POJECT DEVELOPMENT PHASE
DELIVERY OF SPRINT -3

DATE	06 NOVEMBER 2022
TEAM ID	PNT2022TMID05432
PROJECT NAME	STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

Abstract: Healthcare system is becoming a significant factor of every human being. So in this paper we are discussing about the vital organ of the human body i.e. Liver. Early analysis of liver disease prediction is highly challenging. As a result, liver disease has become more common which is gradually improved, enhancing people's troubles. To recognize the causes and the identification phases are more important. For this, we applied a machine learning technique that was highly promising for studies with regard to healthcare and health. To uncover hidden data for precise diagnosis and decision-making, ML can also be used. Any form of liver disease is hazardous to your health. Poisons, narcotics, and excessive drinking are the main causes of this condition. In this study, we will use several associated features and KNN technology to enhance liver disease prediction. The accuracy forecast is provided by this algorithm.

I. INTRODUCTION

Machine learning, a branch of artificial intelligence (AI), is extensively used in data science. A vital internal organ of the human body is the liver, carrying out several key tasks like digestion, metabolism, and vitamin preservation, among others. Liver ailment is one of the crucial/dangerous disease, nowadays liver disorders increased rapidly. Prediction of the disease in an early stage is an important thing, so that it would be increase the survival rate. Effective decision making in medical, ML algorithms plays an important role. For the purpose of predicting liver disease, there are numerous machine learning methods available. Every algorithm learns in a unique way. Here we focuses on KNN algorithm for the better performance which optimises the model, it will

developing the several models step by step. This KNN method was chosen because it classifies diseases by looking for a number of values that are close by. This improves our ability to analyse the various characteristics associated to the prediction process. In this research, we addressed how to analyse and forecast illnesses more precisely through a step-by-step procedure that includes dataset selection, pre-processing, and classification, after which a machine learning model is built using the available data. The analysis and validation of each feature distribution will be done as part of the pre-processing of the data. The dataset is first determined to be balanced or imbalanced, and then the proper techniques are employed to convert the data to a normal distribution and balance the dataset using a variety of techniques

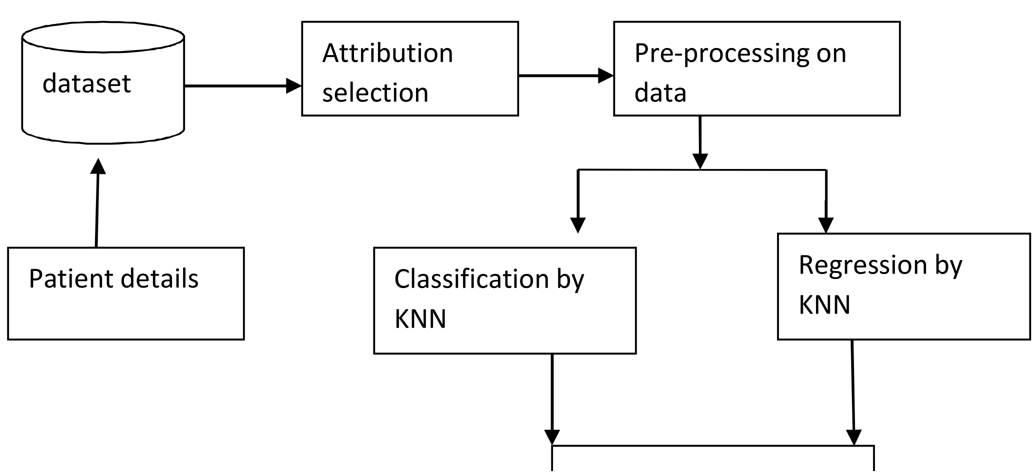
II. LITERATURE SURVEY

L.Aliceauxilia [1] presents Accuracy Prediction for Indian Patient Liver Using Machine Learning Techniques. Here, they employed the "Naive Bayes" Classification, which has an accuracy rate of 37%. The authors of Thirunavukkarasu K et al [2] discuss how to predict liver disease using classification algorithms. In this case, "Logistic Regression" was applied, achieving 73.97 percentage accuracy. In their study, Jagdeep Singha et al. [3] employed the "random forest regression technique". This achieved accuracy to the tune of 71.36 percent. Mehtaj Banu H [4] Presenting Liver Disease Prediction Using Machine-Learning. This analysis makes use of the "Support Vector Machine algorithm". Shobana, G et al [5] suggested a method of "feature reduction using Recursive Feature Elimination" and applying "Machine learning boosting methods" to increase prediction accuracy. Sateesh Ambesang et al [6] report the Liver Diseases Prediction Using KNN with Hyper Parameter Tuning Techniques. In this case, the suggested model with K- NN parameters that have been fine-tuned using grid search performs better with 91 percentage accuracy. The focus of this work is KNN algorithms, procedures to optimise the model's performance, and the gradual development of numerous models. Kanza Hamid et al. [7] is explained the Abstention-based Machine Learning for Automated Liver Disease Diagnosis. In this paper, an unique technique for automatically identifying liver anomalies in ultrasound images is described.

