Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 8s (2024)

A Novel Multi-Stage Feature Selection and Classification Model for Accurate Liver Disease Detection

Gurmeet Kaur Saini[1], Sachin Ahuja[2], Vishal Bharti[3]

[1] Department of Computer Science and Engineering , Chandigarh University, Punjab, India
Email: gurmeetsaini02@gmail.com

[2,] Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India

[3]Department of Computer Science and Engineering, Maharishi Markandeshwar (Deemed to be University) Mullana, Ambala, Haryana

Corresponding Author: gurmeetsaini02@gmail.com

Article History:

Received: *26-04-2024*

Revised: *16-06-2024*

Accepted: *28-06-2024*

Abstract:

Liver disease, comprising a spectrum of conditions that afflict the liver, stands as a significant health challenge with global implications. This paper presents an effective and highly accurate liver disease detection model wherein unique feature Selection and classification techniques are implemented. The novelty contribution of this work is effective feature selection technique in which features are selected in three stages. In the first stage, entropy, Eigenvector centrality and entropy-correlation based techniques are implemented on each feature to determine their relevance score. These features are then combined and their average value is calculated to form the first feature score vector. In the second stage, Randon Forest (RF) classifier is used for analysing the effectiveness of features selected in first stage in terms of their accuracy. Based on this accuracy, the relevance score of features is again updated to form the second feature set. In the third stage, Fuzzy model is used for determining the contextual relevance among various features. The selected feature set is then passed to proposed BELV classification model, wherein techniques like bagging, ensemble learning and voting is applied. Three baseline classifiers i.e., KNN, RF and Decision Tree (DT) are used in ensemble learning to make individual predictions which are then combined before applying majority voting mechanism to make the final prediction. The efficacy of proposed model is tested on ILPD and CPD datasets for binary classification and multi-stage disease detection respectively. Through extensive experiments in MATLAB software proposed model attained an accuracy of 93% and 96.8% for binary and multi-stage disease classifications respectively.

Keywords: Liver Disease Detection, Learning Methods, Deep Learning Methods, Meta-Heuristic Approaches, Optimization Algorithms, Medical Science.

1. Introduction

Liver is considered as the largest and strongest part of human body that plays a crucial role in maintenance of haemostasis and coagulation process [1]. It is surrounded by the rib cage that is located in the upper right region of the belly. Beneath this grand organ in human body, some other small organs like pancreas, intestines and gallbladder are located. The liver works on a variety of intricate body processes and its weight is around 3 pounds. The two main structural divisions of the liver are right lobe and left lobe. Liver is responsible for collaborating with other organs to break down, metabolize and assimilate food. Additionally, the liver produces protein molecules which are necessary for blood

coagulation and various other processes [2,3]. The phrase "liver disease" is broad and encompasses any possible issues that could prevent the liver from functioning normally or from performing its assigned tasks. Each symptom is dependent on the type and severity of the ailment. Figure 1 shows the pictorial form of normal and infected liver.

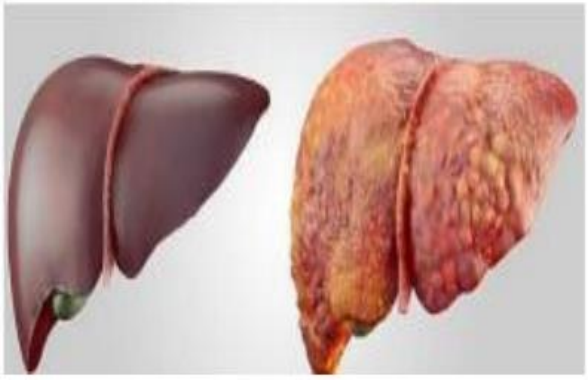

Figure 1. Normal and infected Liver [4]

One of the most prevalent types of liver disease that has been seen around the globe consistently is Fatty Liver Disease or FLD. This FLD is sometimes also known as Fatty liver steatosis and is very difficult to diagnose [5]. It is important to detect FLD disease in earliest possible stages because if its not identified timely it may lead to cirrhosis, cancers, steatohepatitis. Other than this, it may also case liver damage and result in acute failures in liver hence, early diagnosis and therapy are therefore crucial for the management of FLD. However, the complexity of the liver disease detection arises with the rise in covid-19 pandemic. Moreover, because of this pandemic the sensitivity and fragility corresponding to liver disease also increases among patients [7,8].

2. Literature Review

Liver diseases encompass a spectrum of medical conditions affecting the liver's structure and function. Ranging from viral hepatitis and fatty liver disease to cirrhosis and hepatocellular carcinoma, these conditions can lead to serious health complications if not diagnosed and managed promptly. Leveraging AI's ability to analyse vast datasets and recognize complex patterns, researchers have developed predictive models that aid in identifying individuals at risk or in early stages of liver disease. By accessing prominent academic platforms such as IEEE, Springer, and Elsevier, this review seeks to explore the current landscape of AI-based models for predicting liver diseases. Through targeted keyword searches and careful analysis of published research, we aim to gain insights into the state-of-the-art methodologies, challenges, and potential directions in this evolving field.

The results were further enhanced by authors in [19] which developed a ML technique that was based on RF classifiers for predicting disease. Also, they implemented univariate and bivariate analysis for checking the skewness and outliers of data. For further balancing the data, different oversampling and under sampling techniques are implemented. Also, the performance of proposed approach was developed by optimizing the hyper parameters by utilizing grid search and FS which achieves an accuracy of 100%. Again in [20], a hybrid soft computing technique is proposed for detecting liver disease at earliest stages. They used Modified WSO approach for selecting the important features from available feature set. For determining the stage of liver disease in patients they implemented HSSI-

DNN model that was tested on three datasets i.e., BUPA, ILPD and MRRLPD. Results showcased that proposed model achieves an accuracy of 83% on BUPA, 84% on ILPD and 91% on MPRLPD databases. Also, the authors in [21], utilized 5 ML classifiers (SVM, NB. KNN, LDA and CART) for predicting liver diseases. Through extensive experimentation, it was observed that KNN attains highest accuracy of 91.7% while as, it was 92.1% for auto encoder model.

3. Materials And Methods

Here, a novel feature selection technique is proposed for selecting only highly informative and critical features that aid in improving overall detection accuracy of the system. This technique helps to overcome complexity and overfitting issues in current databases that may contain redundant or irrelevant features which has no role in disease detection. All these techniques (Pre-processing, FS and Classification) are implemented on ILPD (Indian Liver Patient Dataset) and CPD (Cirrhosis Prediction Dataset), whose detailed information is given in subsequent sections of this paper. It is

pertinent to mention here that two datasets i.e., ILPD and CPD are used to perform binary and multi-stage disease classifications respectively. Figure 2 demonstrates the architecture of proposed liver disease detection model.

