**LITERATURE SURVEY**

**INTRODUCTION**

Medical diagnoses have important implications for improving patient care, research, and policy. For a medical diagnosis, health professionals use different kinds of pathological methods to make decisions on medical reports in terms of the patients’ medical conditions. Recently, clinicians have been actively engaged in improving medical diagnoses. The use of artificial intelligence and machine learning in combination with clinical findings has further improved disease detection.

The advantage of computers and technologies, one can collect data and visualize many hidden outcomes such as dealing with missing data in medical research. Statistical machine learning algorithms based on specific problems can assist one to make decisions. Machine learning (ML), data-driven algorithms can be utilized to validate existing methods and help researchers to make potential new decisions.

**USE CASE**

The purpose of this study was to extract significant predictors for liver disease from the medical analysis of 615 humans using ML algorithms. Data visualizations were implemented to reveal significant findings such as missing values. Multiple imputations by chained equations (MICEs) were applied to generate missing data points, and principal component analysis (PCA) was used to reduce the dimensionality.

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 most routine tests are urinalysis, complete blood count (CBC), and comprehensive metabolic panel (CMP). These tests are generally less expensive and can still be very informative.

**Existing solutions**

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.

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.

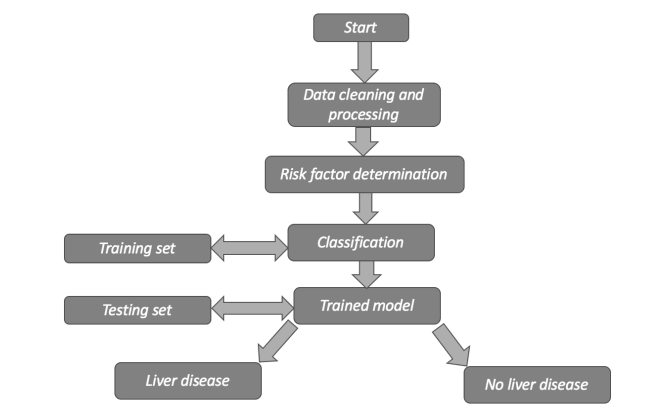
This study aimed to

1. compute missing data using the MICE algorithm.
2. determine variable selection using eigen decomposition of a data matrix by PCA and to rank the important variables using the Gini index
3. 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

1. obtain confusion matrices for comparing actual classes with predictive classes
2. compare several ML approaches to assess a better performance of liver disease diagnosis
3. evaluate receiver operating characteristic (ROC) curves for determining the diagnostic ability of binary classification of liver disease.

**STUDY DESIGN**

****

**Conclusions**

Chronic liver diseases is detected by clinicians who are well trained in identifying significant observations and classifying them as normal or abnormal using background information and other context clues. ML algorithms can be trained to detect the possibility of liver disease in a similar way to assist healthcare workers. Using the correlation of each variable with the risk of liver disease to train the model, ML methods were able to identify which blood donors were healthy and which had liver disease with high accuracy. The PCA results showed five important factors for liver disease diagnosis: AST, ALT, GGT, BIL and ALP.

**References**

1. Wang, Y.; Li, Y.; Wang, X.; Gacesa, R.; Zhang, J.; Zhou, L.; Wang, B. Predicting Liver Disease Risk Using a Combination of Common Clinical Markers: A Screening Model from Routine Health Check-Up. Dis. Markers 2020, 2020, 8460883.

2. Torkadi, P.P.; Apte, I.C.; Bhute, A.K. Biochemical evaluation of patients of alcoholic liver disease and non-alcoholic liver disease. Indian J. Clin.Biochem. 2014, 29, 79–83.

3. Ceriotti, F.; Henny, J.; Queraltó, J.; Ziyu, S.; Özarda, Y.; Chen, B.; Boyd, J.C.; Panteghini, M. Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ-glutamyltransferase (GGT) in serum: Results from an IFCC multicenter study. Clin. Chem. Lab. Med. 2010, 48, 1593–1601.

4.Chalasani, N.; Younossi, Z.; Lavine, J.E.; Charlton, M.; Cusi, K.; Rinella, M.; Harrison, S.A.; Brunt, E.M.; Sanyal, A.J. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018, 67, 328–357.

5. Woreta, T.A.; Saleh, A.A. Evaluation of abnormal liver tests. Med Clin. 2014, 98, 1–16.