Predicting Liver Cirrhosis using Feature Selection Aided by PSO Based Optimized SVM

Puneet Chhina [1], Munish Saini [2], Amit Chhabra [3]

[1, 2, 3] Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, Punjab, India

Orcid : [1] 0009-0004-7174-6528, [2] 0000-0003-4129-2591, [3] 0000-0003-2056-6231

Email: [1] puneetchhina1800@gmail.com, [2] munish.cet@gndu.ac.in, [3] amit.cse@gndu.ac.in

Abstract

The technology of data mining has been at the forefront of medical research. Not only has it achieved remarkable results in assessing patients' risks, but it has also been used in numerous models to make clinical decisions and predict diseases. Using various data mining models, raw data can be transformed into useful information that can be analyzed logically and scientifically to produce accurate predictions and decisions. The selection of features that are more important than others presents a challenge when it comes to disease prediction. To improve the prediction's accuracy, subset feature selection is carried out. The study compared various classification models to select the most important features. Correctness is checked by extracting, loading, transforming, and analyzing the data. Based on the results, the Multi-Layer-Perceptron (MLP) neural network, Support Vector Machine (SVM), random forest, Particle Swarm Optimization (PSO-SVM), and Bayesian network are compared to other data mining models. In order to increase models' correctness and accuracy, they are cross-validated using both 10 fold and 5 fold methods. Exactness is assessed at 86.26%, 66.06%, 75.15, 78.11% and 94.62 % for irregular woodland, Bayesian organization, SVM, MLP-brain organization and PSO-SVM. As indicated by referenced processed rules most elevated precision for anticipating liver cirrhosis sickness is anticipated by half and half PSO-SVM.

Keywords: Data Mining, Liver Cirrhosis, Prediction, Support Vector Machine.

1. Introduction

In this day and age, liver illnesses are quite possibly of the quickest developing sickness; In point of fact, they now rank eleventh among the leading causes of death [1]. Liver diseases account for the deaths of 1.1 million people annually [2]. The quantity of individuals with liver cirrhosis expanded from 71 million out of 1990 to 122 million in 2017 [3]. Some of the circumstances that can result in liver cirrhosis comprise non-alcoholic steatohepatitis (NASH), alcohol-related liver disease, and chronic hepatitis B and C virus (HCV) infections [4]. Over the past ten years, the prevalence of NASH has significantly increased, while the prevalence of the other causes has decreased. There has been an evaluation that non-alcoholic oily liver disorder (NAFLD) will augment from 25% [5] to 33.5% by 2030[6]. In patients with liver cirrhosis, vitamin D deficiency typically ranges from 60% to 96%, rising to 96% in patients undergoing liver transplantation [7]. Additionally, patients with cirrhosis have thin bones. A concentrate on 76 men who drank 216 grams or more each day for over 24 years viewed that as 30% of them had vertebral fractures [8].

The overwhelming majority preinvasive liver cancers occur through hepatocellular carcinoma (HCC), the second-leading cause of cancer-related death in the world [9, 10]. 24,500 passings a year are believed to be brought about by HCC, as per reports [11]. It has been laid out that cirrhosis is the most widely recognized type of hepatocellular carcinoma, and the yearly gamble of HCC advancement is 3 to 8 percent in cirrhosis brought about by HBV and 1 to 7 percent in cirrhosis brought about by HCV [12, 13]. The most common way to stage and diagnose liver cirrhosis is with a liver biopsy. The liver biopsy is more susceptible to sampling error[14] because the specimen is so small. Moreover, interior draining and torment might happen. Some fundamental imaging findings of liver cirrhosis have been developed, including ultrasonography (US), registered tomography (CT), and attractive reverberation imaging (MRI) [15] in order to locate morphological changes earlier. In patients with persistent hepatitis, ultrasound is one of the most cost-effective and secure imaging techniques used to determine whether annual or semiannual testing is necessary [16]. The most delicate analytic device for assessing hepatic morphological change is CT [17]. An excellent evaluation of organ morphology, physiology, and function is made possible by the excellent contrast provided by MRI for the high-resolution image [18]. The CT filter image of both the cirrhotic and typical liver is depicted in Figure 1. It is particularly evident that cirrhotic liver has some surface nodularities while normal liver has a smooth surface with almost no nodules [19]. In individuals with grave cirrhosis, fibrotic septa emerge as hypo- and hyperintense reticulations on T1 and T2 weighted MRIs [21]. Numerous regenerating nodules and fibrotic scars created by hepatocytes are what produce the nodularity that has been noticed on the surface of the liver [20]. One of the morphologic CT manifestations of cirrhosis is the modified caudate right lobe ratio, and cross-sectional scans can be used to assess the prognosis [22].

Additional physical characteristics include the right back hepatic indent sign and the extended gallbladder fossa indication [23]. The most critical factor in the development of ascites is splanchnic vasodilatation, which results in a decrease in the volume of arterial blood [24]. Cirrhosis increases hepatic resistance through the formation of collateral veins and blood shunts, which increases portal flow and pressure [25]. Repaid cirrhosis patients had the option to live for around 10 to 12 years, making it very dangerous to perform liver transfers on them [26]. At five years after the liver transplant, the endurance rate rises to 75% [27].

To get a more exhaustive viewpoint on the assessment of cutting edge persistent liver illnesses, organizing of fibrosis and cirrhosis is required[28]. Histologic characteristics used to stage fibrosis include ECM deposits, changes in lobular architecture, and their location within the liver lobule. A semi-quantitative scoring system is created by combining these attributes, as shown in Table 1 [29]. The staging is performed employing a linear numerical scale with values that vary from 0 (showing no fibrosis) to 4-6 (indicating cirrhosis). The scoring system shows that the Ishak framework, which classifies cirrhosis into discrete and fragmented forms, is more distinct for late-stage fibrosis, resulting in a level of exactness that impacts the factor of repeatability and observer variability[30].

The following is a description of the proposed work's primary contributions:

1. to propose a framework for the hybrid metaheuristic-inspired detection of liver cirrhosis.

Comparing the proposed framework's performance to that of other cutting-edge methods.

The following is the structure of the remaining section of the paper: The related works are demonstrated in Section 2. In Section 3, the proposed method is explained. In area 4 execution model is assessed and examined. Area 5 addresses the aftereffects of the investigation. The conclusion is presented in section 6 at the end.

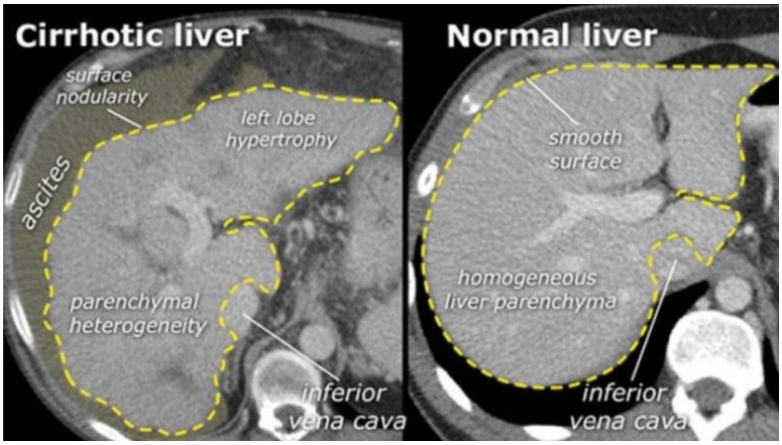


Fig.1 CT scan image of Cirrhotic liver and Normal liver[59].

Table1. Staging of Liver hepatitis to detect condition of liver.

EVALUATION SYSTEM	SHEUER	BATTS AND LUDWIG	METAVIR	ISHAK)
Levels 0	Negative	Na	clear fibrosis	Not at all
Region 1	upgraded PT	Acute fibrosis	portal segmental fibrosis	Several PT have portal fissures that may or might not feature short septa.
Stage 2	Portal septa	A periportal fibrosis	atypical septa and portal fibrosis	Short septa, with or without, are involved in the majority of PT portal fibrosis cases.
Level 3	Fibrosis having the distorsion of architecture	Decay in architecture with septal fiboris	A large number of septa with no cirrhosis symptoms	PT majority of portal fibrosis
Level IV	Cirrhosis (possible or certain)	Enhanced stage of Cirrhosis	Severe Problem of Cirrhosis	Most PT have portal fibrosis with significant bridging.
Stage 5	na	no	nil	Cirrhosis

2. Existing Work

Choi and co. [31] used CT blood vessel portography imaging procedures to find the hepatic nodular sores, supported the location of HCCs, and essentially worked on the injuries' guesses. R. Vivanti and others [32] concluded the divisions of new tumors in longitudinal

liver CT survey and liver developments which have inconvenience estimation using overall convolutional cerebrum network classifier. Ben and colleagues [33] used a three-overlap cross-approval CT technique to assign the location of liver division and liver metastases using the fully convolution network (FCN).

Christ and others [34] describe a method for subsequently dividing liver and injury CT images using flowed completely convolutional brain networks (CFCNs) and thick 3d restrictive irregular Fields (crfs) with quantitative biomarkers for precise clinical determination and PC-aided emotional supportive network selection. Wen and others [35] divided the liver sore from CT pictures utilizing a robotized strategy in light of convolutinal brain organizations and classifiers prepared with handmade highlights that hold mean, change, and context oriented highlights. Ben and colleagues [36] provided decision support for the classification of liver metastases into primary cancer sites using the LOOCV classification and top-n-accuracy (i.e., classes with top n probabilities).

