Liver Disease Prediction using Machine Learning Algorithms

Riya
Department of Computer Science and Engineering
IET Bhaddal Technical Campus Ropar
Punjab, India

Barinderjit Kaur
Department of Computer Science and Engineering
IET Bhaddal Technical Campus Ropar
Punjab, India

ABSTRACT
Early diagnosis is essential to improving patient outcomes and decreasing the costs of healthcare. Liver disease is a big worldwide health problem, and early detection plays a critical role in both of these areas. A data-driven technique that makes use of supervised learning algorithms is presented in this research report as a method for estimating the likelihood of developing liver disease. For the objectives of research, the study makes use of a dataset collected from www.kaggle.com that contains information on the demographics, lifestyle, and medical history of 416 patients who were treated at a hospital in India. In order to construct accurate prediction models for the risk of liver illness, we use three different supervised learning techniques. These are decision trees, random forests, and logistic regressions. Accuracy, specificity, sensitivity, and the area under the receiver operating characteristic (ROC) curve are the metrics that are used in order to assess the performance of these models. According to the findings, the hybrid method surpasses the other two algorithms by obtaining higher levels of accuracy (75.7 percent), sensitivity (71.4 percent), and specificity (77.2 percent), as well as a higher area under the ROC curve (0.80). This work demonstrates the potential of supervised learning algorithms in forecasting the risk of liver disease using patient data, especially in areas where there is a limited availability of resources.

Keywords
Liver Disease, Chronic diseases, data analysis, Machine Learning Algorithms

1. INTRODUCTION

It is becoming clear that the digital technological revolution has the potential to be a really disruptive breakthrough [1]. The rise of cutting-edge medical technologies is marked by developments in nanotechnology and genetics [2] Prognosis, therapy, and healthcare monitoring in the digital age have great promise, and there seems to be no end to the possible applications of this potential. Every time a medical procedure is carried out, a huge quantity of data is processed and dealt with on a frequent basis [3]. These data sets may be inferential, referential, or raw enough to draw conclusions about further valuable medical data sets. There is a wide variety of origins and purposes for this data. They have applications in illness prediction, diagnosis, and therapy [4]. The same research might be studied to speed up similar initiatives. It could be useful for making statistical conclusions about future trends that may aid the process as a whole. Classification methods are widely used in data mining for illness prediction and medical diagnosis [5].

The liver is a crucial organ that sits at the very top of the digestive system. It's important for things like breaking down food, making energy, eliminating toxins, producing antibodies, and storing nutrients [6]. The liver is the body's second largest internal organ, and it's responsible for a wide variety of functions, including digestion, clotting blood, and producing bile. About three pounds is its estimated weight [7]. The liver is responsible for a wide variety of tasks, including digestion, metabolism, immunity, and nutrition storage. For these reasons, the liver is a vital organ; without it, cell death from a lack of oxygen and nutrients would occur rapidly throughout the body [8]

Liver diseases pose a significant health risk and can even lead to death. These diseases are categorized based on their causes and effects on the liver [9]. Improved patient outcomes and lower healthcare costs are possible when liver disease is diagnosed and treated as early as possible. The potential of machine learning to aid in illness risk prediction and clinical decision-making has attracted a lot of interest in recent years [10]. In this study, we use machine learning techniques to patient data in order to create a model that reliably predicts the likelihood of developing liver disease. Alcoholic liver disease, Viral hepatitis, non-alcoholic fatty liver disease (NAFLD), and liver cirrhosis are all examples of liver disease [11]. If these disorders aren't diagnosed and treated in time, they may cause liver dysfunction, compromised liver function, and even liver failure [12]. Therefore, identifying those at high risk for developing liver disease may have a major influence on public health by allowing for early treatments and individualized preventative measures.

This research gives a proposed research strategy to solve this problem by using supervised learning techniques. It is possible to create prediction models with the help of supervised learning algorithms since they can learn patterns and correlations from labelled data. Data from 416 patients' medical records, lifestyle characteristics, and demographics were utilized in this analysis. This data collection was collected from an Indian hospital and uploaded to a trustworthy database for scientific study. The aim of this research paper is to develop an accurate prediction model for liver disease risk using machine learning algorithms and patient data. The objective of the study is to achieve the following:

- Analyze the efficacy of the random forest, decision tree, and the logistic regression supervised learning algorithms in determining the likelihood of liver disease.
- Evaluate the prediction models created by the three algorithms with respect to accuracy, sensitivity, specificity, and area under the ROC curve.
- Based on the results of the assessment, choose the algorithm that produces the best accurate predictions of risk for liver disease.
- Evaluate supervised learning algorithms for their potential use in forecasting the risk of liver disease in low-resource areas.

The remainder of the paper is laid out as follows. Related literature is also discussed in Section 2 The dataset and the technique used are described and analyzed in Section 3. Additionally, in Section 4, we draw conclusions based on the findings of the experiments. In the final section we give the conclusion of the paper.

2. LITERATURE REVIEW

Rahman, Shamrat, Tasnim, Roy, and Hossain, [13]"A comparative study on liver disease prediction using supervised machine learning algorithms" Many people throughout the globe suffer from chronic liver disease, making it a major international mortality. It's caused by things like being overweight, having an undetected hepatitis infection, and drinking excessively. Serious complications from this syndrome include abnormal nerve function, hepatic encephalopathy hemoptysis or vomiting blood, liver failure, renal failure, jaundice. Chronic liver disease is difficult and costly to diagnose. The primary goal of this study is to apply six different supervised machine learning classifiers to develop a reliable method of identifying people who suffer from chronic liver disease. With the hope of lowering the astronomical expenses involved with diagnosing chronic liver disease, this research compares the efficacy of many machine learning algorithms for making such a prediction. In this study, researchers used six different machine learning algorithms.

