Chapter 9: Improving Hepatitis C Diagnosis Using Machine Learning Techniques: An Experimental Analysis

Author(s):

Md. Alif Sheakh (Daffodil International University, Dhaka, Bangladesh alif15-3292@diu.edu.bd orcid.org/0009-0006-5845-5210)

Mst. Sazia Tahosin (Daffodil International University, Dhaka, Bangladesh sazia15-3666@diu.edu.bd orcid.org/0009-0006-2027-7148)

* Taminul Islam (Daffodil International University, Dhaka, Bangladesh taminul15-11116@diu.edu.bd orcid.org/0000-0002-7803-4564)

Rishalatun Jannat Lima (Daffodil International University, Dhaka, Bangladesh rishalatun15-11120@diu.edu.bd orcid.org/0000-0003-2252-6284)

ABSTRACT

Hepatitis C is a major global health problem; the right diagnosis is critical to effective disease management. In recent years, machine learning techniques have shown promise in improving diagnostic accuracy in various medical applications. We aim to improve the diagnosis of hepatitis C by comprehensively analyzing several machine learning algorithms in this study. We compared and evaluated the classification accuracy, precision, and recall of 12 different models, including Random Forest, Gradient Boosting, K-Nearest Neighbors, Extreme Gradient Boost, Extra Trees, AdaBoost, LogitBoost, CatBoost, Support Vector Machine, Naive Bayes, Neural Network (Multilayer Perceptrons) and Gaussian process classifiers. Here, we train and test these machine learning models to determine the most effective way to classify hepatitis C diagnoses. We evaluate the algorithm's accuracy and compare our results with existing literature. The highest model accuracy in this study using LogitBoost was 98%, while Extra Trees achieved 99% accuracy after undersampling the data. This experimental analysis demonstrates the potential of machine learning techniques to improve hepatitis C diagnosis. The high accuracy of the LogitBoost and Extra Tress models highlights their effectiveness in identifying hepatitis C cases. By harnessing the power of machine learning algorithms, we can improve hepatitis C diagnostics, enabling early detection, timely intervention, and improved patient outcomes. The results of this study have major implications for the medical community and may improve the development of more accurate and effective hepatitis C diagnostic tools.

Keywords: hepatitis c, machine learning, hepatitis, liver, logitboost, extra tress, classification

9.1 Introduction

Hepatitis C causes a major global public health issue, changing a significant number of people on a worldwide level. Hepatitis C is a liver problem connected to the Hepatitis C virus [1]. The illness can appear as either acute or chronic. Individuals who catch the Hepatitis C virus and have been infected for a duration of less than six months suffer from a disease known as Acute Hepatitis C [2]. Usually, people don't show any symptoms [2]. Around 20% of the people who have been infected will properly cure the virus within a period of 6 months, which prevents any suffering consequences [3]. It is estimated that the remaining 80% of people infected with the Hepatitis C virus will progress to a chronic state of the disease [4]. Individuals infected will

continue to have the virus indefinitely, thereby maintaining their infectiousness and posing a potential risk to others. It is estimated that 71 million individuals approximately suffer from chronic HCV infection based on data provided by the World Health Organization [5]. Hepatitis C virus (HCV) has the potential to induce various liver ailments, such as chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (HCC) [6]. The precise and prompt identification of hepatitis C is essential for efficiently controlling and mitigating disease advancement. Traditional diagnostic techniques, including antibody tests and polymerase chain reaction (PCR) assays, have garnered extensive utilization [7]. However, these methods possess certain drawbacks about their sensitivity, specificity, and cost-efficiency.
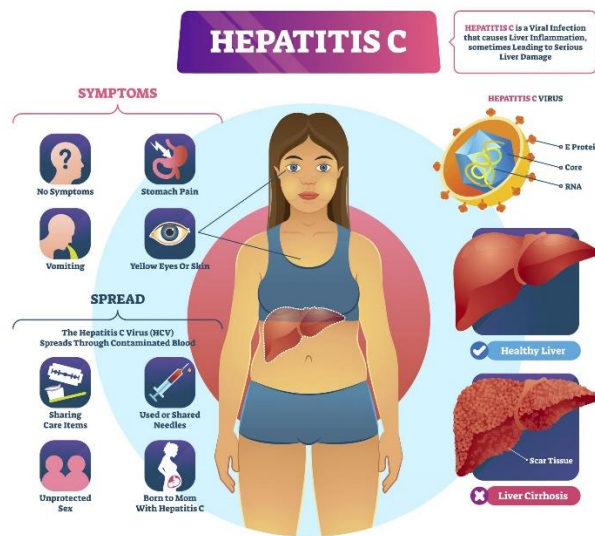

Figure 9.1: Overview of Hepatitis C

The majority of people infected with Hepatitis C have absent symptoms. The start of symptoms may occur within a period that varies from 14 days to 180 days following the initial infection [8]. Yet, it is possible that symptoms might begin for years after initial exposure, possibly suggesting that the liver damage has reached to an advanced stage. Currently, there is a lack of an effective vaccine to prevent Hepatitis C [9]. However, various strategies are present to reduce the possibility of developing the infection [10]:

- For drug use don't share needles, spoons, etc.
- Sharing personal items like razors, toothbrushes, etc. that may come into contact with infectious blood is not recommended.
- Condoms are required to be used during sexual activity.
- Visit authorized tattoo, acupuncture, and body piercing outlets.

Machine learning has become popular for healthcare research and has demonstrated promise in a number of medical fields, such as disease diagnosis [11]. Machine learning algorithms can immediately discover relationships and patterns in large datasets [12]. This allows the creation of predictive models to help detect diseases quickly and accurately. Machine learning models may boost hepatitis C diagnostic accuracy and patient care by analyzing complicated data patterns and combining multiple factors.

This study analyzes potential machine learning may enhance hepatitis C treatment. Our objective is to use our major scientific and clinical information to establish accurate models that can help identify hepatitis C infection efficiently. To be able achieve our research objective, we are going to perform thorough and methodical analysis of the experiment. The dataset of patients will be utilized in our evaluation. The data set provided will be used as a starting point for both training and validating our machine learning models. This study aims to utilize a range of machine learning algorithms, such as random forest [13], gradient boosting [14], k-nearest neighbors [15], extreme gradient boosting [16], extra tree [17], adaboost [18], logitboost [19], catboost [20], support vector machine [21], naïve bayes [22], neural networks (multi-layer perceptron) [23], and gaussian process [24] classifier models, to perform an analysis of the dataset and create predictive models for the diagnosis of hepatitis C. Besides all of these algorithms, CNN [25] is also applied in various medical problem-solving research. The models will be tested by analyzing accuracy, recall, and precision metrics.

The expected findings of the research are predicted to have major impacts for hepatology and how it's used in hospitals. The possible effect on creating reliable machine learning models for diagnosing hepatitis C is significant as it can significantly change existing diagnostic methodologies and supply clinicians with valuable guidance resources. Through applying elegant algorithms and an in-depth review of huge amounts of patient data, our objective is to create exact and strong models that can aid in timely recognition and diagnosis of health problems. Findings of this study possess an opportunity to make a valuable contribution to the progression of medical practices and to better worldwide management of hepatitis C.

9.2 Literature Review

In the research they carried out, Ma et al. [26] done an evaluation of six separate machine learning algorithms to try to predict hepatitis C. In particular, they stated a high accuracy rate of 95% when using support vector machines (SVM) and XGBoost algorithms. They employed the NHANES and UCI datasets, which include clinical data from a group of over 5,000 and 1,000 people diagnosed with hepatitis C, respectively. The main finding of the research active the assessment and contrast of different machine learning methods for predicting the onset of hepatitis C. Results of the learn found that Support Vector Machines (SVM) and XGBoost shown the highest degree of accuracy in the present situation.

