INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 7, Issue 6 - V7I6-1394)
Available online at: https://www.ijariit.com

Jaundice prediction using Machine Learning approach

Sussma S.
sussma.s@gmail.com
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu

Srivignesh S.
sri28vignesh@gmail.com
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu

Kishore V. S.
kichu9100@gmail.com
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu

Dr. M. Marimuthu
mmarimuthu@cit.edu.in
Coimbatore Institute of Technology, Coimbatore, Tamil Nadu

ABSTRACT

Jaundice occurs when rise in level of bilirubin causes the skin, mucus membrane and white part of eyes to appear yellowish. bilirubin is a reddish yellow substance produced when red blood cells break down.it is excreted through liver.the bilirubin level will rise up when abnormally high level of red blood cells breaking down. Any person with liver disease develops jaundice (i.e) when liver does metabolize bilirubin the way it's supposed to do, jaundice is developed. Depending on the underlying cause of jaundice, treatment will be provided. If it is caused by viral hepatitis, it will recover on its own. If the cause is because of other infections, diagnosing will be the appropriate treatment. Objective of this work is to develop a most efficient model for any medical lab to predict jaundice. Any data containing relevant factors to jaundice can be used in this model. Standard dataset is collected containing the components age, gender, total bilirubin, direct bilirubin, total protein level, albumin, sgpt, sgot and etc. Principal component analysis and factor analysis is performed to identify the useful and important factors which helps to determine jaundice. Supervised learning models such as random forest, decision tree, support vector machines,* naive bayes classifier *and other models are used to train the dataset to predict the jaundice with better accuracy.

Keywords: *Bilirubin, Jaundice, Prediction, Supervised Learning, Predictive Analytics.*

1. INTRODUCTION

Jaundice is defined as a yellowing of skin and sclera due to the deposition of yellow-orange bile pigment i.e. bilirubin. The bilirubin is an endogenously synthesized pigment that can be toxic especially in new-born children. The model which is done here is using machine learning algorithms and statistical tools. It explains whether the values of bilirubin are in normal, low or high and confirms whether jaundice is affected or not. The main aim of the work is to predict the jaundice in the person and to predict the chances of occurrence of jaundice in his/her body.

2. LITERATURE REVIEW

Non-Invasive Bilirubin Detection Technique for Jaundice Prediction Using Smartphones was proposed by the authors [1].Jaundice prediction is done using the bilirubin level and the colour of the skin. Colour of the skin can be found by using Smartphone application and if it is yellow that indicates the occurance of jaundice.

Advanced Bilirubin Measurementby a Photometric Method implemented by [2].The bilirubin level in the human body can be found by transmitting a blue monochromatic light of particular frequency.

Diagnosis of Neonatal Jaundice Using Artificial Neural Networks was proposed by Smriti Shrivastava. Neonatal jaundice is a yellowish discolouration in the eyes were the white part of eyes becomes yellowish which is mostly occurs to newborn babies and it is diagnosis using artificial neural networks.

Optical Transcutaneous Bilirubin was developed by James W.Kronberg. A transcutaneous bilirubin is a detector comprising a source of light having spectral components absorbable and not absorbable by bilirubin.

Assay For Bilirubin was implemented by Tai-Wing Wu and Rochester. Increase in the blood level can be detect using bilirubin test. Conjugated bilirubin is formed in the liver when sugars are conjugated in bilirubin.

Noninvasive Detection of Bilirubin Using Pulsatile Absorption developed by Mark McEwen, Karen Reynolds .The absorption spectra of Haemoglobin and bilirubin are measured for different ranges.

Method for Determining Bilirubin Concentration from Skin Reflectance was implemented by[8].The level of bilirubin concentration in the human body is calculated by the skin reflectance.

Transcutaneous Bilirubinometer in Assessment of Neonatal Jaundice in Northern India was proposed by Gajan Maharajan.Correlated transcutaneous bilirubin index and serum bilirubin levels without the help of photohteraphy in humans.

Method And Apparatus For Optical Measurement Of Bilirubin In Tissue developed by Steven L. Et.al.,The Optical Measurement of Bilirubin in tissue can be calculated using the apparatus.

