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

PREDICTION OF LIVER DISEASE USING MACHINE LEARNING

[1]Archana.K.K, [2]Suseendran.S.S, [3]Roganprasath.M, [4]WilsonImmanuel.E,
[1]Assistant Professor, [2,3,4]Student,,
[1,2,3,4]Department of Computer Science and Engineering,
[1,2,3,4]Sri Ramakrishna Institute of Technology,Coimbatore,India.

Abstract : This study explores various machine learning models to classify utilizing the Indian Liver Patient Dataset (ILPD) to study liver disease. Our analysis encompassed data preprocessing techniques such as handling missing values, duplicates, and imbalance using SMOTE. We employed feature selection methods including Pearson correlation, Gain Ratio, and Random Forest feature importance to identify relevant features. Subsequently, we developed and evaluated various models of classification, such as Random Forest, Support Vector Machines (SVM), and Logistic Regression, both with all features and with selected features. Moreover, we conducted hyperparameter tuning for Logistic Regression and Random Forest models to enhance their performance. Additionally, an alternative preprocessing approach using iterative imputer and upsampling was investigated, demonstrating its impact on model performance. Our results accentuate the efficiency of machine learning models in precisely categorizing liver disease, with the ensemble model of Support Vector Machine and Random Forest Classifier achieving a notable accuracy of 91.01% after tuning. This study offers insightful information about the use of machine learning methods for liver disease diagnosis, offering potential avenues for improving healthcare decision-making and patient outcomes.

***IndexTerms* - healthcare,liver disease,prediction,machine learning,data analysis.**

I. INTRODUCTION

Liver disease poses a significant global health challenge, necessitating accurate and early diagnosis for improved medical intervention and patient care. Leveraging machine learning and advanced analytics offers promising avenues for enhancing predictive capabilities in this domain.This project focuses on developing and evaluating predictive models for liver disease using comprehensive patient record datasets. Employing cutting-edge machine learning techniques and rigorous data preprocessing, our methodology aims to enhance prediction reliability and precision.Beginning with meticulous data preprocessing, including addressing missing values and class imbalance using techniques like SMOTE, and scikit-learn's iterative imputer ensures equitable representation of liver disease cases. We evaluate three machine learning techniques, such as Random Forest, Support Vector Machine (SVM), and Logistic Regression, optimizing hyperparameters and feature selection for each model to maximize predictive accuracy.Our results highlight the effectiveness of ensembling the hyperparameter-tuned Support Vector Machine Classifier and hyperparameter-tuned Random Forest Classifier, achieving an impressive accuracy of 91.01%, particularly when using selected features.This research not only contributes to liver disease diagnosis but also advances the broader field of healthcare analytics. By offering practical insights and methodologies, our project supports data-driven decision-making in liver disease diagnosis and management, with the ultimate goal of improving patient outcomes and care.

II. SCOPE OF THE PROJECT

The scope of the "Liver Disease Prediction" project extends beyond predictive modeling to encompass a broader aim of enhancing accessibility and inclusivity in healthcare analytics. By leveraging advanced machine learning techniques, the project seeks to develop predictive models that aid in the early detection and management of liver diseases.Similar to the emphasis on inclusivity in the model project, our scope underscores the importance of democratizing access to healthcare resources. Through the creation of accurate predictive models, this project aims to ensure that individuals, regardless of their healthcare access or geographical location, can benefit from timely diagnosis and intervention for liver diseases.Moreover, the project aligns with current trends in healthcare technology, contributing to the creation of a more accessible digital healthcare environment. By employing rigorous data preprocessing techniques and comparative analyses of machine learning algorithms, we strive to maximize the reliability and precision of our predictive models, mirroring the meticulous approach outlined in the model project.Furthermore, the project's scope extends to the presentation and discussion of results, emphasizing the practical implications of our findings on healthcare practices and patient outcomes. By promoting the integration of predictive analytics into clinical decision-making processes, this project aims to facilitate proactive healthcare management and improve public health outcomes related to liver diseases.Overall, the "Liver Disease Prediction" project shares a common goal with the model project in advancing inclusivity and accessibility, albeit within the realm of healthcare analytics, through the development and application of predictive models for liver disease diagnosis and management.

III. PROPOSED SYSTEM

In the proposed system, data preprocessing has a vital role in ensuring the integrity and reliability of the dataset for liver disease classification. Missing values were addressed by the removal of records containing such values, a strategy deemed appropriate due to the low prevalence of missing data. This approach safeguards the quality of the dataset while preserving valuable information. Additionally, categorical feature encoding was performed using one-hot encoding on the 'Gender' feature, facilitating the transformation of categorical variables into numerical representations suitable for machine learning algorithms.

