

# Comparative Analysis of Supervised Machine Learning Algorithms for Liver Disease Prediction with SMOTE Enhancement

**Anurag Dhyani**

*Computer Science and Engineering  
Graphic Era Hill University  
Dehradun, India  
anuragdhyani06@gmail.com*

**Ashutosh Rawat**

*Computer Science and Engineering  
Graphic Era Hill University  
Dehradun, India  
tosh.rawat@gmail.com*

**Gautam Bisht**

*Computer Science and Engineering  
Graphic Era Hill University  
Dehradun, India  
gbisht2282@gmail.com*

**Satvik Vats**

*Computer Science and Engineering  
Graphic Era Hill University  
Dehradun, India  
svats@gehu.ac.in*

**Vikrant Sharma**

*Computer Science and Engineering  
Graphic Era Hill University  
Dehradun, India  
vsharma@gehu.ac.in*

**Manish Mohan Baral**

*GITAM School of Business  
GITAM (Deemed to be University)  
Visakhapatnam, A.P., India  
manishmohanbaral.31@gmail.com*

**Satya Prakash Yadav**

*Computer Science and Engineering  
GLBITM  
(GLBITM), Greater Noida- 201306, India  
prakashyadav.satya@gmail.com*

**Abstract**—Chronic liver disease is a serious condition that involves ongoing damage to the liver, leading to impaired liver function and potentially life-threatening complications[19]. Detecting and treating this disease at an early stage is vital to prevent irreversible liver damage[2] and enhance patient outcomes. This study aims to compare the effectiveness of various Machine Learning algorithms with SMOTE in diagnosing chronic liver disease[20]. The algorithms evaluated in this study include Random Forest, Logistic Regression, Support Vector Machine, MLP, and other boosting algorithms like Adaboost, CatBoost, etc.

**Keywords:** Machine Learning, Liver Disease, Classification, SMOTE, Boosting.

## I. INTRODUCTION

The liver, being the largest organ in the body, serves crucial functions in metabolism, digestion, and detoxification. It produces bile to aid in fat breakdown, stores glycogen, regulates blood sugar levels, filters toxins, and produces essential proteins for blood clotting. Liver disease encompasses a wide range of conditions that disrupt the liver's normal function, varying in severity from reversible to life-threatening. As a result, liver disease poses a significant global health challenge, with approximately 2 million deaths reported each year worldwide. Extensive research is underway to explore various aspects of liver disease, including its causes, diagnosis, treat-

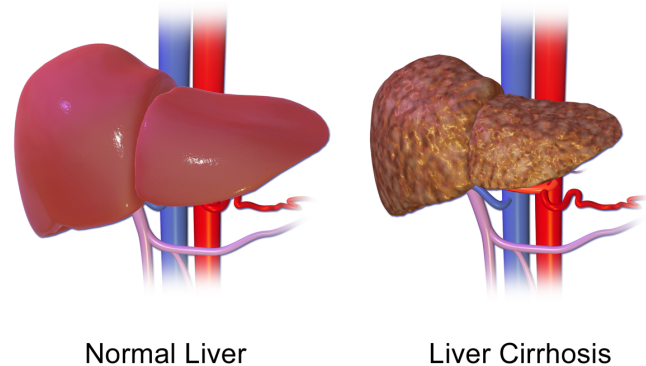


Fig. 1. Normal liver vs Liver Cirrhosis [2]

ment, and prevention. Machine learning (ML) has emerged as a promising tool for predicting chronic liver disease, and numerous studies have investigated the application of ML algorithms in this domain. For instance, a study published in the BMC Medical Informatics and Decision Making journal focused on developing a predictive model for liver cirrhosis using ML algorithms [1]. The researchers utilized demo-

graphic and clinical data from electronic health records of both cirrhosis and non-cirrhosis patients and through training a random forest model, they achieved an impressive accuracy of 86.4% in predicting cirrhosis risk. ML algorithms excel in analyzing large datasets, identifying patterns, and making accurate predictions based on the available information. In fields such as finance, ML models can forecast stock prices[4], in the field of agriculture[3][25][26], and assess investment risks. It has become an indispensable asset in the healthcare industry, allowing for the processing and analysis of extensive medical data. This capability facilitates precise diagnoses, customized treatment strategies, and the early identification of diseases[8][11][13][17][18][20][19][21][23]. These findings highlight the active research in the field of ML-based prediction for liver disease[15], emphasizing the potential of these algorithms to aid in the early identification and monitoring of liver disease progression. By harnessing the power of ML, healthcare professionals and researchers can gain valuable insights into liver disease and work towards more effective prevention and treatment strategies.

