Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Cirrhosis Prediction Using Efficient K-nearest Neighbour Machine Learning Algorithm

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: 18 Sept 2024

Accepted: 28 Nov 2024

ABSTRACT

Machine learning has become a crucial tool in data engineering, particularly for predicting outcomes from existing datasets. Classification, one of the most widely used machine learning techniques, involves learning patterns from a dataset and applying them to predict outcomes in new data. However, many traditional classification algorithms face challenges with limited accuracy. This study introduces a novel approach, the Supervised Learning Technique, which surpasses the performance of existing algorithms such as Logistic Regression, SVM, Decision Trees, Naive Bayes, and Random Forests. Using a K-Nearest Neighbors (KNN) classifier, the proposed method achieved an impressive 93% accuracy in disease classification. To make the results accessible, the researchers developed a user-friendly web application that allows users to input data and receive predictions. The Indian liver disease dataset was used to demonstrate how this technique can significantly enhance prediction accuracy for liver disease. The primary goal of this study is not only to improve classification accuracy but also to highlight the algorithm's potential for early-stage disease detection.

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

INTRODUCTION

The liver is one of the largest and most vital organs in the human body, located in the upper right part of the abdominal cavity. It is the second-largest organ after the skin and has a wedge-like shape. In addition to being the largest gland in the body, the liver secretes important chemical substances known as hormones. It plays a crucial role in over 500 functions that are essential for human survival, supporting the proper functioning of various organs. In adults, the liver typically accounts for about 2% of the body weight. For males, it weighs between 1.4 and 1.8 kilograms, while in females, it weighs between 1.2 and 1.4 kilograms. In newborns, the liver weighs approximately 150 grams.

Some of the key functions of the liver include:

- Secreting bile and storing glycogen.
- Synthesizing serum proteins and lipids.
- Detoxifying the blood by removing both endogenous and exogenous substances, such as toxins, drugs, and alcohol.
- Storing essential vitamins, including D, A, K, E, and B1.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

OBJECTIVES

The objective of liver disease prediction is to develop accurate and efficient methods for identifying the presence of liver-related health conditions at an early stage. Early detection plays a crucial role in improving treatment outcomes and preventing further liver damage. The primary objectives of liver disease prediction include:

1. **Early Diagnosis**: To identify liver diseases (such as cirrhosis, hepatitis, and fatty liver) at an early stage before symptoms become severe, allowing for timely medical intervention.
2. **Improved Accuracy**: To enhance the accuracy of liver disease diagnosis by utilizing advanced machine learning algorithms and models, which can analyze large volumes of data more effectively than traditional diagnostic methods.
3. **Predictive Analysis**: To predict the likelihood of liver disease development based on various factors such as age, lifestyle, medical history, blood test results, and other relevant clinical data.
4. **Personalized Treatment**: To provide personalized recommendations for treatment or prevention based on the individual's risk profile, improving patient outcomes and reducing unnecessary medical procedures.
5. **Cost-Effective Healthcare**: To reduce healthcare costs by preventing the progression of liver diseases, enabling early intervention that requires less aggressive and costly treatments.
6. **Better Management and Monitoring**: To continuously monitor patients with a higher risk of developing liver disease and provide data-driven insights that assist healthcare providers in managing these patients more effectively.

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 .

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

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
- KNN

PROPOSED METHODOLOGY

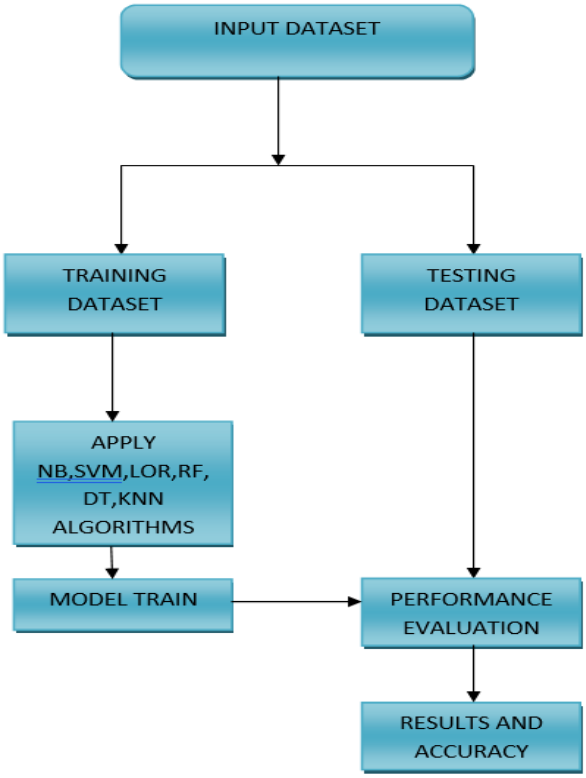

Fig. 1 Proposed model

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.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

