Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

Machine Learning-Based Classification Algorithms for Predicting Hepatitis C: A Comprehensive Analysis

Amna Zulfiqar[1*], Tahira Iqbal[2], and Anum Munir[2]

[1]Department of Computer Science, NFC-IET Multan, Pakistan.
[2]Department of Computing, National University of Modern Languages, Islamabad, Pakistan.
[*]Corresponding Author: Amna Zulfiqar. Email: nisazulfiqar15@gmail.com

Received: July 11, 2024 **Accepted:** September 29, 2024

Abstract: A disease's accurate diagnosis is one of the most important tasks in the medical field. The most dangerous sickness, which continuously affects a many individuals is hepatitis disease, hence there is a need to automate the disease diagnosis. This study evaluates the diagnostic performance in terms of various parameters and optimization techniques using a range of machine-learning algorithms on a hepatitis dataset. A large dataset that included clinical history, lab test results, and demographic data was used. To get the data ready for analysis, preprocessing techniques such as data cleaning, data discretization, and data normalization were used. The algorithms included in this study are Multi-Layer-Perceptron, Support Vector Machine, Naive Bayes and Random Forest, these algorithms were trained and assessed using metrics including accuracy, recall, precision and F1 score. To minimize overfitting, the model's performance was checked using K-fold cross-validation. ReLU activation function was applied to Multi-Layer-Perceptron for solving the vanishing gradient problem. The classification accuracy scores demonstrate promising outcomes, with SVM scoring 91.86%, NB scoring 89.43%, RF scoring 89.43% and MLP scoring 92.68%. Among all algorithms MLP shows highest frequency.

Keywords: Hepatitis; Liver; Hepatitis A (HAV); Hepatitis B (HBV); Hepatitis C (HCV); Hepatitis D (HDV); Hepatitis E (HEV); Machine Learning; Random Forest (RF); Support Vector Machine (SVM); Multilayer Perceptron (MLP); Naïve Bayes (NB); Artificial Neural Network (ANN).

1. Introduction

The name "hepatitis" comes from the Latin word "itis," which means inflammation, and the ancient Greek word "hepar," which means liver [1]. One of the most common viral diseases, hepatitis is described as "liver inflammation" brought on by a virus and causes 1.5 million deaths annually globally [2]. Liver cells are often the target of this virus. The biggest gland in the human body is the liver. Its weights is around 3 pounds (1.36 kg). Hepatitis is further divided into *chronic* and *acute* hepatitis. Acute hepatitis is characterized by a rapid, severe, and painful onset of illness, with much severe symptoms for critical patients, although it lasts for small duration of less than a month or two. In most cases, there is just a slight harm to the liver cells and inflammation in the immune system. Chronic hepatitis can last for up to six months [3]. Hepatitis virus has five variants, each with its own set of characteristics. These five variants include HAV, HBV, HCV, HDV, and HEV [22].

The core purpose of this research is to optimize the automatic detection of hepatitis C using different machine learning algorithms.

1.1. Hepatitis C

Hepatitis C is a global disease that impacts people all over the world. Being in direct contact with infected blood or body fluids containing blood can spread this blood-borne illness. Hepatitis causes chronic illness in almost 71 million people globally, and approximately 290000 people died from this disease. This virus is more common among persons in Africa and Southeast Asia than in industrialized nations in North

America and Europe. Additionally, other nations, such as China, Egypt, and Pakistan, have a higher proportion of people with chronic diseases [19]. Symptoms of the hepatitis-C virus occur quite late in the course of the illness. Eighty out of every 100 individuals infected with the disease do not exhibit any symptoms after contracting it in the initial stages, which leads to increased mortality rates and liver failure. No vaccine is effective against the Hepatitis-C virus. Therefore, assessing the degree of liver damage in the afflicted patient may help doctors treat and successfully manage chronic infections[18],[20]. There is currently no effective therapy for chronic HCV, a chronic infection that can lead to major consequences such cirrhosis, liver fibrosis, and cancer of the liver [4],[21]. Cirrhosis is a late stage of liver fibrosis characterized by alterations in the liver's vascular structure and overall architecture. [5]. Liver fibrosis means scarring of healthy liver tissues, this occurs as a result of the wound healing caused by the chronic hepatitis C. Fibrosis is an ongoing concept that can be reversed if HCV infection is treated early enough [6]. HCV-related hepatocellular carcinoma (HCC) is most commonly linked with cirrhosis or fibrosis of the liver [7]. Even though modern treatment alternatives such as direct-acting antivirals (DAAs) may effectively eliminate viruses from human systems with a success rate of more than 90%, however removal of HCV removal is still a big challenge [23]. The following paper proposes to make many important contributions to the body of current literature:

Comprehensive Data Utilization: We made use of a variety of data that includes clinical history, important laboratory test findings, and demographic data. This comprehensive method enhances model robustness by providing a more comprehensive understanding of the variables affecting Hepatitis C diagnosis.

