Statistical Machine Learning Approaches to Liver Disease Prediction

Team ID: PNT2022TMID48226

Team Leader: P.M.Priyadharshini

Team Member: N.N.Deepika

Team Member: T.G.Yamini

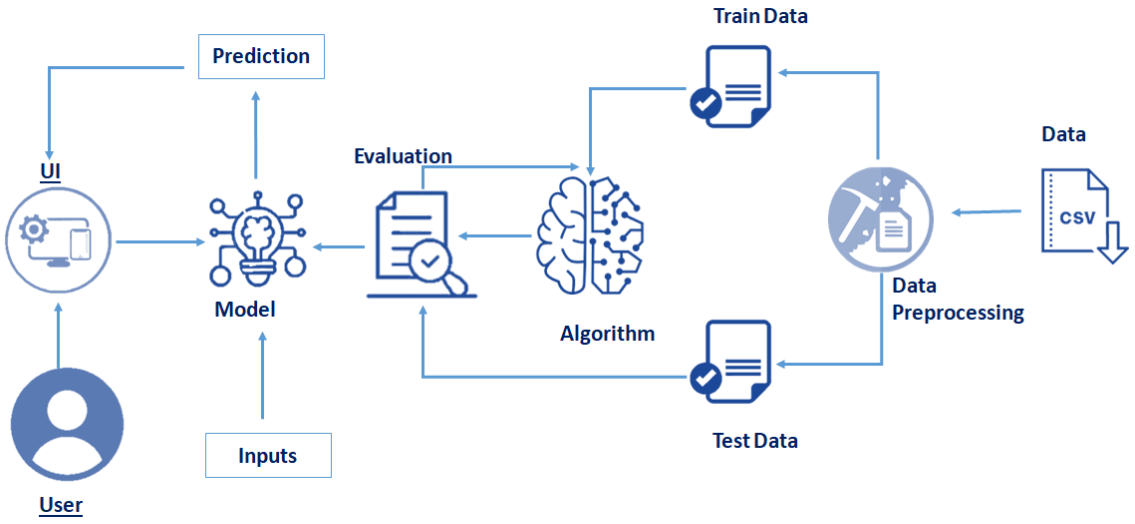
Team Member: S.S.Megha

Learning Outcomes

Statistical Machine Learning Approaches to Liver Disease Prediction:

Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this project is to analyze the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This Project examines data from liver patients concentrating on relationships between a key list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. Here we are building a model by applying various machines learning algorithms find the best accurate model. And integrate to flask based web application. User can predict the disease by entering parameters in the web application.

Technical Architecture:



How to Preprocess Data in Python:

Here's a step-by-step tutorial on data preprocessing implementation using Python, NumPy and Pandas…

In this article, we'll prep a machine learning model to predict who survived the Titanic. To do that, we first have to clean up our data. I'll show you how to apply preprocessing techniques on the Titanic data set.

To get started, you'll need:

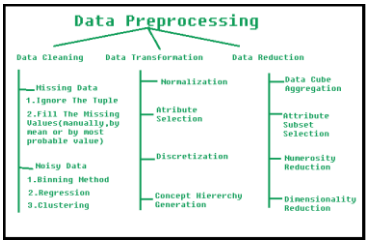
- Python
- NumPy
- Pandas
- The titanic data set

For machine learning algorithms to work, it's necessary to convert **raw data** into a **clean data** set, which means we must convert the data set to **numeric data**. We do this by encoding all the **categorical labels** to column vectors with binary values. **Missing values**, or NaNs (not a number) in the data set is an annoying problem. You have to either drop the missing rows or fill them up with a mean or interpolated values. **Note**: Kaggle provides two data sets: training data and results data. Both data sets must have the same dimensions for the model to produce accurate results.

Data Preprocessing in Data Mining:

Preprocessing in Data Mining
Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

1. Data Cube Aggregation:
2. Attribute Subset Selection:
3. Numerosity Reduction:
4. Dimensionality Reduction:



Data Visualization and Predictive Analysis for Smart Healthcare:

The healthcare industry is one of the most significant sources of Big Data. It is not feasible to manually interpret and understand the huge amounts of data generated by hospitals accurately. This creates the need for a data analytics and visualization tool. Visualizations are intuitive and help interpret the data easily. It would help the hospital to get insights from the data and to provide better service to the society. The aim of this project was to develop a data analysis and visualization tool for a hospital. This was implemented as a web application. The web application is developed using Django, which is a Python-based free and open-source web framework. For the visualizations embedded in the application, the Python library Altair was used. The application supports the upload of files that are the source for the visualizations and provides interactive visualizations based on the analysis performed. The visualizations can be exported as images using the application. Generating visualizations of choice becomes easier using navigation by menu bars in the application rather than writing complicated queries to the database. The tool ultimately attempts to help the hospital optimize time and resources effectively. Prediction using Long Short Term Memory (LSTM) for pharmacy orders and number of orders per patient will further help the hospital predict trends, patterns and outliers. Analysis tools will help analysing past, current and predict future pharmacy and diagnostics in a hospital which ultimately lead to better quality, efficient smart healthcare.

Applying different algorithms according to dataset and based on visualization:

Database Issues for Data Visualization: the word data occurs twice in this title and thus clearly shows the emphasis of these proceedings. Databases and visualizations both deal with data. But each of these fields typically deals with its data independently from the other. Visualization features found in today's database management systems are far from what are provided by visualization systems. And today's visualization systems provide very primitive data access and manipulation tools compared to those that a well developed database management system offers.It is obvious that both groups will benefit significantly from close cooperation. The ever increasing amount of data stored in databases demands new, comfortable ways of access and manipulation. The amount of data that is produced daily by satellites and sent back to Earth to be stored in huge databases require interactive visual support to provide more efficient data management and other tasks. Visual exploration of data stored in databases offers new ways of accessing and interpreting this data. There are many other applications providing examples for the need to integrate databases and visualization. The data that is currently dealt with and visualized in supercomputing environments has grown in step with the ever increasing size of the numerical problems that can currently be computed and solved. The results of time-dependent numerical simulations can easily reach gigabytes in size. To allow a useful interpretation of these results it is important that the data can be accessed and managed efficiently, and thus visualized.On October 28, 1995, the second workshop on Database Issues for Data Visualiza-tion was held during the IEEE Visualization '95 conference in Atlanta, Georgia. Two years after the first and very successful workshop (proceedings of which are available in Lecture Notes in Computer Science Volume 871), researchers from the areas of databases and visualization met again to discuss and share their problems, needs, and goals. The second workshop topics included data modeling and access, object modeling, user interface construction, dataflow and program module storage and retrieval, the composition and manipulation of graphical representations, and the integration of knowledge bases and rule-based systems, all in a visualization context. These proceedings offer a snapshot of current research in this field and together with the proceedings of the first workshop offer a way to see the progress of this emerging area. The proceedings also provide a survey of the problems that must be addressed now and in the future towards the integration of database management systems and data visualization. With these proceedings, the reader is presented with a treatment of a wide range of issues, top to bottom, of the research areas and problems facing the integration of data-base and visualization systems. We hope to further stimulate

research activity in this field, and look forward to the realization of truly integrated systems that accommodate end-user requirements in terms of models, services, displays, and interaction capabilities. We also hope the reader will find the reports and papers as invigorating as the discussion sessions during the workshop.