Durai, V., Ramesh, S. and Kalthireddy, D., [14]"Liver disease prediction using machine learning" Medical diagnosis and prognosis are two areas that have benefited greatly from data mining tools. Data mining methods have been used on a mountain of medical records. Liver-related disorders are becoming more common as the prevalence of obesity and poor lifestyles rises sharply. In this study, researchers use an extensively studied classification algorithm on patient data to determine the likelihood that a certain individual suffers from liver disease. Given that methods for analyzing both patient data and classifier data already exist, the more pressing concern is developing methods that more accurately anticipate the same final outcome. In all, there are 5 stages to this procedure. At first, the min-max method is used on the raw data from the UCI repository's liver patient dataset. The second step makes use of PSO feature selection to identify crucial characteristics. From complete normalized datasets of liver patients, this helps extract the subset of essential data. The next stage, classification algorithms, is used to make comparisons and put things into categories. The fourth stage is called Accuracy Calculation. Root Mean Square value and Root Error value are used. Evaluation is the fifth and last stage. A straightforward assessment procedure is carried out, as required by the research, to protect the credibility of a reliable representation of the results. The J48 algorithm has a 95.04 percent accuracy rate, making it the best performing algorithm for feature selection.

Azam, Rahman, Iqbal, and Ahmed, 2020[15] "Prediction of liver diseases by using few machine learning based approaches" Improvements in healthcare have always been an extremely important part of human civilization. As medical science and technology advance, patients are more required to undergo treatment using cutting-edge methods and tools. Accuracy in the medical sciences is increasingly dependent on the use of machine learning methods. In this study, we developed computational model building strategies for reliable prediction of liver disease. In order to forecast liver illnesses, we employed various powerful classification algorithms including Random Forest, etc. In order to enhance prediction capabilities, our works employ hybrid model design and comparison analysis. At first, the raw datasets of liver patients from the UCI repository are put through categorization algorithms. We compared the performance of other classifiers and evaluated the characteristics to fine-tune our own prediction. We looked into it, and the KNN algorithm proved to be the most effective method for selecting features.

Hartatik, Tamam, and Setyanto, [16] "Prediction for diagnosing liver disease in patients using KNN and Naïve Bayes algorithms". Patients' medical histories may be used as a foundation for making predictions about their likelihood of developing liver disease, and so there is a need to extract relevant data from their medical records and laboratory results. This is helpful for both doctors and patients, especially if they are experiencing similar symptoms. Machine learning is being employed in this research because of the large amounts of data being used to make predictions about the future. It has been shown in the past that assessment outcomes may be somewhat subjective. This research uses training data and model-influencing factors to implement the suggested approach to performance optimization.

3. MATERIALS AND METHODS

Including class balancing and ranking features in the balanced data, we will describe the dataset we utilized and the primary stages of the selected strategy for forecasting the risk of liver sickness. Finally, we detail the ML models that were used in order to make sense of the experimental results.

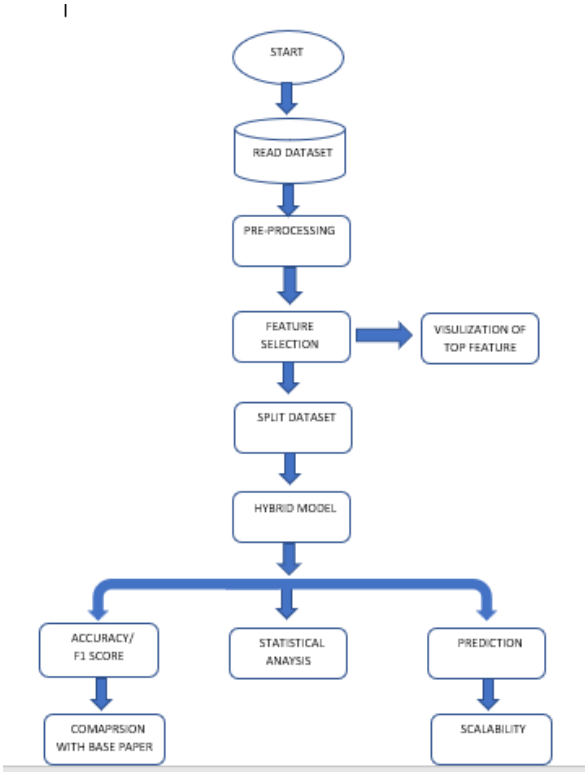
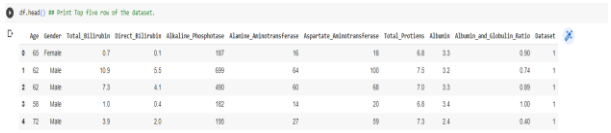


Fig 1: Flowchart of Proposed approach

Dataset

After that, the dataset is "preprocessed," which involves things like fixing any inconsistencies and figuring out how to handle missing information. The specific dataset includes 583 participants. The target class represents whether or not the person has been diagnosed with liver illness. There were 416 people who had been diagnosed with liver illness (71.5 percent

). In Table 1, we can get a summary of the data set's properties.