Liu and co. [37] purposes a PC helped cirrhosis finding structure to investigate cirrhosis spread out ultrasound pictures and tuned on significant convolutional mind network model and are ready on help vector machine(SVM) classifier. Extreme learning machine (ELM) and the standard support vector machine (SVM) protocols that are particular to liver tissue are discussed by Mainak et al. [38], along with the cross-validation standards for deep learning that make use of a 22-layered neural network.

A deep learning approach that makes use of a stacked sparse auto-encoder to represent features was discussed by Hassan and colleagues [39]. This strategy depends on the innovation of profound learning. The stacked sparse auto-encoder is trained in this manner through unsupervised training. Guo and others By choosing CEUS pictures from the blood vessel stage, venous stage, and late stage, 40] fostered a man-made brainpower strategy of two-stage multi-view learning foundation for contrast-improved ultrasound (CEUS) based PC helped finding of liver cancers and following liver cirrhosis.

Felix and co [41] provided a completely automated method using a modified vessel connectivity analysis technique for abstracting and securing hepatic vascular graphs directly from the output of the segmentation model. To develop and validate the dense feature fusion neural network (DFuNN) for automatically admitting separate MRI sequences and phases, the sensitivity, specificity, accuracy, and area under the receiver-under-the-characteristic curve were calculated using the procedure defined by Shu-Hui and colleagues [42].

Aaron and co. [43] proposed using clinical patient data and attractive resonance (MR) imaging to create a man-made reasoning design in the hope of beneficial transarterial chemoembolization outcomes through the use of AI strategies. Merve et al. [44] used CT scans to examine the psoas muscle region at the degree of mid-3 lumbar vertebra to investigate the thickness of sarcopenia in cirrhotic patients, its correlation with standard prognostic scores, and their associations with mortality.

3. Material Used

The dataset for this study came from the repository at the University of California, Irvine (UCI). The data sets included 583 individuals, 416 of whom were liver patients and 167 who were not liver patients. Orientation, age, complete bilirubin, direct bilirubin, soluble

phosphate, absolute proteins, alamine aminotransferase, aspartate aminotransphirase, and yes or no names are among the 11 elements in the dataset.

The extract from the dataset description is shown in Table 2.The extract of a dataset desciption is shown in Table 2.

Table 2. Description of labels in given dataset.

ATTRIBUTE	DESCRIPTION
Gender	Male or Female
Age	Age of Person
Total bilirubin	A blood test is carried out to determine the bilirubin level. The purpose of this test is to determine your liver's functionality.
Directbilirubin	The liver converts bilirubin into an easily eliminated form. This is known as conjugated bilirubin or direct bilirubin.
Alkaline phosphate	The percentage of alkaline phosphatase (ALP) in your blood is assessed by an ALP test. An enzyme called ALP is often found throughout your body.
Aminotransferase	Aminotransferases, also known as transaminases, are a class of enzymes that catalyze the interconversion of oxoacids and amino acids by transferring amino groups.
Aspartate aminotransphirase	You can find the enzyme aspartate aminotransferase (AST) in many different tissues in your body. A compound is a protein that helps your body work by triggering synthetic responses.
Total proteins	The total protein test analyses the total quantity of two distinct types of proteins in the fluid that compose your blood.
Albumin	Albumin is a protein made by the liver
Albumin glogulin ratio	The ratio test counts all the proteins in your blood.
Yes or not	Person is having liver disease or not.

Another dataset of liver cancer divisions comes from LiTS - the Liver Growth Division Challenge (LiTS17), which was coordinated in connection with ISBI 2017 and MICCAI 2017. 130 liver segmentation NII-format CT scans are included. As depicted in Figure2, the NII format is converted into the PNG format for further preprocessing in order to enhance the sharpness and contrast of the images.

Table 3. Normal clinical range for healthy liver according to their attributes.

Attribute	Range
Total bilirubin	0.1-1.2 mg/dl
Directbilirubin	Less than 0.3 mg/dl
Alkaline phosphate	44 to 147 IU/l
Aminotransferase	4-36 U/L
Aspartate aminotransphirase	8-33 U/L
Albumin	3.4-5.4 g/dL
Ratio of albumin to globulin	1 or a little bit more
Complete proteins.	6.0-8.3 g/dL

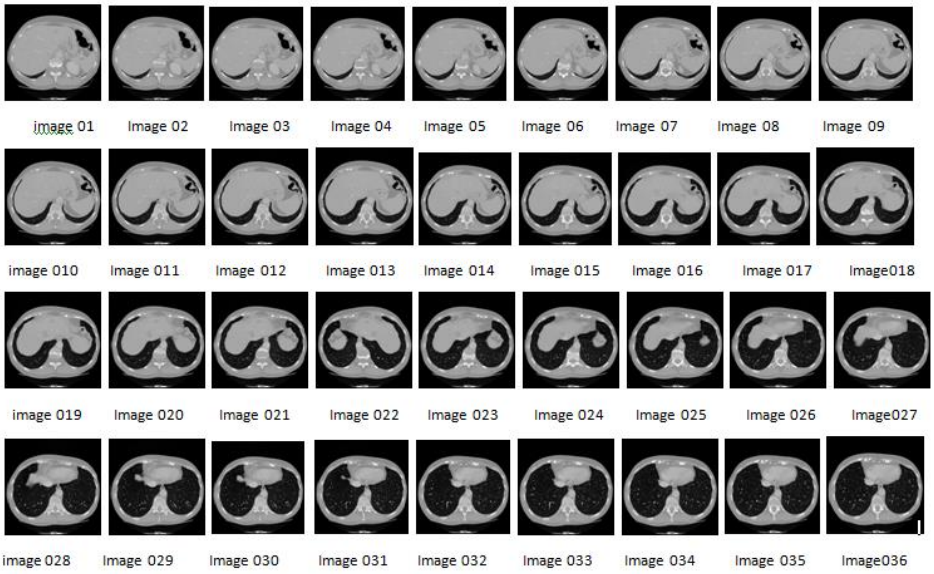


Figure 2. Segmentation of dataset used in proposed system.

3.1 Data Mining

Because there are datasets related to the liver that can be used to extract knowledge, using data mining models in medical research is a hot topic right now. Using a tree structure to present a general schema for data mining is crucial. Figure 3 shows the clinical information mining model which is separated into two classifications unmistakable and prescient.

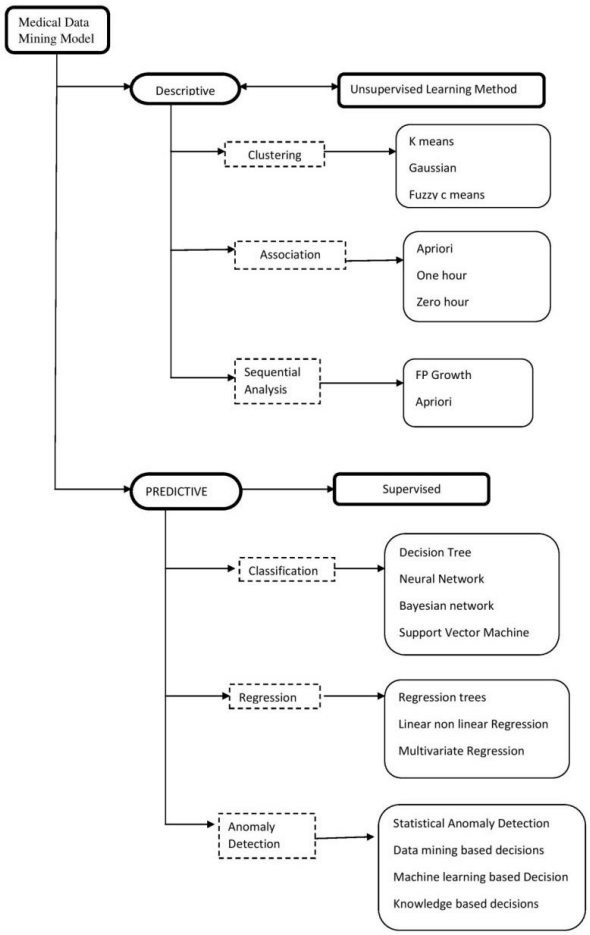


Figure3. Medical Data Mining Model.

3.1.1 Descriptive Methods

Correlation, frequency, cross-tabulation, and other descriptive statistics are all produced using descriptive methods. these techniques decide the likenesses in the information to investigate the current examples. They center around outline and change of information into significant data. Bunching, affiliation rule mining, and successive example disclosure are three models of clear nature in information mining.

3.1.2 Predictive Method.

To predict the value of a certain component or the effects of a certain ailment, prescient methodologies are used. The patient's history and current numbers are used in the approach above to anticipate future data. The most often used techniques in this paradigm are classification, regression, and anomaly detection..

3.3 Definitions of Vital Models

3.3.1 Random Forest

According to Liaw et al., one of the descriptive machine learning models is the Random Forest model. It has gained a tonne of popularity as a general-purpose classification and regression method [45]. This model joins different randomized decision trees and hoards their assumption by averaging. However, it is flexible enough to be used for both small-scale and large-scale learning tasks [46]. The optimal forest is an irregular one when the number of elements exceeds the number of perceptions. The random forest model for the UCI liver data set is shown in Figure 4.

