Ideation Phase
Literature Survey and Information Gathering

DATE	1st october 2022
TEAM ID	PNT2022TMID24994
PROJECT NAME	STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION
MAXIMUM MARKS	

LITERATURE SURVEY

1. **Bendi et al.** [1] authors used two different input dataset and evaluated that the AP dataset was 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. Naïve Bayes, C4.5, KNN, Backward propagation and SVM has 95.07, 96.27, 96.93, 97.47, & 97.07% accuracy respectively.

2. **Bendi et al.** [2] proposed a paper based on Modified Rotation Forest, using 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.

3. **Yugal Kuma & G. Sahoo** [3] proposed a paper based on different classification techniques and used the north east area of Andhra Pradesh (India) liver dataset. And the results show that the Decision tree(DT) algorithm is better than other algorithms and provides accuracy of 98.46%.

4. **S.Dhamodharan** [4] proposed a paper based on two classification techniques 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 that Naïve Bayes is a better algorithm compared to other algorithms.

5. **Heba Ayeldeen et al.** [5] propose a paper for prediction of liver fibrosis stages using decision tree technique and using Cario university data set and result shows that decision tree classifier accuracy is 93.7%.

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

7. **Somaya Hashem et al.** [7] presented a paper for diagnosis of liver disease. In this paper they used two algorithms, SVM & Backpropagation and used UCI machine repository dataset. And concluded SVM has 71% better result than Backpropagation accuracy 73.2%.

8. **Han Ma et al.** [8] in this paper 11 different classifications 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.

9. **Joel Jacob et al.** [9] proposed a paper to diagnose liver disease by using three different algorithms, Logistic regression, K-NN, SVM, and ANN and used Indian Liver Patient Dataset comprising 10 different attributes of 583 patients. Logistic regression, K-NN, SVM,& ANN has 73.23, 72.05, 75.04 & 92.8% accuracy respectively.

10. **Sivakumar D et al.** [10] proposed a paper for prediction of chronic liver disease by using two different techniques K-means and C4.5. UCI repository.

11. **Mehtaj Banu H** [12] in this paper authors study different machine learning techniques, Supervised, unsupervised & reinforcement and also analyse UCI dataset database and conclude that KNN and SVM improved better performance and exactness of liver disease prediction.

Comparison table of various machine learning techniques used to detect liver disease based on performance.

Methods	Accuracy	Specification	Sensitivity
Decision Tree	98	95	95
Bayesian Network	83	87	67
SVM	71		
Naive Bayes	55		
KNN	70		