	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Dataset
0	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.90	1
1	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
2	62	Male	7.3	4.1	490	60	68	7.0	3.3	0.89	1
3	58	Male	1.0	0.4	182	14	20	6.8	3.4	1.00	1
4	72	Male	3.9	2.0	195	27	59	7.3	2.4	0.40	1

Fig 2. Dataset

Figure 2 shows 583 rows and 11 columns from the liver detection dataset. In this table, each row represents a separate instance or sample, and each column represents a different characteristic or attribute of that instance. This dataset seems to have liver detection as its major emphasis, suggesting that its goal is to identify and label occurrences as having or not having features connected to the liver. There are a total of 11 columns in the dataset, all of which indicate different characteristics that may be useful for liver identification. Age, gender, weight, height, body mass index (BMI), liver enzyme levels, and other medical or physiological parameters are all examples of characteristics that are associated with liver health or problems. These features will be used to build a machine learning model for liver detection, and each column gives unique information or measurements about the samples. Researchers and data scientists might benefit greatly from this dataset by examining the connections between the various characteristics and the existence or absence of liver-related disorders. In addition, researchers may use this data to train and test machine learning algorithms and prediction models for effectively categorizing new occurrences based on their liver features [17]. As a result, the dataset is a useful tool for training and assessing such algorithms, which might eventually lead to better patient outcomes via more refined liver detection methods.

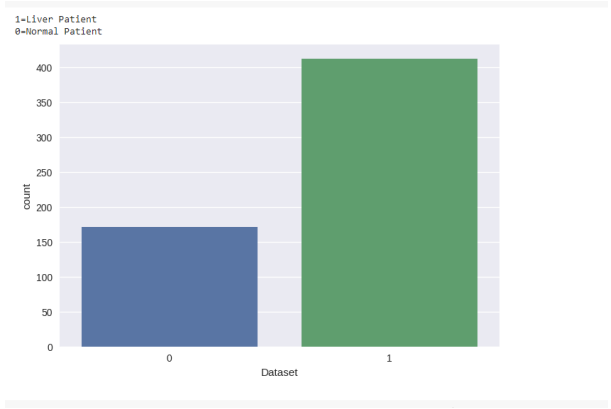


Fig 3 Dataset

Figure 3 shows that the liver detection dataset has 583 rows, or instances. Each instance is a unique record. There are a total of 583 patient records, 171 of which show that the patient did not have liver disease and 412 of which had. This pattern indicates that there is an inequity between the percentage of patients with and without liver disease in the dataset. In machine learning and statistical analysis, imbalanced datasets may be problematic because models may overfit to the dominant class and underfit to the minority class. Understanding the class distribution of the dataset is crucial for developing efficient modeling approaches and evaluation metrics [18]. Oversampling the minority class, undersampling the majority class, or using techniques designed for imbalanced data are all effective ways to deal with class imbalance and ensure equitable representation and improved model performance.

Liver Disease Risk Prediction

These days, physicians and health care providers use machine learning algorithms to create accurate tools for determining the likelihood of illness incidence given a set of risk variables. Here, we frame the issue of predicting the long-term risk of liver disease as a classification problem with two classes: c = "Liver-Disease" (LD) and c = "Non-Liver-Disease" (Non-LD). The probability of liver disease is predicted by the trained ML models, which can classify a new unclassified instance as LD or Non-LD depending on the values of the input characteristics.

Data Preprocessing

The dataset is then subjected to preprocessing, which involves cleaning the data to remove inconsistencies and handling missing values appropriately. Uneven distribution of LD and Non-LD occurrences in the dataset may interfere with their proper identification. Here, we use an oversampling technique, which generates synthetic data on the underrepresented group by means of a 5-NN classifier. Non-LD cases are oversampled to ensure that the two populations are evenly represented. The dataset is now evenly split between LD and NonLD cases, totaling 416. Finally, we give the information of the dataset in figure 1.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    object
 2   Total_Bilirubin             583 non-null    float64
 3   Direct_Bilirubin            583 non-null    float64
 4   Alkaline_Phosphotase        583 non-null    int64
 5   Alamine_Aminotransferase    583 non-null    int64
 6   Aspartate_Aminotransferase  583 non-null    int64
 7   Total_Protiens              583 non-null    float64
 8   Albumin                     583 non-null    float64
 9   Albumin_and_Globulin_Ratio  579 non-null    float64
 10  Dataset                     583 non-null    int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
```

Fig 4. Data.info

The column "Albumin and Globulin Ratio" in the liver detection dataset is missing four values as shown in figure 4. One popular method for dealing with such gaps is to substitute the column's mean for the missing value. The mean is found by adding together all the non-missing values and dividing by the total number of such values. By using the column's mean to fill in missing values, we can preserve the column's statistical features and reduce the effect of missing data on our analyses and models. We may maintain the general features of the data by utilizing the mean value to estimate what the missing values could be. This method relies on the assumption that missing values occur at random and that the current data faithfully represents the missing ones. While the median imputation approach is simple and extensively used, there are several methods that may be used instead. To forecast missing values based on other pertinent factors, for instance, regression imputation or more complex machine learning methods might be used. To keep the dataset whole and usable for analysis and modeling purposes with regards to liver detection, we imputed the missing values in the "Albumin and Globulin Ratio" column with the mean. But it's still smart to check how missing data imputation affects the final analysis findings and switch to a different approach if it works better for your data and goals.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    object
 2   Total_Bilirubin             583 non-null    float64
 3   Direct_Bilirubin            583 non-null    float64
 4   Alkaline_Phosphotase        583 non-null    int64
 5   Alamine_Aminotransferase    583 non-null    int64
 6   Aspartate_Aminotransferase  583 non-null    float64
 7   Total_Protiens              583 non-null    float64
 8   Albumin                     583 non-null    float64
 9   Albumin_and_Globulin_Ratio  583 non-null    float64
 10  Dataset                     583 non-null    int64
dtypes: float64(6), int64(4), object(1)
memory usage: 50.2+ KB
```