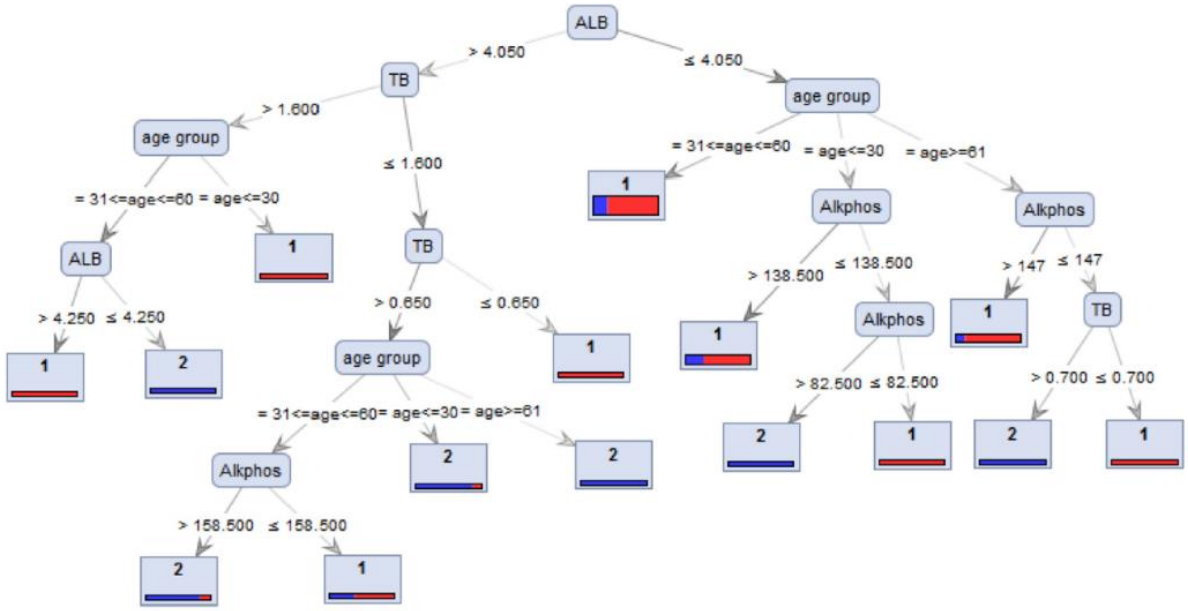


Fig. 4. Dataset from UCI on liver ailments using a random forest model.

3.3.2 Multi-layer Perceptron Neural Network

The definition and construction of a neural network have been influenced by the technique in which real neurons engage with one another [47]. An input layer, a hidden layer, and an output layer are all parts of its node layers. There are additionally weights and thresholds assigned to each linked node. Any hub is implemented if its output exceeds the edge value. To be able to learn and gain accuracy over time, neural networks rely on training data [48]. A neural network with a multi-layer perceptron is displayed in Figure 5.

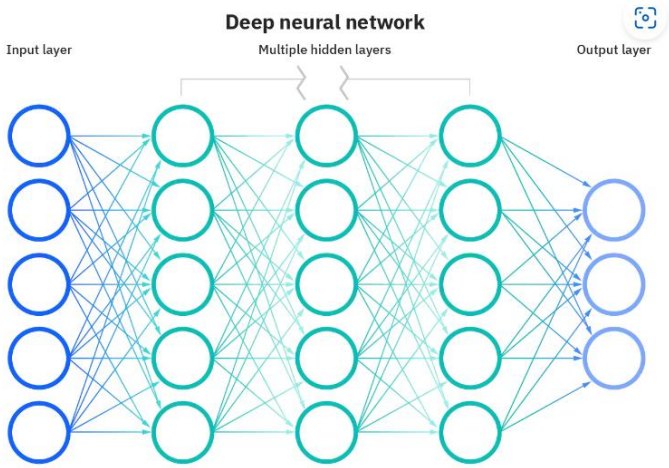


Fig 5. Model for MLP neural network models.

3.3.3 Bayesian Network

Classification is the primary application of the Naive Bayes supervised machine learning algorithm [49]. This model has both information highlights and straight-up yields prepared. The assumption that the inputs this model considers are independent of one another is made. One component change won't influence some other features[50]. Naive Bayes can handle both continuous and discrete data, has a large number of predictors and data points, and is highly scalable. A directed or non-circular graph is used in a Bayesian network, a graphical The structure for showing expected characteristics through a conditional tandem [51]. Figure 6 exhibits the Bayesian association histogram, which was applied to the UCI Liver Disease Innovative Collection.

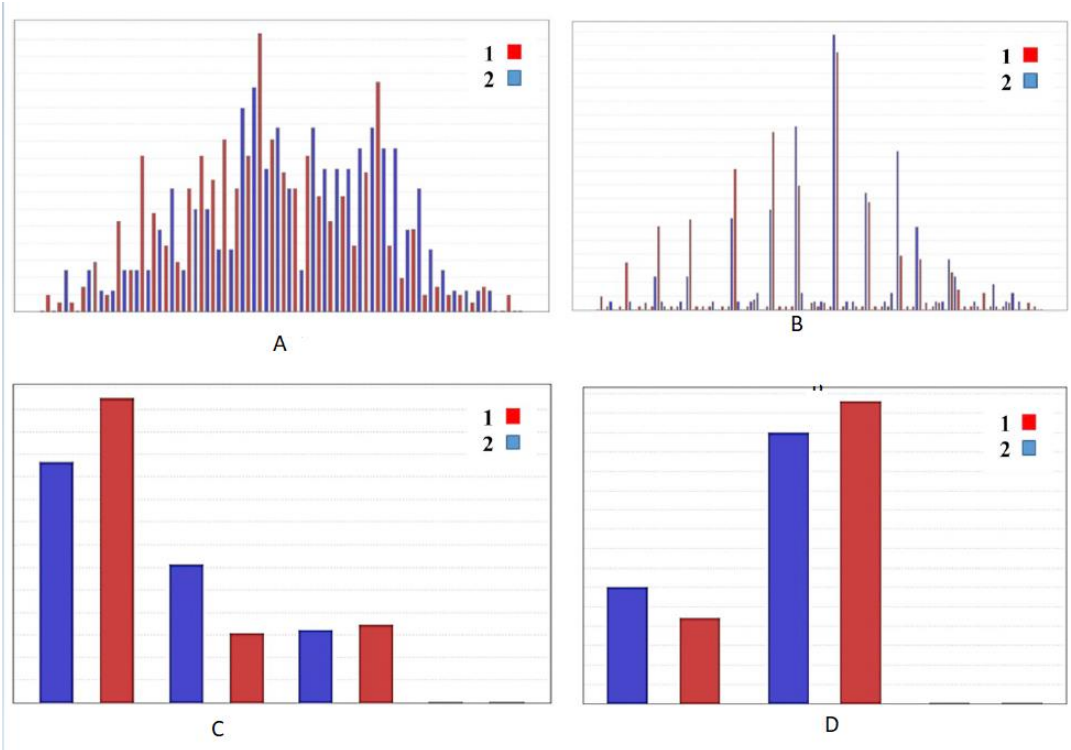


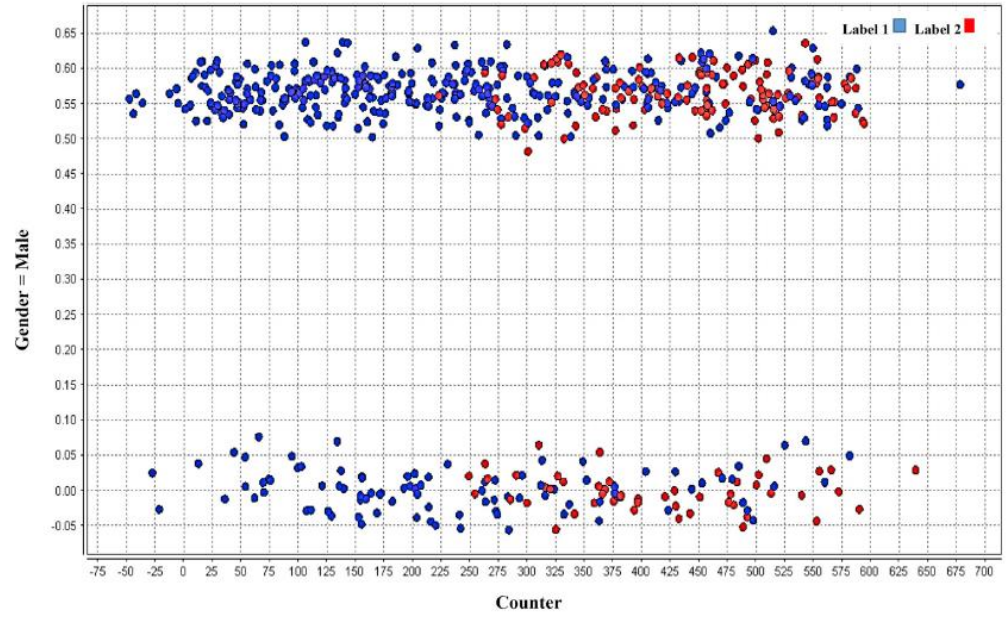
Fig. 6 Highlights within the liver ailment dataset that were communicated by the Bayesian Organisation model include (A) albumin, (B) albumin glogulin ratio, (C) age group, and (D) gender (make a recording density on the Y axis and feature range on the X axis).

