Analysis of Liver Disease Using Multi Perceptron Neural Network and Voting Classifier

Ashutosh Kesharwani[1], Deepanivasini B S[2], Khadar Babu Shaik[3]

MSc Data Science, Department of Mathematics, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India-632014.
ashukesharwani07768@gmail.com, deepanivi18@gmail.com, khadar.babu36@gmail.com

ABSTRACT:

The liver serves as one of the most significant organs in the human body. Nowadays, alcohol consumption, tattoos or body piercings, drugs, and obesity are the main causes of liver disease. The liver is in charge of producing bile, which is a fluid that aids digestion. The red blood cells will disintegrate, balance, and produce the nutrients as the blood exits the stomach and intestine and travels through the liver. These are the main processes that are carried out inside the body. If these fail to happen, then it will lead to many problems inside the body. So, it's important to see that the functions of the liver are at a normal level. Bilirubin, Albumin, Proteins, AST, ALT, and ALP are the liver function tests that help to check whether the liver has a normal range or an abnormal range. We can predict liver disease in a patient at an early stage based on previous predicted values using data from patients with abnormal liver function. This will be helpful for the doctors to make a diagnosis. In this paper, the liver function test is analysed for predicting liver disease, where the input of the patients and the output data are passed into various classifiers such as Support Vector Machine, K-Nearest Neighbor, Hard Voting Classifier, and Deep Neural Network Multilayer Perceptron techniques for predicting the liver health of patients, and optimization techniques such as the Confusion Matrix, Precision Score, Recall, Accuracy, Specificity, and F-score are used to determine which model is the best. The study shows that the voting classifier is the best for this dataset. It is the most accurate and simple method for detecting liver disease in humans.

Keywords: Feed Forward Network, Perceptron Algorithm, Support Vector Machine, Voting Classifier.

1. Introduction

Liver diseases are fast becoming recognised as public health priorities in India, as the liver is the largest organ and gland in the human body. The liver holds about 13% of a pint of the body's blood supply at any given moment. The liver has roles that include detoxification, protein synthesis, and the production of chemicals that help digest food. The liver is also a gland that has functions like bile production, absorbing and metabolising bilirubin, supporting blood clots, fat metabolization, metabolising carbohydrates, vitamin and mineral storage, helping metabolise proteins, filtering the blood, immunological function, production of albumin, and synthesis of angiotensinogen. The liver can experience a range of problems. There are different types of liver disease: fascioliasis, cirrhosis, hepatitis, alcoholic liver disease, PSC, fatty liver disease, Gilbert's syndrome, and liver cancer. The contribution of cirrhosis and its complications, collectively known as chronic disease, as causes of mortality in India has increased progressively since 1980. According to the latest WHO data published in 2022, liver disease deaths in India reached 7,28,476 (or 3.17% of total deaths).

Some statistics show that around 10 lakh patients with liver cirrhosis are newly diagnosed every year in India, and liver disease is the tenth most common cause of death in India as per the World Health Organization. 40% of India suffers from non-alcoholic fatty liver disease. The people between the ages of 40 and 50 have liver diseases in common. Therefore, a feasible and accurate prediction of liver disease is very important. Liver disease does not show noticeable signs and symptoms. The liver function tests will check the levels of enzymes and proteins in the blood. If the levels are not at a normal level, then it will indicate liver problems.

Alanine transaminase (ALT), aspartate transaminase (AST), alkaline phosphatase (ALP), albumin, total protein, and bilirubin are all common liver function tests. These tests are based on several steps that normally identify the reason for liver disease and try to treat the liver infection.

2. Literature Survey

Comprehensive feature analysis and severity prediction of liver disease can be done using machine learning by using the ANOVA test, Principal Component Analysis (PCA), and linear discriminant analysis to analyze the performance of the classifiers by Shyamala Devi et al. (2021). Every year, liver dysfunction is the cause of more than 2.4% of fatalities in India. The early stages of liver problems can be difficult to determine due to their mild symptoms. Often, the signs don't surface until it's too late by S. Sontakke et al. (2017). The utilisation of digital revolutions in medical practise can advance efficient liver disease detection from a late stage to something like an early stage through the utilisation of a machine learning approach with defined parameters that can recognise liver illness in a normal community that is asymptomatic by Samir Hassoun et al. (2020). According to Na Jiang et al. (2021), the best outcomes may be achieved by building the method of assessing chronic liver disease using an improved feedforward neural network with backpropagation that is combined with an improved ant colony optimization.

Machine learning can be used to diagnose well-compensated cirrhosis across various liver disease etiologies, and the Ensemble algorithm outperforms all other machine learning techniques by Soren Sabet Sarvestany et al. (2020). 9 models using machine learning were used to assess laboratory data from 1,453 patients with parameters pertaining to alcohol, diabetes, and fatness. The results demonstrate that the Ensemble Stacker Model produces the most accurate Prediction by Lucy Bennett et al. (2022). FLD might be recognised from the first fatty liver examination of all patients in New Taipei City using classification techniques. We employed the receiver operating characteristic curve to assess four models by Chieh-ChenWu et al. (2019). Numerous machine-learning techniques can be used to forecast heart disease, but only one technique will deliver the greatest accuracy in terms of results by Megha Kamboj (2020) and Keerthika Dulam et al. (2022). Machine learning is one of the skewing advancements that is being used in many other fields, including the pharmaceutical industry's application for illness diagnosis by Yash Jayesh Chauhan (2020). Heart-related illnesses have emerged as a major cause of death worldwide over the past several decades, not just in India by Adeen et al. (2021). The latest discovery of a mutated gene can improve the predictability of non-alcoholic fatty liver syndrome by anna kotronen et al. (2009). Data analysis has made it simple to anticipate diseases in the healthcare industry by Dr.S.Vijayarani et al. (2015).

