

# Introducing Confusion Matrix and Accuracy in Disease Prediction on Liver Using Machine Learning

Abdullah Ali Salman <sup>[1]</sup>, Dr. M. Sampath Kumar <sup>[2]</sup>

<sup>[1],[2]</sup> Department of Information Technology & Computer Applications Andra University College of Engineering Andhra Pradesh-India

## ABSTRACT

The timely prophecy of liver malady is dreadfully essential to accumulate human life and capture appropriate ladder to restrain the malady. Logistic regression algorithms have been alive lucratively used in diverse subject areas, particularly in medical scientific punishment. This traces a communication channel of investigation, study explores the early prophecy of liver infection spending several logistic degeneration techniques. The liver sickness data sets which is selected in place of this scholarship is involving of aspects like partial bilirubin, direct bilirubin, age, gender, total proteins, albumin and globulin ratio. The leading single-mindedness of this work is to estimate the bearing out of various algorithms like Decision tree, SVM, logistic regression technique and put side by side their routine. The laboratory analysis proves that logistic degeneration is giving more accuracy than other techniques.

**Keywords:** Data analytics, machine learning, Svm , Logistic regression, Decision tree.

## I. INTRODUCTION

The liver acts as an significant role in countless physical occupations commencing protein fabrication and blood clotting to fatty acid, glucose (sugar), and iron metabolism. It presents a range of functions, excluding realization rid of toxin from the body, and is crucial to death. The loss of those characters can do generous harm to the target social system. When the liver is disease-ridden with a virus, damaged by elements, or less than attack from own immune method, the simple danger is the same that liver will come to be so worn that it make the sack no longer use to save a person alive. A liver sickness caused by hepatotropic viruses imposes a significant problem on health care wherewithal. Continual infections from the appetites B virus (HBV), appetites C virus, and appetites delta virus result in chronic liver disease. The most basic taxonomy of liver disease is equally keen and staying.

The edge of acute liver disease is based on the interval, with the story of the disease does not stretch over half dozen months. Delicate viral appetites and drug reaction account for the minority of cases of acute liver disease. Liver disease is likewise brought up to as hepatic disease. As a rule queasiness, squeamishness, right upper quadrant ventral pain, fatigue and weakness are classic symptoms of liver disease. Symptoms of liver patient include jaundice, abdominal pain, tiredness, nausea, vomiting, back pain, abdominal swelling, weight loss, fluid in abnormal cavity,

general itching, pale stool, enlarged spleen and gall bladder [1]. Symptoms of liver disease can conform, but they often include growth of the tummy and legs, streak easily, changes in the color of your stool and urine, and jaundice, or bruising of the skin and eyes. Sometimes, at that place are no detectable symptoms. Tests such as imaging tests and liver occupation quizzes can check for liver impairment and help to diagnose liver diseases.

## II. SYSTEM ARCHITECTURE

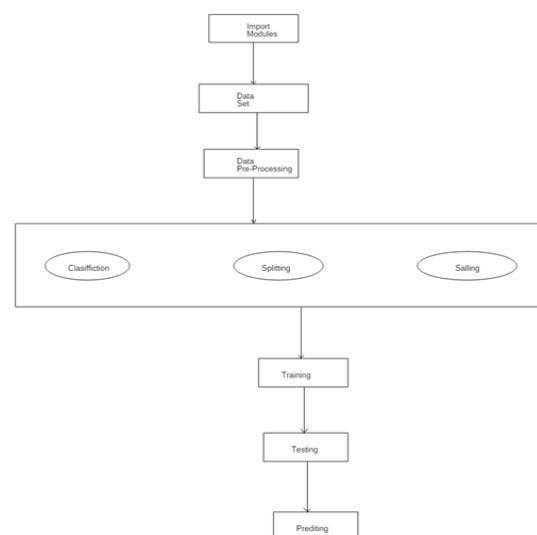


Fig1: System Architecture

### III. PYTHON FRAMEWORK

Intro to Django This book is about Django, a Web improvement framework that saves your time and makes Web growth a joy. Using Django, you can dimensions and keep high- quality Web presentations with minimal fuss. At its Grade A, Web development is an sensational, resourceful act; at its worst, it can be a repetitive, provoking annoyance. Django allows you to effort on the fun material the crux of your Web presentation — while relieving the pain of the repetitive bits. In doing so, it withdraws advanced generalizations of common Web development patterns, timesaver for repeated programming tasks, and clear settlements for how to work out problems. At the same time, Django tries to stay out of your way, authorizing you move outside the scope of the framework as needed. The goal of this book is to make you a Django expert. The focus of this report is two fold. Foremost, we explain, in depth, what Django does and how to ramp up Web applications with it. Second, we confer higher-level brain wave where appropriate, resolving the question “How can I apply these tools effectively in my own projects?” By reading this book, you’ll find out the skills needed to develop powerful Web sites quickly, with code that is clean and comfortable to continue improving.

