TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

Prediction of Liver Diseases using Decision Trees and Machine Learning Algorithms

Debnath Bhattacharyya[1], N. Thirupathi Rao[1], Qin Xin[2]

[1]Department of Computer Science and Engineering
K L Deemed to be University, KLEF,
Guntur - 522502, Andhra Pradesh, India.
[2] Department of Computer Science and Engineering,
Vignan's Institute of Information Technology (A),
Visakhapatnam 530049, AP, India.
[3] Faculty of Science and Technology,
University of Faroe Islands,
FO 100, Faroe Islands, Denmark.
debnathb@kluniversity.in, nakkathiru@gmail.com, QinX@setur.fo

Article Info
Volume 83
Page Number: 8170 - 8178
Publication Issue:
May - June 2020

Article History
Article Received*: 19 November 2019*
Revised*: 27 January 2020*
Accepted*: 24 February 2020*
Publication*: 18 May 2020*

Abstract:

Human body consists of various important organs for processing the entire metabolism of the human being. Some of them are the kidneys, brain, heart and liver. The proper functioning of these organs will lead the person happy life. The identification of problems to these organs is also important at the right point of time. In the current article a GUI model had been developed such that to identify the level of damage happened to liver based on the data given by users. The current model will also provide the details about various problems that causes to the liver through GUI model. The saying in the medical field is always predicting the disease in earlier stages is better than at the crucial period such that to reduce the damage of that problem to that particular organ. The symptoms also become more different and also difficult to analyze at later stages. In the current tool, an attempt had been made to utilize the machine learning algorithms like K-means, ANN and SVM to analyze the liver patients from a group of normal patients. The comparison of the machine learning algorithms for their performance to identify the liver problems based on performance factors also observed. The developed GUI can be a good source for the doctors to analyze liver diseases.

Keywords: *Liver, diseases, GUI, machine learning, ANN, SVM, K-means, doctors.*

1. Introduction

Human body had a versatile number of parts which are playing a key role for the healthier life of any human being. Some of those parts are brain, heart, liver, kidneys, etc,. The successful functioning of these parts will lead the life any human being successful and a happy life [Dimitris Lipara et. Al., Kulwinder Kaur et. Al., Pooja Sharma et. Al.,]. The diseases or the problems occurring with these parts may lead to great damage to the functioning of the human body. The major problem concerned with these components was all these parts working metabolism depends on the functioning of other parts of the body. The identification of problems with these parts and diagnosis at the right intervals at right point of time is more important such that the further damage of the parts and the problem spreading to other parts also can be reduced to some level. When the diseases to these organs are in severe conditions, the symptoms are apparently changes from the regular types of symptoms and it becomes difficult to identify when the diseases are in critical stages. With proper medication and other measures taken by the patients, these problems can be reduced to some level of damage to the other parts of the body. The currently developed application is a graphical user interfacing unit with which the doctors or medical practitioners can utilize to identify the symptoms and give as input and can identify the problems of liver diseases at various stages [Hao Jia, Manmohan Shukla et. Al., V.V.Ramalingam et. Al.,].

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

The current model is also used to identify the number of people with liver problems also identified from the available list of patients. A group of patients and their blood samples and values are made with databases and the data was given as input to the current model and list of patients with liver problems are identified from the available list of total patients in the selected database.

1.1. About Liver

Liver is one of the important organs in the human body [G.Ignisha Rajitha et. Al., Esraa Mamdohi Mai et. Al., Islam MM et. Al., Mei Ying et. Al.,]. The size of the liver is large and its place in the human body is on the right side of the belly of the human being. The colour of the liver is red and brown in colour. It always looks like the feel of a rubber type of material to touch. The tasks performed by the liver in the human body are around 500 tasks [Abhijit singh Putu et. Al., MoonSun Shin et. Al., Jung-euk Ahn et. Al.,]. The best feature of this organ is that it can regenerate at any point of time. Due to this unique property, it can be repaired or the problems occurred to this organ can be easily reduced and can be resolved easily. The normal view of a human liver can be observed in fig 1.