You will be able to know how to find the accuracy of the model:

In this research work we have established a baseline machine learning model using the Sci-Kit learn package, that can be used for classification, and would result into optimum accuracy and efficiency, i.e., prediction time. From the time when different models were found in machine learning, the major question was, which is the best suited model. In the modern times, with rise of demand of this field and with introduction of more and more complex datasets, this question remains an important one. The models which we have taken into consideration in this research for comparison are all in terms of classification. The datasets on which the models have been implemented belong to different fields, namely, banking sector, wine recognition, medical sector, environmental sector and occupancy detection. All of these datasets are chosen in such a way that they have different size of attributes, patterns and classes. We have generalized the baseline model selection by experimenting with various classification models and implementing them over a variety of datasets. This research work can be upscaled by considering more models and comparing with the existing ones or taking more datasets for implementation or by doing both.

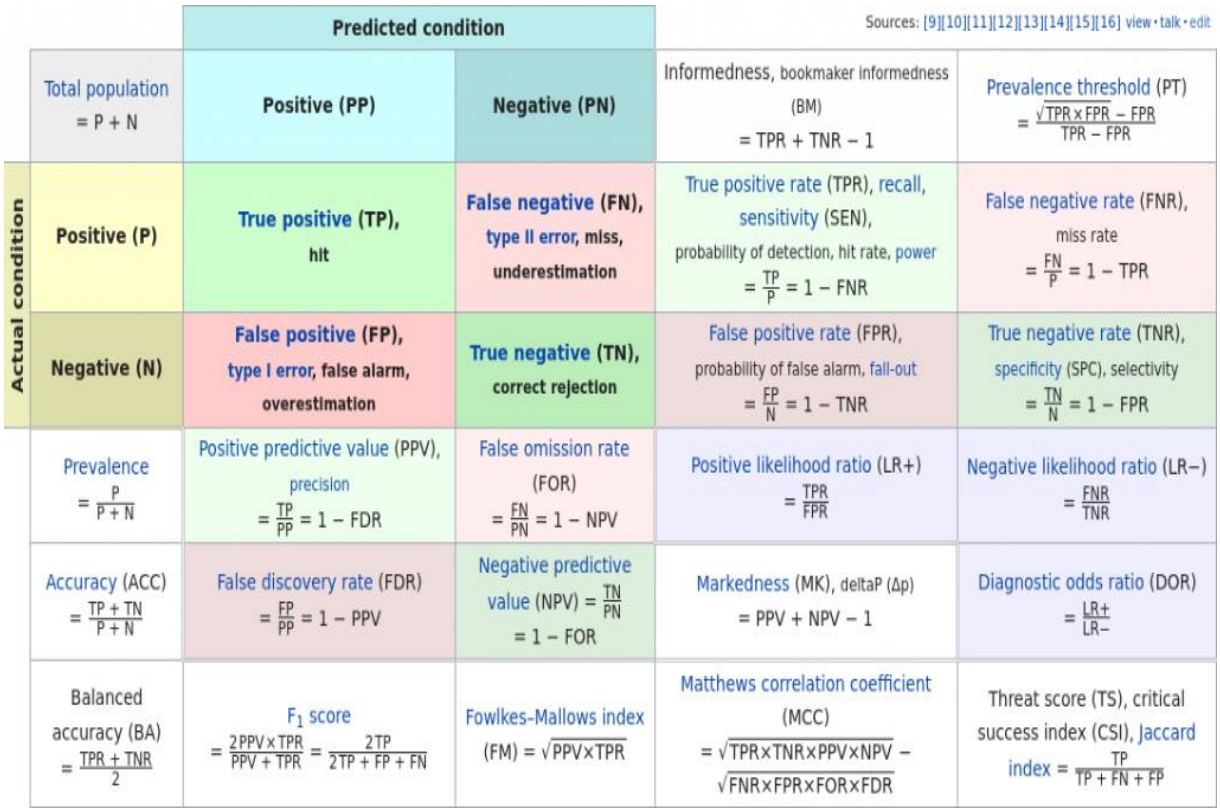
		Actual Class	
		Positive	Negative
Predicted Class	Positive	80	80
	Negative	20	820

Actual Value

Prediction	Positive(1)	Negative(0)
Positive(1)	True Positive(TP)	False Positive(FP)
Negative(0)	False Negative(FN)	True Negative(TN)

In machine learning problems, we usually assume that the validation accuracy is a good estimation of prediction accuracy for datasets without ground truth. In reality, this assumption may not hold. Therefore, we propose an approach to estimate the prediction accuracy of a target model on unlabeled datasets. The proposed approach uses multiple target homogeneous models to assign each unlabeled sample a confidence value, based on the number of models agreeing on the predicted label. With the confidence values, the prediction accuracy of the target model on the datasets can be estimated. In the experiments, the target model is a convolutional neural network (CNN) model, and the homogeneous models

only differ in initial weights. The experiments are conducted with datasets from a wide variety of music genres. The estimation performance of the proposed approach is compared with the reversed testing qualities (RTQ) and the ensemble average qualities (EAQ) approaches. The RTQ approach was proposed to estimate the prediction accuracy of trained models, and the EAQ approach was originally designed for estimating the predictive uncertainty of individual samples. We apply all three compared models to estimate prediction accuracy of datasets by using a linear model. The parameters of the linear model are either computed by using multiple labeled datasets or one labeled dataset. The experimental results show that when compared with the RTQ approach, the proposed approach has much lower estimation errors for some datasets. When compared with the EAQ approach, the proposed approach is more robust for datasets with large distribution shifts. Finally, we show an additional benefit of the proposed approach. In case that the estimated accuracy is unsatisfactory, we may re-train the target model with a new training set, which contains the original training samples plus new training samples with manual labeling from the unlabeled dataset. The experimental results confirm that it is more effective to select (and label) new samples from those with low confidence values than those randomly selected. Overall, the proposed approach is a promising approach for estimating prediction accuracy on unlabeled datasets