Fig 5. After fill Null value

After missing data was filled in, the final results were shown in Figure 5.

```
Outlier Values in each Column
==============================
Age                          0
Gender                       0
Total_Bilirubin             27
Direct_Bilirubin            29
Alkaline_Phosphotase        14
Alamine_Aminotransferase    11
Aspartate_Aminotransferase  19
Total_Protiens               9
Albumin                      7
Albumin_and_Globulin_Ratio   4
Dataset                      0
dtype: int64
```

Fig 6. Outliers Values in Each Column

There are 11 columns in the liver detection dataset; eight of them contain outliers. The results of statistical studies or models may be skewed by outliers as shown in figure 6, or data points that differ dramatically from the norm. Capping is a frequent method used to deal with these extreme cases. Capping, also known as winsorization, is the process of substituting highly improbable numbers with more moderate ones that fall within an established range. The 5th and 95th percentiles are common ways to describe this range. Outlying values are substituted with their respective percentile values when they are found to be outside this range. We expect that by putting a ceiling on the outliers, we can reduce the impact they have on our statistical models and other analyses that use the dataset. The analysis is more robust and trustworthy when outliers are not allowed to have an undue influence by capping. However, care must be used while dealing with extreme cases. However, before deciding to cap outliers, it is important to think about the data and the situation in which the analysis will be used. Blindly eliminating or manipulating outliers might lead to potentially biased or misleading conclusions since they can occasionally hold significant information or indicate true oddities in the data. It is best practice to review the data, get familiarity with the domain, and if feasible, contact domain experts before implementing any outlier management strategy, including capping. In order to make an informed choice on whether or not to utilize capping or another outlier management approach, one must first consider the dataset, the analytic goals, and the possible influence on the final results or conclusions [19]. The eight columns of the liver detection dataset include some outliers, and we use capping to remove them in order to keep the data usable for analysis while minimizing any distortions that could arise from them.

Feature selection

Then we apply feature selection techniques to identify the most relevant features for predicting liver disease. We select the features that have high predictive power and are meaningful from a medical perspective. To gain insights into the selected features, data visualization techniques are employed, allowing for the visualization of distributions, correlations, and trends. The dataset is then split into training and testing sets to facilitate model development and evaluation as seen in figure.

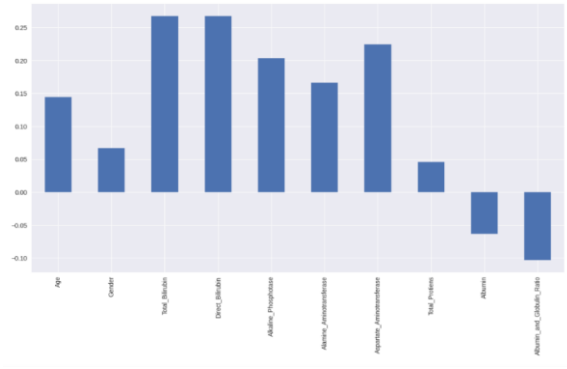


Fig 7. Correlation

As can be seen in Fig. 7, three columns—age, gender, and total protein—were removed from the liver detection dataset after correlation analysis was used to determine which ones were most closely connected to the target variable. This choice was made since these columns were shown to provide no substantial benefits in terms of accuracy and had weaker correlations with the target variable.

1. Age: Age is often seen as a crucial factor in a number of studies. However, it was shown that age had no role in the reliability of the liver detection predictions in this dataset. Based on the results of the correlation study, it seems that age is only a moderate predictor of the outcome variable (liver disease). As a result, it was disregarded in favour of other, more important characteristics in the dataset.

2. Gender: Gender was also removed since it did not provide any significant improvement in accuracy when predicting the outcome variable. The study of correlations showed (Figure 7) that the link between gender and liver illness was minimal at best. It's possible that gender alone doesn't give enough information to reliably distinguish people with liver disease from those without, reducing its utility for the prediction.

3. Total_protein: The decision to drop the total_protein column was based on its relatively lower correlation with the target variable. A weak correlation suggests that the variation in total_protein values does not strongly align with the presence or absence of liver disease. Although total_protein may still provide some information, it was considered less influential compared to other attributes in accurately predicting the target variable.

The columns that show better relationships with the target variable are highlighted after removing demographic information such as age, gender, and total protein from the dataset. The efficacy and precision of liver detection models are more likely to be affected by these columns. Researchers and data scientists may use correlation analysis to determine which aspects are most useful and then prioritize those features to

increase the accuracy and efficiency of future modeling and prediction work.