3. PROPOSED

The **logistic model** is a widely used statistical model that, in its basic form, uses a logistic function to model a binarydependent variable; many more complex extensions exist. In regression analysis, **logistic regression** is estimatingthe parameters of a logistic model; it is a form of binomial regression.

A **Neural network** is enabled with the power of taking a decision by feeding it information about some parameters called independent variables. These parameters are the information bits which are processed inside a neural network and based on the outcome, a decision is taken.Multilayer perceptron model is used here with two hidden layers and the probability of getting affected by jaundice is predicted.

A **Decision Tree** is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.**Random Forest**operates by constructing multitude of decision trees at training time and produces great results with mean score of all the possible constructed trees. These are the supervised learning models build using various factors which are responsible for the occurrence of jaundice which helps to predict the chances of a person getting affected.

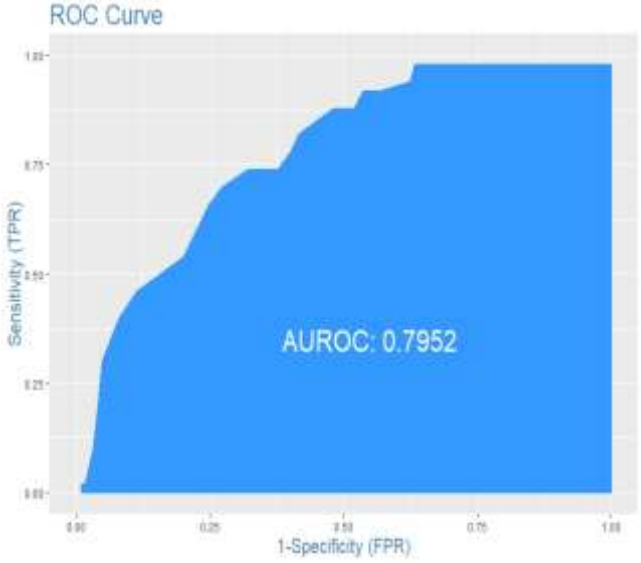
Principal component analysis (**PCA**) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called **components.**
Factor analysis is a statistical method used to describevariabilityamong observed, correlated variablesin terms of a potentially lower number of unobserved variables called **factors.**

4. RESULTS AND DISCUSSION

The data consists of factors such as:
1)Gender
2)Age
3)Total Bilirubin
4)Direct Bilirubin
5)Total Proteins
6)Albumin
7)Ag_Ratio
8)SGPT
9)SGOT

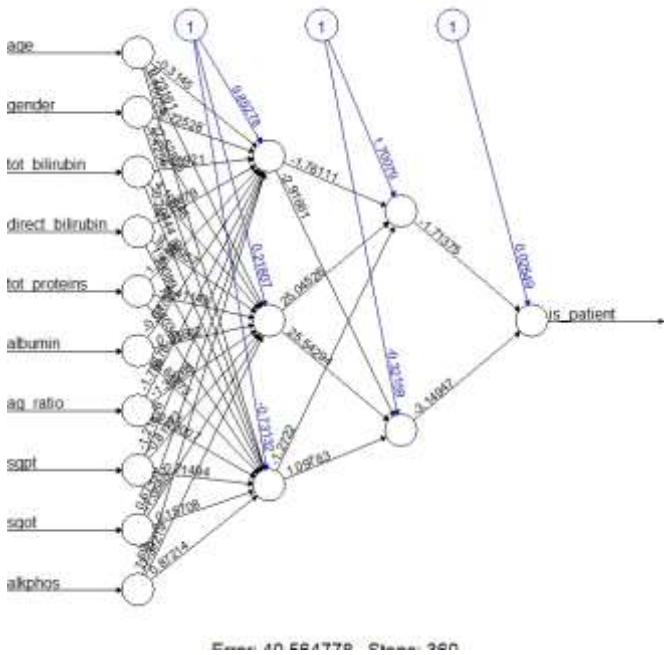
ROC

ROC reflects the diagnostic ability of a binary classifier system as its discrimination threshold is varied.