3.3.4 Particle Swarm Optimization.

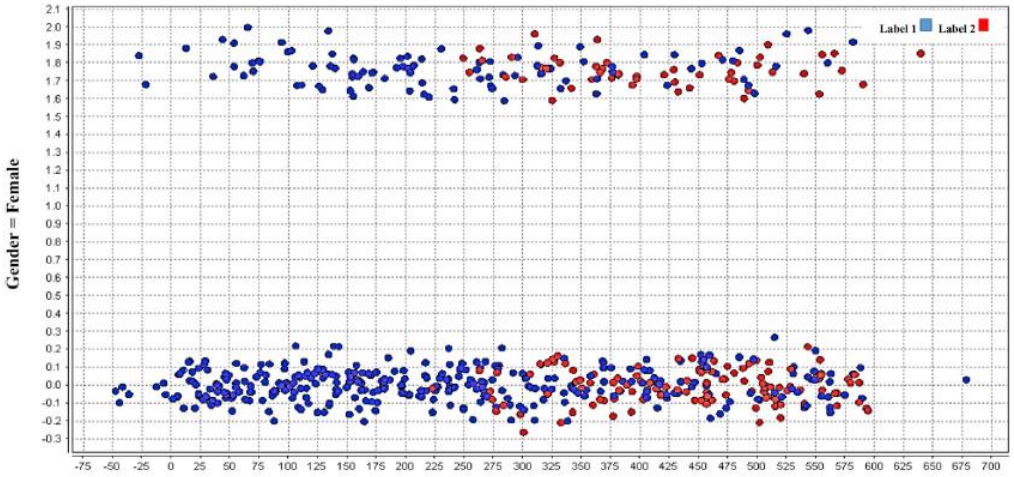
The Molecule Multitude Enhancement method was presented in 1995 and it is an aggregate hunt model upheld on the social ways of behaving of bird categories [52]. By adjusting each particle in the search space in accordance with the foremost space in a neighborhood, this relatively new method has been used to perform well on many optimizations [53]. However, it is more closely associated with evolutionary programming and genetic algorithms [54].

3.3.5 Support Vector Machine

An antiquated AI methodology called the assistance vector machine is used to solve problems with enormous amounts of detail grouping [55]. The goal is to select a decision boundary during the training phase that reinforces the minimal distance to each category [56]. Support vectors, or points, are used to select the boundary. Figure 7 depicts a support vector machine chart with a Counter axis and a Label axis. The dataset is part of label1, and label has characteristics like "Male" and "Female."



A

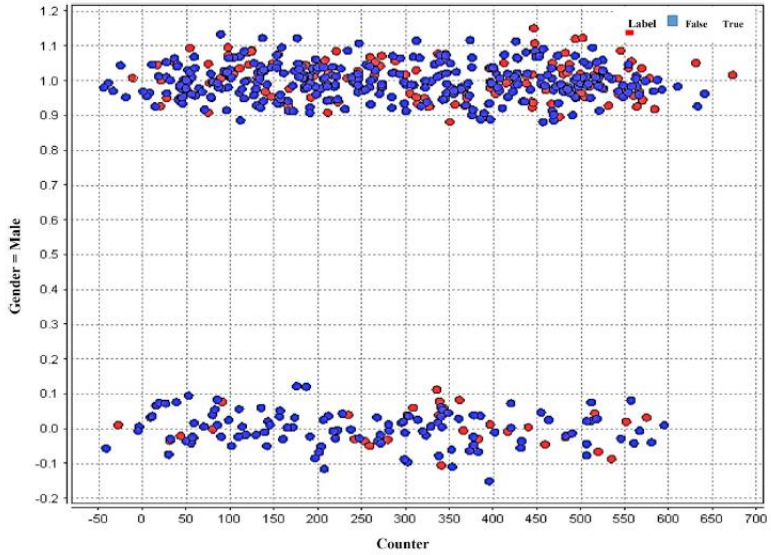


B

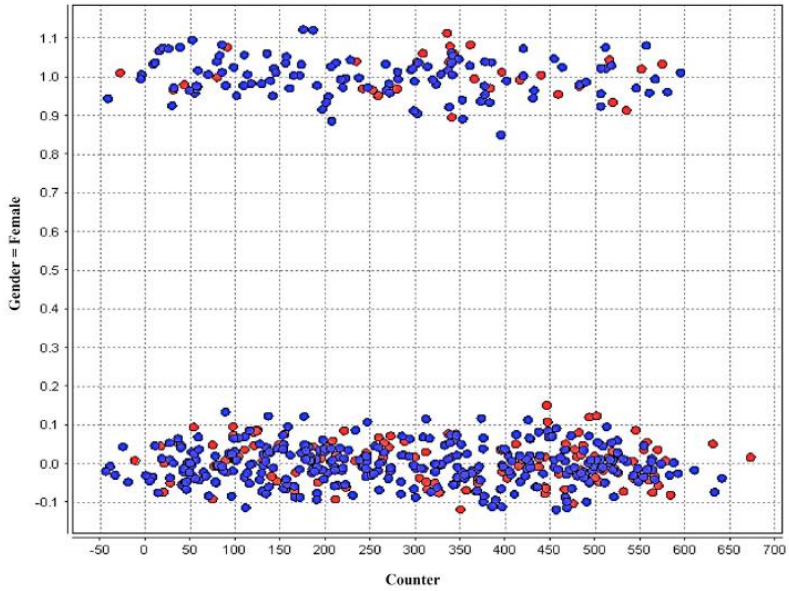
Fig 7. For additionally (A) male and (B) female, SVM distribution.

3.3.6 Particle Swarm Optimization With Support Vector Machine

The review proposed the Molecule Multitude Optimization (PSO) model and the Backing Vector Machine (SVM) model, which combined the PSO and SVM models to improve order exactness with a small and imperfect element subset [57]. Traditional SVM suffers from subjectivity and random city, whereas PSO selects the parameters for SVM automatically, providing a wider selection of parameters [58]. With both true and false labels, Figure 8 shows the PSO-SVM scatter for the two features "Male" and "Female" on a liver dataset.



(A)



(B)

Fig. 8 For the following aspects of PSO-SVM scatter: Male (A), Female (B), or combined

4. Proposed Methodology

This study experimented with the UCI liver disease dataset using Weka (Waikato Environment for Knowledge Analysis) version 3.8.5, which offers an effective visual representation of a predictive analytical diagram. The dataset of patients with liver cirrhosis was initially gathered from UCI[60] before being imported into WEKA. Before the data

mining process begins, feature engineering is followed by preprocessing to select the most crucial features for transformation and classification techniques. All highlights, paying little mind to crease strategies or split rates, are qualified for include choice and demonstrating. The channel based technique brings about the production of a satisfactory model with striking elements.

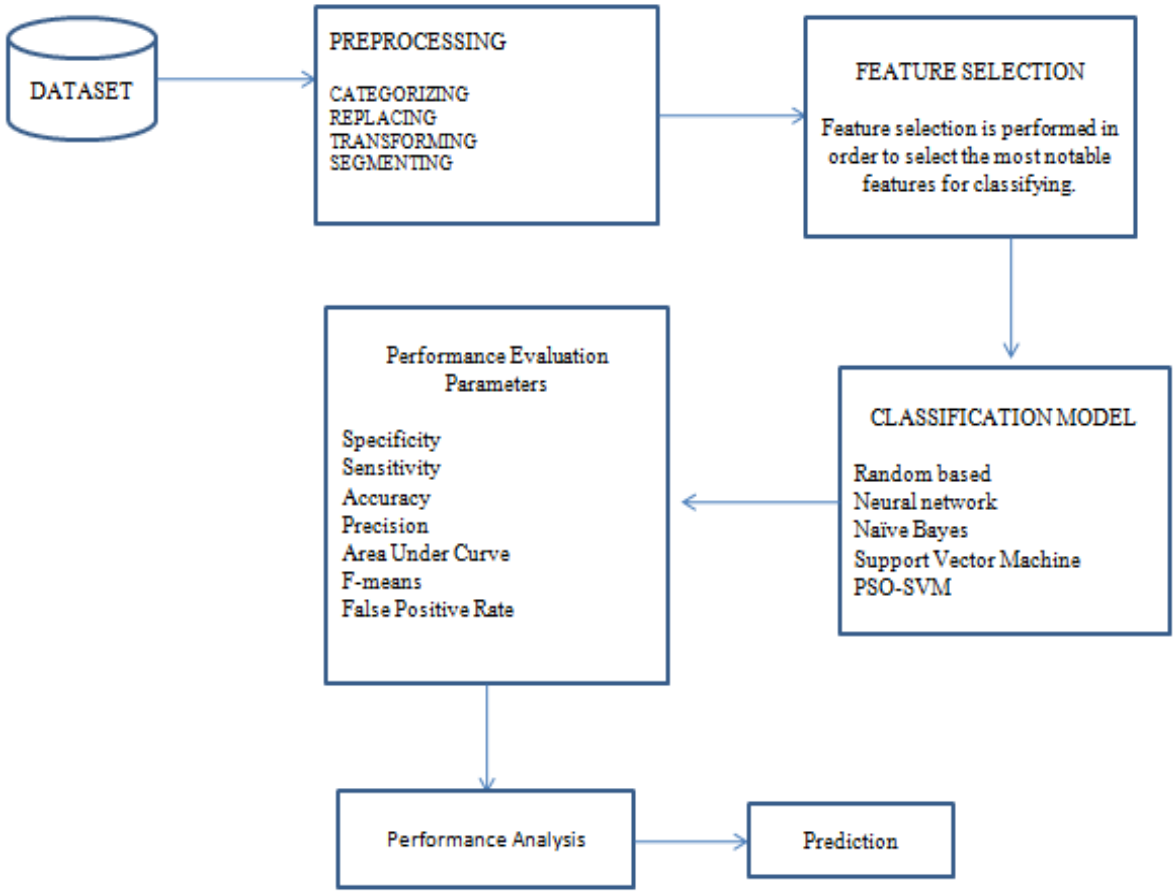


Fig. 9 A diagram showing a recommended approach.

