

LITERATURE SURVEY

Using machine learning algorithms to predict disease is made possible by increasing access to hidden attributes in medical data sets. 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. Singh et al. designed software based on classification algorithms (including logistic regression, random forest, and naive Bayes) to predict the risk of liver disease from a data set with liver function test results. Vijayarani and Dhavanand found that SVM performed better over naive Bayes to predict cirrhosis, acute hepatitis, chronic hepatitis, and liver cancers from patient liver function test results. SVM with particle swarm optimization (PSO) predicted the most important features for liver disease detection with the highest accuracy over SVM, random forest, Bayesian network, and an MLP-neural network. SVM more accurately predicted drug-induced hepatotoxicity with reduced molecular descriptors than Bayesian and other previously used models [38]. Phan and Chan et al. demonstrated that a convolutional neural network

(CNN) model predicted liver cancer in subjects with hepatitis with an accuracy of 0.980 . The ANN model has been used to predict liver cancer in patients with type 2 diabetes . Neural network ML methods can help differentiate between types of liver cancers when applied to imaging data sets . Neural network algorithms have even been trained to predict a patient's survival after liver tumor removal using a data set containing images of processed and stained tissue from biopsies . ML methods can facilitate the diagnosis of many diseases in clinical settings if trained and tested thoroughly. More widespread application of these methods to varying data sets can further improve accuracy in current deep learning methods. This study aimed to (i) impute missing data using the MICE algorithm; (ii) determine variable selection using eigen decomposition of a data matrix by PCA and to rank the important variables using the Gini index; (iii) compare among several statistical learning methods the ability to predict binary classifications of liver disease; (iv) use the synthetic minority oversampling technique (SMOTE) to oversample minority class to regulate overfitting; (v) obtain confusion matrices for comparing actual classes with predictive classes; (vi) compare several ML approaches to assess a better performance of liver disease diagnosis; (viii) evaluate receiver operating characteristic (ROC) curves for determining the diagnostic ability of binary classification of liver disease.

Bendi et al. [1] authors used two different input dataset and evaluate that the AP datasets has better than UCLA dataset for all the different selected algorithms. Based on performance on their classification KNN,

Backward propagation and SVM are giving better results. The AP data set is better than UCLA for the entire selected algorithm. And found out Naïve Bayes, C4.5, KNN, Backward propagation and SVM has 95.07, 96.27, 96.93, 97.47, & 97.07% accuracy respectively.

- Bendi et al. [2] proposed a paper based on Modified Rotation Forest, used two dataset as an input UCI liver dataset and Indian liver dataset. And results show that MLP algorithm with random subset gives better accuracy of 94.78% for UCI dataset than CFS achieved accuracy of 73.07% for Indian liver dataset

- . • Yugal Kuma & G. Sahoo [3] proposed a paper based on different classification technique and used north east area of Andhra Pradesh (India) liver dataset. And the results shows that Decision tree(DT) algorithm has better than other algorithm and provide accuracy of 98.46%. • S.Dhamodharan [4] proposed a paper based on two classification technique naïve Bayes and FT tree and used WEKA (Waikato Environment for Knowledge and Analysis) dataset. Naïve Bayes is 75.54% accuracy and FT Tree is 72.6624% accuracy and concluded Naïve Bayes gas better algorithm compare to other algorithms.

- Han Ma et al. [9] in this paper 11 different classification are evaluated and Demonstrated in China Zhejiang University, College of medicine and concluded Bayesian network accuracy of 83%, specificity 83%, sensitivity of 0.878 and F-measure of 0.655.

- Heba Ayeldeen et al. [5] propose a paper for prediction of liver

fibrosis stages using decision tree technique and used Cario university data set and result shows that decision tree classifier accuracy is 93.7%.

- D.Sindhuja & R. Jemina Priyadarsini [6] survey a paper for classification of liver disease. In this survey different classification techniques of data mining are study and used dataset of dataset of AP liver has better than Dataset of UCLA, and concluded C4.5 achieved better results than other algorithms.