IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning Techniques For The Prognosis Of Liver Disease

[1]Dr. A. Antony Prakash

[1] Department of Information Technology, St. Joseph's college, Trichy, Tamilnadu, India

Abstract: The human body relies on the healthy liver to perform over 500 essential functions, and any dysfunction can have severe consequences, potentially leading to fatality. Timely identification and treatment of liver diseases can significantly enhance the chances of survival. In this regard, machine learning (ML) emerges as a valuable resource that can aid healthcare experts in diagnosing hepatic patients. The conventional ML framework encompasses various techniques such as data pre-processing, feature extraction, and classification, which collectively contribute to accurate diagnoses. Liver disease is a matter of great concern in the global health landscape, impacting a vast number of people across the world. The timely and precise identification of liver disease plays a pivotal role in ensuring successful treatment and averting potential complications. Over the past few years, the advent of machine learning has revolutionized the healthcare sector, empowering the creation of predictive models that aid in diagnosing and forecasting diverse medical ailments, liver disease being one of them. The proposed approach incorporates several ML algorithms, including logistic regression (LR), random forest (RF), and confusion matrix. The results indicate that the suggested system has the potential to complement the diagnosis of liver disease made by a physician.

Index Terms **- Liver disease, Machine learning, Prediction, Data analytics, Healthcare, Autoencoders**

I. INTRODUCTION

The latest data from the World Health Organization reveals that liver disease caused 264,193 deaths in India in 2018, with an age-adjusted death rate of approximately 23.00 per 100,000 people [1]. The liver, which weighs around 1.36 kg, is the largest organ in the body and is made up of four lobes of varying sizes and shapes. It is located behind the diaphragm beneath the abdominal cavity and is supplied with blood by the hepatic artery and the portal vein [2]. The liver's primary function is to remove harmful substances from the bloodstream before they can cause damage to other parts of the body. Liver disease is a serious and often fatal condition that can be caused by a variety of factors, including hepatitis infection, fatty liver, cirrhosis, liver fibrosis, high alcohol consumption, drug exposure, and genetic abnormalities. If the liver fails completely, a liver transplant may be the only option for treatment. Early detection of liver disease is crucial for successful treatment and recovery, as the disease progresses through several stages, including healthy, fibrosis, cirrhosis, and cancer. However, detecting liver disease in its early stages can be challenging, and failure to provide timely treatment can lead to further damage to liver tissue [3].

Additional indications of liver disease may involve abnormalities in the brain and nervous system, such as memory loss, numbness, and fainting, as well as skin issues like yellowing, spider veins, and redness in the feet. To avoid liver diseases, it is recommended to see a doctor frequently, receive vaccinations, reduce soda and alcohol consumption, exercise regularly, and maintain a healthy weight. The emergence of artificial intelligence has resulted in the creation of various machine learning algorithms that improve the precision and efficiency of diagnosing and predicting liver disease [4].

Classification approaches are widely used in numerous automatic medical diagnostic tools. Detecting liver diseases at an early stage is challenging as they do not show symptoms until the organ is partially damaged [5]. However, the presence of certain enzymes in the blood can be utilized to identify liver diseases [5].

Additionally, the use of mobile devices to monitor human health is becoming more prevalent. In such cases, automatic classification algorithms are essential. By employing mobile and online technologies that can automatically identify liver illnesses, patient wait times with liver specialists like endocrinologists can be reduced.

Liver disease is a prevalent and serious global health issue that can lead to various complications if not addressed early on (Dutta et al., 2022) [6]. According to a report by the World Health Organization (WHO) in 2018, liver diseases accounted for approximately one million deaths worldwide, ranking it as the 11th leading cause of mortality (World Total Deaths, n.d.) [7]. Detecting liver disease in its early stages poses a challenge for medical professionals, as symptoms may not be apparent until the condition becomes chronic [8]. Moreover, the conventional diagnostic methods such as sonography, MRI scans, and CT scans are costly, potentially harmful, and associated with numerous side effects. Consequently, healthcare workers face the significant obstacle of predicting liver diseases at an early stage, while also ensuring affordability and an improved healthcare system for their treatment. Severe liver diseases manifest through symptoms such as indigestion, dry mouth, and abdominal pain, yellowing of the skin, numbness, memory loss, and fainting. These symptoms often go unnoticed in the initial stages and only become apparent when the disease progresses to a chronic state. However, even when the liver is partially affected, it can still maintain its functionality [8].