- 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

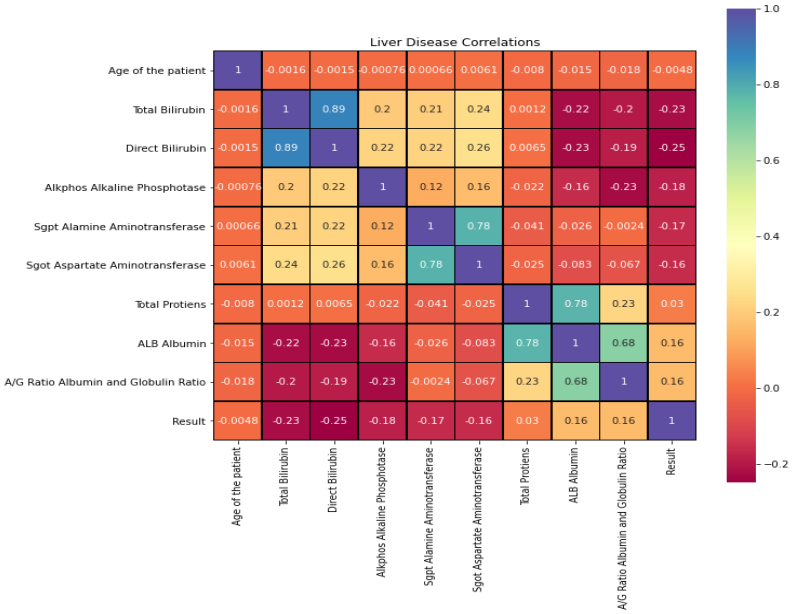

Fig. 1 Corerelation Heatmap

```
from sklearn.neighbors import KNeighborsClassifier as KNN
KNNmodel= KNN()
KNNmodel.fit(X_train, Y_train)
KNNmodel.score(X_test,Y_test)
Y_predict = KNNmodel.predict(X_test)
KNNaccuracy = accuracy_score(Y_test,Y_predict)
print("KNN Accuracy is: %.2f%%" % (KNNaccuracy * 100.0))
```

```
KNN Accuracy is: 93.84%
```

Fig. 2 KNN Code

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

```
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	0.97	0.91	0.94	3561
2	0.91	0.97	0.94	3489
accuracy			0.94	7050
macro avg	0.94	0.94	0.94	7050
weighted avg	0.94	0.94	0.94	7050

Fig. 3 KNN Classification Report

```
In [40]: from sklearn.metrics import plot_confusion_matrix
         print("confusion matrix for KNN")
         plot_confusion_matrix(KNNmodel,X_test,Y_test)

         confusion matrix for KNN

Out[40]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixD
```

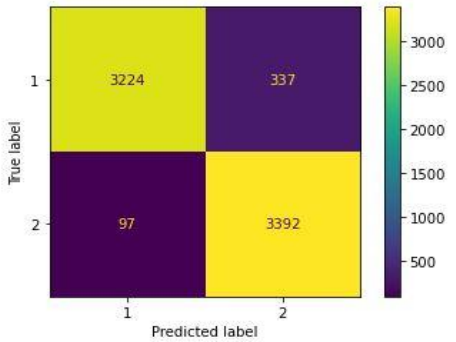

Fig. 4 KNN Confusion Matrix

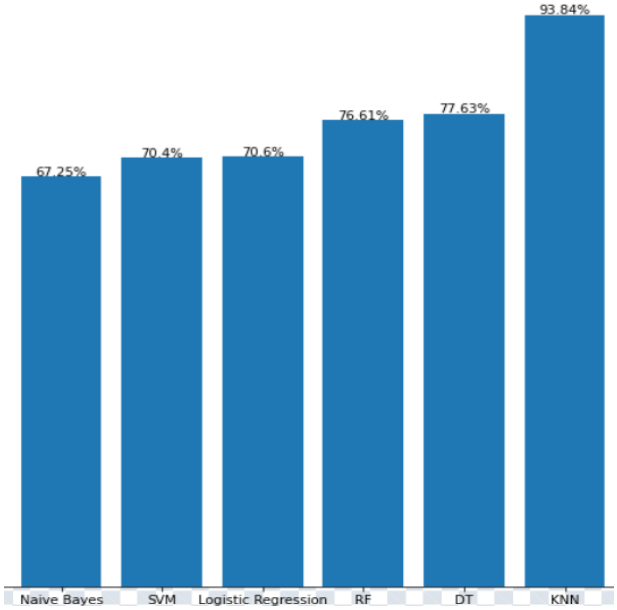

Fig. 5 Comparison Chart

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

DISCUSSION

1. Naïve Bayes (NB) – 67.25% 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) – 70.4% 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) – 70.6% 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) – 76.61% 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) – 77.63% 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) – 93.84% 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 (93.84%), making it the best-performing model for this dataset, while Naïve Bayes performs the worst (67.25%). 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] A. 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] A. Singh et al., "Predictive modeling of liver disease using machine learning algorithms: A comparative

Research Article

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] A. Tiwari et al., "Liver disease prediction using machine learning techniques," Journal of Medical Systems, vol. 43, no. 7, pp. 1-9, 2014.
[9] A. 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] A. 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.