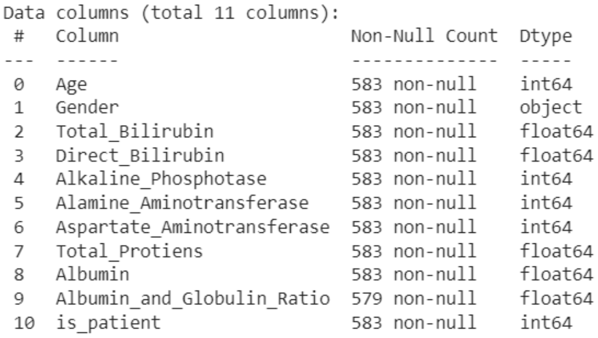
To tackle class imbalance inherent in the dataset, two distinct approaches were adopted. Initially, Synthetic Minority Over-sampling Technique (SMOTE) was applied to ensure a balanced representation of both liver disease and non-liver disease cases. This technique generates synthetic instances of the minority class, thereby mitigating the effects of class imbalance and improving model performance. Subsequently, an alternative strategy involving iterative imputer and upsampling was explored. This innovative approach aimed to enhance the representation of minority class instances by iteratively imputing missing values and upsampling the minority class to match the number of samples in the majority class. This iterative imputation process, coupled with upsampling, resulted in a more balanced dataset, consequently contributing to improved model performance and reliability in predicting liver disease.

When creating machine learning models for liver disease classification, various algorithms were evaluated, including Logistic Regression, Support Vector Machines (SVM), and Random Forest. Hyperparameter tuning using GridSearchCV was employed to optimize the performance of Logistic Regression and Random Forest models, thereby enhancing their predictive capabilities. Furthermore, feature selection techniques such as Pearson correlation, Gain Ratio, and Random Forest feature importance were utilized to identify informative predictors contributing to liver disease classification. These methods provided insights into feature relationships and the relevance of predictors in the classification task.

Model evaluation and validation were conducted rigorously to guarantee the resilience and generalization of the developed models. K-fold cross-validation was employed to assess model performance across multiple iterations, thereby validating their effectiveness in predicting liver disease. Metrics for performance such as F1-score, recall, accuracy, and precision were computed, providing comprehensive insights into the models' predictive capabilities.

The ensemble approach was incorporated during the model evaluation phase to further enhance predictive performance. Given that both the Support Vector Machine (SVM) and Random Forest classifiers achieved the same highest accuracy of 86.82%, an ensemble model was constructed by combining the predictions of these two classifiers. This ensemble approach leverages the strengths of both algorithms, potentially improving the robustness and accuracy of the classification. By aggregating the predictions of multiple models, the ensemble method gives a more comprehensive and reliable prediction of liver disease, contributing to the overall effectiveness of the proposed system.

3.1 Dataset Description



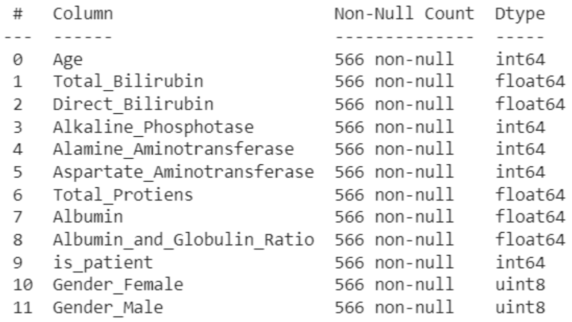
```
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    object
 2   Total_Bilirubin             583 non-null    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  is_patient                  583 non-null    int64
```

Figure 1: Description of the ILPD Dataset

The selected Indian Liver Patient Dataset (ILPD) have a total of eleven columns,with ten features and one target variable "is_patient".

3.2 Data Preprocessing

We can see that in the dataset, the Albumin_and_Globulin column has missing values. It is handled by removing the entire rows which have null or missing values.



```
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         566 non-null    int64
 1   Total_Bilirubin             566 non-null    float64
 2   Direct_Bilirubin            566 non-null    float64
 3   Alkaline_Phosphotase        566 non-null    int64
 4   Alamine_Aminotransferase    566 non-null    int64
 5   Aspartate_Aminotransferase  566 non-null    int64
 6   Total_Protiens              566 non-null    float64
 7   Albumin                     566 non-null    float64
 8   Albumin_and_Globulin_Ratio  566 non-null    float64
 9   is_patient                  566 non-null    int64
 10  Gender_Female               566 non-null    uint8
 11  Gender_Male                 566 non-null    uint8
```

Figure 2: Peek at the dataset after data cleaning

The above figure represents the dataset after removing the records which have missing values and performing One Hot-Encoding for the Gender Column since it has strings as values.

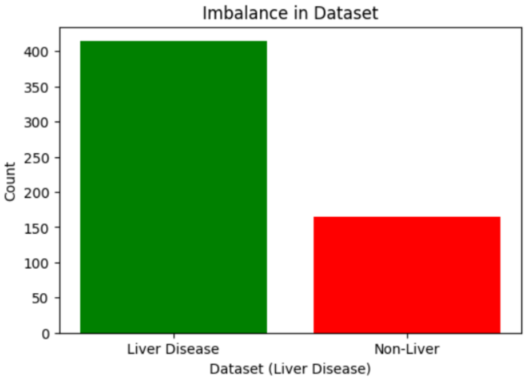


Figure 3: Imbalance in the dataset

The above figure showcases the distribution of number of records for each class representing whether or not the individual has liver disease.It represents 416 Liver Disease instances and 167 Non Liver Disease instances.