II. LITERATURE REVIEW

The prevalence of human diseases has increased significantly in recent decades. When comparing liver diseases to other severe illnesses, the number of individuals affected by liver diseases continues to rise steadily [9]. Liver diseases in their early stages often lack noticeable symptoms, making them difficult to detect. However, with the advent of modern databases, extracting data and gaining valuable insights has become a straightforward process, aiding in the effective treatment of various diseases [10].
The researcher employs multiple strategies to gain insights from the dataset, including the use of ML classifiers for feature selection or extraction. However, not all strategies involve ML. Currently, data is generated and stored without any special considerations. Nevertheless, this data has proven to be useful in solving problems in fields such as medical imaging, finance, genomics, transactions, and more. When dealing with a large volume of data, both necessary and unnecessary, it is important to select and extract the most relevant features to accurately predict diseases or other objects.

Pasha and Mohamed et al. [11] conducted a study on heart disease prediction using various datasets such as Cleveland, Hungarian, Statlog, and Switzerland. They employed a unique feature reduction (NFR) model in their methodology to accurately predict cardiac diseases. The process involved initial dataset processing, followed by the identification of significant features using statistical techniques like weighted least squares (WLS) and correlation matrices. The ML and Data Mining (DM) algorithms were then applied to the reduced set of features, and the AUC and accuracy were measured for each individual feature. In their proposed NRF model, boosted regression trees (BRT) achieved the highest AUC of 96.68% and an accuracy of 93.53% on the Cleveland dataset.
On the Hungarian dataset, LR achieved the highest AUC of 92.51% and an accuracy of 85.06%, followed by BRT, SGB, and SVM. For the Statlog dataset, BRT achieved the highest AUC of 91.79% and an accuracy of 87.65% when compared to SVM and RF. Lastly, on the Switzerland dataset, BRT achieved the highest AUC of 99.20% and an accuracy of 95.52%. The results demonstrate that the proposed NFR model outperforms the model without NFR in terms of AUC and accuracy in predicting heart disease.
Four classification techniques were implemented by Gan et al. [12], namely AdaC-TANBN, TANBN, BN, and SVM. The integrated TANBN using a cost-sensitive method (AdaC-TANBN) yielded an accuracy of 69.03% after experimentation, surpassing the results of the other techniques.

An alternative approach was introduced by Anagaw et al. [13], known as the Compliment Naive Bayesian (CNB) classification method. This method was compared to the naive Bayes classifier as well as several other classifiers. The CNB method achieved a superior accuracy of 71.36% compared to the other methods. Wu et al. (2019) conducted a prediction analysis on patients diagnosed with Fatty Liver Disease (FLD). The researchers obtained a dataset of 700 patient records from New Taipei Hospital, which included screening tests for fatty liver disease. After considering the patient's age and availability of sufficient data, 577 records

were included in the analysis [14]. Out of these 577 patients, 377 were diagnosed with fatty liver disease, while the remaining patients did not have the condition.

The dataset consisted of various patient health details, including age, gender, systolic and diastolic blood pressure, abdominal girth, glucose level, and triglyceride, HDL-C, SGOT-AST, and SGPT-ALT. To pre-process the data, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied, followed by normalization. For the prediction task, four machine learning algorithms were employed: Random Forest, Naïve Bayes, Artificial Neural Network, and Logistic Regression. These algorithms were evaluated using 3, 5, and 10-fold cross-validation. In addition to assessing the accuracies of the models, the area under the receiver operating curve (AUC-ROC) was also examined for each algorithm. Among the four algorithms, Random Forest consistently demonstrated the highest accuracy across all cross-validations.