## II. LITERATURE STUDY

In the literature study, several researchers have explored the use of machine learning algorithms for the prediction of liver disease. The following studies provide valuable insights into the application of different algorithms and techniques: In the seminal work conducted by Dr. S. Vijayarani in 2015 [27], two machine learning techniques, Support Vector Machine (SVM) and Naive Bayes, were utilized to determine whether a patient had liver cirrhosis. The results revealed that SVM had the better performance, achieving a precision score of 76.6%.

In a work Sumedh (2017) [10] employed machine learning algorithms such as backpropagation and support vector machines (SVM) for liver disease prediction. The backpropagation algorithm achieved the highest prediction accuracy of 73.2%.

A.K.M Sazzadur Rahman (2019) [9] conducted a comparative study evaluating six machine learning algorithms (LR, KNN, DT, SVM, NB, and RF) on various metrics. Logistic regression demonstrated the highest accuracy and precision, achieving 75

A study by Nazmun Nahar (2019) [24] explored the use of ensemble learning methods for liver disease prediction. Five ensemble algorithms, including AdaBoost, LogitBoost, BeggRep, BeggJ48, and Random Forest, are compared in terms of accuracy, RMSE, TPR, FPR, and ROC curve. LogitBoost algorithm achieves the highest accuracy of 71.53% and outperforms other methods.

M. Banu Priya (2019) [21] conducted a work that focused on the performance analysis of liver disease prediction using machine learning algorithms. They implemented feature model construction and comparative analysis to improve prediction accuracy. The J48 algorithm, combined with feature selection,

performs better than other classification algorithms, achieving an accuracy of 95.04%

Ritesh Choudhary (2021) [7] utilized the N-Fold Cross-Validation technique and investigated basic machine learning models and boosting algorithms. Their study revealed that logistic regression showed the best overall performance.

G. Shobana (2021) [6] focused on improving model accuracy by employing Recursive Feature Elimination. They used various basic machine learning models and boosting algorithms, resulting in increased efficiency. The Gradient Boosting algorithm achieved a prediction accuracy of 94. These studies highlight the effectiveness of different machine-learning algorithms and techniques in predicting liver disease.

## III. METHODOLOGY

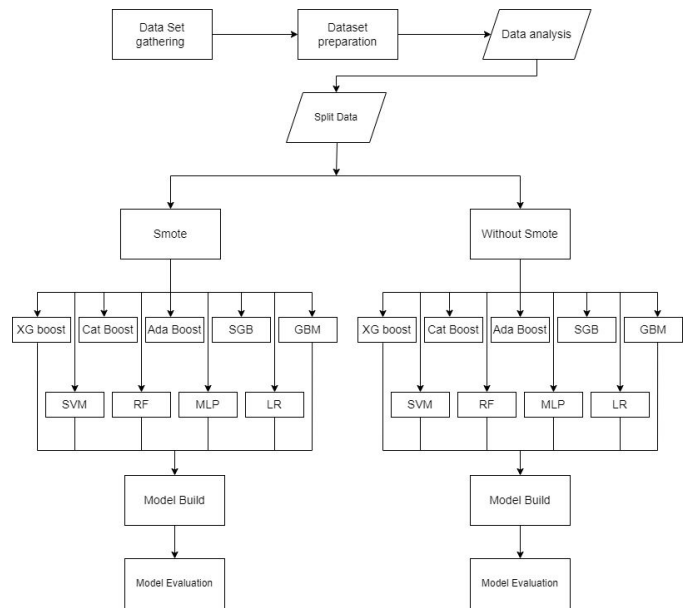


Fig. 2. Flowchart of Methodology proposed

The pre-processing phase of the methodology consists of several steps. First, missing values are removed. Next, categorical variables are transformed into numerical values to facilitate analysis. Then, the correlation among attributes is examined. To handle imbalanced data, the SMOTE enhancement technique is applied. Afterward, basic machine learning models and boosting algorithms are employed[22]. Finally, the performance of the models is evaluated.