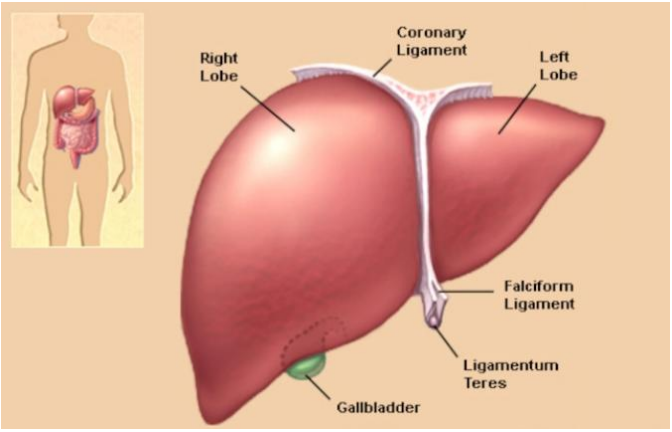

Fig.1: Human Liver [1]

Some of the problems that can occur to these organs are liver cancer, liver failure, gallstones, Ascites are some of the common and more serious issues that were happen to most of the people in terms of liver problems.

1.2. Machine Learning

Machine learning is one of the applications of artificial intelligence and it plays a key role in machines to learn things by observing the previous data and can take decisions based on the previous data [Mei Ying et. Al.]. The data that can be given as input can be considered as the training data from which the machine will learn the facts or the conditions on the data that was supplied and based on such data, the machines tries to take decisions on various scenarios. If the data regarding a particular problem did not exist in the database, then a hypothetical situation will be arid to any machine. Hence, it is the duty of the developers to lookout the data related to all the problems or the issues that the machine is going to face should be there in the training data. Otherwise, the problems of hypothetical errors might arise [Abhijit singh Putu et. Al.,]. Several algorithms and techniques had been developed under this application and are being used highly in the research world today. Some of the important algorithms available today for the better implementation of the current model algorithms are,

a. Support Vector Machine (SVM) [16]
b. Artificial Neural Network (ANN) [16,17]
c. Random Forest

a. Support Vector Machine (SVM)

In machine learning, the supervised algorithms are the important type of algorithms and some important algorithms fall under this group of algorithms. Support vector machine (SVM) algorithm is the important algorithm falls under this category. This algorithm can be used for regression type of problems and for classification type of problems and their analysis [Islam MM et. Al.,]. In order to differ or separate two classes of data, several types of hyper planes are available such that to choose among the available group of hyper planes. In most of the cases, the objective is to identify one plane among the other planes that has the margin values are very high. The hyperplanes example for an SVM algorithm can be seen at the fig.2.

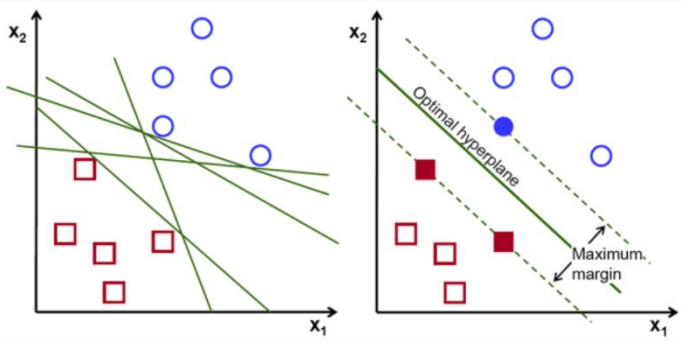

Fig.2: Hyper planes example of an SVM Algorithm [Tom Bromley et. Al.,]

The value is high means it can be taken as that the distance between various points in those classes is

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

having the good margin. As this distance is more, the future classification is also done easily for further future sources. These hyperplanes can also be considered as the boundary decisions to be considered and also for identifying and classifying various pints of data in various databases. When the program is being run, the data points will fall on both sides of the planes, both sides of the planes data points can be considered as the data points belongs to two different classes of data.

b. Artificial Neural Networks

Neural networks are the one of the important applications of machine learning. The most important thing in this model is the current model works on the mechanism of neurons in the human brain such that to resemble the working of a neural systems in the human brain. The major applications or the help that the users are getting from these networks are the cluster and classifying of data. The artificial neural network model can be seen clearly for better understanding at fig.3.

