



Liver Data Feature Selection Using Adaptive Lasso for Liver Disease Prediction

¹Nibin Mathew, ²Dr.R.Rangaraj,

¹Research Scholar, ²Associate Professor & Head,
Department of Computer Science,
Hindusthan College of Arts and Science,
Coimbatore – 641028, Tamilnadu, India.

Abstract - Liver disease is a serious health concern worldwide, and accurate prediction of liver disease is crucial for timely intervention and treatment. In this paper, research proposes a feature selection method using Adaptive Lasso for liver data to enhance the prediction accuracy of liver disease. Adaptive Lasso effectively selects relevant features from a large pool of potential predictors by incorporating penalty terms. Experimental results demonstrate that research proposed method achieves superior performance in liver disease prediction compared to traditional feature selection techniques. The selected features provide valuable insights into the underlying factors contributing to liver disease, enabling more targeted and effective healthcare interventions. This research contributes to the field of liver disease prediction and highlights the importance of feature selection for improved healthcare outcomes.

Keywords: Liver disease, feature selection, Adaptive Lasso, prediction accuracy, healthcare interventions;

1. Introduction

Liver disease is a critical public health problem around the world, with more than 1,000,000 deaths every year. Liver diseases have a few causes, including viral infections, alcohol abuse, metabolic disorders, and genetic mutations. Early detection and exact diagnosis are urgent for successful treatment and worked on patient results. Machine learning (ML) techniques, for example, decision trees, support vector machines, and neural networks, have been broadly applied to anticipate liver disease. Feature selection is a basic move toward ML demonstrating, as it assists with distinguishing the most significant indicators that work on the model's accuracy and diminish its complexity.

Liver disease is a developing concern around the world, influencing a large number of individuals consistently. It is assessed that more than 30 million Americans, or roughly 10% of the

US populace, have some type of liver disease. Early detection and precise prediction of liver disease can help in forestalling serious health difficulties and can build the possibilities of effective treatment. One of the vital difficulties in anticipating liver disease is distinguishing the main features or factors that are connected with the disease. The selection of applicable features is pivotal for building exact and successful prediction models. In this paper, investigate different feature selection techniques for liver disease prediction, and assess their adequacy in further developing prediction accuracy.

The liver is a vital organ in the body, performing functions like bile production, chemical detoxification, and production of important proteins necessary for blood coagulation. The prevalence of various liver ailments has recently increased dramatically over the entire planet. 2.4% of the population in India dies as a result of the illness. There are more than 100 different types of liver illnesses, and cirrhosis is one of them. It is

identified when the liver cells are damaged and replaced with dead scar tissues. One of the most common methods of identifying liver disease (LD) is to have a specialist radiologist check to see if the liver tissue is abnormal. Studies have shown that a simple visual diagnosis of liver illnesses can have an accuracy of about 72%. Early detection of LD is possible for quicker treatment because the vast majority of medical facilities, hospitals, and diagnosis centres are equipped with modern PC-based machines for testing and diagnosis. Using machine learning algorithms on the lab data, a model for a significantly more effective diagnosis can be created. Different categorization methods and examination in light of the information may result in varying accuracy rates.

The Indian Liver Patient Dataset provided by ics.uci.edu is used in this study. A few machine learning techniques for classification can arrange the liver illnesses. The study takes a piece by piece approach to how to optimize the ML module for the LASSO method rather than choosing the algorithm that offers better execution. Robert Tibshirani introduced LASSO in 1996 as a technique for determining a selection of pertinent indications for a specific reaction variable. Since the LASSO technique is built on a few samples of information obtained by dividing information, the model isn't optimized for a lot of specific information. The paper's main goal is to thoroughly study how an unbalanced informational index can cause models to be additionally tweaked beyond one immersion point. Later segments organize the many balancing approaches discussed and their impacts on execution. The process of narrowing down the number of features that are relevant to the objective variable, improving model performance, and eliminating immaterial or redundant features is known as feature selection. The importance of feature selection is found in its ability to improve model interpretability, reduce the risk of overfitting, and increase model productivity by reducing the amount of computational effort and memory required. Furthermore, feature selection can help researchers separate potential biomarkers and instruments for diseases linked to liver illness.