Comparative Analysis of methods: We systematically compared the performance of various machine learning classification techniques, including Random Forest, Support Vector Machines, Naive Bayes and Multiple Layer Perceptron, using predetermined evaluation criteria. A better knowledge of which algorithms are best at predicting Hepatitis C and under what circumstances is made possible by this thorough examination.

Verification Using Cross-Validation Methods: We made sure that our results were reliable and generalizable by using k-fold cross-validation. This rigorous procedure highlights the significance of strong validation procedures in machine learning research and establishes a standard for future work in this field.

Future Research Directions: By identifying key areas that require further investigation, such as the addition of more characteristics like social and economic variables which might improve model performance even more, this study establishes the framework for future research.

1.2. Machine Learning

Among the most widely used fields in data science is machine learning, which has several uses. It refers to the automated detection of patterns in data. The goal of machine learning technologies is to provide algorithms the ability to understand and make decision on its own [8]. Machine learning enables computers to learn existing data and then do prediction on future information. The algorithms used in machine learning, firstly we train them to learn from the input data and forecast future data. A machine learning model is not static, in fact, it is dynamic and it makes itself better when more data is provided.

There are three main categories of machine learning algorithms however we will discuss only two:

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

1.2.1. Supervised Machine Learning

Supervised machine learning algorithms are those in which model is trained through labelled training dataset. Then the trained algorithm is used in the prediction of unlabeled data [24].

Supervised Learning can be further divided into two types: classification and regression. In classification, the output of dataset will be distinct. Data can easily be categorized and model can depict the class of the instance, i.e., either the instance being tested belongs to Class A or class B and more precisely. In our case, the classification algorithm can predict if the person is having hepatitis or not.

In regression, the model will predict continuous values. Continuous values are on-going and non-static, e.g., predicting the house price based on the features of the house.

Enlisted below are few examples of supervised classification algorithms:

1. Support Vector Machine (SVM)
2. Naive Bayes (NB)
3. Random Forest (RF)
4. Multi-Layer Perceptron (MLP)

1.2.2. Unsupervised Learning

Labeled data is not required for unsupervised learning. The algorithm train itself by learning the hidden patterns and information present in the unlabeled dataset. Model is built and trained using training data while to check accurate performance, test data is used. The machine learns from previously presented information to anticipate the factual result by predicting the outcome derived from past occurrences [11]. Clustering, Anomaly detection and Neural networks are few examples of unsupervised learning algorithms.

2. Literature Review

Krishna [8] utilized machine learning models to automatically detect hepatitis along with predicting the life expectation of hepatitis patients. Comparative study was performed on different machine learning algorithm including SVM ,ANN and K-nearest neighbor (KNN). Out of these algorithms ANN performed best with 96% prediction accuracy . Mahmood *et al* [9] made a comparison of previous studies implementing machine learning algorithms for detecting several diseases. It was found that the accuracy of any dataset depends on the size of the dataset and the features selected. Another important finding of this study was that ensemble of algorithms gives better result as compared to a single model.

Yang Chen *et al* [10] proposed a detailed model to forecast hepatic fibrosis in persistent hepatitis patient. Image dataset was fed to four classifiers including SVM, NB, RF and KNN. Best results were provided by Random Forest classifier leading to the fact that sophisticated algorithms are much better for assessing the hepatic fibrosis stage.

Nayeem *et al* [12] used a variety of data mining methods to forecast hepatitis-related illnesses. To enhance the efficiency of machine learning algorithms, observations that had missing data were eliminated. Moreover the selection method named information-gain feature was used to remove the unnecessary features from the feature set. The different classification algorithms applied included KNN, NB, RF, MLP and SVM. The best performance was achieved by Random forest with the prediction accuracy of 92.41%.