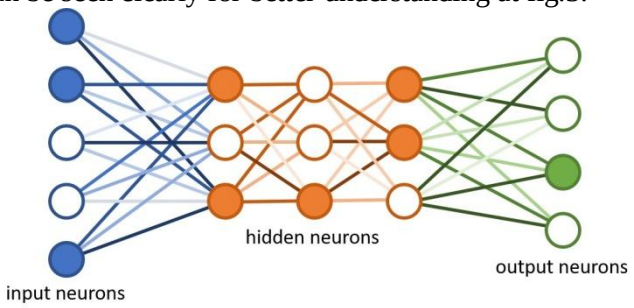

Fig.3: Artificial Neural network model [Tom Bromley]

Some of the major applications of these neural network models are face detection, identifying the objects in images, identifying the expression on faces, identifying the gestures in images and videos, detection of voices, speech translation etc.

d. Random Forest

Classification is one of the major important considerations for any machine learning application. Performing this classification task precisely will give us more accurate and better performances than the expected ones. The major considerations are like predicting a customer for a product purchase or predicting the mood of the customers to check whether the product will be purchased by the person or not. All these sort of applications are key factors for the implementation of these random forest algorithms. Basically this algorithm belongs to the family of decision trees which are used very frequently for the identification of results from these sorts of applications. For better understanding of the random forest example, a model was presented at fig.4.

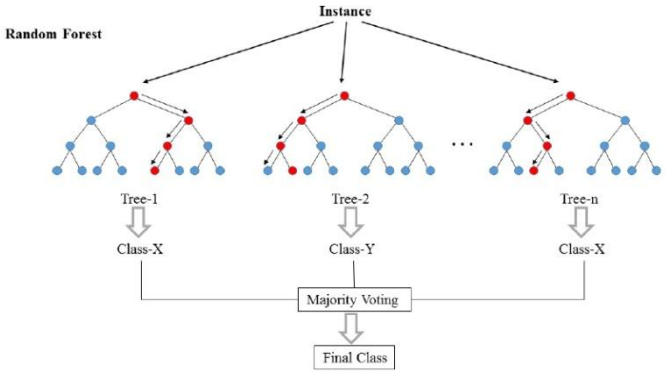

Fig.4: Random Forest Example [4]

In the working of the current algorithm, a various number of decision trees will present, these trees are of two types built. At first, every single tree is built based on the data available or data getting from the original database. And the next step is a subset of features are selected randomly from the each near nodes and the new data set or the new result set of nodes which were classified are given as output.

1.3. The Roc Curve

ROC curve stands for the receiver operating the characteristic curve is basically used to represent the data trends that were showing towards the true false and false true etc. Basically this plot is designed by using the True positive rate and false positive rate at different threshold values.

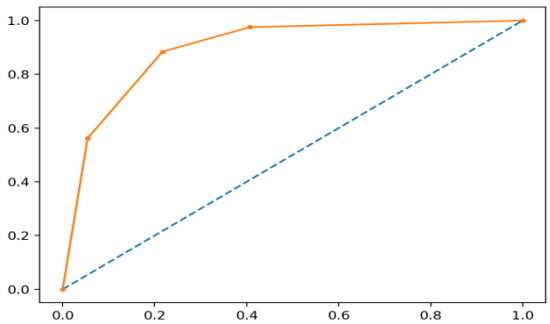

Fig.5: An example of ROC curve model [5]

The mode of classification and the success rate of data classification can also be observed with this curve as an output. Most of the researchers are using the same the mode of curve fro representing various data classification techniques and models now days. In order

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

to have the better understanding of the ROC curve model, an example curve had been presented at fig.5.

2. Literature Review

Several authors had done various works on the area related to the current article. Some of the articles and authors who had done some related works are discussed in the current section to identify the similar set of works,

V.V. Ramalingam, A.Pandian, R. Ragavendran discussed about various algorithms that can be used for identifying the problems of liver diseases in human beings. They had discussed in detail about various algorithms of machine learning and decision trees, and also explained how these algorithms can be used for further analysis of finding diseases in liver and its related components.

G. Ignisha Rajathi and G. Wiselin Jiji discussed in detail about the chronic diseases that might occur with various symptoms and with several causes. All the symptoms and the disease issues were discussed in detail with the help of various algorithms like the Simulated Annealing and Ensemble Classifier. The results are very good and impressive and the identification of these serious diseases can be found easily by utilizing this method of techniques further for achieving the better results.

Esraa MamdouhMai, Mabrouk Mai Mabrouk discussed about diabetes and chronic liver diseases and the reasons for their presence in human bodies are also discussed in detail. SVM algorithm had given some good resulted when their model was implemented with their own dataset.

Islam MM, Wu CC, Poly TN, Yang HC, Li YJ discussed about the fatty liver diseases and their symptoms and the problems and solutions for the same diseases. They had used various machine learning algorithms to identify the symptoms from the pool of items to be noted at various databases and the algorithms for the same. The results and the utilization of the method are described in detail and the results are interesting for further analysis and extensions of the works.

