**AI-Statistical Machine Learning Approaches to**

**Liver Disease Prediction**

Team ID: PNT2022TMID4827

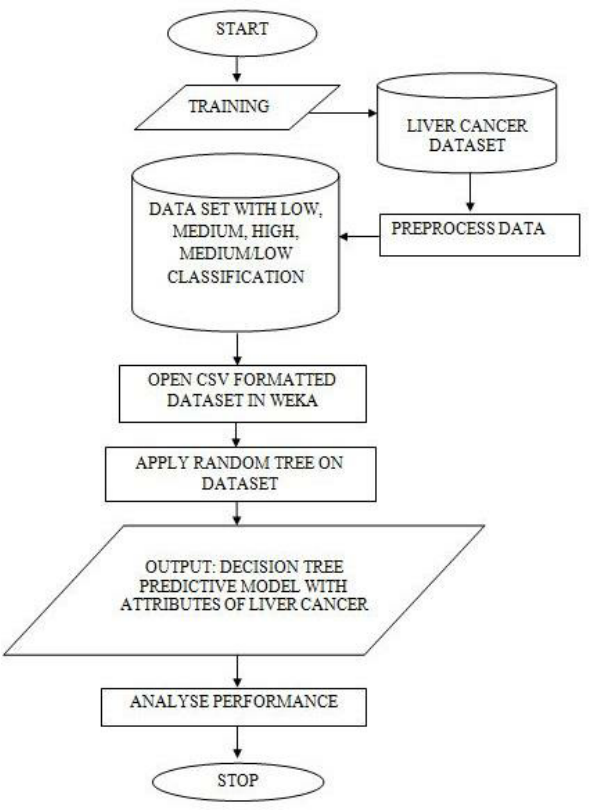
**Faculty Mentor: Team Leader:** G.lydia

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

**Team Member:** U.lavanya sri

**Team Member:**G.nagalakshmi

**Data Flow Diagrams:**



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, 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