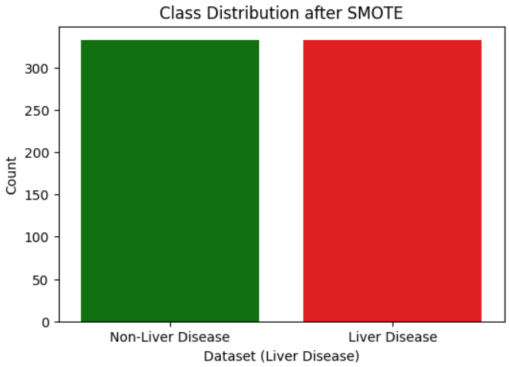


Figure 4: Balanced dataset after applying SMOTE

The above figure showcases the distribution of number of records for each class representing whether or not the individual has liver disease after using SMOTE. It represents 332 Liver Disease instances and 332 Non Liver Disease instances.

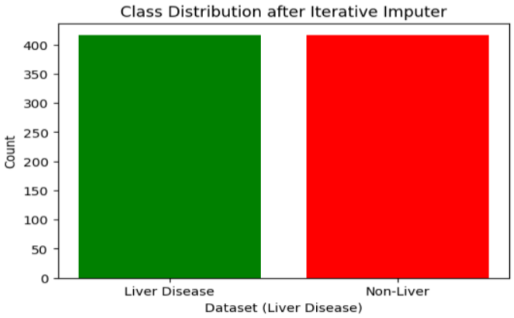
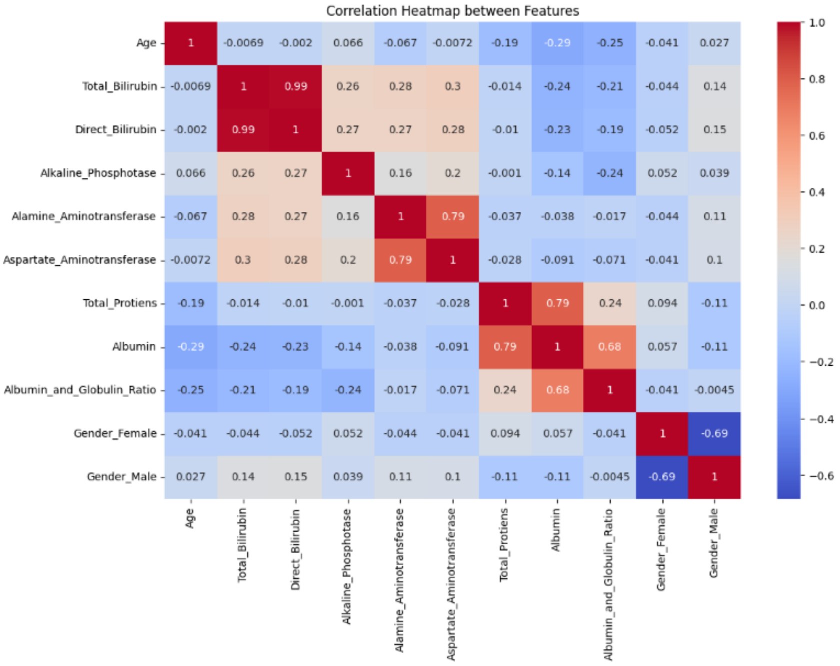


Figure 5: Balanced dataset after using iterative imputer

The above figure showcases the distribution of number of records for each class representing whether or not the individual has liver disease after using iterative imputer. It represents 416 Liver Disease instances and 416 Non Liver Disease instances.

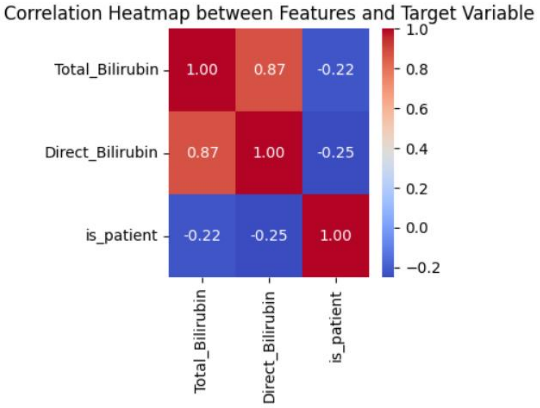
3.3 Pearson Correlation



	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Gender_Female	Gender_Male
Age	1	-0.0069	-0.002	0.066	-0.067	-0.0072	-0.19	-0.29	-0.25	-0.041	0.027
Total_Bilirubin	-0.0069	1	0.99	0.26	0.28	0.3	-0.014	-0.24	-0.21	-0.044	0.14
Direct_Bilirubin	-0.002	0.99	1	0.27	0.27	0.28	-0.01	-0.23	-0.19	-0.052	0.15
Alkaline_Phosphotase	0.066	0.26	0.27	1	0.16	0.2	-0.001	-0.14	-0.24	0.052	0.039
Alamine_Aminotransferase	-0.067	0.28	0.27	0.16	1	0.79	-0.037	-0.038	-0.017	-0.044	0.11
Aspartate_Aminotransferase	-0.0072	0.3	0.28	0.2	0.79	1	-0.028	-0.091	-0.071	-0.041	0.1
Total_Protiens	-0.19	-0.014	-0.01	-0.001	-0.037	-0.028	1	0.79	0.24	0.094	-0.11
Albumin	-0.29	-0.24	-0.23	-0.14	-0.038	-0.091	0.79	1	0.68	0.057	-0.11
Albumin_and_Globulin_Ratio	-0.25	-0.21	-0.19	-0.24	-0.017	-0.071	0.24	0.68	1	-0.041	-0.0045
Gender_Female	-0.041	-0.044	-0.052	0.052	-0.044	-0.041	0.094	0.057	-0.041	1	-0.69
Gender_Male	0.027	0.14	0.15	0.039	0.11	0.1	-0.11	-0.11	-0.0045	-0.69	1

