Hepatitis C Virus Diagnosis and Prognosis Using Machine Learning Algorithms

NAME: P. Pavan **H.T.NO**:2203A51685

Under the Guidence of

Mohammed Ali Shaik

Assist.Professor

Abstract: The Hepatitis C Virus (HCV) is among the most virulent health challenges that persists today as millions and millions of new cases are reported every year. In the early stages, the affected individual may not show any signs or symptoms. Nevertheless, longterm complications may include cirrhosis and hepatocellular carcinoma which are liver late chronic conditions. Succinct diagnosis and correct forecast are the core things to have good disease management, as capable progression of it could lead to severe liver illnesses, including cirrhosis and hepatocellular carcinoma. In this study, ML algorithms are used to diagnose and prognose HCV cases with a goal of identifying the most accurate model so as to arrive at crucial decisions concerning chronic CV. Leveraging the dataset, ML techniques such as K-Nearest Neighbor (KNN), Cat Boost ,Support Vector Machine (SVM), Gaussian Naïve Bayes (NB), Random Forest (RF), Gradient Boost, Multi-Layer Perceptron and Logistic Regression (LR) are implemented and evaluated for their predictive accuracy and error rates. The study aims to identify the most effective ML model for HCV diagnosis by assessing performance metrics like accuracy, precision, recall, F1-score, Support and Error rates. The findings of this research contribute in developing robust predictive tools for early detection and management of HCV, ultimately providing better results for patients and healthcare resource utilization.

Keywords: Hepatitis C Virus (HCV); Machine Learning (ML); HCV Diagnosis; Healthcare management.

Introduction

Hepatitis C virus (HCV) disease is a major health issue across the globe; statistics show that more than 71 million people are chronically infected and suffering from significant liver morbidity and mortality [1]. Hepatitis C is a virus that mainly affects the liver and circulates in the blood. It consequently causes chronic inflammation, fibrosis, cirrhosis, and hepatocellular carcinoma when it is left untreated. Unlike Hepatitis B and D viruses, Hepatitis C is an insidious infection wherein patients may remain asymptomatic for many years thus making it more difficult to detect, diagnose and appropriately manage the disease [3].

The early treatment plan for the patients at risk of disease progression is of the uttermost importance, that if determined timely, it can help them to avoid critical outcomes by the application of personalized care and tailored treatments. Hepatitis C management decision making by clinicians is determined through a multitude of parameters which include patients' demographic profiles, laboratory test results, viral genotype, liver fibrosis, and comorbidities [5]. Nevertheless, the fact for the health providers is that it is often difficult in practice to accurately judge disease progression trajectories and estimate the risk profiles of individual patients, as the variety of their data is complex and diverse.

In this regard, ML methods have come into as a potent instrument to predict models and the risk stratification in the healthcare for the past few years leveraging data from large datasets via pattern recognition algorithms which result in the reveal of titbits necessary for the

decision making [5]. ML algorithms like this have the ability that based on the historical patient data they can identify the pertinent patterns and associations and therefore can be used for the purposes of disease prognosis, treatment response and clinical outcomes prediction.

This paper works on the prevention of hepatitis C spread by predicting progression of the disease with the help of ML techniques and providing the healthcare specialists with the patient's categorization founded on the progress rate of the disease, that consequently enhances treatment program accuracy. The model to be proposed is a complete one so it uses an anonymized data set of patient records including demographic information, lab test results, medical history, and treatment outcomes. A variety of supervised ML algorithms, such as Decision trees, Cat Boost, Random forests, Support vector machines, Gradient boosting and logistic regression performing the hepatitis C progression estimation are studied and tested for their performance.

For selecting the best ML algorithm for the ML-based modelling problem, an evaluation of each algorithm's performance by using various metrics like accuracy, sensitivity, and specificity is performed. Through finding out the difference between the error rates and performance metrics of the algorithms on the dataset, we strive to define the optimal model that can well estimate the hepatitis C progression. The empirical process used here reflects our confidence that the selected algorithm is not only able to provide accurate forecasts but also generalized well to unknown data which, in return, increases its trustworthiness and usability in the in-field practice.

Literature Review

The role of predictive modelling in hepatitis C progression has been the subject several studies in the medical literature concerning risk factors, prognostic markers, and predictive outcome models of the disease. The commonly used statistical methods, including logistic regression and survival analysis, have already had a wide range of practical applications in hepatitis C-related liver complication diagnosis and prediction [7]. Nevertheless, these approaches hinge on too many simplifications and they might not manifest the entire trajectory and dynamics of the disease process progress.

In the past few years, uses of this type of ML in the health research increased because the methods are able to process the large volume, complicated data and develops optimal models [8]. In several studies using ML models have been applied to build customized models for hepatitis C prognosis, treatment response, and disease progression [9]. The research made it clear that ML methodologies can help to rise the exactness and the reliability of prognostic modelling in hepatitis C treatment.

In a recent study on hepatitis C virus (HCV) prediction using machine learning, Satish CR Nandipathi et al. [T3] used the dataset of 1385 instances with 29 attributes, including multiclass labels for baseline histological staging and binary class labels for fibrosis levels. The study to compared seven machine learning methods performance, which yielded 54.5% accuracy for the RF binary class label in both Python and R tools. A total of 12 key features

that exhibited similar performance to the full datasets, enabling HCV detection are identified by feature selection methods. The study also noted some common symptoms of hepatitis C in the selected features, suggesting the need for further examination of symptom-based prognostic models

In a study done by Fergie Joanda Kaunang, a dataset consisting of 589 instances with 13 features was analyzed by using Support Vector Machine, Random Forest and Logistic Regression methods that yielded an accuracy of 88.3% using Logistic Regression algorithm. The study featured data visualization components, where the relationship between attributes was understood and possible machine learning was presented in disease prediction and risk factor identification that was non-invasive or invasive. These points reveal the useful perspective on the problems that current healthcare system suffers from and explore ways to improve the efficiency of the given process.