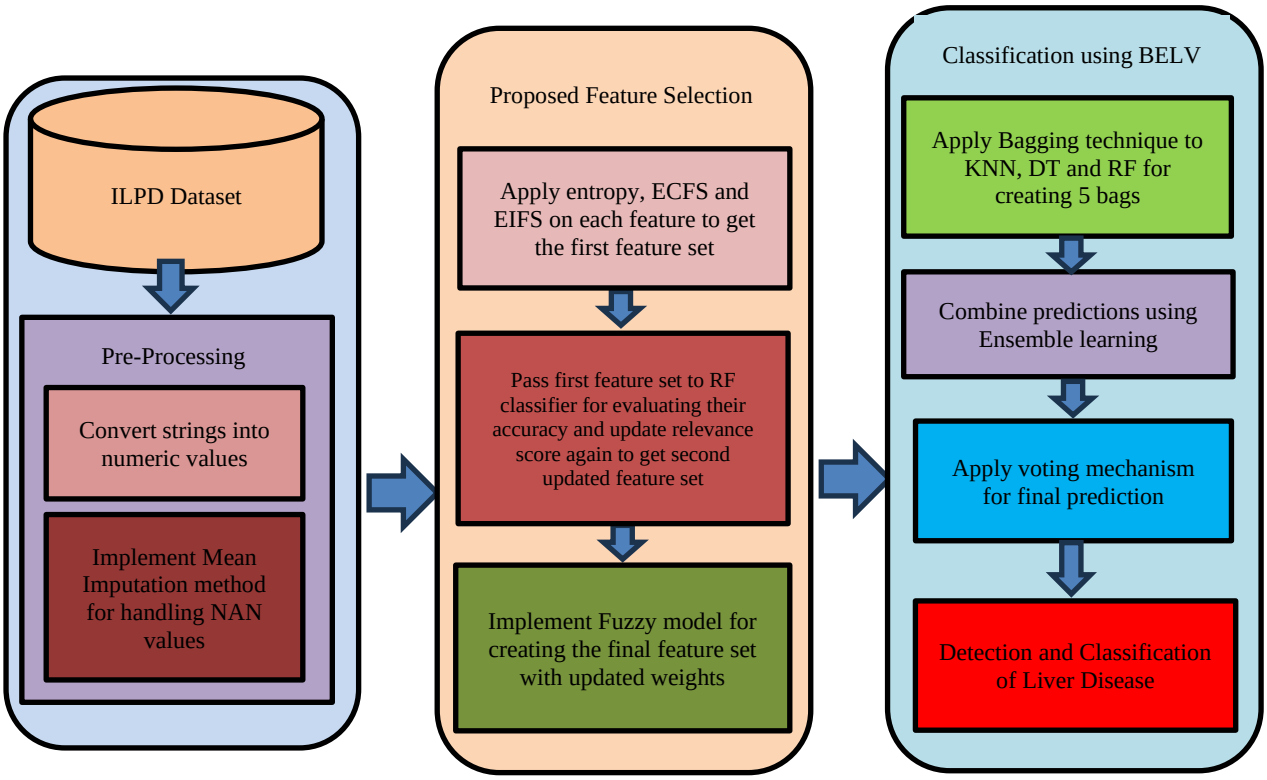

Figure 2. Proposed Architecture for Liver disease detection

3.1 Dataset And Its Preparation

In our work, we have considered two datasets i.e., ILPD and CPD for detecting liver disease and determine its stages respectively. The brief description of these two datasets is given below along with their attribute information.

3.1.1 ILPD Dataset

ILPD is one of the frequently utilized datasets for detecting liver diseases. The information of the dataset is publicly accessible on Kaggle.com via https://www.kaggle.com/datasets/uciml/indian-liver-patient-records. A total of 10 features and one target attribute are present in the given dataset, whose information is given in Table 1.

Table 1: ILPD Dataset Attribute information

S.no	Feature Name	Type	Description
1	Age	Feature	Represents age of individual
2	Gender	Feature	Represents sex of individual
3	TB	Feature	Total Bilirubin
4	DB	Feature	Direct Bilirubin
5	Alkphos	Feature	Alkaline Phosphate
6	Sgpt	Feature	Alamine Aminotransferase
7	Sgot	Feature	Aspartate Aminotransferase
8	TP	Feature	Total Proteins
9	ALB	Feature	Albumin
10	A/G Ratio	Feature	Albumin and Globulin Ratio
11	Selector	Target	Used for separating data into two sets

3.1.2 CPD Dataset

The second dataset used in proposed work is CPD whose data is accessible on https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset.

Table 2: CPD Dataset Sample

ID	N_Days	Bilirubin	Cholesterol	Albumin	Triglycerides	Platelets	Prothrombin	Stage
1	400	14.5	261	2.6	172	190	12.2	4
2	4500	1.1	302	4.14	88	221	10.6	3
3	1012	1.4	176	3.48	55	151	12	4
4	1925	1.8	244	2.54	92	183	10.3	4
5	1504	3.4	279	3.53	72	136	10.9	3
6	2503	0.8	248	3.98	63		11	3
7	1832	1	322	4.09	213	204	9.7	3
8	2466	0.3	280	4	189	373	11	3
9	2400	3.2	562	3.08	88	251	11	2

Since the datasets contains information in string format which is not recognized by our classifiers, therefore, we have first implemented label encoder technique in the pre-processing stage that converts the string attributes like male and female into numeric values of 0 and 1 respectively. This helps in enhancing feature engineering process as the proposed model can efficiently recognize patterns and relationship among various attributes. Secondly, the NAN values present in the dataset are handled by applying a Mean imputation technique, because we are dealing with numeric data now.

4. Proposed Feature Selection Technique

Feature Selection (FS) is one of the crucial steps in the process of liver disease detection, wherein informative feature set is created by selecting only crucial and informative features that aid in enhancing the accuracy of model. Moreover, it also aids in addressing the dimensionality issues which can lead to overfitting when dealing with high dimensional data. Traditionally, various FS techniques were implemented by researchers in their respective works for selecting important features in liver disease datasets, however, the problem is that majority of these techniques selected features based on their weights generated by employing single technique. This makes the model computational complex and time consuming as the model has to undergo through number of iterations to attain a feature subset with best weights. In realm to this limitation, a novel and unique FS technique is proposed in this paper that can not only enhance the accuracy of detection but also addresses above mentioned limitations. The proposed FS technique is different from previous techniques in the fact that it calculates the relevance score of features in three stages by using different techniques (as shown in figure 3). In the first stage, relevance score of each feature is obtained by employing entropy and eigenvector centrality methods. Moreover, an Enhanced Infinite Feature Selection (EIFS) is also proposed at this stage, which is basically an extension of standard IFS method. The standard IFS is enhanced in proposed work by introducing the concept of entropy and correlation in it, for evaluating the relevance score of features present in ILPD and CPD datasets. In the next stage of Feature selection, a ML classifier is utilized for assessing the features selected in first stage and updating their relevance score by calculating their accuracy values. Finally, in the last stage of FS, a fuzzy system is introduced for evaluating the contextual relevance among various features to make the final feature set that will be used for training the classification model. The detailed description of each FS stage is explained below.

Stage 1 of FS

The process of selecting important and informative features in proposed work starts by defining some basic parameters of the model that will be used in subsequent stages. The value of these parameters is mentioned in Table 2. Initially, entropy-based method is implemented on each feature to measure the amount uncertainty in given feature set. The formula used for calculating the feature score entropy of features in proposed work is given in equation 1.

$$H(f) = \sum_{i=1}^{n} p_i \log_2 p_i \qquad (1)$$

By using above given equation, scores of the features were obtained which are then ranked in the descending order to select more effective features based on their weight value. The features with less entropy values are ranked higher and hence are more relevant to our work. After this, the second method i.e., ECFS is applied on each feature of dataset to form another set of features with different weights or relevance scores. ECFS is basically a graph-based approach that determines the impact of each feature in the model. Considering a featured graph G= (V, E), having V vertices and A=($a_{v,t}$) be its adjacent matrices, with $a_{v,t}$= 1. On the other hand, of vertex v is connected to "t" vertex and value of $a_{v,t}$ is zero, then eigenvector centrality can be calculated by using equation 2.

$$x_v = \frac{1}{\lambda}\sum_{t\in M(v)} x_t = \frac{1}{\lambda}\sum_{t\in G} a_{v,t}\, x_t \qquad (2)$$

Wherein, M(v) depicts the neighbor set of v and λ is constant.