Wang et al. [27] identified HBV and HCV status utilizing four machine learning models. Notably, their HCV status analysis displayed a 98.1% accuracy rate. The NHANES dataset, which consists of clinical data from over 5,000 adults in the US was used by the researchers. This study employed five algorithms based on machine learning to predict hepatitis C status. This study revealed that the K-Nearest Neighbors (KNN) model was incredibly accurate.

Ahammed et al. [28] employed feature selection and predictive machine learning algorithms to forecast hepatitis C. Their K-Nearest Neighbors (KNN) algorithm reached 83% accuracy. They evaluated 1,801 patients and 12 features. This work applied feature selection to enhance hepatitis C prediction machine learning models.

Based on clinical information, Abdelrahman et al. [29] constructed a structure for machine learning for identifying hepatitis C. Their architecture has a 94.88% accuracy rate using the Random Forest (RF) method. Clinical data from regarding 850 people suffering from hepatitis C have been utilized in the Egyptian HCV dataset. A machine learning conduct that accurately

foretold real-world hepatitis C cases was its primary contribution. The most significant contribution included a machine learning framework that can predict real-world hepatitis C infections.

Singh et al. [30] put forward a method of ensemble learning to predict hepatitis C disease, reaching a high accuracy of 95.69% employing the Ensemble Model. The scientists utilized the UCI dataset, which covers clinical data obtained from a group of over 1,000 people suffering from hepatitis C. The main work consisted of the creation of an ensemble learning structure that was able to increasing the accuracy of machine learning models in the field of hepatitis C estimation.

Priya et al. [31] evaluated machine learning approaches for identifying hepatitis C. The logistic regression (LR) methodology had 97.9% accuracy. Clinical data from more than 1000 hepatitis C patients were utilized in the UCI dataset. Evaluating machine learning techniques for hepatitis C prediction were the study's major contribution. The research found that the logistic regression (LR) approach was the most accurate.

To predict hepatitis C, Sahoo et al. [32] evaluated multiple and a binary labeling of a dataset. The AdaBoost algorithm provided the researchers 54.23% accuracy. Clinical data from over 1,000 hepatitis C patients were employed in the PROMISE dataset. This study examined the efficacy of multi-class and binary classifications to predict hepatitis C. The research conducted found that binary class labels appeared more accurate.

In their research, Syafa'ah et al. [33] done a comparison of four machine learning classification algorithms, specifically k-nearest neighbors, naïve Bayes, neural network, and random forest, to predict hepatitis C. A dataset that make up 1,200 patients identified as having hepatitis C was utilized in the research. The findings showed that the neural network displayed the greatest accuracy, specifically 95.12%.

Alotaibi et al. [34] utilized machine learning methods to identify the presence of liver disease for people suffering from hepatitis C. The researchers employed a dataset comprising 2038 Egyptian patients from the UCI Machine Learning Repository. Following that, they trained four different machine learning algorithms on this dataset, including Random Forest, Gradient Boosting Machine, Extreme Gradient Boosting, and Extra Trees model. The Extra Trees model shown better outcomes compared to the other models, getting an accuracy rate of 96.92%.

In their research, Abd El-Salam et al. [35] use machine learning approaches to project the likelihood of esophageal varices for humans diagnosed with chronic hepatitis C. A dataset making up 4962 patients from Egypt had been used, and six different machine learning algorithms have been used for learning purposes. The algorithms covered Neural Networks, Naïve Bayes, Decision Tree, Support Vector Machine, Random Forest, and Bayesian Network. The Bayesian Network algorithm displayed excellent results, with an accuracy rate of 74.8%.

In their study, Chen and Ji [36] created a customized machine learning structure to identify hepatitis C in particular patients. The model is trained utilizing patients' individual clinical data, helping it make predictions concerning the possibility of hepatitis C in a given patient. The model's efficiency was examined using a dataset with 615 patients, which gave an accuracy rate of 99%.

Table 9.1: Comparison with related works.

Ref	Contribution	Year	Dataset	Model	Accuracy
This Work	An experiment with twelve machine learning model to predict hepatitis C	2023	UCI	RF, KNN, GB, XGBoost, Extra Trees, AdaBoost, LogitBoost, CatBoost, SVM, NB, NN, GP	**99% (Extra Trees)**
[26]	Compared six machine learning algorithms for hepatitis C prediction	2023	NHANES and UCI	SVM, KNN, LR, DT, XGBoost, and ANN	95% (SVM, XGBoost)
[27]	Used four machine learning models to predict HBV and HCV status	2023	NHANES	KNN, SVM, LR, CART	98.1% (KNN)
[28]	Used a combination of feature selection and machine learning to predict hepatitis C	2023	Dataset of 1,801 patients with 12 features	KNN, LR, DT, RF, NN	83% (KNN)
[29]	Developed a machine learning framework to predict hepatitis C using clinical data	2023	Egyptian HCV dataset	NB, RF, KNN, LR	94.88% (RF)
[30]	Developed an ensemble learning model for predicting hepatitis C	2022	UCI	MLP, Bayesian Network, Quest, Ensemble	95.69% (Ensemble Model)
[31]	Compared the performance of different machine learning models for predicting hepatitis C	2022	UCI	k-NN, SVM, RF, NN, NB, LR	97.9% (LR)
[32]	Comparisons between multi and binary class labels of the same dataset	2020	PROMISE dataset	KNN, RF, SVM, GNB, NN, Bagging, AdaBoost	54.23% (AdaBoost)

9.3 Methodology

The proposed methodology for this research uses machine learning algorithms in the early diagnosis of Hepatitis C, which involves several phases, as shown in Figure 9.2.
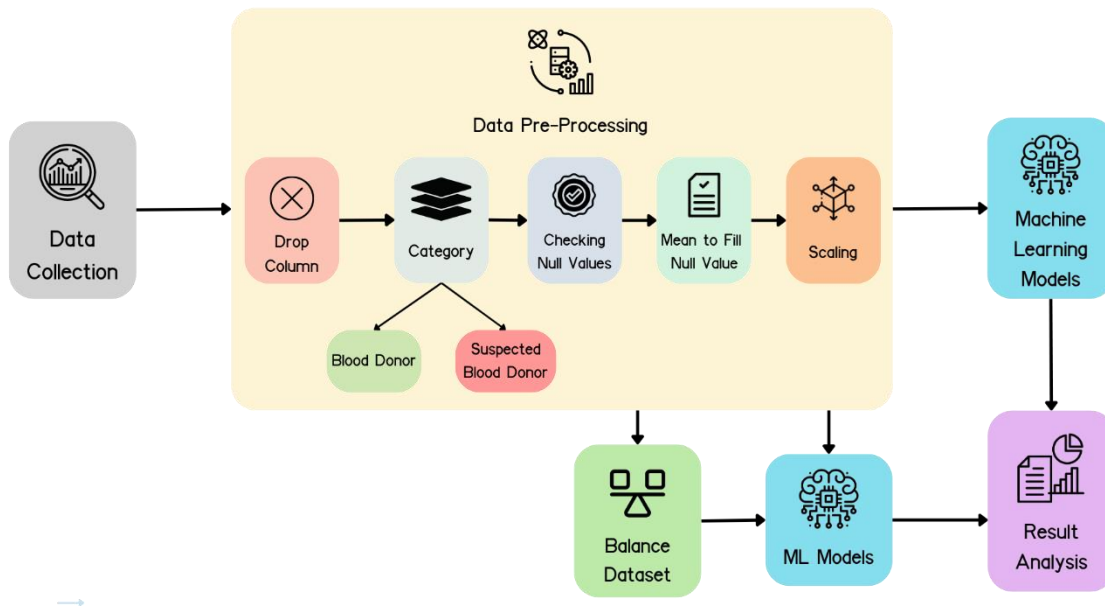
Figure 9.2: Proposed methodology