Steps involved as shown in Figure 2 :

- Obtain the Indian Liver patient's data set from the UCI repository [5].
- Pre-process data by filling in all the missing values and changing categorical values into numerical values.
- Splitting data into sets (75% and 25%) for training and testing the data models.
- Using SMOTE to balance the dataset and enhance the performance of models used.

- Apply data to basic and boosting ml models.
- Observe and choose the model with the highest precision and accuracy.

#### A. Dataset Description

The University of California, Irvine's Machine Learning Repository provided the data set for this work. The provided dataset contained information on 583 Indian patients, 416 of which suffered from chronic liver disorders, and 167 of whom were in good health. This created an imbalance in the dataset which can lead to biased models that may perform poorly on the minority class (healthy patients) and thus to balance the dataset SMOTE was used.

This dataset has 11 specific parameters, however, we used 10 of them for our research and 1 of them as a target class. Such as,

- Patient's Age
- Patient's Gender
- TB: Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphatase
- Alamine Aminotransferase
- Aspartate Aminotransferase
- Total Proteins
- Albumin
- Albumin to Globulin Ratio
- Using a selector field, the data were divided into two sets and labeled by experts.

to identify patterns and relationships among the data. The attributes that are highly correlated are Total Bilirubin and Direct Bilirubin, Alamine Aminotransferase and Aspartate Aminotransferase, Total Proteins and Albumin, and Albumin to Globulin Ratio Ratio.

#### B. Using SMOTE

Working with imbalanced datasets requires the use of SMOTE (Synthetic Minority Over-sampling Technique), a method that is often employed in machine learning. An uneven dataset might provide a biased model since there are significantly fewer examples of one class than the other.

It functions by using a set of k-nearest[18] neighbour criteria to generate synthetic data [4]. SMOTE starts by randomly choosing information from the minority class, after which the statistics are used to determine the k-nearest acquaintances. Then, synthetic records might be created between the randomly chosen k-nearest neighbour and the random records.

#### C. Classification Models

- Random Forest (RF)

It works on using ensembled decision trees to make predictions. To predict the outcome, we trained the Random Forest model on a dataset comprising various features, including bilirubin levels, albumin levels, etc. The model's training phase involved constructing multiple decision trees and aggregating their predictions to make a final prediction. Each decision tree in the Random Forest was trained on a randomly sampled subset of the data, making them diverse and capable of capturing different aspects of the data. By combining the predictions from multiple trees, the Random Forest algorithm was able to accurately predict whether an individual has liver disease or not.

- Multilayer Perceptron (MLP)

A type of ANN, or artificial neural network[16], is made up of linked layers of synthetic neurons known as perceptrons. Information passes from the source layer through hidden layers to the resultant layer in a feedforward fashion. In this, it consists of an input layer containing 10 nodes which takes all features such as ( Total\_Bilirubin, Albumin) and, one or more hidden layers, and an output layer with 2 nodes. Each layer is composed of multiple artificial neurons, also known as nodes or units.

- Logistic Regression (LR)

Logistic regression is a statistical model used for binary classification tasks. It assumes a linear relationship between all of the input features and the log-odds of the target variable which stores the patient's disease status. The log-odds are transformed into probabilities using the sigmoid function. The model estimates the weights that

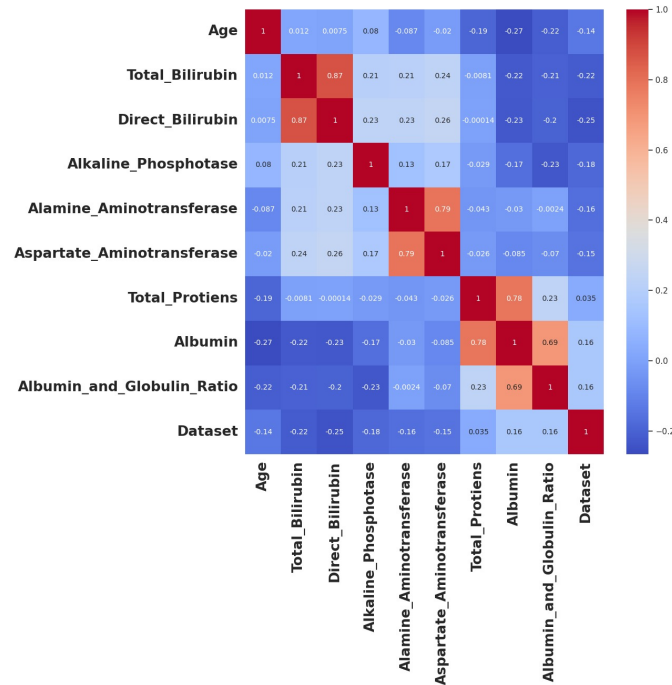


Fig. 3. Correlation heatmap

[3]Figure 2 provides a visual representation of the correlation between variables in the data frame, allowing us

minimize the difference between predicted probabilities and actual labels. New instances can be classified by calculating the weighted sum of features and applying the sigmoid function.