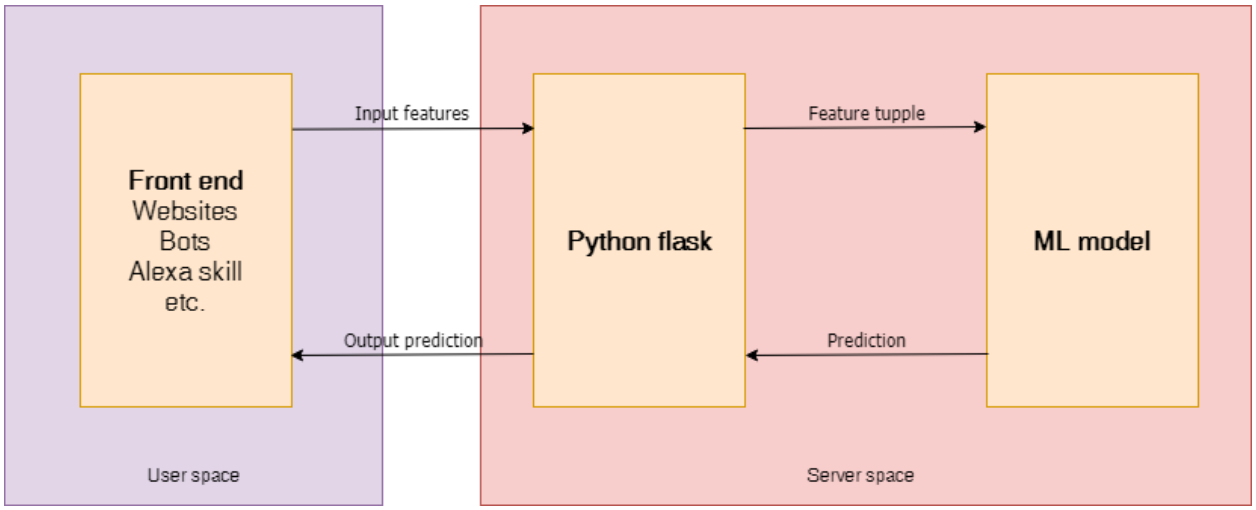
		Predicted condition			
	Total population = P + N	Positive (PP)	Negative (PN)	Informedness, bookmaker informedness (BM) = TPR + TNR − 1	Prevalence threshold (PT) = $\frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	Positive (P)	True positive (TP), hit	False negative (FN), type II error, miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power = $\frac{TP}{P}$ = 1 − FNR	False negative rate (FNR), miss rate = $\frac{FN}{P}$ = 1 − TPR
	Negative (N)	False positive (FP), type I error, false alarm, overestimation	True negative (TN), correct rejection	False positive rate (FPR), probability of false alarm, fall-out = $\frac{FP}{N}$ = 1 − TNR	True negative rate (TNR), specificity (SPC), selectivity = $\frac{TN}{N}$ = 1 − FPR
	Prevalence = $\frac{P}{P+N}$	Positive predictive value (PPV), precision = $\frac{TP}{PP}$ = 1 − FDR	False omission rate (FOR) = $\frac{FN}{PN}$ = 1 − NPV	Positive likelihood ratio (LR+) = $\frac{TPR}{FPR}$	Negative likelihood ratio (LR−) = $\frac{FNR}{TNR}$
	Accuracy (ACC) = $\frac{TP+TN}{P+N}$	False discovery rate (FDR) = $\frac{FP}{PP}$ = 1 − PPV	Negative predictive value (NPV) = $\frac{TN}{PN}$ = 1 − FOR	Markedness (MK), deltaP (Δp) = PPV + NPV − 1	Diagnostic odds ratio (DOR) = $\frac{LR+}{LR-}$
	Balanced accuracy (BA) = $\frac{TPR+TNR}{2}$	F_1 score = $\frac{2PPV \times TPR}{PPV+TPR} = \frac{2TP}{2TP+FP+FN}$	Fowlkes–Mallows index (FM) = $\sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) = $\sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index = $\frac{TP}{TP+FN+FP}$

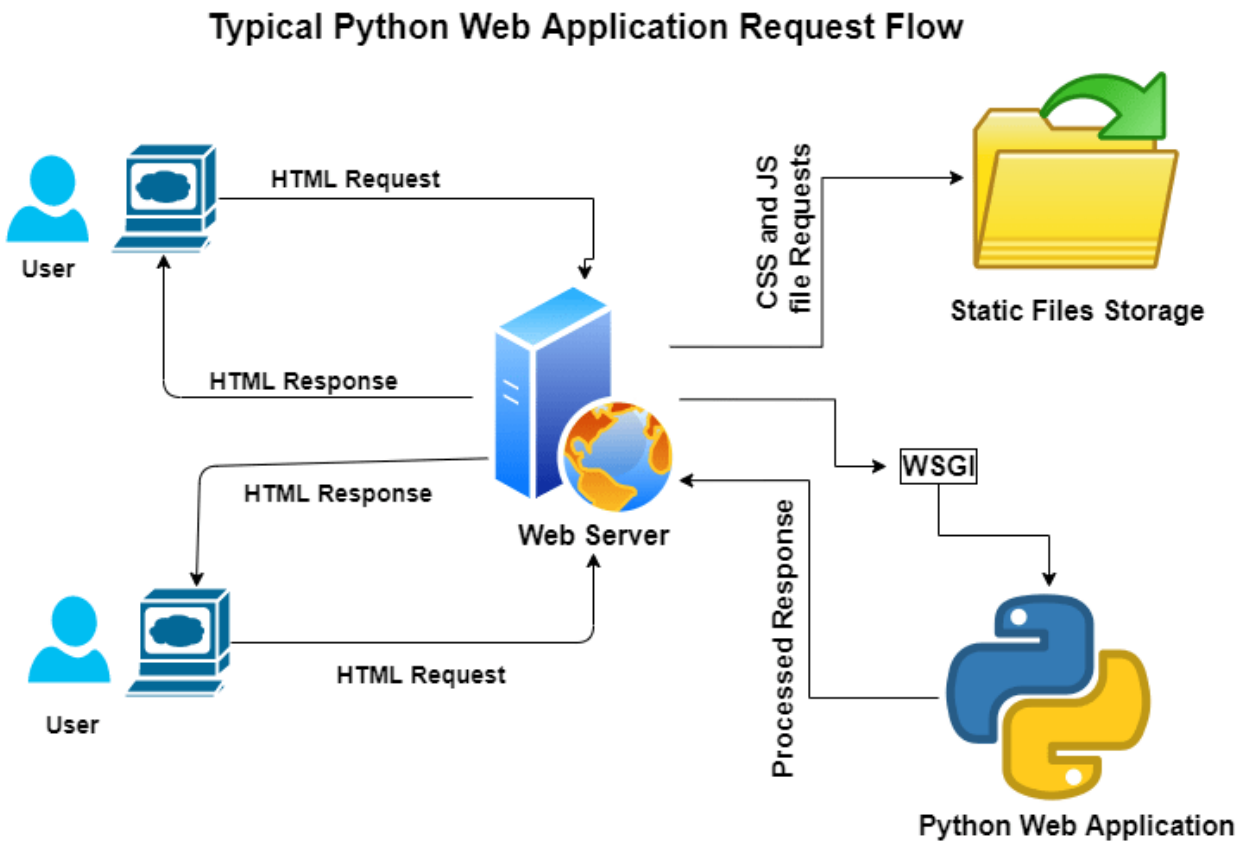
Sources: [9][10][11][12][13][14][15][16] view • talk • edit

