Comparative Analysis of Liver diseases by using Machine Learning Techniques

ABSTRACT

In a human body function of the liver is important. Many persons are suffering from liver disease, but they don't know it. The identification of liver diseases in the early stage helps a patient get better treatment. If it is not diagnosed earlier, it may lead to various health issues. To solve these issues, physicians need to examine whether the patient has been affected by liver disease or not, based on the multiple parameters. In this paper, we classify the patients who have liver disease or not by using different machine learning algorithms by comparing the performance factors and predicting the better result. The liver dataset is retrieved from the Kaggle dataset.

**EXISTING SYSTEM**

Nazmun Nahar, et al. [3] implemented by using various decision trees techniques like LMT, J48, Hoeffding Tree, Decision Stump, and Random tree. for calculating expected time predication of disease affected to liver finally, the Decision Stump gives the highest accuracy results among other techniques.

A Saranya, et al. [4] explained the applications in data mining techniques and also used Medical Data Mining (MDM) to diagnose liver diseases. This technique includes prediction in the early stage, the existence and also complexity of the disease which helps partial assistance to the physicians.

S. Dhamodharan [5] considers three major liver diseases like cirrhosis, hepatitis, and liver cancer. The fundamental purpose of this forecast is to find the type of disease by using classifications techniques such as cirrhosis, hepatitis, liver cancer, and "no disorders." Then compare the accuracy of the FT and Naive Bayes tree algorithms and shows that the Naive Bayes algorithm accuracy is significantly higher than that of the other methods.

Kemal Akyol, Yasemin Gultepe [6] by using the dataset which has shown a balanced result by using sampling technique for getting accuracy. the Stability Selection technique is used for selection based on attributes. For improving the performance, a blend of Stability Selection and Random Forest methods is used.

Shambel Kefelegn, Pooja Kamat [7] for getting better results different data mining classification techniques are compared with the earlier liver prediction methods. The accuracy is measured with the help of confusion matrices for getting the better performance of the accuracy Fadl Mutaher Ba-Alwi, et al. [8] using various machine learning algorithms compared the Hepatitis prognostic data among them. In that Naive Bayes, technique gave good accuracy and also takes less time to build a model.

K. Thirunavukkarasu, et al. [9] Used different classification techniques for predicting liver diseases. They compare the results of accuracy score and confusion matrix with Logistic Regression, SVM, and K-Nearest Neighbour. Bendi Venkata, et al. [10] used different classifications algorithms, they checked the accuracy, precision, sensitivity, and specificity on liver datasets.

Tapas Ranjan Baitharua et al. [11] has proposed an Intelligent medical decision support system to help physicians diagnose liver disorders through a learning pattern technique. In this, several classification techniques are used to compare the effectiveness, correction rate, and also accuracy for the data is analysed with different scenarios.

A diagnostic support system [12] was developed with the support of a number of models with help of neural, which is helped to the physicians for diagnosis on the liver in the medical field. M. Banu Priya P, et al. [13] Using a root mean error value, root mean square value, the accuracy is calculated, and better accuracy is produced with the support of the PSO features selection technique.

Dietterich, Thomas G [14] states that the ensemble learning technique produces a better performance than the other single classifier techniques with the Bayesian averaging, error-correcting output coding, boosting, and bagging. In this paper, the author analyses existing ensemble approaches with some novel experiments to determine why Adaboost does not overfit quickly.

**Disadvantages**

* An existing methodology doesn’t implement an effective Machine Learning Classifiers.
* The system not implemented Decision Trees which leads complexity in testing and training datasets.

Proposed System

To analysis about liver disease, the data is retrieved from the Indian Liver Patient from the Kaggle. Here the patient has been characterized based on diseases as either 1 or 2. The values used in the dataset are given in Table I. The gender attribute is transformed to an integer value during the data pre-processing stage (0 and 1). The overall procedure of the proposed system is depicted in this system. The workflow of the proposed system is as follows a collection of data sets, Handling Categorical values, Splitting the data for Training, and Testing. Perform feature selection and apply the machine learning techniques and compare the predicted result and find better accuracy.

**Advantages**

* It is a supervised Machine learning technique applied for both classification and regression kinds of problems, but it is used for classification types of problems. This model is applied to predict the categorical dependent variable with support of independent variables the output should be 0 or 1.
* This Classifier technique is effective when only a small amount of training data is required to derive approximation parameters. With highly scalable model creation, it can tackle a wide range of challenging real-world problems.
* This is a pattern recognition system that involves the training datasets for finding the k closest relatives in new conditions. When using k-NN for classification, we must calculate the location data within the nearest neighbor’s category.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).