Advanced Artificial Intelligence
COMP 520

Project

Liver Disorder Prediction

Group Members

Madhu Gaurav
Dhruv Mongia

Introduction

Liver disorders are universal problem throughout the world now a days. Liver is essential for digesting food and it helps in detoxification of our body. Liver disorders can be genetic or caused by a number of external factors like alcohol consumption. It is imperative to know about the factors that might play a role in detecting liver disorders. To establish the relation between such factors the liver disorders, Artificial Intelligence can be taken advantage of.

ILPD (Indian Liver Patient Dataset) Data Set has been used for the pyurpose of this project. The results of back-propagation Random forest and Backpropagation on the ILPD data set will be compared. The data set has 10 attributes namely - age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. This data set has 416 liver patient records and 167 non liver patient records.The data set was collected from north east of Andhra Pradesh, India. Selector is a class label used to divide into groups(liver patient or not). This data set contains 441 male patient records and 142 female patient records. Any patient whose age exceeded 89 is listed as being of age "90".

Attributes and Types

Gender	Categorical
Age	Total bilirubin
Direct bilirubin	Real Numbers
Total protiens	Real number
Albumin	Real number
A/G ratio	Real number
SGPT	Integer
SGOT	Integer
Alkphos	Integer

ILPD Dataset

The attributes are obtained by simple blood tests used to measure the levels of enzymes, proteins and bilirubin levels in the blood that helps to predict the liver damage. Proteins are large molecules that are needed tor the overall health. Enzymes help to carry our chemical reactions that occur in the body. Bilirubin helps the body to break down and digest fats. ALT (SGPT), AST (SGOT), ALP and GGT are the enzymes produced by the liver. The ALT, AST, ALP and GGT are the liver enzyme tests that measure the level of ALT, AST, ALP and GGT in the blood respectively. High levels of ALT and AST in the blood can be a symptom of liver damage. High levels of ALP and GGT can be sign of bile duct damage.

Background

Several works related to liver disorder prediction has been done. Two datasets were evaluated using analysis of varianace and multivariate analysis of variance. First dataset is taken from University of California at Irvine and it has 345 records with 6 attributes. Second dataset is taken from Andhra Pradesh, India which has 583 records with 10 attributes. 3 attributes, namely, Alkphos, SGPT and SGOT were common between these two datasets. These common attributes were considered for comparison purpose. In the UCI dataset, 145 patients were labelled as having liver diorder and in the Indian dataset, 416 patients were having liver disorder.

One-way Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) are applied to evaluate the significance between two populations for better classification. Bendi Venkata Ramana compared popular Classification Algorithms for evaluating their classification performance in terms of Accuracy, Precision, Sensitivity and Specificity in classifying liver patients dataset. Accuracy, Precision, Sensitivity and Specificity are better for the Indian Liver Dataset compared to UCLA liver datasets with all the selected algorithms. The reason can be attributed to more number of useful attributes like Total bilirubin, Direct bilirubin, Indirect bilirubin He also proposed Bayesian Classification for diagnosis of liver diseases. The Bayesian Classification is combined with Bagging and Boosting for better accuracy. The following figure presents how accuracy varies with different classifiers on different dataset.

Accuracy										
Diseases → Classifiers ↓	**Breast Cancer Data**	**Chronic Kidney Disease**	**Cryother apy**	**Hepatitis**	**Immunot herapy**	**Indian Liver Patient Dataset**	**Liver Disorders**	**Pima diabetes**	**Risk factors cervical cancer**	**Statlog (Heart) Data Set**
Bagging	95.85	98.75	88.89	64.52	84.44	69.3	69.57	75.78	96.15	80
IBK	95.14	95.75	90	66.45	70	64.49	62.9	70.18	94.41	75.19
J48	94.56	99	93.33	58.06	82.22	68.78	68.7	73.83	95.1	76.67
JRip	96.28	97.75	87.78	63.23	82.22	66.38	66.67	76.04	96.15	80.74
MP	95.85	99.75	87.78	62.58	80	68.95	71.59	75.39	94.76	77.41
NB	95.99	95	83.33	71.61	76.67	55.75	55.36	76.3	88.69	83.7

Fig. Accuracy with different classifiers on different datasets.

