Machine Learning Approach Analysis for Early-Stage Liver Disease Prediction in the Context of Bangladesh and India

Md. Alif Sheakh, Taminul Islam, Md. Rezwane Sadik, and Md. Masum Rana

Abstract Liver disease can cause death in millions of individuals worldwide. Early detection and accurate diagnosis improve patient outcomes and reduce mortality. Machine learning algorithms can detect and diagnose liver diseases more accurately and affordably. To test machine learning methods for early liver disease diagnosis, this study created a prediction model that distinguished those with liver illnesses from those without. Machine learning algorithms studied included Extra Tree, XGBoost, Random Forest, CatBoost, LogitBoost, Gradient Boosting, AdaBoost, and KNN. By using oversampling methodologies and assessing metrics like accuracy, recall, precision, and $F1$-score, the Extra Tree algorithm was found to be the most effective way for early liver problem detection with 99% accuracy. Undersampling, oversampling, and SMOTE reduced class imbalance. This problem is widespread in many machine learning applications. Machine learning algorithms may improve liver disease detection, early intervention, and healthcare costs. However, data protection, ethics, and healthcare professional training and understanding must be adequately addressed. This research is a major step toward a more accurate and practical liver disease diagnostic tool that could help individuals stay healthy and prevent liver illnesses.

Keywords Liver disease · Machine learning · Disease prediction · Random forest · Liver illnesses prediction

Md. Alif Sheakh (✉) · T. Islam
Daffodil International University, Dhaka, Bangladesh
e-mail: alif15-3292@diu.edu.bd

T. Islam
e-mail: taminul15-11116@diu.edu.bd

Md. Rezwane Sadik · Md. Masum Rana
University of South Dakota, Vermillion, SD, USA
e-mail: mdrezwane.sadik@coyotes.usd.edu

Md. Masum Rana
e-mail: mdmasum.rana@coyotes.usd.edu

A. Iglesias et al. (eds.), *Proceedings of World Conference on Information Systems for Business Management*, Lecture Notes in Networks and Systems 834,
https://doi.org/10.1007/978-981-99-8349-0_23

1 Introduction

Liver disease is a serious public health concern around the world, affecting millions of individuals and leading to potentially fatal disorders such as cirrhosis and liver cancer [1]. The early detection and precise identification of liver disorders are crucial to improving patient outcomes and lowering mortality rates. Machine learning (ML) algorithms have developed as potent tools in the biomedical industry in recent years, giving promising possibilities for predicting and detecting liver illnesses with increased efficiency and cost-effectiveness.

The liver, the body's largest organ, performs about 500 critical activities, including bile generation, glycogen synthesis, blood cleansing from chemicals and poisons, and vitamin storage [2]. With such vital roles, any disruption in liver function can have major consequences for overall health. To tackle the difficulties associated with early identification of liver illnesses, researchers have turned to machine learning technologies. Machine learning, an area of artificial intelligence, employs algorithms that learn from input–output data pairs to equip computers with the ability to learn without explicit programming.

In the healthcare area, machine learning [3] has shown considerable promise in terms of improving diagnostic accuracy, facilitating early intervention, and lowering healthcare expenditures. This project's purpose is to investigate the use of machine learning algorithms in the early diagnosis of liver disorders. We hope to construct a predictive model that can accurately identify patients with and without liver disorders by leveraging the strength of categorization algorithms. Several machine learning methods [4] will be assessed in this study. By comparing the performance of several algorithms, we seek to determine the most effective strategy for the early detection of liver problems. A few performance evaluation criteria that will be used to analyze how well the models can predict the future are accuracy, recall, precision, and $F1$-score. Machine learning algorithms can detect trends and forecast probable consequences by analyzing enormous databases of medical information [5]. With this knowledge, healthcare providers can make informed decisions about appropriate interventions and treatments for patients. One of the primary advantages of machine learning algorithms in liver disease diagnostics is their capacity to detect disorders earlier than traditional approaches [6].

Early diagnosis of liver illnesses can considerably improve patient outcomes since it allows for earlier intervention and successful therapy before the disease advances further. The early diagnosis of liver illnesses using machine learning algorithms is a promising technique for improving patient outcomes and lowering mortality rates. The development of precise and efficient methods for early identification of liver illnesses can provide healthcare practitioners with dependable tools that can improve diagnosis, improve patient outcomes, and potentially save lives. As the field of machine learning in healthcare continues to advance at a rapid pace, there are various problems involved with its implementation. These include data privacy concerns, ethical constraints, and the requirement for specialized training and understanding among healthcare providers. However, if these issues are adequately handled, machine learning algorithms have the potential to revolutionize liver disease diagnostics and open the way for more personalized and effective healthcare solutions.