Web Application Implementation with Machine Learning:

Every Service nowadays has applications. If we want to order food but don't like to talk to someone, we simply go for the web apps or mobile apps and ordered food and it is the more convenient and easy method. so basically, the web application provides a virtual platform where we do lots of tasks from anywhere all around the globe. Here we make a responsive college community web application that helps the college students, faculty, and alumni to interact on one platform. As the objective of this paper is to elaborate the web application working functionality like the creation of fronted part using React Js, backend part by using Django framework, use of the database in web apps to store data in the database, fetch and serve data in the user interface, deployment of a website in the cloud, all components integration and the main part is the use of machine learning

where we create a Machine learning NLP Model for text analysis and using Scikit learn library for using python and using flask to deployed in Web apps.





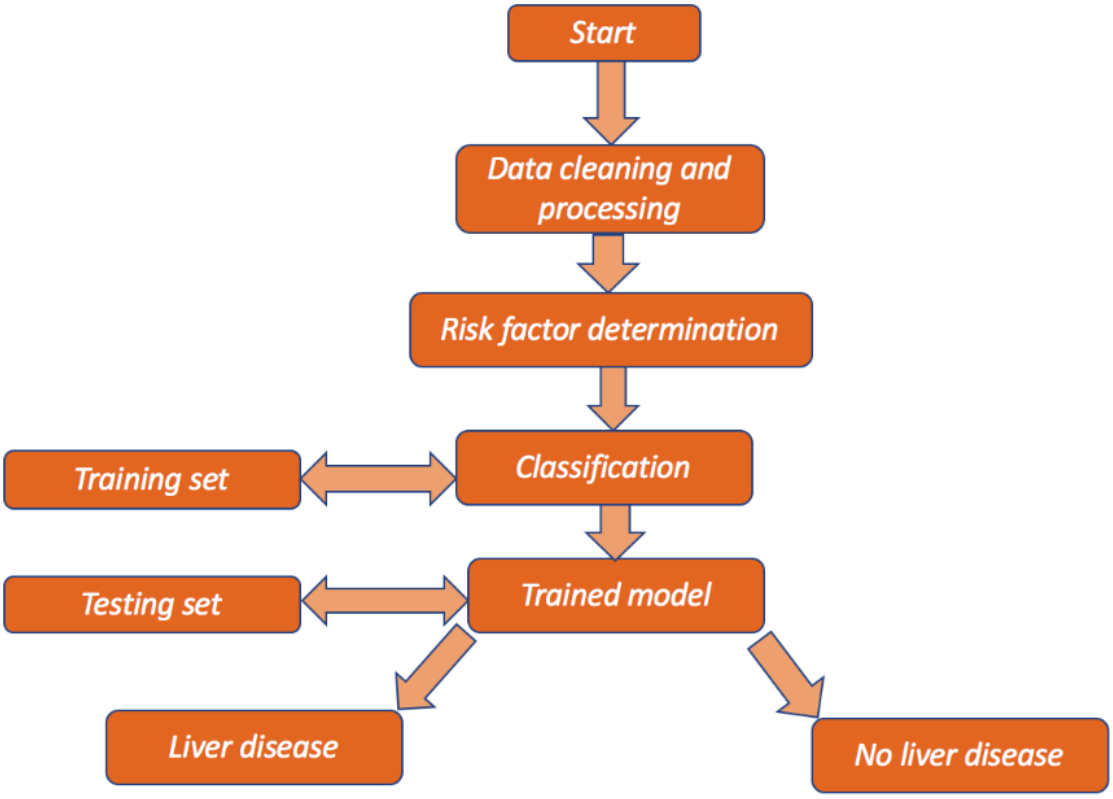
By the end of this project:

Medical diagnoses have important implications for improving patient care, research, and policy. For a medical diagnosis, health professionals use different kinds of pathological methods to make decisions on medical reports in terms of the patients' medical conditions. Recently, clinicians have been actively engaged in improving medical diagnoses. The use of artificial intelligence and machine learning in combination with clinical findings has further improved disease detection. In the modern era, with the advantage of computers and technologies, one can collect data and visualize many hidden outcomes such as dealing with missing data in medical research. Statistical machine learning algorithms based on specific problems can assist one to make decisions. Machine learning (ML), data-driven algorithms can be utilized to validate existing methods and help researchers to make potential new decisions. The purpose of this study

was to extract significant predictors for liver disease from the medical analysis of 615 humans using ML algorithms. Data visualizations were implemented to reveal significant findings such as missing values. Multiple imputations by chained equations (MICEs) were applied to generate missing data points, and principal component analysis (PCA) was used to reduce the dimensionality. Variable importance ranking using the Gini index was implemented to verify significant predictors obtained from the PCA. Training data ($n_{train} = 399$) for learning and testing data ($n_{test} = 216$) in the ML methods were used for predicting classifications. The study compared binary classifier machine learning algorithms (i.e., artificial neural network, random forest (RF), and support vector machine), which were utilized on a published liver disease data set to classify individuals with liver diseases, which will allow health professionals to make a better diagnosis. The synthetic minority oversampling technique was applied to oversample the minority class to regulate overfitting problems. The RF significantly contributed ($p < 0.001$) to a higher accuracy score of 98.14% compared to the other methods. Thus, this suggests that ML methods predict liver disease by incorporating the risk factors, which may improve the inference-based diagnosis of patients. Keywords: liver disease; demographic variables; prognostic/biochemical variables; statistical learning for variable selection andclassification The liver has many functions such as glucose synthesis and storage, detoxification, production of digestive enzymes, erythrocyte regulation, protein synthesis, and various other features of metabolism. Chronic liver diseases include chronic hepatitis, fibrosis, and cirrhosis. Hepatitis can occur from viral infection (e.g., hepatitis c virus) or auto-immune origin. Inflammation from hepatitis infection can cause tissue damage and scarring to occur in the liver. Moderate scarring is classified as fibrosis, while severe liver damage/scarring