In 3D positron emission tomography, liver tumours can be found and segmented for analysis by Weimin Huang et al. (2013). Several study teams received assistance in determining cancer patients as high- or low-risk by Shweta et al. (2022). For the purpose of diagnosing the illness, numerous knowledge techniques usually are integrated, generating numerous probability.by Sandhya Kumari et al. (2018). Naive Bayes, decision trees, and neural network algorithms were employed by Monika Gandhi et al. (2015) to analyse the medical dataset. There are different attributes at play. Therefore, it is necessary to decrease the number of capabilities. The process will involve feature engineering. They claim that the point is lost as a result. They used neural network models and decision trees.

Mohit Agrawal (2021) employs a wide range of machine learning methods, including KNN, LR, NB, and SVM, and finds that KNN, which boasts a 98% predictive analysis rate, is the most efficient. Both developed and developing nations rank breast cancer as the leading cause of mortality. Machine learning algorithms are utilised to analyse breast cancer and provide accuracy to determine the risk that it will emerge in 8% of women by Tolga Ensari et al.(2018). And Keshk (2022). When using machine learning algorithms to predict liver damage, Ketan Gupta et al. (2022) found that the RF, Light GB, and Ada booting categorization algorithms performed with the highest degree of accuracy. A confusion matrix was used to estimate the

classification performance of each machine learning approach, which provided the foundation for the comparison by S. Gupta et al. (2021), and P. Porwal et al. (2018).

3. Material and Methods

3.1 Data Collection

The datasets for the Indian Liver Patient Data of the 600 individuals who suffer from liver disorders are provided by the directorate of non-medical fields of science and engineering, the National Science Foundation, and the UCI Machine Learning Repository. The study classifies liver illnesses based on age and gender to determine whether a person has liver disease or not by evaluating liver function test criteria such as total proteins, alkaline phosphate, total bilirubin, albumin, aspartate aminotransferase, direct bilirubin, A/G ratio, and alamine aminotransferase.

3.2 Methodology

3.2.1 Data Processing

Pre-processing is required to add some necessary features, such as scaling, eliminate some irrelevant columns, and fill in some blanks in the original database. After loading the data, we'll attempt to visualise some data. Additionally, we will change the data's format to the appropriate one using a few modifications that could be useful for categorization and visualisation. It comprises data standardisation and data cleaning.

a) Data Cleaning

The datasets contain a variety of impurities that must be eliminated before building the model in order to make the predictability of the models more accurate. In this dataset, there is a feature that is a nominal categorical variable, i.e., gender, which is transformed into a numerical variable. There are also a lot of empty fields and null values in the data sets. These were filled in with KNNimputer imputation techniques to make sure that the accuracy and correctness of the models were not impacted during the model building.

b) Data Standardization

The method of standardisation involves scaling each property to ensure that it resembles a conventional normal distribution with a mean of 0 and a standard deviation of 1.

$$\theta = \frac{\alpha - \mu}{\sigma}$$

As there are significant disparities across the ranges of the features in the input data set, standardisation is used to eliminate the differences and scale the data down to lower values. So the data can be made easier for building the model, and it also helps in choosing the proper activation function for the perceptron algorithm.

The objective feature of this Indian Liver Patient Dataset follows the binary classification function, which was visualised using the pair plot. It was discovered that the data points are highly overlapped and scattered, so they can’t be linearly separable, so we have used various non-linear separability models to ensure that the model would produce an accurate result.

3.2.2 K-Nearest Neighbours (KNN)

K-Nearest Neighbors (KNN) come under supervised learning. It is used to analyse the new data entered and determine which category it belongs to. As a classification algorithm, it will take the number of the nearest point in the new data, which is the k value, check the maximum distance using Euclidean and Manhattan, and allot it to the maximum category it belongs to. Outlier and unbalanced data will be impacted in KNN. Using the right K-neighbors value can increase accuracy. For liver disease analyses, we have used the neighbor value of 4.

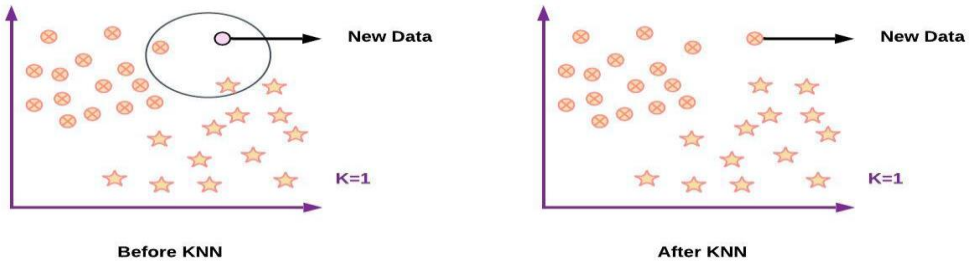

Figure 1: K-Nearest Neighbours Classification

3.2.3 Support Vector Machines (SVM)

Support vector machines (SVM) learn from supervised, labelled datasets and can be used for both forecasting and prediction-based problems.SVM is used to separate two categories with a hyperplane and also create two margins that will pass through the nearest positive and negative points in a classification, which is called the support vector. If the data is non-linearly separable, then the SVM kernel is used to separate the two categories, where it will convert lower dimensions to lower dimensions, so it will be easy to classify the points.