From the above works, almost all the works are done with the various machine learning algorithms. Very few works are done with the combination of both machine learning and decision tree algorithms. Hence, in the current work an attempt has been made to analyze our new dataset with the combination of both these types of algorithms and tried to analyze the performance of these algorithms. The results and representations are discussed in detail in the form of ROC curve and other values.

3. Proposed System

Several authors had considered several algorithms and methods to identify the presence of liver related issues and diseases in human beings with the presence of various symptoms. If the detection of these diseases were done at right intervals of time, the disease and its related issues can be reduced to a great level of relief to these patients. In the current work, an attempt had been made to analyze the performance of the two set of algorithms like the machine learning algorithms and the decision trees algorithms. From the pool of the machine learning algorithms, we had considered the SVM algorithm and the ANN algorithm and from the pool of decision trees algorithms Random Forest algorithm had been chosen. The performance was analyzed by collecting the various set of samples from various hospitals in Visakhapatnam and the results were displayed in the results section.

4. Implementation and Results of the Current Methods

The implementation of the current model considered was done in the following process and the detail process had explained as follows. For the better submission of data to analyze and to process, an integrated graphical user interface had been developed such that to submit the data and to process the data. The graphical interface had been developed with the PHP model and the algorithms which were using in the current model are connected with the current GUI such that user can select the type of algorithm he needs to implement or he can check the data with a particular algorithm. A drop down menu had been created for the three set of algorithms and the user can select his choice of algorithms and also the users can select all the three algorithms individually and can check the results with three algorithms one after the other algorithm.

The disease symptoms and its values may vary sometimes to some levels in males and females. Hence, it is always better to first classify the data for both male patients and female patients and then apply the algorithm from the above selected three algorithms and try to get

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

the results and conclude with those set of results obtained from the above interfacing unit with each algorithm implementation. The detailed architecture of the entire process had been presented in fig.6.

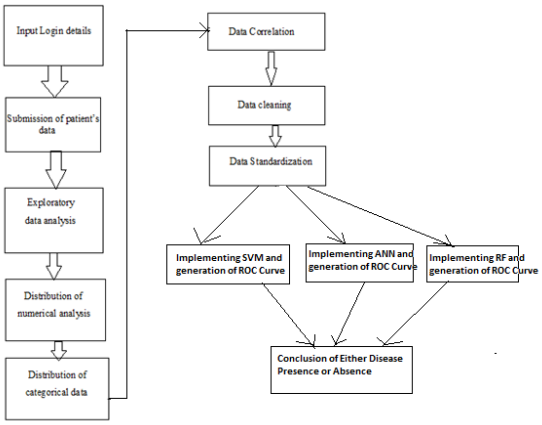

Fig.6: Architecture model of the current method and flow of execution

Whenever the data was being submitted by the user for checking with the currently developed model, the machine will collect the knowledge or expertise from the database that we had already submitted with previous results and symptoms. The machine learning algorithms needs to work on some existing data of patients such that to train from that particular data and take the decisions. Hence, a set of patients data was collected from various hospitals and laboratories in and around the Visakhapatnam City and a dataset of patients with both male and female with various ages had prepared. Once the data is collected and dataset is prepared, the cleaning of the data and correlation of the data is to be done and then the errors and noise in the database was removed for fast and easy working of the machine with these algorithms. Once the entire process of this data analysis had completed, the standard format of the data was prepared and the data needs to be analyzed for further algorithms to get more accurate results based on the various symptoms and their values present. Whenever the new choice of patient's data is to be verified, the data is being verified and then the current record of patient's data was added to the existing dataset. Once the data is cleaned and ready for checking the data had been distributed on the categorical basis. Then the data correlation and cleaning of data had been completed and the algorithms are implemented and the data results are represented in the form of ROC curve and the results presentation.

When the database is present with various patients' data with various attributes, the dataset is to be distributed with various numerical features. Some of those features are like age, albumin, alamine_Aminotransfarese, Albumin and Globulin Ratio, Alkaline Phosphate, direct bilirubin, total bilirubin, etc. The user interface of the current model was presented at fig.7 for better understanding.

