**AI-Statistical Machine Learning Approaches to**

**Liver Disease Prediction**

Team ID: PNT2022TMID4827

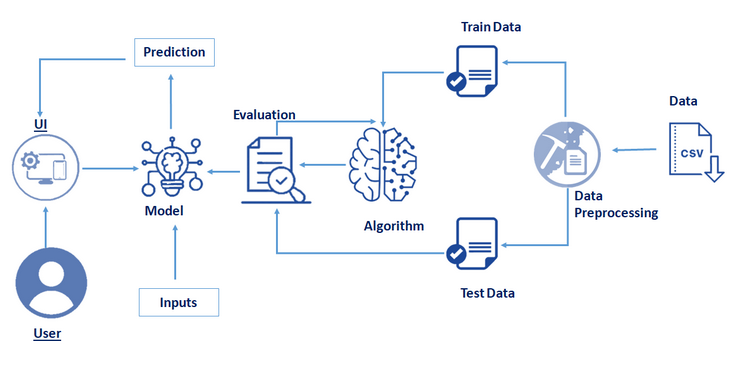
**Faculty Mentor: Team Leader:** G.lydia

D.**Pradhiba Team Member:** R.priya

**Team Member:** U.lavanya sri

**Team Member:**G.nagalakshmi

**Technology Architecture**



Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this project is to analyze the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This Project examines data from liver patients concentrating on relationships between a key list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. Here we are building a model by applying various machine learning algorithms find the best accurate model. And integrate to flask based web application. User can predict the disease by entering parameters in the web application. 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. In the modern era, with 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. 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. Variable importance ranking using the Gini index was implemented to verify significant predictors obtained from the PCA. Training data (ntrain = 399) for learning and testing data (ntest = 216) in the ML methods were used for predicting classifications. The study compared binary classifier machine learning algorithms (i.e., artificial neural network(ML), data-driven algorithms can be utilized to validate existing methods and help researchers to make potential new decisions. 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. Variable importance ranking using the Gini index was implemented to verify significant predictors obtained from the PCA. Training data (ntrain = 399) for learning and testing data (ntest = 216) in the ML methods were used for predicting classifications. The study compared binary classifier machine learning algorithms (i.e., artificial neural network, random forest (RF), and support vector machine), which were utilized on a published liver disease data set to classify individuals with liver diseases, which will allow health professionals to make a better diagnosis. The synthetic minority oversampling technique was applied to oversample the minority class to regulate overfitting problems. The RF significantly contributed (p < 0.001) to a higher accuracy score of 98.14% compared to the other methods. Thus, this suggests that ML methods predict liver disease by incorporating the risk factors, which may improve the inference-based diagnosis of patients. Keywords: liver disease; demographic variables; prognostic/biochemical variables; statistical learn- ing for variable selection and classification 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. Hepatitis can occur from viral infection (e.g., hepatitis c virus) or auto-immune origin.