A study by Ashfaq Ali Kashif (Rashid University) concerned the classification methods, carried out for the non-linear relationship between the drug reaction and severity of the disease. It used various classifiers such as the K Nearest Neighbour, k*, Naive Bayes, Random Forest. On evaluating the performance of these classifiers the KNN, k* performed best in terms of the training accuracy, and the Decision Tree performed better with testing dataset accuracy. Such findings draw a big line under agreement on the relevance of the correct response to treatment and on the possibility to apply machine learning methods in the medical sphere to improve the patient health.

Mamdouh Farghaly, Heba (2021), analyzed a dataset of 859 patient records kept in the National Liver Institute of Egypt. The data was related with Hepatitis C Virus (HCV) infection prevalence among healthcare workers and the data included 12 features. Implemented multiple machine learning classifiers which had accuracy in the range of 83.01% to 89,88 %. Data preprocessing techniques were adopted to ensure that dataset is up to date. Such framework included data preprocessing, splitting, feature selection, machine learning classifiers, and then performance evaluation, the eventual purpose being to help in detection of HCV amongst healthcare workers at the early instance. Besides that, information collected during the research reveals ML techniques applicability for the prediction and managing of chronic Hepatitis C infections in healthcare institutions.

The researchers of Duraid Y. Mohammed also conducted work on hepatitis disease diagnosis using machine learning techniques and found that models that do not follow a fixed sequence, such as non-linear iterative partial least squares (NLPALS) in combination with clustering-based approach like the self-organizing map (SOM), accomplished high classification accuracies. The work of Ahmadi et al. has been able to show a good classification rate which are 83.06% and the results of Sartakhti and Mozafari demonstrated great classification rates, that is, 86.25%. Doyle et al presented the Models and they compared them to a dataset of HCV patients through the Stacked Ensemble model which achieved the highest precision of 86%. From all the algorithms in this study, the one that shows the best result is Logistics Regression, with 86.9% accuracy and then the Random Forest algorithm that shows the most result of 87.3% accuracy. The purpose of the research is to mark out the best approaches to enhance the early diagnosis of hepatitis and other

common diseases in Iraq, stressing on a holistic end-to-end research approach for health service improvement.

References	Dataset Utilized	No. of instances	Method	Performance Metrics (Accuracy)
[1]	UCI machine learning repository dataset	1385	RF SVM NB NN KNN	50.34% 50.34% 49.52% 48.43% 50.70%
[2]	UCI machine learning repository Dataset	615	KNN SVM NN LR NB	86.5% 87.1% 87.8% 88.3% 85.9%
[3]	Patient's dataset from a hospital in city of Lahore, Pakistan	512	K* DT KNN MLP RF SVM	78.16% 85.91% 78.169% 79.57% 82.74% 84.15%
[4]	HCV dataset from the NLI institute, founded at Menoufiya University (Menoufiya, Egypt)	859	RF KNN LR NB	89.88% 83.75% 83.01% 84.08%
[5]	UCI repository of machine learning databases	155	RF DT SVM	86.9% 84.3% 82.2%

Table 1:Summary of the Literature Review of hepatitis C prediction

Methodology

Dataset:

In this study, a real-world multivariate dataset originating from the UCI Machine Learning Repository is used which offers valuable insights for Hepatitis C diagnosis research. The dataset contains a total of 615 instances along with 14 attributes. It integrates demographic data like age and gender with various blood test results for a collection of individuals. The target attribute for classification is Category which presumably signifies the patient's status,

distinguishing healthy blood donors from those with Hepatitis C. Table 2 shows the information about attributes of the dataset.

Attribute Name	Type	Description	Missing Values
ID	Integer	ID of a particular Patient	no
Age	Integer	Age of the Patient	no
Sex	Binary	Patient Gender (Male or Female)	no
ALB	Float	Albumin levels (*low albumin levels gives a sign for liver or kidney disease*)	yes
ALP	Float	Alkaline phosphatase levels (*higher ALP levels may indicate a damage/disease in liver*)	yes
AST	Float	Aspartate aminotransferase levels (*higher AST levels may be a sign of conditions like hepatitis, cirrhosis, or other liver diseases*)	yes
BIL	Float	Bilirubin levels (*high bilirubin levels may mean your liver is not working right*)	no
CHE	Float	Cholinesterase levels (*high level CHE may indicate nephrotic syndrome, hyperthyroidism, fatty liver etc..*)	no
CHOL	Float	Cholesterol level (*low cholesterol levels may be observed in individuals with advanced liver disease, including those with chronic hepatitis C*)	yes
CREA	Float	Creatinine levels (*low creatinine may result in inflamed liver*)	no
CGT	Float	Gamma-glutamyl transferase levels (*High GGT levels in blood may be a sign of liver disease or liver damage*)	no
PROT	Float	Protein levels (*low protein level persons may have a liver or kidney problem*)	yes
Category	Categorical	The diagnosis (*values: '0=Blood Donor', '0s=suspect Blood Donor', '1=Hepatitis', '2=Fibrosis', '3=Cirrhosis'*)	no
ALT	Float	Alanine transaminase levels (*The higher levels in ALT are an indicator of serious liver disease*)	no

Table 2: Information about the attributes of dataset

Data Processing and Transformation:

After looking into the dataset, in the 615 instances, a total of 31 values are missing. These missing cells in the dataset are filled with the median of their respective attribute. The attribute category is having five different types of categories, and to make things easier, this attribute is then divided into two groups, one which determines non-hepatitis (blood

donors, suspect blood donors) and the other which determines hepatitis (hepatitis, fibrosis, cirrhosis).