Method

Preprocessing

The labels were binary encoded 0 for healthy patients and 1 for patients having liver disorder. The missing values in each column were filled with the mean of each column. Binary encoding for the gender column was also used - 0 for female and 1 for male. All the attributes were normalised column-wise

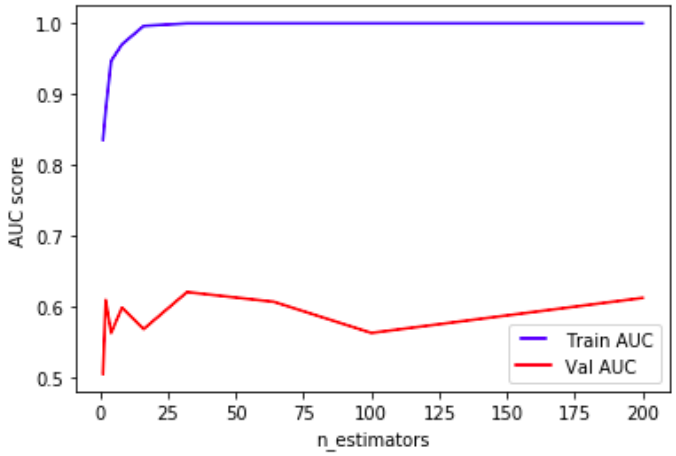
Splitting

For back propagation, 20% of data was taken as test set and 24% of the data was taken as validation set.
For random forest, 20% of data was taken as test set and 10-fold cross validation was used on the training data.

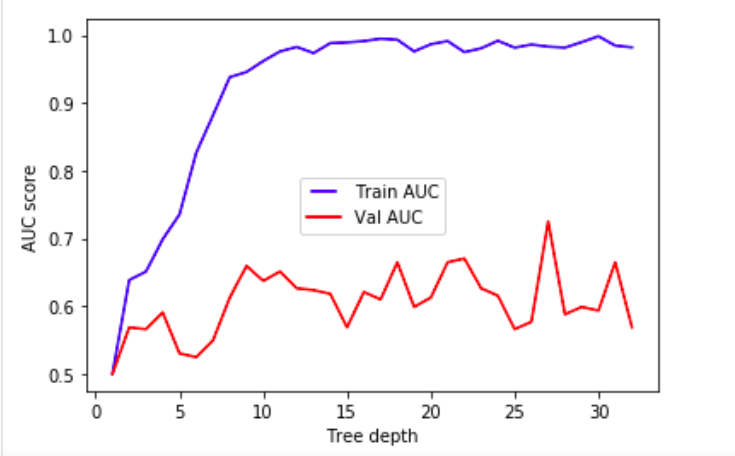
Random Forest

- Scikit-Learn package with Python 3.7 was used to implement Random Forest Algorithm.
- The seed value as a base to generate a random number was 50
- Gini impurity was used to measure the quality of the split
- Maximum square root of available features were used to branch the trees at each step
- The hyper parameters such as number of decision trees and maximum depth were found by enumerating all possible combinations and validating it using 10-fold cross validation.
- It was found that the best number of decision trees to be used was 32 and the maximum depth allowed should be 27