There are many types of SVM kernels available, but for our problem, the kernel RBF (Radial Basis Function) suits the best as our data points are highly coincident, and to determine the closeness of one point to the other, mathematical formulas for the radial basis function (RBF) were used.

$$f(\chi_1, \chi_2) = \exp\left(-\frac{||\chi 1 - \chi 2||^2}{2\sigma^2}\right)$$

Where, σ- variance and ||χ1-χ2|| - Euclidean distance between the points

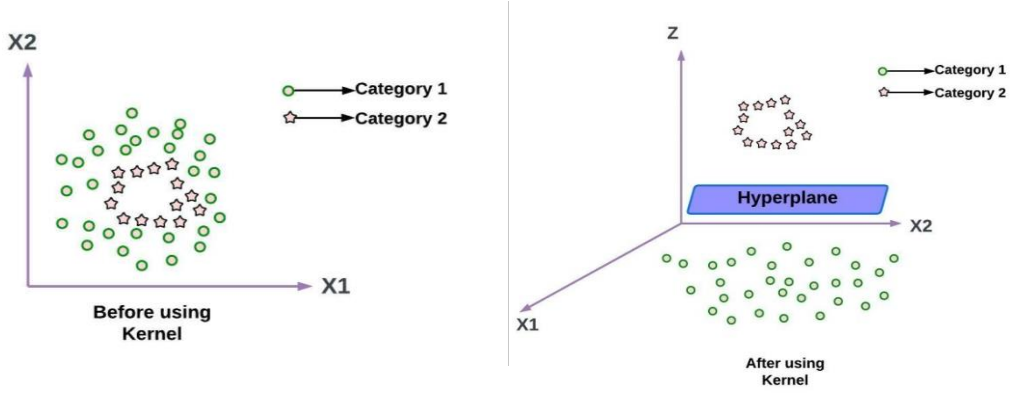

Figure 2: SVM Kernel Plotting.

3.2.4 Multilayer Perceptron Neural Network (MLP)

A multilayer perceptron neural network is a fully connected neural network that has three or more hidden layers and a non-linear activation function that causes a node in the network to change the state of a signal coming in to a signal going out. It also describes the state of neurons inside their cells.

Since our datasets have been standardised and scaled down to be between 1 and -1, we chose a 3-layer neural network with the Tanh and Sigmoidal activation functions to handle the highly overlapping and non-linear nature of the data so that it can produce better results.

Hyperbolic Tangent activation function:

Tanh is a sigmoid, but it has more enriched functionality than a sigmoid as it maps zero inputs close to zero and aggressively maps negative inputs as negatives. Its function ranges from -1 to 1.

$$\boldsymbol{Tanh} = \frac{e^{\omega}-e^{-\omega}}{e^{\omega}+e^{-\omega}}$$, ω is the input parameters

Sigmoidal activation function:

The sigmoid function represents a function that is not linear by mapping any real input to a result that is between 0 and 1.

$$\boldsymbol{Sigmoid} = \frac{1}{1+e^{-\aleph}}$$, $\aleph$ is the input parameters

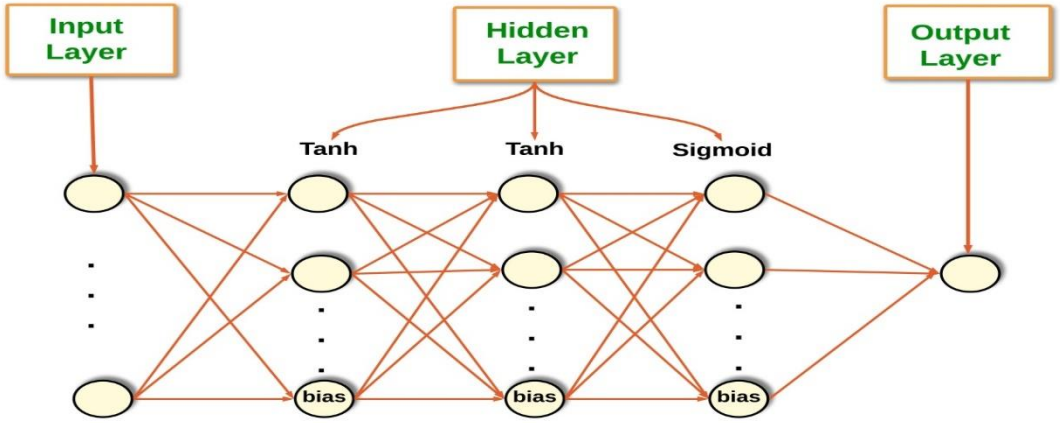

Figure 3: Multilayer Perceptron implementation

Procedure for Multilayer Perceptron

Step1: Create a perceptron with n+1 input neurons

$X=(x_0, x_2, x_3, x_4, x_5, x_{6.....} x_n)$

Step2: Where $\alpha = 1$,Which is a learning rate with the biased input.

Step3: Initialized ω = (W_0, W_1, W_2, W_3,... W_n) with the random weights. “Here, the Perceptron Convergence Theorem guarantees that we can start with any weights."

Step4: Iterate through the input patter $\emptyset_j$ of the training set using the weight set (Compute the weight sum of input for) $\sum_{i=0}^{n} Xi\omega i$

Step5: Following this calculation, the output is sent to an activation function Ƒ. which will result in the perceptron's output.

Step6: Now send the result to the next layer and continue steps 4 and 5 until the last layer has been reached.

Step7: Now compute the output. $\hat{y}j$ using the step function

$$\gamma_j = Ƒ(\emptyset j) = f(x) = \begin{cases} 1, & \emptyset j > \tau \\ 0, & Otherwise \end{cases}$$

Where a threshold parameter is τ

Step8: For each input pattern $\emptyset j$, compare the computed output γ_j with the desired output Yj , if all the input patterns were successfully identified with the output, the weight is represented.

Step9: Otherwise update the weights as follows:

- If the computed output γ_j is 1 but should have been 0, update the weight as $\omega i = \omega i - \alpha * \emptyset i$, where i iterate from 0 to n.
- And if the computed output γ_j is 0 but should have been 1, update the weight as $\omega i = \omega i + \alpha * \emptyset i$, where i iterate from 0 to n.

Where α is learning rate, ω is the weight and $\emptyset$ is input parameter

Step 10: Go to Step 4 at the first hidden layer.

3.2.5 Hard Voting Classifier (HVC)

The Hard Voting Classifier is a classifier that predicts an output based on the class that has the highest likelihood of becoming the output. It is trained using a large ensemble of many models. It just takes the average of the results from each classifier that was put into the voting classifier to predict the output class with the most votes.