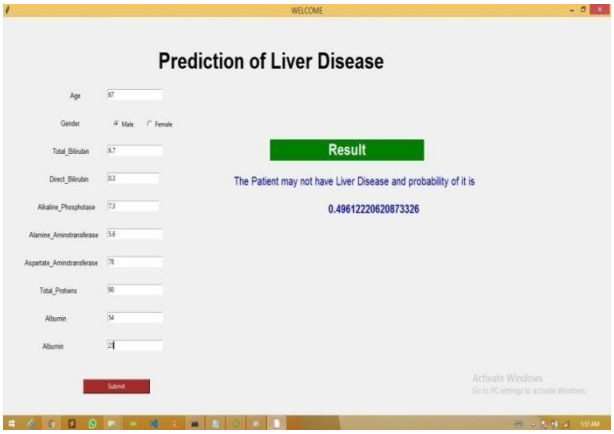

Fig.7: The user interface of the current developed model

In order to understand the data and its components perfectly, exploratory data analysis and the distribution of numerical features had been done such that for the better understanding of the trends of the data. The exploratory data analysis model was presented in fig.8 and distribution of numerical values is also observed in fig.9 in detail.

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Asparta
0	65	Female	0.7	0.1	187	16	
1	62	Male	10.9	5.5	699	64	
2	62	Male	7.3	4.1	490	60	
3	58	Male	1.0	0.4	182	14	
4	72	Male	3.9	2.0	195	27	

Fig.8: Exploratory Data Analysis

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

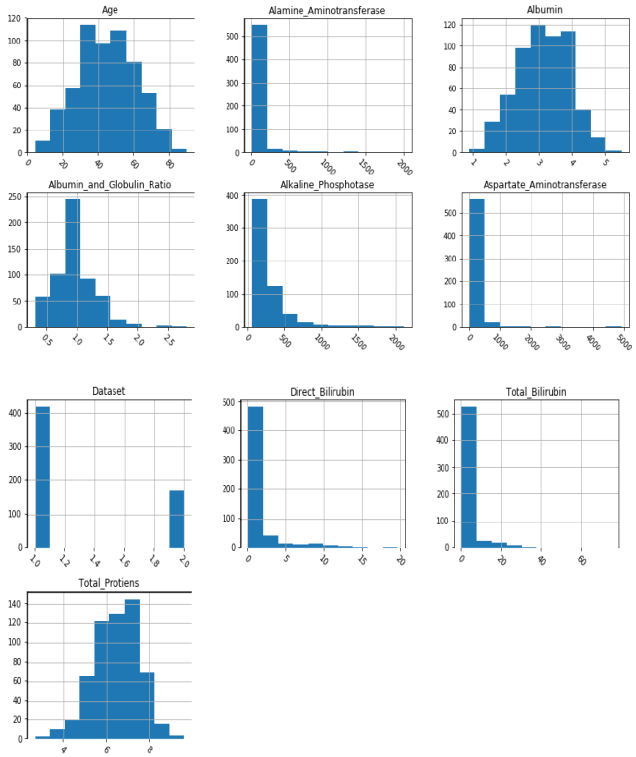

Fig.9: Distribution of Numerical Features

The resultant values after distribution of parameters and their components are shown and presented in fig.10.

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspar
count	583.000000	583.000000	583.000000	583.000000	583.000000	
mean	44.746141	3.298799	1.486106	290.576329	80.713551	
std	16.189833	6.209522	2.808498	242.937989	182.620356	
min	4.000000	0.400000	0.100000	63.000000	10.000000	
25%	33.000000	0.800000	0.200000	175.500000	23.000000	
50%	45.000000	1.000000	0.300000	208.000000	35.000000	
75%	58.000000	2.600000	1.300000	298.000000	60.500000	
max	90.000000	75.000000	19.700000	2110.000000	2000.000000	

Fig.10: Resultant values after the proper distribution of values or the symptoms

It seems there is outlier in Aspartate_Aminotransferase as the max value is very high than mean value. Dataset i.e the output value has '1' for liver disease and '2' for no liver disease so let's make it 0 for no disease to make it convenient.

From the above table, the distribution of category and the features are plotted in the form of Bar graph for better understanding of the values and the symptoms are presented in fig.11 and fig.12.

