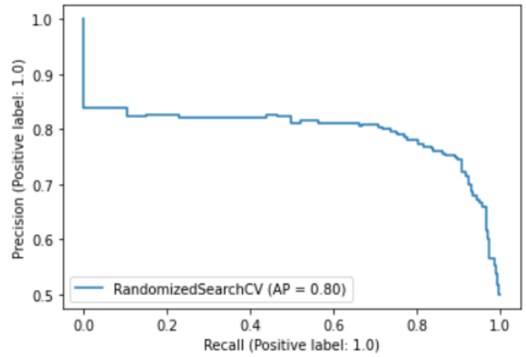


Figure 25: Precision-Recall Curve

Figure 25 represents a Precision-Recall curve of the GB classifier with a randomized search CV. In this figure, the AP value of the gradient boosting classifier is 0.80.

4) Ada Boost Classifier with Randomized Search CV:

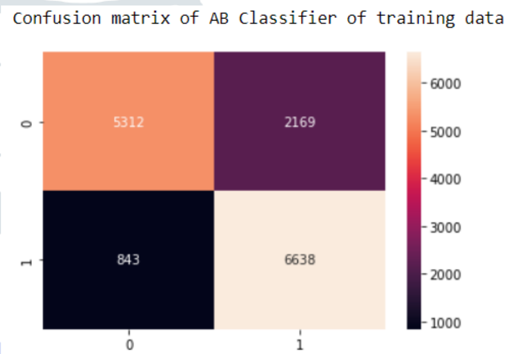


Figure 26: Confusion Matrix of AB classifier

Figure 26 represents a confusion matrix for the AB classifier with a randomized search CV. In this matrix, the values are given as 5312 true negative and 843 true positive while 2169 false negative, and 6638 false positive, respectively.

Classification report GradientBoostingClassifier of training data

	precision	recall	f1-score	support
0	0.86	0.71	0.78	7481
1	0.75	0.89	0.82	7481
accuracy			0.80	14962
macro avg	0.81	0.80	0.80	14962
weighted avg	0.81	0.80	0.80	14962

Figure 27: Classification Report of Ada Boost Classifier

Figure 27 shows the classification report for the performance of the Ada Boost Classifier with Randomized Search CV. Precision, recall, and f1 scores are 86%, 71%, and 78% for the data class with zero labels. And for one label, the values are 75%, 89%, and 82%, respectively. Also, different values of the weighted avg and the macro average are given in this figure, and the model accuracy is 80%, respectively.

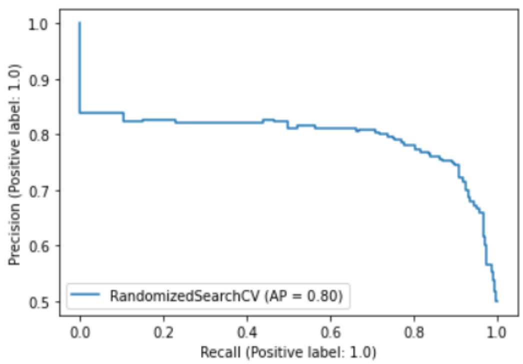


Figure 28: Precision-Recall Curve

Figure 28 represents a Precision-Recall curve for the AB classifier with a randomized search CV for training. In this graph, the AP value of the AB classifier randomized search CV is 0.80.

Table 1: Testing Results of Proposed Models Based on Performance Measures

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
GB Classifier	79.70	0.79	0.79	0.79
GB with RSCV	0.79	0.79	0.79	0.79
AB Classifier	0.78	0.78	0.78	0.78
AB with RSCV	0.75	0.78	0.78	0.78

Table 2: Training Results of Proposed Models Based on Performance Measures

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
GB Classifier	0.80	0.87	0.87	0.87
GB with RSCV	0.79	0.80	0.80	0.80
AB Classifier	0.82	0.82	0.82	0.82
AB with RSCV	0.86	0.80	0.80	0.80

The testing results and training of the suggested models are shown in Tables 1 or 2 above based on performance metrics such as accuracy, precision, recall, & f1-score.

D. Comparative Results

Table 3: Base and proposed testing models comparison table

Model	**Base**				**Propose**			
	XGBoost	**LGBM**	**XGBoost with Grid**	**LGBM with Grid**	**Gradient Boosting**	**AdaBoost**	**GB with RSCV**	**AB with RSCV**
Accuracy	75	0.75	76.73	0.77	79.70	0.78	0.79	0.75
Precision	0.33	0.40	0.35	0.15	0.79	0.78	0.79	0.78
“Recall”	0.25	0.50	0.30	0.75	0.79	0.78	0.79	0.78
“F1-Score”	0.28	0.44	0.30	0.26	0.79	0.78	0.79	0.78

Table 3 shows the comparative results between the base and proposed models for testing. The results of the studies show that, compared to the other methods, the gradient boosting classifier delivers the best results empirically. The gradient boosting model, the best classification strategy overall, had the highest testing accuracy of all classification methods, at 79.70 percent.

Table 4: Comparison Table Between Base and Propose Models for Training

Model	**Base**				**Propose**			
	XGBoost	**LGBM**	**XGBoost with Grid**	**LGBM with Grid**	**Gradient Boosting**	**AdaBoost**	**GB with RSCV**	**AB with RSCV**
Accuracy	0.86	0.67	1.0	0.86	0.80	0.82	0.79	0.86

Table 4 shows the comparative results between the base and proposed models for training. The results of the studies show that the Adaboost classifier with randomized search CV produces the best accuracy compared to the other techniques. Model AB with RSCV achieved a training accuracy of 0.86 percent, which was the highest.

