A Comparative Study For Liver Disease Prediction Using Machine Learning (ML)

[1]Tejas Chavda, [2]Saket Swarndeep

[1]Student of ME, [2]Assistant Professor
[1]Computer Engineering,
[1]LJ University, Ahmedabad, India
[1]tc4220@gmail.com, [2]sanket.swarndeep@ljinstitutes.edu.in

Abstract— **Liver is Vital internal Organ of Human Body which performs 500 functions in our human body. Its main task is to Digest Food, Eliminate waste and store nutrition and vitamins in form of energy called glycogen. It is the only organ which functions well even if it is 50% damaged hence Liver disease symptoms are subtle in nature which makes it for difficult to identify. Early Diagnosis and treatment of patient can help to reduce the risk. Due to Lethal nature of liver disease it's diagnosing process is quite expensive and sophisticated. Hence we are trying to implement machine learning algorithms which are going to classify the liver disorder.**

Keywords: Liver disease prediction, Liver disorder, Machine Learning, smart system to diagnose liver disease, Prediction system, classifiers

INTRODUCTION

Liver diseases cause millions of deaths worldwide every year. Viral hepatitis alone causes 1.34 million deaths every year. It is expected that by 2025 India may become the World Capital for Liver Diseases[1]. Liver disorders can be recognized earlier by analyzing the levels of enzymes in the blood. Liver is a vital & largest body organ, located in upper right position of abdomen. Its weight is nearly 1.36 kgs. It plays crucial role in digestion, detoxification (clearing blood), decomposition of red blood cells etc. It is cone shape organ performing more than 500 vital functions. Gallbladder, Pancreas & Intestine work along with Liver to digest, absorb and process food. It is the only organ in the body which can regrow or regenerate itself even after removing 90% out from body.

Liver converts excess glucose into glycogen for storage and also regulates the blood clotting. The main issue with liver disease is it is not discovered at earlier stage because liver keeps on function properly even when it is 50% damaged.

There are various types of of Liver disorders occurring due to infections or uneven habits such as

- Over Drinking of liquor
- Smoking
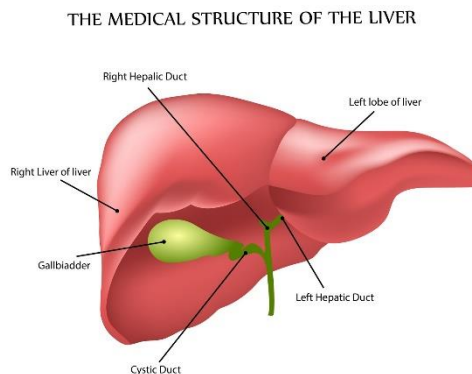- Genetic Disorder of DNA
- Viral infections
- Diabetes

Fig - 1: Structure of Liver [2]

Presently Liver Diseases are diagnosed using Blood testing, Imaging Tests(CT Scan, Ultra sound), & Checking tissue samples (biopsy) which are very time consuming process. Hence We are proposing smart Machine learning based system which provide accurate & Higher accuracy for Liver diseases. Some of ML Classification algorithms such as decision trees, Extra Trees, Naïve Bayes, Knn etc are giving accuracy of 70% approx. However not even 95-99% afterall its related to human health. No Risk can be taken apart.

Delay in Diagnosing may result delay in medications. Our system will play crucial role in delivering efficient services to physicians so that better treatments can be offered to the patients

The main Objective of Machine learning is to improve the Disease Diagnosing accuracy of Healthcare sector to the next level which can ultimately help more and more people to save there and lives. The Goal is to create such system where just few information about patient will help to know if he/she has liver disease or not. If he/she has disease the process for medical treatment can be started as early as possible.

Liver Disorders:
Disorders represents the disruption of organ from proper functioning. There are nearly 100 types of liver infection[1]. Liver disorders take place due to some disease as below:

1. **Hepatitis:** Its state where Inflammation of the liver is usually caused by viruses like hepatitis   A, B, and C. In-some non-infectious cases it can be through Drugs, Allergic reaction, excessive hard   drinking or obesity

2. **Cirrhosis:** It's a Last- stage chronic liver disease where a healthy liver tissue is replaced with scar tissue and the liver is permanently damaged. Scar tissue usually block the flow of blood through the liver and slows the ability to process nutrients, drugs and toxin.