At the very beginning, a dataset that includes people suffering from Hepatitis C is collected. The dataset is then subjected to multiple data preprocessing methods. The data preparation procedures involve removing unnecessary columns, changing the "Category" column into a binary class representation (Blood Donor, Suspected Blood Donor), and encoding the "Sex" column as binary values (0 for Male and 1 for Female). The dataset contains rows with mean values that replace null values. The Robust Scaler is used to connect all attributes to a common scale. The dataset that received preprocessing is split into two subsets: training and testing sets. After that, a comprehensive set of twelve different machine learning models is applied, namely Random Forest [13], K-Nearest Neighbors [15], Gradient Boosting [14], Extreme Gradient Boost [16], Extra Trees [17], AdaBoost [18], LogitBoost [19], CatBoost [20], Support Vector Machine [21], Naive Bayes [22], Neural Network (Multilayer Perceptrons) [23], and Gaussian process [24] classifiers. The comparative evaluation of each model contains an evaluation of its classification accuracy, precision, and recall. The dataset is balanced by implementing underestimating techniques to reduce potential class imbalance issues. This entails the undersampling of the majority class to achieve a balanced distribution of samples with the minority class. Then, we evaluate the model by conducting multiple statistics, including accuracy, precision, recall, and confusion matrix [37]. The results of the research are subjected to evaluation and translation. This methodology offers an approach for predicting Hepatitis C and related liver diseases by implementing machine learning algorithms. This research aims to identify the most efficient model for early detection of Hepatitis C by evaluating various models and performance metrics.

9.3.1 Data Collection

We take this research's dataset from the UCI Machine Learning Repository [38]. The data set consists of 615 records of patients, 75 of whom are healthy and 540 uncertain of hepatitis C. The research sample contained patients across various regions. The dataset has been compiled

in Table 9.2. The dataset's target variable, classification class, and sex distribution are shown in Figures 9.3 and 9.4.

Table 9.2: Summary of the dataset.

Classes	Values
Category	Healthy Patients (0), (Suspected Patients (1)
Age	19 - 77
Sex	Male (0), Female (1)
Albumin (ALB)	14.9 - 82.2
Alkaline Phosphatase (ALP)	11.3 - 414.6
Alanine Aminotransferase (ALT)	0.9 - 325.3
Aspartate Aminotransferase (AST)	10.6 - 324
Bilirubin (BIL)	0.8 - 254
Cholinesterase (CHE)	1.42 – 16.41
Cholesterol (CHOL)	1.43 – 9.67
Creatinine (CREA)	8 – 1079.1
Gamma-Glutamyl Transferase (GGT)	4.5 – 650.9
Protein (PROT)	44.8 - 90
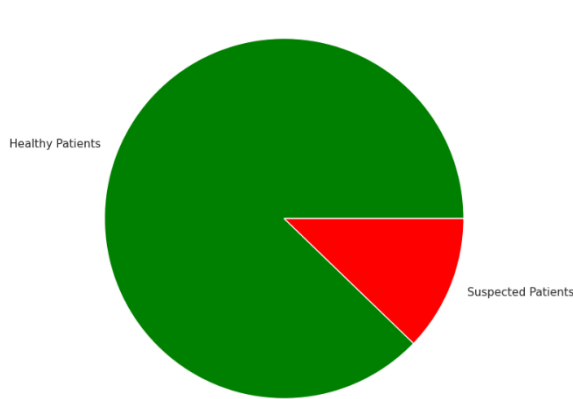
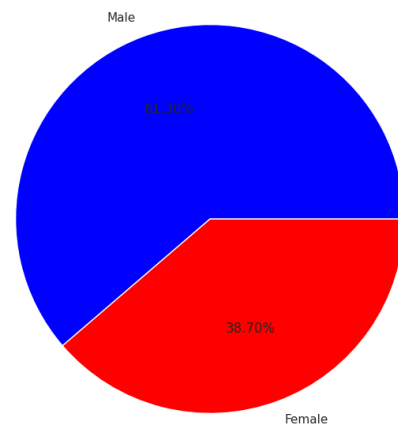**Figure 9.3: Distribution of Category class.**

Figure 9.4: Distribution of Sex class.

9.3.2 Data Preprocessing

To simplify the process, we remove unnecessary columns from the dataset that do not assist with the classification task. In the "Category" column, there were multiple categories identified, including "Blood Donor," "Hepatitis," "Fibrosis," and "Cirrhosis." The classes were transformed into binary classes, with "Blood Donor" being categorized as "Healthy Patients"

and all other three classes being classified as "Suspected Patients". Similarly, the "Sex" column, first denoted as "Male" and "Female", is encoded using binary values (0 and 1), where 0 means "Male" and 1 means "Female". After that, we check the dataset to see if any null value exists or not, after finding they are handled through using of the mean value related to the related feature.
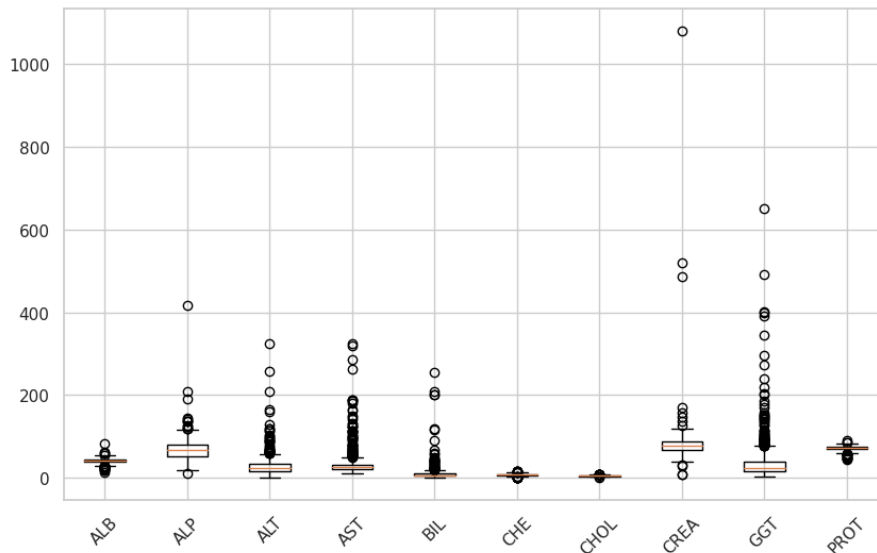

Figure 9.5: Box plot for each column

Some of the columns contain outliers, as shown in Figure 9.5. The robust scaler scales data because it is less dependent on dataset outliers. Figure 9.6 shows age data from our dataset. In Figure 9.7, we use a boxplot to explore the variation in age within different groups to identify the number of healthy and insufficient patients.
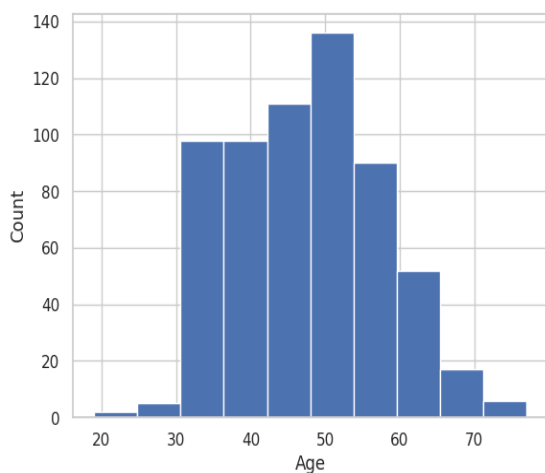
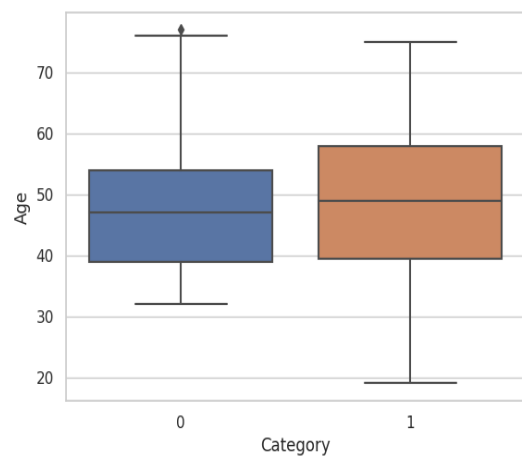