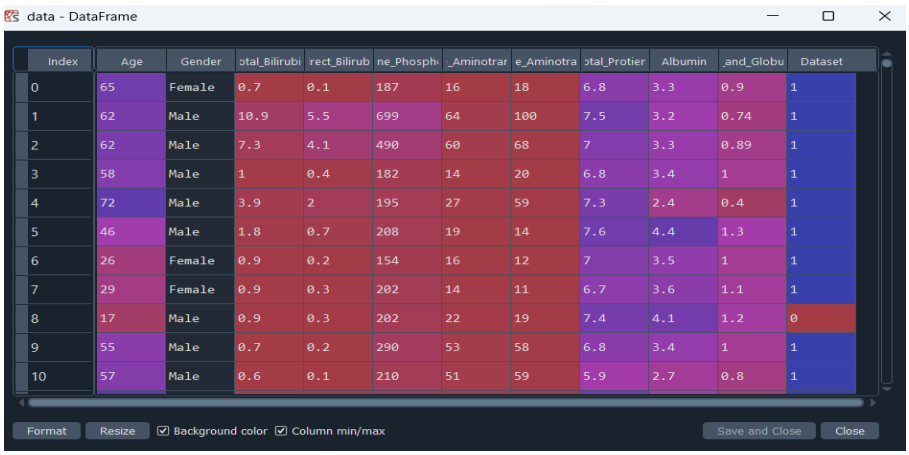
Singh and colleagues (2020) conducted a study on liver disease prediction using various classification techniques and feature selection, along with the development of software for easy prediction [15]. The research was carried out on the Indian Liver Patient Records dataset, where some attributes were eliminated during the feature selection phase using the Correlation-based Feature Selection Subset Evaluator with the Greedy Stepwise search method in WEKA. Only five attributes were selected, namely Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, Alamine Aminotransferase, and Aspartate Aminotransferase. Six different classification methods were applied, including Logistic Regression, Naïve Bayes, Sequential Minimal Optimization (SMO), Random Forest, Instant based Classification (IBk), and Logistic Regression provided the highest accuracy of 74.36%. On the other hand, Naïve Bayes produced the least accuracy of 55.9%.

Geetha et al. aimed to enhance the perceived nature of liver disease by employing machine learning techniques. The primary focus of their work was on classification using the random forest model, which involved utilizing various pre-processing techniques to address the issue of unbalanced data. To refine the model, hyper parameter tuning was conducted through grid search and feature selection [16]. However, it should be noted that the study does not definitively determine the superiority of the selected model. The majority of prior research has primarily focused on the analysis aspect rather than the pre-processing phase of the Indian Liver Patient Records dataset. Consequently, this study aims to address this gap by emphasizing the significance of pre-processing as a crucial stage in data analysis. Furthermore, several other algorithms have been employed in this research.

III. Research Methodology

3.1 Dataset

The number of individuals suffering from liver disease has been on a steady rise due to the overindulgence of alcohol, exposure to hazardous fumes, and consumption of tainted food, pickles, and medications. This particular dataset was utilized to assess forecasting algorithms with the aim of alleviating the workload of medical professionals. The ILPD dataset was collected from the North East region of Andhra Pradesh, India, comprising of 583 observations with ten features and one target output.



Index	Age	Gender	otal_Bilirubi	rect_Bilirub	ne_Phosph	_Aminotrar	e_Aminotra	otal_Protier	Albumin	_and_Globu	Dataset
0	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	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	3.3	0.89	1
3	58	Male	1	0.4	182	14	20	6.8	3.4	1	1
4	72	Male	3.9	2	195	27	59	7.3	2.4	0.4	1
5	46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3	1
6	26	Female	0.9	0.2	154	16	12	7	3.5	1	1
7	29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1	1
8	17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2	0
9	55	Male	0.7	0.2	290	53	58	6.8	3.4	1	1
10	57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8	1

Table 1. ILPD dataset features description.

For a more detailed overview of the dataset, refer to Table 1. Our research involved retrieving the dataset from the UCI ML repository for evaluation purposes. UCI hosts a wide range of databases, domain theories, and data generators, making it a central hub for machine learning and information systems. This valuable resource is utilized by various members of the machine learning community, including students, experts, researchers, instructors, and others, as the primary and essential source for evaluating machine learning problems. To Utilized the make classification command from the Python scikit-learn library to generate simulation data and validate our strategic plan on ILPD's data.

3.2 Content

There are a total of 416 liver patient records and 167 non-liver patient records included in this dataset. These records were collected from the North East region of Andhra Pradesh, India. The "Dataset" column serves as a class label, categorizing the patients into two groups: those with liver disease (liver patient) and those without any disease (no disease). Additionally, this dataset comprises 441 records of male patients and 142 records of female patients. For patients whose age surpasses 89, their age is recorded as "90".

3.3 Columns

- ✓ Age of the patient
- ✓ Gender of the patient
- ✓ Total Bilirubin
- ✓ Direct Bilirubin
- ✓ Alkaline Phosphatase
- ✓ Alamine Aminotransferase
- ✓ Aspartate Aminotransferase
- ✓ Total Proteins
- ✓ Albumin
- ✓ Albumin and Globulin Ratio
- ✓ Dataset: field used to split the data into two sets (patient with liver disease or no disease)

IV. Performance Analysis

The Indian Liver Patient Dataset (ILPD) from the UCI machine learning repository is utilized for liver disease classification. It comprises 11 columns consisting of 10 features and a target variable. These features include age, gender, total bilirubin (TB), direct bilirubin (DB), total proteins (TP), albumin (ALB), albumin and globulin ratio (A/G), alamine aminotransferase (SGPT), aspartate aminotransferase (SGOT), and alkaline phosphotase (Alkphos). It presents the characteristics of these features for the patients. The output variable distinguishes between patients with liver disease and those without, with two classes. The dataset encompasses 583 patient records collected from the North East region of Andhra Pradesh, India. The distribution of patients with and without liver disease is depicted in Figure 1.