#### What Is a Web Framework?

Django is an outstanding supporter of a new age group of Web frameworks. And so what exactly fixes that term mean? To answer that enquiry, let’s look at the role of a Web application written using the Common Gateway Interface (CGI) regular, a standard way to write Web presentations around 1998. In those days, when you wrote a CGI application, you did everything yourself the the same of freezing a cake from scrape. For example, here’s a simple CGI script, written in Python, that displays the ten most recently brought out records from a database:

```
import MySQLdb

print "Content-Type: text/html"
print
print "<html><head><title>Books</title></head>"
print "<body>"
print "<h1>Books</h1>"
print "<ul>"

connection = MySQLdb.connect(user='me', passwd='letmein', db='my_db')
cursor = connection.cursor()
cursor.execute("SELECT name FROM books ORDER BY pub_date DESC LIMIT 10")
for row in cursor.fetchall():
    print "<li>%s</li>" % row[0]

print "</ul>"
print "</body></html>"

connection.close()
```

This code should be comparatively straight. First, it publishes a “Content-Type” line, accompanied by a

downright line, as required by CGI. It prints just about basic HTML, associates to a database and finishes a query that repossesses the latest ten books. Looping over those records, it brings an HTML unordered list. In the final stage, it prints the closing HTML and closes the database connection.

With a one-off dynamic page such as this one, the write-it-from-scratch approach isn’t necessarily bad. For ace thing, this code is simple to comprehend even a novice developer can read these 16 lines of Python and understand all it does, from commencement to finish. There’s nothing else to memorize; no other code to translate. It’s also simple to install: just save this program in a file called latestbooks.cgi, upload that file to a Web server, and visit that page with a browser. But as a Web relevance matures beyond the minor, this approach opportunities down, and you face a number of problems:

Should a developer really have to worry about striking the “Content-Type” line and screaming back to shut the database connection? This variety of received idea reduces programmer as and gives opportunities for mistakes. These setups and teardown related tasks would best be handled by some common infrastructure.

- What takes place when this code is reused in multiple environments, each with a separate database and password? At this stage, any environment-specific configuration becomes essential.
- What occurs when a Web exclusive who has no understanding coding Python wishes to remodel the page? Ideally, the logic of the page — the salvage of records from the database — would be attached from the HTML presentation of the page, so that a upscale could edit the latter without affecting the anterior.
- These problems are exactly what a Web framework intends to work. A Web framework provides a programming infrastructure for your applications, so that you can concentrate on writing clear, maintainable code without having to reinvent the bicycle. In a nutshell, that’s what Django does.

### IV. ALGORITHMS

#### Support vector machines:

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which ever can be applied for both categorization and regression challenges. Yet,

it is generally recycled in sorting harms. In this algorithm, we plan each one data point as a point in n-geometric planetary (where n is the number of structures you have) with the worth of each feature animation the value of a exacting bring together. And so, we perform ordering by discovery the hyper-plane that discriminate the two sessions very easily (look at the below snap). The SVM algorithm is put through in practice using a substance. The erudition of the hyper plane in linear SVM is done by translating the problem using some linear algebra, which is out of the reach of this entry to the SVM. A great insight is that the right SVM can be restated using the interior manufactured article of any two set observations, instead than the observations themselves. The inner invention between two vectors is the essence of the multiplication of each span of input values. For instance, the internal produce of the directions [2, 3] and [5, 6] is 2\*5 + 3\*6 or 28. The balance for making a extrapolation for a new effort using the spot product between the response (x) and each support vector (xi) is planned as follows:

$$f(x) = B0 + \text{sum}(ai * (x,xi))$$

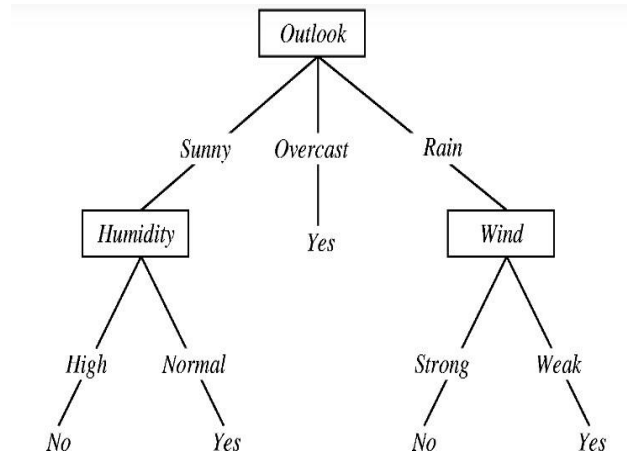
**Logistic Regression:**

A popular statistical technique to predict binomial outcomes (y = 0 or 1) is Logistic Regression. Logistic regression predicts categorical outcomes (binomial / multinomial values of y). The predictions of Logistic Regression (henceforth, LogR in this article) are in the form of probabilities of an event happening, i.e. the probability of y=1, set sure values of input variables x. Therefore, the results of longer range between 0-1. LogR models the data points using the standard logistic function, which is an S- shaped curve also called as sigmoid curve and is given by the equation:

$$\frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

**Decision Tree:**

A decision tree is a flow diagram like building in which each internal node signifies a test for a characteristic (e.g. Whether a coin flip comes up heads or tails), each leaf node symbolizes a category label (decision taken after computing all features) and outlets characterize conjunctions of organizations that contribute to those class labels. The paths from source to leaf represent classification rules. Below diagram illustrate the basic stream of decision trees for decision making with labels (Rain (Yes), No Rain (No)).



**V. Screen Shorts**

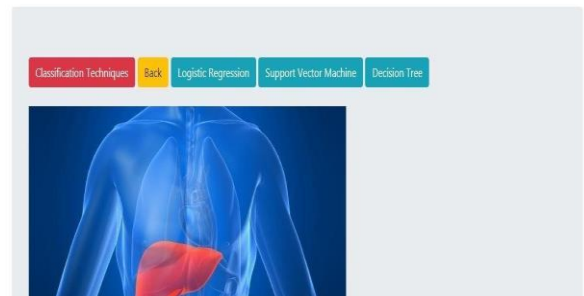


Fig2: Home Page

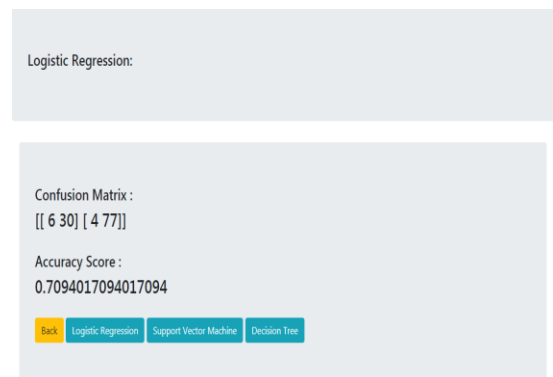


Fig3: Logistic Regression

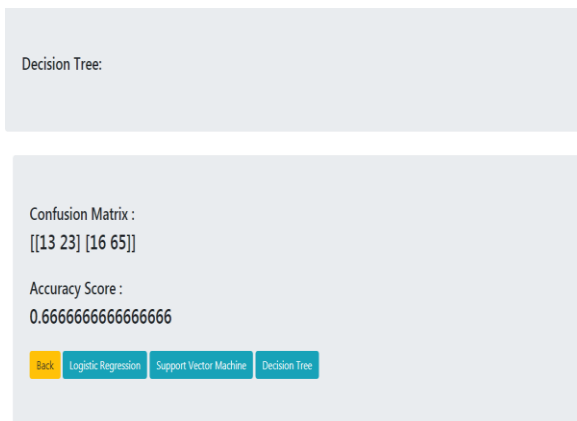


Fig4: Support Vector Machine

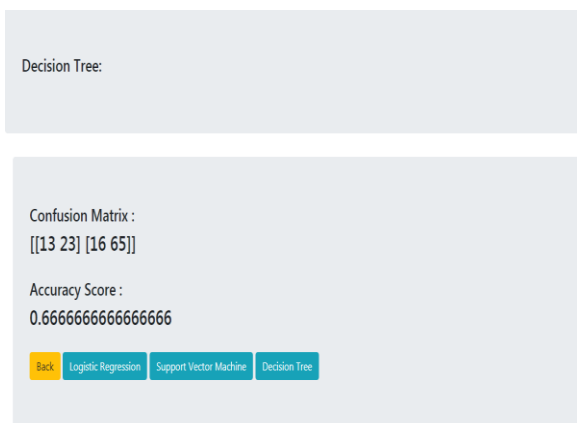


Fig5: Decision Tree

age	gender	total_bilirubin	direct_bilirubin	alkaline_phosphatase	alanine_aminotransferase	aspartate_aminotransferase	total_protein	albumin	albumin_glb	
60	F	1.1	0.1	101	191	101	8.5	3.8	0.4	
62	M	1.0	0.5	699	44	180	7.5	3.2	0.74	
62	M	1.3	0.4	490	68	68	7.3	3.0	0.9	
59	M	1.0	0.4	102	18	29	8.0	3.2	0.4	
70	M	1.0	0.2	195	37	59	7.3	3.2	0.8	
46	M	1.1	0.8	7	280	19	14	7.8	4.1	0.3
26	F	0.8	0.0	2	154	16	12	7.3	5.1	1
29	F	0.8	0.0	2	102	16	11	6.7	5.1	1.1
17	M	0.9	0.3	282	22	19	7.4	4.1	1.2	
59	M	0.7	0.2	290	59	58	6.8	3.4	1.1	
57	M	0.6	0.1	239	55	59	6.2	3.7	0.8	
72	M	2.7	1.3	269	71	56	7.4	3.9	0.6	
64	M	0.9	0.3	318	61	58	7.3	4.0	0.9	
79	F	0.8	0.1	214	22	39	6.8	3.4	1.1	
61	M	0.7	0.2	145	51	41	6.2	3.7	0.8	
29	M	0.6	0.1	183	91	53	5.2	3.0	0.7	
69	M	1.0	0.0	343	168	44	7.6	4.1	0.3	

Fig6: Data Sets

## VI. FUTURE ENHANCEMENT

Some more algorithm we can use and more parameter also can find in liver disease prediction or this method can be use in other disease accuracy finding. The future methodology is used to analyze the liver region into separable compartments i.e. liver etc. However, the method requires further improvement mostly regarding feature selection of the liver into multiple components: renal cortex, renal column, renal medulla and renal

pelvis. Apart from that, it is planned to expand the database on which the system will be tested. And also the proposed method in this thesis can be employed for detecting the heart diseases in future with the heart dataset and classification of the diseases.

## VII. CONCLUSION

This research work used classification algorithms namely Logistic Regression and Support Vector Machine (SVM) for liver disease prediction. Comparisons of these algorithms are done and it is based on the performance factors classification accuracy and execution time. From the experimental results, this work concludes, the Logistic regression classifier is considered as a best algorithm because of its highest classification accuracy. On the other hand, while comparing the execution time, the logistic regression classifier needs minimum execution time.

