Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7766223

A COMPARATIVE REVIW ON LIVER DISEASE PREDICTION USING MACHINE ALGORITHM

Mr. Darshan Poonambhai Patel[1], Dr. Dushyantsinh B. Rathod[2]
PhD Scholar[1] ,Professor & HOD[2]
[1,2],Computer Engineering Department ,
Swarnim Startup & Innovation University,Kalol,Gujarat[1],
Ahmedabad Institute of Technology,Ahmedabad,Gujarat[2]

ABSTRACT
Human beings, the liver is the most primary part of the body and performs many functions, including the production of bile, excretion of bile and bilirubin, metabolism of proteins and carbohydrates, activation of enzymes, storage of glycogen, vitamins, and minerals, plasma protein synthesis, and the production of clotting factors. The liver is easily harmed by the consumption of alcohol, pain reliever tablets, and a variety of wired practices. Currently, liver-related diseases are identified by analysing liver function blood test reports and scan reports. It takes more time and is more expensive. While employing different data mining algorithms to ease this process, it is possible to reduce the time for diagnosing liver disease. the machine learning model made before us for prediction for liver disease were a very good models but their accuracy is very low and it was not available for everyone or common people, Also it is the fact that when more data is used, The prediction will be more accurate and to make sure of that we have created multiple machine learning model to get the best accuracy, model like Random Forest, Decision Tree, SDG Classifier, Grid Search, Logistic Regression, XG Boost all this models gave accuracy between 60% -70%, then we also used SVM which gave the highest accuracy of 80% and deployed this model so that every common people can you it..
Keywords: Machine Learning, Liver Disease, SDG Classifier, XG Boost, Accuracy, Random Forest, Decision Tree, Support Vector Machine.

I. INTRODUCTION

With a growing trend of sedentary and lack of bodily activities, diseases associated with liver have become a common come upon these days. Viral hepatitis alone reasons 1.34 million deaths every year. An early diagnosis of liver problems will boom affected person's survival rate. Liver screw ups are at high fee of danger amongst Indians. It's far expected that by way of 2025 India may emerge as the sector capital for liver sicknesses. The sizeable occurrence of liver contamination in India is contributed due to deskbound life-style, extended alcohol consumption and smoking. There are approximately a hundred varieties of liver infections. With such alarming figures, it is essential to have a subject in the direction of tackling those sicknesses. Afterall, we cannot assume an advanced and wealthy state, with unhealthy youths. Symptoms of liver disease include:

- Pain in the abdomen (or belly) (especially on the right side).
- prone to bruising.
- alterations in the color of your stools or urine.
- Fatigue.

- indigestion or vomiting
- swelling in your legs or arms

Your risk of developing liver cancer can be increased by certain types of liver disease. Others continue to harm your liver if left untreated. Scarring, or cirrhosis, develops.

Over the long haul, a harmed liver will not have sufficient sound tissue to work. Failure of the liver can result from untreated liver disease.

Your doctor will also recommend one or more tests in order to accurately diagnose and determine the cause of liver disease. Some examples include:

Blood tests: The levels of liver enzymes in your blood are measured by liver enzymes. The international normalized ratio (INR), a blood-clotting test, is one more way to measure liver function. Abnormal levels may indicate liver function issues.Tests for images: To check for liver damage, scarring, or tumors, your doctor might use ultrasound, MRI, or CT scans. The degree of scarring and fat deposition in the liver can be assessed with the help of fibro scan, a different kind of specialized ultrasound.

Liver biopsies: A tiny amount of liver tissue is taken out of your body with a thin needle during a liver biopsy. They look at the tissue to see if there are any signs of liver disease.

so to save your life and your expenses in our proposal we are developing a machine learning model that works as a liver disease prediction model With The Accuracy Of 80%, which can make your conscious about your liver disease way before it too late. and deployed the model so it is available for everyone to used.