Proposed Model:

This Study aims to propose a suitable model in diagnosing of Hepatitis C. The framework of the proposed model is shown in Fig 1. After selecting a hepatitis C dataset, a pre-processed data is obtained using data cleaning, missing values completion, balancing and normalization. Then data are trained through nine different models. After testing the data through different evaluation metrics, a best model/classification is chosen to diagnosis the patients.

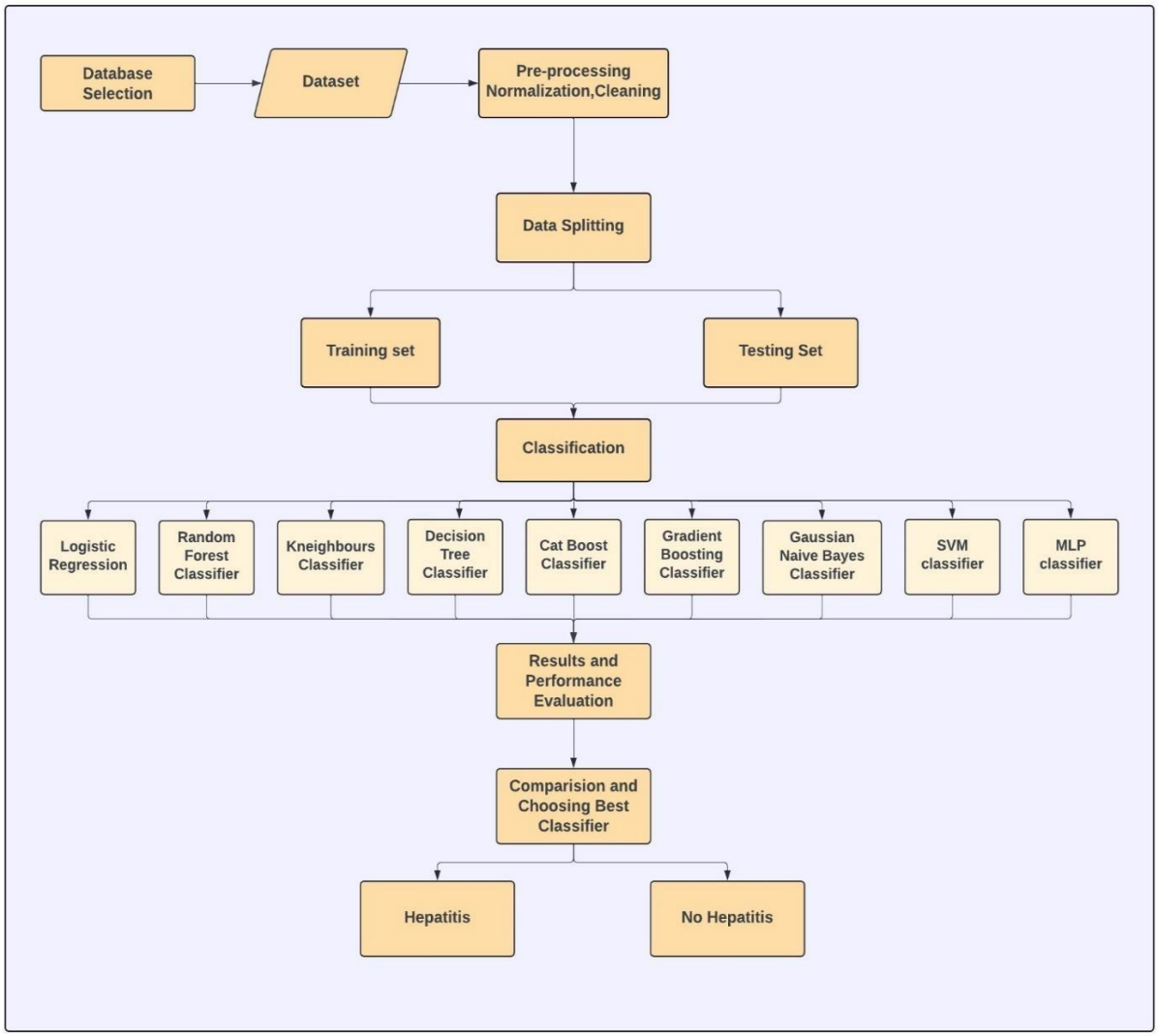

Figure 1: Diagnosis of Hepatitis C using Machine Learning algorithms

Classification Models:

Logistic Regression: Logistic groups two entities by finding the possibility of a certain event, occurrence or observation. For LR, predictors are mapped along with their likelihoods with the help of sigmoid function. A Sigmoid function maps the real value to a range from 0 to 1

using the shape of an S curve. Because of its simplicity and straightforward interpretation, logistic is essential for hepatitis C diagnostics. It helps doctors find vulnerable individuals and act quickly to improve patient care and results.

Decision Tree: A decision tree is widely accepted and used as means of supervised nonparametric machine learning where recursive partitioning of input variables is applied to allocate data into various subsets. This is a decision tree type Hepatitis C forecasting model which will provide an insight into what factors do have the most significant influence upon the disease course and develop a hierarchical structure as a fragmentation tool for the examples. Thus, by applying the method of horizontal splitting of data set on the features that are more integral to the cause Hepatitis C one can deduce the reasons behind the spread.