3.   **Liver Failure:** It can occur due to some kind of infections, genetic disease from DNA or Excessive alcohol consumption.

4. **Inherited Diseases:** Some liver problems might be happening due to certain genetically Designed structure of DNA. Inherited diseases are received to child from genes of parents by birth

I. MACHINE LEARNING

Machine Learning is branch of computer science which uses certain datasets and algorithms to copy humans learning methodology. Nowadays with the increase in internet users in India and worldwide there has been tremendous increase in data volume created, stored, maintained worldwide. World Generated volume about 64.5 Zettabytes in year 2020.[3] Machine learning is main reason for storing such large datasets it usually helps to get certain information from those datasets by finding hidden Patterns. It Plays vital Role in making predictions, Fraud detection, Classifications, Recommendations, Clustering, Decision making, problem solving, reasoning and lot more.

Fig – 2: Understanding of Machine Learning [4]

Nowadays Machine learning in various real time sectors such as banking, E-commerce, Recommendation systems, Sales and marketing, Financial services and Health sector as well.

Machine Learning Techniques:

1. **Supervised Learning:** It is process where dataset feed to the data model consists of labelled column representing the output. Supervised Algorithms are Naïve Bayes, Logistic Regression, SVM, Random Forest and many more
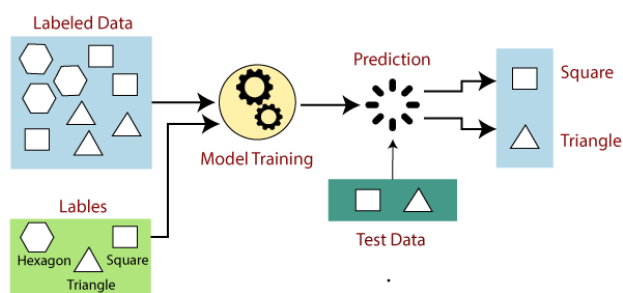

Fig – 3: Supervised Learning Model [5]

2. **Unsupervised Learning:** It is process where dataset feed to the data model don't consist of labelled column representing the output. It find the output from hidden patterns. Unsupervised Algorithms are Association & Clustering
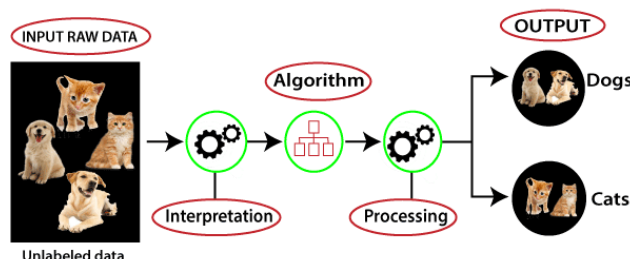

Fig - 4: Unsupervised Learning Model [6]

3. **Reinforcement Learning:** It is process where machine gets some positive or negative rewards based on action or decisions taken. Machine makes observations of rewards and acts accordingly.
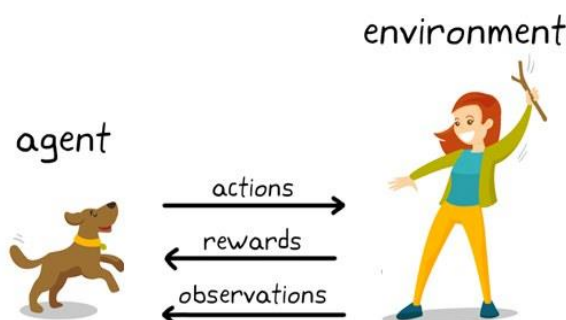

Fig - 5: Reinforcement Learning Model [7]

Machine Learning Algorithms:

1. **Logistic Regression Algorithm**

   Logistic regression is a very popular Supervised Learning algorithm works on probability of an event. It is mainly predicts the binary outcome such yes , no, true or false etc. The output of this algorithm can be only between 0 & 1. It uses the activation function called Sigmoid. The equation for LR is as below:

   $$\log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

   The curve obtained is called as sigmoid curve or S-curve. Consider the below image:
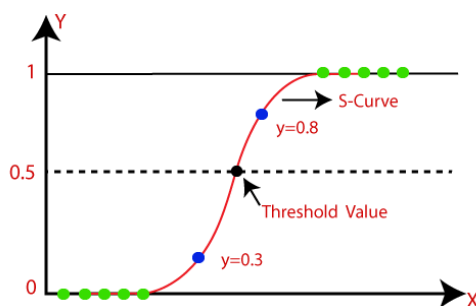

   Fig - 6: S Curve of LR [8]

2. **SVM Algorithm with RBF Kernel**

   SVM is most stable and popular supervised Learning Algorithm. SVM stands for support Vector   Machine. It is used for classification problems. The ultimate goal of this algorithm is to create best line diffentiates n-dimensions in classes. The best decision boundary is called hyperplane.
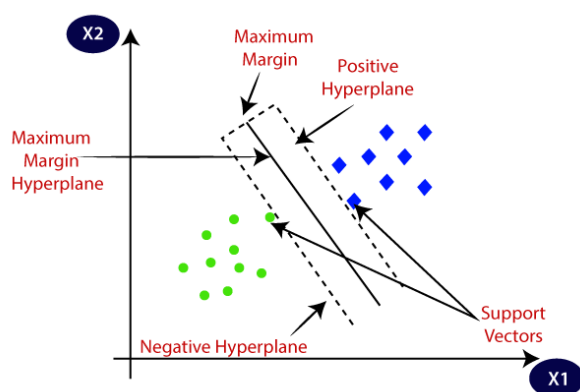   It can be depicted from diagram shown:

   Fig - 7: SVM Working [9]

SVM supports various kernels such as Gaussian, Sigmoid, polynomial, etc. Kernels are pcs of code functioning as library. It takes the data as input and converts into required processing format.
   RBF stands for Radial Basis Function it is on of SVM kernel. RBF method uses radial basis method to improve the processing of data. The equation for RBF kernel is as below:

$$K(x,y) = e^{-}(\gamma||x-y||^2)$$
$$K(x,x1) + K(x,x2)(Simplified - Formula)$$
$$K(x,x1) + K(x,x2) > 0(Green)$$
$$K(x,x1) + K(x,x2) = 0(Red)$$

3. **LGBM Classifier**

   LGBM stands for Light Gradient Boost Machine. It is Fast, Distributed gradient boosting framework based on Decision Tree algorithm. It is used for various purposes in Machine Learning such as ranking, classification etc. LGBM is Faster than XGBoost. It needs less amount of storage space.

$$\bar{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b}\sum_{x_i \in B_l} g_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b}\sum_{x_i \in B_r} g_i\right)^2}{n_r^j(d)}\right)$$

4. **Random Forest**

Random Forest is supervised learning Ensemble based technique. It can be used for both classification and Regression. It is kind of classifier that contains certain number of decision trees, the higher the number of trees leads to higher accuracy and prevents overfitting. It is a set of decision tree from a randomly selected subset of the training set, it collects the votes from different decision trees to decide the final prediction result. The fig (8) depicts the working of basic decision tree.
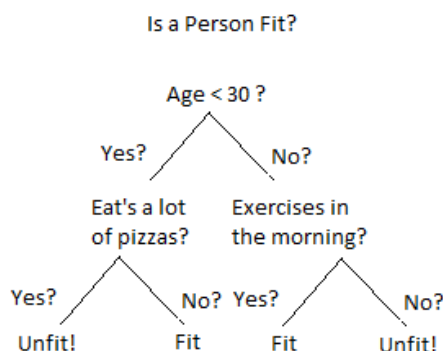

Fig – 8: Decision Tree Model [10]

II. LITERATURE SURVEY

Jagdeep Singh a.[11] in there literature work authors have used 6 classification algorithms such as Logistic Regression, SMO, Random Forest algorithm, Naive Bayes, J48 & K-nearest neighbor are implemented on Liver patient dataset to find the accuracy. There derived results were giving 72.50 % accuracy max. through LR.