Shivangi Gupta et al [8] present A Web-Based Framework Using Combined Machine Learning Models for Liver Disease Diagnosis. This study combines three optimised machine learning algorithms, including weighted K-Nearest Neighbors, decision trees, and artificial neural networks, to categorise the provided data and produce results. A.Sivasangari et al [9] Discuss Machine Learning Models for Liver Disease Diagnosis. In order to forecast liver sickness more correctly, precisely, and consistently, Random Forests, Decision Trees, and Support Vector Machines (RF) techniques are recommended. Golmei Shaheamlung et al [10] talk about a survey on the use of machine learning for diagnosing liver illness. In order to aid the medical community in the diagnosis and prognosis of liver disease, this study seeks to provide a thorough overview and comparative analysis of all machine learning methodologies.

III. PROPOSED METHODOLOGY

This initiative seeks to foresee the liver disease. For the classification of Indian Liver Patient Disease (ILPD), the suggested approach leverages the idea of a machine learning algorithm; in this case, the models are trained before being tested. After that the most accurate model will forecast the outcome.



IV. RESULT ANALYSIS

Output of the project is prediction of Indian Liver Patient Dataset (ILPD) in which accuracy will be predicted and the average of the accuracy will be printed

```
[[1.0, -0.5, 0.2, 72.0, 140.0, 2.5, 1]] predicted --> 1
[[1.0, -0.5, 2.3, 32.0, 92.0, 2.6, 2]] predicted --> 2
[[1.0, -0.5, 0.2, 53.0, 58.0, 3.4, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 15.0, 18.0, 4.7, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 24.0, 20.0, 3.2, 2]] predicted --> 2
[[1.0, -0.5, 0.1, 28.0, 51.0, 4.2, 2]] predicted --> 2
[[1.0, -0.5, 2.4, 126.0, 202.0, 2.9, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 30.0, 28.0, 1.8, 2]] predicted --> 2
[[1.0, -0.5, 0.2, 28.0, 35.0, 3.4, 2]] predicted --> 2
[[1.0, -0.5, 0.1, 26.0, 29.0, 4.2, 2]] predicted --> 2
[[1.0, -0.5, 1.7, 950.0, 1500.0, 3.8, 1]] predicted --> 1
[[1.0, -0.5, 0.1, 37.0, 29.0, 2.6, 1]] predicted --> 1
[[1.0, -0.5, 1.2, 55.0, 95.0, 3.7, 1]] predicted --> 1
[[1.0, -0.5, 8.8, 43.0, 185.0, 2.4, 1]] predicted --> 1
[[1.0, -0.5, 0.4, 60.0, 103.0, 3.0, 2]] predicted --> 2
```

Fig 2: Predicted values

This above fig showing the prediction result 1 or 2 along with dataset.

```
[[-1.0, -0.5, 0.2, 35.0, 32.0, 3.9, 2]] predicted --> 2
[[-1.0, -0.5, 0.3, 26.0, 15.0, 3.8, 1]] predicted --> 1
[[1.0, -0.5, 1.3, 23.0, 29.0, 0.9, 1]] predicted --> 1
[[1.0, -0.5, 0.1, 25.0, 27.0, 3.1, 1]] predicted --> 1
[[1.0, -0.5, 0.7, 10.0, 21.0, 2.1, 2]] predicted --> 2
[[1.0, -0.5, 1.4, 48.0, 89.0, 3.0, 1]] predicted --> 1
[[1.0, -0.5, 0.5, 20.0, 30.0, 1.9, 2]] predicted --> 2
[[1.0, -0.5, 0.3, 28.0, 25.0, 3.0, 1]] predicted --> 1
[[1.0, -0.5, 0.7, 10.0, 21.0, 2.1, 2]] predicted --> 2
[[1.0, -0.5, 0.8, 53.0, 58.0, 3.8, 1]] predicted --> 1
[[1.0, -0.5, 4.9, 140.0, 350.0, 2.1, 1]] predicted --> 1
[[-1.0, -0.5, 0.8, 28.0, 44.0, 2.9, 2]] predicted --> 2
[[-1.0, -0.5, 0.2, 13.0, 21.0, 3.2, 2]] predicted --> 2
[[1.0, -0.5, 0.5, 30.0, 42.0, 3.8, 2]] predicted --> 2
93.63636363636364
```