Random Forest: Random forest (RF) is an advanced plural cluster learning approach which is commonly applied for classification and regression problems. The model works by the construction of various decision trees during the training. Here, random samples of the training data are selected, after which a number of random subsets are employed. This technique yields flexibility in the model, thus, significantly minimizing the risk of overfitting on the surface which consequently increases the model strength. RF as a non-parametric and nonlinear machine learning technique is widely used for complex data sets in the fields of finance, healthcare and ecology among others. Besides this, this advantageous algorithm generally comes up with the important feature scoring and sometimes restrictions. Facilitates Relevant Interpretation The RF‟s strength of getting to the variables that are actually important, the high accuracy with which it picks them up and the interpretability that it has as a machine learning model makes it a very handy tool in machine learning research and applications.

K-Nearest Neighbors: KNN, is among the simplest supervised ML technique where it finds use in both, classification and regression. Concerning Hepatitis C progression, KNN works by locating the K closest data points that corresponds with given input feature and classifying it based on the majority among neighboring K points. Taking this route implies missing an opportunity to find the patterns in the patient data by identifying correlations and simulating unknown cases with the help of the similar ones, when dealing with a non-linear or intricate decision boundaries. Even though it looks simple to implement, KNN's is able to provide better even with larger datasets. Nevertheless, it remains a valuable tool in healthcare for making personalized treatment decisions and assessing risks in the management of Hepatitis C.

CatBoost: CatBoost Classifier, an advanced algorithm developed by Yandex’s researchers, stands out for its proficiency in classification tasks. It handles categorical variables efficiently, eliminating the need for manual preprocessing steps such as one-hot encoding. The unique methods for handling categorical features and missing data contribute to its high accuracy and training speed. With robust hyperparameter tuning support and scalability over large

data sets, CatBoost is a formidable asset for predictive models in finance, healthcare, and ecommerce, representing a major breakthrough in machine learning

Gradient Boosting: Gradient boosting is an effective cluster learning technique which is able to combine a pair of small Decision trees, to create a robust prediction model. In ML, gradient boosting works by adding new images sequentially to correct errors in the previous fashion, where each new image highlights a memory or error caused by existing sets so is so. By optimizing default loss characteristics through gradient descent, gradient boosting can better cope with complex relationships in the data and capture nonlinear patterns The main strength of gradient boosting lies in its ability to generate strength of males or females easy startup processes and blends them scientifically to generate incredibly accurate models Management of multiple data sets and predictive functions is widely incorporated in many packages including health, economics and herbal language applications because of due to its complexity and versatility

Gaussian Naive Bayes: Gaussian Naive Bayes Rooted in Bayes' theorem is a classification algorithm, especially effective for continuous data and can be widely used in various areas like information classification, medical research, spam filtering etc. Calculating probabilities of class membership based on feature distributions within each class it works. A Gaussian distribution is used to model processes. Despite its simplicity, Gaussian Naive Bayes exhibits strong performance, especially with large data sets, due to its speed and efficiency. Its ease of use and ability to handle high-dimensional data make it a popular choice for real-world segmentation projects.

SVM: SVM is a powerful supervise algorithm used in distribution, regressive functions. Its main goal is to find the optimal hyper-plane. At its best, it separates the data point from class in high-dimensional space. At its core, the goal of SVM is to find decision boundaries that maximize the mean difference Lessons learned. This decision boundary is defined by a hyperplane in a feature space, where distance b/w the hyperplane, each class, and neighboring data points, is known support vectors. The main power of the SVM is its ability to do the high-quality data and be efficient Overloading is prevented, especially where quantity exceeds quantity of the specimens. Furthermore, it is not significantly affected by the presence of redundant features. Its focus on supporting vectors, which are critical data points that influence decision making boundaries.

MLP: MLP classification is an important feature of artificial neural network methods, and excels in numerous tasks of classification. Its structure of layered connected nodes enables the identification of complex relationships in data. With the ability to use various generating functions such as ReLU and sigmoid, MLP captures micropatterns efficiently. Reducing prediction errors through iterative training methods such as backpropagation and gradient descent to improve classification accuracy In modern applications in domains such as image recognition and economic analysis, MLP classification has a major part in development of machine learning and pattern recognition