- **Support Vector Machine (SVM)**

SVM is a powerful machine learning technique that effectively classifies data by finding an optimal boundary between different classes. By training the SVM model on a dataset of clinical and laboratory features, such as bilirubin levels, albumin levels, and other relevant parameters. The model's training phase involved finding an optimal hyperplane that maximally separates the two classes (liver disease positive and negative) while maximizing the margin between their separations.

- **BOOSTING METHODS** Boosting is an ensemble modeling strategy that aims to construct a strong classifier from a collection of weak classifiers.

- **Gradient Boosting:** It is an augmentation procedure that generates the final model by combining numerous weak learning algorithms that have been trained on the same dataset.
- **XGBoosting (Extreme Gradient Boosting):** This technique excels at situations involving structured data and information. It is a gradient-boosting framework implementation that tries to create a set of weak decision trees that train and combine their predictions to form a strong prediction model.
- **AdaBoost:** It is a machine learning technique that is part of the ensemble learning method family. It combines several weak learners (often basic decision trees) to create one strong learner.
- **CatBoost :** (Categorical Boosting) designed for handling categorical features and variables in machine learning tasks.
- **Stochastic Gradient Boosting (SGB):** This boosting technique that incorporates unpredictability into the training process. We integrate the gradient boosting theory with the idea of random sampling to improve the performance and generalization abilities of ensemble models.

#### IV. RESULT

##### A. PERFORMANCE EVALUATION

In our study, we incorporated objective metrics to evaluate the performance of various classification algorithms in terms of test execution. The evaluation procedures employed included accuracy, Recall, specificity, precision, and F1 measure. These metrics[13] were used to assess the effectiveness and efficiency of the classification methods.

- **Accuracy** Accuracy is a performance metric in machine learning that evaluates the overall correctness of predictions. It measures the proportion of correctly

predicted results, considering both positive and negative instances, out of the total instances.

- **Precision** Precision is a performance metric in machine learning that measures the proportion of correctly predicted results out of all the predicted results. It quantifies the accuracy of the model by focusing on the correctness of positive predictions.
- **Recall / Sensitivity** It quantifies the proportion of correctly predicted results out of all the actual results.

##### B. Analysis of the Result

In this work, we considered different ml models for the classification of liver diseases which includes basic Machine Learning models and boosting algorithms.

**Algorithms without Smote**

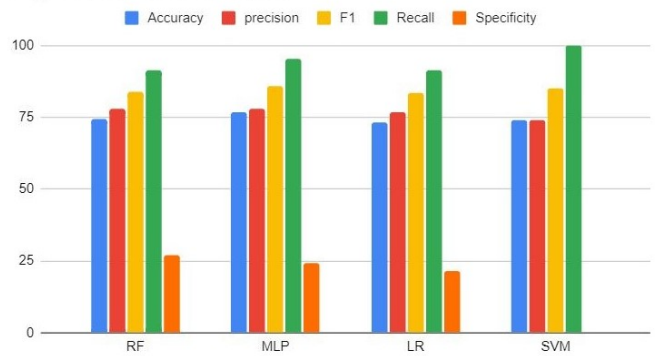


Fig. 4. Performance of ML models Without SMOTE

TABLE I  
ALGORITHMS WITHOUT SMOTE

Algorithms	Accuracy	Precision	F1	Recall	Specificity
<b>RF</b>	74.6	78	84	91.4	27
<b>MLP</b>	76.7	78.1	85.8	95.2	24.3
<b>LR</b>	73.2	76.8	83.4	91.4	21.6
<b>SVM</b>	73.9	73.9	85	100	0
<b>ADA</b>	76.7	78.5	85.7	94.2	27
<b>GBM</b>	73.9	76.5	84.1	93.9	18.9
<b>SGB</b>	73.9	75	84.6	97.1	8.1
<b>CAT</b>	76	80.8	84.5	88.5	40.5
<b>XG</b>	71.8	77.7	81.9	86.6	29.7

Figure 4 illustrates the performance of basic models without using the SMOTE technique to balance the datasets.