4.1 Data Pre-processing

After the dataset is gathered, the information is pre-handled to target mark values, track down missing qualities, and fragment the information. The pre-processing steps include the following procedure:

1. Four records with missing values had the most appropriate value, which is related to the Albumin globulin ratio, added to them. Additionally, age feature can be broken down into three age groups: between 30 and 61, over 61, and below 30
2. The target input data, which consists of 583 records, and the liver disease target data are the primary divisions of the data. The patient's liver disease status is indicated by the additional grouping of data in the target data for liver disease.
3. From that point onward, the dataset is parted into two gatherings, one for preparing and the other for testing, with 60% of the dataset utilized for preparing and 40% for testing.
4. The classification models are currently preparing the data for use. The stage for classifying the model begins in this step.

4.2 Choosing the features for the classification model.

The significance of predicting liver disease is one important aspect of this study. Only two of the eleven features—"gender" and "age"—include particulars about each patient. The rest are purely clinical aspects. The weighted approach of categorising the items has been utilised in this experiment evaluation.

Therefore, in order to identify the worthy characteristics for predicting liver disease, their performance ought to be estimated and quantified in conjunction with their notable features. The combination of different features together is a helpful strategy for predicting liver sickness. Weight by information gain is a huge method to find the components that can predict the sickness effectively. After providing an unwavering value for the component in the associated model, the acquired data lists the dataset's change in entropy.

Info $gain(S, X) = \sum_{i=1}^{N \in C1}$ Pi X Info entropy(S1)

Info *entropy(S)*$(S) = \sum_{j=1}^{N \in Ci} Pilog2Pi$

Essential phosphate is the most critical marker factor with a value of 0.71 and the most second pointer factor is age with a value of 0.26. All loads acquired through data gain should be picked for drive expectation utilizing the Bayesian organization. The SVM chooses all of its features based on weight. The weights are updated by the multi-layer neural network using backpropagation after each iteration. The weight indicates the degree to which the features are interconnected. The hidden layer, the input layer, and the hidden layer all have random weights in a neural network. The output layer and hidden layer's weights are determined analytically at random [61]. If the weights are close to zero, the change in input will have no effect on the output. Seven components were selected for the weight by Data gain PSO-SVM model. The most basic stage for PSO-SVM is picking the data for growing the accuracy for assumption and administering weight to features by information gain.

4.3 Creating a Classification Model Following data processing, WEKA is used to implement the preferred models. Using Weka, this investigation applies SVM, arbitrary woods, Bayesian organization, PSO-SVM, and MLP-brain network to the arrangement model.

5. Performance Evaluation

The confusion matrix is used to evaluate and scrutinise the model. As was brought up in the data pretreatment section, the dataset is separated into two groups: training and testing. Consider that the models get both test and preparing information, with 60% filling in as preparing information and 40% filling in as test information. In addition, the average Weka accuracy was determined by computing the model using the 5-fold and 10-fold cross-validation methods. In order to evaluate the production of models, specificity, accuracy, and sensitivity—also known as FPR, F measure, and precision—are enumerated. The preparation dataset is utilized to fabricate models, and the testing dataset is utilized to make sure that the model is precise.

Exactness is equal to True positive + True Negative/True Positive + True Negative + False Positive + False Negative.

Genuine Negative Rate (True Negative Rate) = True Negative/True Negative + False Positive= Explicitness

Accuracy corresponds to TP/TP + FP

Genuine Positive Rate (TPR) or Responsiveness = TP/TP + FN

F measure = recall/accuracy + review + 2*precision

FPR= 1 - Explicitness

Here, TP represents "True Positives," whereas TN indicates for "True Negatives." In addition, the terms FP and FN refer to false negatives. The following are the guiding principles:

As indicated in Table 4, a confusion matrix is used to assess the standard for confusion matrix via offering visualization of the model's performance by contrasting how miscategorized samples differ from genuine ones.

FN = The number of records that are actually favorable though are anticipated to be negative.

TP = The proportion of anticipated positive records that genuinely emerge as positive records.

TN = the percent of records that are certainly forecasted to be negative.

Table 4. Confusion matrix of testing dataset.

Models	TP	FN	FP	TN
Bayesian network	142	20	12	59
SVM	133	30	28	42
Random Forest	142	20	12	59
Neural MLP network	137	29	22	45
PSO- SVM	150	8	5	70

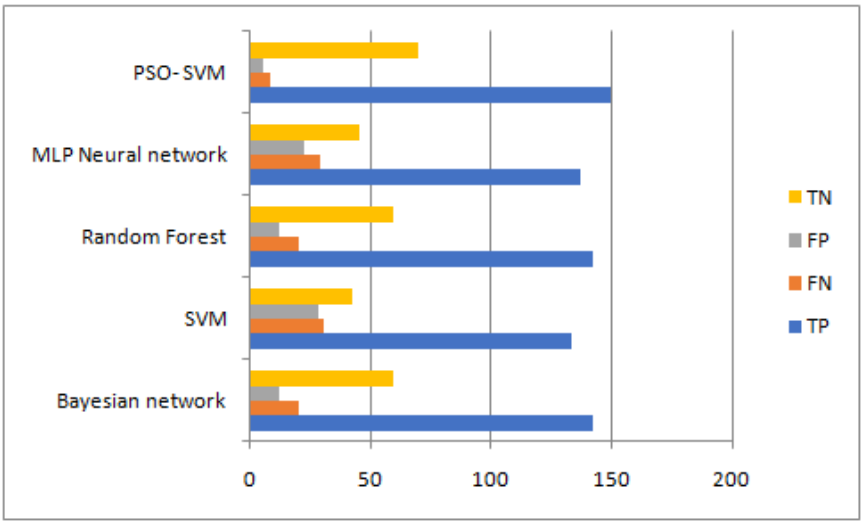


Fig 10. Bar graph of Confusion matrix of dataset.

6. Metaheuristic Approach

Metaheuristic solves difficult optimization issues by employing computational intelligence. Metaheuristics are additionally assembled into two sections analogy based and non-similitude based [62]. The majority of performance-related algorithm research is quantitative, and performance validation metrics like mean error, standard deviation, and correlations are utilized. The metaheuristics are taken on because of their efficient exhibition, the elevated number of references, determined developmental administrators, interesting correspondence systems between people, dealing with ideas, and stagnation counteraction techniques. Hybrid metaheuristics and parallel metaheuristics have emerged as a result of the foundation of new research opportunities. Metaheuristic techniques can computationally systematically precede a near-optimal solution, but they do not guarantee finding the exact optimal solution.

6.1 Particle Swarm Optimization

Swarm-based optimization algorithms are shown to be effective methods that are inspired by how social animals act together. A swarm of particles running through the parameter space and defining trajectories that are guided by their neighbor's own best performance is what is meant by the term "particle swarm optimization" (PSO). This group of potential solutions to the optimization problem is referred to as the group of candidate solutions. In the search space, PSO is used to create new sets of attributes, and semblance is used to calculate divergent possible sets of attributes. The attribute with the greatest coherence is assumed to have the best fitting one.

$x_{i,j}^{(k+1)} = x_{I,j}^{(k)} + V_{i,j}^{(k+1)}$

$v_{i,j}^{(k+1) = NV}{}_{i,j}^{(k)} + c1r1^{(k)} (p_{i,j}^{(k)} – x_{I,j}^{(k)} + c2r2^{(k)}(g_j^{(k)} - x_{i,j}^{(k}$

$V_{i,j}$ is is the velocity of each particle in the dimensions I and j, with K is the iterations index,

$X_{I,j}$ constitute for each particles position,

N represents the inertia of weight which balance global or local search,

$P_{i,j}$ implies each particle's best prior position,

G_j reveals world's best,

Two positive constants named C1 and C2 possess a value equals 2.,

r1 and r2 represents two random values between 0 and 1 in each iteration

6.2 Hybrid Metaheuristic Approach (PSO-SVM)

A general algorithm framework that, when applied to a different optimization problem, requires relatively fewer modifications to adapt to particular problems is one definition of a hybrid metaheuristic [63]. Hybrids perform better than standalone systems when there is synergy. Defining up piece limits In SVM getting ready and picking the fitting components increase the course of action accuracy. The PSO-SVM F-score strategy and further developed highlight determination both make it easier to select the precise component. By starting with a particular variable and then creating their subset to improve the classification problem, feature selection in the hybrid metaheuristic approach focuses on the subset of the variable with the best way to classify the problem's features [64]. The metaheuristic method of examining the entire space yields great arrangements.