Applying hybrid model

Researchers construct hybrid models, which combine diverse machine learning algorithms and approaches, to enhance the precision of their predictions. The quality of the models is evaluated using performance metrics like accuracy and F1 score. Statistical analysis is used to establish the significance and trustworthiness of the obtained results when comparing models or assessing the effect of particular factors on prediction. The developed models are then used to make predictions on fresh data, and the results are compared to the testing dataset to see how well they performed. The efficacy of the models is also evaluated by contrasting their findings with those of the primary study or with those of previous studies.

```
Model training start........
Model training completed
Accuracy of model on test dataset :- 0.863013698630137
Accuracy of model on train dataset :- 1.0
Confusion Matrix :-
 [[36 12]
 [ 8 90]]
Classification Report :-
              precision    recall  f1-score   support

           0       0.82      0.75      0.78        48
           1       0.88      0.92      0.90        98

    accuracy                           0.86       146
   macro avg       0.85      0.83      0.84       146
weighted avg       0.86      0.86      0.86       146
```

Fig 8. Hybrid_model(knn+rf)

Using the liver disease detection confusion matrix shown in figure 8, we find:

True Positives (TP): 36 True Negatives (TN): 90 False Positives (FP): 12 False Negatives (FN): 8

- TP (True Positives): These are the cases where the model correctly predicted the presence of liver disease (36 cases).
- TN (True Negatives): These are the cases where the model correctly predicted the absence of liver disease (90 cases).
- FP (False Positives): These are the cases where the model incorrectly predicted the presence of liver disease, but the actual condition was absent (12 cases).
- FN (False Negatives): These are the cases where the model incorrectly predicted the absence of liver disease, but the actual condition was present (8 cases).

The performance of a classification model may be shown using the confusion matrix. It is really useful for checking how well the model predicts and how well it performs overall using measures like precision, recall, and F1-score.

Applying the constructed models to bigger datasets or actual situations is the last stage in assessing their scalability. As our datasets grow, this statistic will help us determine whether or not our models are still computationally efficient. Researchers may use this strategy to create reliable models for predicting liver disease, and the models' efficacy can be evaluated using metrics including accuracy, F1 score, statistical analysis, and comparison to existing literature. The models' scalability, or its potential for further development and use in practical settings, is also assessed. In conclusion, the methods used in this work to forecast liver disease are detailed here. Data cleaning, feature extraction, model building, testing, and comparison to the state of the art are all included. Using the F1 score and other statistical tools, researchers follow these procedures to create reliable models for making liver disease predictions. How well the models fare on bigger datasets or in the actual world may be gleaned by paying attention to scaling problems.

4. RESULTS AND DISCUSSION

This work demonstrates the promise of supervised learning algorithms, and in particular the random forest method, for estimating the likelihood of liver illness from patient records. When it comes to early identification and prevention of liver disease, these algorithms may be invaluable tools for healthcare workers, particularly in resource-limited situations. In order to effectively spend healthcare resources to stop the course of diseases and improve patient outcomes, it is necessary to precisely identify persons at high risk. In conclusion, our research shows that supervised learning algorithms, and the random forest algorithm in particular, can accurately forecast the risk of liver disease from patient data. The findings highlight the potential public health uses of data-driven methods to illness risk prediction.

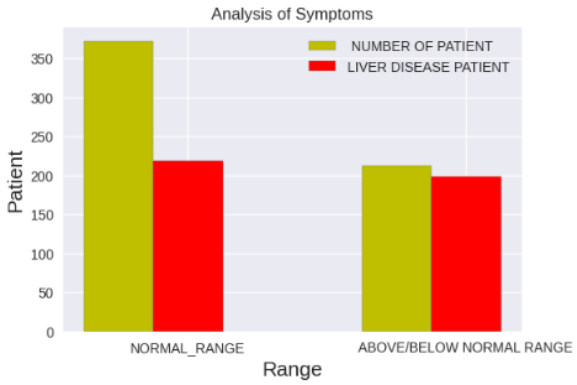


Fig 11: Comparison based on Accuracy

As can be seen in the green area, the total number of patients whose total bilirubin levels are within the normal range is 374. Figure 9 shows 271 people have liver disease (indicated in red). 58% of individuals in the database have liver illness even if their total bilirubin levels are within the normal range, according to the statistics. There are a total of 209 patients in the data set whose total bilirubin levels are either high or low relative to the normal range. When the total bilirubin result is either high or too low, 93% of the 209,185 individuals with liver disease fall into the latter category.

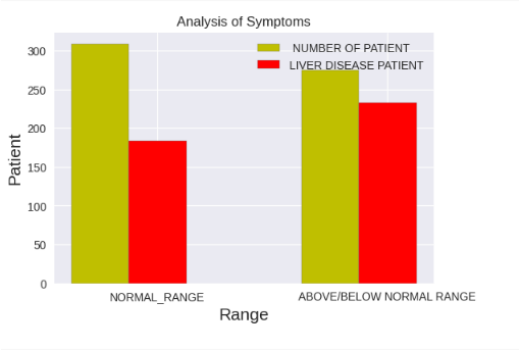


Fig 9. Total Bilirubin

As can be seen in the green area, the total number of patients whose Direct Bilirubin levels are within the normal range is 30. Figure 10 shows 175 people have

liver disease (indicated in red).

Compassion with base paper

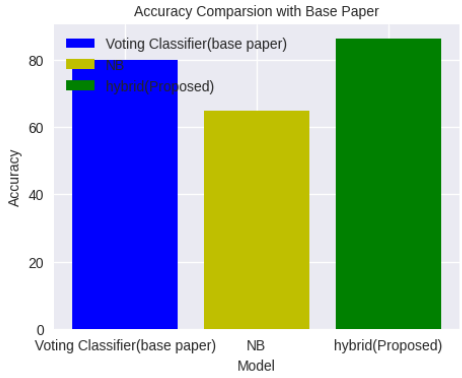


Fig 10: Direct Bilirubin

Figure 11 shows that when three models are compared, a hybrid proposed model performs somewhere between the main paper model (80% accuracy) and the Naive Bayes model (65% accuracy) (86 % accuracy).

