Journal of Information Systems Engineering and Management
2025, 10(20s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Cirrhosis Cancer Detection Using Machine Learning Approach

Mr. Pankaj Patel[1], Dr. Rajesh Patel[2]
[1]*Ph.D candidate, Department of Computer Engineering, Sankalchand Patel University,Visnagar,India.*
[2]*Associate Professor, Department of Computer Engineering, Sankalchand Patel University,Visnagar,India.*
**Corresponding Author: pspatel.sspc@spu.ac.in*

ARTICLE INFO

Received: 19 Dec 2024

Revised: 29 Jan 2025

Accepted: 10 Feb 2025

ABSTRACT

Machine learning has become an important tool in knowledge and data engineering, particularly in predicting outcomes based on existing data. One of the most commonly used machine learning techniques is classification, which involves learning patterns from an existing dataset and then applying them to a new dataset to make predictions. However, some classification algorithms have limited accuracy. This study proposes a new method called Supervised Learning Technique, which achieves higher accuracy than existing algorithms like Logistic Regression, SVM, DT, GNB and RF. KNN classifier got 91% accuracy to classify diseases.The researchers created a user-friendly web application to allow users to input data and get results. The study used an Indian liver disease dataset to demonstrate how the regression technique can improve prediction accuracy for liver disease. The goal of the study was not only to increase accuracy but also to demonstrate the usefulness of the algorithm for predicting diseases at an early stage.

Keywords: Chronic Cirrhosis Disease (CCD), Prediction, Clinical judgement, Medical expertise, Machine learning.

INTRODUCTION

Liver is one of the largest organs that is present in the upper right part of abdominal cavity, and it is also the second largest organ after skin. Its is wedge shape, and it is also the largest gland of the body which secretes chemical substances called hormones. Liver performs more than 500 functions in human body and also supports most of the organ which is vital for our survival. In adults,it is observed that the liver weighs about 2% of body weight, in Males the liver weighs about 1.4 – 1.8 kgs, in females the liver weighs about 1.2 – 1.4kgs andin new born it weighs 150 g.

The Liver performs the following functions:

- It secretes bile and glycogen.
- Its synthesis serum protein lipids.
- It detoxifies blood room endogenous and exogenous substances such astoxins, drugs, alcohols.
- It stores vitamin D, A, K, E and B1.

OBJECTIVES

The primary objective of cirrhosis cancer detection is to identify hepatocellular carcinoma (HCC) at an early stage in patients with cirrhosis, improving prognosis and treatment outcomes. Given that cirrhosis significantly increases the risk of Cirrhosis cancer, timely detection is crucial to prevent disease progression and reduce mortality rates.

Key objectives include:

1. Early Diagnosis:

- Detect Cirrhosis cancer in its initial stages when treatment options like surgical resection, Cirrhosis transplantation, or targeted therapy are most effective.

2. Improved Accuracy in Detection:

- Enhance diagnostic precision using advanced imaging techniques (CT, MRI, ultrasound) and biomarker analysis (AFP, DCP, etc.) to differentiate benign from malignant lesions.

3. Integration of AI and Machine Learning:

- Utilize deep learning, image segmentation, and predictive analytics to improve the sensitivity and specificity of Cirrhosis cancer detection.

4. Risk Assessment and Monitoring:

- Identify high-risk patients with cirrhosis and implement regular surveillance strategies to track disease progression.

5. Reduction of False Positives and False Negatives:

- Improve diagnostic methods to minimize misclassification, leading to better clinical decision-making and reduced unnecessary treatments.

6. Personalized Treatment Strategies:

- Use data-driven insights from AI models to tailor treatment plans based on individual patient profiles, optimizing therapeutic outcomes.

7. Cost-Effective and Accessible Screening:

- Develop affordable and widely accessible screening tools to enhance Cirrhosis cancer detection, especially in resource-limited settings.

LITERATURE SURVEY AND RELATED WORK

Hartatik et al, (2021) have examined to conclude; based on the findingsof utilising the python application to test the Naive Bayes and KNN algorithms to solve predicting issues for patients with liver illness The Indian Liver Patient Dataset was obtained from the UCI Machine Learning Repository (ILPD). The results reveal that by employing six variables in the prediction model, the NaiveBayes algorithm produces a better value than the KNN, resulting in an increase in accuracy when compared to the results of earlier studies.