Figure 9.6: Visualization of Age Category

Figure 9.7: Visualization of Age with Category

In Figure 9.8, a scatter plot illustrates the relationship between Aspartate Aminotransferase (AST) and the variable type within the dataset. After plotting the data, we can observe the range of aspartate aminotransferase (AST) levels across both categories of hepatitis C.
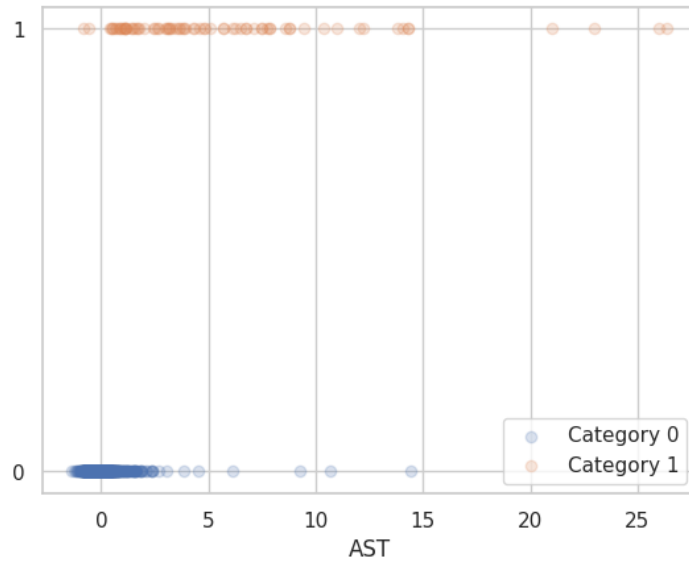

Figure 9.8: Scatter plot of Category separately

The dataset is processed to identify the correlation between all columns, shown in Figure 9.9. Analyzing a dataset's variables' connection and interdependence allows us to understand.
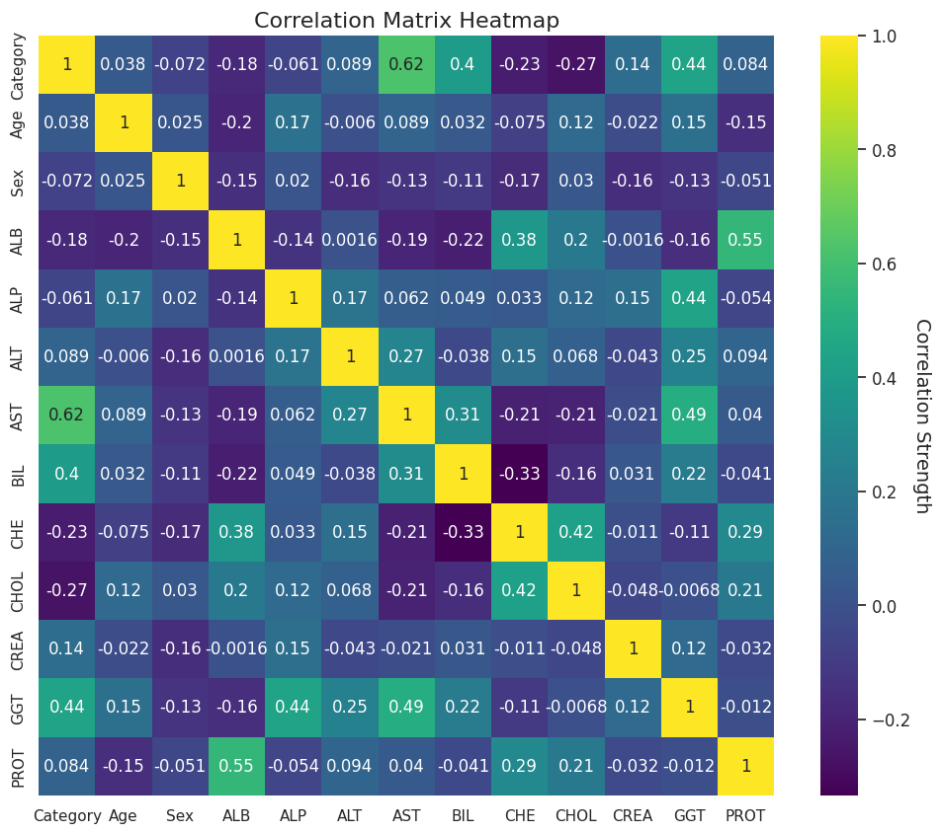

Figure 9.9: Correlation between all columns

In the end, in Figure 9.10, our focus moves toward a graph of the distribution of various blood test results through both groups. Histogram plots are used for visualization in this context. This method makes identifying major differences or overlaps in category values easier.
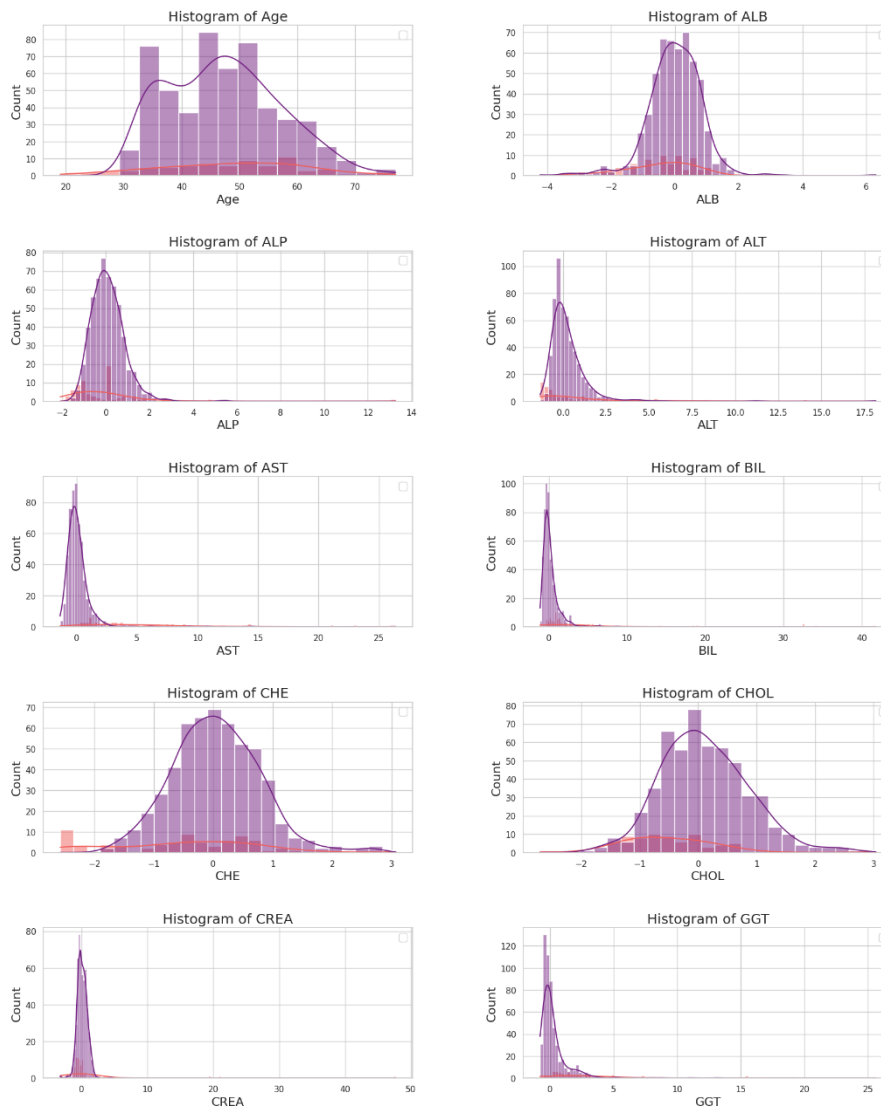

Figure 9.10: Distribution of various blood test Results by Category

After completing each of the previous procedures, we move on to implement a total of 12 machine learning models. The dataset was split into an 80:20 ratio for training and testing. As shown in Figure 3, it is clear that our dataset displays an imbalance. In order to achieve a balanced dataset, the undersampling technique is employed. Undersampling is a method that reduces the number of samples in the majority class in line with the number of samples in the minority class [39]. This process aims to create a more equal distribution of data for training machine learning models. This approach minimizes potential bias to the dominant class and improves the overall accuracy of classification. Next, we execute a comparative analysis of the results obtained from the machine learning models by evaluating the accuracy, recall, and precision metrics for both the before and after balanced datasets.