It utilises the ensemble bagging learning technique by merging various machine learning models to achieve superior results. When compared to using a specific model, this method makes it possible to make better predictions.

A voting classifier is used for our datasets because the single model's individual results are unsatisfactory. As a result, we integrated all the three models to enhance their predictive capability and performance.

1. Logistic Regression
2. Decision Tree
3. Support Vector Machine (SVM)

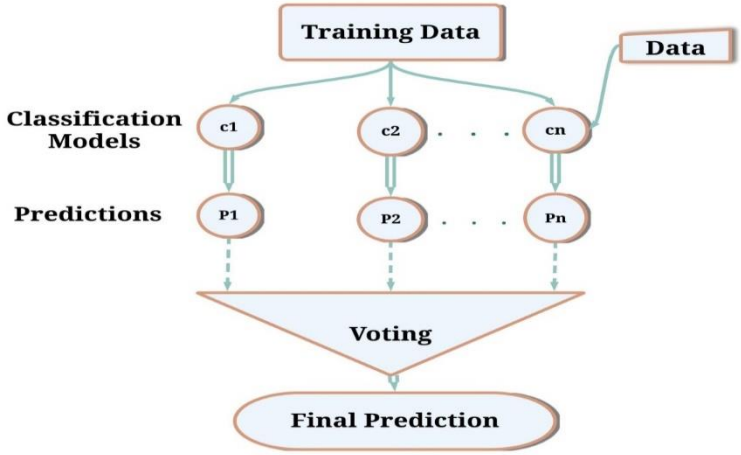

Figure 4: Voting Classifier Procedure

3.2.6 Model Evaluation Criteria

a) Confusion Matrix

The confusion matrix can be employed to assess the accuracy of the proposed model. It will utilise just 1 and 0 values to characterise the real and anticipated values in matrix form.

- If the model predicted a value of 1 and its actual value was 1, the end result would be a true positive.
- If the model predicted a value of 1 and its actual value was 0, the end result would be a false positive.
- If the model predicted a value of 0 and its actual value was 1, the end result would be a false negative.
- If the model predicted a value of 1 and its actual value was also 1, the end result would be a true positive.

b) Precision Score

Precision works on only positive results where it is an Information retrieval. precision will return Zero if the true positive + False positive becomes 0. The Precision formula can be written as

$$P\,\omega = \frac{TP}{TP+FP}$$

c) Recall

Recall works on the positive where it will check the True positive by true positive and False negative. The recall also called the True positive rate gives the percentage of the positives predicted and the formula is given as

$$R\omega = \frac{TP}{TP + FN}$$

d) Accuracy

Accuracy gives how well the model is predicted, by calculating the True Prediction by the total number of data in the Liver Disease data. The best accuracy value is 1.0 and the worst accuracy value is 0.1

$$A\omega = \frac{TP + TN}{TP + FP + TN + FN}$$

e) Specificity

Specificity works on the result of the negative or False prediction where it will checks the True Negative by True Negative and False Positive.

$$S\omega = \frac{TN}{TN + FP}$$

f) F-score

For comparing the performance of two classifiers F- score validation is used. F-score uses both precision and recall where the harmonic mean of precision and recall is the F-score value. F-score can be determined using both 0 and 1. If the F-score value is near 1 then the model is good and accurate.

$$\text{F-score} = 2 * \frac{P\,\omega \times R\omega}{P\,\omega + R\omega}$$

4. Result and Discussion

The Indian patient data were applied to each of these K-Nearest Neighbors (KNN), Multilayer Perceptron Neural Networks (MLP), Hard Voting Classifiers (HVC), and Support Vector Machines

(SVM) models to evaluate their performance and predicting ability, and their effectiveness is determined by using model evaluation criteria. We put this whole algorithm into action with Python programming, and the result found is shown below. Figure 5 shows the histogram depicting the number of patients divided into various age groups, while the pie chart displays the overall percentage of male and female participants in the dataset.

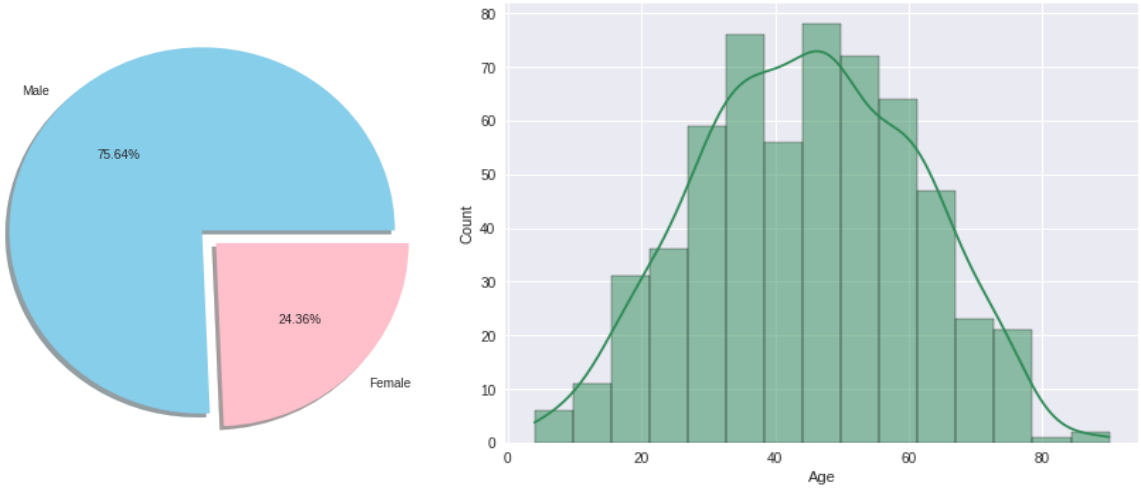

Figure 5: Patient Age group and gender count

Figure 6 shows the histogram of a number of patients in the age range, suffering from a liver disorder or not. Here, 1 indicates that the patient has an unhealthy liver, and 2 indicates that their liver is healthy and they don't have any diseases. Total datasets consist of 583 people, where 71%, or 416 people, have an unhealthy liver and 28.64%, or 167 people, have a healthy liver. The graph states that

people between the age groups of 40 and 60 are more highly affected by liver diseases than the other age groups.

