A Review of Liver Patient Analysis Methods using Machine Learning

Team ID : NM2023TMID32697

Team Size : 4

Team Leader : DURGADEVI S

Team member : JANANI S

Team member : SARITHA S

Team member : VEMBU R

INTRODUCTION

Liver is the largest internal organ in the human body, is essential for digesting food and releasing the toxic element of the body and plays a major role in metabolism and serving several vital functions. The liver is the largest glandular organ of the body. It weighs about 3 lb (1.36 kg) .The liver's main job is to strain the blood coming from the digestive tract, before passing it to the rest of the body. The liver also detoxifies chemicals and metabolizes drugs. As it does so, the liver hides bile that ends up back in the intestines. The liver also makes proteins important for blood clotting and other functions. The liver supports almost every organ in the body and is vital for our survival. Liver disease may not cause any symptoms at earlier stage or the symptoms may be vague, like weakness and loss of energy. Symptoms partly depend on the type and the extent of liver disease. Liver diseases are diagnosed based on the liver functional test[1].

Several diseases states can disturb the liver. Some of the diseases are Wilson's disease, hepatitis (an inflammation of the liver), liver cancer, and cirrhosis (a chronic inflammation that progresses ultimately to organ failure). Alcohol alters the metabolism of the liver, which can have on the whole detrimental effects if alcohol is taken over long periods of time. Hemochromatosis can cause liver problems [2].

Common Liver Disorder

- Fatty liver is a revocable condition where large vacuoles of triglyceride fat acquire in liver cells via the process of limit. It can occur in people with a high level of alcohol consumption as well as in people who never had alcohol.
- Hepatitis(usually caused by a virus spread by excess contamination or direct contact with infected body fluids).
- Cirrhosis of the liver is one of the most serious liver diseases. It is an action used to indicate all forms of diseases of the liver characterized by the significant loss of cells. The liver gradually contracts in size and becomes leathery and hard. The regenerative action continues under liver cirrhosis but the progressive loss of liver cells exceeds cell replacement.
- Liver cancer. The risk of liver cancer is higher in those who have cirrhosis or who had valid types of viral hepatitis; but more often, the liver is the site of secondary (metastatic) cancers spread from other organs.

LITERATURE REVIEW

Health care and medicine handles huge data on daily basis. This data comprises of information about the patients, diagnosis reports and medical images. It is important to utilize this information to decipher a decision support system. To achieve this it is important to discover and extract the knowledge domain from the raw data. It is accomplished by knowledge discovery and data mining (KDD) [3]. The implementation of data mining techniques is widespread in biological domain. In recent years, liver disorders have excessively increased and liver diseases are becoming one of the most fatal diseases in several countries. In this study, liver patient datasets are investigate for building classification models in order to predict liver disease. Several feature model construction and comparative analysis are implemented for improving prediction accuracy of Indian liver patients. Different studies have been conducted for classification of liver disorders, they are discussed briefly.

Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification accuracy.

METHODOLOGY

1. Data Collection

For this study, the Indian Liver Patient Dataset (ILPD) was selected from the UCI Machine Learning repository. It is a sample of the whole Indian population taken from the area of Andhra Pradesh. There were 583 instances based on ten different biological parameters in the dataset. Based on these criteria, the class value was stated as either yes (416 cases) or no (167 cases), reflecting the liver.

2. Pre-processing and Feature selection

To normalize the missing values, pre-processing techniques have been introduced. The missing values were replaced by null values along with their instances. Feature selection was followed to classify the appropriate attribute for classification. Using both filter and wrapper approaches, feature selection was carried out. The attributes with more than 70% correlation were initially excluded by correlation analysis from the dataset. The algorithm was implemented to estimate the value of different features in a dataset on the basis of random forests [9].

3. Randomization and splitting of dataset

To build classification models, the features selected in the preceding phase were accepted. The dataset was initially randomized to produce an arbitrary sample permutation. Splitting of the dataset into training (70 percent of the dataset) and test (30 percent) sets was followed. The training set consisted of 389 cases and the evaluation set consisted of the remaining 194 cases.

4. Classification algorithms

Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification accuracy. Different data mining algorithms like Naïve Bayes, Decision Tree, Logistic Regression, Random forest and Support vector machine (SVM) were implemented.

The application mainly consists of the following tasks:

➢ **Building and training the system: The phase is Totally worked upon by developer of the system, And end user has nothing to do with it. In this Phase, we split the dataset into training dataset And test dataset, and then trained the models Using training dataset.**

➢ **Testing the models: In this phase we tested the Accuracy of the models with the test dataset that Was formed in previous phase and the most Accurate model is figured out.**

➢ **Entering details and prediction: In this phase, The end user comes into picture. He/she enters The details of blood test report using GUI of the Application. The application then matches the Details with the training dataset of the most Accurate model, and then predicts final result Displaying, 'Risk' or 'No Risk' on the screen.**

CONCLUSION

This work presents an approach that will be used for Hybrid model construction of community health Services. These classification algorithms can be Implemented for other dominant diseases also like Cardiac and diabetes prediction and classification. More than one dataset may be used for better Approach and comparison (Ahmed et al., 2020). Another scope is to see whether by applying new Algorithms will result any improvements over Techniques which are used in this work in future. More techniques for accuracy increment may be Applied. Wrapper method may be applied for Removing noise in the dataset.

Classification rules and disease identifying Techniques may also be generated by using different Efficient algorithms. More than one database for Comparative analysis may also be used. Our works Has certain limitations as the model has Underperformed having less accuracy than Expectations. So, in future, inclusion of deep learning Methods may improve our results further.

Future work

Diseases related to liver and heart are becoming more and More common with time. With continuous technological Advancements, these are only going to increase in the future. Although people are becoming more conscious of health Nowadays and are joining yoga classes, dance classes; still the Sedentary lifestyle and luxuries that are continuously being Introduced and enhanced; the problem is going to last long. So, in such a scenario, our project will be extremely helpful To the society. With the dataset that we used for this project, We got 100 % accuracy for SVM model, and though it might Be difficult to get such accuracies with very large datasets, From this project's results, one can clearly conclude that we Can predict the risk of liver diseases with accuracy of 90 % or More.

Today almost everybody above the age of 12 years has Smartphones with them, and so we can incorporate these Solutions into an android app. Also it can be Incorporated into a website and these app and website will be Highly beneficial for a large section of society.

THANK YOU!