Results and Comparisons

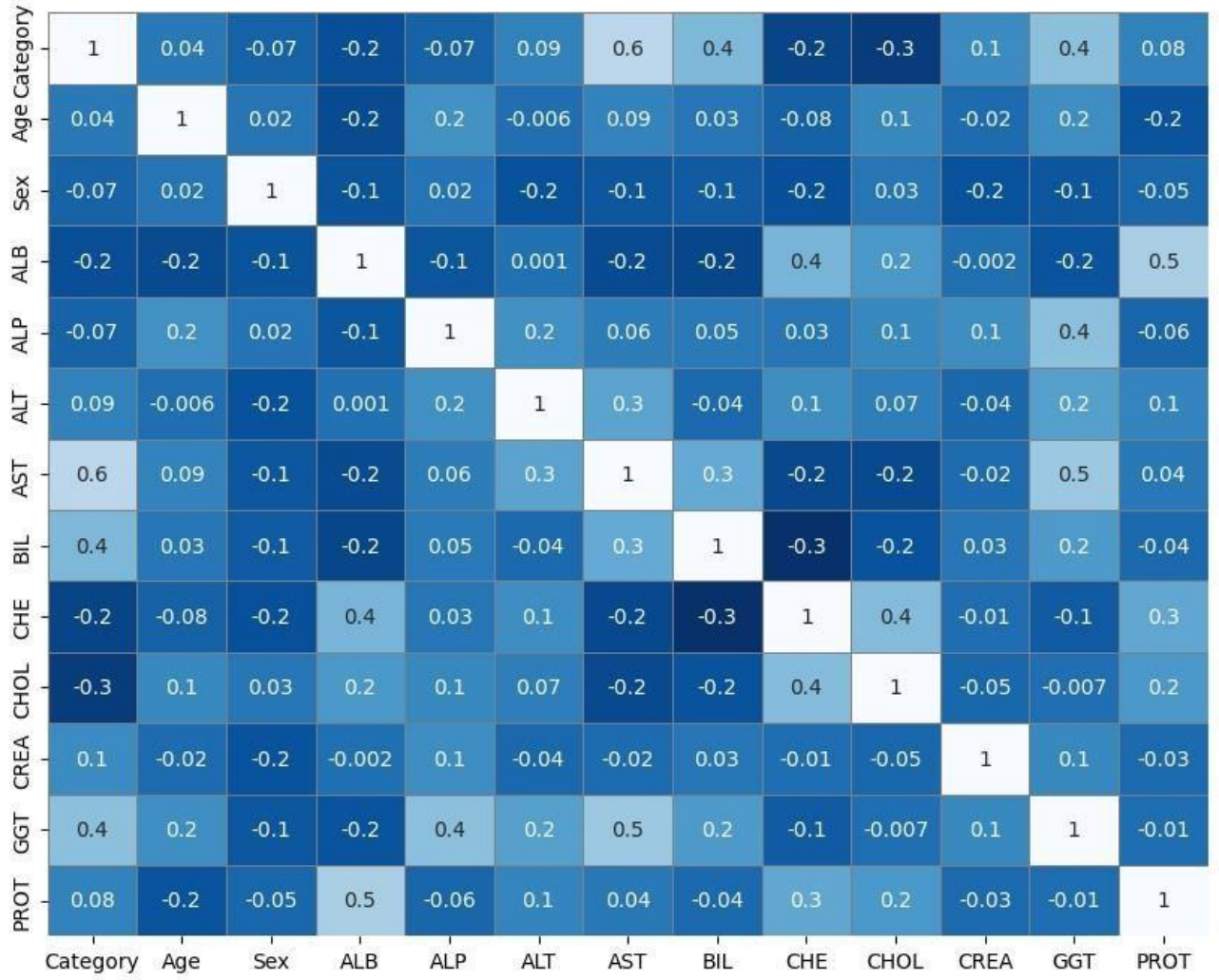

Figure 2:Correlation Matrix for hepatitis c dataset

This represents the correlation matrix between the various features of the hepatitis c dataset. To have fast and easy interpretation, features and corelation coefficients have been colored. The features include category, CHE, Sex, AST, ALB, PROT, ALP, CREA, Age, ALT, GGT, BIL, CHOL.

Here the intensity of the color represents the strength of the correlation. The color varies from dark blue to white. Here Dark Blue indicates a strong negative correlation between two features. As one variable increases, the other one decrease significantly.
White represents a correlation coefficient close to 1 which indicates having a strong positive correlation between two features. As the one variable increase or decrease, the other also does the same. Here variables increase or decrease together. The color light blue suggests a weak or no correlation between the features.

Classification Report:

Performance evaluation of ML algorithms in the context of Hepatitis C prediction when a real patient is involved needs to involve precision, recall, and accuracy as metrics. Via these metrics, the models' ability to discern patients who might progress with the disease can be

empirically confirmed. This technique offers a significant advantage over manual methods due to its increased efficiency. It is possible by this way researchers are able to determine any algorithms which show how well it is the patient at risk of being identified, therefore help a lot in model selection for prediction tasks. The table 3 shows results of the different metrics used to predict hepatitis c.

Classifier	Precision	Recall	F1-Score	Support	Accuracy
Linear Regression	0.89	0.98	0.93	99	88.61
Decision Tree	0.93	0.97	0.95	99	92.68
Random Forest	0.92	0.99	0.95	99	92.68
K Nearest Neighbors	0.85	1	0.92	99	85.36
Cat Boost	0.91	1	0.95	99	91.87
Gradient Boosting	0.93	1	0.97	99	94.31
Gaussian Naïve Bayes	0.90	0.96	0.93	99	87.8
Support Vector Machine	0.92	0.97	0.95	99	91.06
Multilayer perceptron	0.91	1	0.95	99	91.87