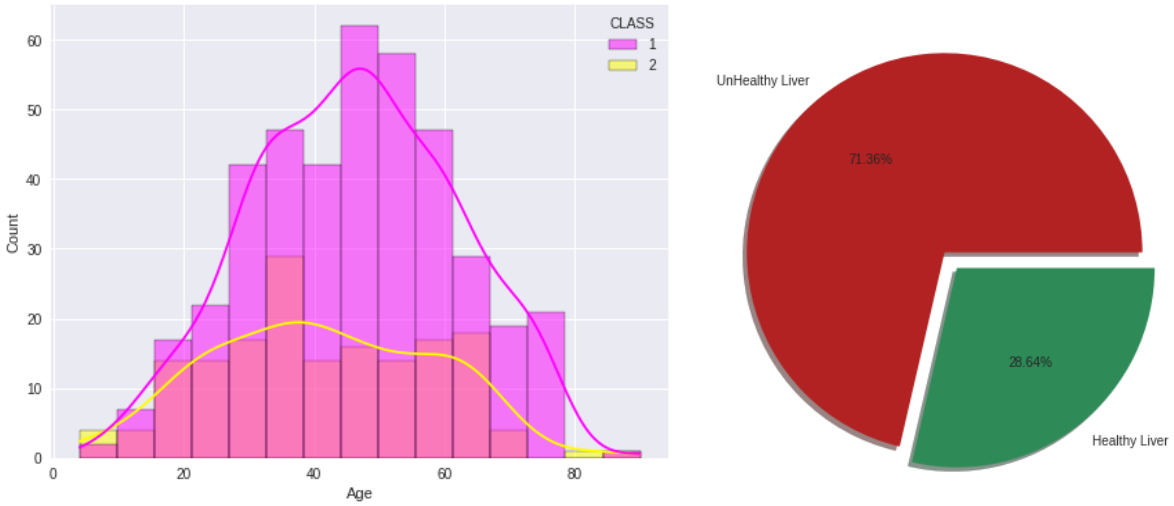

Figure 6: Classification of diseased person based on age & Healthy and Unhealthy liver ratio.

Figure 7 shows the correlation coefficient matrix, which is the correlation between all the features of the datasets. Our dataset has 10 attributes. This matrix shows how these 10 variables are correlated to each other

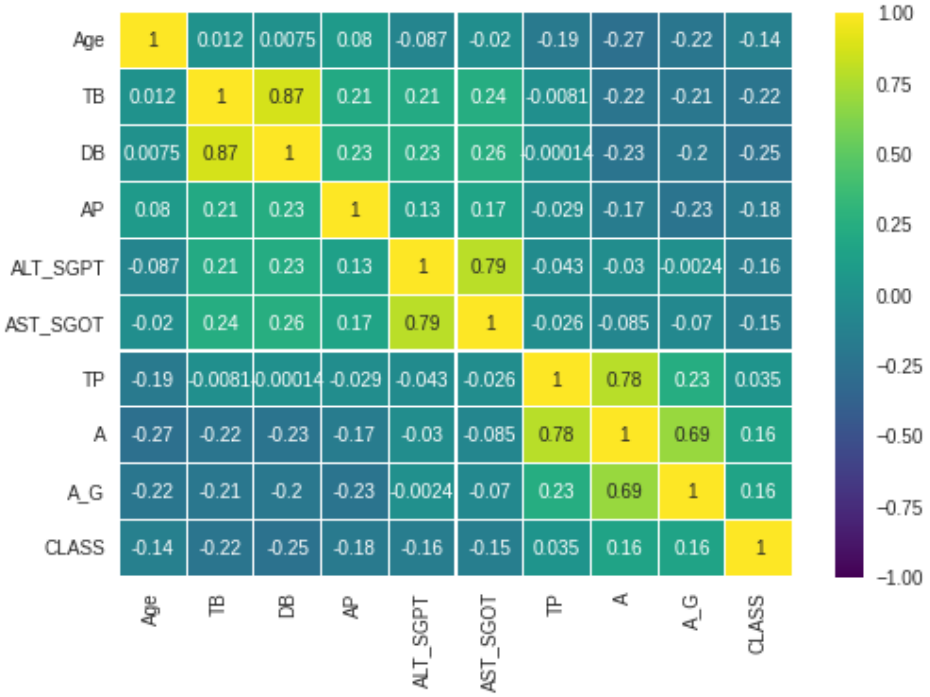

Figure 7: Correlation matrix

Figure 8 represents the pair plot graph, which graphically depicts the relationship between each and every feature of the datasets. This also provides information about the model that will produce the best results on this dataset. As our problem statement for our datasets is binary in nature, and the graphs shows that points are highly overlapped and scattered, they can't be separated using a single linear line, so we have to choose a non-linear model function for this dataset.