Figure 6: Correlation Matrix of the dataset

The above figure represents Correlation Matrix depicting the relationships between the features in the dataset. From the above Correlation matrix, it is observed that the features Direct_Bilirubin and Total_Bilirubin are highly correlated above 80 percentage.

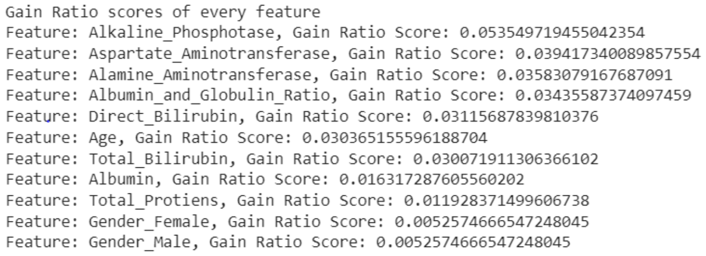


	Total_Bilirubin	Direct_Bilirubin	is_patient
Total_Bilirubin	1.00	0.87	-0.22
Direct_Bilirubin	0.87	1.00	-0.25
is_patient	-0.22	-0.25	1.00

Figure 7: correlation matrix to identify individual correlation

Based on the correlation matrix above, it is observed that the “Direct_Bilirubin” feature exhibits a stronger correlation with the target variable 'is_patient' than the “Total_Bilirubin”feature.

3.4 Gain Ratio Method

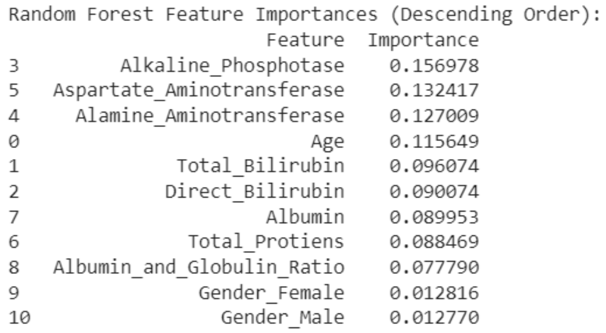


```
Gain Ratio scores of every feature
Feature: Alkaline_Phosphotase, Gain Ratio Score: 0.053549719455042354
Feature: Aspartate_Aminotransferase, Gain Ratio Score: 0.039417340089857554
Feature: Alamine_Aminotransferase, Gain Ratio Score: 0.03583079167687091
Feature: Albumin_and_Globulin_Ratio, Gain Ratio Score: 0.03435587374097459
Feature: Direct_Bilirubin, Gain Ratio Score: 0.03115687839810376
Feature: Age, Gain Ratio Score: 0.030365155596188704
Feature: Total_Bilirubin, Gain Ratio Score: 0.030071911306366102
Feature: Albumin, Gain Ratio Score: 0.016317287605560202
Feature: Total_Protiens, Gain Ratio Score: 0.011928371499606738
Feature: Gender_Female, Gain Ratio Score: 0.0052574666547248045
Feature: Gender_Male, Gain Ratio Score: 0.0052574666547248045
```

Figure 8: Gain Ratio scores of each feature

The above figure represents the Gain Ratio scores of each feature in the ILPD dataset,from which it is observed that Direct_Bilirubin has more Gain Ratio score than the Total_Bilirubin feature.

3.5 Random Forest Method



```
Random Forest Feature Importances (Descending Order):
                         Feature  Importance
3           Alkaline_Phosphotase    0.156978
5    Aspartate_Aminotransferase    0.132417
4      Alamine_Aminotransferase    0.127009
0                            Age    0.115649
1                Total_Bilirubin    0.096074
2               Direct_Bilirubin    0.090074
7                        Albumin    0.089953
6                 Total_Protiens    0.088469
8     Albumin_and_Globulin_Ratio    0.077790
9                  Gender_Female    0.012816
10                   Gender_Male    0.012770
```

Figure 9: Random Forest feature importance scores of each feature in the dataset

The above figure represents the Random Fores Feature Importances from which it is observed that both the Direct_Bilirubin and Total_Bilirubin equal scores.Hence from the observations from Pearson Correlation, Gain Ratio and Random Forest,the feature Total_Bilirubin dropped and Direct_Bilirubin is selected.