7. Experimental Results

The confusion matrix provides results by selecting the most prominent features. The show is being checked out and the precision of Sporadic boondocks, MlP cerebrum association, Bayesian association, and SVM model was 86.26%, 78.11%, 66.09%, and 75.10% exclusively. PSO-SVM is accurate to 94.62 percent. PSO-SVM has the best prescient qualities, as well as the most elevated F-measure, accuracy, awareness, and particularity. Likewise, the Bayesian organization, SVM, MLP brain organization, Irregular timberland, and PSO-SVM models have normal assessed upsides of 66.78%, 76.51%, 78.91%, 87.35%, and 95.17% while utilizing 10-overlap and 5-overlay cross approval techniques on a liver dataset in this review. However, only PSO-SVM achieved 95.21 % accuracy during 5-fold cross-validation, outperforming the 10 fold cross-validation method. Table 5 records the processed outcomes for exactness, explicitness, accuracy, responsiveness, F-measure, and FPR for different characterization models.

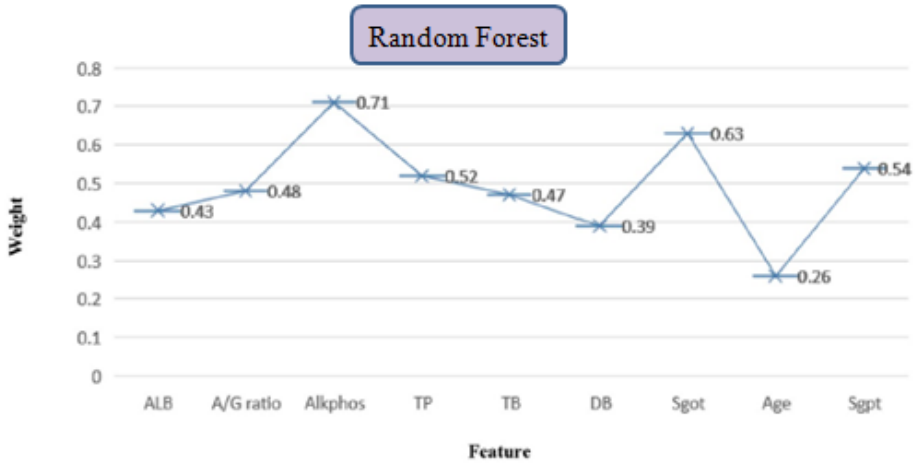


Fig 11. Significant feature selected for random forest.

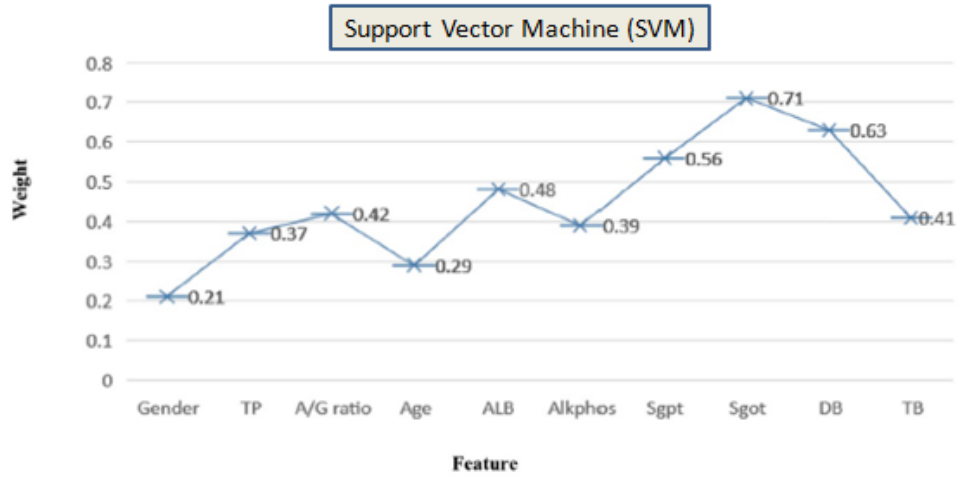


Fig 12. Significant feature selected for Support Vector Machine.

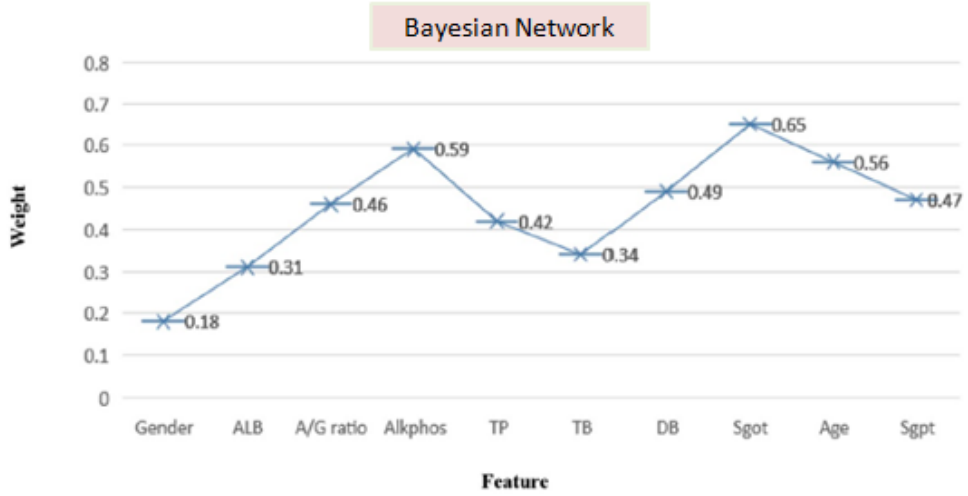


Fig 13. Significant feature selected for Bayesian Network

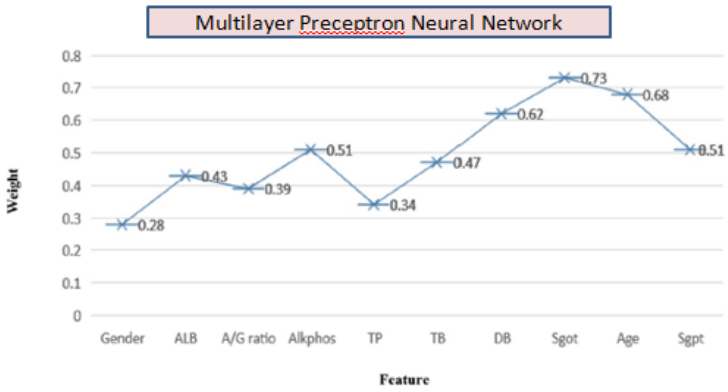


Fig 14. Significant feature selected for Multi Layer Preceptron Neural network.

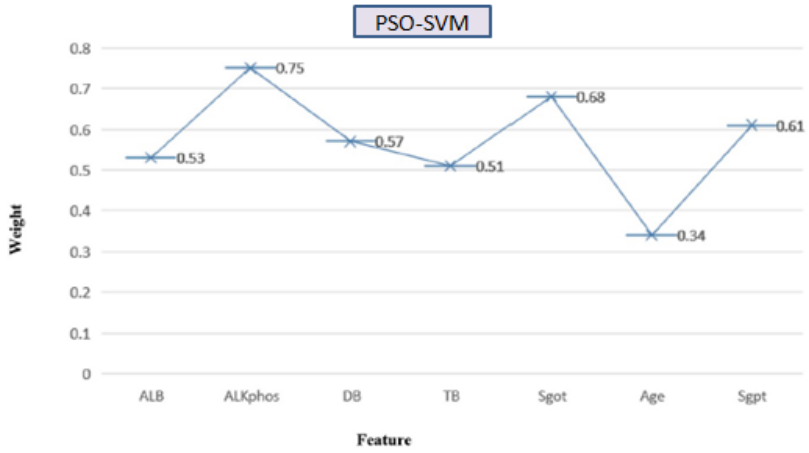


Fig 15. Significant feature selected for PSO-SVM.

Table 5. Result of evaluation criteria for classification models.

Classification Models	Accuracy	Specificity	Precision	Sensitivity	F-measure	FPR
Naives Bayes	66.09	44.28	75.92	75.46	75.68	55.72
Random Forest	86.26	83.09	92.20	87.65	89.86	16.91
MLP Neural Network	78.11	67.16	86.16	82.53	84.30	32.84
SVM	75.10	60.00	82.60	81.59	82.09	40.00
PSO-SVM	94.62	93.33	96.77	94.43	95.84	6.67

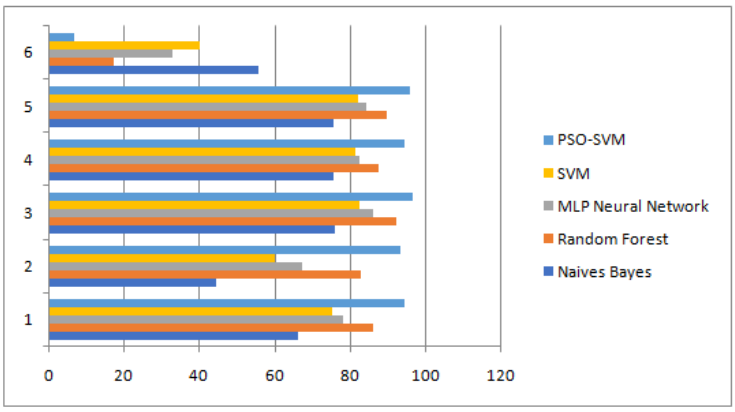


Fig16. Bar graph for evaluation criteria for classification model.