OBJECTIVE OF RESEARCH

The main objective of this Research is:

- Increased convenience for predicting a liver disease: As of now there is no accurate model that can tell you if you are in danger of having a liver disease from home just by giving some attributes. and the prediction comes to be danger or probability of having the liver disease you may consult our doctor as soon as possible
- Reduction in number of deaths due to liver diseases: Viral hepatitis alone causes 1.34 million deaths every year, Our Main Aim To Save As Many Life As We Can Just By make People Conscious About Their Liver.
- More accurate diagnosis of liver disease by the doctors: As People Are Now Concert About Their Liver Now After Using Our Machine Learning Model Their Will Contact Their respective Doctors about there Concern and that diagnosis will be more accurate.

II. LITERATURE SURVEY/STUDY OF EXISTING SOLUTIONS/PRODUCTS

We have done couples of surveys and research for the Causes of liver diseases causes: In the research paper for the causes of liver diseases we come to know about the various parameter or the causes of the liver diseases.

The causes of various types of liver disease vary. Liver disease can be brought on by:

Infections with virus: A viral infection is the root cause of hepatitis A, hepatitis B, and hepatitis C.

Deficiencies in your immune system: Autoimmune liver diseases can arise when your immune system mistakenly attacks your liver. Autoimmune hepatitis and primary biliary cholangitis are two examples.

Diseases inherited: A genetic condition—one that you inherit from your parents—causes some liver issues. Hemochromatosis and Wilson disease are two examples of liver diseases that are inherited.

Cancer: Tumors may form in your liver as abnormal cells multiply. There are benign (noncancerous) and malignant (liver cancer) types of these tumors.

Excessive exposure to toxins: Consumption of alcohol is the cause of fatty liver disease linked to alcohol. Consuming too much fat is the cause of non-alcohol-related fatty liver disease (NAFLD).As diabetes and obesity rates rise, so does the prevalence of NAFLD.

Non-alcohol fatty liver disease is one type of liver disease that rarely causes symptoms. The most common symptom of other conditions is jaundice, which causes the whites of your eyes and skin to turn yellow. When your liver can't get rid of a substance called bilirubin, you get jaundice.

Machine learning models before us: We have done the research on the various machine learning models made before us and all of them have done an excellence work but their liver disease prediction in around 65% - 70%. Also their model is not available for every one as they have not deployed them.

RELATED WORKS

The liver disease was predicted using J48, LMT, Random Tree, Random Forest, REPTree, Decision Stump, and Hoeffding Tree. These algorithms have also been the subject of a comparison study. By measuring their accuracy, precision, recall, mean absolute error, F-measure, kappa statistic[14], and run time, the system evaluates the performance of each algorithm. The analysis revealed that, in comparison to other algorithms, the Decision Stump algorithm performs well and has an accuracy rate of 70.67 percent.

Classification is yet another method that can be used to predict accuracy and differentiate between different kinds of data [19,20,21].The process of grouping similar abstract objects into classes is known as clustering. The process of finding rules that may govern associations and causal objects between a set of items is known as association rule mining. Sindhujaetal conducted a survey of various classification methods for predicting liverdiseases., [22]. Speed, accuracy, performance, and cost were used to compare and evaluate C4.5, Naive Bayes, Decision Tree, SVM, Classification and Regression Tree, and Back Propagation Neural Network algorithms. The C4.5 algorithm was found to be superior to other algorithms.

Based on the comparison, accuracy, and the ROC curve of performance. It is inferred that Arbitrary woodland model showed better execution RN(87.48) when contrasted with other order models could assist specialists with arranging greasy liver patients for early medicines.

Alfisahrin, Sadiyah Noor Novita, and others[24] created a model for the WEKA tool that uses classification algorithms like Decision Tree, Navies Bayes, and NBTree to predict liver disease, taking into account factors like age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, total proteins, albumin as part ate amino transferees, ratio albumin, and globulin, and total bilirubin. Additionally, the Chai-squared ranking method was utilized to evaluate the impact of various characteristics. Execution time is used to evaluate each algorithm's performance. Disarray grid was used to gauge the precision. The experiment demonstrates that the Navies Bayes algorithm has the quickest computation time while NBTree has the highest accuracy. Alice and her companions Using an R tool, [25] developed a model for analyzing various liver disease disorders. Using machine learning methods like the decision tree, support

vector machine, and Naive Bayes algorithm, the datasets are well trained and tested. The decision tree is more accurate than other classification algorithms because Pearson correlation is used to measure each algorithm's accuracy, specificity, and sensitivity.