V. CONCLUSION AND FUTURE WORK

The primary objective of this research is to develop a method that is effective for identifying individuals with chronic liver infections using specialized machine learning classifiers. The main problem with classifiers used to study liver disease has been their accuracy in predicting or detecting it. Adaboost has an 86% diagnostic accuracy in predicting liver illness, whereas the GB classifier obtains the highest precision, recall, and f1-score (87%). GB classifier achieves the best test results of 79.70% accuracy, 79% precision, f1-score, and recall. The decision support system and diagnosis of chronic diseases will now use the outperform classification technique.

Our study has a few implications for future research in this area. We have only looked at a few well-known machine learning algorithms; other algorithms can be chosen to build a more accurate model for predicting liver illness, and performance can be gradually enhanced. Additionally, this research is poised to play a significant role in healthcare, particularly concerning preventative measures against liver infection.

References

[1] G. Borroni *et al.*, "Comparison of simple tests for the non-invasive diagnosis of clinically silent cirrhosis in chronic hepatitis C," *Alimentary Pharmacology and Therapeutics*. 2006, doi: 10.1111/j.1365-2036.2006.03034.x.

[2] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," *Journal of Hepatology*. 2019, doi: 10.1016/j.jhep.2018.09.014.

[3] J. A. Udell *et al.*, "Does this patient with liver disease have cirrhosis?," *JAMA*, 2012, doi: 10.1001/jama.2012.186.

[4] M. Gupta and H. C. Kaplan, "Measurement for quality improvement: using data to drive change," *J. Perinatol.*, 2020, doi: 10.1038/s41372-019-0572-x.

[5] J. Benneyan, "The design, selection, and performance of statistical control charts for healthcare process improvement," *International Journal of Six Sigma and Competitive Advantage*. 2008, doi: 10.1504/IJSSCA.2008.021837.

[6] C. Duguay and F. Chetouane, "Modeling and Improving Emergency Department Systems using Discrete Event Simulation," *Simulation*, 2007, doi: 10.1177/0037549707083111.

[7] M. Subramaniyan, A. Skoogh, M. Gopalakrishnan, H. Salomonsson, A. Hanna, and D. Lämkull, "An algorithm for data-driven shifting bottleneck detection," *Cogent Eng.*, 2016, doi: 10.1080/23311916.2016.1239516.

[8] R. Couronné, P. Probst, and A. L. Boulesteix, "Random forest versus logistic regression: A large-scale benchmark experiment," *BMC Bioinformatics*, 2018, doi: 10.1186/s12859-018-2264-5.

[9] A. B. Musa, "Comparative study on classification performance between support vector machine and logistic regression," *Int. J. Mach. Learn. Cybern.*, 2013, doi: 10.1007/s13042-012-0068-x.

[10] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inform.*, 2002, doi: 10.1016/S1532-0464(03)00034-0.

[11] B. H. Al Telaq and N. Hewahi, "Prediction of Liver Disease using Machine Learning Models with PCA," 2021, doi: 10.1109/ICDABI53623.2021.9655897.

[12] M. S. Salman *et al.*, "Data Mining Technique for Prediction System of Heart Disease Using Associative Classifications," 2021, doi: 10.1109/ICCCNT51525.2021.9579825.

[13] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," 2020, doi: 10.1109/B-HTC50970.2020.9297949.

[14] H. Hartatik, M. B. Tamam, and A. Setyanto, "Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms," 2020, doi: 10.1109/ICORIS50180.2020.9320797.

[15] M. F. Rabbi, S. M. Mahedy Hasan, A. I. Champa, M. Asifzaman, and M. K. Hasan, "Prediction of liver disorders using machine learning algorithms: A comparative study," 2020, doi: 10.1109/ICAICT51780.2020.9333528.

[16] V. Vats, L. Zhang, S. Chatterjee, S. Ahmed, E. Enziama, and K. Tepe, "A Comparative Analysis of Unsupervised Machine Techniques for Liver Disease Prediction," 2019, doi: 10.1109/ISSPIT.2018.8642735.

[17] S. Hashem *et al.*, "Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, 2018, doi: 10.1109/TCBB.2017.2690848.

[18] K. Hamid, A. Asif, W. Abbasi, D. Sabih, and F. U. A. A. Minhas, "Machine Learning with Abstention for Automated Liver Disease Diagnosis," 2017, doi: 10.1109/FIT.2017.00070.

[19] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, 2013, doi: 10.3389/fnbot.2013.00021.

[20] S. Rahman, M. Irfan, M. Raza, K. M. Ghori, S. Yaqoob, and M. Awais, "Performance analysis of boosting classifiers in recognizing activities of daily living," *Int. J. Environ. Res. Public Health*, 2020, doi: 10.3390/ijerph17031082.

[21] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, 1997, doi: 10.1006/jcss.1997.1504.

[22] P. Gnip, L. Vokorokos, and P. Drotár, "Selective oversampling approach for strongly imbalanced data," *PeerJ Comput. Sci.*, 2021, doi: 10.7717/PEERJ-CS.604.