Therefore, by implementing the formula 2 on each feature, we calculated the relevance score of features and only those features which are central in the feature graph are considered relevant.

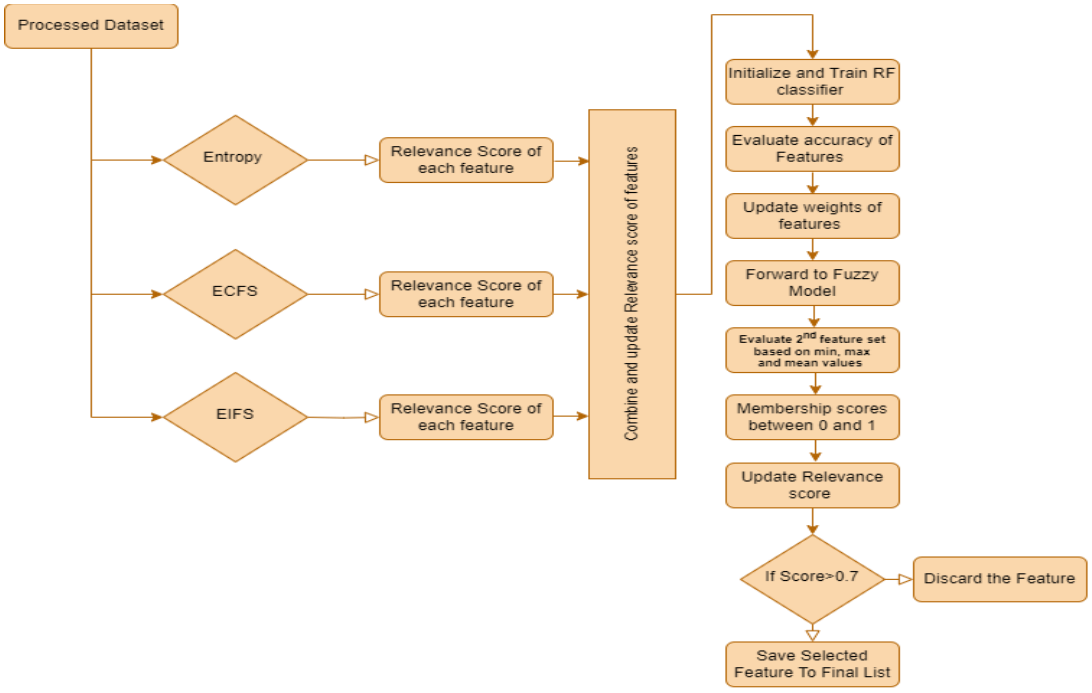

Figure 3. Flowchart of proposed Feature Selection Technique

Furthermore, we have calculated the relevance score of features by enhancing the standard IFS method. Conventionally, we have observed that features in IFS were selected either on entropy based or correlation based, but in our work, both entropy and correlation are used for calculating weights of features. In proposed EIFS, weights are assigned to features by combining their entropy and correlation values. The formula for entropy is same as given in equation 1. However, the correlation of features is computed by using equation 3. It must be noted that correlation among features and target variables is measures by their mutual information factor (I).

$$I = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \qquad (3)$$

Once the scores of each feature is obtained through entropy, ECFS and EIFS method, we proceed towards next step of updating relevance score of features. For this, we considered the initial scores of entropies, ECFS and EIFS and calculated their average value by using equation 4.

$$RelevanceScore\ 1 = \frac{Entropy\ score + ECFS\ score + EIFS\ score}{3} \qquad (4)$$

Wherein relevance score 1 represents the feature set obtained at first stage of FS. This feature set determines the importance of features based on multiple criteria of entropy, ECFS and EIFS. After this, the feature with highest relevance scores is grouped together to form an updated feature list. It must be noted here, that this feature set is basically ordering the features as per their relevance scores obtained by using equation 5.

Stage 2 of FS

In the second stage of proposed FS method, a ML classifier is utilized for calculating the weight of features. The main reason for doing so is to make available feature set accuracy dependent and enhance reliability of system. Here, we have used Random Forest (RF) classifier which is one of the widely used ML algorithm that can be used for both classification and regression tasks. The classifier is initialized by defining the parameters given in Table 3.

Table 3: RF initialization parameters

Parameter	Values
Alpha	0.5
No of Bagger	100
No of trees	10
Method to work	Classification
Prediction method	OOB

Once the RF classifier is initialized, the ordered feature set formed in first stage is passed to it for training purpose. The RF classifier analyzes features and starts getting trained by employing the Tree Bagger function. After this, "OOB" prediction method is employed for retrieving the feature scores. These scores are calculated based on out-of-bag samples and are a measure of how much each feature contributes to the accuracy of the RF model. The process keeps on repeating iteratively till we get the best accuracy values. After each iteration, the current accuracy value of RF is compared with previous one and if it comes out to be better, than accuracy value is updated, otherwise it remains same. The equation for updating the relevance score of features is given in equation 5.

$$RSu = RSp + Nacc * ENwgt \quad (5)$$

Wherein, Rsu and Rsp represents updated relevance score and previous relevance score. While as, Nacc and ENwgt represents the accuracy and entropy-based weights of features respectively. This RSu represents the second feature set that contains more relevant and important features than first feature subset.

Stage 3 FS

In the third and final stage of our work, we have improved the RSu score further by making it dependent on whole set of features using Fuzzy system. Fuzzy system is a model wherein a set of features are given as input to the model which are then evaluated by per some defined rules to get the single outcomes determining membership degree of a feature. In our case, the RSu feature weights serves as input to the proposed fuzzy system, which defines membership functions of features weights based on their minimum, maximum and mean values to introduce contextual relevance among features. This signifies that membership functions are sensitive to overall distribution and spread of feature weights. This means that features with significantly different weights than mean or close to min or max values will receive distinctive membership scores to depict their relevancy. Moreover, another reason for introducing fuzzy in our work is to handle uncertainties or noises present in datasets. The fuzzy model allows features to have partial membership in multiple membership functions to receive

a moderate membership score, even if its weight is extremely low or high. In other words, we can say that, fuzzy system enhances feature weighting by incorporating context and uncertainty into the process. It considers how each feature's weight relates to the distribution of weights across all features, resulting in a more context-aware and interpretable evaluation of feature importance. This can lead to more meaningful and reliable feature selection, particularly in situations where the importance of features may vary or is difficult to determine using traditional methods.