G. Shobana [12] in there literature work authors compares issue with ML methods & use Methods such as Gradient Boosting, XGBoost, LGBM Classifier, CAT Boost and tries to improve prediction accuracy by adding Feature selection technique such as RFE (Recursive Feature Elimination). Using LGBM Classifier they derived highest accuracy of 94%

A.Sivasangari [13] in there literature work authors using machine learning algorithms i.e SVM, Decision Tree & Random Forest to predict or classify. The dataset of liver disease is gathered from UCI Machine Learning Library. Using SVM they got the accuracy of 95%

Rong-Ho Lin [14] in there literature work author is using machine learning algorithms i.e CART (Classification and Regression Tree) & CBR (Case Based Reasoning). Actual implementation shows that the intelligent diagnosis model can be of integrated to examine liver diseases with considerable accuracy. CART gave the accuracy of 92.94% and CBR gave 90% accuracy.

Shivangi Gupta [15] in there literature work the authors develop system using machine learning ensemble method. It combines the results KNN, Decision Tree and Neural Networks and tries to predicts the results. It provides the accuracy of 93%.

Thirunavukkarasu K. [16] in there literature work authors uses ML algorithms such as LR, SVM & KNN. Out of that LR & SVM gave similar accuracy of 73.97%.

The Above Researches points out that the despite of using high-end algorithms the results are max to max 95%. We conclude that using right method and right preprocessing can help to get better accuracy.

III. DATASET DESCRIPTION

Dataset is being used which is downloaded from UCI Repository, consists of 583 instances and 11 Data Columns & records were collected from North East of Andhra Pradesh, India. It contains 416 liver patient records and 167 non liver patient records.

The data-set consists of 11 columns, each attributes listed below:

1. Age of the patient
2. Gender of the patient
3. Total Bilirubin
4. Direct Bilirubin
5. Alkaline Phosphotase
6. Alamine Aminotransferase
7. Aspartate Aminotransferase
8. Total Protiens
9. Albumin
10. Albumin and Globulin Ratio
11. Dataset: field used to split the data into two sets (patient with liver disease, or no disease)

IV. PROPOSED SYSTEM

Our Proposed System uses various algorithms such as Logistic Regression, LGBM Classifier, SVM Classifier with RBF Kernel & Random Forest Algorithm. Our Proposed system has several phases in-order to improve the Accuracy of results
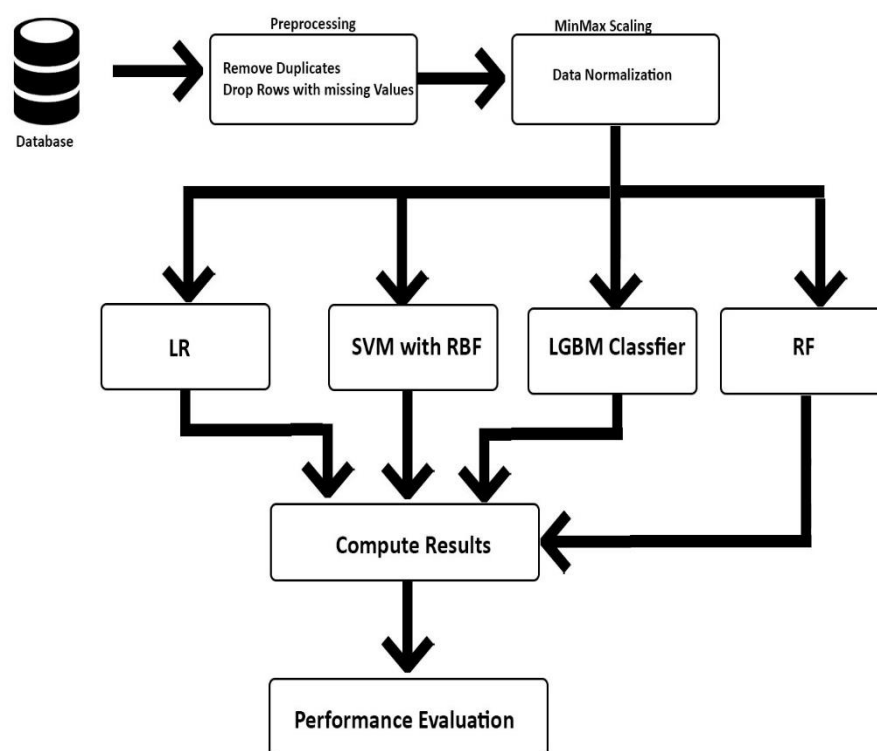