In terms of accuracy, the MLP (Multi-Layer Perceptron) model achieved the highest accuracy of 76.7%. On the other hand, Logistic Regression exhibited the worst performance among all the models, with an accuracy of 73.2%.

When considering precision, the MLP model once again performed the best, achieving a precision of 78.1%. Random Forest came in second place with a precision that was 0.1 lower, and SVM (Support Vector Machine) had the lowest precision of 73.9%. The F1 scores of the algorithms were

fairly similar. SVM and MLP achieved the highest scores of 85, while Logistic Regression had the lowest score of 83. In Figure 5, the performance of models utilizing the SMOTE technique is depicted. Regarding accuracy, both Random Forest and MLP attained the highest accuracy of 76%. Conversely, Logistic Regression (LR) exhibited the lowest accuracy among all models, with a value of 70.4%.

When examining precision, LR obtained the highest value of 90.9, indicating the highest precision level among the models. On the other hand, SVM had the lowest precision of 75.1.

In terms of F1 scores, SVM achieved the highest score of 85.1, indicating the best overall performance. LR obtained the lowest score of 76.9 among the models.

Algorithms with smote

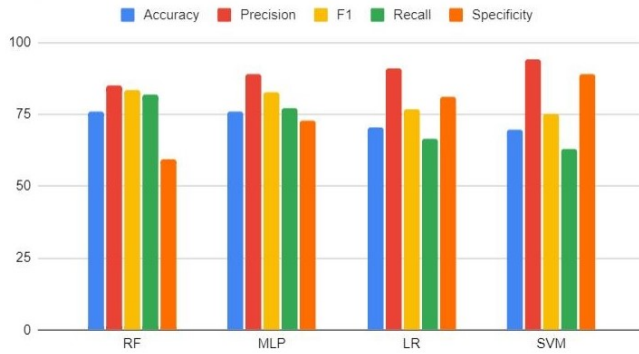


Fig. 5. Performance of ML models With SMOTE

The results from Figure 6, which presents the performance of boosting algorithms without using SMOTE: Adaboost got the best result of 76.6%, while XGBoost had the worst accuracy of 71.83%. CatBoost had the best precision score of 80.8, while SGB had the worst precision score of 75. The F1 scores of all boosting algorithms were fairly similar. AdaBoost had the highest F1 score of 85.7, while XGBoost had the lowest F1 score of 81.9.

Boosting Algo without smote

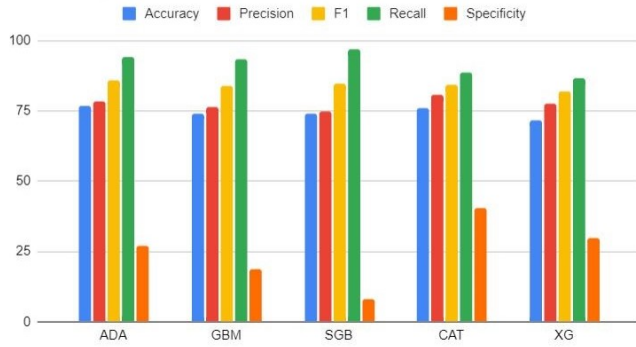


Fig. 6. Boosting algorithms without SMOTE

TABLE II  
ALGORITHMS WITH SMOTE

Algorithms	Accuracy	Precision	F1	Recall	Specificity
RF	76	85.1	83.4	81.9	59.4
MLP	76	89	82.6	77.1	72.9
LR	70.4	90.9	76.9	66.6	81
SVM	89.1	62.8	75.4	62.8	89.1
ADA	69.7	88.7	76.7	67.6	75.6
GBM	71.1	78	81.2	84.7	32.4
SGB	73.9	83.3	82.1	80.9	54
CAT	75.3	85	82.9	80.9	59.4
XG	70.4	86.2	78.1	71.4	67.5