The fuzzy system starts evaluating each feature weight as per the value of membership functions to obtain a single output of membership score. The value of this membership score ranges from 0 to 1, which determines the importance or relevancy of each feature with previous feature sets. After this, the feature weights were combined with the previous feature set to get the final set. For this, each features membership score is multiplied by corresponding feature weight obtained in previous step. Therefore, the final feature weights are obtained by using equation 6.

$$Feature\ Weight = membership\ degree\ or\ score\ \times feature\ weights\ of\ RSu \quad (6)$$

This feature weight depicts the final score of features but they are not selected entirely in the proposed model. Instead, we have set a threshold value of 0.7, which acts as a filter for selecting final features. The value of threshold is 0.7 because we have observed from literature survey that the feature should have at least 70% relevancy for predicting the liver disease with high accuracy. The feature weight obtained in equation 6 are then analyzed based on this threshold value and only those features having relevancy more than 0.7 are selected, rest are discarded. Equation 7 shows the mathematical equation for selecting the final features in proposed model.

$$FfeatSet = Feture\ Weight > thershold\ 0.7 \quad (7)$$

Wherein, FfeatSet represents the final feature set obtained after processing it through three stages of selection. This final feature set is then passed to classifier for making the final prediction regarding the presence or absence of liver disease among patients. The process of proposed Feature selection is defined in Algorithm 1.

Algorithm 1: Feature Selection process

Input

- Training dataset with features and labels
- Set parameters like alpha for ECFS and EIFS

Calculate Relevance Score

- Calculate entropy of feature H(f) by using below equation

$$H(f) = \sum_{i=1}^{n} p_i \log_2 p_i$$

- Rank H(f) features in descending order of their entropy values

Calculate relevance score by ECFS

- Calculate EC of features by using ECFS method

$$x_v = \frac{1}{\lambda}\sum_{t\in M(v)} x_t = \frac{1}{\lambda}\sum_{t\in G} a_{v,t}\, x_t$$

Entropy-correlation based IFS

- Calculate the relevance score of features by using EIFS

$$I = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

Update Relevance score

- Relevance score of features is updated by using below equation

$$RelevanceScore\ 1 = \frac{Entropy\ score + ECFS\ score + EIFS\ score}{3}$$

Iterative Feature Selection

- **For each iteration**

- Identify the feature with highest relevance score:
  Max_score_feature=arg $\max_{\text{feature}}$ relevance_score
- Add this feature to new list of features (ordered by relevance score):
Selected feature=add max features values

ML based feature weighing

- Initialize best accuracy: best_accurcay=0.0
- Initialize best weight: best_weights=initial weight

While True:

- Train ML model with selected features: model=trainRFModel(selected_features_labels)
- Evaluate model and calculate accuracy:
Accuracy=evaluate Model (model, selected features, labels)
- Check if new accuracy is better than previous best accuracy:
if accuracy> best accuracy
- Update best accuracy: best_accuracy=accuracy
- Update best weights from the model:
best weight=getFeatureWeights(model)

else

exit loop

Accuracy dependent Weight Assignment

- Update relevance score RSu for each feature

$$RSu = RSp + Nacc * ENwgt$$

Fuzzy Membership Functions

- For each weight, define fuzzy membership functions based on min, max and mean values
- For each feature, evaluate its membership in each defined fuzzy membership function, to form set of membership score for each feature.
- Multiply each feature membership score with corresponding feature weight

$$Feature\ Weight = membership\ degree\ \times feature\ weights$$

Apply threshold-based FS

- FfeatSet=Feature Weights>Th

Where,

- FfeatSet is final feature set and TH is threshold with value of 0.7

Return **FfeatSet**

4.1 Disease Detection Using BELV

The features extracted in the previous step are categorized into two categories of training and testing data in the proportion of 70:30 respectively. In this section of paper, we are going to identify and determine the stage of liver disease in patients by employing bagging, Ensemble Learning and Voting methods, (BELV) on ILPD and CPD datasets. Figure 4 demonstrates the classification mechanism opted in our work. Bagging which is sometime also known as Bootstrap Aggregating creates multiple datasets of the training data through random sampling with replacement and training individual models on featured data subsets. On the other hand, ensemble learning is the classification method in which two or more classifiers are used together for making prediction. The final prediction is made by using the majority voting mechanism of ensemble methods. Traditionally, bagging technique was implemented on single classifiers to make the final prediction. However, to introduce the novelty concept in our model and enhance its accuracy rate, we have not only implemented bagging method but have also used ensemble learning method and Mod voting mechanism for making the final prediction. There are ample number of reasons why ensemble learning is preferred over standard prediction models, but one of the major reasons is that single classifiers do not have the capacity to generate reliable results which degrade their overall accuracy rate. Moreover, the mechanism of combining outputs generated by different classifiers not only reduces errors of individual models but also increases the overall accuracy rate of the model. In our work, we have used three baseline classifiers i.e., KNN, DT and RF. The reason why specifically these three classifiers have been selected in the proposed work is that they generate more effective results individually as seen from the literature. However, we tend to increase their individual performance by combining them into ensemble learning method.

K-Nearest Neighbor (KNN)

KNN is a machine learning algorithm that can be applied in liver disease detection to classify patients based on their medical data It is a non-parametric algorithm which means it doesn't make any assumptions for given data distribution and is based on the principle of finding k-nearest data points to make the prediction.

Decision Tree (DT)

DT is yet another ML classifier that is widely used in classification tasks. It organizes data into a hierarchical tree-like structure, with each internal node representing a feature and a decision based on that feature, and each leaf node indicating the predicted class label.

Random Forest (RF)

The Random Forest (RF) classifier is a powerful and versatile machine learning algorithm commonly used for both classification and regression tasks. It belongs to the ensemble learning category and is built upon the concept of Decision Trees.

After the three baseline models are initialized, process of training is started. However, before making prediction we have implemented bagging technique on each classifier. The total number of bags in our work is 5, which means that 5 copies of features dataset are created and each classifier is trained on each data subset to generate 5 individual outcomes. After this, the role of ensemble learning comes into play wherein the predictions made by each classifier for 5 bags is combined for making the final prediction.

The process of combining data is performed in such a way that first prediction of KNN, DT and RF are combined to form first output and then second output of three baseline classifiers is combined to form second output. This process keeps on going till we got the five predictions made by combining the data of three classifiers. Once this process is completed, MOD majority voting mechanism is applied to make the final binary and multi-class classification in which presence or absence of disease as well as stage of disease is determined. The MOD based voting mechanism gives that output as final prediction for a particular dataset, which possess highest frequency or got highest votes.

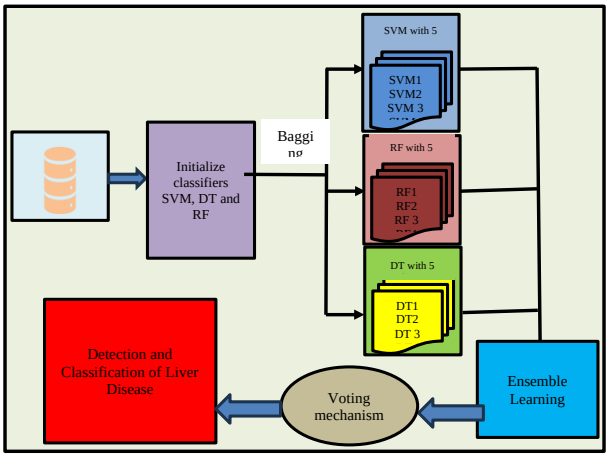

Figure 4. Proposed BELV classification model

4.2 How Proposed NeuroOPT is Better?

The proposed NeuroOPT model stands out as a highly effective method for liver disease detection, addressing key limitations in existing approaches. Our model's superiority lies in its comprehensive design and practical implementation, which surpasses current methods commonly found on the Kaggle site, where the ILPD and CPD datasets are sourced. Unlike the majority of existing models that perform binary classification using a single dataset, NeuroOPT incorporates a multi-stage classification system and leverages both ILPD and CPD datasets, significantly enhancing its versatility and applicability.