Fig - 9: Proposed Model

Step 1: The first step is to clean the data. Filling the missing values, Remove the duplicates & transform the data from nominal attribute to binary attribute.

Step 2: The Next Step is Normalize the data by implementing MinMax Scaling on given data. MinMaxscaler is a type of scaler that scales the minimum and maximum values to be 0 and 1 respectively

Step 3: In this Step the scaled and Non Scaled Data is pass to Various Algorithms.

Step 4: Results Computed and Performance of Each Algorithm is evaluated to know the best model for prediction

V. CONCLUSION

Machine Learning is not just limited to something it has wide scope to resolve real time problems we just need to understand its techniques and find the place where it can be implemented. It has capability to automate the task which are currently performed by humans. Machine learning models can be trained to act well as we humans do and learn from nature.

REFERENCES

[1] Rakshith D B , Mrigank Srivastava , Ashwani Kumar, Gururaj S P, 2021, Liver Disease Prediction System using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 06 (June 2021)

[2] Narayana Health(2015, November). https://www.narayanahealth.org/blog/liver-function-tests-a-healthy-liver-for-a-healthy-life/

[3] Foresttech(2021,April). https://foresttech.events/the-world-generated-64-2-zettabytes-of-data-last-year-but-where-did-it-all-go/

[4]   Ori Abramovsky(2021,August).   https://spectralops.io/blog/how-to-create-a-code-dataset-for-ai/

[5] Javatpoint. https://www.javatpoint.com/supervised-machine-learning

[6] Javatpoint.https://www.javatpoint.com/unsupervised-machine-learning

[7] KDnuggest https://www.kdnuggets.com/2019/10/mathworks-reinforcement-learning.html

[8] Javatpoint. https://www.javatpoint.com/logistic-regression-in-machine-learning

[9] Javatpoint. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[10] Chirag Sehra (2018, January). https://chirag-sehra.medium.com/decision-trees-explained-easily-28f23241248

[11] Singh, J., Bagga, S., & Kaur, R. (2020). Software-based prediction of liver disease with feature   selection   and   classification techniques. *Procedia Computer Science*, *167*, 1970-1980.

[12] Shobana, G., & Umamaheswari, K. (2021, April). Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1223-1229). IEEE.

[13] Sivasangari, A., Reddy, B. J. K., Kiran, A., & Ajitha, P. (2020, October). Diagnosis of Liver Disease   using Machine Learning Models. In *2020 Fourth International Conference on I-SMAC (IoT in   Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 627-630). IEEE.

[14] Lin, R. H. (2009). An intelligent model for liver disease diagnosis. *Artificial Intelligence in Medicine*, *47*(1), 53-62.

[15] Gupta, S., Karanth, G., Pentapati, N., & Prasad, V. B. (2020, September). A Web Based Framework for Liver Disease Diagnosis using Combined Machine Learning Models. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 421-428). IEEE.

[16] Singh, A. S., Irfan, M., & Chowdhury, A. (2018, December). Prediction of liver disease using classification algorithms. In *2018 4th international conference on computing communication and automation (ICCCA)* (pp. 1-3). IEEE.

[17] ML Science https://www.ml-science.com/confusion-matrix