9.4 Experimental Result

The tables in this research show the results of the machine learning algorithms used to detect Hepatitis C.

The results of the machine learning algorithms on the imbalanced dataset are shown in Table 3. Before performing the balancing process, the LogitBoost algorithm displayed a level of accuracy reaching 0.98. Additionally, it demonstrated a precision of 0.99 and a recall of 0.92. The Extreme Gradient Boost (XGBoost) algorithm showed outstanding results, achieving an accuracy of 0.97, precision of 0.99, and recall of 0.83. More algorithms, namely Gradient Boosting, AdaBoost, and Random Forest displayed important levels of accuracy and precision.

Table 9.3: Before balancing the dataset the result of machine learning algorithms

Algorithms	Accuracy	Precision	Recall
Random Forest	0.93	0.99	0.62
K-Nearest Neighbors	0.84	0.83	0.21
Gradient Boosting	0.94	0.99	0.71
Extreme Gradient Boost	0.97	0.99	0.83
Extra Trees	0.91	0.93	0.58
AdaBoost	0.95	0.99	0.75
LogitBoost	**0.98**	**0.99**	**0.92**
CatBoost	0.94	0.99	0.71
Support Vector Machine	0.84	0.99	0.17
Naive Bayes	0.88	0.76	0.54
Neural Network (Multilayer Perceptrons)	0.93	0.99	0.67
Gaussian process	0.85	0.88	0.29

Table 9.4 shows what was determined after dataset balancing. The Extra Trees algorithm got 0.99 precision, recall, and accuracy. The Random Forest approach performed higher with an accuracy of 0.97, precision of 0.93, and recall of 0.99. The Support Vector Machine algorithm performed low compared to other algorithms, having an accuracy rate of 0.83 and a recall rate of 0.64.

Table 9.4: After balancing the dataset the result of machine learning algorithms

Algorithms	Accuracy	Precision	Recall
Random Forest	0.97	0.93	0.99
K-Nearest Neighbors	0.90	0.99	0.79
Gradient Boosting	0.90	0.87	0.93
Extreme Gradient Boost	0.90	0.87	0.93

Extra Trees	**0.99**	**0.99**	**0.99**
AdaBoost	0.93	0.88	0.99
LogitBoost	0.90	0.87	0.93
CatBoost	0.93	0.93	0.93
Support Vector Machine	0.83	0.99	0.64
Naive Bayes	0.87	0.86	0.86
Neural Network (Multilayer Perceptrons)	0.97	0.93	0.99
Gaussian process	0.87	0.92	0.79

Figures 9.11, 9.12, and 9.13 present a graphical representation showing the contrasting evaluation of the algorithms' accuracy, precision, and recall metrics, each before and after following the dataset being balanced. The given figures represent the effect of balancing on the performance metrics, highlighting the observed enhancements after the execution of balancing methods.
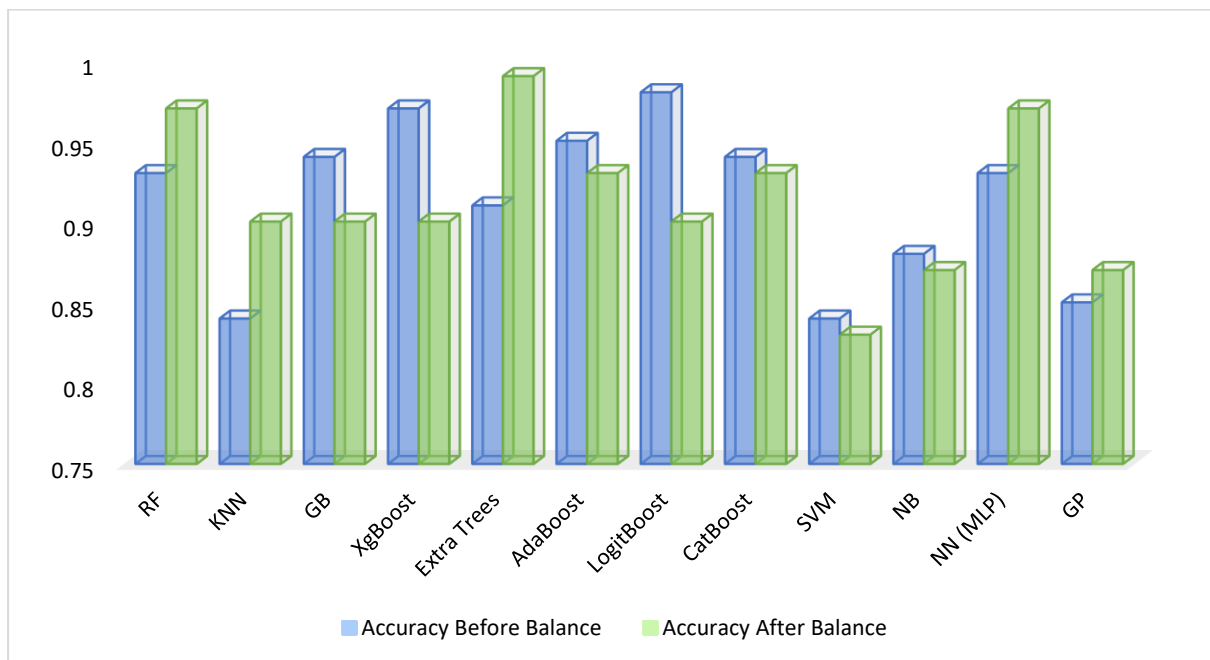