	Gender
count	583
unique	2
top	Male
freq	441

Fig.11: Distribution of categorical data

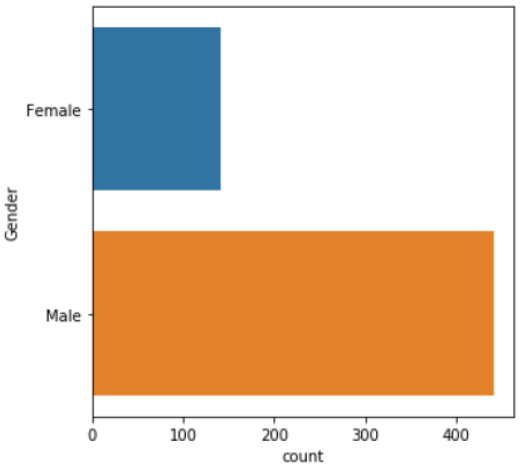

Fig.12: Bar plots for categorical Features

Number of patients that are male: 441
Number of patients that are female: 142

There are more male patients than female patients, Label Male as 0 and Female as 1. Once the data labelling format is decided, the represented format of the data in the below dataset table I,

Table I. Dataset format with the representation of the labelling

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspart
0	65	1	0.7	0.1	187	16	
1	62	0	10.9	5.5	699	64	
2	62	0	7.3	4.1	490	60	
3	58	0	1.0	0.4	182	14	
4	72	0	3.9	2.0	195	27	
5	46	0	1.8	0.7	208	19	
6	26	1	0.9	0.2	154	16	
7	29	1	0.9	0.3	202	14	
8	17	0	0.9	0.3	202	22	
9	55	0	0.7	0.2	290	53	
10	57	0	0.6	0.1	210	51	
11	72	0	2.7	1.3	260	31	
12	64	0	0.9	0.3	310	61	
13	74	1	1.1	0.4	214	22	
14	61	0	0.7	0.2	145	53	

The correlations between the various parameters and the symptoms and their values are represented in the form of

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

a table format with their represented values which are correlated with other. The parameters considered with various symptoms and their values are represented at table II as,

Table II. Correlated values representation of symptom values

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	A
Age	1.000000	-0.056560	0.011763	0.007529	0.080425	
Gender	-0.056560	1.000000	-0.089291	-0.100436	0.027496	
Total_Bilirubin	0.011763	-0.089291	1.000000	0.874618	0.206669	
Direct_Bilirubin	0.007529	-0.100436	0.874618	1.000000	0.234939	
Alkaline_Phosphotase	0.080425	0.027496	0.206669	0.234939	1.000000	
Alamine_Aminotransferase	-0.086883	-0.082332	0.214065	0.233894	0.125680	
Aspartate_Aminotransferase	-0.019910	-0.080336	0.237831	0.257544	0.167196	
Total_Protiens	-0.187461	0.089121	-0.008099	-0.000139	-0.028514	
Albumin	-0.265924	0.093799	-0.222250	-0.228531	-0.165453	
Albumin_and_Globulin_Ratio	-0.216408	0.003424	-0.206267	-0.200125	-0.234166	
Dataset	0.137351	-0.082416	0.220208	0.246046	0.184866	

Once the data in the dataset was stabilized and values are represented, now it is the time to clean the data and tries to remove the outliers from the dataset and the process is implemented and the resultant data can be shown in the form of the graph at fig.13 as,

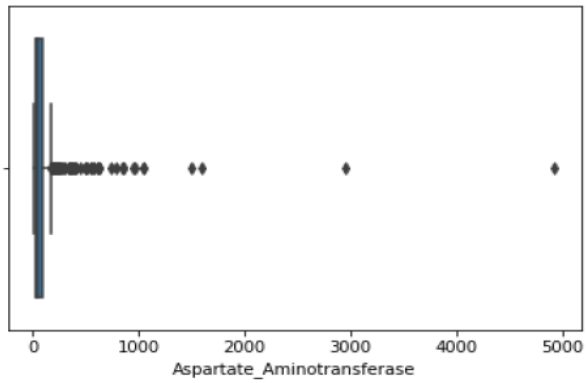

Fig.13: Data representation after cleaning the data

Once the data is finalized for testing and implementing from the dataset, now the standard procedure of data standardization needs to be implemented and the algorithms had also been used on the same dataset for further verification of the performance of the considered algorithms and their performance. The training dataset that was been done before going for the implementation can be seen at table III

.

Table III. Standardizing the data for implementation with algorithms

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminot
count	4.510000e+02	4.510000e+02	4.510000e+02	4.510000e+02	4.510000e+02	4.5
mean	8.517454e-17	2.067821e-17	-1.839499e-16	-1.493878e-15	3.938707e-17	-5.
std	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.0
min	-2.459710e+00	-5.980292e-01	-4.613674e-01	-5.013974e-01	-9.009611e-01	-3.
25%	-7.595886e-01	-5.980292e-01	-3.996190e-01	-4.658144e-01	-4.636908e-01	-3.
50%	2.975347e-02	-5.980292e-01	-3.687448e-01	-4.302314e-01	-3.270438e-01	-2.
75%	7.583769e-01	1.668451e+00	-1.063141e-01	-7.440091e-02	2.042996e-02	-9.
max	2.762091e+00	1.668451e+00	1.105471e+01	6.472879e+00	7.090935e+00	9.7

In the next level of the process, the standard deviation method is also applied to standardize the test data and the test data in the dataset such that to get better results. The training process for the above table 3 had been done and the resultant table results are as follows at table IV.