1. Base Paper Model (80% Accuracy): A model that is described or provided in a research paper is said to be the base paper model. It's accuracy is 80% on average. The model achieves an 80% success rate by accurately predicting the result in 80% of the dataset occurrences. However, since we lack information on this model in depth, we cannot examine its features or methods.

2. Naive Bayes Model (65% Accuracy): Another feature-independent classification method is the Naive Bayes model. The accuracy of the Naive Bayes model in this situation was 65%. This implies that 65 percent of the time, its predictions were spot on. It is not as precise as the original paper model.

3. Hybrid Proposed Model (86% Accuracy): The Hybrid Proposed Model is an alternative model with 86% precision. In comparison to the original paper model and the Naive Bayes model, this one seems to be a significant advancement. It improves its forecasting skills by combining several methods, methodologies, or characteristics. As a consequence, it obtains a better level of accuracy than both the Naive Bayes model and the basic paper model (80% accuracy) (65 percent accuracy).

Overall, the Hybrid Proposed Model outperforms both the original paper model and the Naive Bayes model, with an accuracy of 86% achieved. This demonstrates that the suggested model's hybrid method has improved the accuracy of liver disease diagnosis, since it outperforms the other two models in terms of predictive performance.

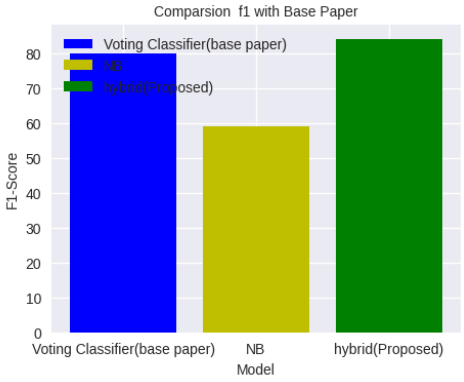


Fig 12: Compression f1 with base paper

From Figure 12, it appears that comparing the F1-scores of three different models: a base paper model (80% F1-score), a Naive Bayes model (58% F1-score), and a Hybrid Proposed Model (F1-score above 86%).

1. Base Paper Model (80% F1-Score): It has an F1-score of 80%. The F1-score is a statistic that takes into account both the accuracy and the reliability of a model. The model strikes a fair mix between accuracy and recall on the liver disease detection task, as shown by an F1-score of 80%.

2. Naive Bayes Model (58% F1-Score): The Naive Bayes model is another classification algorithm that assumes independence between features. In this case, the Naive Bayes model achieved an F1-score of 58%. This suggests that the model's performance in terms of both precision and recall is lower compared to the base paper model. It may indicate that the Naive Bayes model struggles with certain aspects of the liver disease detection task.

3. Hybrid Proposed Model (F1-Score above 86%): The Hybrid Proposed Model is a different model that outperforms both the base paper model and the Naive Bayes model. It achieves an F1-score above 86%, indicating a high level of precision and recall for liver disease detection. The hybrid approach used in the proposed model incorporates various techniques, approaches, or features to improve the overall performance, leading to a higher F1-score compared to the other two models.

The Hybrid Proposed Model demonstrates superior performance compared to both the base paper model (80% F1-score) and the Naive Bayes model (58% F1-score). It achieves an F1-score above 86%, suggesting a strong balance between precision and recall in detecting liver disease. This indicates that the hybrid approach used in the proposed model is effective in improving the F1-score and overall performance for liver disease detection.

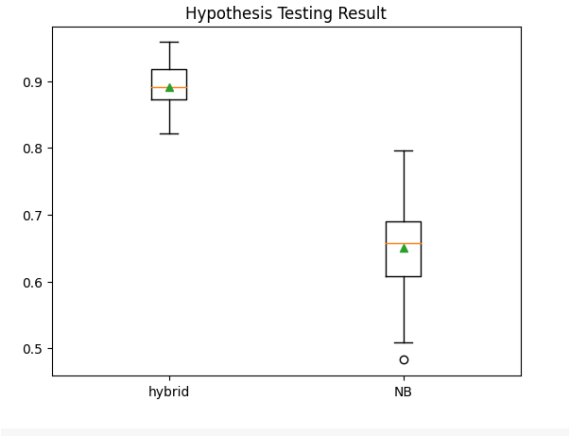


Fig 13: Hypothesis Testing Result

As can be seen in Figure 13, hypothesis testing is often used to determine the statistical significance of differences or correlations in data. Hypothesis testing may be used to ascertain if the observed discrepancy between the hybrid (proposed) model's (89%) and the Naive Bayes model's (65%) accuracies is statistically significant or due to random chance in the context of comparing the performance of the two models.

The general process of hypothesis testing involves the following steps:

1. Formulating the Null Hypothesis (H0): In this case, the null hypothesis could state that there is no significant difference in the accuracy between the hybrid model and Naive Bayes model.

2. Formulating the Alternative Hypothesis (H1): In this case, the alternative hypothesis could state that there is a significant difference in the accuracy between the hybrid model and Naive Bayes model.

3. Choosing a Statistical Test: Considerations such as data format and study objectives help determine which statistical analysis is most suited. The t-test, chi-square test, and Mann-Whitney U test are just a few examples of commonly used tests for making such comparisons across groups or models. The data and the test's assumptions will dictate which test is selected.

