International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 2558–2563

Investigation On Performance Analysis Of Available Machine Learning Techniques For Liver Diseases Predictions

[1]Aishwarya R

Research Scholar, Department of CSE,

Annamalai University, Chidambaram, Tamil Nadu, India

aishwaryar.rajendran@gmail.com

[2]R. Bhargavi

Assistant Professor, Department of CSE,

Sridevi Women's Engineering College, Hyderabad, Telangana, India

bhargavi.rudraraju@gmail.com

Abstract

Diagnosing the types of liver diseases in clinical is very difficult. Based on the laboratory test analyzing different types of liver disease is very difficult. To make it simple we have ruled based classification model to classify and predict liver disease. Rule based model and machine learning combination is used to predict liver disease.

Keywords— liver diseases, classification model.

INTRODUCTION

World Health Organization (WHO) reports that liver fault is the major problem amongst Humans globally. Liver diseases is major cause for the death among women in the world. Usually, liver diseases can be easily detected if specific symptoms appear. However, many women suffer from liver diseases have no symptoms. Hence, regular liver diseases screening is very important for early detection.

Early detection of liver diseases aids for early diagnosis and treatment, because the prognosis plays major role for long time survival. Since early detection, diagnosis, and treatment of liver diseases can reduce the risk of death, it plays a significant role in saving the life of the patient. Any delay in finding and detecting liver diseases in early stages leads to disease progression and complication of treatment, therefore long time waiting before diagnosis of liver diseases and starting the treatment process is of screening is a major concern. Previous studies on the investigation of the consequences of a late diagnosis of liver diseases confirm that it is strongly associated with progression of the disease to more advanced stages, consequently less chance to save the patient's life.

- A major review is conducted by Prof MA Richards et al. for analysis of 87 studies strongly concluded that female patients with b liver diseases who start their therapy less than 3 months after the appearance of symptoms significantly have a higher chance of survival compare to those people waiting for more than 3 months. There were many discussions from the previous studies that conformed–that early the detection more the chance for survival, because it prevents the spreading of malignant cells throughout the entire body.

- The main focus in paper reveals the role of early detection. This is done using machine learning techniques and artificial intelligence which is applied in detecting, diagnosis and empower accuracy.For example, deciding on the biopsy results for detecting Liver Diseases if

ISSN: 2233-7857 IJFGCN

the patient needs surgery or not. Currently, Mammograms test are used which shows abnormal cells that leads to unnecessary biopsies and surgeries.

- Sometimes surgery is done to remove lesions reveals that it is benign which is not harmful. This makes unnecessary painful and expensive surgery. ML Algorithms were introduced with many features such as effective performance on healthcare related dataset which involve images, x-rays, blood samples, etc. Few methods are appropriate only for small datasets whereby others are suitable for huge datasets. However, noise can be a problematic concern in some methods .

LITERATURE SURVEY

Support vector machine (SVM) evaluates chronic hepatitis along with liver fibrosis.-Resource Allocation Mechanism with a classifier was proposed with evolution by comparing classifiers based on accuracy, time and resource. Kemal Polat proposed resource allocation mechanism with AIRS (Artificial Immune recognition system) for prediction of liver disorders. This approach is called Fuzzy-AIRS which uses as classifiers. The accuracy of classification is evaluated by comparing classifiers based on accuracy, number of resource and time. ANN was proposed by [2] to predict Liver cirrhosis and other related liver diseases. Susceptibility of liver using machine learning and decision to predict hepatitis is proposed to evaluated liver fibrosis with hepatitis using support vector machine (SVM)[4]. Latter decision tree along with machine learning is proposed for detecting liver disease [4].Resource allocation mechanism is proposed based on the classifiers. This method is used for predicting the susceptibility of liver diseases such as hepatitis, cirrhosis. Nucleotide polymorphism (SNP) data. They also used to identify a set of SNPs which were relevant to liver disease. Evaluation by comparing them with reported classifier's accuracy, time and resources is proposed [5].Resource allocation mechanism of AIRS was changed with a new one decided Fuzzy-Logic. This approach called as Fuzzy- AIRS was used as a classifier in the diagnosis of Liver Disorders.