2 Related Work

Several studies have looked into the use of machine learning algorithms in the detection of liver disease in the early stages.

Behera et al. [7] proposed a hybrid machine learning technique for heart and liver disease prediction that combines modified particle swarm optimization and support vector machine. They used a dataset from the UCI Machine Learning Repository. The construction of this hybrid model, obtained 97.41% accuracy when compared to other models such as SVM, PSOSVM, CPSOSVM, and CCPSOSVM.

Rahman et al. [8] conducted a comparison study on liver disease prediction using supervised machine learning algorithms. The goal of the study was to evaluate how effectively alternative algorithms may reduce the high cost of detecting chronic liver disease. The UCI Machine Learning Repository provided data on 583 liver patients for the study. Six algorithms, including LR, DT, KNN, SVM, Naive Bayes (NB), and RF, were evaluated using a variety of approaches, including accuracy, precision, recall, $f1$ score, and specificity. The logistic regression strategy was the most accurate (75%).

Using ensemble machine learning techniques, Md et al. [9] created a superior preprocessing strategy for diagnosing liver illness. They employed the Indian Liver Patient Dataset and data preparation procedures such as feature scaling, feature selection, and data balancing. The GB, XGBoost, Bagging, RF, Extra Tree (ET), and Stacking ensemble algorithms were utilized to train the required features. The suggested model attained the highest testing accuracy of 91.82% and 86.06% using the Extra Tree classifier and Random Forest, respectively.

Dritsas et al. [10] used machine learning techniques to predict the prevalence of liver disease. Using the Indian Liver Patients' Records Dataset, they examined various supervised machine learning models and ensemble techniques. With tenfold cross-validation, the voting classifier outperformed other SMOTE-following models, with accuracy, recall, and F-measure of 80.1%, precision of 80.4%, and AUC of 88.4%. The report underlines the importance of early detection of liver disease while ignoring any limitations.

Khan et al. [11] used a variety of classification approaches to detect liver illness at an early stage. The parameters and anticipated accuracies of RF, LR, and the Separation Algorithm were compared. Datasets on liver disease were gathered from various sources. In terms of predicting liver illness, RF outperformed the other algorithms. The paper contributes to better healthcare services and early disease identification. It does, however, have the limitation of focusing exclusively on related research and not taking other contributing factors in liver disease prediction into account.

Assegie et al. [12] created a machine learning model for predicting liver illness using the SVM and KNN algorithms. They tested the model with the Indian Liver Dataset. SVM was shown to be 82.90% accurate in the outcomes. Based on experimental test findings, SVM outperformed KNN in predicting liver illness. The research helps to improve the detection and treatment of liver disease.

Sivasangari et al. [13] propose using Decision Tree, Support Vector Machine, and Random Forest models to accurately predict liver illness. The study makes use of the Indian Liver Dataset, which has 583 patient records and ten features. SVM has an accuracy of 95.18%, DT has an accuracy of 87.95%, and RF has an accuracy of 92.77%. The models have high accuracy and reliability for early detection of liver disease.

3 Proposed Methodology

Several phases are involved in the suggested methodology for investigating the application of machine learning algorithms in the early diagnosis of liver disorders. Figure 1 depicts the proposed model workflow for this study.

At beginning, we combed through reliable sources to compile a database of patients suffering from liver illness. Following the completion of the data collection, we examined the dataset for any areas that contained null values and then inserted median values into such areas. Then, we used label encoding to convert categorical data to numerical data. When it comes time to investigate the connections between the features, the correlation matrix will be put to use. We are going to get rid of any aspects that have a strong link with one another or that include the same information. After that, we scaled the data to ensure that all of the attributes