is classified as cirrhosis. Fibrosis and cirrhosis can also occur from alcoholism and non- alcoholic fatty liver disease. When liver disease is diagnosed at an earlier stage, in between infection and fibrosis but before cirrhosis, liver failure can be avoided. Tests, such as a CMP and biopsy, can be conducted to diagnose all forms of liver disease. A CMP with a liver function panel can detect albumin (ALB), alkaline phosphatase (ALP), alanine amino-transferase (ALT), aspartate amino-transferase (AST), gamma glutamyl-transferase (GGT), creatine (CREA), total protein (PROT), and bilirubin (BIL). Diagnosis of a certain liver disease and discovery of its origin are made by interpreting the patterns and ratios of circulating liver-associated molecules measured with the CMP test and compared to values normalized with a patient's age, sex, and BMI. Aminotransferases, AST, and ALT are enzymes that participate in gluconeogenesis by catalyzing the reaction of transferring alpha- amino groups to ketoglutaric acid groups. AST is found in many tissue types and is not as specific to the liver but may denote secondary non-hepatic causes of liver malfunction. ALT is found in high concentrations in the cytosol of liver cells. Liver cell injury cancause the release of both aminotransferases into circulation. When ALT is significantly increased in proportion to ALP, the liver disease is likely from an inflammatory origin(acute or chronic viral hepatitis and autoimmune disease)

Study Design:

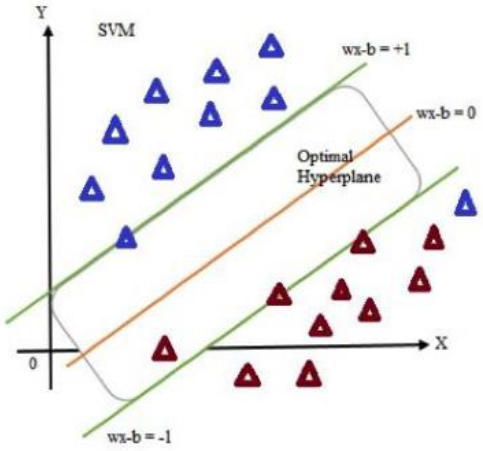


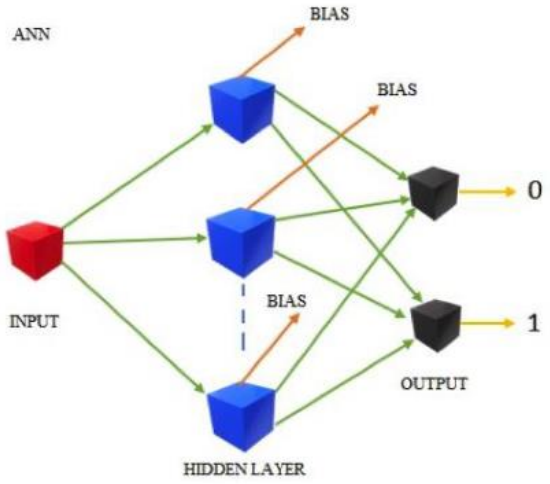
Multiple Imputation by Chained Equations for Missing Data:

Multiple imputation was used via the chained equations method to generate the missing data. For multivariate missing data, the R-package [22] known as "MICE" was used for multiple imputations. This function auto-detects certain variables with missing values. It basically uses predictive mean matching (PMM), which is a semi-parametric imputation. It is very close to regression except missing items are randomly filled by regression prediction. The algorithms for MICE are given below.**Step1**: Start with imputing the mean. Mean imputations are considered "position holders";
Step 2: the "position holder" presents imputations for one variable ("Var") which are impeded to the missing items; **Step 3:** "Var" is the response variable where the other variables are predictor variables in the linear

regression model (under the same assumption); **Step 4**: the missing values for "Var" are then replaced with imputed values from the regression model;
Step 5: Repeat steps 2–4 and produce the missing data.One iteration is needed for each variable and, finally, the missing values. Ten such cycles were performed Exploratory data analysis is used to determine the hidden attributes of a data set. Examples of exploratory methods are correlation heatmaps and box plots to help visualize the data. One of the main goals of this study was to determine the most important metrices that describe almost all of the data set but, at the same time, keeps the loss of information to a minimum. The need for multiple tests per patient increases the cost associated with liver disease but may be required for accuracy in diagnosis. Reducing the dimensions can be helpful for clinicians to determine which biochemical markers are most important for diagnosis and pattern evaluation, therefore, reducing the number of tests for patients in the future. Statistical learning methods, such as PCA, assist in reducing the dimensionality of a data set.. Principal Component Analysis for Dimension Reduction Let $X \in R^{n \times p}$ be the data matrix with the integer k (with $0 < k < p$). PCA [26] can be determined through singular value decomposition (SVD). For this, let $X = (x^T)^T \in R^{n \times p}$ (where $\tilde{x}_i = x_i - x$) be the original and centered data matrices, respectively. Then, the square matrix C is a symmetric and positive semidefinite, which is defined as follows:
The principal direction of the data set is given by the top eigenvectors for the corre-sponding eigenvalues of the covariance matrix. Thus, the dimension of C is p × p. More mathematically, it is the right singular vector where columns of V are the principal directions, Σ is the singular values, where $\lambda_i = \sigma^2_i$ is the eigenvalue by each principal direction, and columns of UΣ are different principal components of the data set X.Training and Testing DataAccording to statistical machine learning techniques, the data set collected from was divided into training and testing data sets, where the trainingset was applied to fit the parameters. A part of training set was used for validationIn fact, it was split into training and test data sets based

on 5-fold cross-validation on the mis-classification error.

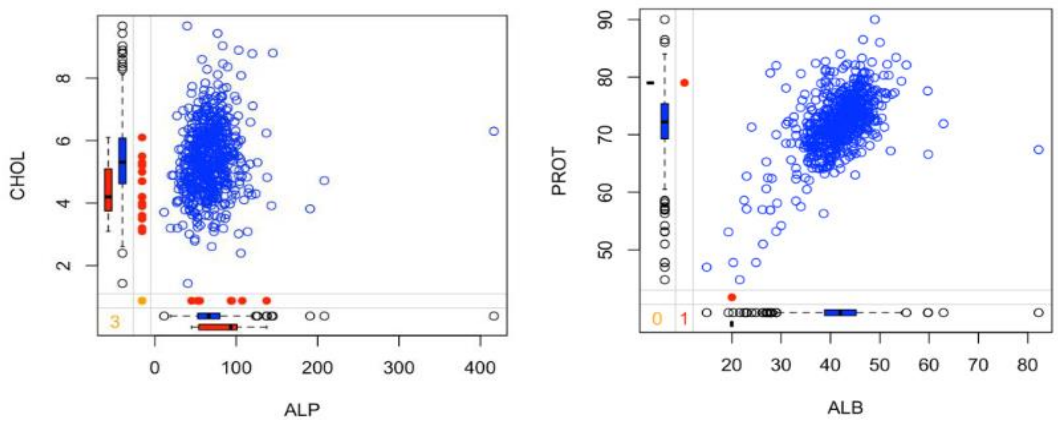




Results and Discussion:

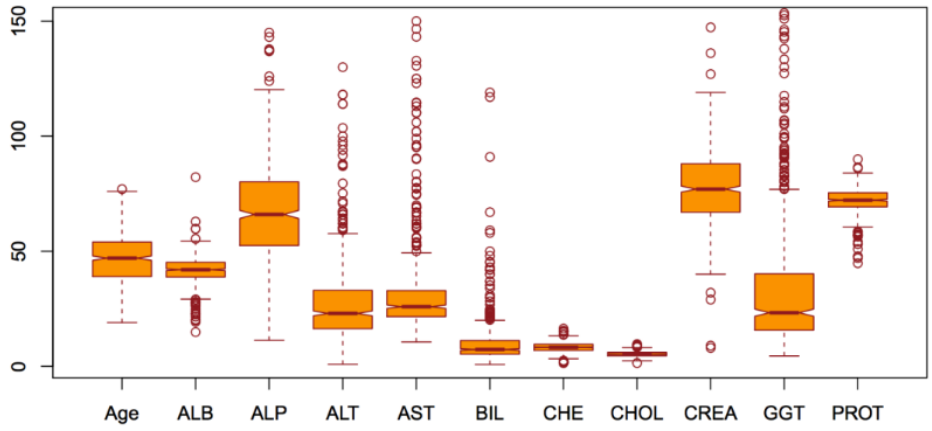
Data visualization techniques were used to plot the summary of the input variables. Using MICE, the missing values were imputed.

Figure 4 investigates the pattern as well as the distribution of incomplete and complete observations of missing input variables. From the exploratory data analysis, ALP, ALB,CHOL, and PROT had missing values. There were 2.22% missing values in total. UsingMICE, missing values were estimated and filled in the data set.Livers 2021, 21, x FOR PEER REVIEW 11 of 19ALB, CHOL, and PROT had missing values. There were 2.22% missing values in total.Using MICE, missing values were estimated and filled in the data

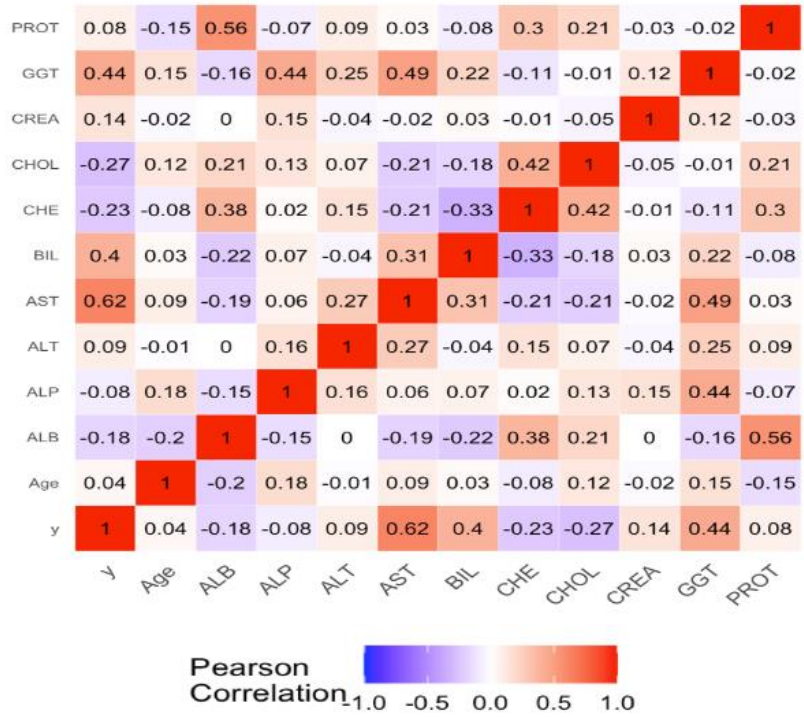


Margin plot for the pattern and distribution of complete and incomplete observations in missing features: (left panel) CHOL versus ALP; (right panel) PROT versos ALB. The blue dots represent observations. In the left and bottom margins, blue box plots are non-missing, and red box plots are the marginal distribution of these observed values. The completed data set was visualized using the box plots in Figure 5. There were several discrepancies between the range and variation of predictors with many outliers.There were some extreme outliers for the following variables: ALT, AST, CREA, and GGT. Figure 5 indicates that some individuals had high amounts of ALT, AST, CREA, and GGT in their blood. Some blood donors may have had elevated ALT, AST, CREA, or GGT due to the fact of a secondary non-hepatic cause. It is also possible that laboratory errors occurred during the initial data collection., 21, x FOR PEER REVIEW 11 of 19 ALB, CHOL, and PROT had missing values. There were 2.22% missing values in total. Using MICE, missing values were estimated and filled in the data Margin plot for the patte rn and distribution of comple te and incomple te observations in missing fe atures: (left panel) CHOL ve rsus ALP; (right panel) PROT ve rsos ALB. The blue dots re pre sent observations. In the le

