

An Enhanced Grey Wolf Optimization and Random Forest Techniques for Predicting Chronic Hepatitis

Parameswari¹, Samraj Lawrence², Manohar³, Malaiarasan⁴

¹PG Scholar, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, India.

²Associate Professor, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, India.

^{3,4}Assistant Professor, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, India.

Abstract - Hepatitis is one of the dangerous disease compare HIV [2], still the patient's cant able to notice any symptoms of hepatitis disease in early stages, when patients noticed symptoms; the disease could be on advanced stage. Based on the 2017 medical survey fifty lakh people affected hepatitis disease in Chennai. Cost plays a major role in hepatitis treatment, and traditional biopsy also has some issues. The proposed algorithms Grey Wolf Optimization and Random Forest Classification to overcome drawbacks of biopsy, low cost and high accuracy.

Key words: Biopsy, Gwo, Hepatitis, optimization, prediction, Rfc.

1. INTRODUCTION

Now a day's machine learning approaches such as regression, classification and optimization are used in prediction, classification and act as diagnostic tools. Machine learning approaches used in various field especially in medical field, the machine learning techniques to overcome the invasive methods drawbacks [10] and detection of disease severity[1][2][3]. Hepatitis implies aggravation of the liver. Viral hepatitis is a noteworthy general wellbeing the issue around the world. Distinctive infections including hepatitis A, B, C, D and E infections cause viral contaminations of the human liver. Hepatitis B and C are comparable sorts of liver contamination, which spread for the most part through blood and blood items [1]. Liver biopsy conveys potential confinements including inspecting mistakes furthermore, inter-observer variation [2]. Biopsy is one of the gold standards [4] but biopsy is one of the invasive methods and it also have sampling error and complicated calculation, to overcome these disadvantages, propose machine learning techniques. Machine learning techniques are one of the Non-Invasive methods. Based on the METAVIR scoring system the biopsy report evaluated but machine learning is in-direct method, in this paper following features are used such as albumin, tot-protein, bilirubin, A/g_Ratio, Sgpt, Sgot, Alkphos. The following figure 1 explains the comparison of the both invasive and non-invasive methods for finding liver fibrosis. Biopsy is invasive and direct method based on METAVIR scoring system. Machine learning approaches based on blood markers and indirect method.

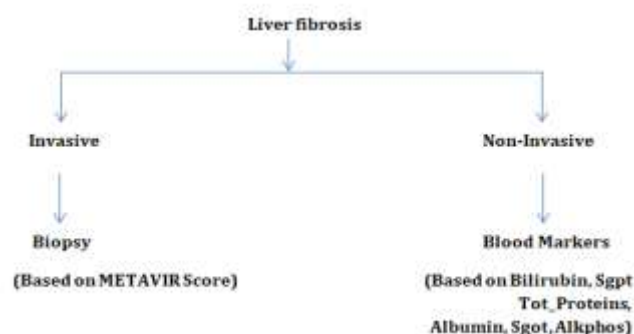


Figure 1. Comparison chart for liver fibrosis methods.

2. RELATED WORK

A. PRE-PROCESSING

In these study patients records are collected from machine learning repository. Database contains 1230 male patients and 540 female patients. The lab tests also conducted at a same time of liver biopsy. The datasets are randomly divided into testing and training sets. In this time 80 % of data are trained and 20% of data should be tested based on the following

table. Table 1 explain the liver function test. Based on the table 1 the patient’s data are divided into three categories that is normal, mild-moderate and Advanced stage of liver fibrosis.

Table 1: Liver Function Test

Liver Function Test			
	Normal	(Mild – Moderate)	Advanced
Bilirubin	0.3 – 1.0	1.0 - 2.5	> 2.5
Tot_Proteins	6 - 8.3	8.3 - 12	> 12
Albumin	3.5 - 4.5	3.0 - 3.5	< 3.0
A/g_Ratio	0.8 - 2.0	< 0.8	> 2.0
Sgpt	7- 56	56 – 200	> 200
Sgot	5 – 40	40 – 200	> 200
Alkphos	37 - 116	< 37	> 116

B. STATISTICAL FEATURE EXTRACTION

In statistical feature extraction data were analysed using Mat-Lab software and Excel while software performed the optimization and classification learning. That data should reported as mean and standard deviation values. The relationship between the two larger variables and the presence of significance of fibrosis has been analysed by using the KRUSKAL-WALLIS test and one way ANOVA table [2].