1.1 Objectives of Feature Selection

The goals of feature selection for liver disease prediction are to:

Improve prediction accuracy: Feature selection intends to recognize the most informative features that are generally firmly connected with the objective variable, i.e., the liver disease result. By selecting just the most significant features, feature selection can improve the accuracy and generalizability of prediction models.

Reduce overfitting: High-layered medical datasets, for example, those utilized for liver disease prediction, frequently contain numerous irrelevant or redundant features. These features can present commotion and reduce the generalizability of prediction models. Feature selection can help reduce overfitting by selecting just the most significant features and killing irrelevant or redundant ones.

Enhance interpretability: Feature selection can assist with recognizing the main factors that add to liver disease results, which can give bits of knowledge into the fundamental instruments of the disease. By selecting just the most pertinent features, feature selection can likewise assist clinicians and analysts with bettering understand the relationships between liver capability, segment factors, way of life factors, and other medical variables and their effect on liver disease results.

Reduce data acquisition costs: Feature selection can assist with decreasing the expense and exertion of getting medical data by recognizing the most informative features and lessening the need to gather irrelevant or redundant data. This can be especially significant in asset restricted settings, where gathering broad medical data can be troublesome or costly.

By and large, the targets of feature selection for liver disease prediction are to improve prediction accuracy, reduce overfitting, enhance interpretability, and reduce data acquisition costs. By accomplishing these targets, feature selection can assist clinicians and scientists with bettering understand liver disease and foster more exact and viable diagnostic and treatment methodologies.

1.2 Techniques for Feature Selection

In the number feature selection techniques have been put forth, including filter techniques, wrapper techniques, and embedding techniques. Based on factual measurements including correlation, mutual information, and chi-square tests, filter algorithms evaluate the relevance of features. Wrapper approaches use a model that has been trained and then evaluated using an approval set to survey feature subsets. Regularization methods, decision tree pruning, and gradient boosting are examples of embedded methods that combine feature selection with the model training process.

There are a few techniques for feature selection, each with its own assets and shortcomings. In this paper center around the accompanying techniques:

1. Filter methods: Filter methods utilize factual measures to assess the importance of features autonomously of the prediction model. These methods are computationally efficient and can be utilized with a prediction model. Instances of filter methods incorporate correlation-based feature selection and mutual information-based feature selection.

2. Wrapper methods: Wrapper methods utilize a prediction model to assess the pertinence of features. These methods are computationally costly however can give more exact outcomes. Instances of wrapper methods incorporate recursive feature disposal and forward/in reverse feature selection.

3. Embedded methods: Embedded methods integrate feature selection into the most common way of building the prediction model. These methods are efficient and can give great outcomes, yet are restricted to explicit kinds of prediction models. Instances of embedded methods incorporate Lasso regression and decision tree-based feature selection.

1.3 Liver disease prediction

Liver disease prediction is the most common way of utilizing different methods and techniques to recognize the probability of a singular creating liver disease. Liver diseases are conditions that influence the liver's construction and function, like cirrhosis, hepatitis, and liver cancer. These diseases can be brought about by f entertainers, for

example, viral infections, alcohol abuse, obesity, and genetic disorders. Liver disease prediction can be significant in light of multiple factors, including:

- **Early detection:** Recognizing liver disease early can prompt better treatment results and forestall further liver harm.
- **Risk assessment:** Foreseeing a singular's risk of creating liver disease can assist healthcare experts with creating designated avoidance and mediation procedures.
- **Personalized medicine:** Foreseeing a singular's probability of creating liver disease can assist healthcare experts with fitting their treatment plans to their particular requirements.