Table 3:Results of Evaluation Metrics

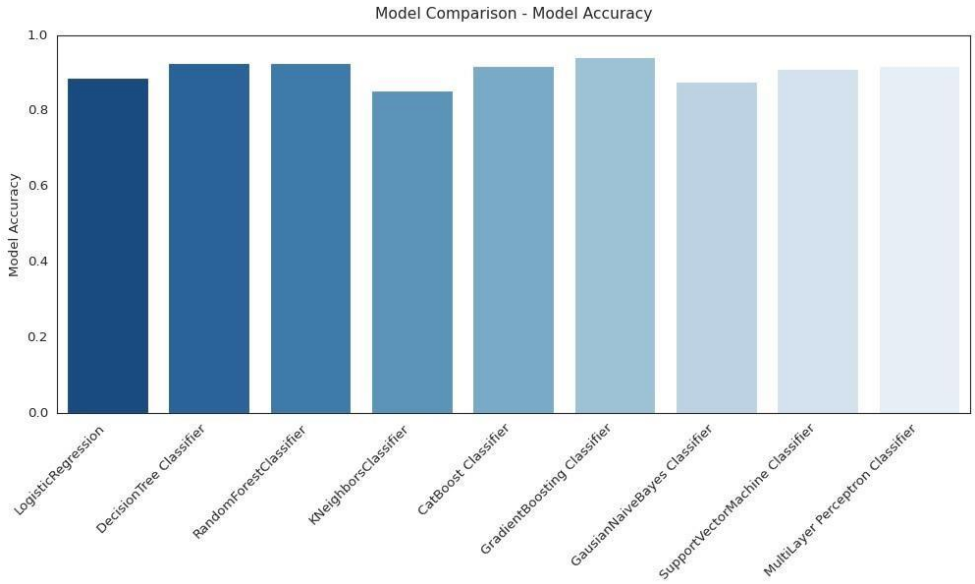

Figure 3:Comaparison Graph of accuracies of all classifiers

To have better view and analysis over, a comparison bar graph is generated by plotting all the classifiers accuracies. We have all the Classifiers producing different accuracies. By comparing the accuracy results we have Gradient boosting with 94.7%, giving the highest accuracy of all the classifiers. An accuracy of 92.68%-Random Forest, 92.68%-Decision Tree, 91.87%-MLP, 91.87%-CatBoost,91.06%-SVM,87.8%-Gaussian NB, 88.61%-LR,85.36%- KNN is produced.

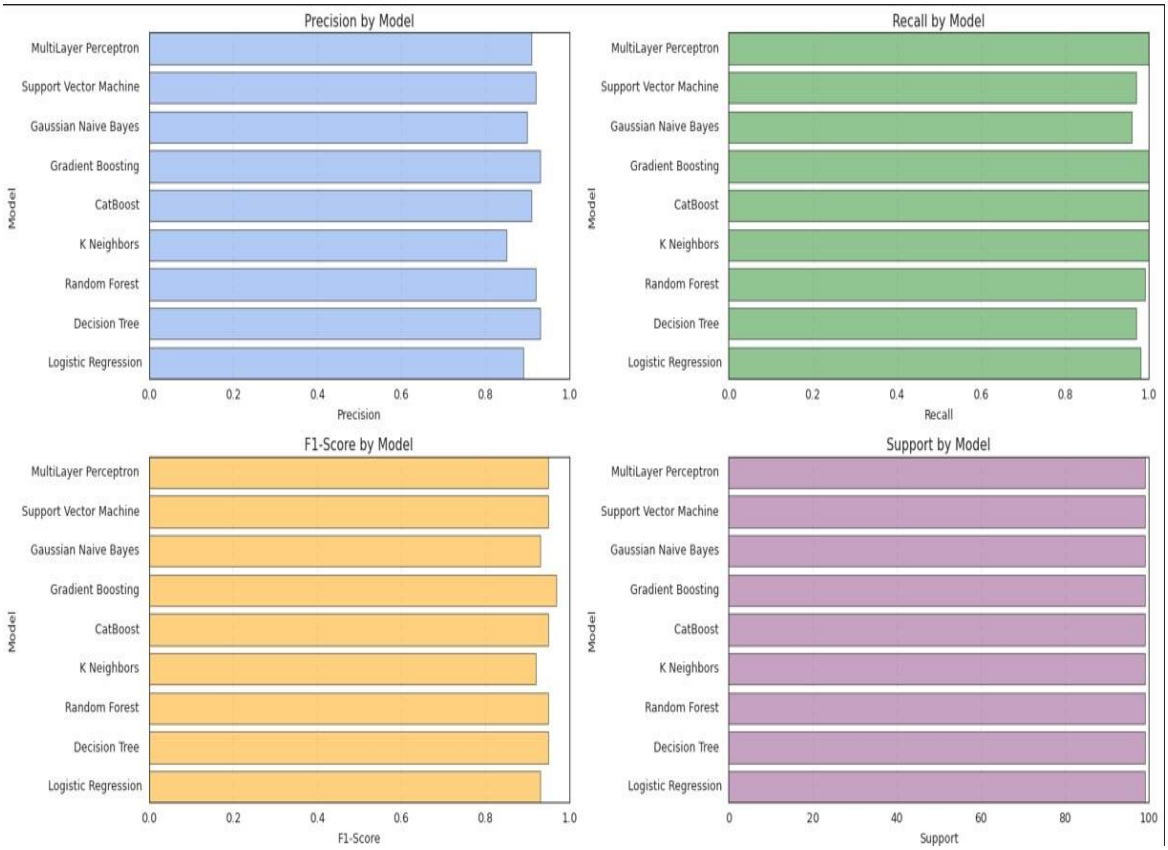

Figure 4:Comarision Graphs of Precision, Recall, F1-score and Support

These plotted graphs precision, recall, f1-score, and support help to provide a better, understandable study about difference of values in those measures. To get a deeper understanding of trade-offs between these metrics, the analysis of these metrics is done. In predicting Hepatitis C progression, models with a balance of high precision, recall, and accuracy are preferred as they demonstrate overall strong performance. By highlighting areas where adjustments are needed, these metrics guide the fine-tuning and improvement of models. The Recall, precision, F1-score and support are calculated by using the below formulas:

$$\text{Precision} = \frac{TruePositive}{TruePositive+FalsePoitve}$$

$$\text{F-Measures} = \frac{Precision*Recall*2}{(Precision+Recall)}$$

$$\text{Recall} = \frac{TruePositive}{TruePositive+FalseNegativies}$$

$$\text{Accuracy} = \frac{TruePositivess+TrueNegativies}{TruePositives+TrueNegatives+FalsePositives+FalseNegatives}$$

Error Rates:

The performance of ML algorithms should be evaluated by narrowing in on error rates such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Logarithmic Error (MALE), when hepatitis C progression is being predicted. These metrics enable the detection of the anomalies in models through the comparison of the forecasts and the actual outcomes. Because of this, researchers can evaluate quantitatively one model against another and identify specific algorithmic attributes and strong and weak points, leading to the selection of the best algorithm among them and finally refining of the software in hepatitis C prognosis and diagnosis support.