The area ROC curve is the measure of usefulness of a model in general, where a greater means a more useful model. In this curve, the area under ROC curve is 79%. This indicates the fair condition of the model.

Neural network:

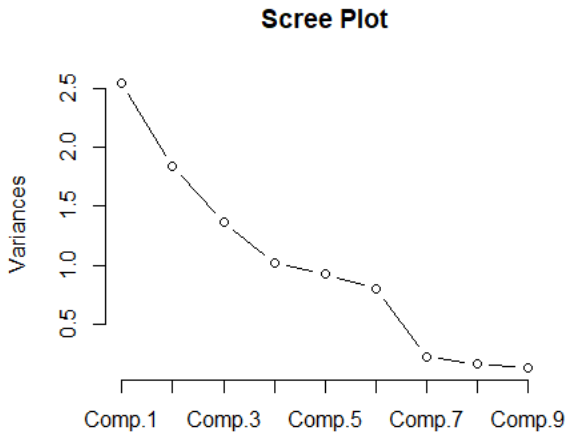


Principle Component Analysis

The Eigen values are
`

2.5502496

1.8413674
1.3634946
1.0205023
0.9202288
0.7988174
0.2214005
0.1602988
0.1236405



In these results, the first four principal components have eigenvalues greater than 1. These three components explain **75.8%**of the variation in the data. The scree plot shows that the eigenvalues start to form a straight line after the fourth principal component. If **75.8%**is an adequate amount of variation explained in the data, then you should use the first four principal components.

Factor Analysis:

The chi-square statistic and p-value in factanal are testing the hypothesis that the model fits the data perfectly. The p-value is **0.649.**

As the p-value is high, we can accept the hypothesis and this model fits the data perfectly

Uniqueness's:

Factor	Uniqueness
tot bilirubin	0.976
direct bilirubin	0.035
tot proteins	0.851
Albumin	0.334
ag_ratio	0.316
Sgpt	0.159
Sgot	0.118
Gender	0.276

When uniqueness is low, the relevance of the variable is more i n the factor model. In this model, direct bilirubin is more releva nt to the factor model.

5 CONCLUSION AND FUTURE WORK

From analysis of the data, bilirubin is the significant component that influences the model. Accuracy of the data to fit the model is 0.7952.From the regression analysis, we are able to find the persons who are affected from jaundice.

6. REFERENCES

[1] Nainika Saini, Ashok kumar, Preeti Khera, Non-Invasive Bilirubin Detection Technique for Jaundice Prediction Using Smartphones,International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 8, August 2016.

[2] M. Penhaker, V.Kasik, B. Hrvolova, "Advanced Bilirubin Measurement by a Photometric Method" lektronika Ir Elektrotechnika, Vol.19, 2014, pp. 47-50.

[3] Smriti Shrivastava,"Diagnosis of Neonatal Jaundice Using Artificial Neural Networks,"International Indexed & Refereed Research Journals, vol. 4, 2013, pp. 69-72.

[4] James W.Kronberg. "Optical Transcutaneous Bilirubin," US Patent 5 259 382, November 9, 1993.

[5] Tai-Wing Wu, Rochester, N.Y., "Assay For Bilirubin", United State Patent 4 069 016, January 17, 1978 pp.1-32.

[6] Mark McEwen, Karen Reynolds, "Noninvasive Detection of Bilirubin Using Pulsatile Absorption,"

[7] David P. Dewitt, Robert E.Hannemann, John F. wiechel "Method for Determining Bilirubin Concentration from Skin Reflectance". United States Patent. 4 029 085, june 4, 1977 pp. 1-8.

[8] Steven L. Jacques, David G. Oelberg,Iyad Saidi,"Method And Apparatus For Optical Measurement Of Bilirubin In Tissue". United States Patent.5, 353, 790, 1994, pp. 1-79.

[9] Gagan Mahajan "Transcutaneous Bilirubinometer in Assessment of Neonatal Jaundice in Northern India". INDIAN PEDIATRICS.Vol. 42, 2005, pp. 41-45.

