IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 August 2023, accepted 13 September 2023, date of publication 18 September 2023,
date of current version 22 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3316515

RESEARCH ARTICLE

Multi Perceptron Neural Network and Voting Classifier for Liver Disease Dataset

VICTOR ANTHONYSAMY AND S. K. KHADAR BABU
Department of Mathematics, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu 632014, India
Corresponding author: S. K. Khadar Babu (khadar.babu36@gmail.com)

This work was supported by the Vellore Institute of Technology, Vellore, India.

ABSTRACT The liver is one of the most significant organs in the human body. We can predict liver disease in a patient at an early stage based on previously predicted values using data from patients with abnormal liver function. Which helps the doctors to make a diagnosis. In this paper, the liver function test is analyzed for predicting liver disease, where the input of the patient's details and 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. Model Evaluation Criteria such as the Confusion Matrix, Precision Score, Recall, Accuracy, Specificity, and F-score are used to determine the best model. A dataset of 583 individuals suffering from liver disease is analyzed and we found that Hard Voting Classifier (HVC) is the best for this dataset. Additionally, this Voter Classifier prediction algorithm gives higher accuracy, which will help to diagnose liver disease.

INDEX TERMS Feed forward network, perceptron algorithm, support vector machine, voting classifier.

I. INTRODUCTION

Liver diseases are fast becoming recognized 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 [1]. The liver has roles that include detoxification, protein synthesis, and the production of chemicals that help digest food. The liver is also a gland with functions like bile production, absorbing and metabolizing bilirubin, supporting blood clots, fat metabolizing, metabolizing carbohydrates, vitamin and mineral storage, helping metabolize proteins, filtering the blood, immunological function, production of albumin, and synthesis of angiotensinogen [2]. As a result, 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, Gilberts 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) [3].

According to some data, over 10 lakh new cases of liver cirrhosis are reported each year in India, and the World Health Organization (WHO) lists liver disease as the tenth most prevalent killer there. 40% of India suffers from non-alcoholic fatty liver disease. People between the ages of 40 and 50 have liver diseases in common. Therefore, a feasible and accurate prediction of liver disease is crucial. 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 range, then it will indicate liver problems. Alanine transaminase (ALT), aspartate transaminase (AST), alkaline phosphatase (ALP), albumin, total protein, and bilirubin are known as standard liver function tests. These tests are based on several steps that generally identify the reason for liver disease and try to treat the liver infection.

Almost every other organ in the body depends on the liver, which is an important organ. The liver is vulnerable to a variety of disorders due to its enviable position and complex tasks, including

- Hepatitis is a common liver inflammatory disease. Hepatitis A, B, C, D, and E are the most prevalent viral diseases and are the most frequent cause of this.
- Toxins that the liver would typically eliminate from the bloodstream build up and induce hepatic encephalopathy. This disorder has the potential to be deadly and cause coma.

The associate editor coordinating the review of this manuscript and approving it for publication was Tony Thomas.

- Blockage of the hepatic veins, including thrombosis, which drains the liver, resulting in Budd-Chiari syndrome. The characteristic trio of abdominal discomfort, ascites, and enlarged liver characterizes its first presentation [4].
- An autoimmune condition of the liver is known as primary biliary cholangitis. Small bile ducts in the liver slowly degrade over time, with the intralobular ducts (Canals of Hering) being most severely impacted early on in the disease.
- There are numerous paediatric liver conditions as well, including as hepatic hemangiomas, Langerhans cell histiocytosis, alagille syndrome, biliary atresia, alpha-1 antitrypsin deficiency, and alagille syndrome. a benign growth the most prevalent kind of congenital liver tumour [5].

II. LITERATURE SURVEY

Comprehensive feature analysis and severity prediction of liver disease can be made using machine learning using the ANOVA test, Principal Component Analysis (PCA), and linear discriminant analysis by analyzing the performance of the classifiers [6]. Every year, liver dysfunction is the cause of more than 2.4% of fatalities in India. The early stages of liver problems can be challenging to determine due to their mild symptoms. Often, the signs don't surface until it's too late [7]. The utilization of digital revolutions in medical practice can advance efficient liver disease detection from early to late stage through a machine learning approach with defined parameters that can recognize liver illness in a normal community that is asymptomatic [8]. The best accurate prediction may be achieved by building the method of assessing chronic liver disease using an improved feed-forward neural network with back-propagation that is combined with an improved ant colony optimization [9].