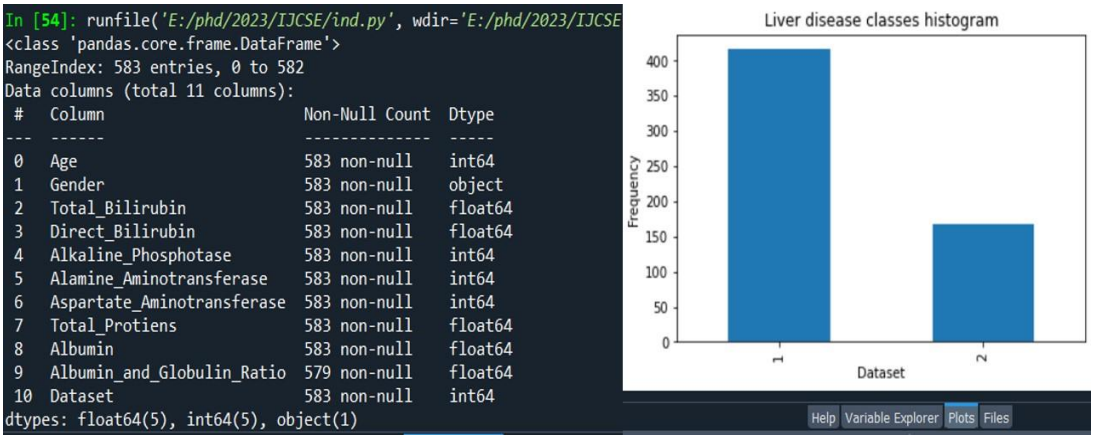


Figure 1 Distribution of patients with and without liver disease

The "Albumin_and_Globulin_Ratio" feature is not fully complete as it is missing 583 values. Hence, it is necessary to tackle this issue during the data pre-processing stage. Our next step involves evaluating the data balance by generating a histogram visualization. In order to convert the categorical data into numerical values, we made use of the conventional pandas function known as "get_dummies". Given that there is only a single column that needs encoding, this function proved to be adequate for the task.

Index	Age	Total_Bilirubin	Direct_Bilirubin	aline_Phosphota	ne_Aminotransf	ate_Aminotrans	Total_Protiens	Albumin	n_and_Globulin	ender_Fema	ender_Ma
0	1.2521	-0.418878	-0.493964	-0.426715	-0.354665	-0.318393	0.29212	0.198969	-0.123797	True	False
1	1.06664	1.22517	1.43042	1.68263	-0.0915993	-0.0343326	0.937566	0.0731566	-0.612077	False	True
2	1.06664	0.644919	0.931508	0.821588	-0.113522	-0.145186	0.476533	0.198969	-0.154315	False	True
3	0.819356	-0.370523	-0.387054	-0.447314	-0.365626	-0.311465	0.29212	0.324781	0.181378	False	True
4	1.68484	0.0969023	0.183135	-0.393756	-0.294379	-0.176363	0.753153	-0.93334	-1.64967	False	True
5	0.0775138	-0.241578	-0.280143	-0.340199	-0.338224	-0.33225	1.02977	1.5829	1.0969	False	True
6	-1.15889	-0.386642	-0.458327	-0.562669	-0.354665	-0.339178	0.476533	0.450593	0.181378	True	False
7	-0.97343	-0.386642	-0.42269	-0.364918	-0.365626	-0.342642	0.199913	0.576405	0.486552	True	False
8	-1.71527	-0.386642	-0.42269	-0.364918	-0.321782	-0.314929	0.84536	1.20547	0.791727	False	True
9	0.633896	-0.418878	-0.458327	-0.00237437	-0.151885	-0.179827	0.29212	0.324781	0.181378	False	True
10	0.757536	-0.434996	-0.493964	-0.331959	-0.162846	-0.176363	-0.53774	-0.555904	-0.428972	False	True

Figure 2 get dummies in one column

Correlation is employed to ascertain the connection between the features or the output variable. It quantifies the linear association between variables. The correlation coefficient may be positive (the output variable value rises as one feature value increases), negative (the output variable value declines as one feature value increases), or zero (indicating no relationship between variables).A valuable approach is to utilize the "corr()" function and generate a heatmap in order to examine the relationships between the features. This enables a visual representation of the correlations between the features.