Optimizing the classification accuracy is proposed and used when analysing some medical datasets [3]. Work done by new meta-heuristic approach, called the Homogeneity-Based Algorithm (or HBA). Approach used to predict error rates and associated penalty costs. Medical applications as the implications of having a false-positive and false-negative case may be tremendously.

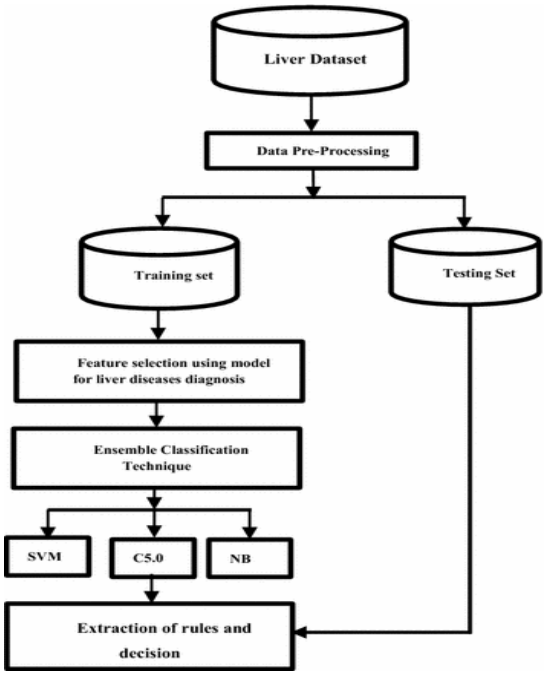
SYSTEM STUDY

In Supervised learning, where dataset is labeled on both input and output parameters which helps in both training and validation.

Bagging Aggregation:

Bagging algorithm, bootstrap aggregation is used for reduction of high variance. An algorithm which has high variance are decision trees, which is similar to classification and regression tree(CART). Decision trees classify datasets based on training data which is feed.- When bagging algorithm is used with decision tree, efficiency of individual trees are grown deep. This decision tree has high variance and low bias. Parameters which need to be checked during bagging decision tree is number of samples. These samples can be chosen by increasing the number of samples in the dataset. Accuracy improves as we use the tree number of times. This takes long time but does not over fit.

ISSN: 2233-7857 IJFGCN



Dataset Selection:
Data selection is a basic steps in machine learning algorithm. Since selected tool makes the values for instances. If the instance is huge in number then the performance becomes less [1]. In our paper the data sets includes ten attribute parameters. These data are collected from kaggle.

Exploration &Pre-processing the Dataset:
The first stage of evaluation of an algorithm in machine learning process is data exploration. The dataset can be selected in many file formats .The file format is called CSV is used in this paper. Preprocessing tools are used for data discretization, normalization, resampling and selection of attributes. Preprocessing can also be called as filtering tool.

All the attributes undergoes data discretization. Data discretization groups the distinct values and then converts all the distinct attribute values into nominal values from the integer values. After this step of converting numeric into nominal values filters are also performed.

S.No	ATTRIBUTE NAME	ATTRIBUTE TYPE
1.	Age	Numeric
2.	Sex	Nominal
3.	Total Bilirubin	Numeric
4.	Direct Bilirubin	Numeric
5.	Alkaline Phosphatase	Numeric
6.	Alamine Phosphatase	Numeric
7.	Total Proteins	Numeric

ISSN: 2233-7857 IJFGCN

8.	Albumin	Numeric
9.	Albumin and Globulin Ratio	Numeric
10.	Result	Numeric (1,2)

Table 1 Dataset Description

Classification techniques:

1. SVM

Support vector machine finds optimal hyperplanes which separates different data into different classes. Hyperplanes are constructed using high dimensional space. Separation is achieved by hyperplane, larger the distance then nearer the training point is present. This means larger the margin lower the generalization error of classifier.

2. Logistic regression

Logistic regression is a simpler classification model based on the parameter interpretation when experiment want to look at relationship within variables. Parametric model can be described using vector parameter (1, 0) .Entire model can be recreated using a single known model. In logistic regression parametric model parameters are coefficient to the predictor variable.

3. K-NN

K-NN algorithm will search through the dataset which is used for training to find the most similar k instances. Similarity is measured using Euclidean distance then fetches data based on the type of data in real dataset. KNN algorithm is based on the instances, competitive learning and lazy algorithm. Lazy algorithm means the model uses data instances for predictive decision.