Preprocessing and investigation are the two phases of the information mining process, and the best arrangement is achieved by combining PSO and SVM to improve execution. Age is the least important distinctive scoring 0.34, while alkphos scores 0.75, making it the most significant feature. The best model for choosing and voting on seven aspects is PSO-SVM. According to Table 6, PSO-SVM outperforms all other models in studies on liver data sets due to its highest accuracy. Table 7 displays significant feature selection for various classification models and their accuracy on the liver dataset set that was previously used in this study.

When the K Star classification model was applied to a liver dataset with co-relation feature selection, an accuracy of 73.07 percent was achieved[65]. Research on liver data sets from India and the United States[65] revealed only three significant K-nearest neighbor features. Ten relevant characteristics were chosen for this classification approach pursuant to the ranking methodology [66], and the Nb-tree's calculated accuracy is 67.01. Although the decision tree's accuracy was 69.40, no noteworthy traits were picked [67]. The accuracy of the ANN classification model at k=3 was 68.49% [68].

Table6. Comparison of different models with their significant features.

Classification Models	Feature selection method	Significant feature Values	Accuracy in %
NEURAL NETWORK concerning MLP	Information-based weight gain	A/g;0.39,TP:0.34,DB;0.62,Sgpt:0.51, Gender:0.28,SGOT:0.73,ALB:0.44,Alkph:0.51,AGE=0.68, Tb:0.47,	78.11
RANDOM FOREST ALGORITHM	Gaining weight via information	ALB:0.43,Sgot:0.54,Alkph:0.71,TP:0.52,DB:0.39A/G ratio:0.49, TB:0.47,	86.26
PSO-SVM	Development in weight thru information	Sgpt:0.61,Alkp:0.75,Tb:0.51,Age:0.34, Sgot:0.68. Alb:0.53, Db:0.57.	94.62
Bayesian Network	Weight earned with information	Sgot:0.65,Gender:0.18,Age:0.56,Alb:0.31,Sgpt:0.47, Alkph:0.60,Tp:0.42, Db:0.49, Gender: 0.18, Tb:0.34, A/gratio:0.47.	66.09
SVM	Weight employing SVM	SGPT: 0.56, SGOT:0.72, Gender: 0.21, TP: 0.38, A/G ratio:0421, dB:0.63, tb:0.42, ALB:0.48,Alkphos:0.59, Age:0.29,	75.10

Table 7. Accuracy for previous research conducted on liver dataset on different classification models.

Classification models	Outcome for selected features	Accuracy in %
K nearest neighbor	Sgpt, Sgot, and Alkphos	-------
K Star	----------	73.07
Decision tree	nil	69.40
ANN	na	68.49
In this Study	Age, ALB, Tb, Db, ALB, Sgot, and Alkphos	94.62
NB tree	Age, Tb,Alkphos, Sgot, Db , A/g, Sgpt, Ratio,Tp,Alb and gender	67.01

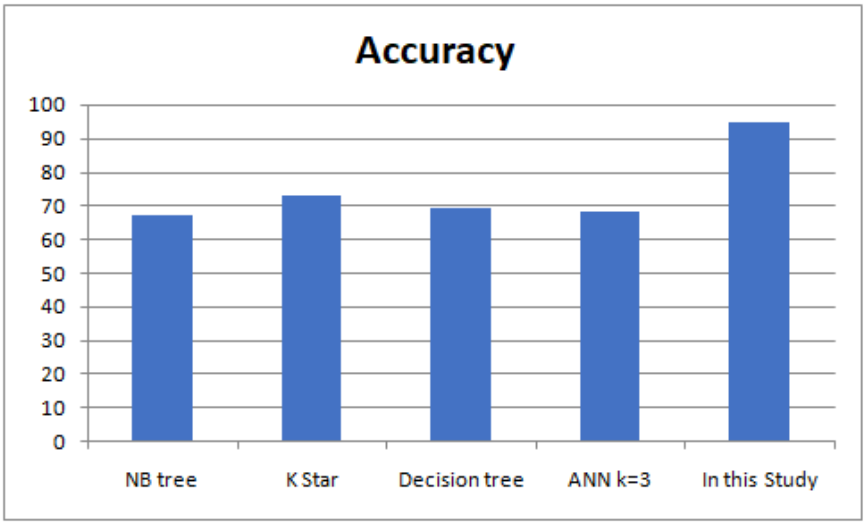


Fig 17. Comparison of previous classification models with current research.

8. Conclusion

This study uses the UCI dataset for exploring medical data mining for liver illness. Predicting liver disease with great accuracy is often favorable. The primary objective of this study is to accurately predict liver disease by selecting the primary risk factor for liver infection. MLP-brain organization, PSO-SVM, irregular backwoods, Bayesian organization, and SVM are the five models viable for characterization. With the greatest amount of accuracy, the seven most significant elements are really long. The PSO-SVM model outperforms all of the others when precision, FPR, accuracy, specificity, AUC, sensitivity, and F-measure models are compared. Iteration ranges of up to 100 iterations and 10-fold cross-validations are used to evaluate the models. When contrasted with different models, PSO-SVM has the most noteworthy typical precision. These findings will be helpful in determining whether people have liver cirrhosis in the real world. The meta-heuristic approach has the potential to enhance classification models like genetic models, ant colony systems, and Bat-inspired algorithms in the field of liver diseases. Additionally, the accuracy of liver cirrhosis disease prediction would be enhanced by significant feature selection techniques in deep neural networks.

References

[1] Sepanlou, S. G., Safiri, S., Bisignano, C., Ikuta, K. S., Merat, S., Saberifiroozi, M., ...&Padubidri, J. R. (2020). The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet gastroenterology &hepatology*, *5*(3), 245-266.

[2] Decharatanachart, P., Chaiteerakij, R., Tiyarattanachai, T., &Treeprasertsuk, S. (2021). Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis. *BMC gastroenterology*, *21*(1), 1-16.

[3] Duncan, B. B., Schmidt, M. I., & Global Burden of Disease 2017 Cirrhosis Collaborators. (2020). The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The lancet. Gastroenterology &hepatology. Amsterdam. Vol. 5 (2020), p. 245-266.*

[4] Sepanlou, S. G., Safiri, S., Bisignano, C., Ikuta, K. S., Merat, S., Saberifiroozi, M., ...&Padubidri, J. R. (2020). The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet gastroenterology &hepatology*, *5*(3), 245-266.

[5] Bellentani, S., Scaglioni, F., Marino, M., &Bedogni, G. (2010). Epidemiology of non-alcoholic fatty liver disease. *Digestive diseases*, *28*(1), 155-161.

[6] Estes, C., Razavi, H., Loomba, R., Younossi, Z., &Sanyal, A. J. (2018). Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*, *67*(1), 123-133.

[7] Collier, J. (2007). Bone disorders in chronic liver disease. *Hepatology*, *46*(4), 1271-1278.

[8] Peris, P., Guanabens, N., Pares, A., Pons, F., Del Rio, L., Monegal, A., ...& Muñoz-Gómez, J. (1995). Vertebral fractures and osteopenia in chronic alcoholic patients. *Calcified tissue international*, *57*(2), 111-114.

[9] Sempoux, C., Jibara, G., Ward, S. C., Fan, C., Qin, L., Roayaie, S., ...&Thung, S. N. (2011, February). Intrahepatic cholangiocarcinoma: new insights in pathology. In *Seminars in liver disease* (Vol. 31, No. 01, pp. 049-060). © Thieme Medical Publishers.

[10] Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., ...& Forman, D. (2012). Bray FGLOBOCAN. v1. 1, Cancer incidence and mortality worldwide: IARC CancerBase no. 11 [internet]. *Lyon. France: International Agency for Research on Cancer*, 2014.

[11] Mantry, P. S., Mehta, A., Madani, B., Mejia, A., &Shahin, I. (2017). Selective internal radiation therapy using yttrium-90 resin microspheres in patients with unresectable hepatocellular carcinoma: a retrospective study. *Journal of Gastrointestinal Oncology*, *8*(5), 799.

[12] El–Serag, H. B., & Rudolph, K. L. (2007). Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*, *132*(7), 2557-2576.

[13] Fattovich, G. (2003). Natural history and prognosis of hepatitis B. In *Seminars in liver disease* (Vol. 23, No. 01, pp. 047-058). Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662.

[14] Rockey, D. C., Caldwell, S. H., Goodman, Z. D., Nelson, R. C., & Smith, A. D. (2009). Liver biopsy. *Hepatology*, *49*(3), 1017-1044.

[15] Yeom, S. K., Lee, C. H., Cha, S. H., & Park, C. M. (2015). Prediction of liver cirrhosis, using diagnostic imaging tools. *World journal of hepatology*, *7*(17), 2069.

[16] Zardi, E. M., &Caturelli, E. (2007). May sonography distinguish between liver fibrosis and liver steatosis?. *Digestive and Liver Disease*, *39*(8), 790.

[17] Kudo, M., Zheng, R. Q., Kim, S. R., Okabe, Y., Osaki, Y., Iijima, H., ...&Kojiro, M. (2008). Diagnostic accuracy of imaging for liver cirrhosis compared to histologically proven liver cirrhosis. *Intervirology*, *51*(Suppl. 1), 17-26.