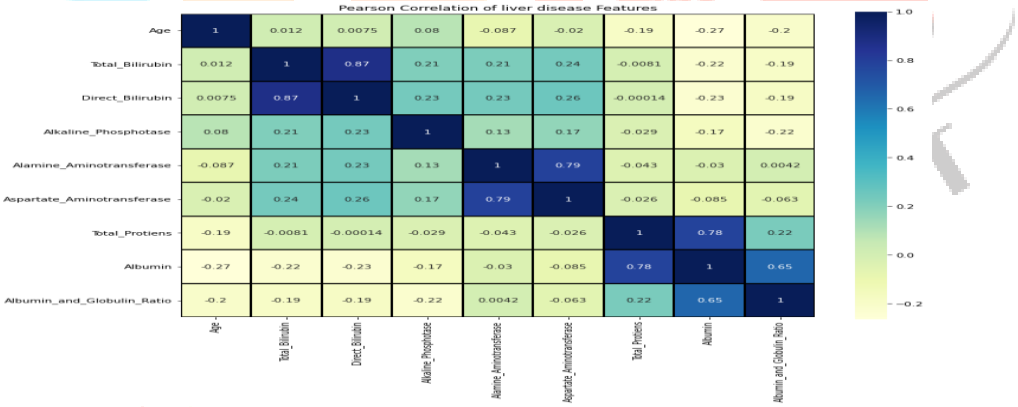
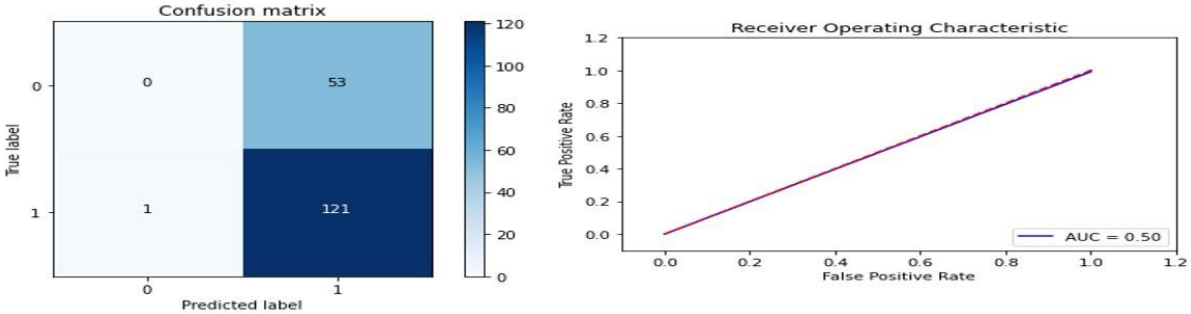


Figure 3 Pearson correlation of liver disease feature

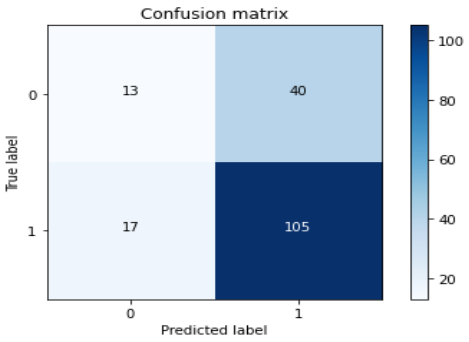
The performance of the Support Vector Classifier (SVC) will now be evaluated on the dataset without using any sampling techniques. This evaluation will be based on the heatmap analysis, which clearly indicates a strong correlation between certain pairs of features. Specifically, there is a high correlation between "Direct_Bilirubin" and "Total_Bilirubin," "Alamine Aminotransferase" and "Aspartate Aminotransferase," and "Total Protiens" and "Albumin."



The confusion matrix reveals that the algorithm has produced no true negatives, indicating an erroneous outcome. This implies that the algorithm is unbalanced and consistently predicts the presence of liver disease in patients. Therefore, it is imperative to fine-tune the model.