Boosting Algo with smote

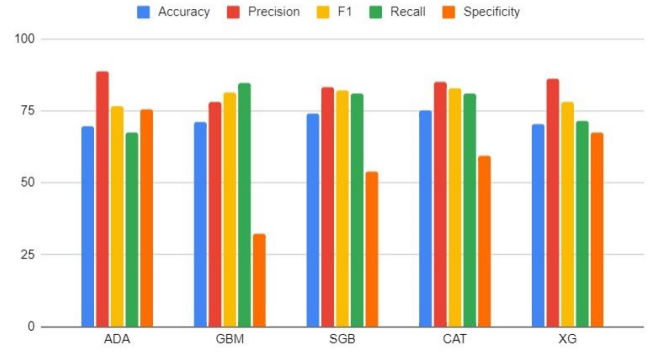


Fig. 7. Boosting algorithms with SMOTE

The results from Figure 7, which compares the performance of boosting algorithms after using the SMOTE technique:

In terms of accuracy, CatBoost attained the best score of 75.3%, while AdaBoost got the lowest at 69.7 %. There was a significant increase in precision. AdaBoost achieved the most precision score of 88.75, while GBM had the lowest score of 78. Regarding F1 scores, CatBoost scored the highest 82.9, while AdaBoost had the lowest score of 76.7.

The utilization of Smote resulted in lower recall scores compared to when it was not employed. Without using SMOTE SVM got the highest score of 100 but after SMOTE enhancement recall dropped and SVM reduced to 62.8 lowest among the algorithm used.

## V. CONCLUSION

Imbalanced datasets present a challenge for machine learning algorithms, particularly in medical disease predictions where precision is crucial. To overcome this challenge, SMOTE was employed to balance the dataset by generating synthetic samples and oversampling the minority class. This effectively improved the precision of all algorithms, including Logistic Regression, which achieved a significant increase to 90.9%. While there was a slight improvement in overall accuracy, there was a gradual decrease in recall.

Precision and recall are vital in medical disease predictions, with higher precision reducing the risk of false positives. The MLP algorithm demonstrated the best performance in terms of



accuracy and precision after applying SMOTE, highlighting its potential in medical applications.

In conclusion, combining SMOTE with feature reduction techniques shows promise for enhancing the performance of algorithms in medical disease prediction. This approach can benefit the medical field by improving the accuracy and precision of predictions, ultimately leading to more accurate diagnoses and improved patient care.