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 [10]. A study on machine learning methods for detecting liver illness done by [11]. About Nine machine learning models 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 [12]. FLD might be recognized 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 [13]. Numerous machine learning techniques can be used to forecast heart disease, but only one technique will deliver the greatest accuracy in terms of results [14] and [15]. 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 [16]. Heart-related illnesses have emerged as a major cause of death worldwide over the past several decades, not just in India [17]. The latest discovery of a mutated gene can improve the predictability of non-alcoholic fatty liver syndrome [18]. Data analysis has made it simple to anticipate diseases in the healthcare industry [19].

In 3D positron emission tomography, liver tumors can be found and segmented for analysis [20]. Several study teams received assistance in determining cancer patients as high or low-risk [21]. For the purpose of diagnosing the illness, numerous knowledge techniques usually are integrated, generating numerous probabilities [22]. Naive Bayes, decision trees, and neural network algorithms were employed [23] to analyse the medical datasets. 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.

Employing 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 [24], [25]. Both developed and developing nations rank breast cancer as the leading cause of mortality. Machine learning algorithms are utilized to analyse breast cancer and provide accuracy to determine the risk that it will emerge in 8% of women [26] and [27]. When using machine learning algorithms to predict liver damage, [28] 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 [28], [29]. Three feature selection algorithms like Relief, mRMR, and LASSO were combined with seven well-known classifiers, including K-NN, ANN, SVM, NB, DT, and random forest, to identify the most crucial characteristics. The system was validated using the K-fold cross-validation method [30].

In this paper, the liver function test is analyzed for predicting liver disease, where the input of the patient's details 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.

III. METHODOLOGY

A. DATA COLLECTION

The datasets for the Indian Liver Patient Data (ILPD) of the 583 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 [31]. The study classifies liver illnesses based on age and gender to determine whether a person has liver disease or not by evaluating liver

TABLE 1. An outline of the patient dataset for liver disease.

Sl. No	Attribute Name	Attribute Type	Attribute Description
1	Age	Numeric	Age of the patient
2	Gender	Nominal	Gender of the patient
3	Total Bilirubin	Numeric	Quantity of total bilirubin in patient
4	Direct Bilirubin	Numeric	Quantity of direct bilirubin in patient
5	Alkphos Alkaline Phosphotase	Numeric	Amount of ALP exzyme in patient
6	Alamine Aminotransferase - SGPT	Numeric	Amount of SGPT in patient
7	Aspartate Aminotransferase - SGOT	Numeric	Amount of SGOT in patient
8	Total Protiens	Numeric	Protein content in patient
9	Albumin	Numeric	Amount of albumin in patient
10	Albumin and Globulin Ratio	Numeric	Fraction of albumin and globulin in patient
11	Status	$\{1, 2\}$	Status of liver disease in patient

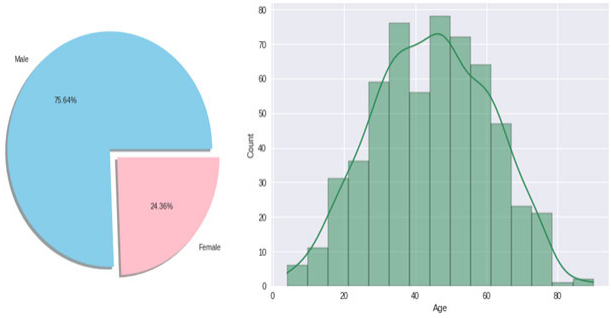

FIGURE 1. Patient gender count(Left), Grouping by age wise(Right).

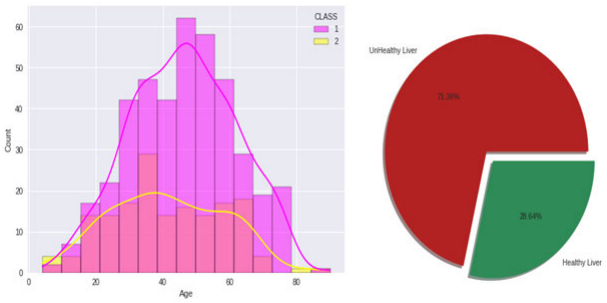

FIGURE 2. Classification of diseased person based on age(Left), Healthy and Unhealthy liver ratio(Right).

function test criteria such as Total Bilirubin, Direct Bilirubin, Alkphos Alkaline Phosphotase, Alamine Aminotransferase-SGPT, Aspartate Aminotransferase-SGOT, Total Proteins, Albumin, and Albumin and Globulin Ratio. An outline of the patient's dataset for liver disease has 583*11 attributes. Table 1 provides detailed information and descriptions for 583 instances of 11 features in the sample.