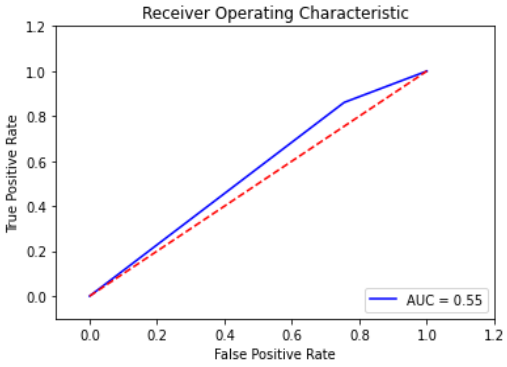
The ROC curve and confusion matrix analysis clearly indicate the importance of reducing false positives as they signify inaccurate predictions. To enhance the model's performance, we have employed the GridSearchCV technique.

```
Unoptimized model
------
Accuracy score on testing data: 0.6914
F-score on testing data: 0.9139

Optimized Model
------
Final accuracy score on the testing data: 0.6743
Final F-score on the testing data: 0.8294
SVC(C=100, random_state=0)
```



The ROC curve is anticipated to exhibit enhanced performance with the incorporation of true negative cases.

```
Unoptimized model
------
Accuracy score on testing data: 0.6914
F-score on testing data: 0.9139

Optimized Model
------
Final accuracy score on the testing data: 0.6743
Final F-score on the testing data: 0.8294
SVC(C=100, random_state=0)
Recall metric in the testing dataset:  0.860655737704918
0.5529693782864212
```

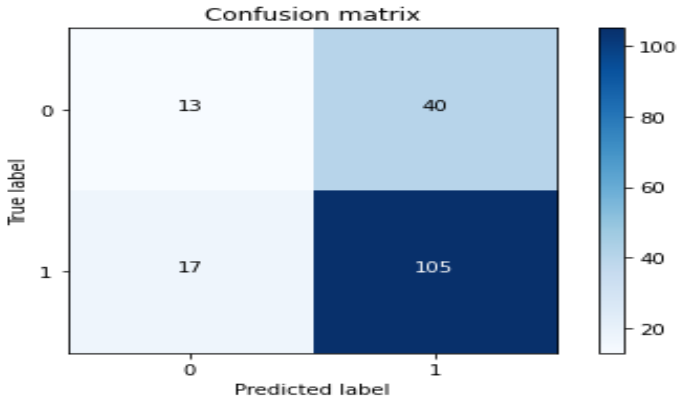


The AUC of 0.58 on the ROC curve is an improvement from the unoptimized model, but it still falls short of being a highly effective model. The dataset's unbalanced nature limits the AUC improvement, and the model's performance is also restricted by the dataset's small size. To address this, I will use the oversampling technique to balance the dataset and increase the data volume.

Support Vector Machines (SVM) is widely recognized as a classification technique that can be utilized for both classification and regression tasks. It possesses the capability to effectively handle numerous continuous and categorical variables.

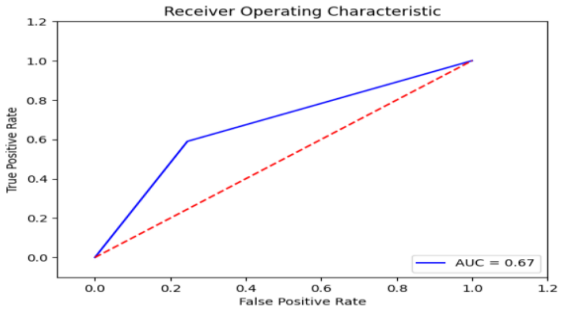
```
SVC
SVC(random_state=0)
```

By constructing a hyperplane in a multidimensional space, SVM is able to separate distinct classes. Through an iterative process, SVM generates an optimal hyperplane that minimizes errors. The fundamental concept behind SVM is to identify a maximum marginal hyperplane (MMH) that optimally divides the dataset into classes. After applying SVM random state=0.



The recall measurement exhibits a suboptimal figure, signifying the necessity to enhance the model for betterment.

```
Optimized Model
------
Final accuracy score on the testing data: 0.6743
Final F-score on the testing data: 0.8294
SVC(C=100, random_state=0)
Recall metric in the testing dataset:  0.860655737704918
0.5529693782864212
Recall metric in the testing dataset:  0.860655737704918
0.4959016393442623
```

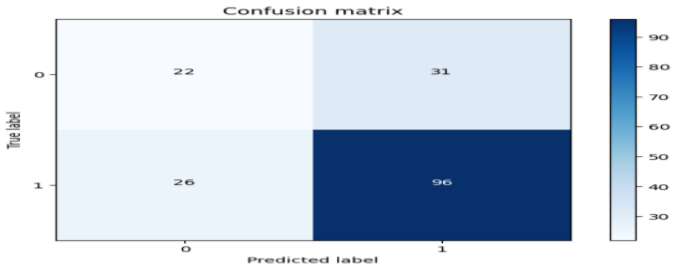


The utilization of the SMOTE technique did not yield satisfactory results in terms of the performance of SVC. The recall metric and AUC score both hover around 0.60, which does not meet the desired level of performance. Consequently, our decision was to investigate the RandomForestClassifier as an alternative approach.

```
RandomForestClassifier
RandomForestClassifier(random_state=0)
```