Figure 9.11: Comparison between accuracy before and after balancing the dataset
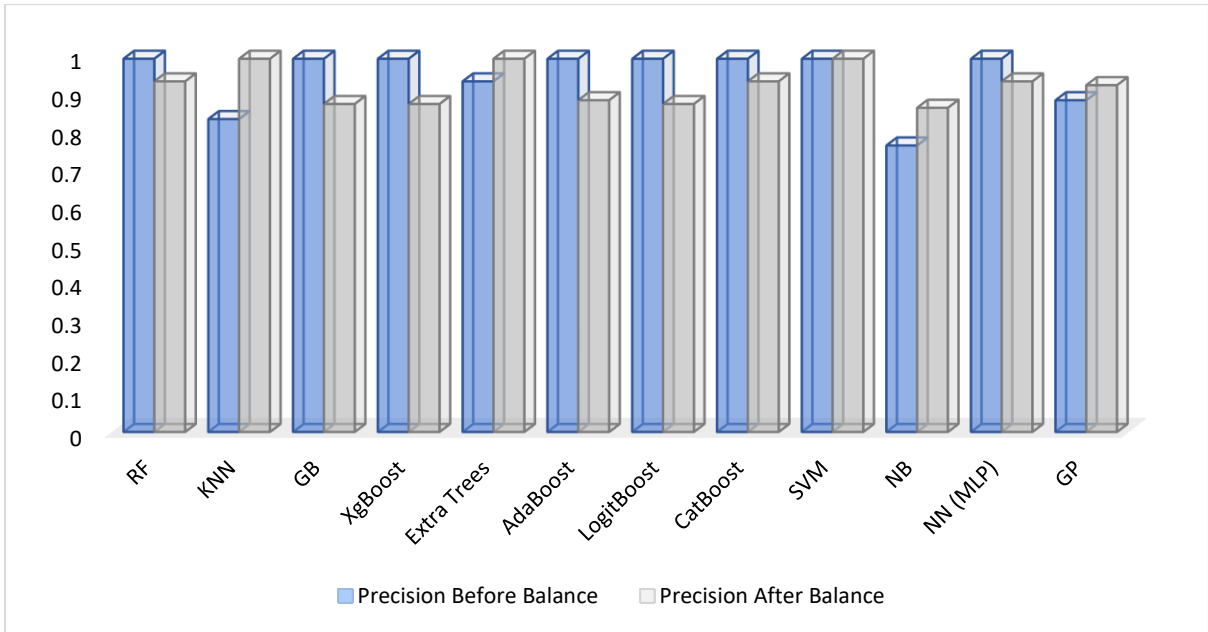
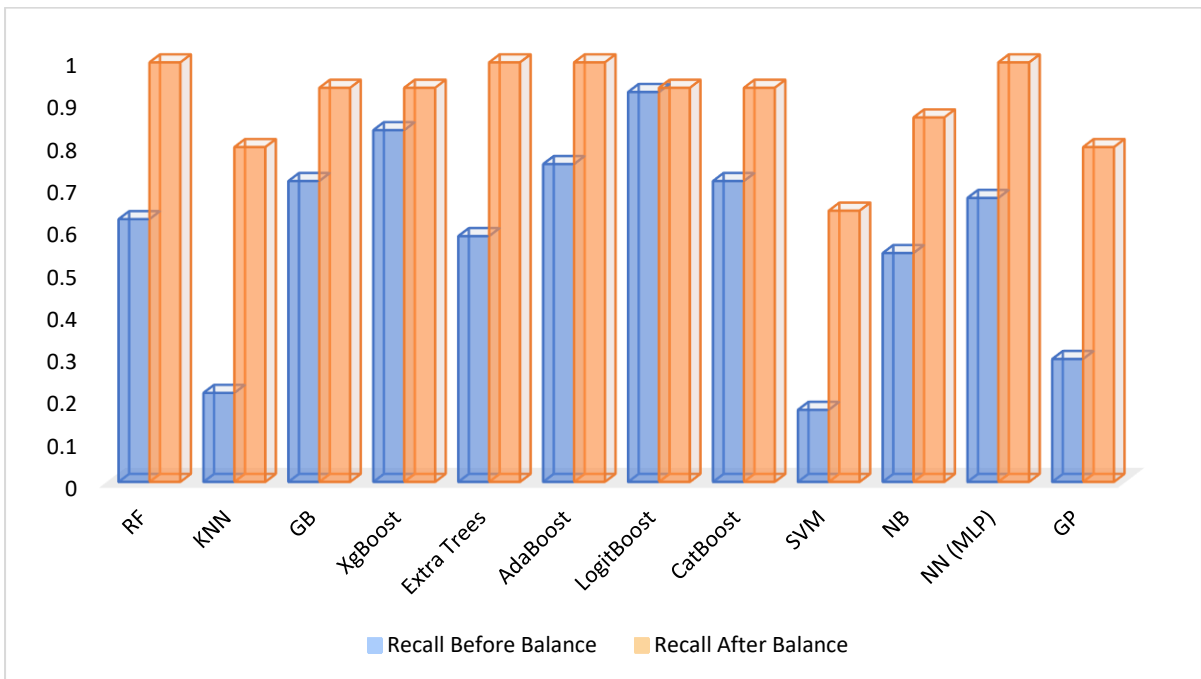**Figure 9.12: Comparison between precision before and after balancing the dataset.**

Figure 9.13: Comparison between recall before and after balancing the dataset

In addition, give the confusion matrices both before and after to the dataset has been balanced. The matrices provided offer valuable insights into the classification performance of the algorithms, displaying the accurate identification of true positives and true negatives and the incorrect identification of false positives and false negatives.
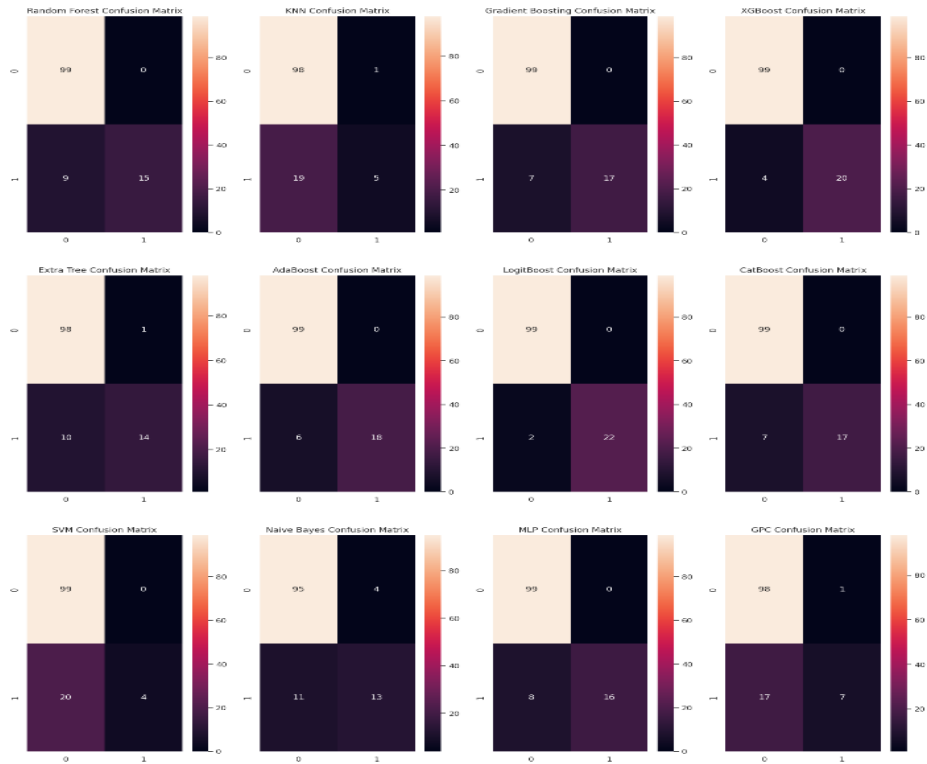
Figure 9.14: Before balancing the confusion matrix
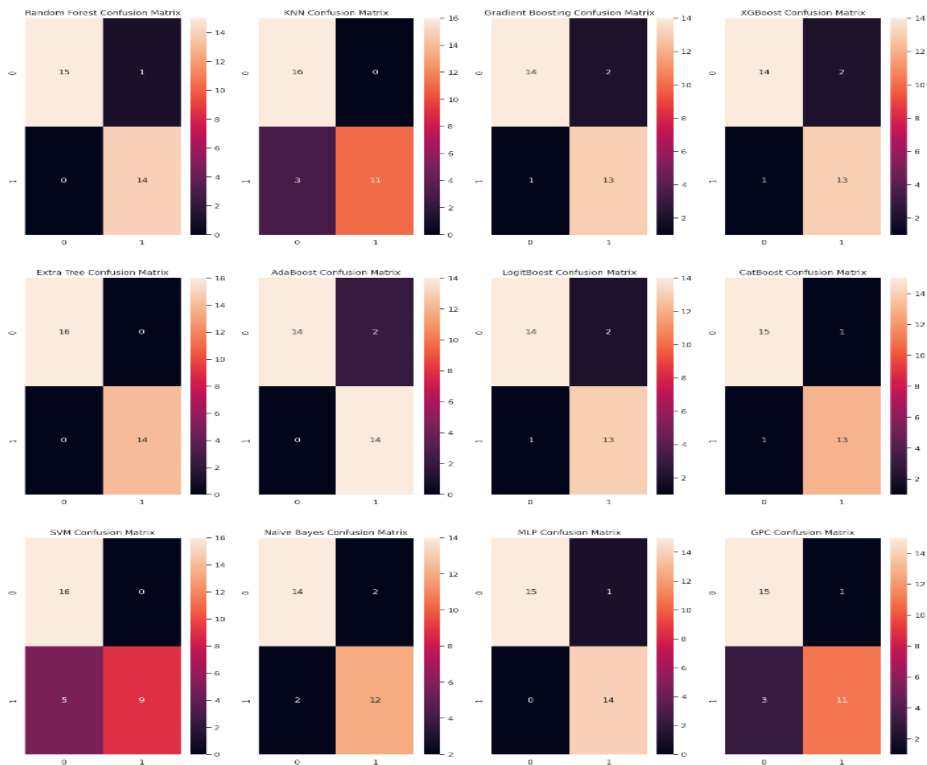