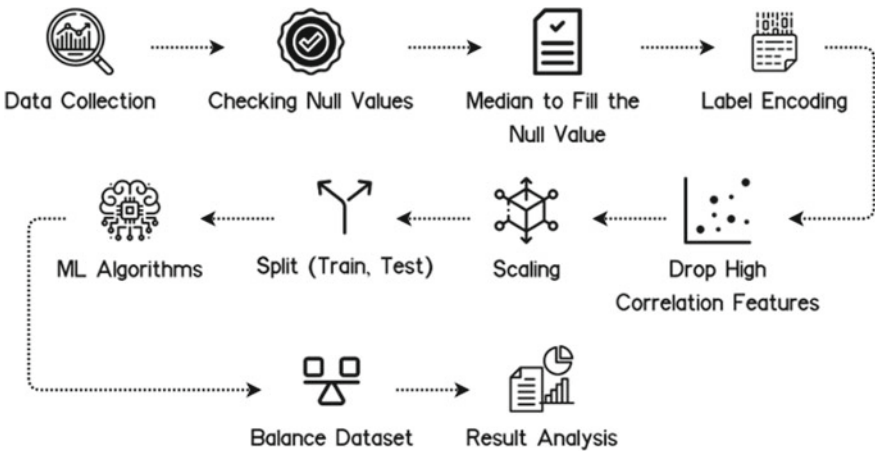


Fig. 1 Proposed model workflow of this research

were measured on the same scale. After that, the dataset is going to be divided into a training set and a testing set. Then, in order to forecast liver disease, we utilized a total of eight different machine learning strategies. These strategies included the Gradient Boosting Classifier [14], the Random Forest Classifier [15], the Extra Tree Classifier [16], the K-Nearest Neighbors [17], the Extreme Gradient Boosting Classifier [18], the AdaBoost Classifier [19], the LogitBoost Classifier [20], and the CatBoost Classifier [21]. In order to arrive at a conclusive prediction, we employed a voting classifier to combine the results of all of the different models. In order to increase the performance of the model, we first attempted to balance the dataset by either undersampling the majority class, oversampling the minority class, or utilizing SMOTE. After that, the performance of the model will be assessed using a wide variety of metrics, such as accuracy, precision, recall, $F1$-score, confusion matrix, and many others. The findings of the investigation are broken down and examined. In general, this methodology offers a methodical way to predict liver disease through the utilization of machine learning algorithms while simultaneously preserving the reliability and quality of the data.

3.1 Data Collection and Preprocessing

The dataset for this study was gathered from a variety of sources, including the UCI ML Repository [22] and from Dhaka Medical College. The dataset contains 1924 patient records, 1370 of which have liver disease and 554 of which do not. Patients were collected from diverse areas. The information includes age, gender, alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, total bilirubin, albumin, direct bilirubin, total proteins, and albumin and globulin ratio. These characteristics are critical for determining the patients' liver health. The collection contains 1402 male patient records and 456 female patient records. To ensure confidentiality, every patient who is older than 89 is labeled as 90. Before using the dataset for this research, a full evaluation will be performed to assure its reliability and accuracy, including confirming the data sources and undertaking data quality checks.

The initial step in the data preprocessing stage is interacting with missing values in the dataset. We discovered null values in the albumin and globulin ratio feature in this scenario. Because the data in this feature is normally distributed but slightly right-skewed, as seen in Fig. 2, we decided to fill in the null values using the median approach. We then looked at two more characteristics, total bilirubin and direct bilirubin, and observed that the records had massive outliers. This was obvious in Figs. 3 and 4, which indicated that the highest values in these features were much greater than the typical range. To overcome this issue, we decided to delete the outliers, which would improve the accuracy of our model.

Following that, we ran label encoding on the dataset to check for feature correlations. We discovered a strong relationship between total bilirubin and direct bilirubin, as well as Alamine Aminotransferase and aspartate aminotransferase, total proteins and albumin, in Fig. 5. We opted to remove one feature from each associated pair to