One of the most difficult aspects of medical data mining is automating disease prediction and diagnosis. Sina Bahramirad, etal.[26] used eleven algorithms, including Logistic, Linear Logistic Regression, Gaussian Processes, Logistic Model Trees, Multilayer Perceptron, K-STAR, RIPPER, Neural Net, Rule Induction, Support Vector Machine, and Classification and Regression Trees, to construct a classification model based on two real liver patient datasets. Using this algorithm, a comparison is made between two kinds of datasets: the AP dataset from Andhra Pradesh in India and the BUPA dataset from California in the United States. The accuracy, precision, and recall of these datasets were compared to see how well they performed. Consequently, the AP dataset outperforms the BUPA dataset in terms of accuracy, but the BUPA dataset is more accurate in terms of precision and recall.

Ashwani Kumar and others In order to select the best classification algorithm, [27] used the info-gain feature selection method in C4.5, Random forest, CART, Random tree, and REP. To improve accuracy, the datasets were divided into two sets of 70-20% and 80-20% of training to testing ratios, respectively. The performance is evaluated by comparison. Consequently, we conclude that Random forest with a training-testing data partition of 6 features at 80-20 percent achieves an accuracy of 79.22 percent.

Guliaetal Anju[28] constructed a hybrid model employing a variety of algorithms, including J48, MLP, SVM, Random Forest, and BayesNet, and compared the various algorithms for improved accuracy. There were three phases to the model. In the first phase, the original dataset is classified using a classification algorithm; in the second phase, features that influence liver disease are chosen; and in the third phase, the results of the original dataset with and without features are compared to one another. The accuracy of the algorithms was measured to assess performance based on the experiments. In light of this, the SVM algorithm is regarded as the best before feature selection. The Random Forest algorithm is regarded as having superior performance to other algorithms following feature selection.

Sanjay Kumar The liver function test attributes of age, gender, DB, Alkphos, total bilirubin, SgptTP, ALB, A/G Ratio, Sgot, and Selector field were taken into consideration with classification algorithms such as Naive-Bayes, Random forest, K-means, C5.0, and K-Nearest Neighbors (KNN) in the development of models using real patient data [29].Prior to the adaptive boosting algorithm's implementation, the Random Forest algorithm provided high accuracy. However, the C5.0 algorithm now provides better accuracy.

III. PROPOSED METHODOLOGY/ SYSTEM ARCHITECTURE

We have divided our Research in 5 step/part to make the work more organized.

- Data Mining / Data Collection And Data Exploration
- Data Visualization
- Data Processing
- Model Training and Testing
- Model Deployment.

DATA COLLECTION

Since we obtained the Indian Liver Patient Dataset (ILPD) from the UCI Machine Learning Repository, this data is secondary. There are 416 records for liver patients and 167 for non-liver patients in this data set. India's north-east part of Andhra Pradesh was where the data set was gathered. The Selector is a class label that is used to divide people into groups (liver patients or not). There are 441 male and 142 female patient records in this data set.
The term "90" refers to any patient who has reached the age of 89 or older.

DATA VISUALIZATION

The use of common graphics, such as charts, plots, infographics, and even animations, to represent data is known as data visualization. In a way that is simple to comprehend, these informational visual displays convey complex data relationships and data-driven insights. Therefore, we utilized the Python libraries for data visualization, Matplotlib and Seaborn, for our dataset's data visualization.
Min Max Scaler To Deal With Outliers
Min-Max scaling is a normalization technique that uses the minimum and maximum values of each feature to scale data within a dataset to a specific range.
The min-max scaler scales the data series using the minimum and maximum values of each column, in contrast to standard scaling, which uses the standard normal distribution (mean = 0 and standard deviation = 1) to scale the data.
Our models' performance may be harmed if the data scales of some features are significantly different from those of others. This is especially true for algorithms like KNN and neural networks that use distance as a measurement.
Additionally, it speeds up convergence and helps optimize gradient descent and other machine learning processes.
Algorithm execution speed and performance can both be improved. Because the data have already been scaled down, the more complicated calculations that are needed to improve algorithms are done faster.
It can also be useful when comparing the performance of various datasets or models.