IV. MODEL DEVELOPMENT

4.1 Logistic Regression

Logistic Regression is a fundamental and commonly employed machine learning and statistical technique that is particularly suited for binary classification problems. The fundamental concept underlying logistic regression is the logistic function (or sigmoid function), which maps any real number to a value between 0 and 1. This function is pivotal in converting the output of a probability created by combining predictor variables in a linear fashion, which can then be interpreted as the probability of an event occurring.

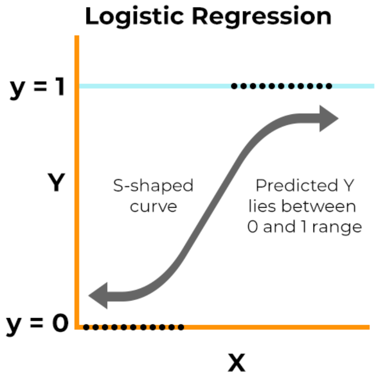


Figure 10: Logistic regression curve

For a single d-dimension input sample,

$$y(x) = w_0 + w_1x_1 + \cdots + w_nx_n \tag{1}$$

Where:

- y(x) represents the log-odds (logit) of the probability of the binary outcome.
- $x_1, x_2, \ldots x_d$ are the independent variables or features.
- $w_0, w_1, \ldots w_d$ are the corresponding coefficients.

For n-input d-dimension samples in matrix form,

$$\hat{y} = Xw \tag{2}$$

Where:

- $\hat{y}$ represents the estimated probabilities of a binary outcome.
- X represents the matrix of independent variables.
- w is the vector of coefficients used in the linear combination of X.

By applying sigmoid transform on $\hat{y}$,

$$p = \frac{1}{1+e^{-\hat{y}}} \tag{3}$$

Where:

- p represents the probability of a binary outcome.
- $\hat{y}$ represents the coefficients and independent variables linear combination.
- e is the base of the natural logarithm.

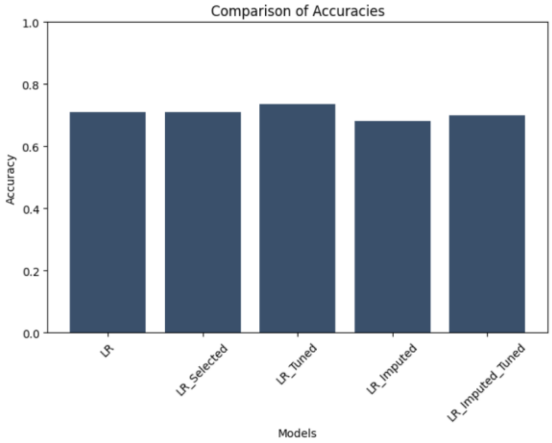


Figure 11: Comparison of Accuracies of logistic regression models

The above figure illustrates the comparison of five logistic regression models: LR, LR_Selected, LR_Tuned, LR_Imputed, and LR_Imputed_Tuned. These models were trained with different sets of features and dataset preprocessing methods, including SMOTE and iterative imputer.

The hyperparameter-tuned logistic regression model performs well with the SMOTE-preprocessed dataset, achieving an accuracy of 73.68%. This is followed by logistic regression models using all features and selected features, which achieved accuracies of 71.05%. Additionally, the iterative imputed logistic regression model achieved an accuracy of 68.26%, while its hyperparameter-tuned version yielded an accuracy of 70.05%.

4.2 Support Vector Machine(SVM)

Support Vector Machine (SVM) is a method that is frequently employed for binary classification tasks, focusing on maximizing the margin between classes. It employs the kernel trick to handle non-linear relationships, enhancing its flexibility. In SVM, the hyperplane with the biggest margin is chosen to divide the points in the input variable space according to their class. The term “support vectors” refers to the closest data points that define the margin.

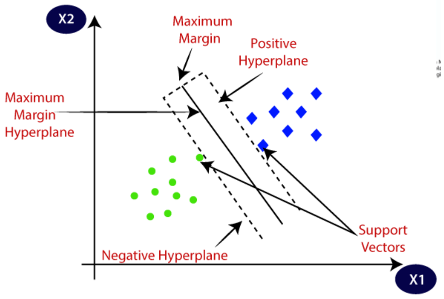


Figure 12: SVM Hyperplane

For a single d-dimensional input sample, the decision function can be represented as

$$f(x) = w_0 + w_1 x_1 + \cdots + w_d x_d \tag{4}$$

Where:

- f(x) represents the decision function.
- $x_1, x_2, \ldots x_d$ are the independent variables or features.
- $w_0, w_1, \ldots w_d$ are the corresponding coefficients.

For n-input d-dimension samples in matrix form,

$$f(X) = Xw \tag{5}$$

Where:

- f(X) represents the decision function for the entire dataset.
- X represents the matrix of independent variables, with each row corresponding to a sample and each column corresponding to a feature.
- w is the vector of coefficients used in the linear combination of X.

After applying a sigmoid transformation to f(X), we obtain the probability p of a binary outcome,

$$p = \frac{1}{1+e^{-f(x)}} \tag{6}$$

Where:

- p represents the probability of a binary outcome.
- e is the base of the natural logarithm.