Fig 3: Showing the first fold accuracy

The above figure shows the accuracy of the first fold

```
[[1.0, -0.5, 0.8, 60.0, 84.0, 3.5, 2]] predicted --> 2
[[1.0, -0.5, 0.4, 41.0, 38.0, 2.1, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 36.0, 32.0, 4.0, 2]] predicted --> 2
[[1.0, -0.5, 0.3, 16.0, 24.0, 3.0, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 72.0, 140.0, 2.5, 1]] predicted --> 1
[[1.0, -0.5, 0.3, 18.0, 21.0, 3.7, 2]] predicted --> 2
[[1.0, -0.5, 9.5, 46.0, 52.0, 2.0, 1]] predicted --> 1
[[1.0, -0.5, 2.0, 27.0, 59.0, 2.4, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 33.0, 57.0, 3.5, 2]] predicted --> 2
[[-1.0, -0.5, 0.2, 166.0, 397.0, 3.0, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 27.0, 28.0, 1.6, 2]] predicted --> 2
[[-1.0, -0.5, 0.2, 23.0, 28.0, 2.9, 2]] predicted --> 2
[[1.0, -0.5, 0.2, 18.0, 15.0, 4.2, 2]] predicted --> 2
[[1.0, -0.5, 3.6, 198.0, 143.0, 3.2, 1]] predicted --> 1
[[-1.0, -0.5, 11.4, 30.0, 37.0, 3.4, 1]] predicted --> 1
[[1.0, -0.5, 0.1, 26.0, 27.0, 3.6, 2]] predicted --> 2
[[-1.0, -0.5, 0.1, 95.0, 127.0, 2.1, 1]] predicted --> 1
[[-1.0, -0.5, 0.2, 15.0, 12.0, 4.7, 1]] predicted --> 1
92.95454545454545
```

Fig 4: Showing the second fold accuracy

In this above figure which shows the accuracy of the second fold accuracy

```
[[1.0, -0.5, 2.3, 32.0, 92.0, 2.6, 2]] predicted --> 2
[[1.0, -0.5, 0.2, 24.0, 46.0, 3.1, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 50.0, 28.0, 4.0, 1]] predicted --> 1
[[-1.0, -0.5, 0.2, 17.0, 24.0, 4.0, 1]] predicted --> 1
[[-1.0, -0.5, 0.1, 20.0, 33.0, 1.9, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 28.0, 35.0, 3.4, 2]] predicted --> 2
[[1.0, -0.5, 1.5, 102.0, 630.0, 3.3, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 19.0, 22.0, 3.6, 2]] predicted --> 2
[[1.0, -0.5, 4.3, 94.0, 92.0, 3.1, 1]] predicted --> 1
[[1.0, -0.5, 0.1, 25.0, 34.0, 2.8, 2]] predicted --> 2
[[1.0, -0.5, 9.0, 22.0, 79.0, 2.7, 1]] predicted --> 1
[[-1.0, -0.5, 0.1, 16.0, 18.0, 3.3, 1]] predicted --> 1
[[-1.0, -0.5, 0.2, 25.0, 18.0, 3.1, 1]] predicted --> 1
[[-1.0, -0.5, 0.2, 25.0, 58.0, 3.4, 1]] predicted --> 1
[[1.0, -0.5, 0.2, 46.0, 57.0, 4.3, 2]] predicted --> 2
[[-1.0, -0.5, 0.4, 22.0, 30.0, 4.1, 1]] predicted --> 1
92.95454545454545
 average=
93.18181818181819
```

Fig 5: Showing the third fold accuracy and average of accuracy.

In this above figure it is showing the accuracy of the third fold accuracy and average of the accuracy

V. CONCLUSION

Liver illness is a condition that is on the rise right now. Future increases in these liver illnesses will be caused by ongoing technological developments. Even if everyone is enrolling in yoga and dancing courses and people are becoming more health conscious, Luxuries and a sedentary lifestyle are constantly being increased to and enhanced. With the dataset we used for this research, our project will be very beneficial to society in such a situation. Therefore, we have presented the technology that uses machine learning to forecast liver illness. The system is implemented by KNN algorithm which trains the machine. The machine is able to predict the outcome for the testing dataset.