Numerous machine learning applications, including computer vision, natural language processing, and speech recognition, have utilized the Min-Max scaler, which is implemented in Sklearn libraries.

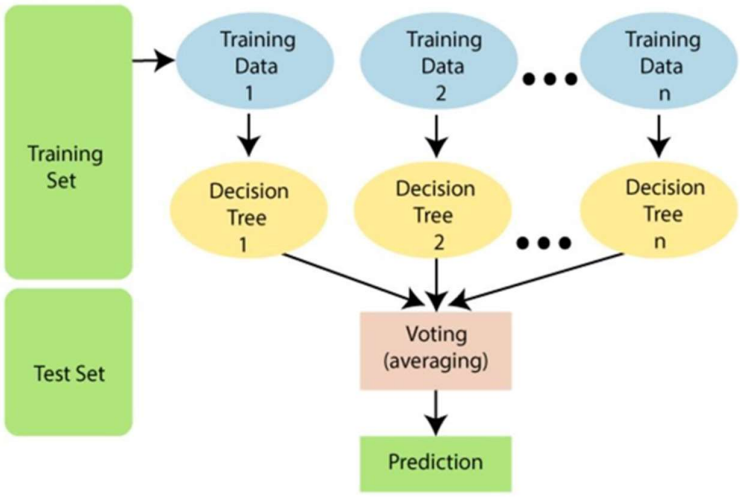
IV. MODEL TRAINING AND TESTING

Random Forest:

The supervised learning method includes the well-known Random Forest machine learning algorithm. In ML, it can be utilized for both regression and classification issues. It is based on the idea of ensemble learning, in which multiple classifiers are combined to solve a complex problem and boost the model's performance.

Random Forest is a classifier that uses the average of a number of decision trees on various subsets of a given dataset to increase that dataset's predictive accuracy.as the name suggests, Random Forest is a classifier. The random forest predicts the final result based on the majority of predictions from each tree rather than relying on a single decision tree.

Accuracy increases and the issue of overfitting is avoided when there are more trees in the forest.



How Random Forest Works

The creation of the random forest by combining N decision trees is the first step in the Random Forest process, followed by the prediction of each tree created in the first phase.

The following steps and diagram provide an illustration of the working procedure: Step-1: From the training set, choose K data points at random.

Step-2: Establish the subsets of decision trees that are associated with the selected data points.

Step-3: For the decision trees you want to build, select the number N.

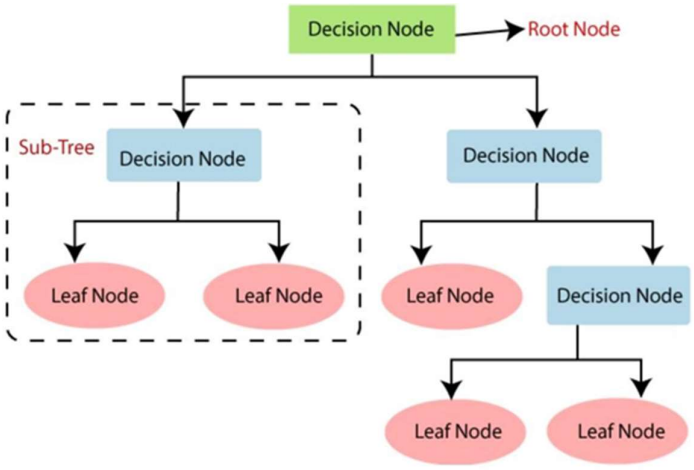
Step-4: Re-do Steps 1 and 2.

Step-5: Find the predictions for each decision tree for new data points, and then put the new data in the category that gets the most votes.