## REFERENCES

- [1] E. Dritsas and M. Trigka, "Supervised Machine Learning Models for Liver Disease Risk Prediction," *Computers*, vol. 12, no. 1, p. 19, Jan. 2023, doi: 10.3390/computers12010019. [Online]. Available: <http://dx.doi.org/10.3390/computers12010019>
- [2] Fig. 1. Normal liver vs Liver Cirrhosis [https://surgery.ucsf.edu/media/9244862/Liver\\_Cirrhosis.png](https://surgery.ucsf.edu/media/9244862/Liver_Cirrhosis.png)
- [3] A. Agarwal, S. Vats, R. Agarwal, A. Ratra, V. Sharma and A. Jain, "Efficient NetB3 for Automated Pest Detection in Agriculture," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1408-1413.
- [4] A. Agarwal, S. Vats, R. Agarwal, A. Ratra, V. Sharma and L. Gopal, "Sentiment Analysis in Stock Price Prediction: A Comparative Study of Algorithms," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1403-1407.
- [5] Indian Liver Dataset. [Online]. Available: Kaggle. [Accessed: May 4, 2023]. <https://www.kaggle.com/datasets/uciml/indian-liver-patient-records>.
- [6] G. Shobana, K. Umamaheswari Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021) IEEE Xplore Part Number: CFP21K25-ART
- [7] Ritesh Choudhary, T. Gopalakrishnan, D. Ruby A. Gayathri, Vishnu Srinivasa Murthy and Rishabh Shekhar An Efficient ML Model for Predicting Liver Disease Using Machine Learning
- [8] M. Bajaj, P. Rawat, C. Bhatt, R. Chauhan and T. Singh, "Heart Disease Prediction using Ensemble ML," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 680-685, doi: 10.1109/ICSCDS56580.2023.10104770.
- [9] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain "A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms" INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 8, ISSUE 11, NOVEMBER 2019.
- [10] Sumedh Sontakke, Jay Lohokare, Reshul Dani Diagnosis of Liver Diseases using Machine Learning 2017 International Conference on Emerging Trends & Innovation in ICT (ICEL) Pune Institute of Computer Technology, Pune, India, Feb 3-5, 2017.
- [11] P. Rawat, M. Bajaj, S. Vats and V. Sharma, "ASD Diagnosis in Children, Adults, and Adolescents using Various Machine Learning Techniques," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 625-630, doi: 10.1109/DICCT56244.2023.10110166.
- [12] <https://www.javatpoint.com/confusion-matrix-in-machine-learning#:~:text=The%20confusion%20matrix%20is%20a,related%20terminologies%20may%20be%20confusing>
- [13] P. Rawat, M. Bajaj, S. Mehta, V. Sharma and S. Vats, "A Study on Cervical Cancer Prediction using Various Machine Learning Approaches," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 1101-1107, doi: 10.1109/ICIDCA56705.2023.10099493.
- [14] Rong-Ho Lin, "An Intelligent Model for Liver Disease Diagnosis," *Artificial Intelligence in Medicine*, 2009".
- [15] P. Rawat, M. Bajaj, P. Prerna, S. Vats, V. Sharma and P. Das, "A Study on Liver Disease Using Different Machine Learning Algorithms," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 721-727, doi: 10.1109/CICTN57981.2023.10141325.
- [16] M. Bajaj, P. Rawat, A. Bhatt, V. Sharma, A. Jain and N. Kumar, "Classification And Prediction of Brain Tumors and its Types using Deep Learning," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 705-710, doi: 10.1109/CICTN57981.2023.10140647.
- [17] S. Vats and B. B. Sagar, "Performance evaluation of K-means clustering on Hadoop infrastructure," *J. Discret. Math. Sci. Cryptogr.*, vol. 22, no. 8, 2019, doi: 10.1080/09720529.2019.1692444.
- [18] S. Vats, B. B. Sagar, K. Singh, A. Ahmadian, and B. A. Pansera, "Performance evaluation of an independent time optimized infrastructure for big data analytics that maintains symmetry," *Symmetry (Basel)*, vol. 12, no. 8, 2020, doi: 10.3390/SYM12081274.
- [19] A. Agarwal, S. Vats, R. Agarwal, A. Ratra, V. Sharma and A. Jain, "Efficient NetB3 for Automated Pest Detection in Agriculture," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1408-1413.
- [20] M. Bajaj, P. Rawat, C. Bhatt, R. Chauhan and T. Singh, "Heart Disease Prediction using Ensemble ML," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 680-685, doi: 10.1109/ICSCDS56580.2023.10104770.
- [21] Priya M.B., Juliet, P.L. and Tamilselvi, P.R., 2018. Performance analysis of liver disease prediction using machine learning algorithms. *Int. Res. J. Eng. Technol.*, 5(1), pp.206-211.
- [22] S. Vats, S. Singh, G. Kala, R. Tarar, and S. Dhawan, "iDoc-X: An artificial intelligence model for tuberculosis diagnosis and localization," *J. Discret. Math. Sci. Cryptogr.*, vol. 24, no. 5, pp. 1257-1272, 2021.
- [23] M. Bajaj, P. Rawat, Diksha, S. Vats, V. Sharma and L. Gopal, "Prediction of Mental Health Treatment Adherence using Machine Learning Algorithms," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 716-720, doi: 10.1109/CICTN57981.2023.10141520.
- [24] N. Nahar, F. Ara, M. A. I. Neloy, V. Barua, M. S. Hossain and K. Andersson, "A Comparative Analysis of the Ensemble Method for Liver Disease Prediction," 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICIET48527.2019.9290507.
- [25] Dolli, P. Rawat, M. Bajaj, S. Vats and V. Sharma, "An Analysis of Crop Recommendation Systems Employing Diverse Machine Learning Methodologies," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 619-624, doi: 10.1109/DICCT56244.2023.10110085.
- [26] P. Rawat, M. Bajaj, V. Sharma and S. Vats, "A Comprehensive Analysis of the Effectiveness of Machine Learning Algorithms for Predicting Water Quality," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 1108-1114, doi: 10.1109/ICIDCA56705.2023.10099968.
- [27] Dr. S. Vijayarani, Mr.S.Dhayanand "Liver Disease Prediction using SVM and Naïve Bayes Algorithms" *International Journal of Science, Engineering and Technology Research (IJSETR)* Volume 4, Issue 4, April 2015s