AUC score V/S Number of Decision Trees



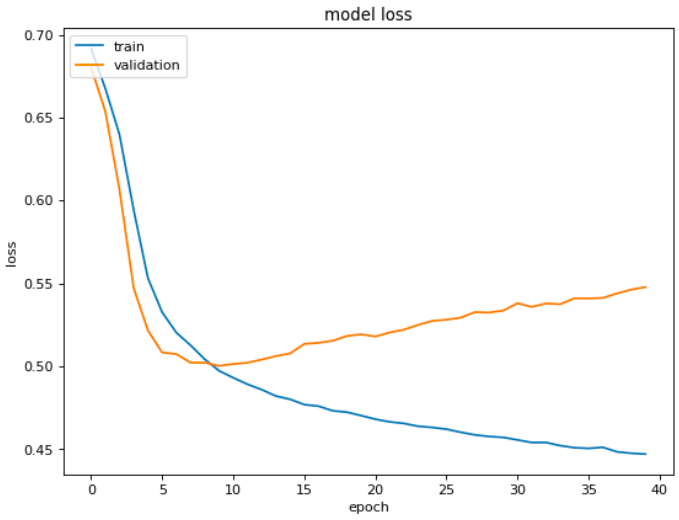
AUC score V/S Tree Depth



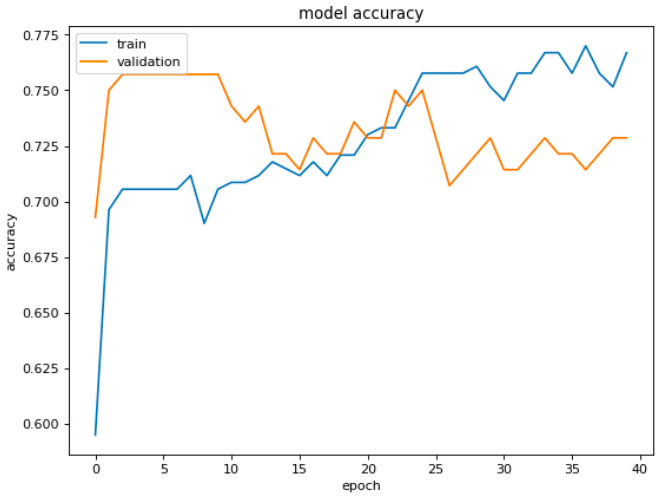
Back Propogation

- **Layers used:** The input layer had a dimension of 10 because of 10 attributes. 2 hidden layers with ReLu activation were used with a random normal kernel initialiser. The hidden layer connected to the input layer had an outpace space of 8th dimension. The second hidden layer had an outpace space of 4th dimension and the final output had one dimension because of the binary classification.
- **Functions used:** The optimiser used was Adam Optimizer and the loss function used was binary cross entropy because of binary classification.
- **Epochs and Batch Sizes:** 40 epochs with a batch size of 2 were used first to plot the graph between the training model accuracy v/s validation model accuracy and training model loss v/s validation model loss to understand where over fitting was taking place.
- It was found out that the final model must be trained up to 9th epoch to prevent over fitting

Model Accuracy v/s Epochs for Training and Validation for Back Propagation



Model Loss v/s Epochs for Training and Validation for Back Propagation



Result

In this study ILPD data set was taken from UCI Machine Learning repository. Selected classifiers Feed Forward Back Propogation and Random Forest are considered for performance evaluation based on accuracy, sensitivity, precision, specificity and ROC Area. The performance evaluation is depicted based on confusion matrix and it shows the ways in which the classification model is confused when it makes predictions. Example of confusion matrix presented in Fig. 1.

Predicted → /*Actual* ↓	**No**	**Yes**
No	TN	FP
Yes	FN	TP

- TP: Predicted yes and they do have the disease.
- TN: Predicted no and they do not have the disease.
- FP: Predicted yes, but they don't have the disease.
- FN: Predicted no, but they actually do have the disease.

Accuracy: The accuracy of a classifier is the percentage of tuples that are correctly classified by the classifier.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

Sensitivity: Sensitivity is the proportion of positive tuples that are correctly identified.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Precision: precision is the proportion of the true positives against all the positive results.

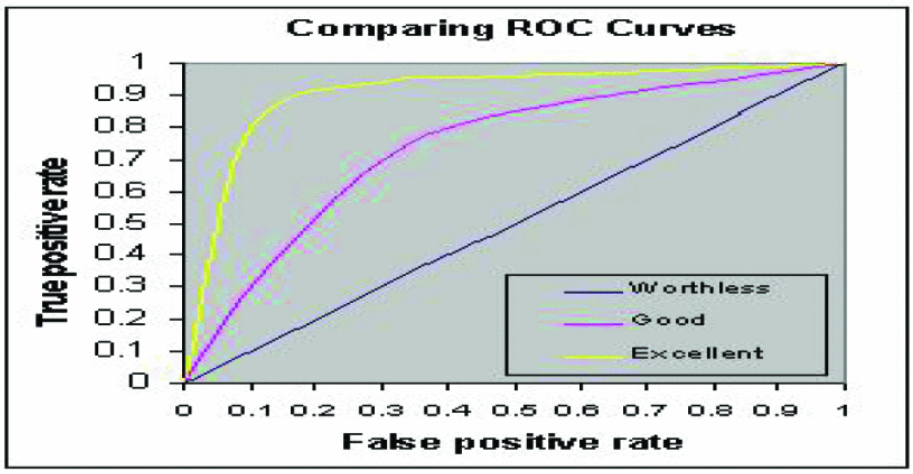
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Specificity: Specificity is the proportion of negative tuples that are correctly identified.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

ROC: The Receiver Operating Characteristic (ROC) curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection. The false positive rate is also known as the fall-out (1-specificity). The Fig. 2 shows three ROC curves representing excellent, good, and worthless tests plotted on the same graph. The accuracy of the test depends on how well the test separates the group being tested into those with and without the disease in question. Accuracy is measured by the area under the