Decision Tree:

Although Decision Tree is a method of supervised learning that can be applied to both classification and regression problems, it is typically used to solve classification problems. It is a classifier with a tree structure, with internal nodes representing a dataset's features, branches representing the decision rules, and each leaf node representing the result.

The Decision Node and the Leaf Node are the two nodes in a Decision tree. Leaf nodes are the results of decisions and do not contain any additional branches, whereas Decision nodes are used to make any decision and have multiple branches.
The features of the given dataset are used to make decisions or conduct tests.
It is a graphical representation that can be used to get all of the possible solutions to a problem or decision based on the conditions that have been given.
Because it begins with the root node and builds a tree-like structure with additional branches, it is referred to as a decision tree.
We employ the CART algorithm, which stands for Classification and Regression Tree algorithm, to construct a tree.
A decision tree simply asks a question and divides the tree further into subtrees based on the answer (yes/no).



How Decision Tree Works:
The algorithm in a decision tree begins its prediction of the dataset's class from the root node. Based on the comparison, this algorithm follows the branch and moves on to the next node. It does this by comparing the values of the root attribute with those of the record (the actual dataset).
The algorithm continues by comparing the attribute value to that of the other sub-nodes for the subsequent node. It continues doing so until it reaches the tree's leaf node. The following algorithm can help clarify the entire procedure:
Step-1: Start the tree with the root hub, says S, which contains the total dataset.
Step-2: Using the Attribute Selection Measure (ASM), select the best attribute from the dataset.
Step-3: Partition the S into subsets that contains potential qualities for the best credits.
Step-4: Create the best attribute-containing decision tree node.
Step-5: Utilizing the subsets of the dataset created in step 3, construct new decision trees in a recursive fashion. Keep going through this procedure until you reach a point where you are unable to further categorize the nodes and identify the final node as a leaf node.
SDG Classifier:
A random probability-linked system or procedure is referred to as stochastic. As a result, rather than selecting the entire data set for each iteration in Stochastic Gradient Descent, a small number of samples are chosen at random. The total number of samples from a dataset that are used to calculate the gradient for each iteration is referred to as a "batch" in Gradient Descent.

Similar to Batch Gradient Descent, the batch is regarded as the entire dataset in typical gradient descent optimization. Even though it is really helpful to use the entire dataset to get to the minima in a less noisy and less random way, when our dataset gets big, the problem comes up. If your dataset contains one million samples, a standard Gradient Descent optimization method will require you to use all one million samples to complete one iteration of the gradient descent, and this must be done for each iteration until the minima are reached. As a result, its execution becomes extremely costly in terms of computing.

Stochastic Gradient Descent provides a solution to this issue. Each iteration in SGD only requires a single sample, or a batch size of one. The sample is shuffled at random before being chosen for the iteration.

Logistic Regression Algorithm:

One of the most widely used Machine Learning algorithms is logistic regression, which is a Supervised Learning technique. Using a particular set of independent variables, it is used to predict the categorical dependent variable.

The outcome of a categorical dependent variable is predicted using logistic regression. As a result, the outcome must be a discrete or categorical value. It can be true or false, yes or no, 0 or 1, etc. However, rather than specifying a precise value between 0 and 1, it specifies probabilistic values between 0 and 1.

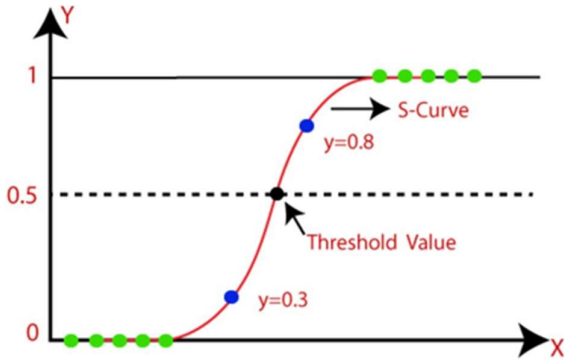
Except for how they are used, Logistic Regression and Linear Regression are very similar. Logistic regression is used to solve classification problems, while linear regression is used to solve regression problems.