Shergill *et al* [13] summarized the existing literature for the automatic detection of hepatitis using Artificial Intelligence along with highlighting the future outlook of AI in automation of disease detection. Hira [14] performed an experiment to find out better performing classification algorithm in automatic detection of hepatitis. The two classification algorithms utilized include MLP and SVM. The class-imbalance dataset was balanced using the SMOTE approach. Overall SVM outperformed MLP in automatic prediction of hepatitis patients. Parisi et al [15] investigated a hybrid AI based model ,their main focus was to predict if the person having hepatitis will survive or not. Multi-Layer Perceptron (MLP) were primarily employed for classification, while Lagrangian Support Vector Machines (LSVM) were also utilized for feature reduction. The model depicted 100% reliability in predicting whether a chronic hepatitis patient will survive or not.

3. Research Question

a. What are the major concerns regarding overfitting in the dataset?
b. How to predict hepatitis disease by using machine learning algorithms?
c. How to implement a model based on multi layers perceptron (MLP) for hepatitis disease prediction?

4. Data Set

The dataset of Hepatitis-C was collected from UC Irvine Machine learning repository which is a public repository [25]. It is important to mention here that the dataset contained some missing values and to handle missing values, there are different ways. For our research, we handled the missing values by replacing them with Zero. The hepatitis dataset consists of 615 entries and 13 attributes shown in **Table 1**.

Table1. Dataset description

Attribute	Description
Category	Tells if the patient is having hepatitis, fibrosis or cirrhosis
Age	It tells in years the age i.e., 26 ,40
Sex	Male or female
ALB	Acronym for Albumin
ALP	Acronym for Alkaline Phosphatase
CHE	Acronym for Cholinesterase
ALT	Acronym for Alanine Aminotransferase
AST	Acronym for Aspartate Aminotransferase
BIL	Acronym for Bilirubin
GGT	Acronym for Gamma Glutamyl
CREA	Acronym for Creatinine
CHOL	Acronym for Cholesterol
PROT	Acronym for Protein

5. Proposed Methodology

In our work, the data was retrieved from the repository and then missing values were filled. After that categorical data was converted to numerical data, i.e., numbers were assigned to different attributes of the data. Following this, the data is categorized into two groups: 20% testing data and 80% training data. Additionally, the K-fold cross validation resampling procedure was used, with K set to 10. Different classifiers including SVM, NB, RF and MLP were implemented in which MLP gave the best results. The vanishing gradient issue was resolved in MLP by using the ReLU activation function. The study's workflow is illustrated in **Figure 1**.

The classification algorithms experimented in our study include Support Vector Machine (SVM), Random Forest (RF), Multilayer perceptron (MLP) and Naïve Bayes (NB).

5.1. Support Vector Machine

SVM, a supervised learning method, is used for both regression and classification. The SVM classifier's primary goal is to distinguish between samples belonging to two different classes in the training data by identifying the optimal classification model [17]. A perfect and distinct hyperplane is created by SVM classifier for classification in high dimensional space.

The aim os to maximize the margin between different classes. If margin is minimized result accuracy will not be good and if it is maximized it means result is more accurate. As a result, the category of data can easily be determined [11]. **Figure 2** explains the working of SVM.

5.2. Naïve Bayes

NB classifier, which is also a Bayes Decision theory, is a straightforward Bayesian classifier. This classifier assumes that each characteristic has a statistically independent impact on each class. Assuming that the existence of one feature in a data collection is unrelated to the existence of another feature, the word naïve refers to a conditional independence across features characteristics.

Figure 3 describes the working of Naïve Bayes algorithm. Assume that we have two categories of items i.e., white and black. It can be seen in **Figure 3** that the ratio of white items is twice of black so new item have more chances to fall in white category. This belief is referred to as prior Probability in Bayesian analysis. Even though prior priority suggest that new item belongs to white class, likelihood indicates that there are more chances to fall in black class. By integrating both prior probability and likelihood the final classifier in a Bayesian study is created.