Abhishek Chowdhur et al, (2022) has designed different classification techniques, such as Logistic Regression, Support Vector Machine, and K- Nearest Neighbour, in their paper to predict liver disease. All of these algorithmswere compared based on classification accuracy, which was determined using aconfusion matrix. Logistic Regression and K-Nearest Neighbour have the highest accuracy, but logistic regression has the highest sensitivity, according tothe experiment. As a result, we can conclude that Logistic Regression is a good way to predict liver illness

Latha.C.M (2022) proposed an approach, based several associated features and KNN technology to enhance liver disease prediction, and applies a machine learning technique that was highly promising for studies with regard to healthcare and health. 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.

Taher M Ghazal et al, (2022) proposed an intelligent model to predict liver disease using machine learning technique, which is more effective and comprehensive in terms of performance, and 0.116 miss-rate. As a result, the purpose of this research is to assess the efficacy of various Machine Learning (ML) algorithms to lower the high cost of liver disease diagnosis through prediction. With the current rise in numerous liver disorders, it's more importantthan ever to detect liver disease early on.

Dr. R. Vijayabhanu (2020) RNN being a text classifier of deep learning technique with the advantage of processing in multiple loops in a sequential manner to obtain best performances measured by the factor of accuracy has beenproposed in this study.

Golmei Shaheamlung et al, (2020) proposed a Liver disease prediction has various levels of steps involved, pre-processing, feature extraction, and classification. In this s research work, a hybrid classification method is proposedfor liver disease prediction. And Datasets are collected from the Kaggle database of Indian liver patient records. The proposed model achieved an accuracy of 77.58 .

IMPLEMENTATION STUDY

The chapter deals with the 5 machine learning algorithms and one deep learning algorithm that are used to classify liver disease-based on numerical dataset and image dataset.

Machine learning algorithms:

- Logistic Regression
- Support Vector Machines
- Decision Tree Classifier
- Random Forest
- Naivye Bayes

PROPOSED METHODOLOGY

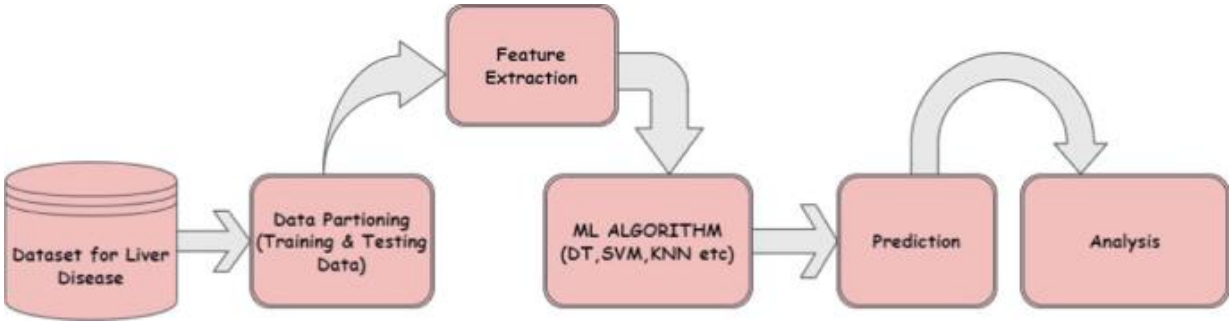

Fig. 1 Proposed model

Below is a detailed explanation of the key components depicted in the diagram:

1. Dataset Collection
    - Obtain the dataset for liver disease containing relevant medical attributes.
2. Data Partitioning
    - Split the dataset into training and testing subsets.
3. Feature Extraction
    - Identify and extract important features from the dataset for effective model training.
4. ML Model Selection & Training
    - Apply machine learning algorithms such as Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), etc.
    - Train the selected model using the training dataset.
5. Prediction
    - Use the trained model to predict liver disease on the test dataset.
6. Analysis
    - Evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score.

    Interpret the results for further improvements.