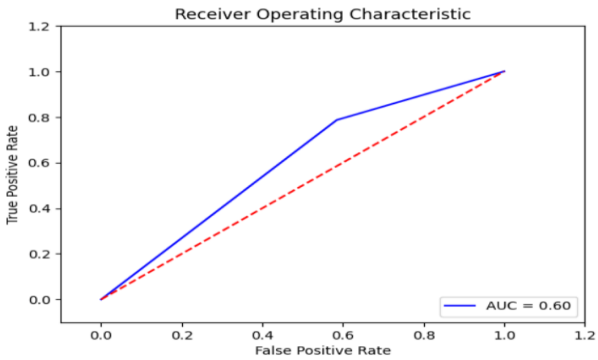
The recall metric has demonstrated enhancement following the utilization of the RandomForestClassifier in comparison to SVC.

```
Recall metric in the testing dataset:  0.7868852459016393
```

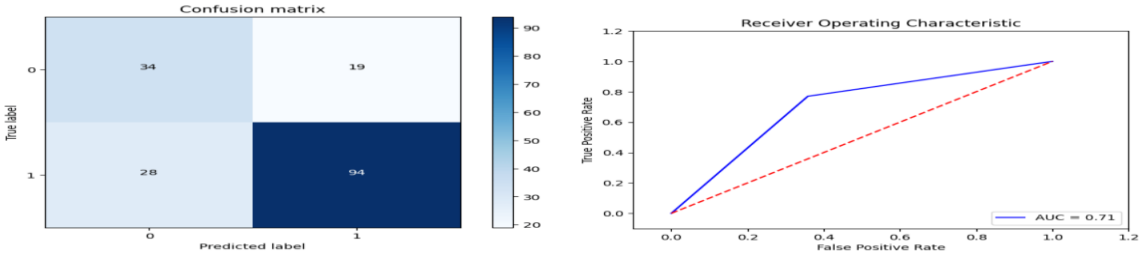


Nevertheless, additional refinement is still necessary to optimize the model's performance.

```
Unoptimized model
------
Accuracy score on testing data: 0.6914
F-score on testing data: 0.9139

Optimized Model
------
Final accuracy score on the testing data: 0.6743
Final F-score on the testing data: 0.8294
SVC(C=100, random_state=0)
Recall metric in the testing dataset:  0.860655737704918
0.5529693782864212
0.4959016393442623
```



The RandomForestClassifier model, after undergoing optimization using GridSearchCV, attained a recall metric of 0.76 and an AUC of 0.69 on the ROC curve.



V. RESULT ANALYSIS

The ensemble algorithms employed for liver disease classification are evaluated in comparison to previous studies that utilized the same dataset and evaluation methods. The outcomes of the proposed research surpass a significant number of the prior works. The approach presented in this study employs advanced pre-processing techniques and ensemble machine learning, outperforming numerous other research studies. The majority of these studies rely on basic machine learning models.

Among them, Support Vector Classifier obtained the accuracy level of 0.4959%, ROC curve and confusion matrix obtained the accuracy level of 0.5529%, ROC curve exhibits an improved AUC obtained the accuracy level of 0.672% and random forest classifier obtained the accuracy level of 0.706%. So the random forest accuracy 0.706 is and is much better than the results obtained than SVC and ROC. Overall, the extra tree classifier, which has not been used for liver disease classification before, surpasses all the other works with an accuracy 0.71%.

5.1 FINDINGS

Model	Random Forest Model	ROC curve and confusion matrix	Support Vector Machine (SVM)	ROC curve exhibits an improved AUC obtained
Accuracy level	0.706	0.5529	0.4959	0.672

VI. CONCLUSION

Machine learning has become an invaluable asset in predicting liver diseases, providing substantial advantages in terms of precision, early identification, and tailored medical treatments. Nevertheless, there are obstacles that must be overcome, including the availability of data, the interpretability of models, and ethical concerns. The potential for future progress in machine learning techniques is vast, promising even more precise and effective liver disease prediction. Through the utilization of machine learning, we can enhance patient outcomes and make substantial progress in the global fight against liver diseases. Liver disease is on the rise worldwide, primarily due to changes in lifestyle and unhealthy eating and drinking habits. Detecting the disease early can significantly improve survival rates. In order to tackle this problem, various ensemble models have been employed for liver disease diagnosis and their effectiveness has been compared to other models. The results revealed that the suggested model, which incorporates an enhanced pre-processing approach with an extra tree classifier, achieved the highest testing accuracy of 91.82%. The random forest model followed closely with an accuracy of 70.06%. This will help in increasing the training data and may improve the model accuracy further.