While reviewing existing liver disease detection models revealed that most achieve an accuracy range of 70% to 93% on the ILPD dataset, which is suboptimal for a reliable binary classification system. Although a few models approach 99% accuracy, they lack dynamism and are limited to predicting disease presence or absence within a single dataset. Recognizing these shortcomings, we aimed to develop a more dynamic and practical model by integrating two datasets and addressing both binary and multi-stage classification challenges.

However, the CPD dataset, in particular, posed a significant challenge for existing models, which demonstrated an accuracy range of only 40% to 67% for multi-class classification, highlighting their ineffectiveness for categorizing multiple stages of liver diseases. To overcome these issues, we introduced a three-stage feature selection technique, enhancing the model's ability to extract relevant features efficiently. Furthermore, we employed the NeuroOPT model, which utilizes a neural network optimized with $GWOA_2$ to fine-tune hyperparameters and improve predictive accuracy. By implementing these advancements, our proposed NeuroOPT model not only increases the accuracy rate for both binary and multi-class classification but also offers a dynamic and practical solution that addresses the limitations of current liver disease detection models. It is pertinent to mention here that $GWOA_2$ tuned FFNN is proven out to be highly effective than conventional approaches on both kaggle datasets, as proven by accuracy rates discussed in results section of this paper.

5. Results And Discussions

This section discusses the results obtained for the proposed liver disease detection approach over other similar models for both binary and multi-stage disease classifications on ILPD and CPD datasets respectively. Moreover, we will also cover the experimental settings used for the proposed approach during the testing phase.

5.1 Simulation setup

The effectiveness of the proposed approach is examined and compared with traditional liver disease detection models in MATLAB software. This software was operated on a system with i5 Core processor and 8GB RAM. Moreover, the OS we used was windows 10 Pro with 500GB HDD. With this configuration, we aim to analyze the performance of various conventional models for binary and multi-class classifications, elaborately discussed in this section of manuscript.

5.2 Binary Classification Results

The binary classification is performed on ILPD dataset using proposed model. To begin with, we have firstly analyzed the performance of our liver disease detection approach by observing its confusion matrix table. The main reason for doing so is that this matrix helps us in evaluating other important

parameters easily. Figure 5 represents the confusion matrix of the proposed model with target class and output class on x and y-axis respectively. Liver disease detection is a binary classification problem where the goal is to distinguish between patients with liver disease (positive class) and those without liver disease (negative class). The confusion matrix helps to evaluate the model's performance in such scenarios. By observing the trend in confusion matrix, we observed our model attains an overall accuracy of 93.2%.

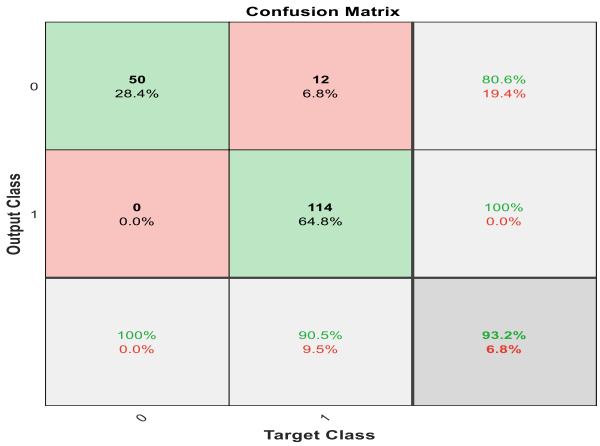

Figure 5. Confusion matrix of proposed model for binary classification

Moreover, we have also evaluated the performance of proposed approach in terms of its ROC graph. It is used to assess and visualize the trade-off between the True Positive Rate (TPR) on y-axis and the False Positive Rate (FPR) on x-axis of a classification model across different threshold values, as shown in Figure 6. The curve showcases how the model's performance varies as we adjust the threshold for classifying positive and negative instances. A higher TPR typically comes at the cost of a higher FPR. The graph reveals that curve starts at the bottom-left corner (0, 0) and moves upward and to the right. The point on the ROC curve that is closest to the top-left corner represents the ideal balance between true positives and false positives. By analysing this curve, we aim in in selecting an appropriate classification threshold and also assessed how well our model separates the classes.

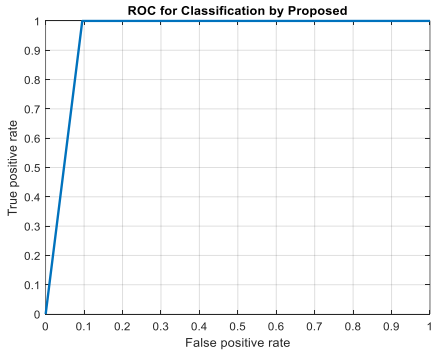

Figure 6. ROC in proposed model

5.3 Comparative Results For Binary Classification

To further validate the efficacy of proposed approach, we compared it with few traditional models in context of their respective accuracy rates. The comparative graph attained for accuracy is showcased in figure 7, with different models and their accuracy values on x and y-axis respectively. Results

showcased that author who used NB classifier for predicting liver disease achieves accuracy rate of just 0.71, while as, models like SVM, LR, MLP and 1-NN achieved accuracy rate of 0.70,

0.7, 0.69 and 0.688 respectively. While comparing the proposed approach with more techniques wherein authors have used j48, RF, RT, RepTree, RotF, AdaBoostM1, stacking, bagging, and voting mechanisms for predicting liver diseases, the accuracy rates were 0.73, 0.79, 0.71, 0.70, 0.77, 0.79, 0.794, 0.78 and 0.80 respectively. On the contrary, our proposed approach achieved a classification accuracy rate of 0.93, which is significantly higher than all previously discussed models. This increased accuracy rate is achieved in proposed model because of implementing effective processing, FS techniques which improves performance of classifiers.

Likewise, the proposed model's effectiveness was also evaluated and compared with other similar traditional approaches in terms of their precision rates. Figure 8 depicts the comparative graph for the same. Upon carefully examining the graph, we observed that 1-NN model was giving worst precision results of just 0.69, whereas, it was 0.70 in MLP, 0.71 in RepTree, RT, and LR, 0.72 and 0.73 in NB and j48, 0.74 in SVM, 0.78 in RotF, 0.79 in RF, AdaBoostM1, Stcking and Bagging and 0.80 in voting methods respectively. These values were good but there was still scope of improvement. The proposed model depicts a precision rate of 0.99 which signifies that out of all the instances that are predicted as positive, 99% of them are actually true positive cases.