Fig. 2 Albumin and globulin ratio feature histogram

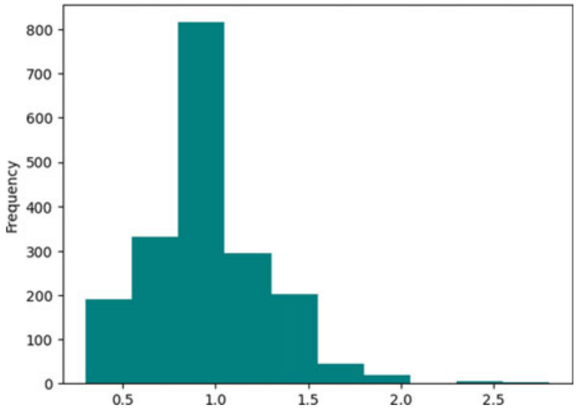


Fig. 3 Boxplot of total bilirubin

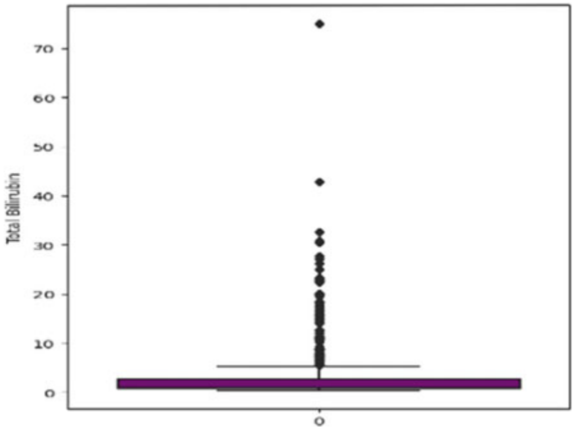
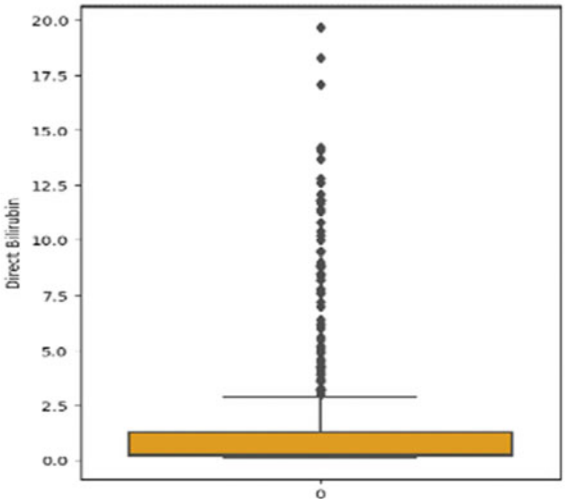


Fig. 4 Boxplot of direct bilirubin



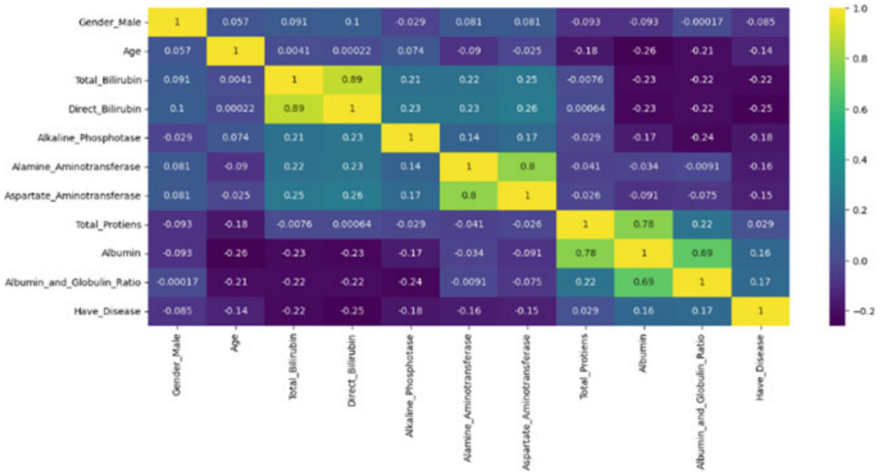


Fig. 5 Correlation between features

improve the training speed and accuracy of our model. As a result, we took direct bilirubin, aspartate aminotransferase, and albumin out of our model.

Furthermore, we discovered imbalanced classes in the dataset. To address this issue, we resampled our data to reduce errors. The data were then scaled, and eight machine learning models were used to calculate accuracy, precision, recall, and confusion matrix scores. After that, we used a voting classifier to boost the accuracy of our model even further. Finally, undersampling, oversampling, and SMOTE techniques [23] were used to balance the majority and minority classes. The outcomes were analyzed to establish the efficacy of our strategy.