## REFERENCES

1. D. Sindhuja and R. J. Priyadarsini, “A survey on classification techniques in data mining for analyzing liver disease disorder”, International Journal of Computer Science and Mobile Computing, Vol.5, no.5 (2016), pp. 483-488.
2. B. V. Ramana, M. R. P. Babu and N.B. Venkaeswarlu, “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis”, International Journal of Database Management Systems (IJDBMS), Vol.3, no.2, (2011), pp. 101-114.
3. A.S.Aneeshkumar and C.J. Venkateswaran, “Estimating the Surveillance of Liver Disorder using Classification Algorithms”, International Journal of Computer Applications (0975 – 8887), Vol. 57, no. 6, (2012), pp. 39-42.
4. S.Dhamodharan, “Liver Disease Prediction Using Bayesian Classification”, 4th National Conference on Advanced Computing, Applications & Technologies, Special Issue, May 2014. International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.8, No.2, March 2018 9
5. P.Rajeswari and G.S. Reena, “Analysis of Liver Disorder Using data mining Algorithms”, Global Journal of Computer Science and Technology, Vol.10, no. 14 (2010), PP. 48- 52.

6. G. Selvara and S. Janakiraman, “A Study of Textural Analysis Methods for the Diagnosis of Liver Disease from Abdominal Computed Tomography”, *International Journal of Computer Applications* (0975-8887), Vol. 74, no.11 (2013), PP.7-13.
7. H. Sug, “Improving the Prediction Accuracy of Liver Disorder Disease with Oversampling”, *Applied Mathematics in Electrical and Computer Engineering, American-MATH 12/CEA12 proceedings of the 6th Applications and proceedings on the 2012 American Conference on Applied Mathematics* (2012), PP. 331-335.
8. R.H.Lin, “An Intelligent model for liver disease diagnosis”, *Artificial Intelligence in Medical*, Vol. 47, no. 1 (2009), PP. 53-62.
9. B. V. Ramanaland and M.S. P. Babu, “Liver Classification Using Modified Rotation Forest”, *International Journal of Engineering Research and Development* ISSN: 2278-067X, Vol. 1, no. 6 (2012), PP.17-24.
10. H.R. Kiruba and G. T. arasu, “An Intelligent Agent based Framework for Liver Disorder Diagnosis Using Artificial Intelligence Techniques”, *Journal of Theoretical and Applied Information Technology*, Vol. 69 , no.1 (2014), pp. 91-100.



ABDULLAH ALI SALMAN  
Masters of Science in Computer  
Science Department of Information  
Technology & Computer  
Applications Andra University  
College of Engineering  
(AUTONOMOUS  
VISAKHAPATNAM - 530003) ,

AP, INDIA.



Dr. M. SAMPATH KUMAR  
M.E (Computer Engg), Ph.D.  
Professor Professor of Department  
of Information Technology &  
Computer Applications Andra  
University College of Engineering,  
Visakhapatnam, AP, INDIA