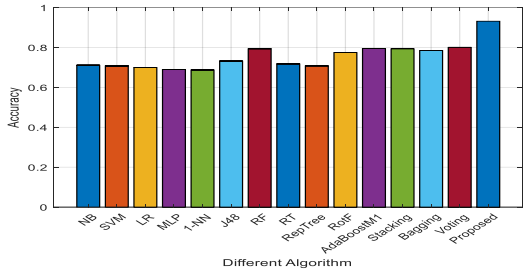

Figure 7. Accuracy comparison graph

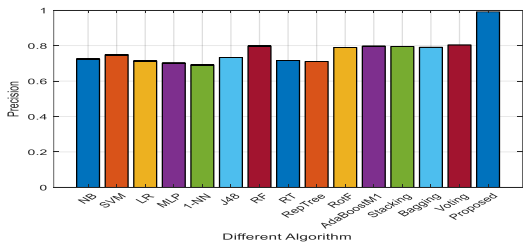

Figure 8. Precision comparison graph

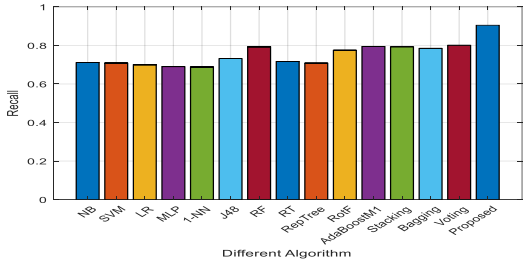

Figure 9. Recall Comparison Graph

Furthermore, we have also evaluated the performance of proposed approach with conventional techniques in terms of their recall rates. The comparative graph obtained for the same is shown in Figure 9. From the given graph, we concluded that out of all the models, the bar of proposed recall rate is highest at 0.90, depicting its supremacy. On the other hand, among traditional models recall value was highest in voting method with 0.80 while as, it was lowest in 1-NN model with only 0.688 respectively. The recall score in other models were in between 0.68 and 0.80. Standard models like NB, SVM, LR, ML, J48, RF and RT attained a recall rate of 0.71, 0.70, 0.7, 0.69, 0.73, 0.79 and 0.71

respectively. Similarly, the recall score was only 0.70 in RepTree, 0.77 in RotF, 0.79 in AdaBoostM1 and stacking, and 0.78 in bagging methods.

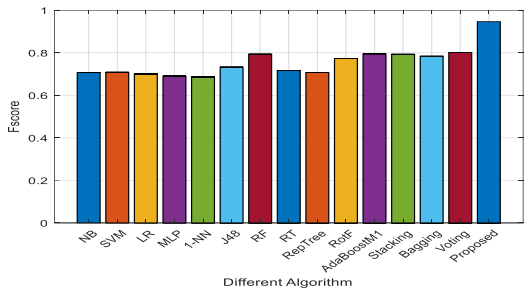

Figure 10. F1-Score comparison graph

Moreover, the efficacy of the proposed system is tested and validated by putting it in comparison with traditional models in terms of their F1-Scores. Figure 10 shows the comparative graph for the same. The graph reveals that traditional 1-NN and MLP are two worst performing models with an F1-Score of 0.68 and 0.69 respectively, while as, AdaBoostM1 and Voting approach are the two best performing models with a F1-Score of 0.80 and 0.79, among all the given standard models. The authors who used NB, SVM, LR, J48, RT, RepTree, RotF and bagging attained F-Score of 0.70, 0.708, 0.7, 0.73, 0.71, 0.70, 0.77 and 0.78 respectively, while as, RF and stacking models attained an F-score of 0.793 respectively. On the contrary side, when F1-Score value was observed in proposed model, it came out to be 0.94 which is significantly around 14% higher than traditional best performing model.

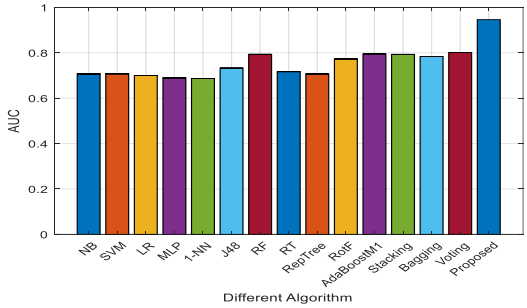

Figure 11. AUC comparison Graph

Finally, to prove the supremacy of proposed approach over other similar approaches, we analyzed its performance in terms of AUC, whose graph is shown in Figure 11. The x and y-axis of the given graph calibrates to the different classification models and their AUC values respectively. As mentioned previously, the model shows best performance when its AUC value is close to 1. From the given graph, it is observed that among all the models 1-NN models is exhibiting worst performance by attaining an AUC of just 0.69 while as, AUC value was better in voting method with 0.88 value. For all the remaining models, the AUC score was 0.70 in SVM, 0.71 in RT, 0.73 in J48, 0.74 in MLP, 0.75 in LR, 0.76 in RepTree, 0.77 in NB, 0.86 in RotF, 0.87 in RF, AdaBoostM1 and bagging, and 0.881 in stacking respectively. While as, this is not the case in proposed model, wherein we were able to achieve an AUC of 0.95 which is very close to 1, when compared with other similar models. The specific values for all the mentioned parameters is represented in Table 7.

Table 7: Comparative values of different parameters

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC
NB	0.711	0.724	0.711	0.707	0.771
SVM	0.708	0.748	0.708	0.708	0.708
LR	0.7	0.713	0.7	0.7	0.754
MLP	0.69	0.701	0.69	0.69	0.742
1-NN	0.688	0.691	0.688	0.687	0.693
J48	0.732	0.733	0.732	0.732	0.735
RF	0.794	0.798	0.793	0.793	0.877
RT	0.718	0.717	0.717	0.717	0.717
RepTree	0.708	0.71	0.708	0.707	0.761
RotF	0.775	0.789	0.775	0.773	0.869
AdaBoostM1	0.795	0.797	0.795	0.795	0.879
Stacking	0.794	0.795	0.793	0.793	0.881
Bagging	0.785	0.791	0.785	0.784	0.872
Voting	0.801	0.804	0.801	0.801	0.884
Proposed	0.93182	0.9913	0.90476	0.94606	0.95238

5.4 Multi-stage Disease Classification Results

In second phase, different stages of liver disease are identified by proposed model through non-ensemble learning, bagging ensemble, and boosting ensemble techniques. The results obtained for three cases were examined and compared with traditional models using CPD dataset. Initially, we have determined the confusion matrix for proposed model for multi-stage disease classification in which four stages of disease are detected, as shown in Figure 12. The given figure demonstrates that proposed approach correctly identified class 1 of disease with 100% accuracy while as, accuracy of 97%, 93% and 95% are attained for classes 2, 3 and 4 respectively. Moreover, the confusion matrix also aids in evaluating other metrics like precision, recall and F1-Score.

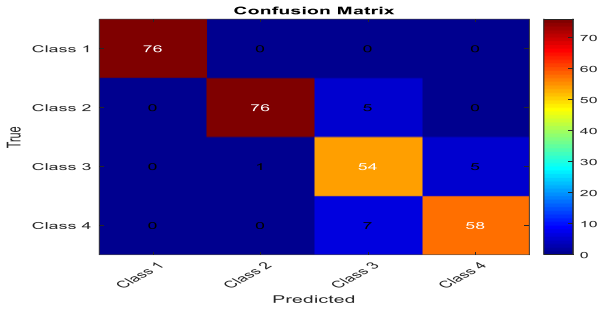

Figure 12. Confusion matrix for multi-stage disease classification

Moreover, to prove the supremacy of proposed approach over other models, we compared its performance with few conventional models for case 1 (non-ensemble learning cross validation) technique. The comparative graph obtained for the same is shown in Figure 13, with different models on x-axis and accuracy, precision, recall and F1-Score metrics on y-axis respectively. After analyzing the given graph, it is observed that proposed model attained higher accuracy of 96.8% whereas, other

models like AdaBoost, DT, ETT, GB, KNN, LGBM, LR, RF and SVC attained only 71%, 61%, 68%, 72.7%, 67%, 69%, 72%, 70% and 69% respectively. Moreover, proposed approach attained high precision, F1-Score values of 92.8% and recall of 92.9% for this case. However, out of all conventional models, LR model attains best precision of 66%, while as, DT and GB attains best recall and F1-Score of 50% and 54% respectively. These results clearly indicate efficacy of proposed approach over other approaches. For precise calculations, refer Table 8.