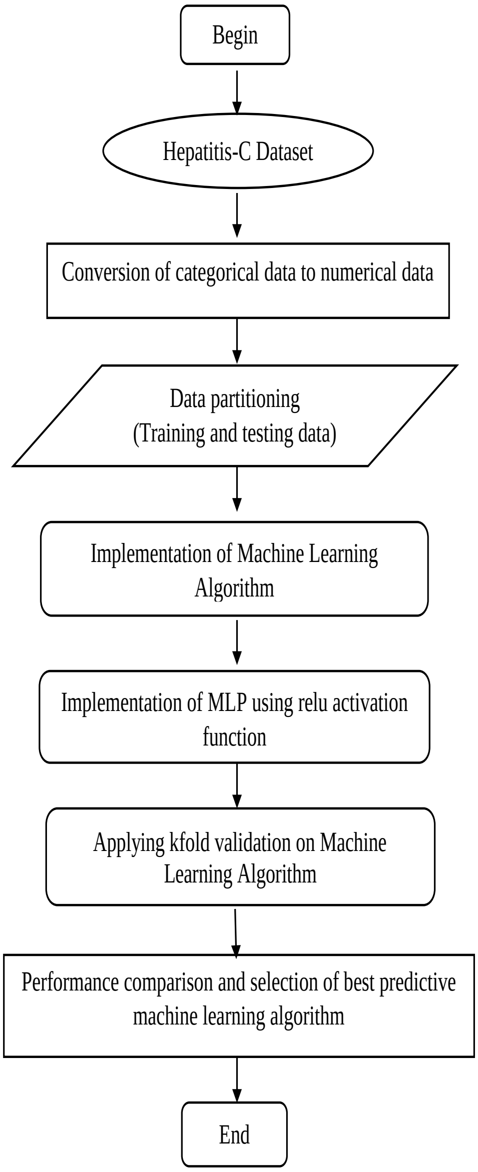


Figure 1. Proposed Methodology

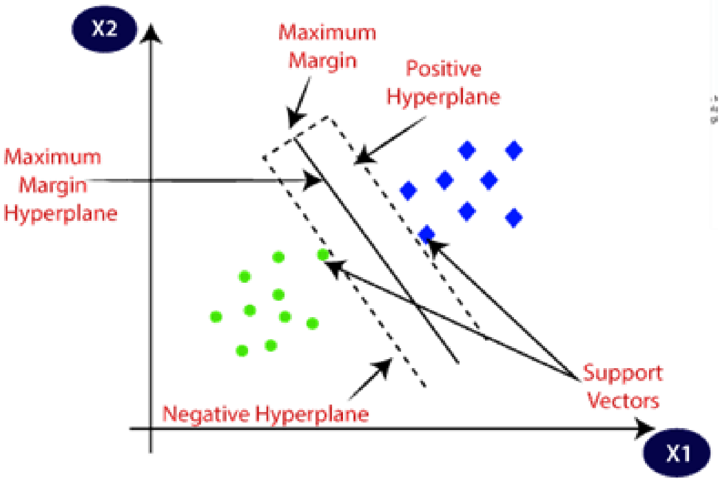


Figure 2. Support Vector Machine

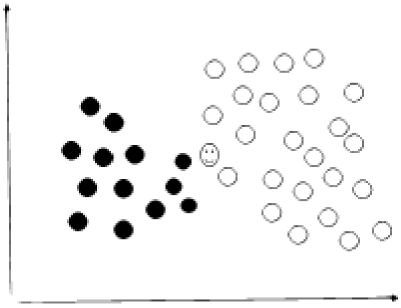


Figure 3. Naive Bayes

5.3. Random Forest

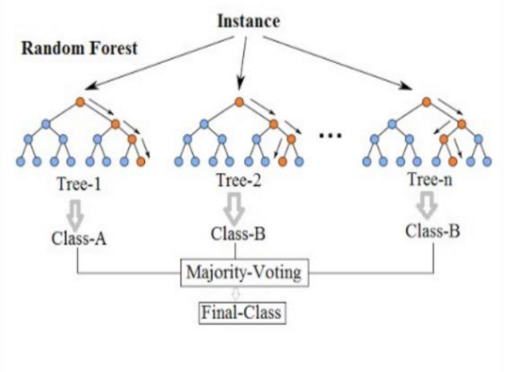


Figure 4. Random Forest

An ensemble learning method called Random Forest uses a large number of decision trees. The development of a strong model is the primary goal of the ensemble technique. In random forest every decision tree has been trained using bootstrap samples. By executing a binary tree, samples are forwarded from root node to leaf node. A certain threshold value is set to compare features. The forest eventually decides based on the 'votes' of all trees. Since the random forest method considers the results of several distinct decision trees, hence it decreases the uncertainty caused by only considering one decision tree for the same dataset. Figure 4 shows the pictorial explanation of Random Forest.