Table IV. The dataset table after implementing the Standard Deviation to the training dataset

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransf
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.0
mean	0.109816	-0.216940	-0.062052	-0.033465	-0.048222	-0.01
std	0.937258	0.851421	0.835512	1.046110	0.734043	0.6
min	-2.338273	-0.598029	-0.445930	-0.501397	-0.768218	-0.3
25%	-0.577433	-0.598029	-0.399619	-0.465814	-0.444170	-0.3
50%	0.090472	-0.598029	-0.384182	-0.430231	-0.342661	-0.2
75%	0.940533	-0.598029	-0.198937	-0.181150	-0.018612	-0.0
max	1.851312	1.668451	4.231512	5.974717	3.147694	4.8

At first, the Support Vector Machine is being implemented with the selected dataset and the results are obtained in the form of both the ROC curve and actual values. The other two algorithms also had followed the same type of process and the results obtained are shown and the final observation about the performance of all the three algorithms are discussed and given the observations.

Case 1.

The results obtained for the SVM algorithm by calculating the ROC curve from the y_test and predictions are observed at fig.14 as,

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

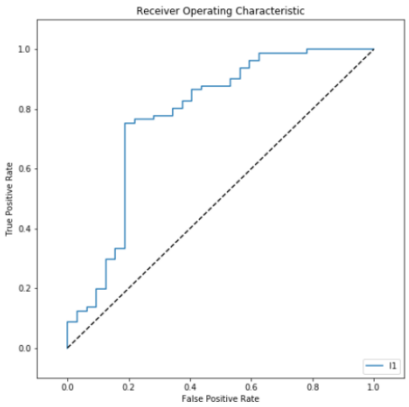

Fig.14: ROC curve for SVM algorithm

The accuracy of the above algorithm is observed as,

```
In [51]:  1 # Calculate AUC for Train
          2 roc_auc_score(y_train, y_train_pred)

Out[51]: 0.7782890007189072

In [52]:  1 print(auc(fpr, tpr))

0.677391975308642
```

Case 2.

The results obtained for the Random Forest algorithm by calculating the ROC curve is presented at fig.15 as,

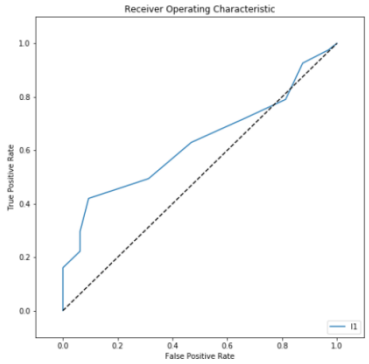

Fig. 15: ROC curve for the second algorithm

The output result from the current algorithm as follows,

```
In [68]:  1 # Calculate AUC for Train
          2 roc_auc_score(y_train, y_train_pred)

Out[68]: 0.6376947040498442

In [69]:  1 # Calculate AUC for Test
          2 print(auc(fpr, tpr))

0.6282793209876544
```

Case 3.

The results obtained for the Artificial Neural Network model algorithm by calculating the ROC curve and value observed at fig.16 as,

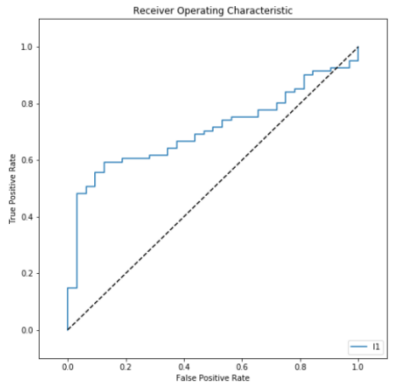

Fig. 16: ROC curve for the ANN model

The output result from the current algorithm are shown in fig.17 as,

```
Neural Score:
 77.83
Neural Test Score:
 69.91
Accuracy:
 0.6991150442477876
[[ 5  7]
 [27 74]]
              precision    recall  f1-score   support

           0       0.42      0.16      0.23        32
           1       0.73      0.91      0.81        81

   micro avg       0.70      0.70      0.70       113
   macro avg       0.57      0.53      0.52       113
weighted avg       0.64      0.70      0.65       113
```

Fig.17: Output results for the model

From the above implemented all the three models with three algorithms, the ANN model is having the better results and better performance compared with the other two algorithms.