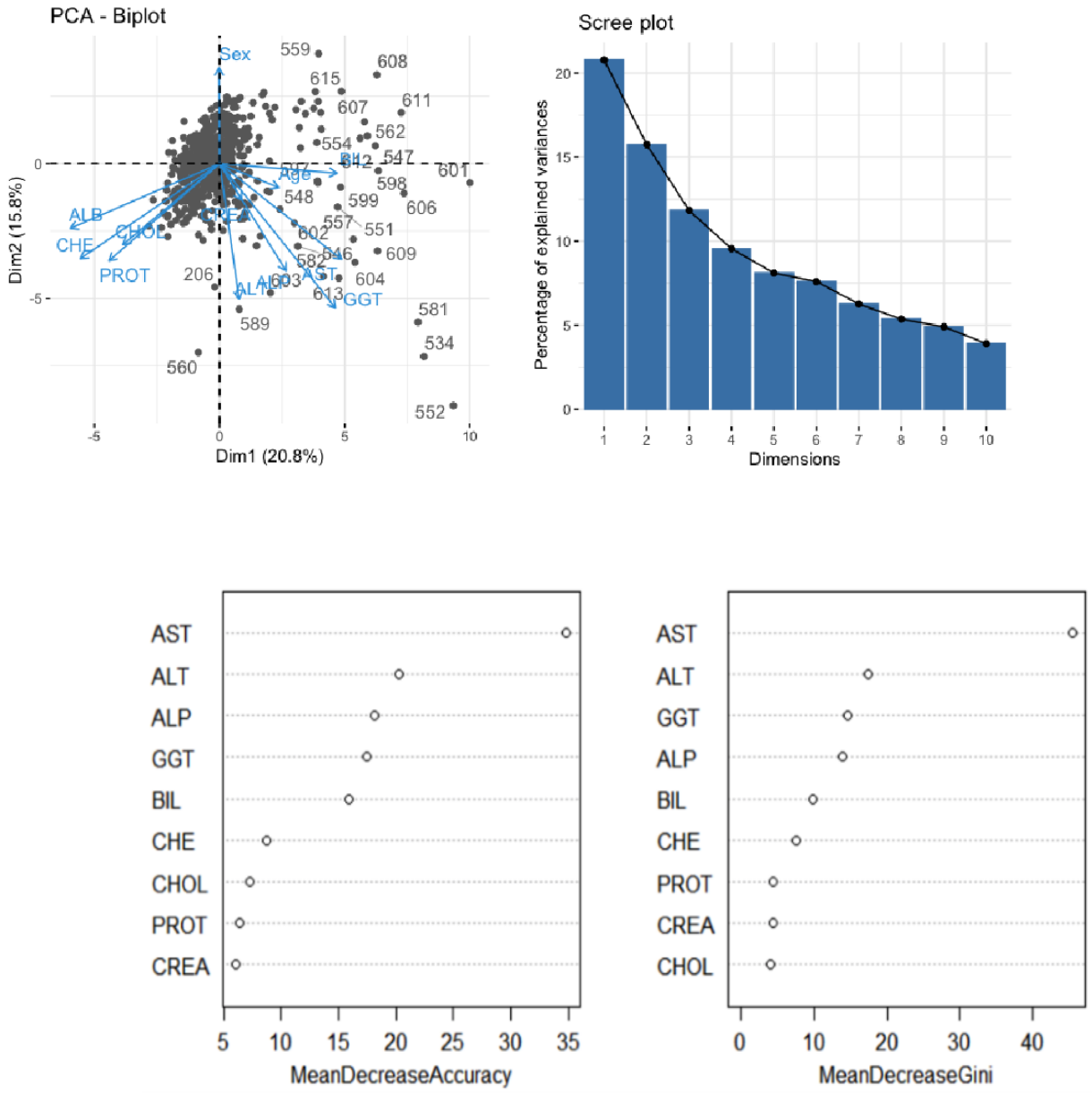
ft and bottom margins, blue box plots are non-missing, and red box plots are the marginal distribution



The binary target variable y had moderately positive relationships with AST, BIL, and GGT. However, y had fairly weak negative relationships with ALP, ALB, CHOL, and CHE. However, the PROT and age variables did not have much of an impact. BIL and GGT were markers of secretion and function specific to the liver; thus, chronic liver disease result edin extremely elevated levels of both values. As such, the correlation between y and BIL was 0.4. The correlation was 0.44 between y and GGT. AST elevation is a significant and commonly occurring risk factor for chronic liver disease, and Figure 6 reflects this with a strong correlation of 0.62 between AST and y.

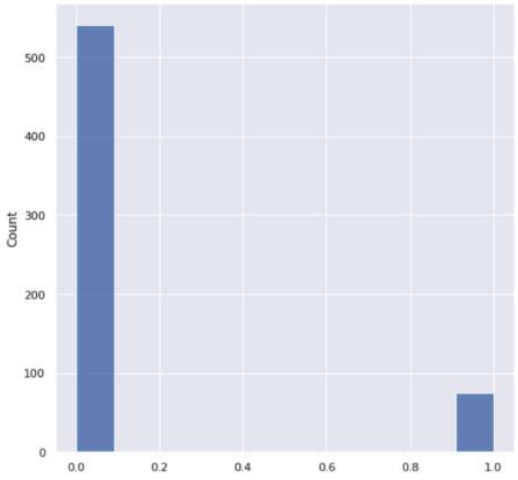


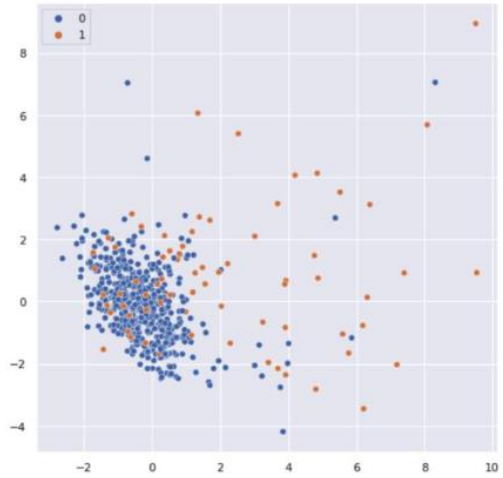
The PCA reduces the dimensionality by projecting each data point into the first few principal components to obtain lower-dimensional data and keep preserving maximum information. In Figure 7, a scree plot (right) is shown to determine the number of input variables. Figure 7 (left panel) shows the principal components with the variables AST,ALT, ALP, BIL, and GGT with almost 85% variability.The variable importance ranking was obtained for RF, and it was measured using the mean decrease in accuracy and mean decrease in Gini as parameters. From Figure 8, AST, ALT, ALP, BIL, and GGT were the most important variables observed in the data set. After comparing with Figure 5 and the results from the PCA, AST, ALT, ALP, BIL, and GGT were used to train the classification models. To confirm the importance of the four variables for disease diagnosis, the following methods determined each variable association with the risk of liver disease development.