B. DATA PROCESSING

Pre-processing is required to add some necessary features, such as scaling, eliminating some irrelevant columns, and filling some blanks in the original database. After loading the data, need to attempt to visualize some data. Additionally, we will change the data format to the appropriate one using a few modifications that could be useful for categorization and visualization. It comprises data standardization and data cleaning.

C. 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 accurately. 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 *KNN-imputer* imputation techniques to make sure that the accuracy and correctness of the models were not impacted during the model building.

D. DATA STANDARDIZATION

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

$$\theta = \frac{\alpha - \mu}{\sigma} \tag{1}$$

As there are significant disparities across the ranges of the features in the input data set, standardization is used to eliminate the differences and scale the data down to a lower values. As a result, the data can be made easier to build the model and it also helps to choose the proper activation function for the perception algorithm.

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

E. K-NEAREST NEIGHBORS (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 allow it to the maximum category it belongs to. Outlier and unbalanced data will be

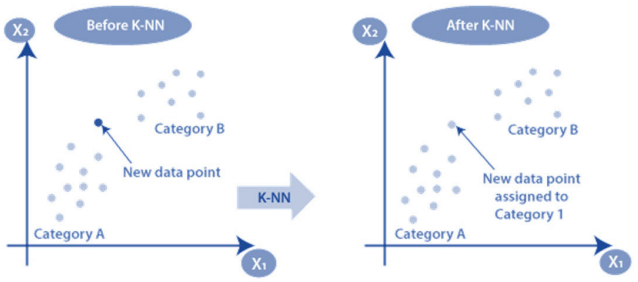

FIGURE 3. K-nearest neighbors classification.

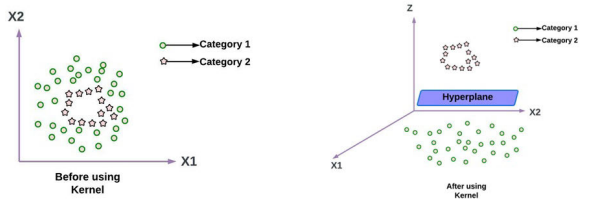

FIGURE 4. SVM Kernel Plotting by category wise.

impacted in KNN. Using the right K-neighbor value can increase accuracy. For liver disease analyses, we used the neighbor value of 4, and its work like Figure 3.

F. SUPPORT VECTOR MACHINES (SVM)

Support vector machines (SVM) learn from supervised, labeled datasets and can be used for both forecasting and prediction-based problems. SVM separates two categories with a hyperplane and creates two margins that will pass through the nearest positive and negative points in a classification, 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 higher dimensions, which makes it easy to clear points [32].

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

$$f(X_1, X_2) = exp\left(-\frac{||X_1 - X_2||^2}{2\sigma^2}\right) \tag{2}$$

where, σ- variance and $||X_1 - X_2||$− Euclidean distance between the points

G. 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 into a signal going out. It also describes the state of neurons inside their cells.

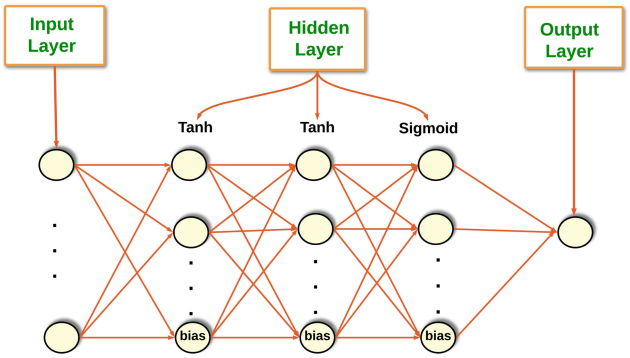

FIGURE 5. Multilayer Perceptron implementation by Layer wise.

Since our datasets have been standardized and scaled down to be between 1 and -1, we chose a 3-layer neural network with the Tan**h** and Sigmoidal activation functions to handle the highly overlapping and non-linear nature of the data which can produce better accuracy. Then its procedures are given as Figure 5.

