STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

TEAM ID: PNT2022TMID15966

ABSTRACT

The improvement of patient care, research, and policy is significantly impacted by medical diagnoses. Medical practitioners employ a variety of pathological techniques to make diagnoses based on medical records and the conditions of the patients. Disease identification has been significantly enhanced by the application of artificial intelligence and machine learning in conjunction with clinical data. Data driven, machine learning (ML) techniques can be used to test current approaches and support researchers in potentially innovative judgments. The goal of this work was to use ML algorithms to derive meaningful predictors of liver disease from the medical data.

INTRODUCTION

The number of patients with liver disease has been steadily rising as a result of excessive alcohol use, exposure to hazardous gases, ingestion of tainted foods such pickles and cucumbers, and drug usage. In an effort to lighten the load on doctors, this dataset was used to assess prediction systems. The data set consists of the patient's age, gender, and total bilirubin. Direct bilirubin, alkaline phosphatase, alamine aminotransferase, aspartate aminotransferase, total proteins, albumin, and the ratio of albumin to globulin are other examples. Set: the field that was utilized to divide the data into two sets (patient with liver disease, or no disease). This study attempts to find an appropriate machine learning algorithm that can determine whether a person has liver disease or not given a dataset containing biological and diagnostic data of Indian patients.

Relevance of the Project

Using certain characteristics such as total bilirubin, direct bilirubin, alkali nephosphatase, total protein,albumin, and globulin,this software can determine whethera patient has liver disease or not.

Scope of the Project

- It is necessary to use supervised learning
- To resolve this binary classification issue. Each data point has ten attributes , and thereis a label that indicates if the patient has liverdisease or not. In order to find the answer,our goal should be to train a variety of supervised learning models on this dataset in order to create a high – performing model that can accurately identify any new data point as positive or negative and outperform the benchmarks.

PROPOSED MODEL