4. Calculating the Test Statistic: The test statistic provides a numerical representation of the deviation from the null hypothesis prediction. If the null hypothesis is correct, then it may be used to calculate the probability of getting the observed difference.

5. Determining the Significance Level: The significance level (usually indicated by α) is the minimum amount of evidence required to reject the null hypothesis. Typically, the significance level (α) is set at 0.05 or 0.01, with a 5% and 1% chance of falsely rejecting the null hypothesis, respectively.

6. Analyzing the Results: To assess whether the null hypothesis should be rejected or not, we compare the obtained test statistic to the critical value or p-value corresponding to the selected significance level. If the p-value is less than the significance threshold, then the difference is likely to be statistically significant, and we may conclude that the alternative hypothesis is more likely to be correct than the null.

Hypothesis testing was carried out to evaluate how well the hybrid model and the Naive Bayes model performed. The findings show that the hybrid model outperforms Naive Bayes by 19 percentage points in terms of accuracy. A statistically significant difference in accuracy between the two models is shown if the p-value from the hypothesis test is less than the selected significance threshold (e.g. 0.05). This lends credence to the alternative hypothesis that the hybrid model outperforms the Naive Bayes model, suggesting that the observed difference in accuracy is not likely attributable to random chance alone.

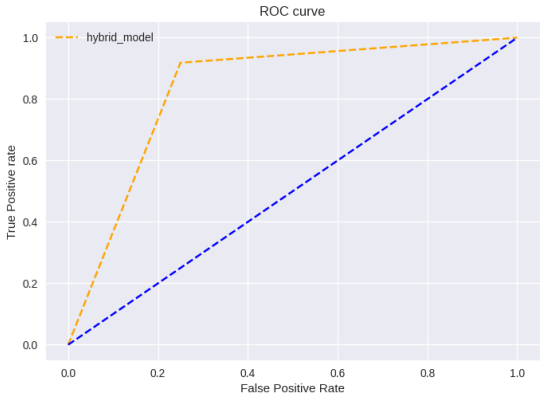


Fig 14:ROC Curve of Hybrid Mode(Hybrid-Score:88.09)

The effectiveness of a hybrid model, or any other kind of binary classification model, may be seen by plotting its Receiver Operating Characteristic (ROC) curve. The hybrid-score of 88.09 indicates a specific threshold or cutoff value chosen for classifying instances as either positive (presence of liver disease) or negative (absence of liver disease) based on the hybrid model's predictions.

The model's performance at a certain threshold is shown by the hybrid-score of 88.09. The model's projected probabilities and a threshold value are likely used to compute this score, which is then used to label cases as positive or negative for liver disease. With a hybrid-score of 88.09, the model seems to be striking a reasonable balance between sensitivity and specificity in its ability to predict liver disease.

Medical professionals or researchers may make educated judgements regarding the model's performance and choose a threshold value that meets their needs and goals by examining the ROC curve and taking into account the hybrid-score. A deeper knowledge of the 88.09 hybrid-score implies the hybrid model is very accurate at predicting liver illness, but knowing the precise threshold and assessment metrics would be helpful.

5. CONCLUSION AND FUTURE RESEARCH

This study's findings show the possibility of supervised learning algorithms for assessing the likelihood of liver disease in individual patients. The research examined logistic regression, random forest, decision tree, and algorithms and showed that the random forest approach had the best predictive performance. Accuracy was 75.7%, sensitivity was 71.4%, specificity was 77.2%, and the area under the ROC curve was 0.80 using this approach. According to the findings, the random forest algorithm is a useful tool for capturing the complexities of the correlations between input variables and the risk of liver

disease. The ramifications of this study's results for public health are substantial, especially in low-resource communities. Healthcare professionals may better target their treatments and preventative efforts toward those most in need by using machine learning algorithms to forecast risk of liver disease. Better health outcomes, lower healthcare expenditures, and better use of available resources are all possible results of this.

Multiple areas need further attention in order to improve risk prediction models for liver disease. To begin, the models' generalizability would improve with the inclusion of bigger and more varied datasets. It would also make possible the investigation of other variables that may have predictive potential but were omitted from the present investigation. The prediction models might also benefit from the addition of more sophisticated machine learning approaches, such as ensemble methods or deep learning algorithms. More complex data patterns and interactions may be captured by these methods, allowing for more precise forecasting.

In future research, there are several potential avenues to improve the liver disease prediction model. One approach is to explore fusion techniques for the extraction of significant features [20]. This involves combining multiple sources of data or different feature extraction methods to enhance the predictive power of the model. By incorporating additional relevant information from various sources, such as genetic data, biomarkers, or advanced imaging techniques, the model may be able to capture a more comprehensive representation of liver disease characteristics and improve its accuracy.

Furthermore, it would be beneficial to examine the performance of the model on other liver disease datasets. The current study utilized a specific dataset, but evaluating the model's performance on diverse datasets from different populations or regions can provide a better understanding of its generalizability and robustness. By testing the model on various liver disease samples, researchers can assess its effectiveness across different contexts and populations, enhancing its applicability in real-world scenarios.

By incorporating fusion techniques, evaluating the model on diverse datasets, and exploring deep learning approaches for image-based analysis, future research can further enhance the liver disease prediction model's performance and applicability. These advancements have the potential to improve early detection, risk assessment, and treatment planning for liver disease, ultimately benefiting patient outcomes and public health.