1) Hyperbolic Tangent activation function:
Tan h 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. Additionally, Its function ranges from -1 to 1.

$$Tanh = \frac{e^{\omega} - e^{-\omega}}{e^{\omega} + e^{-\omega}} \tag{3}$$

where ω is the input parameters

2) 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.

$$Sigmoid = \frac{1}{1 + e^{-N}} \tag{4}$$

where N is the input parameters

H. PROCEDURE FOR MULTILAYER PERCEPTRON

Step 1: Create a perceptron with $n + 1$ input neurons $X = (x_0, x_2, x_3, ...x_n)$

Step 2: Where $\alpha = 1$,Which is a learning rate with the biased input.

Step 3: Initialized $\omega = (W_0, W_1, W_2, W_3, \ldots W_n)$ with the random weightss. Here, the Perceptron Convergence Theorem guarantees that we can start with any weight.

Step 4: Iterate through the input pattern ϕ_j of the training set using the weight set (Compute the weight sum of input for) $\sum_{i=0}^{n} X_i \omega_i$

Step 5: Following this calculation, the output is sent to an activation function F. which will result in the perceptron's output.

Step 6: Now, send the result to the next layer and continue steps 4 and 5 until the last layer has been reached.

Step 7: Now compute the output. Y_j using the step function

$$Y_j = F(\phi_j) = f(x) = \begin{cases} 1, & \text{if } \phi_j \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

where the threshold parameter is τ

Step 8: For each input pattern ϕ_j, compare the computed output Y_j with the desired output Y_k, if all the input patterns were successfully identified with the output, the weight is represented.

Step 9: Otherwise update the weights as follows:

1. If the computed output Y_j is 1 but should have been 0, update the weight as $\omega_i = \omega_i - \alpha * \phi_i$, where i iterate from 0 to n.

2. if the computed output Y_j is 0 but should have been 1, update the weight as $\omega_i = \omega_i + \alpha * \phi_i$, where i iterate from 0 to n. Where α is the learning rate, ω is the weight and ϕ is the input parameter

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

I. 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. In majority voting, each classifier C votes in the majority (Hard) to determine the class label y.

$$y = mode\{C_1(x), C_2(x), \ldots, C_n(x)\} \tag{5}$$

It utilizes the ensemble bagging learning technique by merging various machine learning models to achieve superior results. In comparison to a specific model, this method makes it possible to make better predictions. The voting classifier procedures are shown in Figure 6. To enhance the classification outcomes, Hard voting system is employed and used by [33].

A voting classifier is used for our datasets because the single model's results are unsatisfactory. As a result, we integrated all three models to enhance their predictive capability and performance. 1. Logistic Regression, 2. Decision Tree, 3. Support Vector Machine (SVM)

J. MODEL EVALUATION CRITERIA

1) CONFUSION MATRIX

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

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

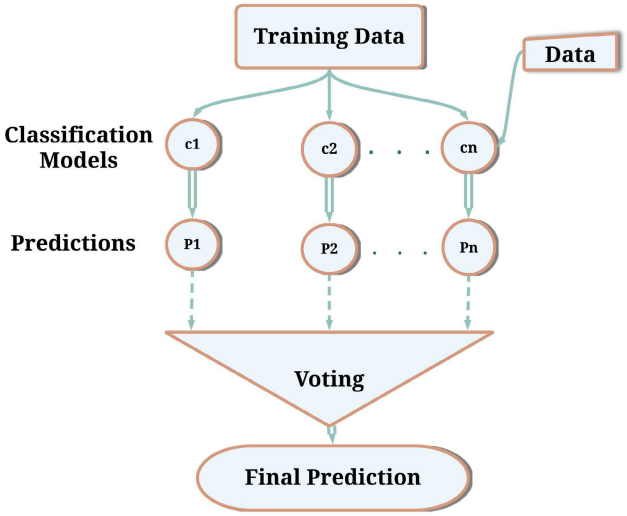

FIGURE 6. Voting classifier procedure.

4) If the model predicted a value of 1 and its actual value was also 1, the end result would be a true positive.