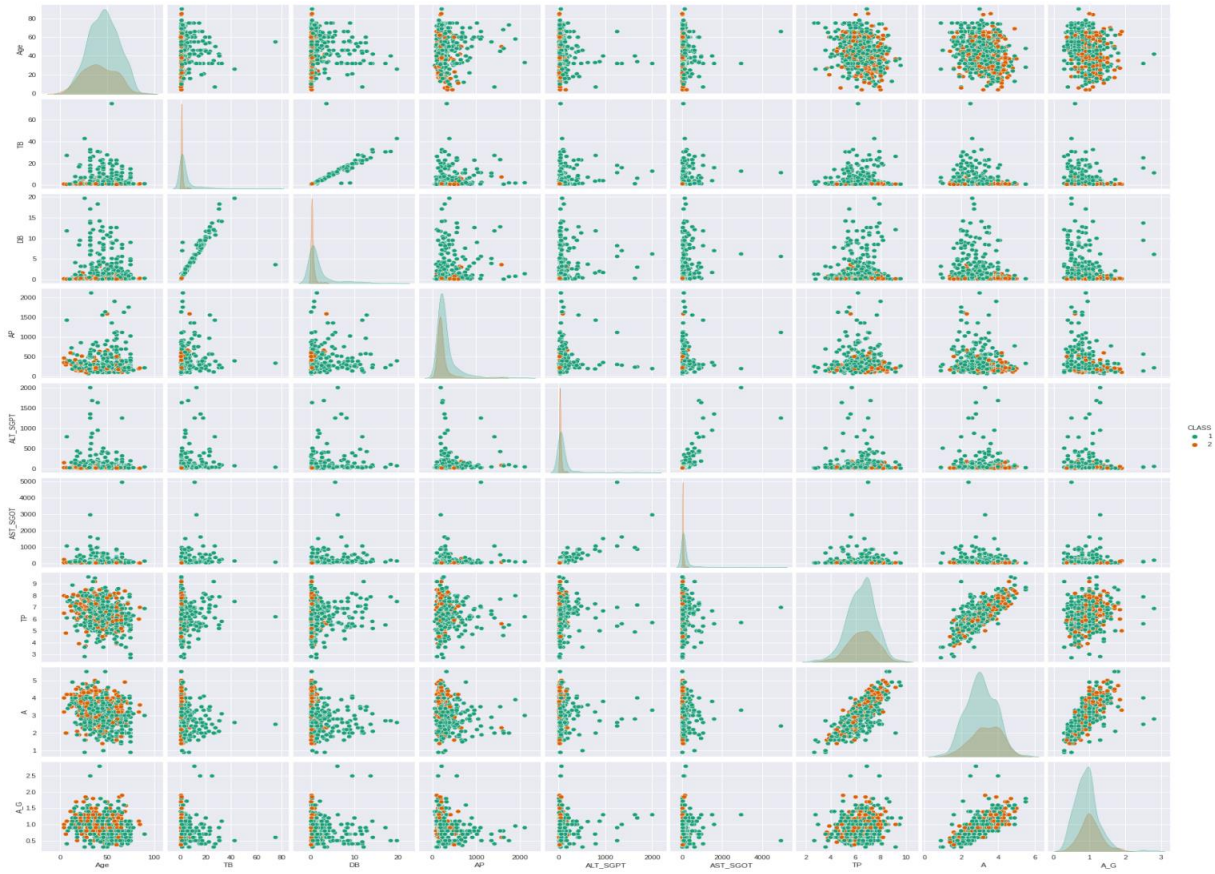

Figure 8: Relationship Graph

Table:1 Result of Models using Evaluating Criteria

Models	***Measures***				
	Accuracy	**Specificity**	**F -score**	**Precision**	**Recall**
MLP	0.77241379	1	0.87159533	0.7724138	1
SVM	0.76551724	0.9642857	0.864	0.7826087	0.964286
KNN	0.73103448	0.8928571	0.83682008	0.7874016	0.892857
HVC	0.7862069	0.9464286	0.87242798	0.8091603	0.946429

Figure 9 table shows the model optimization indices of various models used in the analyses of liver disease. The efficient model is the one that has less specificity but high recall values, and the comparison of all the indices, i.e., accuracy, specificity, F-score, precision, recall, and confusion matrix, helps to select the best appropriate model for the datasets. When all the optimization indices of different MLP, SVM, KNN, and HVC models are compared in this table, the voting classifier comes out on top for all of them.