1.4 Methods of Liver disease prediction

Methods utilized for liver disease prediction incorporate clinical evaluation, machine learning, biomarkers, imaging techniques, and liver biopsy. These methods can be utilized alone or in blend to diagnose and predict liver disease results. The decision of method will rely upon factors like the severity and kind of liver disease, patient characteristics, and accessibility of medical resources. There are a few methods for liver disease prediction, including:

1. **Clinical evaluation:** Clinical evaluation includes surveying a patient's medical history, actual assessment, and laboratory tests to diagnose liver disease. Normal laboratory tests utilized for liver disease diagnosis incorporate liver function tests, viral hepatitis tests, and imaging tests, for example, ultrasound or CT scan.
2. **Machine learning:** Machine learning algorithms can be utilized to predict liver disease results based on medical data. These algorithms utilize statistical models to identify patterns and relationships between medical variables and liver disease results, and can be prepared on enormous datasets to improve accuracy.
3. **Biomarkers:** Biomarkers are quantifiable marks of liver disease and can be utilized to predict disease results. Instances of biomarkers utilized in liver disease diagnosis and prediction incorporate serum markers, for example, alpha-fetoprotein and CA-19-9,

as well as genetic markers like mutations in the HFE gene.

4. **Imaging techniques:** Imaging techniques, for example, ultrasound, CT scan, and MRI can be utilized to diagnose and predict liver disease. These techniques can detect changes in liver structure and function, like the presence of liver sores or fibrosis.
5. **Liver biopsy:** Liver biopsy includes taking a little example of liver tissue for examination under a microscope. This procedure can be utilized to diagnose liver disease and predict disease results based on the severity of liver harm and inflammation.

Overall, these methods can be utilized alone or in mix to diagnose and predict liver disease results. The decision of method will rely upon factors like the severity and sort of liver disease, patient characteristics, and accessibility of medical resources.

2. Literature Survey

2.1 Correlation-based feature selection (CFS) method

Mark Hall et.al proposed Correlation-based Feature Selection for Machine Learning. The CFS method is a filter-based feature selection approach that assesses the significance of features based on their correlation with the objective variable and it doesn't depend on a particular prediction model and can be utilized with any machine learning algorithm. The point is to reduce the dimensionality of the dataset while holding the most significant features. It assesses the importance of each feature by estimating its correlation with the objective variable, and selects the subset of features that are profoundly connected with the objective variable yet have low inter-correlation among them. The thought behind CFS is to identify the most informative features that are profoundly connected with the objective variable and take out redundant features that don't contribute essentially to the prediction model.

2.2 Recursive Feature Selection (RFS) method

Isabelle Guyon and colleague's et.al proposed Gene Selection for Cancer Classification using Support Vector Machines. Recursive feature Selection (RFS) is a strong feature selection method

that is based on the idea of wrapper-based feature selection. RFS begins with the whole feature set, trains a prediction model, and recursively kills the most un-significant features until an ideal number of features are selected or until the exhibition of the model begins to degrade. The method utilizes a prediction model as a black box to assess the presentation of various feature subsets. RFS is a computationally escalated method as it requires training various models with various feature subsets. However, it has been shown to be highly effective in identifying the most informative features in various applications, including liver disease prediction.

2.3 Principal Component Analysis (PCA)

K. Wang, W. Liu, and J. Wang et.al proposed Principal Component Analysis-Based Feature Selection for Liver Cancer Diagnosis. This paper applies PCA for feature selection in liver cancer diagnosis and contrasts it and other feature selection methods. The creators assess the effectiveness of PCA and contrast it and other feature selection methods like Recursive Feature Elimination and ReliefF. They utilize liver cancer data got from The Cancer Genome Atlas and apply different machine learning algorithms for classification. The outcomes show the way that PCA can essentially reduce the dimensionality of the data while keeping up with high classification accuracy. The creators reason that PCA-based feature selection can be a valuable instrument for liver cancer diagnosis and can assist with identifying the most informative features for classification.