2) 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. Then Precision formula can be written as

$$P\omega = \frac{TP}{TP + FP} \tag{6}$$

3) 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 which gives the percentage of the positives predicted and the formula is given as

$$R\omega = \frac{TP}{TP + FN} \tag{7}$$

4) ACCURACY

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

$$A\omega = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

5) SPECIFICITY

Specificity works on the result of the negative or False prediction where it will check the True Negative by True Negative and False Positive. The specificity was calculated as

$$S\omega = \frac{TN}{TN + FP} \tag{9}$$

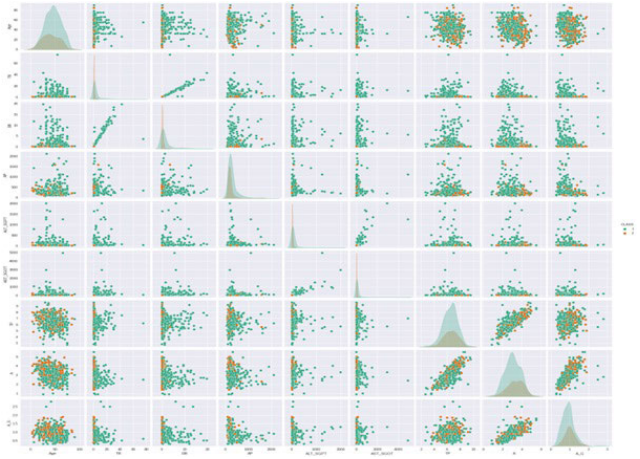

FIGURE 7. Pair-plots among MLP, SVM, KNN, and HVC.

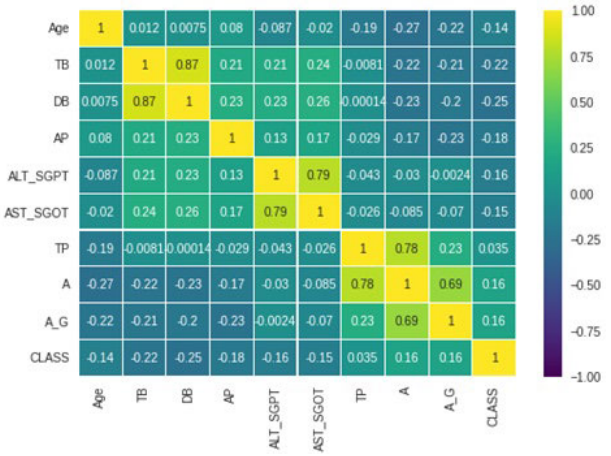

FIGURE 8. Correlation matrix based on the Attribute of the Liver diseased data set.

6) 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.

$$Fscore = 2 * \frac{P\omega * R\omega}{P\omega + R\omega} \tag{10}$$

IV. 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, and the results are predicted.

SVM models evaluate their performance and predictability, and effectiveness is determined by using model evaluation criteria (Figure 4). We put this whole algorithm into action with Python programming and the result is shown Figure 1. And it 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 2 shows the number of patients in the age ranges, 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. The total dataset consist of 583 people, where 71%, or 413 people, have an unhealthy liver and 28.64%, or 170 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 7 represents the pair plot graph, which graphically depicts the relationship between 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, and the graphs shows that points are highly overlapped and scattered, they cannot be separated using a single linear line,

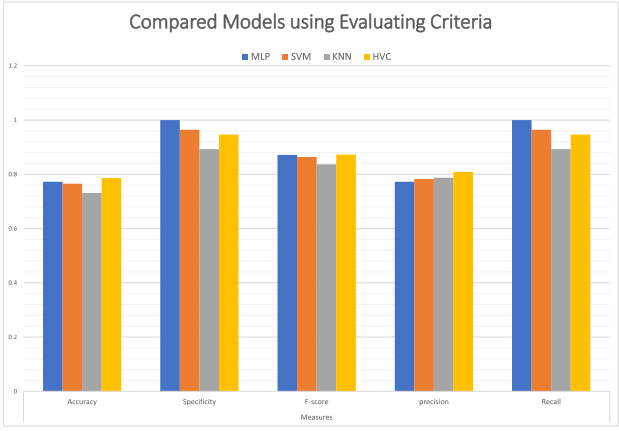

FIGURE 9. Compared models using evaluating criteria.

so we have to choose a non-linear model function for this dataset.

The Pearson correlation coefficient is between −1 and +1, denoting a perfect positive/negative correlation and 0 denoting no correlation. However, figure 8 shows that there is a correlation between the data.

Table 2 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. The voting classifier comes out on top (See Table 2). All the models are compared in Individual Models Criteria Comparison in Figures 9 and 10.

Figure 11 shows the confusion matrix of all the models 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,

TABLE 2. Result of models using various evaluating criteria.