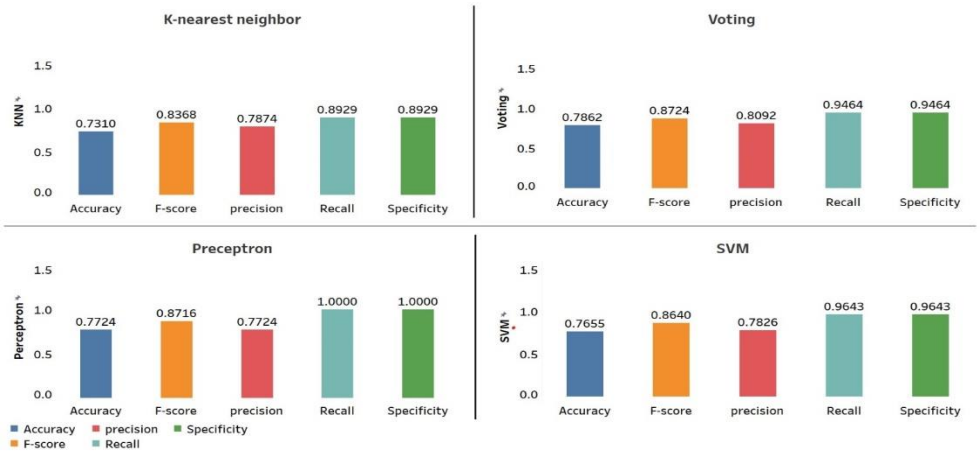

Figure 10: Compared Models using Evaluating Criteria

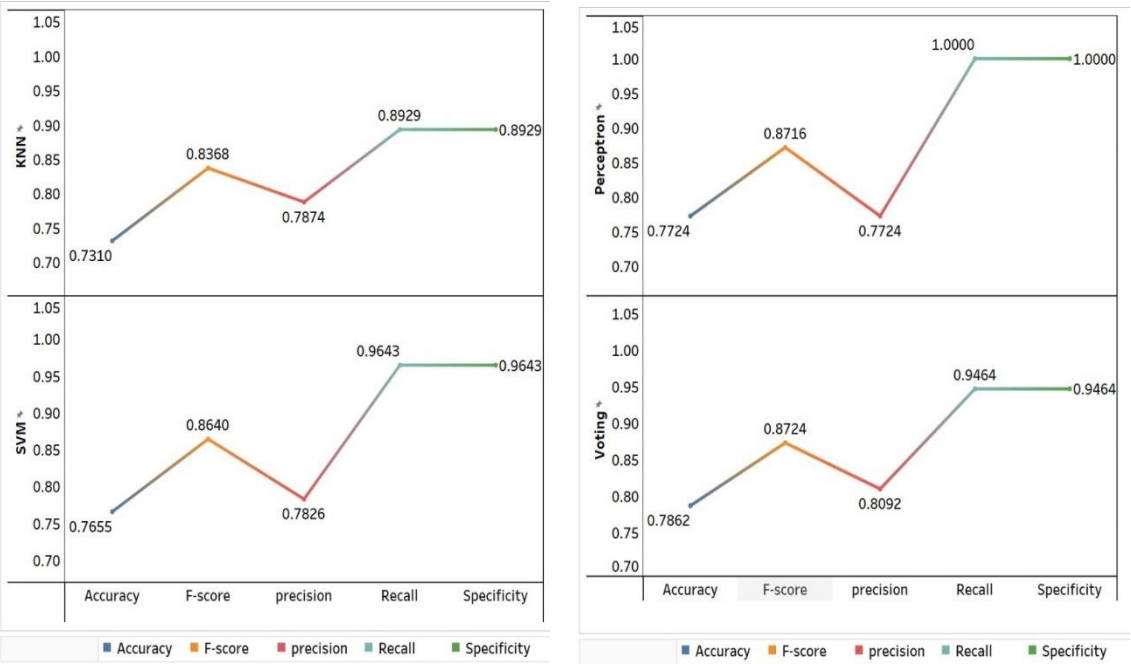

Figure 11: Individual Models Criteria Comparison

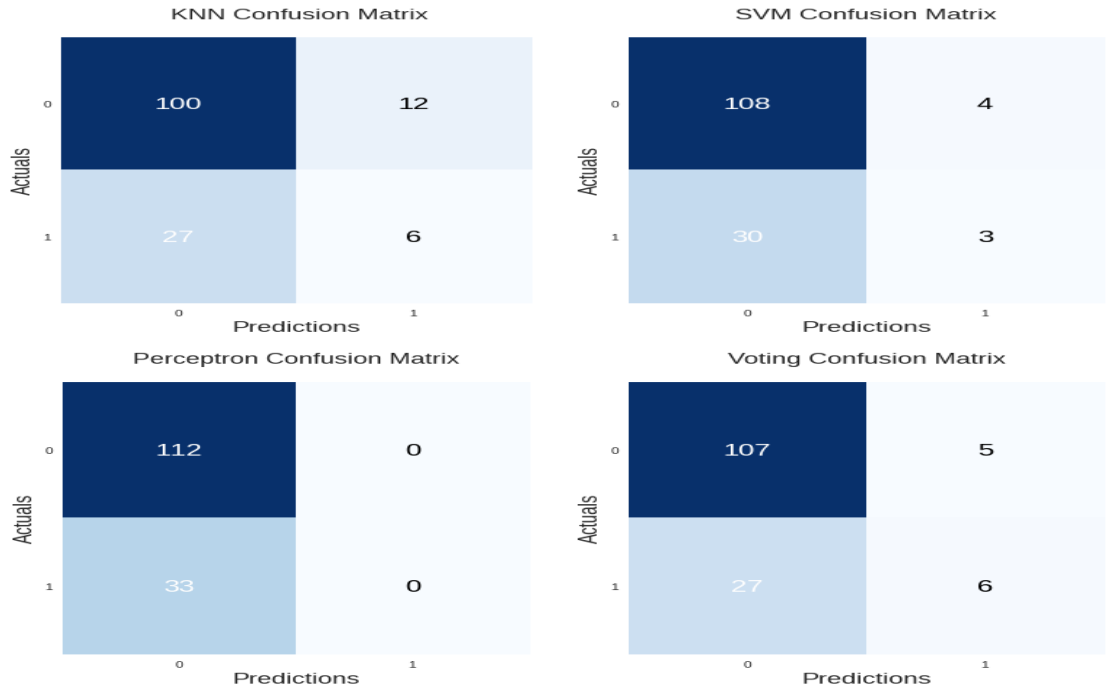

Figure 12: Model Confusion Matrixes

Figure 12 shows the confusion matrix of all the models applied for predicting liver diseases. When the TPR (total positive rate), TNR (total negative rate), FPR (false positive rate), and FNR (false negative rate) of all the models were compared, the voting classifiers' result matrix produced the best result, as the TPR and TNR ranked higher while the FPR and FNR ranked lower. and also that the voting classifier gives a greater number of correct predictions as compared to other models for the liver disease analyses.

5. Conclusion

The death rate due to liver disease has been increasing gradually since 1980 because there are not usually any visible signs or indications of liver disease. If liver symptoms usually emerge, they may include eyes and body parts that seem yellow (jaundice), bloating, and pain in the stomach. People dismiss these symptoms as very normal, preferring to avoid going to the doctor and taking standard medicines to get cures, but these things exacerbate the problem, and people only discover this at the very end, when it is difficult to cure and leads to death. But in this day of cutting-edge technology, several imaging techniques can be used to detect liver disorders. Still, there are some liver conditions that cannot be detected with scans and cause death. This paper considered the possibility of detecting liver diseases using various liver functionality tests from blood work so that each and every liver disease can be identified at the early stages. According to the paper's findings, after applying the various algorithms of machine learning and deep learning, like the Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Multilayer Perceptron (MLP), and Hard Voting Classifier (HVT), we came to the conclusion that the Hard Voting Classifier gave the best result among other models for this dataset. The Accuracy we got was 0.78 with a Specificity of 0.94, an F-score of 0.72, Precision 0.80, and Recall 0.94. Additionally, the Confusion Matrix showed that the Voting Classifier had a higher percentage of accurate predictions than the other models. So we can conclude that, for these datasets, the voting classifier gives the best result for the prediction of liver disease in a patient without taking the scans.