4 Experimental Result

Before balancing the dataset, the results of all machine learning algorithms are displayed in Table 1. Among all algorithms, the Extra Tree algorithm had the highest accuracy, reaching 99%. It had 0.95 precision and 0.96 recall. The CatBoost algorithm came in second place in terms of accuracy, with a score of 97%, a precision of 0.95, and a recall of 0.91. AdaBoost and LogitBoost scored 76% and 98% accuracy, respectively, with precision ratings of 0.70 and 0.97 and recall scores of 0.48 and 0.96. The KNN and XGBoost algorithms both achieved 77% accuracy, with KNN achieving 0.66 precision and 0.61 recall and XGBoost achieving 0.94 precision and 0.95 recall. The Gradient Boost algorithm obtained 84% accuracy, 0.92 precision, and 0.55 recall. Random Forest had the highest accuracy among KNN, XGBoost, and Gradient Boost, but was less accurate than Extra Tree, with 99% accuracy, 0.92 precision, and 0.94 recall. The ensemble model, which combined AdaBoost, Extra

Table 1 The outcome before balancing the dataset

Algorithms	Accuracy	Precision	Recall
RF	0.99	0.92	0.94
KNN	0.77	0.66	0.61
Gradient boost	0.84	0.92	0.55
XGBoost	0.99	0.94	0.95
Extra tree	**0.99**	**0.95**	**0.96**
AdaBoost	0.76	0.70	0.48
LogitBoost	0.98	0.97	0.96
CatBoost	0.97	0.95	0.91
Voting classifier (AdaBoost, extra tree, CatBoost)	0.98	0.97	0.94

Tree, and CatBoost, had an accuracy score of 98%. The precision and recall scores for this ensemble model were 0.97 and 0.94, respectively.

Unbalanced data can undoubtedly hinder machine learning techniques. In order to address this issue, the current study employed undersampling, oversampling, and Synthetic Minority Oversampling Technique (SMOTE) on the dataset under consideration.

Following the use of undersampling, the outputs of several machine learning algorithms were analyzed, and the findings are shown in Table 2. The method with the greatest accuracy score was XGBoost, with a score of 0.95, closely followed by Random Forest (RF), with a score of 0.94. Among the algorithms tested, RF, Extra Tree, and CatBoost had the highest precision scores, ranging from 0.90 to 0.94. When it came to recall, Gradient Boost came out on top, with a score of 0.97. Notably, the voting classifier, which included AdaBoost, Extra Tree, and CatBoost, performed well in terms of recall, matching the maximum score of 0.97. Based on these results, XGBoost and Random Forest have the highest accuracy ratings, while Gradient Boost and the voting classifier have good precision and recall.

Table 2 The outcome after undersampling the dataset

Algorithms	Accuracy	Precision	Recall
RF	0.94	0.90	0.97
KNN	0.74	0.68	0.86
Gradient boost	0.88	0.82	0.97
XGBoost	**0.95**	**0.94**	**0.97**
Extra tree	0.94	0.93	0.93
AdaBoost	0.76	0.69	0.88
LogitBoost	0.86	0.83	0.91
CatBoost	0.93	0.89	0.97
Voting classifier (AdaBoost, extra tree, CatBoost)	0.92	0.88	0.97

Table 3 The outcome after oversampling the dataset

Algorithms	Accuracy	Precision	Recall
RF	0.99	0.97	0.98
KNN	0.91	0.85	0.97
Gradient boost	0.91	0.87	0.96
XGBoost	0.97	0.95	0.97
Extra Tree	**0.99**	**0.98**	**0.99**
AdaBoost	0.78	0.76	0.81
LogitBoost	0.97	0.95	0.92
CatBoost	0.98	0.94	0.97
Voting classifier (AdaBoost, extra tree, CatBoost)	0.99	0.94	0.95

After oversampling the dataset, the outputs of several machine learning algorithms were evaluated, and the findings are shown in Table 3. Random Forest (RF) had the best accuracy score of 0.99. It was closely followed by Extra Tree (0.92), XGBoost (0.90), and CatBoost (0.86). Extra Tree achieved the greatest precision score of 0.89 among the defined algorithms. KNN, on the other hand, achieved the greatest recall score of 0.76. These results emphasize the greater accuracy of Random Forest and Extra Tree, with Extra Tree earning the greatest precision score and KNN achieving the highest recall score among the defined algorithms.

