International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-08, Issue-01, APR 2022

Survey on Machine Learning Techniques for The Diagnosis of Liver Disease

[1]Prof. Satish Manje, [2]Miss.Tannaz Inamdar, [3]Miss.Shivani Kamalekar, [4]Miss. Akanksha Shaha.

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India. [1]satishmanje93@gmail.com, [2]tannazinamdar50@gmail.com, [3]shivanikamalekar@gmail.com, [4] shahaakanksha15@gmail.com

Abstract- Machine Learning Algorithm focuses on building the model using sample data which is known as training of the data for making decisions or predictions. The project focuses to give a survey and providing a comparative survey of the entire ML techniques for diagnosing and predicting liver disease in Medical Areas, which are been already used for predicting liver disease by several authors, and analyses are based on Sensitivity, Accuracy, Precision, and Specificity. The project focuses on giving the basic idea of past published papers for detection and liver disease diagnosis based on different ML Algorithms such as Decision tree, and ANN which provide better Accuracy on detecting and predicting liver disease. Different Algorithms has different performance which is based on different scenario but most importantly the feature selection and dataset is also important for getting better prediction results.

Keywords – Machine Learning, Diagnosis, Liver Disease.

I. INTRODUCTION

Artificial Intelligence have several branches which includes expert systems, fuzzy systems, machine learning, etc. Machine Learning is of great use in the areas of engineering. Similar to expert systems, fuzzy systems also store experts knowledge and uses it in their system for processing the input and generating outputs. IN, a Fuzzy system is used to control the two state variables using some membership functions which are defined by experts. IN a Fuzzy Cluster Means (FCM) method for diagnosing Liver Disease which is the global health problem, is presented. FCM performs an important role for classification, and evaluation. The Liver is a organ which plays an vital role in all human being and presently there is no way

For restoring the lack of liver function. Liver disease patient are rapidly increasing due to the consumption of alcohol, intake of contaminated food and drugs that is widespread global. The opportunity of using the data science and technology for personalized health care.

Machine learning uses AI for generating predictive models more effectively and efficiently than conventional methods through detection of hidden patterns within the large data sets. Naive Bayes classifiers predicts the class labels by computing the likelihood of the observed features under each class, returning the class with maximum likelihood. K-nearest determines output based on the value of the classes of K-nearest training samples.

II. AIMS AND OBJECTIVE

a) Aim

The motive of project is to improve the liver disease diagnosis using ML Approaches. The project also focuses to compare the various classification algorithms depending upon the performance factors. The objective of the research is to use classification algorithms for identifying the patients having liver disease from healthy individuals.

b) Objective

The objective of this paper is to use various algorithms to identify the liver patients from healthy ones. In this learning, four different classification algorithms Logistic Regression, SVM, KNN, and ANN, have been taken under the consideration for comparing their performance based on the liver patient data. Detection of liver problems at early stage will increase survival rate of the patient.

III. LITERATURE SURVEY

Diagnosis of liver disorder is actually based on previous history and actual examination. Medical history, drug or alcohol intake, history of liver disease needs to be evaluated in detail for liver disease detection.

Paper 1: A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis:

This paper evaluates the methods for the classification of liver patient datasets. The classification algorithms considered here are the Naïve Bayes classifier, C4.5, Backpropagation, Neural Network algorithm, and Support Vector Machine. Liver disorder can be diagnosed by analyzing enzymes level in blood.

Paper 2: Prediction of Different Types Of Liver Diseases Using Rule-Based Classification Model:

The main focus of this paper is to present a classification model with ML techniques for prognosis of various types of

 DOI : 10.35291/2454-9150.2022.0174

Liver disorders. SVM, Rule Induction, Decision Tree, Naive Bayes, and Artificial Neural Network data mining methods with the K-cross fold method are utilized with the proposed methodology for liver disease prediction.

Paper 3: A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder:

This paper describes study on classification methods in data mining to analyze liver disorder. Data mining is extraction method that is systematic and gives meaningful information from the large database. This survey aims to assess different ML techniques in divination of advanced fibrosis by blending the serum bio-markers and medical information to create the classification models.