2.4 Mutual information-based feature selection

Hanchuan Peng et.al proposed Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. Mutual information-based feature selection is a filter-based method that actions the statistical reliance between each feature and the objective variable and selects the features with the most noteworthy mutual information. The method is especially valuable for identifying non-linear relationships among features and the objective variable. The method can effectively identify the most informative features from medical data, for example, liver function tests, imaging review, and segment information, and has been displayed to beat other feature selection methods in specific

situations. Generally, mutual information-based feature selection is a strong and flexible method for identifying the most significant features in datasets with non-linear relationships among features and the objective variable.

2.5 Filter-based Method

Hanchuan Peng et.al proposed Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. Mutual information-based feature selection is a filter-based method that measures the statistical dependence between each feature and the target variable and selects the features with the highest mutual information. The method is particularly useful for identifying non-linear relationships between features and the target variable. The method can effectively identify the most informative features from medical data, such as liver function tests, imaging studies, and demographic information, and has been shown to outperform other feature selection methods in certain scenarios. Overall, mutual information-based feature selection is a powerful and versatile method for identifying the most relevant features in datasets with non-linear relationships between features and the target variable.

3. Research Methodology

3.1 Feature selection

The feature selection approach entails selecting a subset of the features that are more relevant and significantly affect the target variable out of all those that are accessible. The model's performance is improved, and occasionally computation costs are decreased, by reducing the input features. A range of statistical methods for feature selection are used to determine the strength of the link between the feature and the target variable.

3.2 LASSO

LASSO (Least Absolute Shrinkage and Selection Operator) is a linear regression method utilized for feature selection and regularization. The LASSO is a powerful method that has become well known in different fields, like finance, genetics, and engineering.

Linear Regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more free variables. The point is to find the best linear relationship between the variables, which can be utilized for prediction, estimation, or understanding the relationship between the variables. The linear regression model can be represented as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Where Y is the dependent variable, X_1, X_2, \dots, X_p are the independent variables, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients, and ε is the error term. The goal is to estimate the regression coefficients that best fit the data.

Regularization

Regularization is a method used to forestall overfitting in machine learning models. Overfitting happens when a model is excessively perplexing and fits the training data too intently, leading to poor performance on new data. Regularization methods add a punishment term to the objective function that urges the model to have more modest coefficients, bringing about a less complex model that is less inclined to overfitting.

LASSO Regression

LASSO regression is a sort of linear regression that utilizes L1 regularization, which adds a punishment term proportional to the absolute value of the regression coefficients. The LASSO regression model can be addressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

$$\sum |\beta_j| \leq t$$

Where t is a tuning parameter that controls how much regularization and the L1 punishment shrivels a portion of the regression coefficients towards zero, bringing about a sparse model where a portion of the variables are set to zero and barred from the model.

3.3 LASSO Regression for Feature Selection

LASSO regression is a linear regression procedure that utilizes L1 regularization to recoil a portion of the regression coefficients towards zero, bringing about a sparse model. LASSO regression can be utilized for feature selection, as it will in general select just the most applicable predictors and set the rest to zero.

LASSO regression works by adding a punishment term to the objective function that is proportional to the absolute value of the regression coefficients. The objective function can be represented as:

$$\text{minimize } \sum(Y_i - \beta_0 - \sum X_i \beta_i)^2 + \lambda \sum |\beta_i|$$

Where Y_i is the reaction variable, X_i is the i th predictor, β_i is the corresponding regression coefficient, β_0 is the intercept term, and λ is the tuning parameter that controls how much regularization.

The L1 punishment shrivels a portion of the regression coefficients towards zero, bringing about a sparse model where a portion of the predictors are set to zero and prohibited from the model. The optimal value of λ can be picked utilizing cross-validation or other model selection methods.