[18] Faria, S. C., Ganesan, K., Mwangi, I., Shiehmorteza, M., Viamonte, B., Mazhar, S., ...&Sirlin, C. B. (2009). MR imaging of liver fibrosis: current state of the art. *Radiographics*, *29*(6), 16.

[19] Smith, A. D., Branch, C. R., Zand, K., Subramony, C., Zhang, H., Thaggard, K., ...& Zhang, X. (2016). Liver surface nodularity quantification from routine CT images as a biomarker for detection and evaluation of cirrhosis. *Radiology*, *280*(3), 771-781

[20] Kudo, M. (2012). Japan's successful model of nationwide hepatocellular carcinoma surveillance highlighting the urgent need for global surveillance. *Liver cancer*, *1*(3-4), 141.

[21] Okazaki, H., Ito, K., Fujita, T., Koike, S., Takano, K., & Matsunaga, N. (2000). Discrimination of alcoholic from virus-induced cirrhosis on MR imaging. *American Journal of Roentgenology*, *175*(6), 1677-1681.

[22] Awaya, H., Mitchell, D. G., Kamishima, T., Holland, G., Ito, K., & Matsumoto, T. (2002). Cirrhosis: modified caudate-right lobe ratio. *RADIOLOGY-OAK BROOK IL-*, *224*(3), 769-774.

[23] Ito, K., & Mitchell, D. G. (2004). Imaging diagnosis of cirrhosis and chronic hepatitis. *Intervirology*, *47*(3-5), 134-143.

[24] Schrier, R. W., Arroyo, V., Bernardi, M., Epstein, M., Henriksen, J. H., &Rodés, J. (1988). Peripheral arterial vasodilation hypothesis: a proposal for the initiation of renal sodium and water retention in cirrhosis. *Hepatology*, *8*(5), 1151-1157.

[25] Iwakiri, Y. (2012). Endothelial dysfunction in the regulation of cirrhosis and portal hypertension. *Liver international*, *32*(2), 199-213.

[26] D'Amico, G., Garcia-Tsao, G., &Pagliaro, L. (2006). Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *Journal of hepatology*, *44*(1), 217-231.

[27] Pomfret, E. A., Fryer, J. P., Sima, C. S., Lake, J. R., & Merion, R. M. (2007). Liver and intestine transplantation in the United States, 1996–2005. *American journal of transplantation*, *7*, 1376-1389.

[28] D'Amico, G., Garcia-Tsao, G., &Pagliaro, L. (2006). Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *Journal of hepatology*, *44*(1), 217-231.

[29] Ishak, K. (1995). Histological grading and staging of chronic hepatitis. *J hepatol*, *22*, 696-699.

[30] Nagula, S., Jain, D., Groszmann, R. J., & Garcia-Tsao, G. (2006). Histological-hemodynamic correlation in cirrhosis—a histological classification of the severity of cirrhosis. *Journal of hepatology*, *44*(1), 111-117.

[31] Choi, B. I., Takayasu, K., & Han, M. C. (1993). Small hepatocellular carcinomas and associated nodular lesions of the liver: pathology, pathogenesis, and imaging findings. *AJR. American journal of roentgenology*, *160*(6), 1177-1187.

[32] Vivanti, R., Szeskin, A., Lev-Cohain, N., Sosna, J., &Joskowicz, L. (2017). Automatic detection of new tumors and tumor burden evaluation in longitudinal liver CT scan studies. *International journal of computer assisted radiology and surgery*, *12*(11), 1945-1957.

[33] Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M., & Greenspan, H. (2016). Fully convolutional network for liver segmentation and lesions detection. In *Deep learning and data labeling for medical applications* (pp. 77-85). Springer, Cham.

[34] Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., ...&Menze, B. H. (2016, October). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In *International conference on medical image computing and computer-assisted intervention* (pp. 415-423). Springer, Cham.

[35] Li, W. (2015). Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. *Journal of Computer and Communications*, *3*(11), 146.

[36] Ben-Cohen, A., Klang, E., Diamant, I., Rozendorn, N., Raskin, S. P., Konen, E., ...& Greenspan, H. (2017). CT image-based decision support system for categorization of liver metastases into primary cancer sites: initial results. *Academic radiology*, *24*(12), 1501-1509.

[37] Liu, X., Song, J. L., Wang, S. H., Zhao, J. W., & Chen, Y. Q. (2017). Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. *Sensors*, *17*(1), 149.

[38] Biswas, M., Kuppili, V., Edla, D. R., Suri, H. S., Saba, L., Marinhoe, R. T., ...&Suri, J. S. (2018). Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. *Computer methods and programs in biomedicine*, *155*, 165-177.

[39] Hassan, T. M., Elmogy, M., &Sallam, E. S. (2017). Diagnosis of focal liver diseases based on deep learning technique for ultrasound images. *Arabian Journal for Science and Engineering*, *42*(8), 3127-3140.

[40] Guo, L. H., Wang, D., Qian, Y. Y., Zheng, X., Zhao, C. K., Li, X. L., ... &Xu, H. X. (2018). A two-stage multi-view learning framework based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images. *Clinical hemorheology and microcirculation*, *69*(3), 343-354.

[41] Thielke, F., Kock, F., Hänsch, A., Georgii, J., Abolmaali, N., Endo, I., ...& Schenk, A. (2022, April). Improving deep learning based liver vessel segmentation using automated connectivity analysis. In *Medical Imaging 2022: Image Processing* (Vol. 12032, pp. 886-892). SPIE.

[42] Wang, S. H., Du, J., Xu, H., Yang, D., Ye, Y., Chen, Y., ...& Yang, Z. H. (2021). Automatic discrimination of different sequences and phases of liver MRI using a dense feature fusion neural network: a preliminary study. *Abdominal Radiology*, *46*(10), 4576-4587.

[43] Abajian, A., Murali, N., Savic, L. J., Laage-Gaupp, F. M., Nezami, N., Duncan, J. S., ...&Chapiro, J. (2018). Predicting treatment response to intra-arterial therapies for hepatocellular carcinoma with the use of supervised machine learning—an artificial intelligence concept. *Journal of Vascular and Interventional Radiology*, *29*(6), 850-857.

[44] Erkan, M., Ahmetoglu, A., Cansu, A., &Erkut, M. (2021). Evaluation of Sarcopenia and Investigation of Prognostic Value OfSarcopenia Using Psoas Muscle Area on

Computed Tomography in Patients with Liver Cirrhosis. *Electronic Journal of Medical and Educational Technologies*, *14*(3), em2111.

[45] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

[46] Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197-227.

[47] Du, K. L. (2010). Clustering: A neural network approach. *Neural networks*, *23*(1), 89-107.

[48] Yasaka, K., Akai, H., Abe, O., & Kiryu, S. (2018). Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology*, *286*(3), 887-896.

[49] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).

[50] Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *403*.

[51] Zhang, H., Jiang, L., & Su, J. (2005, July). Hidden naive bayes. In *Aaai* (pp. 919-924).

[52] Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. *Swarm intelligence*, *1*(1), 33-57

[53] Eberhart, R., & Kennedy, J. (1995, November). Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks* (Vol. 4, pp. 1942-1948).

[54] Kennedy, J., & Eberhart, R. (1995, November). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942-1948). IEEE

[55] Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567

[56] Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101-121). Academic Press.

[57] Huang, C. L., & Dun, J. F. (2008). A distributed PSO–SVM hybrid system with feature selection and parameter optimization. *Applied soft computing*, *8*(4), 1381-1391.

[58] Jiang, H., Tang, F., & Zhang, X. (2010, December). Liver cancer identification based on PSO-SVM model. In *2010 11th International Conference on Control Automation Robotics & Vision* (pp. 2519-2523). IEEE.

[59] https://t.co/jHt3LL3pZO.

[60] https://www.kaggle.com/datasets/uciml/indian-liver-patient-records.

[61] Gaier, A., & Ha, D. (2019). Weight agnostic neural networks. *Advances in neural information processing systems*, *32*.

[62] Abdel-Basset, M., Abdel-Fatah, L., & Sangaiah, A. K. (2018). Metaheuristic algorithms: A comprehensive review. *Computational intelligence for multimedia big data on the cloud with engineering applications*, 185-231.

[63] Blum, C., Puchinger, J., Raidl, G. R., & Roli, A. (2010). A brief survey on hybrid metaheuristics. *Proceedings of BIOMA*, 3-18.

[64] Yusta, S. C. (2009). Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*, *30*(5), 525-534.

[65] Ramana, B. V., Babu, M. S. P., & Venkateswarlu, N. B. (2012). A critical comparative study of liver patients from USA and INDIA: an exploratory analysis. *International Journal of Computer Science Issues (IJCSI)*, *9*(3), 506.

[66] Alfisahrin, S. N. N., & Mantoro, T. (2013, December). Data mining techniques for optimization of liver disease classification. In *2013 International Conference on Advanced Computer Science Applications and Technologies* (pp. 379-384). IEEE.

[67] Jin, H., Kim, S., & Kim, J. (2014). Decision factors on effective liver patient data prediction. *International Journal of Bio-Science and Bio-Technology*, *6*(4), 167-178.

[68] Abdar, M., Zomorodi-Moghadam, M., Das, R., & Ting, I. H. (2017). Performance analysis of classification algorithms on early detection of liver disease. *Expert Systems with Applications*, *67*, 239-251.