Models	Measures				
	Accuracy	F -score	Precision	Recall	Specificity
MLP (11, 9, 1)	0.7724	0.8715	0.7724	1	1
SVM (ker = RBF, c=100, g=0.0001)	0.7655	0.864	0.7826	0.9643	0.9643
KNN (K-NN, K=4)	0.7310	0.8368	0.7874	0.8929	0.8929
HVC (ker = RBF)	0.7862	0.8724	0.8092	0.9464	0.9464

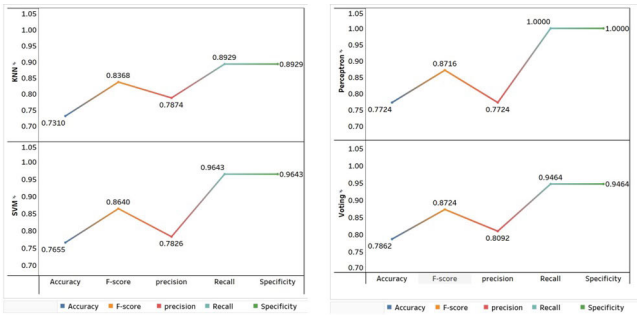

FIGURE 10. Individual models vs criteria comparison.

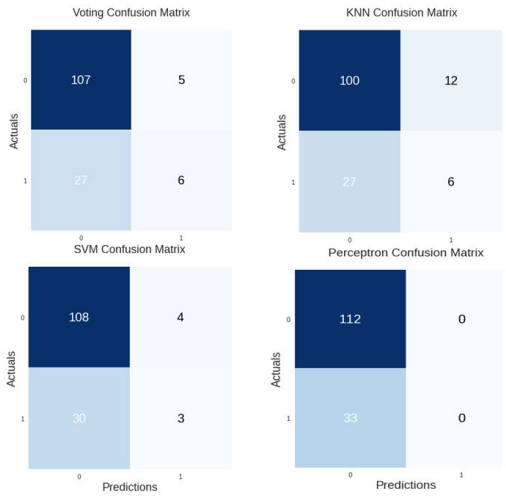

FIGURE 11. Model confusion Matrices among MLP, SVM, KNN and HVC.

the voting classifiers produced the best result, as the TPR and TNR ranked higher while the FPR and FNR ranked lower. Additionally, the voting classifier gives a greater number of correct predictions as compared to other models for liver disease analyses.

The Hard Voting Classifier (HVC) performed better than the other four classifiers in terms of specificity, precision, F-score, recall, and accuracy (94%, 80%, 87%, and 78%, respectively). The Multilayer Perceptron (MLP) produces output after being trained on various input and hidden layers. The result is then produced with 11 inputs, 9 hidden neuron units, and 1 unit in the final layer. The MLP was the second-best classifier, achieving 77% accuracy, 100% specificity, 87% F-Score, 77% precision, and 100% recall.

The SVM kernel RBF achieves 96% specificity, 76% accuracy, and 96% recall at C = 100 and g = 0.0001. The K-Nearest Neighbor (KNN), which possesses a specificity 89%, recall 89%, F-score 83%, precision 78%, and accuracy 73%.

V. 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 study 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 research findings, after applying the various algorithms of machine learning and deep learning, like the Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP), and Hard Voting Classifier (HVC), Concluded 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, a F-score of 0.87, Precision of 0.80, and Recall of 0.94. Additionally, the Confusion Matrix showed that the Voting Classifier has more accurate predictions than the other models. Therefore, we can conclude that, for these datasets, the voting classifier gives the best result for the prediction of liver disease. In the future, In order to enhance the Hard Voting classifier's performance for identifying a different disease, we'll carry out additional experiments in the future using other feature selection algorithms and optimization strategies.