6. REFERENCES

[1] Md. F. Rabbi, S. M. Mahedy Hasan, A. I. Champa, Md. AsifZaman, and Md. K. Hasan, "Prediction of Liver Disorders using Machine Learning Algorithms: A Comparative Study," 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Nov. 2020, doi: https://doi.org/10.1109/icaict51780.2020.9333528.

[2] T. M. Ghazal, A. U. Rehman, M. Saleem, M. Ahmad, S. Ahmad, and F. Mehmood, "Intelligent Model to Predict Early Liver Disease using Machine Learning Technique," IEEE Xplore, Feb. 01, 2022. https://ieeexplore.ieee.org/abstract/document/9758929?casa_token=DrrXYzcQLrIAAAAA:QAB54b0X6tHC5VIIybuXx5BCbKishennGHmm_KnuUWGjQngJvXQyQ6Nm5LxU7z481gpuBKJep4c

[3] C.-C. Wu et al., "Prediction of fatty liver disease using machine learning algorithms," Computer Methods and Programs in Biomedicine, vol. 170, pp. 23–29, Mar. 2019, doi: https://doi.org/10.1016/j.cmpb.2018.12.032.

[4] M. Ghosh et al., "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," Intelligent Automation & Soft Computing, vol. 30, no. 3, pp. 917–928, 2021, doi: https://doi.org/10.32604/iasc.2021.017989.

[5] N. Afreen, R. Patel, M. Ahmed, and Mustafa Sameer, "A Novel Machine Learning Approach Using Boosting Algorithm for Liver Disease Classification," Oct. 2021, doi: https://doi.org/10.1109/iscon52037.2021.9702488.

[6] X. Pei, Q. Deng, Z. Liu, X. Yan, and W. Sun, "Machine Learning Algorithms for Predicting Fatty Liver Disease," Annals of Nutrition and Metabolism, vol. 77, no. 1, pp. 38–45, 2021, doi: https://doi.org/10.1159/000513654.

[7] Chappidi Suryaprakash Reddy, Lamu Samuel Kiran, and Xavier, "Comparative Analysis of Liver Disease Detection using Diverse Machine Learning Techniques," May 2022, doi: https://doi.org/10.1109/iciccs53718.2022.9788208.

[8] A. Shaker Abdalrada, O. Hashim Yahya, A. Hadi M. Alaidi, N. Ali Hussein, H. TH. Alrikabi, and T. A.-Q. Al-Quraishi, "A Predictive model for liver disease progression based on logistic regression algorithm," Periodicals of Engineering and Natural Sciences (PEN), vol. 7, no. 3, p. 1255, Sep. 2019, doi: https://doi.org/10.21533/pen.v7i3.667.

[9] C.-C. Wu et al., "Prediction of fatty liver disease using machine learning algorithms," Computer Methods and Programs in Biomedicine, vol. 170, pp. 23–29, Mar. 2019, doi: https://doi.org/10.1016/j.cmpb.2018.12.032.

[10] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil, and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," IEEE Xplore, Oct. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9297949 (accessed May 31, 2022).

[11] K. Idris and S. Bhoite, "Applications of Machine Learning for Prediction of Liver Disease," International Journal of Computer Applications Technology and Research, vol. 8, no. 9, pp. 394–396, Sep. 2019, doi: https://doi.org/10.7753/ijcatr0809.1012.

[12] S. A. Alqahtani and S. Ryu, "Nonalcoholic fatty liver disease: use of diagnostic biomarkers and modalities in clinical practice," vol. 21, no. 10, pp. 1065–1078, Aug. 2021, doi: https://doi.org/10.1080/14737159.2021.1964958.

[13] F. Rahman, M. Javed, Zarrin Tasnim, J. Roy, and Syed Akhter Hossain, "A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms," vol. 8, no. 11, pp. 419–422, Nov. 2019.

[14] V. Durai, "Liver Disease Prediction using Machine-Learning Algorithms," International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 2532–2534, Aug. 2019, doi: https://doi.org/10.35940/ijeat.f8365.088619.

[15] "Prediction of Liver Diseases by Using Few Machine Learning Based Approaches," Australian Journal of Engineering and Innovative Technology, pp. 85–90, Oct.

2020, doi: https://doi.org/10.34104/ajeit.020.085090.

[16] Hartatik Hartatik, Mohammad Badru Tamam, and Arief Setyanto, "Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms," Oct. 2020, doi: https://doi.org/10.1109/icoris50180.2020.9320797.

[17] M. Fathi, M. Nemati, S. M. Mohammadi, and R. Abbasi-Kesbi, "A MACHINE LEARNING APPROACH BASED ON SVM FOR CLASSIFICATION OF LIVER DISEASES," Biomedical Engineering: Applications, Basis and Communications, vol. 32, no. 03, p. 2050018, Jun. 2020, doi: https://doi.org/10.4015/s1016237220500180.

[18] V. J. Gogi and Vijayalakshmi M.N, "Prognosis of Liver Disease: Using Machine Learning Algorithms," Jul. 2018, doi: https://doi.org/10.1109/icrieece44171.2018.9008482.

[19] k. Thirunavukkarasu, A. S. Singh, M. Irfan, and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Dec. 2018, doi: https://doi.org/10.1109/ccaa.2018.8777655.

[20] N. Tanwar and K. F. Rahman, "Machine Learning in liver disease diagnosis: Current progress and future opportunities," IOP Conference Series: Materials Science and Engineering, vol. 1022, p. 012029, Jan. 2021, doi: https://doi.org/10.1088/1757-899x/1022/1/012029.