In logistic regression, we fit an "S"-shaped logistic function that predicts two maximum values (0 or 1) rather than a regression line.

The likelihood of something, such as whether a cell is cancerous or not based on its weight, whether a mouse is obese, etc., is shown by the logistic function curve.

Because it can classify new data based on continuous and discrete datasets and provide probabilities, logistic regression is a significant machine learning algorithm.

Using various types of data, Logistic Regression can easily classify the observations and identify the most effective classification variables. The logistic function is depicted in the picture below:



The equation for linear regression can be used to calculate the logistic regression equation. The following are the mathematical procedures for obtaining equations for logistic regression:

We are aware that the straight-line equation can be written as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

Since y can only be between 0 and 1 in Logistic Regression, let's divide the previous equation by (1-y):

$$\frac{y}{1-y}\text{; 0 for y= 0, and infinity for y=1}$$

However, we require a range of -[infinity] to +[infinity], and if we take the logarithm of the equation, it will be as follows:

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

The final equation for logistic regression is the one above.

XG Boost:

The XGBoost Python library is used to implement gradient boosted decision trees, which are intended for speed and execution—the most crucial aspects of machine learning (ML).

XgBoost:

Researchers at the University of Washington introduced the Python XgBoost (Extreme Gradient Boosting) library. It is a C++ module for Python that trains Gradient Boosting algorithms to help ML models.

Gradient boosting:

This is an AI technique that is used, among other things, in assignments like classification and regression. It provides an expectation model in the form of a collection of weak forecast models, also known as decision trees.

What are the workings of Fundamental Gradient Boosting?

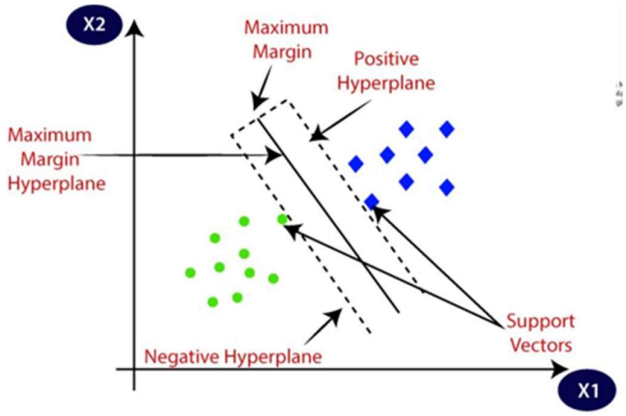
A loss function should be improved, which means reducing it more than the result. To make assumptions, powerless students are utilized in the model

Choice trees are used in this, and they are used in a desirous way, which implies picking the best-partitioned shines considering Gini Pollution, etc or to restrict the misfortune capability

The added substance model is used to assemble all of the delicate models, restricting the misfortune capability. Each tree is added to ensure that the decision tree's existing trees do not change. The best hyper boundaries are frequently identified using the angle plummet method, after which loads are further refreshed.

Using XgBoost We got accuracy of 70%.

SVM



Support Vector Machine, also known as SVM, is one of the most widely used supervised learning algorithms. It can be used to solve regression and classification problems. However, in Machine Learning, it is primarily utilized for Classification issues.

The SVM algorithm's objective is to find the most effective line or decision boundary for classifying n- dimensional space, allowing us to quickly place a new data point in the appropriate category in the future. A hyperplane is the name given to this best decision boundary.

The extreme points and vectors that aid in the creation of the hyperplane are selected by SVM. The algorithm is referred to as a Support Vector Machine because these extreme cases are referred to as support vectors. Take a look at the diagram Above, which shows two distinct categories that are separated by a decision boundary or hyperplane

V. HARDWARE / SOFTWARE SPECIFICATION

Hardware Development –

- There No Special Hardware Required,

Just A i5 generation Laptop With Chrome Brower. Software Development –

IDE:

- Jupyter / Sypder / Colab Data Exploration:
- Numpy
- Pandas

Data Visualization:

- Matplotlib
- Seaborn Data Processing:
- Sklearn.Preprocessing
- Imblearn

Model Training And Testing:

- Sklearn

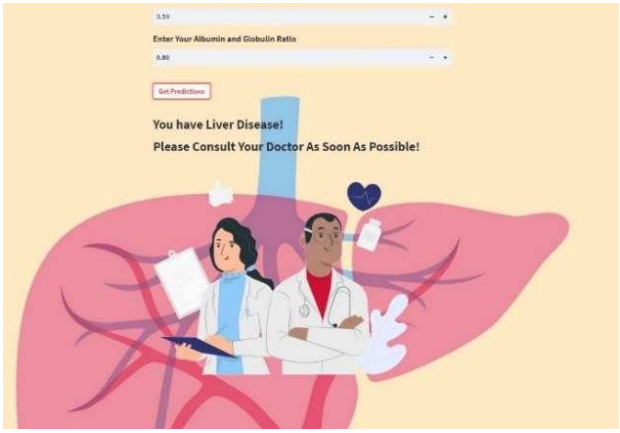
Model Deployment:

- Heruko
- Streamlit

VI. IMPLEMENTATION / RESULT

We Had Divided Our Research Into 5 Parts Where In Data Exploration In Done Using Numpy And Pandas, We Have Also Implemented Graphs For Better Visualization Of Data Using

Matplotlib And Seaborn After This We Done Data Processing Where Dealing With Duplicate Values, Null Values, Imbalance Data And Scaling Data Was The Main Task And Was Complete Using Sklearn, Imblearn And Scaling Was Done Using Min Max Scaler With Range Of -1 to 1, After That We Have Split The Data And Trained Our Machine Learning Model Using Sklearn And The Algorithm Used Is SVM Which Gives Accuracy Of 85%. And At The Last We Have Deployed Our Model Using Heruko For Easy Access To Anyone In Need.



VII. CONCLUSION

Despite the fact that people are participating in dance and yoga classes and becoming more health conscious in today's society; The issue will continue for a long time due to the sedentary lifestyle and luxuries that are constantly introduced and improved. Therefore, in such a scenario, our Research will greatly benefit society.

We will attempt to increase the machine learning model's prediction rate and accuracy with the dataset we will use for this Research . Since nearly everyone over the age of 12 now has a smartphone, we can incorporate these solutions into an iOS or Android app.

Additionally, it can be incorporated into a website, and both the app and the website will benefit a large portion of society greatly.

VIII. REFERENCES

[1] W. Yu, T. Liu, R. Valdez, M. Gwinn, and M. J. Khoury, ‘‘Application of support vector Machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes,’’ BMC Med. Inform. Decis. Making, vol. 10, no. 1, p. 16, 2010.

[2] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, ‘‘Predicting diabetes mellitus with machine learning techniques,’’ Frontiers Genet., vol. 9, p. 515, Nov. 2018.

[3] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, and H. Benali, ‘‘Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI,’’ Neuroradiology, vol. 51, no. 2, pp. 73–83, Feb. 2009.

[4] I. S. F. Dessai, ‘‘Intelligent heart disease prediction system using probabilistic neural network,’’ Int. J. Adv. Comput. Theory Eng., vol. 2, no. 3, pp. 2319–2526, 2013.

[5] Z.-H. Zhou, Y. Jiang, Y.-B. Yang, and S.-F. Chen, ‘‘Lung cancer cell identification based on artificial neural network ensembles,’’ Artif. Intell. Med., vol. 24, no. 1, pp. 25–36, 2002.

[6] B. E. Bejnordi et al., ‘‘Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,’’ JAMA, vol. 318, no. 22, pp. 2199–2210, Dec. 2017

[7] Y. Cao, Z.-D. Hu, X.-F. Liu, A.-M. Deng, and C.-J. Hu, ‘‘An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters,’’ Disease Markers, vol. 35, no. 6, pp. 653– 660, 2013.

[8] W. S. McCulloch and W. Pitts, ‘‘A logical calculus of the ideas immanent in nervous activity,’’ Bull. Math. Biophys., vol. 5, no. 4, pp. 115–133, 1943.