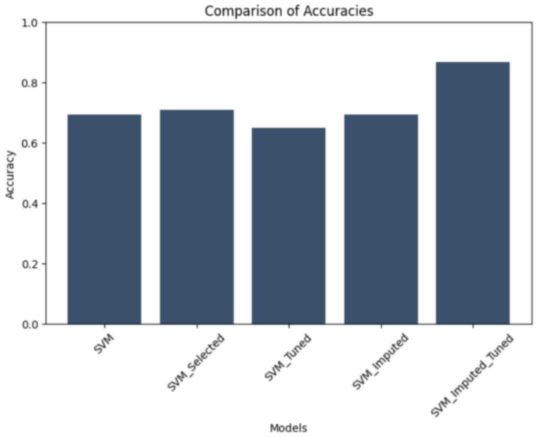


Figure 13: Comparison of Accuracies of SVM models

The above figure illustrates the comparison of five svm models: SVM, SVM_Selected, SVM_Tuned, SVM_Imputed, and SVM_Imputed_Tuned. These models were trained with different sets of features and dataset preprocessing methods, including SMOTE and iterative imputer.

The hyperparameter-tuned svm model performs well with the dataset preprocessed using the iterative imputer, achieving an accuracy of 86.82%. This is followed by svm models trained using all features, selected features, and hyperparameter tuning with the SMOTE-preprocessed dataset, which achieved accuracies of 69.29%, 71.05%, and 64.91%, respectively. Additionally, the iterative imputed random forest model and its hyperparameter-tuned version yield accuracies of 69.46% and 86.82%, respectively.

4.3 Random Forest

Random Forest Classifier is a robust and versatile ensemble learning technique suitable for tasks involving regression and classification. It works especially well in handling complex data and is known for its high predictive accuracy. Decision trees in the Random Forest ensemble were trained using various subsets of the data., and it combines their outputs for the purpose of prediction. The core concept behind the Random Forest Classifier is the combination of multiple decision trees to generate a more precise and stable prediction model.

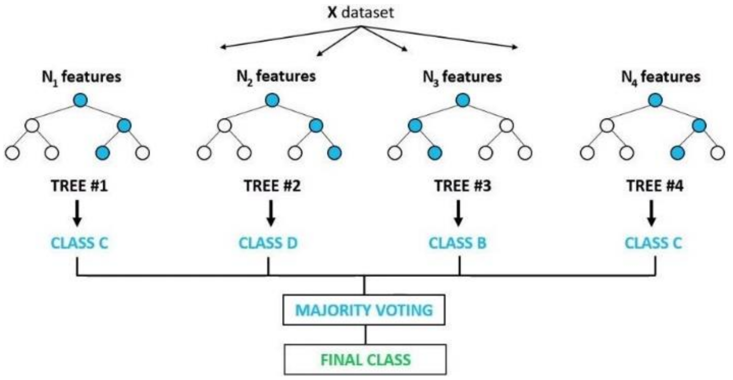


Figure 14: Random Forest decision trees

For a random forest classifier, the decision function for a single sample can be represented as a fusion of several decision trees. Each decision tree provides a classification decision, and the final prediction is often determined by a voting mechanism or averaging of the predictions from individual trees.

Let's denote the decision function for a single sample as

$$f(x) = \frac{1}{N}\, \Sigma_{i=1}^{N} T_i(x) \tag{7}$$

Where:

- f(x) represents the decision function.
- $T_i(x)$ represents the decision made by the i^{th} decision tree.
- N is the total number of decision trees in the random forest.

For a random forest classifier, instead of explicit coefficients $w_0, w_1,\ldots w_d$ we have decision trees $T_1, T_2,\ldots, T_n$ contributing to the final prediction.The probability of a binary outcome can then be estimated as a proportion of trees in the forest that predict the positive class in a classification scenario:

$$f(x) = \frac{1}{N} \Sigma_{i=1}^{N} I(T_i(x) = Positive) \tag{8}$$

Where:

- p represents the probability of a binary outcome.
- I(·) is the indicator function that returns 1 in the case that its argument is true and 0 otherwise.

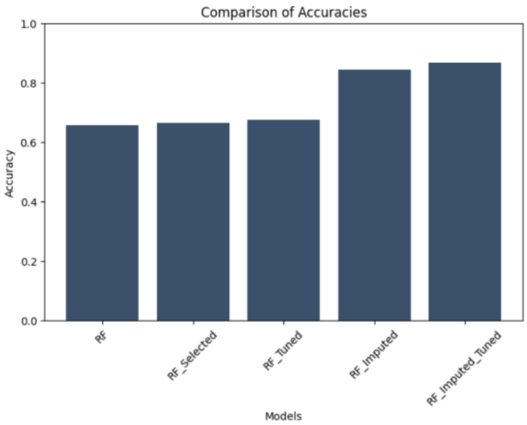


Figure 15: Comparison of Accuracies of Random Forest models

The above figure illustrates the comparison of five random forest models: RF, RF_Selected, RF_Tuned, RF_Imputed, and RF_Imputed_Tuned. These models were trained with different sets of features and dataset preprocessing methods, including SMOTE and iterative imputer.