We implemented several optimization and classification techniques such as GWO, ANN, regression, Svm, random forest. The ROCs accuracy and predictive values were evaluate the performance of both training and test data sets.

C. GREY WOLF OPTIMIZATION

The GWO calculation imitates the authority chain of command and lead instrument of Wolf’s package. In these package four sets of wolves utilized for reproducing the initiative chain of importance, such as Alpha (α), Beta (β), Omega (ω) and Delta (δ). The GWO has been successfully applied for solving various engineering optimization problems [11].

Pseudo Code For GWO Algorithm Is As Follows

- 1: Generate initial search agents G_i ($i=1, 2, \dots, n$)
- 2: Initialize the vector’s a, A and C
- 3: Estimate the fitness value of each hunt agent
 G_α =the best hunt agent
 G_β =the second best hunt agent G_δ =the third best hunt agent
- 4: Iter=1
- 5: repeat
- 6: for $i=1: G_s$ (grey wolf pack size)
 Renew the location of the current hunt agent.
- End for
- 7: Estimate the fitness value of all hunt agents
- 8: Update the value of $G_\alpha, G_\beta, G_\delta$
- 9: Update the vectors a, A and C
- 10: Iter=Iter+1
- 11: until Iter>= maximum number of iterations {Stopping criteria}
- 12: output G_α
- End

D. RANDOM FOREST CLASSIFIER

Random Forest is an adaptable, simple to utilize machine learning calculation that produces, even without hyper-parameter tuning, an incredible outcome more often than not. It is additionally a standout amongst the most utilized calculations since its straightforwardness and the way that it tends to be utilized for both grouping and relapse assignments. Going to realize, how the irregular backwoods calculation works and a few other essential things about it.

Irregular Forest is a managed learning calculation. Would already be able to see from its name, it makes a backwoods and makes it by one way or another arbitrary. The "woodland" it manufactures, is a group of Decision Trees, more often than not prepared with the "stowing" technique. The general thought of the packing technique is that a mix of learning models builds the general outcome.

To state it in basic words: Random woodland constructs various choice trees and combines them to get a progressively exact and stable forecast.

One major favourable position of random forest is that it tends to be utilized for both order and relapse issues, which frame the lion's share of current machine learning frameworks. Talk about irregular woods in characterization, since order is in some cases considered the building square of machine learning. Underneath can perceive how a random forest would look like with two trees: The figure 4 shows the simple random forest classification graphical representation.

3. RESULTS AND DISCUSSION

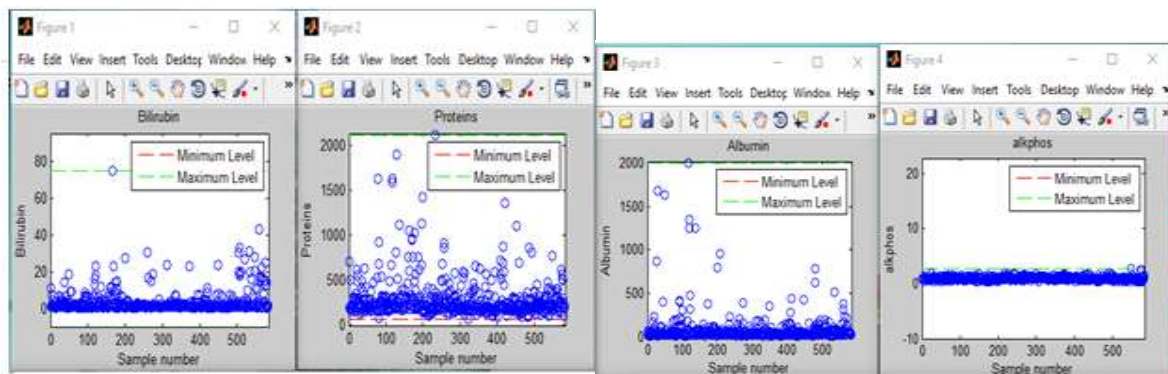


Figure 2: Pre-Processing

In this figure 2 explain the pre-processing of the features such as bilirubin, proteins, albumin, alkphos and display the minimum and maximum value of the features based on the liver function test table 1. Red colour and green colour dotted lines indicates the maximum and minimum level.