5.4. Multi-layer Perceptron

Comprising three distinct layer types—an input layer, one or more hidden layers, and an output layer—it is among the most widely used neural network models [16]. A perceptron, which is an artificial neuron like the human brain and is also known as a node, transmits information from one node to another. To acquire the goal value, the output of one node's activation function is passed on to the subsequent node layer. When data in the MLP algorithm is fed forward and propagated through several layers to produce a result that is different from the expected output, then that difference is referred to as an error. A supervised process called backpropagation accompanies MLP which repeats the process several times while modifying the weights in order to get the desired result [15].

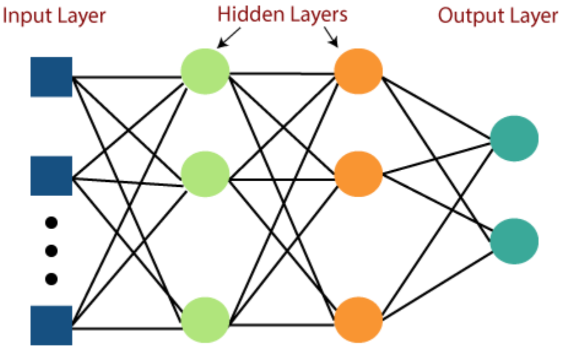


Figure 5. Multi-layer Perceptron Model

5.5. k-fold cross validation

K-fold cross validation is used to assess the model's performance to a greater extent. The dataset is split into K folds for K-fold cross validation, in which K is the total number of groups into which the data samples are divided. At any stage throughout the procedure, each fold serves as a testing set. For instance, we may refer to this as a 10-fold cross-validation if the k-value is 10. It means that the dataset is divided into 10 groups out of which 9 are used for training and 1 for testing. This process of training and testing is repeated 10 times so that every instance is utilized once for testing and k-1 times for training.

6. Experimental Setup and Performance Measures

6.1. Experimental Setup

Python programming language was used to carry out the experiments. It is robust and quick programming language which is simple to learn. It is widely used in development of web and desktop applications and especially in the domain of artificial intelligence. It provides several libraries for different tasks. The libraries incorporated in our study are enlisted below:

- Panda
- NumPy
- SciPy
- Matplotlib
- Scikit
- Seaborn

6.2. Performance Metrics

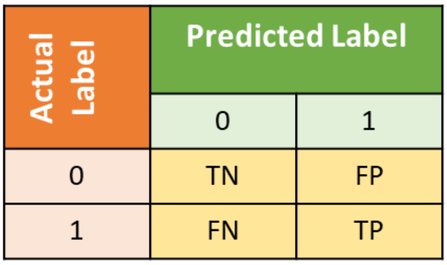
The suggested methodology's performance is assessed using a number of performance measures which are briefly explained in this section. Here it is important to mention that mostly performance metrics were calculated using the confusion matrix. So firstly, we will briefly discuss about the confusion matrix.

The confusion matrix displays the algorithm's performance. It shows how the classifier becomes bewildered when making predictions. The columns show the predicted class label, while the actual class label is shown by rows in **Figure 6.** Here True Positive (TP) tells how many patients are correctly diagnosed having hepatitis disease.

The number of accurately recognized healthy patients is shown by True Negative (TN).

The number of persons who have been mistakenly diagnosed with a disease when, in fact, they are healthy is known as False Positives (FP).

False Negative (FN) symbolizes the patients who were incorrectly classed as healthy but were actually ill.



Actual Label	Predicted Label	
	0	1
0	TN	FP
1	FN	TP

Figure 6. Confusion Matrix

6.2.1. Precision

The percentage of cases that are truly positive and are categorized as positive is known as precision. It displays how much estimated values match actual values.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

6.2.2. Accuracy

The accuracy of the model is defined as the proportion of correctly classified occurrences to all instances.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

6.2.3. F1- score

The combination of precision and recall while intending to create a balance between the two is termed as F1 score.

$$F1\ score = \frac{2*precision*recall}{precision+recall} \tag{3}$$

6.2.4. Recall

Recall is the proportion of positive occurrences that are correctly classified as positive. Sensitivity is often used to describe recall.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

7. Results

In this dataset, there were total 615 patients belonging to age group of 19 to 74. There were 238 females and 377 males. Out of 615 patients, the healthy control patients are 533 and patients that are diagnosed by hepatitis are 82. Hepatitis-C is further divided into liver fibrosis and cirrhosis.