6. References

1. Devi, M. & Rahul, Kamma & Chowdary, Ambati & Chowday, Jampani & Manubolu, Satheesh. (2021). Principal Integrant Analysis Based Liver Disease Prediction Using Machine Learning. In Computational Vision and Bio-Inspired Computing. Springer, Singapore. 1318. 71-80. https://doi.org/10.1007/978-981-33-6862-0_6.
2. Sontakke, S., Lohokare, J., & Dani, R. (2017). Diagnosis of liver diseases using machine learning. In 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI) (pp. 129-133). IEEE. doi: 10.1109/ETIICT.2017.7977023.
3. Hassoun, S., Bruckmann, C., Ciardullo, S., Perseghin, G., Di Gaudio, F., & Broccolo, F. Setting up of a Machine Learning Algorithm for the Identification of Severe Liver Fibrosis Profile in the Asymptomatic Adult Population. Available at SSRN 4201355. http://dx.doi.org/10.2139/ssrn.4201355
4. Jiang, N., Zhao, Z., & Xu, P. (2021). Predictive analysis and evaluation model of chronic liver disease based on BP neural network with improved ant colony algorithm. Journal of Healthcare Engineering. doi: 10.1155/2021/3927551.
5. Sarvestany, Soren & Kwong, Jeffrey & Azhie, Amirhossein & Dong, Victor & Cerocchi, Orlando & Ali, Fuad & Karnam, Ravikiran & Kuriry, Hadi & Shengir, Mohamed & Candido, Elisa & Duchen, Raquel & Sebastiani, Giada & Patel, Keyur & Goldenberg, Anna & Bhat, Mamatha. (2020). Effective Detection of Compensated Cirrhosis Using Machine Learning. SSRN Electronic Journal. 10.2139/ssrn.3705267.
6. Bennett, Lucy & Mostafa, Mohamed & Hammersley, Richard & Purssell, Huw & Patel, Manish & Street, Oliver & Athwal, Varinder & Hanley, Karen & Hanley, Neil & Morling, Joanne & Guha, Indra. (2022). Using a Machine Learning Model to Risk Stratify for the Presence of Significant Liver Disease in a Primary Care Population. SSRN Electronic Journal. 10.2139/ssrn.4173873.
7. Wu, C. C., Yeh, W. C., Hsu, W. D., Islam, M. M., Nguyen, P. A. A., Poly, T. N., ... & Li, Y. C. J. (2019). Prediction of fatty liver disease using machine learning algorithms. Computer methods and programs in biomedicine, 170, 23-29. doi.org/10.1016/j.cmpb.2018.12.032.

8. Megha Kamboj, "Heart Disease Prediction with Machine Learning Approaches", International Journal of Science and Research (IJSR),Volume 9 Issue7,July 2020, pp.1454- 1458.
9. Keerthika Dulam, N. V. S., Sai Koushik, K., Sakhamuri, S., Hyndavi, C., & Sindu, G. (2022). Heart Disease Prediction with Machine Learning Approaches. In ICCCE 2021 (pp. 1155-1166). Springer, Singapore.
10. Chauhan, Y. J. (2018). Cardiovascular disease prediction using classification algorithms of machine learning. Int. J. Sci. Res. 9(5), 194-200. ISSN, 2319-7064.
11. Adeen, Preeti Sondhi. (2021). Random Forest Based Heart Disease Prediction. International Journal of Science and Research (IJSR), 10 (2). 1669-1672.
12. Kotronen, A., Peltonen, M., Hakkarainen, A., Sevastianova, K., Bergholm, R., Johansson, L. M., ... & Yki–Järvinen, H. (2009). Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. Gastroenterology, 137(3), 865-872.
13. Vijayarani, S., & Dhayanand, S. (2015). Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816-820.
14. Huang, W., Li, N., Lin, Z., Huang, G. B., Zong, W., Zhou, J., & Duan, Y. (2013). Liver tumor detection and segmentation using kernel-based extreme learning machine. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. pp. 3662-3665.
15. Shweta, Riya, Abhay Kumar. (2022). Cancer Prediction Using Machine Learning Algorithm", International Journal of Science and Research (IJSR), 11(5), 873-875, https://www.ijsr.net/ get_abstract.php?paper_id=MR22511160749
16. Sandhya Kumari, Dr. R. Viswanathan. (2020). Heart Disease Prediction System. International Journal of Science and Research (IJSR), 9(7), 425-427. https://www.ijsr.net/get_ abstract.php?paper_id=SR2053 017 1710
17. Gandhi, Monika & Singh, Shailendra. (2015). Predictions in heart disease using techniques of Machine Learning. 2015 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE 2015. 520-525. 10.1109/ABLAZE.2015.7154917.
18. Mohit Agrawal. (2020). Cancer Prediction Using Machine Learning Algorithms. International Journal of Science and Research (IJSR), 9(8), pp. 281-286, https://www.ijsr.net/get_abstract.php?paper_id=SR20801 174131
19. Tolga Ensari, Saliha, Gagaoua Oukid, Ikram and Mariam Amrane. (2018). Breast cancer classification using ML Algorithm. Biomedical Engineerings' Meeting (EBBT).
20. Keshk, A. (2022). Breast Cancer Classification Using Ml Algorithms. Kafrelsheikh Journal of Information Sciences, 3(1), 1-7.
21. Shaheamlung, Golmei & Kaur, Harshpreet & Kaur, Mandeep. (2020). A Survey on machine learning techniques for the diagnosis of liver disease. 337-341. 10.1109/ICIEM48762.2020.9160097.
22. S. Gupta, M. Sameer, and N. Mohan. (2021). Detection of Epileptic Seizures using Convolutional Neural Network," in 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 786–790, doi: 10.1109/ESCI50559.2021.9396983.
23. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data, 3(3), 25.