The hyperparameter-tuned random forest model performs well with the dataset preprocessed using the iterative imputer, achieving an accuracy of 86.82%. This is followed by random forest models trained using all features, selected features, and hyperparameter tuning with the SMOTE-preprocessed dataset, which achieved accuracies of 65.78%, 66.66%, and 67.54%, respectively. Additionally, the iterative imputed random forest model and its hyperparameter-tuned version yield accuracies of 84.43% and 86.82%, respectively.

V. RESULT AND DISCUSSION

A total of 16 models were developed and analyzed for better performance. These models were trained on different sets of features and dataset preprocessing methods, including SMOTE and iterative imputer. Out of these, two models achieved a high accuracy of 86.82%.Since both the hyperparameter-tuned SVM and hyperparameter-tuned Random Forest models achieved a high accuracy of 86.82% when trained on the dataset preprocessed using iterative imputer, it was decided to ensemble them. The resulting ensemble model achieved the highest accuracy of all, at 91.01%.

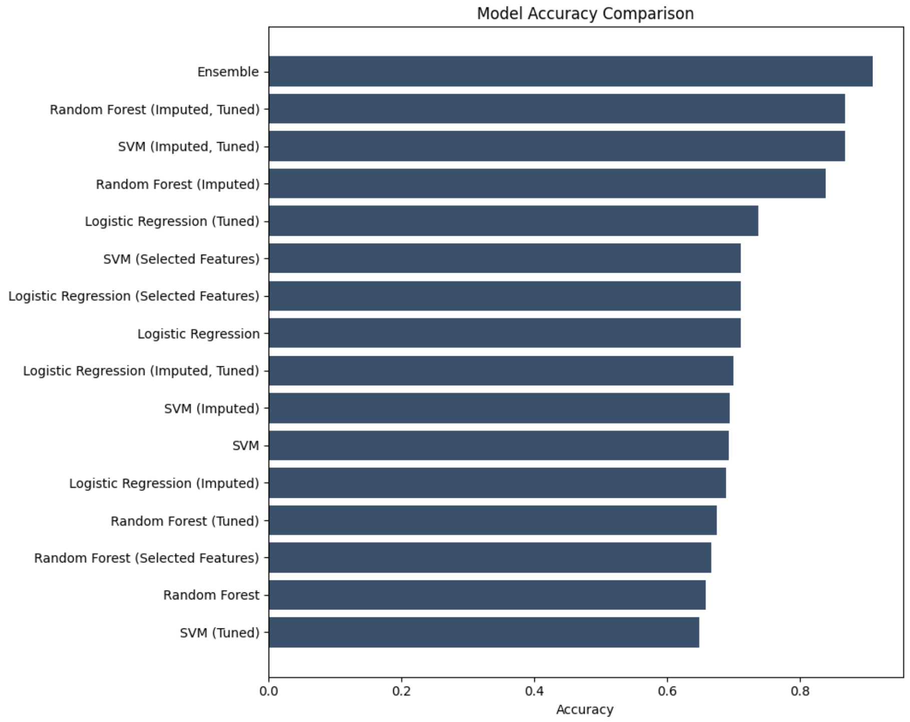


Figure 16: Comparison of Accuracies of all models

VI. CONCLUSION

In conclusion, this study presents a comprehensive framework for liver disease prediction leveraging machine learning techniques and advanced analytics. Through meticulous data preprocessing, including handling missing values and addressing class imbalance, we ensured the integrity and reliability of the dataset, establishing a strong basis for precise predictive modeling. By employing cutting-edge machine learning techniques such as Logistic Regression, Support Vector Machines (SVM), and Random Forest, and optimizing their hyperparameters, we demonstrated the effectiveness of these models in accurately classifying liver disease cases.

Furthermore, this study highlights the importance of feature selection techniques in identifying relevant predictors for liver disease classification. By employing Pearson correlation, Gain Ratio, and Random Forest feature importance, we gained valuable insights into the relationships between features and their impact on disease prediction. This enabled us to develop more interpretable and effective models, enhancing our understanding of the underlying factors contributing to liver disease.

The ensemble approach, combining the strengths of Support Vector Machine (SVM) and Random Forest classifiers, emerged as a powerful strategy to further boost predictive performance. Achieving a notable accuracy of 91.01%, the ensemble model underscores the benefits of integrating multiple classifiers to harness their collective predictive power. Overall, this project contributes to the advancement of liver disease diagnosis and management, offering valuable insights and methodologies that can inform healthcare decision-making and ultimately improve patient outcomes

REFERENCES

[1] Abdul Quadir Md, Sanika Kulkarni ,Christy Jackson Joshua, Tejas Vaichole, Senthilkumar Mohan and Celestine Iwendi"Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease" doi: https://doi.org/10.3390/ biomedicines11020581 (2023).

[2] Anil Utku "Deep Learning Based Cirrhosis Detection" Vol. 6, Issue 1, 2023, pp. 95-114 doi: https://doi.org/10.31181/oresta/060105 (2022).

[3] Ashaq Hussain Bhat, Syed Musthafa, B.Latha, T Yogameera, Dr D.Shanthi, Fathima H "Modern detection methodology for liver continual diseases with hepatitis and cirrhosis avoidance by estimating the Euclidean distance using machine learning algorithm" Vol. 38 (2) doi:10.5281/zenodo.776966 (2023).