REFERENCES

[1] P. J. Thuluvath, A. Saraya, and M. Rela, "An introduction to liver disease in India," *Clin. Liver Disease*, vol. 18, no. 3, pp. 105–107, Sep. 2021, doi: 10.1002/cld.1149.
[2] J. C. Ozougwu, "Physiology of the liver," *Int. J. Res. Pharmacy Biosciences*, vol. 4, no. 8, pp. 13–24, 2017.
[3] H. Devarbhavi, S. K. Asrani, J. P. Arab, Y. A. Nartey, E. Pose, and P. S. Kamath, "Global burden of liver disease: 2023 update," *J. Hepatology*, vol. 79, no. 2, pp. 516–537, Aug. 2023.
[4] G. M. Hirschfield and M. E. Gershwin, "The immunobiology and pathophysiology of primary biliary cirrhosis," *Annu. Rev. Pathol., Mech. Disease*, vol. 8, no. 1, pp. 303–330, Jan. 2013, doi: 10.1146/annurev-pathol-020712-164014.
[5] R. Saxena and N. Theise, "Canals of hering: Recent insights and current knowledge," *Seminars Liver Disease*, vol. 24, no. 1, pp. 43–48, 2004, doi: 10.1055/s-2004-823100.
[6] M. Devi, K. Rahul, A. Chowdary, J. Chowday, and S. Manubolu, "Principal integrant analysis based liver disease prediction using machine learning," in *Computational Vision and Bio-Inspired Computing*. Singapore: Springer, 2021, pp. 71–80.
[7] S. Sontakke, J. Lohokare, and R. Dani, "Diagnosis of liver diseases using machine learning," in *Proc. Int. Conf. Emerg. Trends Innov. ICT (ICEI)*, 2017, pp. 129–133, doi: 10.1109/ETIICT.2017.7977023.
[8] S. Hassoun, C. Bruckmann, S. Ciardullo, G. Perseghin, F. Di Gaudio, and F. Broccolo, "Setting up of a machine learning algorithm for the identification of severe liver fibrosis profile in the general US population cohort," *Int. J. Med. Informat.*, vol. 170, Feb. 2023, Art. no. 104932.
[9] N. Jiang, Z. Zhao, and P. Xu, "Predictive analysis and evaluation model of chronic liver disease based on BP neural network with improved ant colony algorithm," *J. Healthcare Eng.*, vol. 2021, pp. 1–7, Nov. 2021, doi: 10.1155/2021/3927551.
[10] S. S. Sarvestany, J. C. Kwong, A. Azhie, V. Dong, O. Cerocchi, F. A. Ali, R. S. Karnam, H. Kuriry, M. Shengir, E. Candido, R. Duchen, G. Sebastiani, K. Patel, A. Goldenberg, and M. Bhat, "Development and validation of an ensemble machine learning framework for detection of all-cause advanced hepatic fibrosis: A retrospective cohort study," *Lancet Digital Health*, vol. 4, pp. e188–e199, 2022, doi: 10.1016/S2589-7500(21)00270-3.
[11] G. Shaheamlung, H. Kaur, and M. Kaur, "A survey on machine learning techniques for the diagnosis of liver disease," in *Proc. Int. Conf. Intell. Eng. Manag.*, 2020, pp. 337–341, doi: 10.1109/ICIEM48762.2020.9160097.
[12] L. Bennett, H. Purssell, O. Street, K. Piper Hanley, J. R. Morling, N.A. Hanley, V. Athwal, and I. N. Guha, "Health technology adoption in liver disease: Innovative use of data science solutions for early disease detection," *Frontiers Digit. Health*, vol. 4, 2022, Art. no. 737729.
[13] C. C. Wu, W. C. Yeh, W. D. Hsu, M. M. Islam, P. A. Nguyen, T. N. Poly, and Y. C. J. Li, "Prediction of fatty liver disease using machine learning algorithms," *Comput. methods Programs Biomed.*, vol. 170, pp. 23–29, Mar. 2019.
[14] M. Kamboj, "Heart disease prediction with machine learning approaches," *Int. J. Sci. Res. (IJSR)*, vol. 9, no. 7, pp. 1454–1458, 2020.
[15] N. V. S. K. Dulam, K. S. Koushik, S. Sakhamuri, C. Hyndavi, and G. Sindu, "Heart disease prediction with machine learning approaches," in *Proc. 4th Int. Conf. Commun. Cyber Phys. Eng.* Cham, SwitzerlandSingapore: Springer, 2022, pp. 1155–1166.
[16] Y. J. Chauhan, "Cardiovascular disease prediction using classification algorithms of machine learning," *Int. J. Sci. Res*, vol. 9, no. 5, pp. 194–200, 2018.
[17] P. S. Adeen, "Random forest based heart disease prediction," *Int. J. Sci. Res. (IJSR)*, vol. 10, no. 2, pp. 1669–1672, 2021.