All these algorithms were implemented in python. Training and testing data split out of the dataset. 20% of the data was utilized for testing, while the remaining 80% was used for training. A method of resampling, i.e., K-fold cross validation was also applied while keeping K=10.

Table 3 - Table 6 shows how well various classifiers performed on the dataset**.** Note that we have five categories in the dataset represented by different class notations while 2 represents the data of hepatitis patients.

Table3. SVM Performance Evaluation

Class	precision	recall	f1-score	support
0	1.00	0.98	0.99	106
1	0.33	1.00	0.50	1
2	0.40	0.80	0.53	5
3	0.60	0.33	0.43	9
4	1.00	0.50	0.67	2
accu			0.92	123
mac avg	0.67	0.72	0.62	123
weigh avg	0.94	0.92	0.92	123

Table 4. Naïve Bayes Performance Evaluation

Class	precision	recall	f1-score	support
0	0.97	0.94	0.96	106
1	0.25	1.00	0.40	1
2	0.25	0.40	0.31	5
3	0.86	0.67	0.75	9
4	1.00	0.50	0.67	2
accu			0.89	123
mac avg	0.67	0.70	0.62	123
weigh avg	0.93	0.89	0.91	123

Table 5. Random Forest Performance Evaluation

Class	precision	recall	f1-score	support

0	0.95	0.98	0.96	106
1	1.00	1.00	1.00	1
2	0.50	0.60	0.55	5
3	0.25	0.11	0.15	9
4	0.50	0.50	0.50	2
Accu			0.89	123
mac avg	0.64	0.64	0.63	123
weigh avg	0.87	0.89	0.92	123

Table 6. MLP Performance Evaluation

Class	precision	recall	f1-score	Support
0	1.00	0.98	0.99	106
1	0.00	0.00	0.00	1
2	0.50	0.80	0.62	5
3	0.62	0.56	0.59	9
4	0.50	0.50	0.50	2
Accu			0.93	123
mac avg	0.53	0.57	0.54	123
weigh avg	0.94	0.93	0.93	123

The accuracy obtained on training and testing set is summarized in **Table 7** for the different classifiers.

Table 7. Classifiers Accuracy on Train and Test Set

Classifiers	Training Accuracy	Test Accuracy
SVM	99.79	91.86
NB	93.04	89.43
RF	100 (highest)	89.43
MLP	99.59	**92.68(highest)**

The result of 10-fold cross validation applied on different classification models is shown in **Table 8**:

Table 8. K-fold Cross Validation Result for Different Classifiers

Models	1-f	2-f	3-f	4-f	5-f	6-f	7-f	8-f	9-f	10-f
SVM	0.796	0.91	0.314	0.677	0.963	0.849	0.864	0.472	0.653	0.472
NB	0.65	0.534	0.458	0.515	0.9	0.773	0.909	0.608	0.351	0.412
RF	0.883	0.839	0.458	0.605	0.933	0.864	0.713	0.713	0.834	0.547
MLP	0.898	0.898	0.506	0.748	0.833	0.924	0.989	0.872	0.702	0.472

8. Discussion

The UCI Machine Learning Repository provided the data.It consisted of 13 attributes and 615 entries. Different classifiers were implemented in this work i.e., SVM, NB, MLP, RF. The research was done using python programming language.

After preprocessing, the dataset was split into two groups: 20% for testing and 80% for training. Training data was used to train the model initially, and testing data was used to evaluate it. (the performance is shown in **Table 3 – Table 6**).