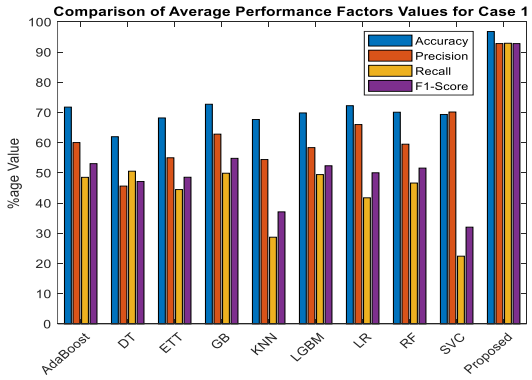

Figure 13. Comparative results for Case 1 multi-stage disease classification

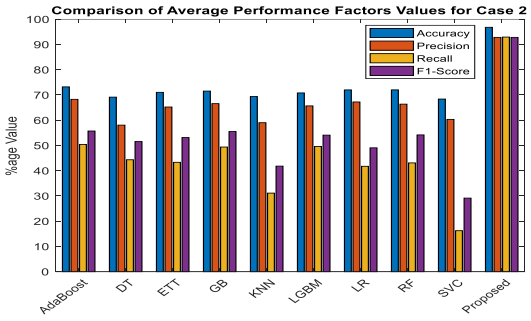

Figure 14. Comparative results for Case 2 multi-stage disease classification

Furthermore, we analyzed the performance of proposed approach with other conventional models on CPD dataset for determining the stage of liver disease through bagging ensemble cross validation technique (Case 2). The comparative graph obtained for the same is shown in figure 14, which shows proposed model attained highest accuracy (96.8%), precision (92.8%), recall (92.9%) and F1-Score (92.8%) respectively. However, this is not the case in traditional models wherein highest accuracy of 73% was attained AdaBoost and lowest accuracy of 68% is attained by SVC model. Similarly, for other metrics like precision, recall and F1-measure AdaBoost is continuously showing better results than other conventional models, however, it was still far from performance of proposed approach. Table 9 depicts precise values obtained for each metric in this case.

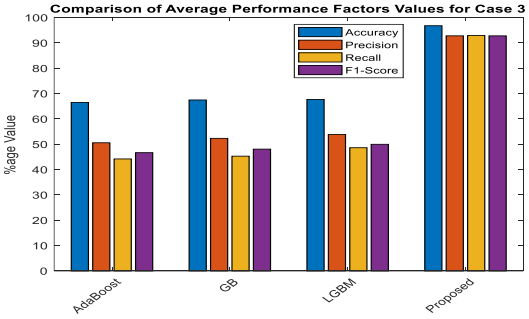

Figure 15. Comparative results for Case 3 multi-stage disease classification

In addition to this, we have also analyzed and compared the performance of proposed approach with traditional models for performing multi-stage disease classification task on CPD dataset by employing Boosting ensemble cross validation technique (case 3). Figure 15 depicts the comparative graph obtained for the same. The results showcased that proposed model attains highest accuracy of 96% whereas, it was only 66% in AdaBoost and 67% in GB and LGBM models respectively. Similarly,

precision values are evaluated which came out to be only 50%, 52% and 53% in conventional AdaBoost, GB and LGBM models and 92.8% in proposed model. Also, proposed model overpowers three traditional models in recall and f-score values as well to prove its supremacy. The specific values obtained for different models for this case is shown in Table 10.

Table 8: Comparative Results for Case 1 multi-stage disease classification on CPD dataset

Parameters	AdaBoost	DT	ETT	GB	KNN	LGBM	LR	RF	SVC	Proposed
Accuracy	71.79	61.98	68.2	72.74	67.69	69.86	72.25	70.1	69.36	96.80851
Precision	60.04	45.64	55.01	62.84	54.42	58.37	66	59.51	70.17	92.82107
Recall	48.54	50.57	44.48	49.9	28.71	49.44	41.75	46.65	22.43	92.93981
F1-Score	53.06	47.17	48.57	54.83	37.1	52.36	50.04	51.6	32.04	92.81078

Table 9: Comparative Results for Case 2 multi-stage disease classification on CPD dataset

Parameters	AdaBoost	DT	ETT	GB	KNN	LGBM	LR	RF	SVC	Proposed
Accuracy	73.21	69.15	71.05	71.54	69.37	70.81	72.01	72.02	68.39	96.80851
Precision	68.28	58.06	65.23	66.56	59.01	65.67	67.23	66.37	60.33	92.82107
Recall	50.39	44.34	43.33	49.38	31.13	49.62	41.75	43.09	16.24	92.93981
F1-Score	55.73	51.55	53.12	55.51	41.82	54.1	49.06	54.18	29.15	92.81078

Table 10: Comparative Results for Case 3 multi-stage disease classification on CPD dataset.

Parameters	AdaBoost	GB	LGBM	Proposed
'Accuracy'	66.51	67.5	67.72	96.80851
'Precision'	50.6	52.35	53.89	92.82107
'Recall'	44.22	45.32	48.65	92.93981
'F1-Score'	46.69	48.1	50	92.81078

5.5 Result Outcomes

During the analysis of results, we gleaned significant insights pertaining to both conventional and novel classification models. Our experimentation revealed that among the traditional models we considered, the 1-NN model consistently exhibited the least favourable performance across all provided parameters. Our observations indicate that our approach led to an improvement in accuracy by approximately 0.13 when contrasted with the standard voting method for binary classification on ILPD dataset. Likewise, the precision rate of our proposed model saw an enhancement of 0.187 over the top-performing traditional model. Additionally, our proposed model exhibited advancements in the F1-Score and AUC by margins of 0.14 and 0.068, respectively. For multi-stage disease classification, the proposed model continuously outperformed all standard models under three cases. The results showed that our approach increased accuracy by around 24.55% than LR model for Case 1, while as, for case 2 and 3, accuracy was improved by around 23.59% and 29.08%, when compared to AdaBoost and LGBM standard models respectively. These findings collectively underscore the superiority of our proposed model across all considered evaluation metrics for binary as well as multi-class classification to prove its supremacy over the mentioned traditional models.

Conclusion

Detecting liver disease is of paramount importance due to the critical role that the liver plays in maintaining overall health and well-being. Keeping this in mind, an effective liver disease detection

approach is presented in this paper wherein work has been done on feature selection and classification part. The simulation of the proposed liver disease detection model is performed in MATLAB software under different evaluation parameters for binary and multi-stage disease classification. The experimental results of proposed model were compared with conventional models to prove its supremacy. For binary classification, our method attained highest accuracy of 0.93 whereas, it was only 0.80 in voting method (traditional best), showing an increment of 0.13. Similarly, the proposed model was outperforming all other conventional models by achieving a precision rate of 0.99, while as it was only 0.80 in standard best performing model. Not only this, the proposed model showed outstanding results for recall and F1-Score as well, by depicting an increment of around 0.10 and 0.14 respectively. Additionally, the AUC curve of proposed model is closest to 1, depicting that model is able to detect and classify liver diseases effectively. Likewise, for multi-stage disease classification, the proposed model is outperforming all traditional models for three cases by achieving an accuracy of 96.8%. Also, it is observed that throughout the three cases the values of other factors like precision, recall and F1-Score doesn't change and is continuously better than all standard models. In future, the performance of liver disease detection models can further be enhanced by implementing nature inspired optimization algorithm along with ML and DL classifiers.