IV. EXISTINGSYSTEM

In the existing system of the medical system for treatment of liver disease has been useful to the community, thus easy detection and the disease can be easily predicted with the use of the expert methodology. With the rapid improvement in the area of Artificial intelligence, different types of ML algorithm have been developed; this will help to enhance the quality and accuracy of the detection or prediction of the liver disease. LIMITATIONS OF EXISTING SYSTEM: Most of the people use the term expert system because of compactness, while there may be no experience and expertise in an expert system, and they can only include general knowledge. Medical expert systems have attribute that differentiate them from other medical software, including that these systems imitate the arguments of an expert physician to achieve medical results. Expert systems require a huge number of medical science laws and regulations in the sector of disease and conditions so that they can provide accurate results.

V. COMPARTIVE STUDY

SR NO.	PAPER TITLE	AUTHOR NAME	METHOD	ADVANTAGE	DISADVANTAGE
1.	A Critical Study Of Selected Classification Algorithms For Liver Disease Diagnosis	Ramana,Bendi Venkata,M. SurendraPrasad Babu, and N.B. Venkateswarlu	Naïve Bayes classifier, C4.5, Back propagation Neural Network algorithm, and Support Vector Machines	Simple to implement.	Scarcity of automatic classification algorithm.
2.	Prediction Of Different Types Of Liver Diseases Using Rule Based Classification Model	Kumar,Yugaland G. Sahoo	Support Vector Machine (SVM), Rule Induction (RI), Naive Bayes (NB) and Artificial Neural Network (ANN)	Good Approach Explained	Difficult to adapt.
3.	Prediction Of Liver Fibrosis Stages by Machine Learning Model	Ayeldeen, Heba, Olfat Shaker,Ghad a Ayeldeen, and haled M. Anwar	Decision Tree (DT)	Good Approach Explained	Time Consuming

Table 1 : Comparative Analysis

VI. PROBLEM STATEMENT

To create a project using ML techniques for the diagnosis of liver diseases using various algorithms such as SVM, Naive Bayes, ANN, K-Means Algorithm. Liver disease detection at a early stage is very crucial for better treatment. It is a challenging work for medical expertise to predict the disease in the prior stages owing to basic precise symptoms. Often this indication becomes apparent when it's too late. For overcoming this issue, we need to improve liver disease diagnosis using machine learning approaches.

VII. PROPOSED SYSTEM

Detection of liver disease in early stages is very essential and crucial because it will help in early treatment and recovery of the disease. And it is very challenging to detect in early stages of the disease with great accuracy recovery of the disease. Direct Bilirubin, Total Bilirubin, Alkaline Phosphatase, Alanine Aminotransferase, Aspartate Aminotransferase, Total Proteins, Albumin, Albumin and Globulin Ratio attributes will take for the Machine learning algorithms. We will train this attribute to our models and finding accuracy. ML is a part of Artificial Intelligence, which help the computer to think like human and can take their own decision without human intervention. Due to rapidly development in Artificial Intelligent, Machine learning has lots of advancement in diagnosis of difference types of disease. Moreover, Machine learning algorithm gives us more accurate prediction and performance.

VIII. ALGORITHM

Step 1 : Start
Step 2 : Loading data set
path ← settings.MEDIA_ROOT +
"\\"+"Indian_Liver_Patients_Dataset.csv"
df ← pd.read_csv(path)
Step 3: Decision Tree classifier implementation
dt ← DecisionTreeClassifier()
dt.fit(self.X_train, self.y_train)
y_pred ← dt.predict(self.X_test)
dt_acc ←accuracy_score(self.y_test, y_pred)
dt_precc ←precision_score(self.y_test)

 DOI : 10.35291/2454-9150.2022.0174

dt_recall ← recall_score(self.y_test, y_pred)
dt_f1Scr qq ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)

Step 4: K-neighbours classifier

knn ← KNeighborsClassifier()
knn.fit(self.X_train, self.y_train)
y_pred ← knn.predict(self.X_test)
knn_acc←accuracy_score(self.y_test,y_pre)
knn_prec←precision_score(self.y ,y_pre)
knn_recall←recall_score(self.y_test,y_pred) knn_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)

step 5 : Logistic regresion

reg ← LogisticRegression()
reg.fit(self.X_train,self.y_train)
y_pred ← reg.predict(self.X_test)
lg_acc ← accuracy_score(self.y_test,y_pred)
lg_precc←precision_score(self.y_test,y)
lg_recall ← recall_score(self.y_test,y_pred)