[4] Deepika Bhupathi, Christine Nya-Ling Tan, Sreenivas Sremath Tirumula and Sayan Kumar Ray "Liver disease detection using machine learning techniques" doi: https://www.researchgate.net/publication/363857071 (2022).

[5] Deivendran.P, S. Selvakanmani, S. Jegadeesan, V. Vinoth Kumar "Liver Infection Prediction Analysis using Machine Learning to Evaluate Analytical Performance in Neural Networks by Optimization Techniques" Volume 71 Issue 3, 377-384 doi: https://doi.org/10.14445/22315381/IJETT-V71I3P240 (2023).

[6] Elias Dritsas, Maria Trigka "Supervised Machine Learning Models for Liver Disease Risk Prediction" https://doi.org/10.3390/computers12010019 (2023).

[7] Farkhondeh Razmpour, Reza Daryabeygi-Khotbehsara, Davood Soleimani, HamzehAsgharnezhad, Afshar Shamsi, Ghasem Sadeghi Bajestani, Mohsen Nematy, Mahdiyeh Razm Pour, Ralph Maddison & Sheikh Mohammed Shariful Islam"Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices" doi: https://doi.org/10.1038/s41598-023-32129-y (2021).

[8] Ganty, Jamila, Gregory Msksha Wajiga, Yusuf Musa Malgwi, Abba Hamman Maidabara "A Diagnostic model for the Prediction of Liver Disease Cirrhosis Using Machine Learning Techniques" Volume 3, Issue 1, P.36-51 doi: 10.51594/csitrj.v3i1.296 (2022).
[9] Ilkyu Park, Nakyoung Kim, Sugi Lee, Kunhyang Park, Mi-Young Son, Hyun-Soo Cho, Dae-Soo Kim"Characterization of signature trends across the spectrum of non-alcoholic fatty liver disease using deep learning method" doi: https://doi.org/10.1016/j.lfs.2022.121195 (2023).
[10] Keerthana PSM, Nimish Phalinkar, Riya Mehere, Koppula Bhanu Prakash Reddy, Nidhi Lal"A Prediction Model of Detecting Liver Diseases in Patients using Logistic Regression of Machine Learning"(2020).
[11] Latha C M "Prediction of Liver Disease using Machine Learning" Volume 2, Issue 1 doi: 10.48175/IJARSCT-5673 (2022).
[12] Mandakini Priyadarshani Behera, Archana Sarangi, Debahuti Mishra , Shubhendu Kumar Sarangi "A Hybrid Machine Learning Algorithm For Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine" (2023) .
[13] Mohammad Alauthman, Amjad Aldweesh, Ahmad Al-qerem, Faisal Aburub, Yazan Al-Smadi, Awad M. Abaker, Omar Radhi Alzubi and Bilal Alzubi "Tabular Data Generation to Improve Classification of Liver Disease Diagnosis" doi: https://doi.org/10.3390/app13042678 (2023).
[14] Ruhul Amin, Rubia Yasmin, Sabba Ruhi, Md Habibur Rahman, Md Shamim Reza "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms" (2023).
[15] Saima Afrin, F. M. Javed Mehedi Shamrat, Tafsirul Islam Nibir, Mst. Fahmida Muntasim, Md. Shakil Moharram, M. M. Imran, Md Abdulla "Supervised machine learning based liver disease prediction approach with LASSO feature selection" Vol. 10, No. 6, pp. 3369-3376 doi: 10.11591/eei.v10i6.3242 (2021).
[16] Sayalee Deshmukh, Anushka Sawant, Manasi Khopade, Pratiksha Kawale, Yashika Palan "Liver Disease Diagnosis Using Machine Learning" Vol. 6, Issue 12 (2022).
[17] Shibam Ch. Karmakar, Subham Pratihar, Shreeja Roy, Sulekha Das, Avijit Chaudhuri"Early Detection of Liver Disease by using Machine Learning" Volume 7, Issue 2 doi: 10.46647/ijetms.2023.v07i02.031 (2023).
[18] Venkateshwara Rao.M, Rajesh Saturi,B.Sushanth, C. H. Vaishnavi, Ch.Dinesh "Prediction of Liver Disease Using Machine Learning" doi: 10.31838/ecb/2023.12.si6.549 (2023).
[19] Vijay Panwar,Naved Choudhary, Sonam Mittal, Gaurav Sahu "Review of Liver Disease Prediction Using Machine Learning Algorithm" Volume 8, Issue 2 (2021).
[20] Yixin Hou, Hao Yu, Qun Zhang, Yuying Yang, Xiaoli Liu, Xianbo Wang and Yuyong Jian "Machine learning-based model for predicting the esophagogastric variceal bleeding risk in liver cirrhosis patients" doi: https://doi.org/10.1186/s13000-023-01293-0 (2023).
[21] Figure10:https://images.spiceworks.com/wp-content/uploads/2022/04/11040522/46-4.png
[22] Figure12: Support Vector Machine (SVM) Algorithm - Javatpoint
[23] Figure14:https://miro.medium.com/v2/resize:fit:1200/1*hmtbIgxoflflJqMJ_UHwXw.jpeg
[24] dataset : Indian Liver Patient Dataset (kaggle.com)