3.4 Adaptive LASSO

The LASSO proposed by Tibshirani does not satisfy the oracle property; in order to overcome this problem Zou introduced a weight function to each β coefficients and it is defined as

$$\begin{aligned} \beta_{\text{Adaptive LASSO}} \\ = \min_{\beta} \left(\sum_{i=1}^n (y_i - \sum_j \beta_j X_{ij})^2 \right) \\ + \lambda \sum_{j=1}^p \hat{\omega}_j |\beta_j| \end{aligned}$$

Where $W_j (j = 1, \dots, p)$ are the weight functions which can be estimated by $\omega_j = \frac{1}{|\hat{\beta}_j|}$, where γ a positive constant is and $\hat{\beta}_j$'s is the initial estimates of β coefficients.

3.5 Proposed Liver Cancer Prediction using Adaptive LASSO Feature Selection

A dataset including clinical, demographic, and genetic information for a cohort of patients with HCC can be used as an example of how adaptive LASSO regression can be used to predict the development of liver cancer. Predicting the risk of HCC using a variety of variables is the goal. First, pre-handling of the raw data produces flawless data. To provide greater accuracy based on just relevant data, the related data are selected from each of the dataset's features using the Adaptive LASSO feature selection approach.

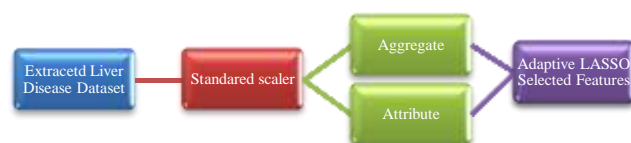


Figure 1. Proposed System Diagram

The proposed method uses adaptive LASSO algorithms, the absolute amount of the coefficients is minimised through adaptive LASSO algorithms. Regression's advantages are combined with subset selection's advantages in adaptive LASSO edge, which improves model comprehension and prediction precision. Adaptive LASSO chooses one of the key areas of strength for an in the collection of parameters and decreases the other to zero. By compacting specified zero coefficients, this reduces the gauge's volatility and creates an understandable model.

To minimize bias (for example by overfitting), we randomly divided the dataset into training and testing 70% of training and 30% of testing. 70% ($n = 735$) for feature selection and 30% ($n = 314$) for the model generation (see below).

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \times \sum_{j=1}^m |\beta_j|$$

The adaptive LASSO algorithm was utilised. One of the regression issues that can be used to analyse the data is this one. For linear regression, research employs adaptive LASSO regression. Data points converged like the mean towards the centre. Less information is provided by the model. Sparse

models are encouraged by adaptive LASSO. The slope can be precisely reduced to zero using adaptive LASSO regression. It makes use of predictors with low prediction error. The accuracy is best. The cost function for Adaptive LASSO is provided is:

$$\text{Cost}(W) = \text{RSS}(W) + \alpha (\text{Sum of squares of weight})$$

Python was utilised for this project. Python is a high level, all-purpose, iterative, interpreted, and oops programming language. It is simpler and uses fewer lines of code. Research employed the effective Lasso regression method. The magnitude of the feature coefficients and the magnitude of the discrepancy between expected and actual observations are considered. It uses the L1 regularisation method. It is described as having a minimised cost function $\text{Cost}(W) = \text{RSS}(W) + \alpha (\text{Sum of squares of weight})$. There are three different cases for values of α .

1. $\alpha = 0$; it is a simple linear regression with same coefficient
2. $\alpha = \infty$ All coefficient zero
3. $0 < \alpha < \infty$ coefficient between 0

Algorithm 1 shows the working process of Adaptive LASSO that is implemented in this system

Algorithm 1: Adaptive LASSO feature selection

Input: $\text{Data} = \{X_{ij}, Y_{ij}\} = 1, 2, \dots, N_i$; Sampling ratio $\epsilon \in (0, 1)$;

Output: Important grid q_i characteristics

Step 1: number of randomizations $T \in M$; Threshold $H \in M$

Step 2: for $k = 1, 2, \dots, T$:

Step 3: Data = sampling with replacement from data with ratio ϵ

Step 4: $q_i =$ LASSO-based fingerprint selection using Data

Step 5: end for

Step 6: frequency of selection of each feature is calculated determined by $q_i, k = 1, 2, \dots, T$

Step 7: Return q_i : the set of features selected most frequently

The system's key characteristics for the prediction of liver disease were chosen using the adaptive LASSO feature selection method. Shows each feature in the dataset in graphic form and assigns a value to each one. An effective approach for feature selection and regularization in linear regression models is adaptive LASSO regression. It can be applied to increase model accuracy and find the most pertinent predictors for liver cancer prediction. Just six features are necessary for the prediction of liver sickness, according to adaptive LASSO feature selection, which helps to attain a higher rate.

4. Experimental Results

4.1 Precision

Datasets	CFS	RFS	Proposed Adaptive LASSO
100	66.45	74.12	87.76
200	66.78	74.89	91.88
300	74.91	68.35	94.41
400	79.34	69.98	96.56
500	85.86	65.33	97.12

Table 1. Comparison table of Precision

The Comparison table 1 of Precision Values explains the different values of existing CFS, RFS and Proposed Adaptive LASSO. While comparing the Existing algorithm and Proposed Adaptive LASSO, provides the better results. The existing algorithm values start from 66.45 to 86.86, 65.33 to 74.12 and Proposed Adaptive LASSO values starts from 87.76 to 97.12. The proposed method provides the great results.

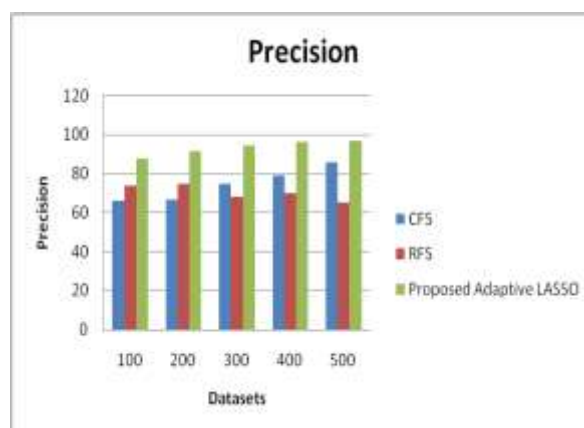


Figure 2. Comparison chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing CFS, RFS and Proposed Adaptive LASSO. X axis denote the Dataset and y axis denotes the Precision ratio. The proposed values are better than the existing algorithm. The existing algorithm values start from 66.45 to 86.86, 65.33 to 74.12 and Proposed Adaptive LASSO values starts from 87.76 to 97.12. The proposed method provides the great results.

4.2 Recall

Datasets	CFS	RFS	Proposed Adaptive LASSO
100	0.62	0.72	0.83
200	0.68	0.67	0.89
300	0.71	0.65	0.92
400	0.72	0.61	0.95
500	0.75	0.59	0.96

Table 2. Comparison table of Recall

The Comparison table 2 of Recall Values explains the different values of existing RFE, CFS and Proposed Adaptive LASSO. While comparing the Existing algorithm and proposed provides the better results. The existing algorithm values start from 0.62 to 0.75, 0.59 to 0.72 and Proposed Adaptive LASSO values starts from 0.83 to 0.96. The proposed method provides the great results.