Step 6 : SVM Implementation

cls ← SVC()
cls.fit(self.X_train, self.y_train)
y_pred ← cls.predict(self.X_test)
svm_acc←accuracy_score(self.y_test,y)
svm_precc←precision_score(self.y_test,y)
svm_recall←recall_score(self.y_test,y_pred)
svm_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)

step 7: Naïve bayes

nb ← GaussianNB()
nb.fit(self.X_train, self.y_train)
y_pred ← nb.predict(self.X_test)
nb_acc ←accuracy_score(self.y_test, y)
nb_precc ←precision_score(self.y_test,y)
nb_recall ← recallscore(self.y_test, y_pred)
nb_f1Scr ← f1_score(self.y_test, y_pred)
cm ← confusion_matrix(self.y_test, y_pred)

Step 8 : Stop

IX. MATHEMATICAL MODEL

1. Support Vector Machine (SVM):

Supervised Learning calculation can be utilized for both Grouping and Relapse issues. For Classification type 1 of SVM, training includes the minimization of the error function:

$$0.5 + \sum \zeta$$

In conflict with SVM, Type 1 model, Type 2 model minimizes the error function:

$$0.5 - vp + w - vp + \sum.$$

2. Mean Absolute Error (MAE):

MAE is used for model valuation based on regression models. The MAE of technique or model regarding an assessment set is the mean of the absolute estimation of the discrete desire error on inclusive events in the assessment set that is the distinction between the predicted error and the true error for overall events. MAE is calculated as follow:

$$MAE = \frac{\sum_{i=1}^{n} abs\left(y_i - \lambda(x_i)\right)}{n}$$

3. Relative Absolute Error (RAE)

RAE is moreover the same as the relative squared error, it is comparative with a simple predictor, which is only the mean of the literal values. For this situation, however, the error is only the total absolute error rather than the absolute squared error. The RAE of a single instance i can be calculated using the following equation:

$$RAE = \frac{\Sigma_{j=1}^{n}\left|P_{(ij)} - T_j\right|}{\Sigma_{j=1}^{n}\left|T_j - \bar{T}\right|}$$

4. Accuracy

Accuracy: A classifier's accuracy is described as the proportion of the test set tuples correctly categorized by the classifier.

$$Accuracy = \frac{no.of\ TP + no.of\ TN}{no.of\ TP + FP + FN + TN}$$

5. Sensitivity

True Positive Rate or the fraction of correctly detected positive tuples, is another name for sensitivity.

$$Sensitivity = \frac{no.of\ TP}{no.of\ TP + no.of\ FN}$$

6. Precision

Precision: The ratio of true positives to all positive outcomes (including True Positives and False Positives) is known as accuracy

$$Precision = \frac{no.of\ TP}{no.of\ TP + no.of\ FP}$$

7. Specificity

Specificity: The True Negative Rate or the fraction of correctly detected negative tuples, is a measure of specificity.

$$Specificity = \frac{no.of\ TN}{no.of\ TN + no.of\ FP}$$

X. SYSTEM ARCHITECTURE

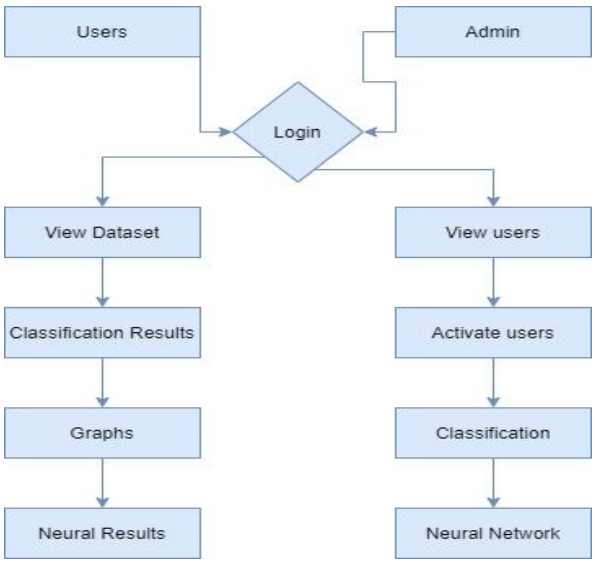

Fig.1: System Architecture

 DOI : 10.35291/2454-9150.2022.0174

Description:

There are 2 modules

1) User: The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the customer. Once admin activated the customer then user can login into system. In the neural result ANN and backpropagation algorithm accuracy will be displayed.