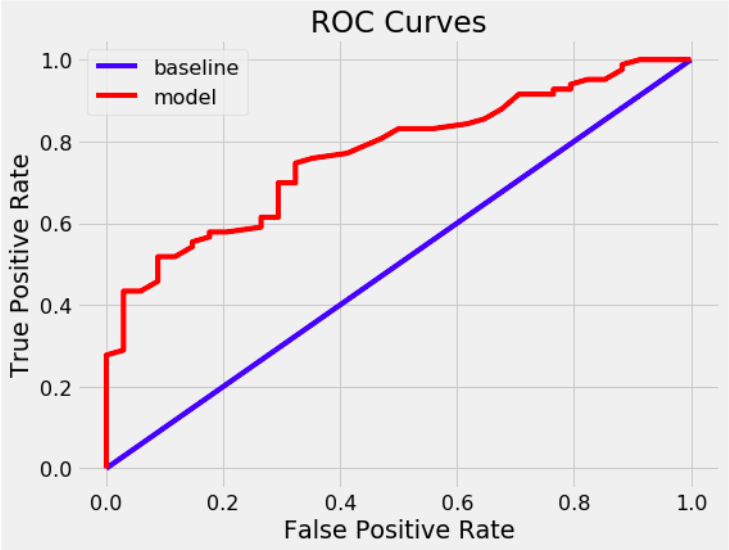
ROC curve. An area of 1 represents a perfect test and an area of .5 represents a worthless test.



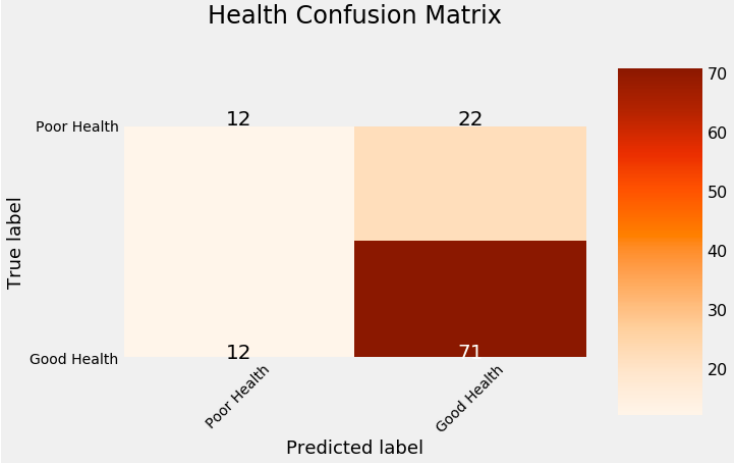
Final Result Comparison

	Random Forest	Back Propagation
Accuracy	70.94%	70.08%
Sensitivity	85.54%	85.0%
Precision	76.34%	74.72%
Specificity	35.29%	37.83%

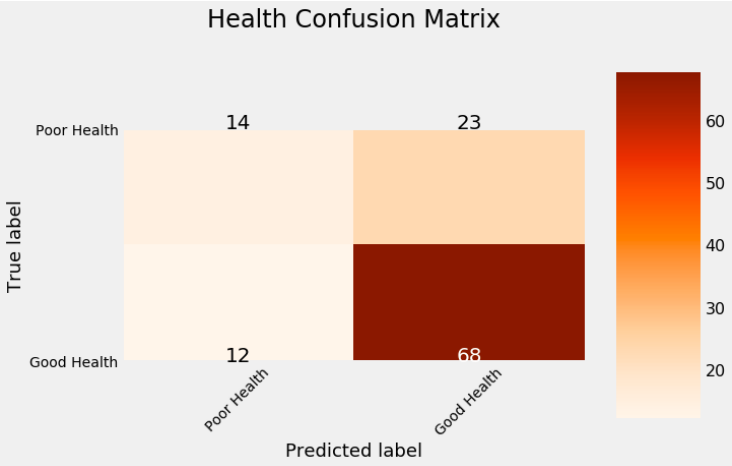
ROC Curve for Random Forest



Confusion Matrix For Random Forest



Confusion Matrix for Back Propagation



Conclusion

The accuracy of Random Forest was found out to be 70.94% whereas the accuracy of Back Propagation was 70.08%.
In medical application, sensitivity usually plays an important role. The sensitivity for Random forest was found out to be 85.84% whereas the sensitivity of Back Propagation was 85.0%.
Therefore, both the classifiers performed equally well. The precision for Random forest was found out to be 76.34% whereas the precision of Back Propagation was 74.72%.
The specificity for Random forest was found out to be 35.29% whereas the specifcity of Back Propagation was 37.83%.
More work can be done to improve the performance of both the classifiers by adjusting more hyper parameters.

References

- Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis, International Journal of Computer Science Issues, ISSN :1694-0784, May 2012.

- Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, âœA Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, ISSN : 0975-5705, PP 101-114, May 2011.