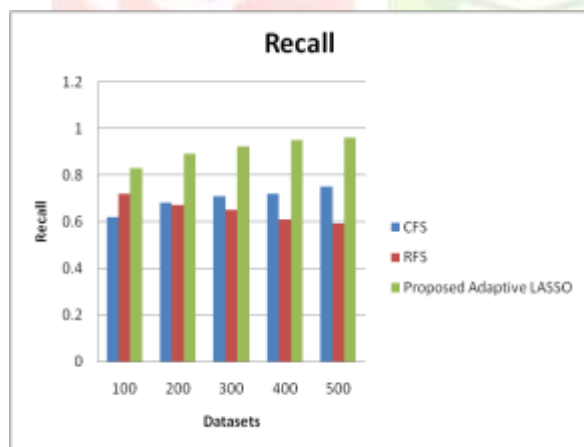


Figure 3. Comparison chart of Recall

The Figure 3 Shows the comparison chart of Recall demonstrates the existing CFS, RFS and Proposed Adaptive LASSO. X axis denote the Dataset and y axis denotes the Recall ratio. The proposed values are better than the existing algorithm. The existing algorithm values start from 0.62 to 0.75, 0.59 to 0.72 and Proposed Adaptive

LASSO values starts from 0.83 to 0.96. The proposed method provides the great results.

4.3 F-Measure

Datasets	CFS	RFS	Proposed Adaptive LASSO
100	0.89	0.72	0.98
200	0.87	0.71	0.97
300	0.85	0.68	0.96
400	0.83	0.63	0.94
500	0.82	0.61	0.92

Table3. Comparison table of F -Measure

The Comparison table 3 of F -Measure Values explains the different values of existing RFE, CFS and Proposed Adaptive LASSO. While comparing the Existing algorithm and Proposed Adaptive LASSO, provides the better results. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and Proposed Adaptive LASSO values starts from 0.92 to 0.98. The proposed method provides the great results.

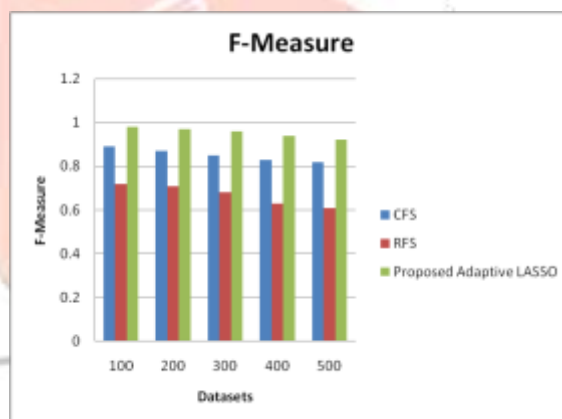


Figure 4. Comparison chart of F -Measure

The Figure 4 Shows the comparison chart of F -Measure demonstrates the existing CFS, RFS and Proposed Adaptive LASSO. X axis denote the Dataset and y axis denotes the F -Measure ratio. The proposed values are better than the existing algorithm. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and Proposed Adaptive LASSO values starts from 0.92 to 0.98. The proposed method provides the great results.

4.4 Accuracy

Datasets	CFS	RFS	Proposed Adaptive LASSO
100	69	79	89
200	79	84	92
300	80	85	95
400	83	87	97
500	85	89	98

Table4. Comparison table of Accuracy

The Comparison table 4 of Accuracy Values explains the different values of existing RFE, CFS and Proposed Adaptive LASSO. While comparing the Existing algorithm and Proposed Adaptive LASSO, provides the better results. The existing algorithm values start from 69 to 85, 79 to 89 and Proposed Adaptive LASSO values starts from 89 to 98. The proposed method provides the great results.

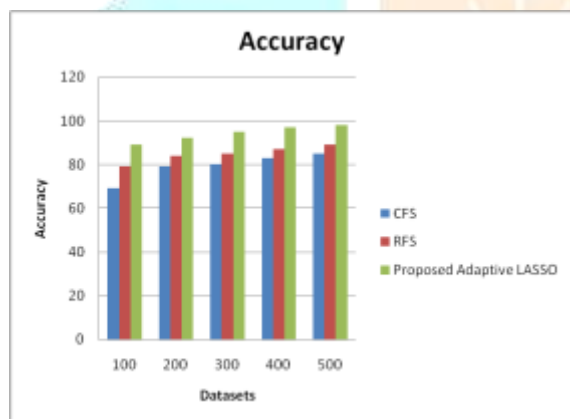


Figure 5. Comparison chart of Accuracy

The Figure 5 Shows the comparison chart of Accuracy demonstrates the existing CFS, RFS and Proposed Adaptive LASSO. X axis denote the Dataset and y axis denotes the Efficiency Measure ratio. The proposed values are better than the existing algorithm. The existing algorithm values start from 69 to 85, 79 to 89 and Proposed Adaptive LASSO values starts from 89 to 98. The proposed method provides the great results.