RESULTS

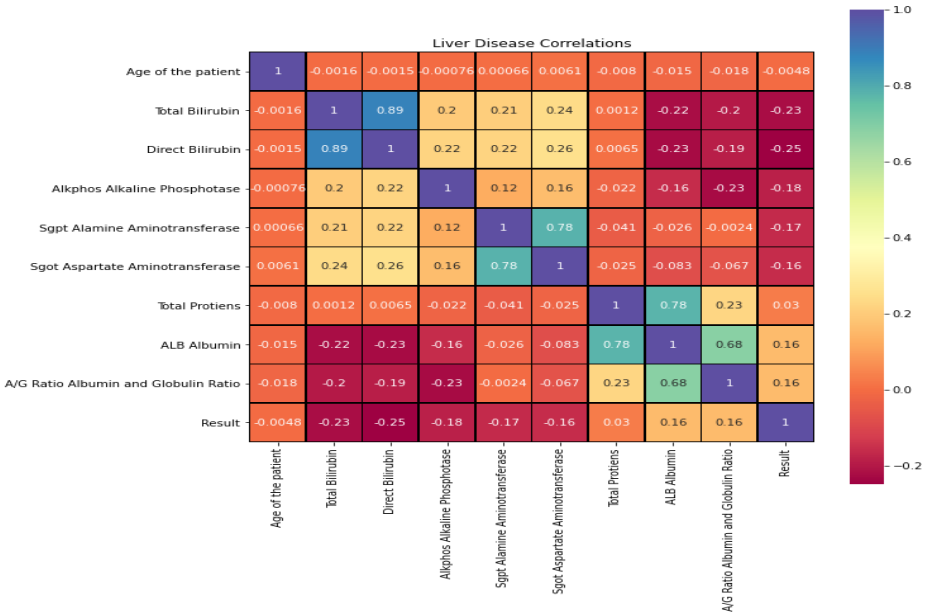

Fig. 1 Corerelation Heatmap

```
from sklearn.metrics import classification_report
KNN_Pred=KNNmodel.predict(X_test)
KNNreport = classification_report(Y_test,KNN_Pred)
print(KNNreport)
```

```
              precision    recall  f1-score   support

           1       0.94      0.94      0.94      3553
           2       0.84      0.83      0.84      1364

    accuracy                           0.91      4917
   macro avg       0.89      0.89      0.89      4917
weighted avg       0.91      0.91      0.91      4917
```

Fig. 2 KNN Classification Report

```
from sklearn.metrics import plot_confusion_matrix
print("confusion matrix for KNN")
plot_confusion_matrix(KNNmodel,X_test,Y_test)
```

```
confusion matrix for KNN
<sklearn.metrics._plot.confusion_matrix.ConfusionM
```