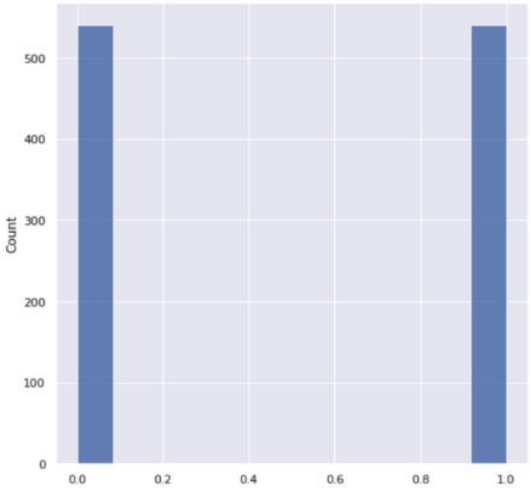


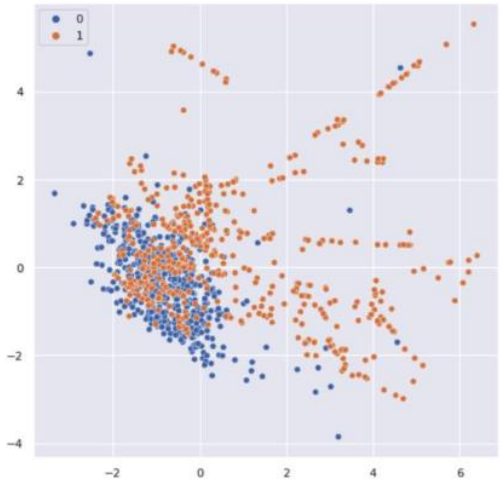
A logistic regression method was performed to calculate the odds ratios and 95%confidence intervals and to determine an association between risk factors and the occur-rence of a liver disease. To reduce the effect of multicollinearity, a correlation analysis among independent variables was conducted, and those which had a variance inflation factor (VIF) greater than three were removed. Considering liver disease with "no" as a reference group, AST (OR = 1.080, 95% CI: 1.050–1.111, p < 0.0001), ALT (OR = 0.981, 95% CI: 0.967–0.995, p = 0.010), ALP (OR = 0.954, 95% CI:

0.935–0.972, p < 0.0001), BIL (OR = 1.080, 95% CI: 1.032–1.130, p = 0.001), and GGT (OR = 1.023, 95% CI: 1.014–1.032, p < 0.0001) were found to be significant risk factors for liver disease. The logistic regression results validated the importance of AST, ALT, ALP.









Before applying binary classification ML methods on the testing data set, the models were trained, and before the training, the hyperparameter of

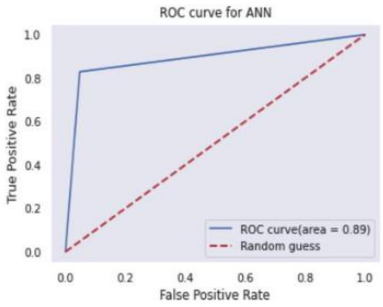
each model was optimized. Hyperparameter tuning was exercised based on different combination s of parameters us-ing a trial-and-error method to obtain the best model with less entropy. Hyperparameters were optimized for the neural network to determine the network's structure as well as learning rate of the network. The sigmoid function was used in the output layer to obtain binary predictors. There were six neurons in the input layer to reduce the overfitting. Adaptive moment estimation (Adam) was used to obtain the model by minimizing the cost function, and in the fitting of the ANN model, the batch size was 32 and the epochs was 30. RF is a meta-estimator that fits several decision tree classifiers various subsamples of the data set using averaging to improve the predicted accuracy and to control the overfitting. Thus, the number of trees in the forest was 10 using a trial-and-error basis. Gini criteria were used to obtain trees, which measured the quality of the optimal split from a root node. The SVM was trained with the radial basis function (RBF) kernel with two parameters (i.e., C and gamma), where the tuning parameter C was chosen as 10, and gamma was defined by how much influence a single training example had was 30. RF is ameta-estimator that fits several decision tree classifiers on various subsamples of the data set using averaging to improve the predicted accuracy and to control the overfitting. Thus, the number of trees in the forest was 10 using a trial-and-error basis. Gini criteria were used to obtain trees, which measured the quality of the optimal split from a root node. The SVM was trained with the radial basis function (RBF) kernel with two parameters and gamma), where the tuning parameter C was chosen as 10, and gamma was defined by how much influence a single training example had shown. The results from the correlation matrix described each model's ability to correctly classify the data. In Table 2, three samples fall in the false positive group and one sample falls in the false negative group for RF classifier model. An interesting result was observed for RF: there were no FP samples in the testing data set. From the matrix, we can determine each model's sensitivity, which evaluates the described model's ability to predict true positives for each available category and the specificity which evaluates the model's ability to predict true negatives for each available category. A summary of model evaluation

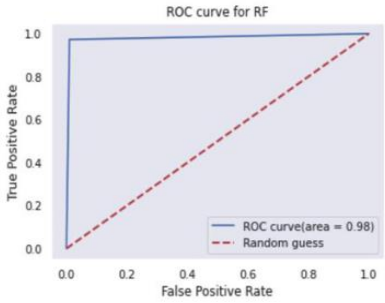
is
given below.

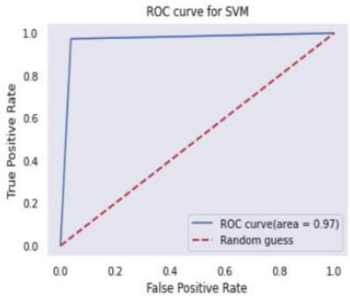
Confusion Matrix	Actual Class			Actual Class			Actual Class		
	ANN	0	1	SVM	0	1	RF	0	1
Predicted Class	0	100	5	0	101	4	0	104	1
	1	19	92	1	3	108	1	3	108

Model	ANN	RF	SVM
Sensitivity	0.9523	0.9904	0.9619
Specificity	0.8288	0.9729	0.9729
Precision	0.9484	0.9908	0.9642
Accuracy	0.8889	0.9814	0.9675
F_1	0.8845	0.9817	0.9685

The AUC–ROC curves [46] of the classification validates the applied techniques for a good accuracy level. Moreover, 0–1 loss function supports the above results, where RF showed the lowest expected loss of 1.86%, which is very low. The diagnostic performance of the implemented tests or the accuracy of the tests for differentiating liver disease patients from normal healthy controlled individuals was evaluated using ROC analysis. This curve was used to compare the diagnostic performance of two diagnostic tests [47]. From all the machine learning methods performed well, because the value of the area under the ROCs were 0.98 for RF and exactly 0.97 for SVM. However, for the ANN, the area under the curve was approximately 0.89. The 95% confidence intervals for the ANN, SVM, and RF were 0.87 and 0.91, 0.96 and 0.98, and 0.96 and 0.99, respectively. The p-values for all of the ML methods were very small ($p < 0.001$). Thus, it was concluded that the area under the ROC was significantly different from the value of 0.5. A significance level of $\alpha = 0.05$ was assumed for rejecting the null hypothesis that both algorithms performed equally well on the data set and conducted the five-fold cross-validated-test.


ROC curve for ANN
True Positive Rate
False Positive Rate
ROC curve(area = 0.89)
Random guess
0.0
0.2
0.4
0.6
0.8
1.0


ROC curve for RF
True Positive Rate
False Positive Rate
ROC curve(area = 0.98)
Random guess
0.0
0.2
0.4
0.6
0.8
1.0


ROC curve for SVM
True Positive Rate
False Positive Rate
ROC curve(area = 0.97)
Random guess
0.0
0.2
0.4
0.6
0.8
1.0