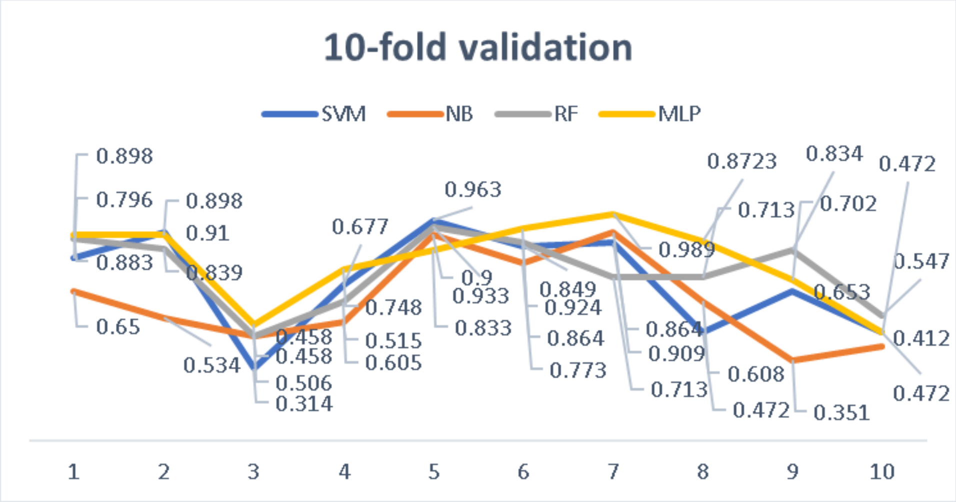


Figure 7. Graph showing K-fold performance

10-fold cross validation was also implemented, which divided the dataset into 10 portions, in each iteration 9 parts were used for training and 1 part for testing. In MLP, ReLU activation function was used which solved the problem of vanishing gradient. Different classifiers were implemented and performance was evaluated with the help of accuracy, , f1-score ,recall and precision. The findings demonstrate that out of different classifiers, the highest accuracy was obtained by MLP, i.e., 92.68% on the test set. The same is represented in **Figure 8.**

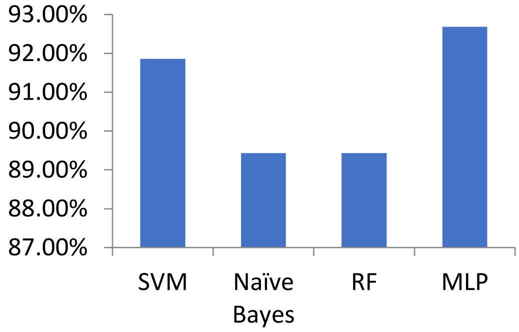


Figure 8. Performance Comparison for test set

9. Conclusion

In this paper, different classification models were utilized to determine if hepatitis disease exist or not, with the aim of automating the disease diagnosis. The models that were implemented in this paper include SVM, NB, RF and MLP. SVM and MLP outperformed others while MLP showed best result with achieved accuracy of 92.68%. The vanishing gradient issue was resolved by using the ReLU activation function. Finally, we can say that ANN can be used to solve a number of complex problems when combined with other effective techniques. The evolution of deep machine learning systems will surely influence the future of artificial intelligence in automating the diseases diagnosis.