After applying the SMOTE to the dataset, multiple machine learning algorithms' results are shown in Table 4. Accuracy, precision, and recall were 0.99 for the Random Forest algorithm. With accuracy, precision, and recall of 0.99, the Extra Tree algorithm likewise scored well across all criteria. CatBoost performed well, with accuracy of 0.99, precision of 0.97, and recall of 0.98. The KNN algorithm scored 0.91 with a precision of 0.84 and a recall of 0.99. Gradient Boost's accuracy was 0.86, precision 0.80, and recall 0.94. XGBoost accuracy, precision, and recall were 0.98, 0.98, and 0.99, respectively. However, AdaBoost scored 0.76 for accuracy, 0.72 for precision, and 0.83 for recall. Logit Boost's accuracy, precision, and recall were greater. The voting classifier's accuracy, precision, and recall were 0.98, 0.96, and 0.98, respectively. These data show how SMOTE affects machine learning algorithms. The dataset was handled well by algorithms like Extra Tree, Random Forest, and CatBoost, which had great accuracy, precision, and recall. Meanwhile, algorithms like AdaBoost had lower ratings, indicating the need for future improvements or other methods. The voting classifier performed well by combining the algorithms' strengths.

Table 4 The outcome after SMOTE the dataset

Algorithms	Accuracy	Precision	Recall
RF	**0.99**	**0.98**	**0.99**
KNN	0.91	0.84	0.99
Gradient boost	0.86	0.80	0.94
XGBoost	0.98	0.98	0.99
Extra Tree	**0.99**	**0.99**	**0.99**
AdaBoost	0.76	0.72	0.83
LogitBoost	0.98	0.97	0.99
CatBoost	**0.99**	**0.97**	**0.98**
Voting classifier (AdaBoost, extra tree, CatBoost)	0.98	0.96	0.98

5 Discussion

The results of our study are especially noteworthy since they demonstrate how successful a variety of machine learning methods were then applied to a balanced dataset. The utilization of methods such as undersampling, oversampling, and SMOTE is a good practice since it guarantees that the dataset that is used for training is indicative of the actual data distribution. This is one of the many reasons why this approach is considered advantageous. This technique helps to relieve problems connected with class imbalance, which is a common impediment in many applications of machine learning. Class imbalance is one of the challenges that this strategy addresses. It is essential to point out that the findings of our study showed that increasing the number of classifiers available in each domain resulted in the highest possible level of accuracy. Ultimately, our findings show how critical it is to carefully select appropriate machine learning algorithms and to ensure that datasets are well-balanced in order to gain the greatest potential performance from models. In order to make accurate predictions, it is essential to handle class imbalance, which should be taken into account while selecting the appropriate algorithms to use. Our findings can be useful to researchers and practitioners working in this sector since they provide insights into best practices and techniques for getting superior results.

6 Conclusion and Future Work

We aimed to create a method for the accurate classification of individuals with early detection of liver illness by making use of a variety of supervised machine learning classifiers. Following an in-depth investigation into the patient information parameters and the application of eight distinct algorithms, we discovered that the Extra Tree classifier achieved the greatest accuracy score across the board, with Random Forest techniques coming in a close second. In addition, we put an algorithmic voting classifier made up of AdaBoost, Extra Tree, and CatBoost through its paces, and the results showed that it was successful in raising overall performance. Our research has major implications for the research and development of healthcare, particularly in low-income nations with restricted access to medical resources. Our method, which involves early diagnosis and the provision of guidance on the upkeep of healthy livers, has the potential to assist in the prevention of chronic diseases and to contribute to the improvement of public health. The dataset is one of the limitations of this work. In future, it can be created more effectively with more attributes. As we move forward, there is still room for additional study to be done in this area, including the investigation of other algorithms and the refining of the prediction model to improve its accuracy. In general, this work offers an important step toward establishing a diagnostic tool for liver illness that is more trustworthy and effective.