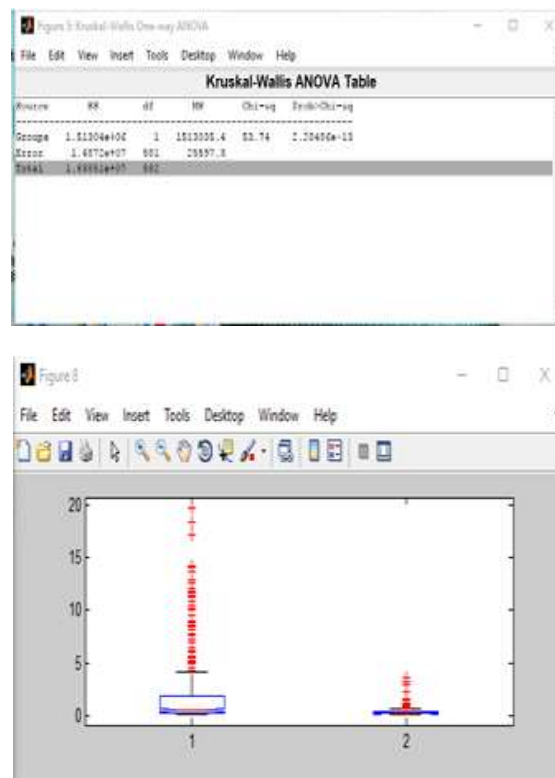


Figure 3: Statistical Feature Extraction

Figure 3 explain the statistical analysis using kruskal-Wallis test and one way ANOVA table. These calculations done by chi square distribution test. The kruskal-wallis test used for finding the difference between two large data sets.

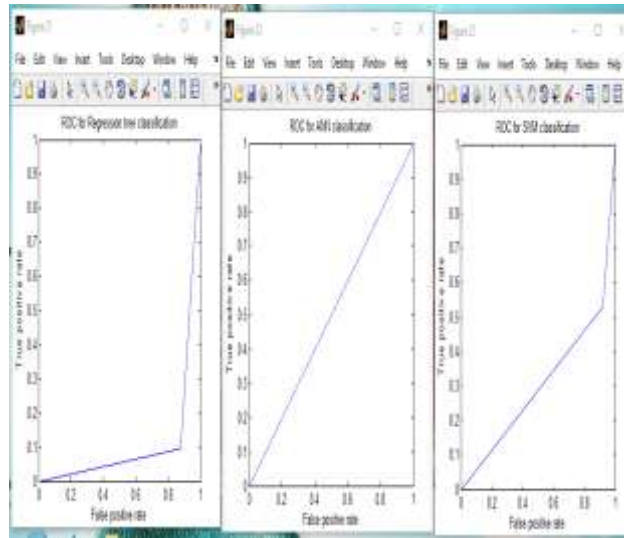


Figure 4: Roc curves for CART, ANN, SVM.

Figure 4 explain the Roc curves the true positive rate is plotted in function of the false positive rate for different cut off points. Compare three classifier CART techniques gives the high accuracy than SVM and ANN.

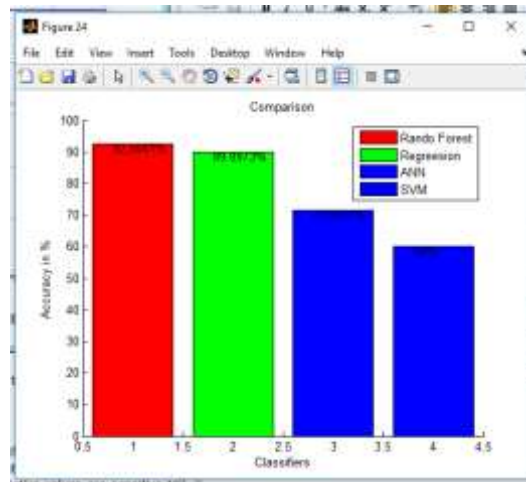


Figure 5 : Comparison Accuracy Between Random Forest, ANN, SVM And Regression.

In figure 5 represents the comparison accuracy between the Random forest, regression, ANN and SVM classifier. Following table 2 explain the accuracy rate, error rate and AUC. The performance of the machine learning algorithms were analysed with the help of confusion matrix parameters where the accuracy is considered as a valid parameter to measure the performance.