REFERENCES

[1] Liver Disease in India," World Life Expectancy. [(Accessed on 14 April 2022)].
[2] Sindhuja D.R.J.P., Priyadarsini R.J. A survey on classification techniques in data mining for analyzing liver disease disorder. Int. J. Comput. Sci. Mob. Comput. 2016; 5:483–488. [Google Scholar]
[3] Shaheamlung G., Kaur H., Kaur M. A Survey on machine learning techniques for the diagnosis of liver disease; Proceedings of the 2020 International Conference on Intelligent Engineering and Management (ICIEM); London, UK. 17–19 June 2020.
[4] Sun Q.-F., Ding J.-G., Xu D.-Z., Chen Y.-P., Hong L., Ye Z.-Y., Zheng M.-H., Fu R.-Q., Wu J.-G., Du Q.-W., et al. Prediction of the prognosis of patients with acute-on-chronic hepatitis B liver failure

using the model for end-stage liver disease scoring system and a novel logistic regression model. J. Viral Hepat. 2009;16:464–470. doi: 10.1111/j.1365-2893.2008.01046.x.

[5] Liu K.-H., Huang D.-S. Cancer classification using Rotation Forest. Comput. Biol. Med. 2008;38:601–610. doi: 10.1016/j.compbiomed.2008.02.007.

[6] Dutta, K., Chandra, S., & Gourisaria, M. K. (2022). Early-Stage detection of liver disease through machine learning algorithms. Lecture Notes in Networks and Systems, 318, 155–166. https://doi.org/10.1007/978-981-16-5689-7_14.

[7] World Total Deaths. (n.d.). World life expectancy. Retrieved October 28, 2021, from https://www.worldlifeexpectancy.com/world-rankings-total-deaths

[8] Devikanniga, D., Ramu, A., & Haldorai, A. (2020). Efficient diagnosis of liver disease using support vector machine optimised with crows search algorithm. EAI Endorsed Transactions on Energy Web, 7(29). https://doi.org/10.4108/EAI.13-7-2018.164177.

[9] J. Hassannataj, H. Saadatfar, A. Dehzangi, S. Shamshirband Informatics in Medicine Unlocked Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection

[10] S.J. Pasha, E.S. Mohamed , "Novel Feature Reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction", IEEE Access, 8 (2020), pp. 184087-184108, 10.1109/ACCESS.2020.3028714

[11] S.J. Pasha, E.S. Mohamed , "Advanced hybrid ensemble gain ratio feature selection model using machine learning for enhanced disease risk prediction", Inform Med Unlocked, 32 (June) (2022), Article 101064, 10.1016/j.imu.2022.101064

[12] D. Gan, J. Shen, B. An, M. Xu, N. Liu, "Integrating TANBN with cost-sensitive classification algorithm for imbalanced data in medical diagnosis", Comput Ind Eng, 140 (January) (2020), Article 106266, 10.1016/j.cie.2019.10626

[13] A. Anagaw, Y.L. Chang , "A new complement naïve Bayesian approach for biomedical data classification", J Ambient Intell Hum Comput, 10 (10) (2019), pp. 3889-3897, 10.1007/s12652-018-1160-1

[14] Wu, C.-C., Yeh, W.-C., Hsu, W.-D., Islam, M. M., Nguyen, P. A. A., Poly, T. N., Wang, Y.-C., Yang, H.-C., & Li, Y.-C. J. (2019). Prediction of fatty liver disease using machine learning algorithms. Computer Methods and Programs in Biomedicine, 170, 23–29. https://doi.org/10.1016/j.cmpb.2018.12.032

[15] 1Singh, J., Bagga, S., & Kaur, R. (2020). Software-based prediction of liver disease with feature selection and classification techniques. Procedia Computer Science, 167, 1970–1980. https://doi.org/10.1016/j.procs.2020.03.226

[16] Lin R.-H. An intelligent model for liver disease diagnosis. Artif. Intell. Med. 2009;47:53–62. doi: 10.1016/j.artmed.2009.05.005.