4. Artificial neural network:

ANN, backpropagation neural network is used were we use ten input neuron in the input layer and this input represents the number of attributes present in the datasets.

Rectified activation function is used in the input layer. The output layer contains single sigmoid activation function. In order to obtain a required recognition rate that is capable enough to diagnose the liver disorder in a patient. Certain parameters are needed in the neural network models to produce the required optimum result. These parameters are the learning, momentum rates and the hidden neurons. All these parameters are present in backpropagation neural networks. The learning rate is the learning power of the system the momentum rate determines the learning speed of the system. The number of hidden neurons in the network has to be varied to produce the optimal result.

The neurons needed at the hidden layer are experimenting in order to determine neurons which are the best that can represent the features in the input dataset accurately to produce the required optimum result. The different neurons are used based on varying experiment in hidden layer. The sigmoid function was used in the output layer because of its soft switching ability and simplicity in derivatives. The implementation is done in neural network using keras package, runs using tensor flow backend in python.

Results and Evaluation

Our objective in this project was to predict liver disease. We predicted using Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbour (K-NN) and Neural Network. All of

ISSN: 2233-7857 IJFGCN

them predicted with better results. By using all these algorithms, we have detected Accuracy, Precision and Specificity which can be defined as follows

Accuracy: The percentage of the experimental result rows that are correctly classified by the classifier.

Precision: The ratio of test result that correctly indicates the presence of a condition against all result.

Specificity: The ratio of test result that not correctly indicates the presence of a condition each of the classification algorithm is summarized in the table shown below.

Algorithm	Overall Result	Accuracy	Precision	Specificity
Logistic Regression	73.23	78.57	88	30.62
K-NN	72.05	80.98	83.74	44.04
SVM	75.04	77.09	79	71.11
ANN	92.8	93.78	97.23	83

Table 2. Results and Evaluation

CONCLUSION

In this project, we have proposed methods for diagnosing liver disease for patients using machine learning techniques. We include SVM, Bagging's algorithm, KNN and Artificial Neural Network four major techniques. This study demonstrates that correct outcomes with technique of Decision Tree will be the major changes among the planned rules based classification which is ideal without procedures for the liver diseases prediction. The good decision can be possible in medical organization if they use this model as a tool.

By applying all the prototypes and its presentation was estimated. Performance evaluation was based on certain performance metrics. The ANN model bring about in the highest accuracy with an accuracy of 98%. Comparing this effort with the earlier study it was discovered that ANN proved highly efficient. A GUI, which will be a medical tool by hospitals and medical staff was implemented using ANN.

FUTURE ENHANCEMENT

The classification algorithms proposed in this work can be enhance to expand the classification accuracy further with additional limitations.

REFERENCES

ISSN: 2233-7857 IJFGCN

[1]PAUL R. HARPER, A REVIEW AND COMPARISON OF CLASSIFICATION ALGORITHMS FOR MEDICAL DECISION MAKING.

[2] P. SUG, ON THE OPTIMALITY OF THE SIMPLE BAYESIAN CLASSIFIER UNDER ZERO-ONE LOSS, MACHINE LEARNING 29 (2–3) (1997) 103–130.

[3] APARNA BEHARA AND M.V. RAGHUNADH, "REAL TIME FACES RECOGNITION SYSTEM FOR TIME AND ATTENDANCE APPLICATIONS", INTERNATIONAL JOURNAL OF ELECTRICAL, ELECTRONICS AND DATA COMMUNICATION, VOLUME- 1, ISSUE- 4, ISSN: 2320-2084.
2009

[4] DR. S.VIJAYARANI, MR.S.DHAYANAND, LIVER DISEASE PREDICTION USING SVM AND NAÏVE BAYES ALGORITHMS, INTERNATIONAL JOURNAL OF SCIENCE, ENGINEERING AND TECHNOLOGY RESEARCH (IJSETR), 2015.

[5] DASH M., LIU H., "FEATURE SELECTION FOR CLASSIFICATION," INTELLIGENT DATA ANALYSIS, ELSEVIER, PP. 131 -156, 1997.

ISSN: 2233-7857 IJFGCN