5. Conclusions

In the current article, an attempt has been made to analyze and diagnose the liver diseases in patients using various machine learning and decision trees algorithms. The machine learning and decision trees algorithms are working perfectly on the selected and prepared dataset of various patients who are suffering with various sliver diseases with various levels of the diseases. From the considered three algorithms, the ANN model is having the better performance when compared to the other two methods of algorithms. The GUI developed for performing the current applications and the tasks is very useful and helpful for the normal people also to work on it. Any educated person also can work on it with minimum knowledge of computer operations.

6. References

[1] https://www.webmd.com/digestive-disorders/picture-of-the-liver#1 [Last Accessed on 10-03-2019]

May – June 2020
ISSN: 0193-4120 Page No. 8170 - 8178

[2] https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47. [Last Accessed on 10-03-2019]
[3] A blog written by Tom Bromley in Xanadu AI entitled, Making a Neural Network, Quantum" on February, 2018. [Last Accessed on 10-03-2019]
[4] Stavros I Dimitriadis, Dimitris Lipara (2018). How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regeneration Research,* 13, 6, 962-970.
[5] A blog written by Jason Brownlee in Python Machine Learning in 2018 entitled "How and When to Use ROC Curves and Precision-Recall Curves for Classification in Python". [Last Accessed on 10-03-2019]
[6] Kulwinder Kaur and Vikas Zandu, (2016). A Secure Data Classification Model in Cloud Computing Using Machine Learning Approach. *International Journal of Grid and Distributed Computing*, 9, 8, 13-22, SERSC Australia.
[7] Pooja Sharma and A K Mishra (2017). Analysis and Classification of Plant microRNAs using Machine Learning Approach. *Asia-Pacific Journal of Neural Networks and Its Applications*, 1, 2, 1-6, GVPress.
[8] Hao Jia (2015). Large-Scale Data Classification Method Based on Machine Learning Model. *International Journal of Database Theory and Application*, 8, 2, 71-80, SERSC Australia.
[9] Manmohan Shukla, and B.K. Tripathi (2017). On the Efficient Machine Learning of the Fundamental Complex-Valued Neurons. *International Journal of Neural Systems Engineering*, 1, 1, 21-26, GVPress.
[10] V.V. Ramalingam, A.Pandian, R. Ragavendran (2018). Machine Learning Techniques on Liver Disease - A Survey. *International Journal of Engineering & Technology*, 7, 4.19, 485-495.
[11] G. Ignisha Rajathi, and G. Wiselin Jiji (2019). Chronic Liver Disease Classification Using Hybrid Whale Optimization with Simulated Annealing and Ensemble Classifier. *Symmetry*, 11, 33, 1-21.
[12] Esraa MamdouhMai, MabroukMai Mabrouk, (2014). A Study of support vector machine algorithm for liver disease diagnosis. *American Journal of Intelligent Systems*, 4, 1, 9-14.
[13] Islam MM, Wu CC, Poly TN, Yang HC, Li YJ (2018). Applications of Machine Learning in Fatty Live Disease Prediction. *Student Health Technological Information*, 247, 166-170.
[14] Mei-ying Ren and Sin-jae Kang (2017). Korean 5W1H Extraction Using Rule-based and Machine Learning Method. *International Journal of Artificial Intelligence and Applications for Smart Devices*, 5, 2, 1-10, GVPress.
[15] Abhijitsingh Putu Gaonkar, Ritwik Kulkarni, Ronnie D. Caytiles and N.Ch.S.N Iyengar (2017). Classification of Lower Back Pain Disorder Using Multiple Machine Learning Techniques and Identifying Degree of Importance of Each Parameter. *International Journal of Advanced Science and Technology*, 105, 11-24, SERSC Australia.
[16] MoonSun Shin, HaYoung Lee, JinSoo Kim and JungHwan Kim (2018). Design of Intelligent Healthcare Support System Applying Machine Learning. *International Journal of Advanced Science and Technology*, 111, 73-84, SERSC Australia.
[17] Jung-euk Ahn, Eun-Surk Yi, Ji-Youn Kim and Byung Mun Lee (2018). A Multiracial Data-Based Machine Learning Model for Exercise Motion Recognition. *International Journal of Control and Automation*, 11, 2, 89-102, SERSC Australia.