Classifier	MAE	MAPE	RMSE	MALE
Linear Regression	0.1138	50.00%	0.3374	0.0547
Decision Tree	0.0732	33.33%	0.2705	0.0352
Random Forest	0. 0732	37.50%	0.2705	0.0352
K Nearest Neighbors	0.1463	75.00%	0.3825	0.0703
Cat Boost	0.0813	41.67%	0.2851	0.0391
Gradient Boosting	0.0569	29.17%	0.2386	0.0273
Gaussian Naïve Bayes	0.122	45.83%	0..3492	0.0586
Support Vector Machine	0.1707	87.50%	0.4132	0.082
Multilayer perceptron	0.0813	41.67%	0.2851	0.0391

Table 4:Results of Error Rates

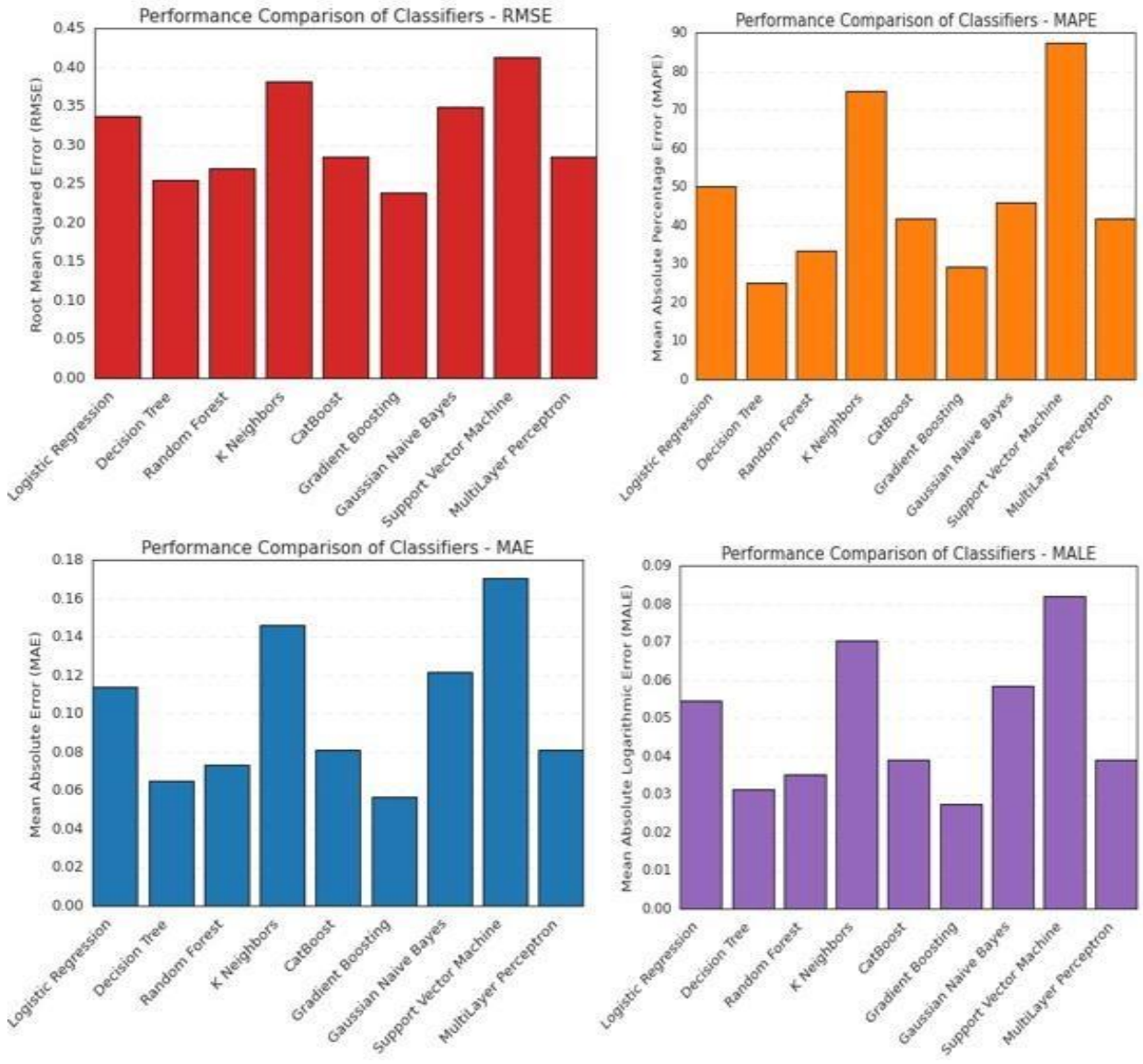

Figure 5:Comparison graphs of MAE, MAPE, MALE and RMSE

The plotted graphs of error rates can guide for a better analysis of the results obtained through accuracy. These helps in choosing a better and most accurate algorithm. By seeing different rates for all the algorithms, the gradient boosting is producing with least error rates in all the cases.