References

1. Asrani SK, Devarbhavi H, Eaton J, Kamath PS (2019) Burden of liver diseases in the world. J Hepatol 70(1):151–171. https://doi.org/10.1016/J.JHEP.2018.09.014
2. Messelmani T et al (2022) Liver organ-on-chip models for toxicity studies and risk assessment. Lab Chip 22(13):2423–2450. https://doi.org/10.1039/D2LC00307D
3. Islam T, Kundu A, Islam Khan N, Chandra Bonik C, Akter F, Jihadul Islam M (2022) Machine learning approaches to predict breast cancer: Bangladesh perspective, pp 291–305. https://doi.org/10.1007/978-981-19-2541-2_23
4. Islam T et al (2023) Review analysis of ride-sharing applications using machine learning approaches. In: Computational statistical methodologies and modeling for artificial intelligence. CRC Press, New York, pp 99–122. https://doi.org/10.1201/9781003253051-7
5. Sheakh MA, Sazia Tahosin M, Hasan MM, Islam T, Islam O, Rana MM (2023) Child and maternal mortality risk factor analysis using machine learning approaches. In: 2023 11th international symposium on digital forensics and security (ISDFS), Chattanooga, TN, USA, pp 1–6. https://doi.org/10.1109/ISDFS58141.2023.10131826
6. Md AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C (2023) Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines 11(2):581. https://doi.org/10.3390/BIOMEDICINES11020581
7. Behera MP, Sarangi A, Mishra D, Sarangi SK (2023) A hybrid machine learning algorithm for heart and liver disease prediction using modified particle swarm optimization with support vector machine. Procedia Comput Sci 218:818–827. https://doi.org/10.1016/j.procs.2023.01.062
8. Rahman S, Javed FM, Shamrat M, Tasnim Z, Roy J, Hossain SA (2019) A comparative study on liver disease prediction using supervised machine learning algorithms. Int J Sci Technol Res 8(11) [online]. Available www.ijstr.org
9. Md AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C (2023) Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines 11(2). https://doi.org/10.3390/biomedicines11020581
10. Dritsas E, Trigka M (2023) Supervised machine learning models for liver disease risk prediction. Computers 12(1):19. https://doi.org/10.3390/computers12010019
11. Khan B, Shukla PK, Ahirwar MK, Mishra M. Strategic analysis in prediction of liver disease using different classification algorithms, pp 437–449. https://doi.org/10.4018/978-1-7998-2742-9.ch022

12. Assegie TA (2021) Support vector machine and k-nearest neighbor based liver disease classification model. Indonesian J Electron Electromed Eng Med Inform 3(1):9–14. https://doi.org/10.35882/IJEEEMI.V3I1.2
13. Sivasangari A, Krishna Reddy BJ, Kiran A, Ajitha P (2020) Diagnosis of liver disease using machine learning models. In: Proceedings of the 4th international conference on IoT in social, mobile, analytics and cloud, ISMAC 2020, pp 627–630, Oct 2020. https://doi.org/10.1109/I-SMAC49090.2020.9243375
14. Blagus R, Lusa L (2017) Gradient boosting for high-dimensional prediction of rare events. Comput Stat Data Anal 113:19–37. https://doi.org/10.1016/J.CSDA.2016.07.016
15. Pal M (2007) Random forest classifier for remote sensing classification, vol 26, no 1, pp 217–222. https://doi.org/10.1080/01431160412331269698
16. Sharaff A, Gupta H (2019) Extra-tree classifier with metaheuristics approach for email classification. Adv Intell Syst Comput 924:189–197. https://doi.org/10.1007/978-981-13-6861-5_17/COVER
17. Zhang Z (2016) Introduction to machine learning: k-nearest neighbors. Ann Transl Med 4(11). https://doi.org/10.21037/ATM.2016.03.37
18. Chang YC, Chang KH, Wu GJ (2018) Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Appl Soft Comput 73:914–920. https://doi.org/10.1016/J.ASOC.2018.09.029
19. An TK, Kim MH (2010) A new diverse AdaBoost classifier. In: Proceedings—International conference on artificial intelligence and computational intelligence, vol 1. AICI 2010, pp 359–363. https://doi.org/10.1109/AICI.2010.82
20. Tehrany MS, Jones S, Shabani F, Martínez-Álvarez F, Tien Bui D (2019) A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. Theor Appl Climatol, 137(1–2):637–653. https://doi.org/10.1007/S00704-018-2628-9/METRICS
21. Postnikov EB, Esmedljaeva DA, Lavrova AI (2020) A CatBoost machine learning for prognosis of pathogen's drug resistance in pulmonary tuberculosis. In: LifeTech 2020—2020 IEEE 2nd global conference on life sciences and technologies, pp 86–87, Mar 2020. https://doi.org/10.1109/LIFETECH48969.2020.1570619054
22. Ramana B, Venkateswarlu N (2012) ILPD (Indian liver patient dataset). UCI Mach Learn Repos. https://doi.org/10.24432/C5D02C
23. Islam T, Hosen MA, Mony A, Hasan MT, Jahan I, Kundu A (2022) A proposed bi-LSTM method to fake news detection. In: 2022 international conference for advancement in technology (ICONAT). IEEE, pp 1–5