Figure 9.15: After balancing the confusion matrix

In general, the experiment results show that the performance of different algorithms used for machine learning was specifically improved by balancing the dataset. The use of a balanced dataset caused raised accuracy, precision, and recall metrics across multiple algorithms, therefore showing enhanced abilities in identifying Hepatitis C.

9.5 Conclusion & Future Work

The present study utilized machine learning algorithms to accurately diagnose hepatitis C. We examined and evaluated twelve different models' classification accuracy, precision, and recall. By analyzing relevant scholarly papers and comparing them with other algorithms, we determined the most effective methods for diagnosing hepatitis C. The empirical results demonstrated that machine learning significantly enhances the accuracy of hepatitis C diagnosis. Specifically, the Extra Trees model achieved an impressive accuracy rate of 99% after implementing undersampling techniques. These remarkable results unequivocally prove that these models can diagnose hepatitis C accurately. It is crucial to note that employing machine learning algorithms in medicine offers several advantages for detecting hepatitis C. The high accuracy levels exhibited by these models enables early detection. Prompt intervention and ultimately improve patient outcomes. Moreover, given the global health threat posed by hepatitis C, utilizing advanced computational methods to enhance illness diagnostics becomes imperative. Therefore, this study contributes substantially to the growing research on machine learning in healthcare. The authors emphasize the significance of advanced computational methods for precise disease diagnosis. Furthermore, these findings underscore the importance of continuous research and development in improving the accuracy and efficacy of diagnostic technologies for hepatitis C. Overall, this study demonstrates that machine learning techniques can effectively improve diagnostic accuracy for hepatitis C. In further study attempts, investigating novel features and data sources, such as genetic data and patient demographics, could yield significant findings for developing more comprehensive diagnostic models. Furthermore, implementing more in-depth study efforts, including diverse and representative datasets, can yield a better understanding of the versatility of models and their practical significance in real-world scenarios.

References

[1]   H. A, M. L, J. AM, M. M, A. GP, and V. S, "Community-Based Assessment and Treatment of Hepatitis C Virus-Related Liver Disease, Injecting Drug and Alcohol Use Amongst People Who Are Homeless: A Systematic Review and Meta-Analysis," *International Journal of Drug Policy*, vol. 96, p. 103342, Oct. 2021, doi: 10.1016/J.DRUGPO.2021.103342.

[2]   S. Frankova, N. Uzlova, D. Merta, V. Pitova, and J. Sperl, "Predictors of Significant Liver Fibrosis in People with Chronic Hepatitis C Who Inject Drugs in the Czech Republic," *Life 2023, Vol. 13, Page 932*, vol. 13, no. 4, p. 932, Apr. 2023, doi: 10.3390/LIFE13040932.

[3]   R. Ahmed *et al.*, "Sofosbuvir/Velpatasvir - A Promising Treatment for Chronic Hepatitis C Virus Infection," *Cureus*, vol. 13, no. 8, Aug. 2021, doi: 10.7759/CUREUS.17237.

[4]   M. P. Manns and B. Maasoumy, "Breakthroughs in hepatitis C research: from discovery to cure," *Nat Rev Gastroenterol Hepatol*, vol. 19, no. 8, pp. 533–550, Aug. 2022, doi: 10.1038/s41575-022-00608-8.

[5]   J. F. H. Eijsink, M. N. M. T. Al Khayat, C. Boersma, P. G. J. ter Horst, J. C. Wilschut, and M. J. Postma, "Cost-effectiveness of hepatitis C virus screening, and subsequent monitoring or treatment among pregnant women in the Netherlands," *The European*

Journal of Health Economics, vol. 22, no. 1, pp. 75–88, Feb. 2021, doi: 10.1007/s10198-020-01236-2.

[6]   S. J. Park and Y. S. Hahn, "Hepatocytes infected with hepatitis C virus change immunological features in the liver microenvironment," *Clin Mol Hepatol*, vol. 29, no. 1, p. 65, Jan. 2023, doi: 10.3350/CMH.2022.0032.

[7]   I. Gow, N. C. Smith, D. Stark, and J. Ellis, "Laboratory diagnostics for human Leishmania infections: a polymerase chain reaction-focussed review of detection and identification methods," *Parasites & Vectors 2022 15:1*, vol. 15, no. 1, pp. 1–24, Nov. 2022, doi: 10.1186/S13071-022-05524-Z.

[8]   M. W. Tenforde *et al.*, "Protection of Messenger RNA Vaccines Against Hospitalized Coronavirus Disease 2019 in Adults Over the First Year Following Authorization in the United States," *Clin Infect Dis*, vol. 76, no. 3, pp. e460–e468, Feb. 2023, doi: 10.1093/CID/CIAC381.

[9]   L. Parlati, C. Hollande, and S. Pol, "Treatment of hepatitis C virus infection," *Clin Res Hepatol Gastroenterol*, vol. 45, no. 4, p. 101578, Jul. 2021, doi: 10.1016/J.CLINRE.2020.11.008.

[10]   C. R. Balsom, A. Farrell, and D. V. Kelly, "Barriers and enablers to testing for hepatitis C virus infection in people who inject drugs - a scoping review of the qualitative evidence," *BMC Public Health*, vol. 23, no. 1, p. 1038, Dec. 2023, doi: 10.1186/S12889-023-16017-8/TABLES/3.

[11]   T. Islam, A. Kundu, N. Islam Khan, C. Chandra Bonik, F. Akter, and M. Jihadul Islam, "Machine Learning Approaches to Predict Breast Cancer: Bangladesh Perspective," *Smart Innovation, Systems and Technologies*, vol. 302, pp. 291–305, 2022, doi: 10.1007/978-981-19-2541-2_23/COVER.

[12]   Md. A. Sheakh, Mst. Sazia Tahosin, M. M. Hasan, T. Islam, O. Islam, and M. M. Rana, "Child and Maternal Mortality Risk Factor Analysis Using Machine Learning Approaches," in *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*, IEEE, May 2023, pp. 1–6. doi: 10.1109/ISDFS58141.2023.10131826.

[13]   T. A. N. Saputra, K. I. Arizona, M. R. Andrian, F. I. Kurniadi, and B. Juarto, "Random Forest in Detecting Hepatitis C," *Proceedings - 2022 9th International Conference on Information Technology, Computer and Electrical Engineering, ICITACEE 2022*, pp. 299–302, 2022, doi: 10.1109/ICITACEE55701.2022.9924074.

[14]   H. Park *et al.*, "Machine learning algorithms for predicting direct-acting antiviral treatment failure in chronic hepatitis C: An HCV-TARGET analysis," *Hepatology*, vol. 76, no. 2, pp. 483–491, Aug. 2022, doi: 10.1002/HEP.32347.