[18] A. Kotronen, M. Peltonen, A. Hakkarainen, K. Sevastianova, R. Bergholm, L. M. Johansson, N. Lundbom, A. Rissanen, M. Ridderstråle, L. Groop, M. Orho-Melander, and H. Yki-Järvinen, "Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors," *Gastroenterology*, vol. 137, no. 3, pp. 865–872, Sep. 2009.
[19] S. Vijayarani and S. Dhayanand, "Liver disease prediction using SVM and Naive Bayes algorithms," *Int. J. Sci., Eng. Technol. Res. (IJSETR)*, vol. 4, no. 4, pp. 816–820, 2015.
[20] W. Huang, N. Li, Z. Lin, G.-B. Huang, W. Zong, J. Zhou, and Y. Duan, "Liver tumor detection and segmentation using kernel-based extreme learning machine," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2013, pp. 3662–3665.
[21] R. Shweta and K. Abhay, "Cancer prediction using machine learning algorithm," *Int. J. Sci. Res. (IJSR)*, vol. 11, no. 5, pp. 873–875, 2022.
[22] K. Sandhya and R. Viswanathan, "Heart disease prediction system," *Int. J. Sci. Res. (IJSR)*, vol. 9, no. 7, pp. 425–427, 2020.
[23] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," in *Proc. Int. Conf. Futuristic Trends Comput. Anal. Knowl. Manag. (ABLAZE)*, Greater Noida, India, Feb. 2015, pp. 520–525, doi: 10.1109/ABLAZE.2015.7154917.
[24] M. Agrawal, "Cancer prediction using machine learning algorithms," *Int. J. Sci. Res. (IJSR)*, vol. 9. no. 8, pp. 281–286, 2020.
[25] K. Babu, C. Chesneau, V. Anthonysamy, S. Ravichandaran, and S. C. Vasan, "Efficacy of the rice crop growth using different smoothing methods," *J. Appl. Anal. Comput.*, vol. 12, no. 6, pp. 2593–2599, 2022, doi: 10.11948/20220116.
[26] M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," *Eng. Appl. Sci. Res.*, vol. 48, no. 1, pp. 65–72, Jan. 2021.
[27] S. Shehab and A. Keshk, "Breast cancer classification using ML algorithms," *Kafrelsheikh J. Inf. Sci.*, vol. 3, no. 1, pp. 1–7, Jun. 2022.
[28] S. Gupta, M. Sameer, and N. Mohan, "Detection of epileptic seizures using convolutional neural network," in *Proc. Int. Conf. Emerg. Smart Comput. Informat.*, Mar. 2021, pp. 786–790, doi: 10.1109/ESCI50559.2021.9396983.
[29] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, Jul. 2018.
[30] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Inf. Syst.*, vol. 2018, pp. 1–21, Dec. 2018, doi: 10.1155/2018/3860146.
[31] D. Dua and C. Graff, "UCI machine learning repository," University of California, School of Information and Computer Science Irvine, CA, USA, 2019. [Online]. Available: http://archive.ics.uci.edu/ml
[32] J. Joachims, *Learning to Classify Text Using Support Vector Machines*. Norwell, MA, USA: Kluwer, 2002.
[33] X.-B. Kang, G.-F. Lin, Y.-J. Chen, F. Zhao, E.-H. Zhang, and C.-N. Jing, "Robust and secure zero-watermarking algorithm for color images based on majority voting pattern and hyper-chaotic encryption," *Multimedia Tools Appl.*, vol. 79, nos. 1–2, pp. 1169–1202, Jan. 2020.

VICTOR ANTHONYSAMY received the B.Sc. and M.Sc. degrees in mathematics from the Sacred Heart College (Autonomous), Tirupattur, Tamil Nadu, India. He is currently pursuing the Ph.D. degree with the Vellore Institute of Technology (VIT), Vellore, India. He has eight years of experience in teaching and research. His research interests include stochastic processes, time series analysis, forecasting models, and advanced statistical modeling.

S. K. KHADAR BABU is currently an Associate Professor Senior with the Department of Mathematics, School of Advanced Sciences, VIT University, Vellore, Tamil Nadu, India. He has 25 years of teaching and research experience. He has published about 86 research articles in various national and international journals and presented about 120 articles at national and international conferences, seminars, symposiums, and workshops. He organized a faculty development program on probability and statistics for engineers. His research interests include statistical quality control, statistical modeling, time series analysis, computer networks, and digital image processing. He is a member of different societies.