Conclusion

A comparative study on the ML classifiers LR, Random Forest, SVM, Gradient Boosting, Gaussian Naïve Bayes, MLP, Decision Tree is presented by this paper. This paper also presents importance of evaluating the error rates like MAE, MAPE, RMSE and MALE which strengthens the diagnosis of Hepatitis C. On evaluation of all the metrics performance, the Gradient Boosting classifier is giving the best results in accuracy and error rates. Therefore, this data can be used to determine model selection optimization and refinement processes, precision, eventually provides disease management tool predictive operation. With this, the research tackled the issue of the role of machine learning is for Hepatitis C disease prediction, as the precision, recall, and accuracy metrics are fundamental for improving prognostic and diagnostic models, and, in that way, highly important for improving the state of health and patient outcomes.

References

[1] Nandipati, Satish CR, et al. “Hepatitis c Virus (HCV) Prediction by Machine Learning Techniques.” *Applications of Modelling and Simulation*, vol. 4, no. 0, 15 Mar. 2020, pp. 89–100, arqiipubl.com/ojs/index.php/AMS_Journal/article/view/122.

[2] Kaunang, Fergie Joanda. "A Comparative Study on Hepatitis C Predictions Using Machine Learning Algorithms." *8ISC Proceedings: Technology* (2022): 33-42.

[3] Kashif, A. A., Bakhtawar, B., Akhtar, A., Akhtar, S., Aziz, N., & Javeid, M. S. (2021). Treatment response prediction in hepatitis C patients using machine learning techniques. *International Journal of Technology, Innovation and Management (IJTIM)*, *1*(2), 79-89.

[4] Mamdouh Farghaly, Heba, Mahmoud Y. Shams, and Tarek Abd El-Hafeez. "Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt." *Knowledge and Information Systems* 65.6 (2023): 2595-2617.

[5] Ahmed, D. Y. Mohammed, and K. A. Zidan, “Diagnosis of hepatitis disease using machine learning techniques,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 3, p. 1564, Jun. 2022, doi: 10.11591/ijeecs.v26.i3.pp1564-1572.

[6] Ghosh, Mounita, et al. "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease." *Intelligent Automation & Soft Computing* 30.3 (2021).

[7] Y. Fan, X. Lu, and G. Sun, “IHCP: interpretable hepatitis C prediction system based on black-box machine learning models,” *BMC Bioinformatics*, vol. 24, no. 1, Sep. 2023, doi: 10.1186/s12859-023-05456-0.

[8] Sachdeva, Ravi Kumar, et al. "A systematic method for diagnosis of hepatitis disease using machine learning." *Innovations in Systems and Software Engineering* 19.1 (2023): 71-80.

[9] L. Syafaah, Z. Zulfatman, I. Pakaya, and M. Lestandy, “Comparison of machine learning classification methods in hepatitis C virus,” *JOIN (Jurnal Online Informatika)*, vol. 6, no. 1, p. 73, Jun. 2021, doi: 10.15575/join.v6i1.719.

[10] Edeh, Michael Onyema, Surjeet Dalal, Imed Ben Dhaou, Charles Chuka Agubosim, Chukwudum Collins Umoke, Nneka Ernestina Richard-Nnabu, and Neeraj Dahiya. "Artificial intelligence-based ensemble learning model for prediction of hepatitis C disease." *Frontiers in Public Health* 10 (2022): 892371.

[11] Alizargar, Azadeh, Yang-Lang Chang, and Tan-Hsu Tan. "Performance comparison of machine learning approaches on hepatitis C prediction employing data mining techniques." *Bioengineering* 10.4 (2023): 481.

[12] Chown, H. (2019). A comparison of machine learning algorithms for the prediction of Hepatitis C NS3 protease cleavage sites. *The EuroBiotech Journal*, *3*(4), 167-174.

[13] Hoffmann, G., Bietenbeck, A., Lichtinghagen, R., & Klawonn, F. (2018). Using machine learning techniques to generate laboratory diagnostic pathways—a case study. J Lab Precis Med, 3, 58. https://archive.ics.uci.edu/ml/datasets/HCV+data

[14] Butt, Muhammad Bilal, Majed Alfayad, Shazia Saqib, M. A. Khan, Manir Ahmad, Muhammad Adnan Khan, and Nouh Elmitwally. "Diagnosing the stage of hepatitis C using machine learning." *Journal of Healthcare Engineering* 2021 (2021): 8062410.

[15] Orooji, A., and F. Kermani. "Machine learning based methods for handling imbalanced data in hepatitis diagnosis. Front Heal Inform 10: 57." (2021).

[16] Abd El-Salam, S. M., Ezz, M. M., Hashem, S., Elakel, W., Salama, R., ElMakhzangy, H., & ElHefnawi, M. (2019). Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients. *Informatics in Medicine Unlocked*, *17*, 100267.

[17] Baptista, D., Ferreira, P. G., & Rocha, M. (2021). Deep learning for drug response prediction in cancer. Briefings in bioinformatics, 22(1), 360-379.

[18] Bhargav, K. S., Thota, D., Kumari, T. D., & Vikas, B. (2018). Application of machine learning classification algorithms on hepatitis dataset. International Journal of Applied Engineering Research, 13(16), 12732–12737.

[19] Kaunang, F. J., & Rotikan, R. (2018). Students' academic performance prediction using data mining. Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018. https://doi.org/10.1109/IAC.2018.8780547

[20] Ali MMR, Helmy Y, Khedr AE, Abdo A (2018) Intelligent decision framework to explore and control infection of hepatitis C virus. International conference on advanced machine learning technologies and applications. Springer, New York, pp 264–274