PROBLEM STATEMENT

STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

INTRODUCTION

To detect disease, healthcare professionals need to collect samples from patients which can cost both time and money. Often, more than one kind of test or many samples are needed from the patient to accumulate all the necessary information for a better diagnosis. The liver has many functions such as glucose synthesis and storage, detoxification, production of digestive enzymes, erythrocyte regulation, protein synthesis, and various other features of metabolism. Chronic liver diseases include chronic hepatitis, fibrosis, and cirrhosis. When liver disease is diagnosed at an earlier stage, in between infection and fibrosis but before cirrhosis, liver failure can be avoided. Liver-related disease accounts for 70% of deaths worldwide. There is a need to find better ways to detect and diagnose liver disease with more accuracy. Most importantly, tests of liver function need to be available and affordable to patients. To avoid the expensive and invasive tests, we need to go for the different algorithms.

EXISTING SYSTEM

We use Machine Learning approaches to the application of statistical machine learning techniques to CMP results for the extraction of information for a clinician might be helpful for diagnosis. Exploratory data analysis methods are extremely important in healthcare; they can predict patterns across data sets to facilitate the

determination of risk or diagnostic factors for disease with more speed and accuracy. The use of these methods can allow for earlier detection and potentially prevent many cases of liver disease from progressing to the point of needing biopsy or complex treatment.

PROPOSED SYSTEM

Various kinds of data sets, such as blood panels with liver function tests, histologically stained slide images, and the presence of specific molecular markers in blood or tissue samples, have been used to train classifier algorithms to predict liver disease with good accuracy. The ML methods described in previous studies have been evaluated for accuracy by a combination of confusion matrix, receiver operating characteristic under area under curve, and k-fold cross-validation.

Software based on classification of algorithm

Logistic regression

Random forest

Navis Bayes