Confusion matrix	Accuracy rate	Error rate	AUC
Regression tree	89.8973	10.1027	0.1066
Support vector	60	40	0.3032
Neural Network	71.5517	28.4483	0.5000
Random Forest	92.5041	7.4959	0.0301

4. CONCLUSION

In this paper, combining various machine learning techniques on the expectation of advanced liver fibrosis in Chronic Hepatitis patients. Grey Wolf optimization and random forest classification models were produced. The following

parameters (bilirubin, alkphos, albumin, SGPT,SGOT,) observed to be the most imperative highlights in the forecast of the propelled fibrosis as they factually have a noteworthy relationship (P -esteem < 0.0001) and acknowledged connection coefficients ($r > 0.1$) with the nearness of advanced liver fibrosis as appeared in the outcomes. The proposed models could be utilized as an adequate, safe, and ease substituting for foreseeing propelled fibrosis as opposed to generally unsafe elective instruments, (for example, the liver biopsy) in unending hepatitis-infected patients.

5. REFERENCE

- 1) Alodini .Q,(May 2012) "Prevalence of Hepatitis B Virus (HBV) and Hepatitis C Virus (HCV) infections among blood donors at Al-Thawra Hospital Sana'a City-Yemen," Yemeni J. Med. Sci., vol. 6, pp. 16–20,.
- 2) Bedossa .B and Carrat .F, (2009) "Liver biopsy: The best, not the gold standard," J. Hepatology, vol. 50, pp. 1–3,.
- 3) Bonacini .M, Hadi .G, Govindarajan .S, and Lindsay .K.L, (1997) "Utility of a discriminant score for diagnosing advanced fibrosis or cirrhosis in patients with chronic hepatitis C virus infection," Am. J. Gastroenterology, vol. 92, pp. 1302–1304,.
- 4) Castera .L (2012) "Noninvasive methods to assess liver disease in patients with hepatitis B or C," Gastroenterology, vol. 142, pp. 1293–1302,.
- 5) Crisan .D, et al.(2012), "Prospective non-invasive follow-up of liver fibrosis in patients with chronic hepatitis C," J. Gastrointest Liver Dis., vol. 21, pp. 375–382,.
- 6) ElHefnawi .M, et al.(2012), "Accurate prediction of response to Interferon-based therapy in Egyptian patients with Chronic Hepatitis C using machine-learning approaches," in Proc. IEEE/ACM Int. Conf. Advances Social Netw. Anal. Mining, pp. 771–778.
- 7) Gravitz .L,(2011) "A smouldering public-health crisis," Nature, vol. 474, no. 7350, pp. S2–S4,.
- 8) Hashem et al .S.(2014), "A Simple multi-linear regression model for predicting fibrosis scores in chronic Egyptian hepatitis C virus patients," Int. J. Bio-Technol. Res., vol. 4, no. 3, pp. 37–46, Jun..
- 9) Hashem .S, et al., (2016) "Accurate prediction of advanced liver fibrosis using the decision tree learning algorithm in chronic Hepatitis C Egyptian patients," Gastroenterology Res. Practice, vol. 2016, Art. no. 2636390.63
- 10) Kim .M.Y, Jeong .W.K.Y, and Baik .S.K,(2014) "Invasive and noninvasive diagnosis of cirrhosis and portal hypertension," World J. Gastroenterology, vol. 20, no. 15, pp. 4300–4315, Apr. 21,.
- 11) Nashaat .E,(2010) "Lipid profile among chronic hepatitis C Egyptian patients and its levels pre and post treatment," Nature Sci., vol. 8, no. 7, pp. 83–89.
- 12) Parkes .J, Guha .I.N, Roderick .P, and Rosenberg W.,(2006)"Performance of serum marker panels for liver fibrosis in chronic hepatitis C," J. Hepatol, vol. 44, pp. 462–474,.
- 13) Regev .A, et al., (2002)"Sampling error and intra-observer variation in liver biopsy in patients with chronic HCV infection," Am. J. Gastroenterology, vol. 97, pp. 2614–2618,.
- 14) Rosen .H,(2011) "Clinical practice, Chronic Hepatitis C Infection," New England J. Med., vol. 364, no. 25, pp. 2429–2438,.
- 15) Sterling .R.K, et al., (2006) "Development of a simple non-invasive index to predict significant fibrosis in patients with HIV/ HCV co-infection," Hepatology, vol. 43, pp. 1317–1325,.