References

1. Nancy, P., Sudha, V., & Akiladevi, R. (2017). Analysis of feature selection and classification algorithms on hepatitis data. Int. J Advanced Res Comp. Eng. Technol, 6(1), 19-23.
2. Pushpalatha, S., & Pandya, J. G. (2017, February). Designing a framework for diagnosing hepatitis disease using data mining techniques. In 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET) (pp. 1-6). IEEE.
3. Ansari, S., Shafi, I., Ansari, A., Ahmad, J., & Shah, S. I. (2011, December). Diagnosis of liver disease induced by hepatitis virus using artificial neural networks. In 2011 IEEE 14th international multitopic conference (pp. 8-12). IEEE.
4. Khatun, M., & Ray, R. B. (2019). Mechanisms underlying hepatitis C virus-associated hepatic fibrosis. Cells, 8(10), 1249.
5. Sajjad, R., Khan, M. F., Nawaz, A., Ali, M. T., & Adil, M. (2022). Systematic analysis of ovarian cancer empowered with machine and deep learning: a taxonomy and future challenges. Journal of Computing & Biomedical Informatics, 3(02), 64-87.
6. Suk, K. T., & Kim, D. J. (2015). Staging of liver fibrosis or cirrhosis: The role of hepatic venous pressure gradient measurement. World journal of hepatology, 7(3), 607.
7. Salas-Villalobos, T. B., Lozano-Sepúlveda, S. A., Rincón-Sánchez, A. R., Govea-Salas, M., & Rivas-Estilla, A. M. (2017). Mechanisms involved in liver damage resolution after hepatitis C virus clearance. Medicina Universitaria, 19(75), 100-107.
8. Baskic, D., Vukovic, V., Popovic, S., Jovanovic, D., Mitrovic, S., Djurdjevic, P., & Mijailovic, Z. (2019). Chronic Hepatitis C: Conspectus of immunological events in the course of fibrosis evolution. PLoS One, 14(7), e0219508.
9. Yarasuri, V. K., Indukuri, G. K., & Nair, A. K. (2019, December). Prediction of hepatitis disease using machine learning technique. In 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 265-269). IEEE.
10. Mahmood, I., & Mohsin Abdulazeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. J. Appl. Sci. Technol. Trends, 2.
11. Chen, Y., Luo, Y., Huang, W., Hu, D., Zheng, R. Q., Cong, S. Z., & Yan, H. (2017). Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. Computers in biology and medicine, 89, 18-23.
12. Mehmood, R., & Selwal, A. (2020). A Comprehensive Review on Face Recognition Methods and Factors Affecting Facial Recognition Accuracy. Lecture Notes in Electrical Engineering, 597, 455-467.
13. Nayeem, M. J., Rana, S., Alam, F., & Rahman, M. A. (2021, February). Prediction of Hepatitis Disease Using K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Multi-Layer Perceptron and Random Forest. In 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) (pp. 280-284). IEEE.
14. Shergill, A., Narasimhan, V. L., & Budu, E. (2021). S1112 Artificial Intelligence and Machine Learning-Based System for Diagnosis of Chronic Hepatitis C. Official journal of the American College of Gastroenterology| ACG, 116, S524.
15. Hira, N. Comparing MLP and SVM to diagnose Hepatitis C (HCV) the 'silent killer'.
16. Parisi, L., RaviChandran, N., & Manaog, M. L. (2020). A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis. Neural Computing and Applications, 32(8), 3839-3852.
17. Amato, F., Mazzocca, N., Moscato, F., & Vivenzio, E. (2017, March). Multilayer perceptron: an intelligent model for classification and intrusion detection. In 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA) (pp. 686-691). IEEE.
18. Saravanan, R., & Sujatha, P. (2018, June). A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 945-949). IEEE.
19. Siraj, M. A., Rehman, A., Aziz, O., & Khan, M. F. (2021). Systematic Literature Review: Smart Drone for Early Smoke Detection in Forest Using IOT. Journal of Computing & Biomedical Informatics, 2(01), 80-88.
20. M. J. Nayeem, S. Rana, F. Alam, and M. A. Rahman, "Prediction of hepatitis disease using K-nearest neighbors, naive Bayes, support vector machine, Multi-Layer Perceptron and Random Forest," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021.
21. Mohamed, A. A., Elbedewy, T. A., El-Serafy, M., El-Toukhy, N., Ahmed, W., & El Din, Z. A. (2015). Hepatitis C

virus: A global view. World journal of hepatology, 7(26), 2676.

22. Singh, L., Janghel, R. R., & Sahu, S. P. (2021). Classification of hepatic disease using machine learning algorithms. In Advances in Biomedical Engineering and Technology (pp. 161-173). Springer, Singapore.
23. Khatun, M., & Ray, R. B. (2019). Mechanisms underlying hepatitis C virus-associated hepatic fibrosis. Cells 8, 1249.
24. The five types of hepatitis, 09-Dec-2020. [Online]. Available: https://www.creative-diagnostics.com/blog/index.php/the-five-types-of-hepatitis. [Accessed: 13-Feb-2023].
25. Zeng, H., Li, L., Hou, Z., Zhang, Y., Tang, Z., & Liu, S. (2020). Direct-acting antiviral in the treatment of chronic hepatitis C: Bonuses and challenges. International Journal of Medical Sciences, 17(7), 892.
26. "Supervised machine learning - javatpoint," www.javatpoint.com. [Online]. Available: https://www.javatpoint.com/supervised-machine-learning. [Accessed: 15-Dec-2022].
27. Lichtinghagen Ralf, Klawonn Frank 2020. UCI Machine Learning Repository[https://archive.ics.uci.edu/ml/datasets/HCV+data Irvine, CA: University of California, School of Information and Computer Science.
28. Shafiq, M. U., & Butt, A. I. (2024). Segmentation of Brain MRI Using U-Net: Innovations in Medical Image Processing. Journal of Computational Informatics & Business, 1(1).
29. Ahmad, R., Salahuddin, H., Rehman, A. U., Rehman, A., Shafiq, M. U., Tahir, M. A., & Afzal, M. S. (2024). Enhancing Database Security through AI-Based Intrusion Detection System. Journal of Computing & Biomedical Informatics, 7(02).
30. Khan, R., Iltaf, N., Shafiq, M. U., & Rehman, F. U. (2023, December). Metadata based Cross-Domain Recommender Framework using Neighborhood Mapping. In 2023 International Conference on Sustainable Technology and Engineering (i-COSTE) (pp. 1-8). IEEE.