[15]   S. Setianingsih, M. U. Chasanah, Y. I. Kurniawan, and L. Afuan, "IMPLEMENTATION OF PARTICLE SWARM OPTIMIZATION IN K-NEAREST NEIGHBOR ALGORITHM AS OPTIMIZATION HEPATITIS C CLASSIFICATION," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 2, pp. 457–465, Apr. 2023, doi: 10.52436/1.JUTIF.2023.4.2.980.

[16] N. Ali, D. Srivastava, A. Tiwari, A. Pandey, A. K. Pandey, and A. Sahu, "Predicting Life Expectancy of Hepatitis B Patients using Machine Learning," *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2022*, 2022, doi: 10.1109/ICDCECE53908.2022.9793025.

[17] X. Liu, L. Wang, C. H. Liang, Y. P. Lu, T. Yang, and X. Zhang, "An enhanced methodology for predicting protein-protein interactions between human and hepatitis C virus via ensemble learning algorithms," *J Biomol Struct Dyn*, vol. 40, no. 21, pp. 10592–10602, 2022, doi: 10.1080/07391102.2021.1946429/SUPPL_FILE/TBSD_A_1946429_SM8044.DOCX.

[18] D. A. Jadhav, "An enhanced and secured predictive model of Ada-Boost and Random-Forest techniques in HCV detections," *Mater Today Proc*, vol. 51, pp. 186–195, Jan. 2022, doi: 10.1016/J.MATPR.2021.05.071.

[19] C. Geetha and S. Maruthuperumal, "Prediction of Liver Cirrhosis using Ensemble Machine Learning Algorithms," *2022 1st International Conference on Computer, Power and Communications, ICCPC 2022 - Proceedings*, pp. 453–457, 2022, doi: 10.1109/ICCPC55978.2022.10072150.

[20] J. Jangiti, C. G. Paluri, S. Vadlamani, and S. K. Jindal, "Hepatitis C Severity Prognosis: A Machine Learning Approach," *Journal of Electrical Engineering and Technology*, vol. 18, no. 4, pp. 3253–3264, Feb. 2023, doi: 10.1007/S42835-023-01441-Y/METRICS.

[21] K. Keyvan, M. R. Sohrabi, and F. Motiee, "An intelligent method based on feed-forward artificial neural network and least square support vector machine for the simultaneous spectrophotometric estimation of anti hepatitis C virus drugs in pharmaceutical formulation and biological fluid," *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 263, p. 120190, Dec. 2021, doi: 10.1016/J.SAA.2021.120190.

[22] R. Safdari, A. Deghatipour, M. Gholamzadeh, and K. Maghooli, "Applying data mining techniques to classify patients with suspected hepatitis C virus infection," *Intelligent Medicine*, vol. 2, no. 4, pp. 193–198, Nov. 2022, doi: 10.1016/J.IMED.2021.12.003/ASSET/5C896865-8D61-41F6-A794-CDDEF0C58309/ASSETS/GRAPHIC/2096-9376-02-04-002-F003.PNG.

[23] Ç. Suiçmez, C. Yılmaz, H. T. Kahraman, E. Cengiz, and A. Suiçmez, "Prediction of Hepatitis C Disease with Different Machine Learning and Data Mining Technique," pp. 375–398, 2023, doi: 10.1007/978-3-031-09753-9_27.

[24] L. A. Escamilla, O. Akarsu, E. Di Valentino, and J. A. Vazquez, "Model-independent reconstruction of the Interacting Dark Energy Kernel: Binned and Gaussian process," May 2023, Accessed: Jul. 08, 2023. [Online]. Available: https://arxiv.org/abs/2305.16290v1

[25] Md. T. Islam, T. Ahmed, A. B. M. Raihanur Rashid, T. Islam, Md. S. Rahman, and Md. Tarek Habib, "Convolutional Neural Network Based Partial Face Detection," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2022, pp. 1–6. doi: 10.1109/I2CT54291.2022.9825259.

[26] A. Alizargar, Y. L. Chang, and T. H. Tan, "Performance Comparison of Machine Learning Approaches on Hepatitis C Prediction Employing Data Mining Techniques," *Bioengineering*, vol. 10, no. 4, p. 481, Apr. 2023, doi: 10.3390/BIOENGINEERING10040481/S1.

[27] V. Harabor *et al.*, "Machine Learning Approaches for the Prediction of Hepatitis B and C Seropositivity," *International Journal of Environmental Research and Public Health 2023, Vol. 20, Page 2380*, vol. 20, no. 3, p. 2380, Jan. 2023, doi: 10.3390/IJERPH20032380.

[28] D. Zhang *et al.*, "Explainable Machine Learning Approach for Hepatitis C Diagnosis Using SFS Feature Selection," *Machines 2023, Vol. 11, Page 391*, vol. 11, no. 3, p. 391, Mar. 2023, doi: 10.3390/MACHINES11030391.

[29] H. Mamdouh Farghaly, M. Y. Shams, and T. Abd El-Hafeez, "Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt," *Knowl Inf Syst*, vol. 65, no. 6, pp. 2595–2617, Jun. 2023, doi: 10.1007/S10115-023-01851-4/TABLES/7.

[30] M. O. Edeh *et al.*, "Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease," *Front Public Health*, vol. 10, p. 892371, Apr. 2022, doi: 10.3389/FPUBH.2022.892371/BIBTEX.

[31] P. L. Septina and J. I. Sihotang, "A Comparative Study on Hepatitis C Predictions Using Machine Learning Algorithms," *8ISC Proceedings: Technology*, pp. 33–42, Feb. 2022, Accessed: Jul. 04, 2023. [Online]. Available: http://ejournal.unklab.ac.id/index.php/8ISCTE/article/view/684

[32] S. C. Nandipati, C. XinYing, and K. K. Wah, "Hepatitis C Virus (HCV) Prediction by Machine Learning Techniques," *Applications of Modelling and Simulation*, vol. 4, no. 0, pp. 89–100, Mar. 2020, Accessed: Jul. 04, 2023. [Online]. Available: http://arqiipubl.com/ojs/index.php/AMS_Journal/article/view/122

[33] L. Syafaâ€™ah, Z. Zulfatman, I. Pakaya, and M. Lestandy, "Comparison of Machine Learning Classification Methods in Hepatitis C Virus," *Jurnal Online Informatika*, vol. 6, no. 1, pp. 73–78, Jun. 2021, doi: 10.15575/JOIN.V6I1.719.

[34] A. Alotaibi *et al.*, "Explainable Ensemble-Based Machine Learning Models for Detecting the Presence of Cirrhosis in Hepatitis C Patients," *Computation 2023, Vol. 11, Page 104*, vol. 11, no. 6, p. 104, May 2023, doi: 10.3390/COMPUTATION11060104.

[35] S. M. Abd El-Salam *et al.*, "Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients," *Inform Med Unlocked*, vol. 17, p. 100267, Jan. 2019, doi: 10.1016/J.IMU.2019.100267.

[36] L. Chen, P. Ji, and Y. Ma, "Machine Learning Model for Hepatitis C Diagnosis Customized to Each Patient," *IEEE Access*, vol. 10, pp. 106655–106672, 2022, doi: 10.1109/ACCESS.2022.3210347.

[37] T. Islam *et al.*, *Review Analysis of Ride-Sharing Applications Using Machine Learning Approaches : Bangladesh Perspective*. New York: CRC Press, 2023. doi: 10.1201/9781003253051-7.

[38] "HCV   data   -   UCI   Machine   Learning   Repository." https://archive.ics.uci.edu/dataset/571/hcv+data (accessed Jul. 08, 2023).

[39] A. Guzmán-Ponce, J. S. Sánchez, R. M. Valdovinos, and J. R. Marcial-Romero, "DBIG-US: A two-stage under-sampling algorithm to face the class imbalance problem," *Expert Syst Appl*, vol. 168, p. 114301, Apr. 2021, doi: 10.1016/J.ESWA.2020.114301.