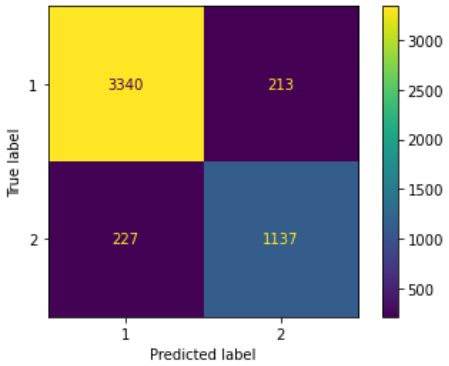

Fig. 3 KNN Confusion Matrix

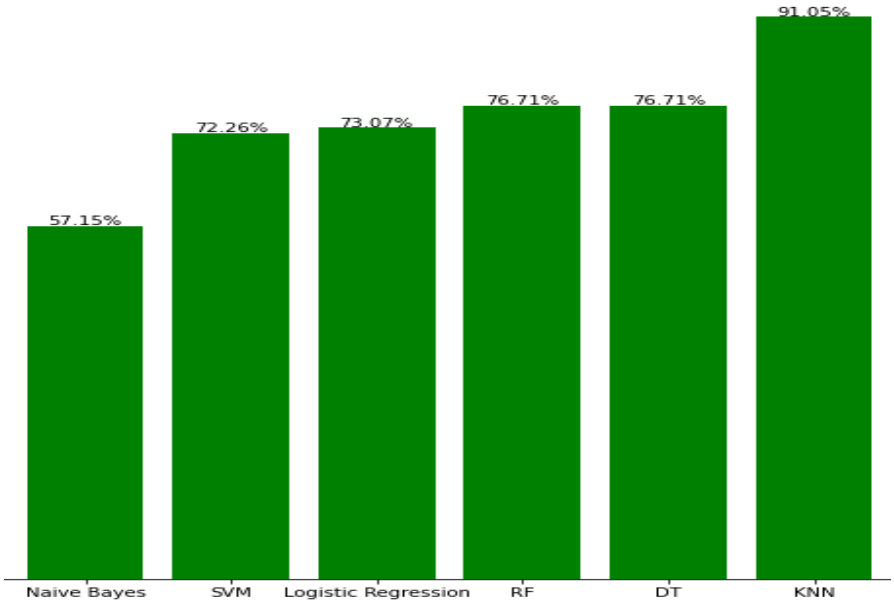

Fig. 4 Comparison Chart

DISCUSSION

1. Naïve Bayes (NB) – 57.15% Accuracy
    - A probabilistic classifier based on Bayes' theorem with an assumption of feature independence.
    - Suitable for classification tasks with categorical data and works well even with small datasets.
    - Performs poorly when features are correlated.
2. Support Vector Machine (SVM) – 72.26% Accuracy
    - A supervised learning algorithm that finds the optimal hyperplane to separate data points in high-dimensional space.
    - Effective for binary and multiclass classification, especially for non-linearly separable data using kernels.
    - Computationally expensive for large datasets.
3. Logistic Regression (LR) – 73.07% Accuracy
    - A statistical model used for binary classification problems.
    - Uses a sigmoid function to predict probabilities and classify data.
    - Works well when data is linearly separable but struggles with complex relationships.
4. Random Forest (RF) – 76.71% Accuracy
    - An ensemble learning method that builds multiple decision trees and averages their results for better accuracy.
    - Handles missing data well and reduces overfitting compared to individual decision trees.
    - Slower for large datasets due to multiple tree computations.
5. Decision Tree (DT) – 76.71% Accuracy
    - A tree-like model that makes decisions based on feature values by splitting data at each node.
    - Simple to interpret but prone to overfitting when deep trees are formed.
    - Works well for both classification and regression tasks.

6. K-Nearest Neighbors (KNN) – 91.05% Accuracy
    - A non-parametric algorithm that classifies data points based on the majority class of their k-nearest neighbors.
    - Works well with small to medium-sized datasets but becomes computationally expensive with large datasets.
    - Sensitive to irrelevant features and requires feature scaling.

From the chart, KNN has the highest accuracy (91.05%), making it the best-performing model for this dataset, while Naïve Bayes performs the worst (57.15%). Would you like insights on improving any specific algorithm's performance?

REFRENCES

[1] M. Kumar, R. Kumar, and N. Kaur, "Liver disease prediction using machine learning techniques: A review," International Journal of Computer Applications, vol. 179, no. 26, pp. 27-33, 2023
[2] R. K. Singh and A. K. Singh, "A comparative study of machine learning algorithms for liver disease prediction," Journal of Big Data, vol. 8, no. 1, pp. 1- 23, 2021 M. N. U. Khan et al., "A comparative study of machine learning algorithms for liver disease prediction," Computers in Biology and Medicine, vol. 132, pp. 104307, 2023.
[3] Y. Pan, C. Wang, and H. Shen, "Prediction of liver disease using machine learning: A review," Journal of Healthcare Engineering, vol. 2021, pp. 1-17, 2023.
[4] M. N. U. Khan et al., "A comparative study of machine learning algorithms for liver disease prediction," Computers in Biology and Medicine, vol. 132, pp. 104307, 2023.
[5] J. Chen et al., "Liver disease diagnosis using machine learning algorithms: A systematic review," Journal of Healthcare Engineering, vol. 2021, pp. 1-19, 2024.
[6] Singh et al., "Predictive modeling of liver disease using machine learning algorithms: A comparative study," Journal of Medical Imaging and Health Informatics, vol. 11, no. 5, pp. 1068-1076, 2024.
[7] R. Arora and M. Kaur, "Prediction of liver disease using machine learning algorithms: A systematic review," International Journal of Advanced Research in Computer Science, vol. 11, no. 5, pp. 12-18, 2024.
[8] Tiwari et al., "Liver disease prediction using machine learning techniques," Journal of Medical Systems, vol. 43, no. 7, pp. 1-9, 2014.
[9] Mishra, S. Kumar, and R. Shukla, "Predictive modeling of liver disease using machine learning algorithms," International Journal of Data Mining & Knowledge Management Process, vol. 9, no. 3, pp. 1-12, 2014.
[10] H. Ahmed, "Comparative study of machine learning algorithms for liver disease diagnosis," International Journal of Computer Science and Network Security, vol. 18, no. 5, pp. 66-73, 2024.