[9] K. Hornik, M. Stinchcombe, and H. White, ‘‘Multilayer feedforward networks are universal approximators,’’ Neural Netw., vol. 2, no. 5, pp. 359–366, 1989.

[10] G. E. Hinton, S. Osindero, and Y.-W. Teh, ‘‘A fast learning algorithm for deep belief nets,’’ Neural Comput., vol. 18, no. 7, pp. 1527–1554, 2006.

[11] A. L. Samuel, ‘‘Some studies in machine learning using the game of checkers,’’ IBM J. Res. Develop., vol. 3, no. 3, pp. 29–210, 1959.

[12] K. Tretyakov, ‘‘Machine learning techniques in spam filtering,’’ Data Mining Problem-Oriented Seminar, MTAT, vol. 3, no. 177, pp. 60–79, 2004.

[13] D. West and V. West, ‘‘Improving diagnostic accuracy using a hierarchical neural network to model decision subtasks,’’ Int. J. Med. Inform., vol. 57, no. 1, pp. 41–55, 2000.

[14] P. Lakhani and B. Sundaram, ‘‘Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks,’’ Radiology, vol. 284, no. 2, pp. 574– 582, Aug. 2017.

[15] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, ‘‘A deep learning approach for Parkinson’s disease diagnosis from EEG signals,’’ Neural Comput. Appl., Aug. 2018. doi: 10.1007/s00521-018-3689-5.

[16] D. S. Kermany et al., ‘‘Identifying medical diagnoses and treatable diseases by image-based deep learning,’’ Cell, vol. 172, no. 5, pp. 1122–1131, 2018.

[17] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, ‘‘Image analysis and machine learning for detecting malaria,’’ Transl. Res., vol. 194, pp. 36–55, Apr. 2018.

[18] Nazmun Nahar and Ferdous Ara, "Liver disease prediction by using different decision tree techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP),Vol.8, No.2, March 2018.

[19] Divya, B, Kalaiselvi, R, Review on Confidentiality of the Outsourced Data, Research Journal of Science and Engineering Systems, vol.1, pp.1-7, 2017.

[20] Rule optimization of boosted c5 is the work of M. Hassoon, M. S. Kouhi, M. Zomorodi-Moghadam, and
M. Abdar.0 characterization involving hereditary calculation for liver disease prediction. The Proceedings of the 2017 International Conference on Computer and Applications (ICCA),299-305, IEEE, 2017.

[21] D.Sindhuja, R. Jemina Priyadarsini," A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder", International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016.

[22] "Prediction of fatty liver disease using machine learning algorithms," Chieh-Chen Wu, Wen-hun Yeh, Wen-Ding Hsu, Md. Mohaimenul Islam, Phung Anh (Alex) Nguyen, Tahmin

Nasrin Posly, and Yao-Chin Wang, Computer Methods and Programs in Biomedicine, Vol. 170, Pages 23-29, March 2019.

[23] Sadiyah Noor Novita Alfisahrin, Teddy Mantoro,"Data Mining Techniques For Optimatization of Liver Disease", International Conference on Advanced Computer Science Applications and Technologies, 2013.

[24] L. Alice Auxilia," Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease", 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018)IEEE Conference, IEEE Xplore, ISBN:978-1-5386-3570-4.

[25] Sina Bahramirad, Aida Mustapha, Maryam Eshraghi, "Classification of Liver Disease Diagnosis: A Comparative Study", 2013 Second International Conference on Informatics & Applications (ICIA),Lodz, Poland, IEEE Explore,2013.

[26] Ashwani Kumar, Neelam Sahu,"Categorization of liver disease using classification techniques", International Journal for Research in Applied Science & Engineering Technology (IJRASET),Vol. 5, No.5, 2017.

[27] Anju Gulia, Rajan Vohra, Praveen Rani, "Liver Patient Classification Using Intelligent Techniques", International Journal of Computer Science and Information Technologies, Vol. 5, No. 4, 2014.

[28] “Effective Analysis and Diagnosis of Liver Disorder by Data Mining," by Sanjay Kumar and Sarvak Katyal, was published in the Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018).