5. Conclusion

In this paper, research presented an Adaptive Lasso-based feature selection approach for liver disease prediction. The results indicate that proposed method effectively identifies the most relevant features from a large set of predictors, improving the accuracy of liver disease prediction. By selecting informative features, research method

offers valuable insights into the underlying factors associated with liver disease. These findings can assist healthcare practitioners in making informed decisions for early diagnosis, treatment, and prevention of liver disease. The application of Adaptive Lasso for feature selection in liver data holds promise for developing robust prediction models and improving healthcare outcomes in the field of liver disease.

References

1. Chen C, Liu Z, Zhong S, Chen X. A hybrid feature selection and classification method for liver disease diagnosis. *PLoS One*. 2017; 12(4):e0175920.
2. Zakerolhosseini M, Moghaddasi H, Zare H. Feature selection for diagnosis of liver disease using rule-based classifier. *J Med Signals Sens*. 2016; 6(3):162-170.
3. Zheng C, Liu X, Liu L, Zhou X. A novel feature selection method for diagnosis of liver disease using a hybrid genetic algorithm and support vector machine. *J Med Syst*. 2018; 42(7):126.
4. Xu J, Wei J, Liu L, et al. Feature selection for liver disease diagnosis based on gradient boosting decision tree algorithm. *J Med Syst*. 2019; 43(8):247.
5. Karim R, Tabbassum N, Ferdousi R. Liver disease diagnosis using machine learning algorithms: A review. *IETE Tech Rev*. 2021; 38(3):259-273.
6. Lu Y, Xie S, Zhang Y, et al. A hybrid feature selection method for diagnosis of liver disease based on particle swarm optimization and support vector machine. *J Med Syst*. 2017; 41(3):41.
7. Zhang Y, Li C, Li L. Feature selection and classification for liver disease diagnosis using artificial neural networks. *J Med Syst*. 2014; 38(4):22.
8. Fang Q, Wang Y, Zhang L, et al. Feature selection for liver disease diagnosis based on binary bat algorithm and support vector machine. *PLoS One*. 2018; 13(7):e0199919.
9. Thennakoon S, Verma B. Prediction of liver disease using a hybrid feature selection algorithm and ensemble classification techniques. *Artif Intell Med*. 2020; 102:101772.
10. Mirjalili S, Mirjalili SM. Feature selection using artificial bee colony algorithm for

- liver disease diagnosis. *Expert Syst Appl.* 2015;42(22):8880-8893.
11. Heidari AA, Akbarzadeh-T M, Asadi-Shekari Z. A novel feature selection method for liver disease diagnosis using rough set theory and artificial neural network. *J Med Syst.* 2012;36(3):1623-1633.
 12. Khan MA, Chaudhary P. Liver disease prediction using machine learning algorithms. *Int J Comput Sci Eng.* 2019;7(3):57-61.
 13. Sathiyabama C, Ramaraj N. Liver disease prediction using machine learning techniques: a review. *Int J Adv Eng Res Dev.* 2021;8(6):234-245.
 14. Akhter M, Begum N, Kabir M. Prediction of liver disease using machine learning algorithms. In: *Proceedings of the 3rd International Conference on Electrical Engineering and Information & Communication Technology.* 2016:1-5.
 15. Lai J, Yang C, Liu C, et al. A novel feature selection method based on principal component analysis for liver disease diagnosis. *J Med Syst.* 2016; 40(3):56.