2) Admin: Admin can login with his credentials. Once he login he can activate the users. The activated user only login in applications. For the classification 80% data will be consider as training data and 20% of data for testing purpose.

XI. ADVANATGES

1) Computer-Aided Diagnosis system is a rapidly developing sector of research in the medical industry.

2) Provides a path toward achieving Artificial General Intelligence someday in the future.

3) As we achieve accuracy, it will increase patients' survival rates.

XII. DESIGN DETAILS

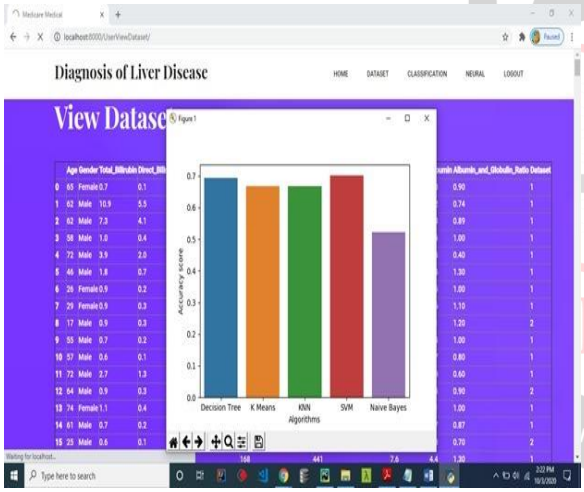

Fig 2: Result

Both SVM and Bayes net provide the accuracy of 83.8%. 81.5% accuracy. By applying the test on 7 best selected features, Bayes Net attained 84.5% of correctness, SVM provides 85.1% accuracy and FT classify 84.5% correctly. Naive Bayes offers 86.419% of accuracy. Naive bayes provides 82.31% accuracy. J48 gives 84.35% of correctness. SVM provide the highest accuracy of 94.60.

XIII. CONCLUSION

Thus, we have tried to implement the paper "Ramana, Bendi Venkata, M. Surendra Prasad Babu and N. B. Venkateswarlu", "*A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis"*, IEEE 2011 and according to the implementation the conclusion is, it has clearly found and observed that various Machine Learning algorithm such as Decision tree, J48 and ANN provides better accuracy in prediction and detection of liver disease. With this survey, it is found that the accuracy and performance can be enhanced by using different combination or hybrid machine learning algorithm, and in the future, can work on more parameters which will help to get better performance than the existing technique. Hence the above project implemented is basically survey of ML techniques for diagnosis of liver disease.

REFERENCE

[1] Ramana, Bendi Venkata, M. Surendra Prasad Babu, and N. B. Venkateswarlu. "A critical study of selected classification algorithms for liver disease diagnosis." International Journal of Database Management Systems 3.2 (2011): 101-114.

[2] Jacob, Joel, Joseph Chakkalakal Mathew, J. Mathew, and E. Issac. "Diagnosis of liver disease using machine learning techniques."Int Res J Eng Technol 5, no. 04(2018)

[3] A.S.Aneeshkumar and C.Jothi Venkateswaran, "Estimating the Surveillance of Liver Disorder using Classification Algorithms", International Journal of Computer Applications (095-8887), Volume 57-No.6, November 2012.

[4] Sontakke, S., Lohokare, J., & Dani, R. (2017, February). Diagnosis of liver diseases using machine learning. In 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI) (pp. 129-133). IEEE.

[5] Kumar, Yugal, and G. Sahoo. "Prediction of different types of liver diseases using rule-based classification model." Technology and Health Care 21, no. 5 (2013): 417-432.

[6] Sindhuja, D., and R. Jemina Priyadarsini. "A survey on classification techniques in data mining for analyzing liver disease disorder." International Journal of Computer Science and Mobile Computing 5.5 (2016): 483-488.

[7] Durai, Vasan, Suyan Ramesh, and Dinesh Kalthireddy. "Liver disease prediction using machine learning." (2019).

[8] D.A. Saleh F. Shebl M. Abdel-Hamid et al. "Incidence and risk factors for hepatitis C infection in a cohort of women in rural Egypt"Trans. R. Soc. Trop. Med. Hyg. vol. 102 pp. 921928 2008.

 DOI : 10.35291/2454-9150.2022.0174