References

1. M. Ghosh, A. B. Smith, C. D. Johnson, D. E. Brown, and E. F. Davis, "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," *Intell. Autom. Soft Comput.*, vol. 30, no. 3, 2021.
2. T. M. Ghazal, M. A. B. Askar, A. S. M. Ali, M. A. El-Masry, and A. A. El-Tawil, "Intelligent Model to Predict Early Liver Disease using Machine Learning Technique," in *2022 Int. Conf. Bus. Anal. Technol. Secur. (ICBATS)*, 2022, pp. 1-6.
3. A. O. D. Ekanem, O. D. T. Omidiran, and O. S. Jesupelumi, "Prediction and Diagnosis of Liver Disease in Human Using Machine Learning," *Int. J. Comput. Trends Technol.*, vol. 68, no. 8, pp. 44-52, Aug. 2020.
4. E. M. Hameed, M. A. Ameen, K. F. Ahmed, and R. I. Ali, "Liver Disease Detection and Prediction Using SVM Techniques," in *2022 3rd Inf. Technol. Enhance e-Learn. Other Appl. (IT-ELA)*, 2022, pp. 1-6.
5. V. Sharma and K. C. Juglan, "Automated classification of fatty and normal liver ultrasound images based on mutual information feature selection," *IRBM*, vol. 39, no. 5, pp. 313-323, 2018.
6. F. Bessone, M. V. Razori, and M. G. Roma, "Molecular pathways of nonalcoholic fatty liver disease development and progression," *Cell. Mol. Life Sci.*, vol. 76, pp. 99-128, 2019.
7. D. Kim, M. E. Cholankeril, A. A. Ahmed, D. V. Tighe, and A. Singal, "Trends in etiology-based mortality from chronic liver disease before and during COVID-19 pandemic in the United States," *Clin. Gastroenterol. Hepatol.*, vol. 20, no. 10, pp. 2307-2316, 2022.
8. B. L. Da, G. Y. Im, and T. D. Schiano, "Coronavirus disease 2019 hangover: a rising tide of alcohol use disorder and alcohol-associated liver disease," *Hepatology*, vol. 72, no. 3, pp. 1102-1108, 2020.
9. G. Cabibbo, F. Enea, F. M. Cammà, V. Petta, M. R. Maida, A. Craxì, and G. D'Antona, "SARS-CoV-2 infection in patients with a normal or abnormal liver," *J. Viral Hepat.*, vol. 28, no. 1, pp. 4-11, 2021.
10. P. Decharatanachart, N. Chaikledkaew, Y. Thakkinstian, and C. Anothaisintawee, "Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis," *BMC Gastroenterol.*, vol. 21, no. 1, pp. 1-16, 2021.
11. S. GM, B. Amuthan, P. M. Muthukumar, and D. Subramanian, "Healthcare Data Analytics Using Artificial Intelligence," in *Artif. Intell. Inf. Manag.: A Healthc. Perspect.*, 2021, pp. 45-85.
12. J. Singh, S. Bagga, and R. Kaur, "Software-based prediction of liver disease with feature selection and classification techniques," *Procedia Comput. Sci.*, vol. 167, pp. 1970-1980, 2020.
13. K. Dutta, S. Chandra, and M. K. Gourisaria, "Early-Stage detection of liver disease through machine learning algorithms," in *Advances Data Inf. Sci.*, Springer, Singapore, 2022, pp. 155-166.

14. G. Shaheamlung and H. Kaur, "The Diagnosis of Chronic Liver Disease using Machine Learning Techniques," *Inf. Technol. Ind.*, vol. 9, 2021, doi: 10.17762/itii.v9i2.382.
15. H. Ding, Z. Fu, X. Zhang, and Y. Tang, "A framework for identification and classification of liver diseases based on machine learning algorithms," *Front. Oncol.*, vol. 12, 2022, Art. no. 1048348.
16. Z. Yao, J. Hu, Y. Lin, and L. Zhang, "Liver disease screening based on densely connected deep neural networks," *Neural Netw.*, vol. 123, pp. 299-304, 2020.
17. P. Kumar and R. S. Thakur, "Liver disorder detection using variable-neighbor weighted fuzzy K nearest neighbor approach," *Multimed. Tools Appl.*, vol. 80, pp. 16515-16535, 2021.
18. M. G. Lanjewar, R. V. Dixit, S. P. Wankhade, and K. G. Parlewar, "CNN with machine learning approaches using ExtraTreesClassifier and MRMR feature selection techniques to detect liver diseases on cloud," *Cluster Comput.*, pp. 1-16, 2022
19. S. Ambesange, V. A, R. Uppin, S. Patil, and V. Patil, "Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques," in *2020 IEEE Int. Conf. Cloud Comput. Emerg. Mark. (CCEM)*, Bengaluru, India, 2020, pp. 98-102, doi: 10.1109/CCEM50674.2020.00030.
20. M. Varchagall and P. A. Yogegowda, "Early detection of liver disorders using hybrid soft computing techniques for optimal feature selection and classification," *Concurrency Comput.: Pract. Exp.*, vol. 35, no. 6, pp. 1-1, 2023.
21. D. Bhupathi, C. N. L. Tan, S. S. Tirumala, and S. Ray, "Liver disease detection using machine learning techniques," 2022.
22. B. H. Al Telaq and N. Hewahi, "Prediction of Liver Disease using Machine Learning Models with PCA," in *2021 Int. Conf. Data Anal. Bus. Ind. (ICDABI)*, Sakheer, Bahrain, 2021, pp. 250-254, doi: 10.1109/ICDABI53623.2021.9655897.
23. R. Amin, S. Ahmed, T. K. Khandaker, and A. S. Khan, "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms," *Informatics Med. Unlocked*, vol. 36, 2023, Art. no. 101155.
24. M. G. Lanjewar, R. V. Dixit, S. P. Wankhade, and K. G. Parlewar, "CNN with machine learning approaches using ExtraTreesClassifier and MRMR feature selection techniques to detect liver diseases on cloud," *Cluster Comput.*, vol. 26, no. 6, pp. 3657-3672, 2023.
25. W. M. Shaban, "Early diagnosis of liver disease using improved binary butterfly optimization and machine learning algorithms," *Multimed. Tools Appl.*, vol. 83, no. 10, pp. 30867-30895, 2024.

BIOGRAPHIES OF AUTHORS

Gurmeet Kaur Saini received B.Tech and M.Tech degrees from CT Group of Institutions Jalandhar and CGC College of Engineering, Landran, India in 2014 and 2016 respectively. She is currentlyworking towards the Ph.D. degree from the Chandigarh University. Her research interests include Digital Image Processing, Machine learning. She can be contacted at email: gurmeetsaini02@gmail.com

Dr. Sachin Ahuja is working as Director in Chandigarh University. He holds a PhD in Data mining. His primary research interests are in the field of educational data mining. Apart from data mining, his teaching interests include big data, relation database and procedural languages. He can be contacted at sachinahuja11@gmail.com.

Dr. Vishal Bharti is working as Professor and Additional Director at Chandigarh University, Mohali, Punjab. He completed his Ph.D. in 2016 in the area of Information Security. His area of specialization is Cyber Security, Network Security and Distributed Computing and Machine Learning. He can be contacted at mevishalbharti11@yahoo.in