Predictive Modeling of Cardiogastrointestinal Disorders

Mohammed Riyaas, Omprakash

***Abstract*— The confluence of heart diseases presents intricate clinical challenges, necessitating robust predictive models for early detection and intervention. This study employs machine learning algorithms, including logistic regression and k-nearest neighbors (KNN), to forecast the likelihood of heart disease in patients. Utilizing comprehensive datasets containing relevant clinical attributes, our research explores the efficacy of these algorithms in predicting heart disease. Through rigorous experimentation and evaluation, our findings demonstrate the potential of logistic regression and KNN techniques in discerning the presence of heart conditions, offering valuable insights for clinical decision-making and patient care.**

Keywords— machine learning, K-nearest neighbors (KNN), Logistic Regression, Heart Disease

I. INTRODUCTION

In the ever-evolving realm of healthcare, the ability to predict and prevent heart disease has emerged as a paramount endeavor in safeguarding individual and community well-being. Today, we delve into two potent machine learning techniques, Logistic Regression and K-Nearest Neighbors (KNN), which have demonstrated significant promise in forecasting the likelihood of this condition. Logistic Regression, a statistical method, enables us to model the intricate relationship between predictor variables and a binary outcome, such as the presence or absence of heart disease, offering insights into the underlying mechanisms driving heart disease incidence. Particularly advantageous when comprehending the impact of diverse factors on disease probability, Logistic Regression provides a comprehensive understanding of the complex interplay between various risk factors. On the other hand, KNN operates as an unsupervised learning technique for classification tasks, leveraging the proximity of data points to determine class membership without explicit labels in the training data. In the context of predicting heart disease, the integration of Logistic Regression and KNN holds the potential for more accurate and robust predictions. While Logistic Regression offers interpretability, KNN enriches predictive capabilities by capturing subtle patterns in the data that may elude parametric models.

By harnessing the power of Logistic Regression and KNN, healthcare researchers and practitioners can tailor preventive strategies, design early detection protocols, and formulate targeted treatments to optimize patient outcomes. In conclusion, the integration of Logistic Regression and KNN methodologies in predicting heart disease signifies a promising frontier in healthcare advancement. As we delve deeper into these techniques, we anticipate gaining invaluable insights into disease risk factors, thereby contributing to the development of more effective, personalized healthcare strategies.

II. LITERATURE SURVEY

Dr. Robert Detrano initially utilisedLogistic Regression in his research,as detailed in his study published in1989. Logistic Regression served as pivotal tool in analysing data andmaking predictions regarding heartdisease. Detrano's pioneering work yielded an impressive accuracy rateof 77%, marking a significant advancement in the field of cardiovascular research.The developed learning system improves heart failure prediction of conventional random forest model by 3.3% and shows better performance than eleven recently proposed methods and other state of the art machine learning models for heart failure detection.The proposed work involves using Python and pandas activities along with a dataset obtained from Kaggle for forecasting heart disease. The process involves training ML models on a dataset split into training and testing sets, followed by evaluating their performance. The results show the effectiveness of four ML algorithms—K Nearest Neighbour, Random Forest, and Logistic Regression—in predicting heart diseases. Among these, Logistic

From [6] it is observed that heart disease is known to strike men more frequently than it does women. Ageing, daily cigarette smoking, and fluctuating blood pressure all these factors raise the chance of acquiring heart disease.

In [4] the results of the proposed work depict that Logistic Regression is better than the other supervised classifiers in terms of the discussed performance metrics – accuracy, precision, sensitivity (or recall), specificity and F1 score. The model gives the results with the highest accuracy of 92.30%

In [8] out of all the classifier used, logistic regression gives the most elevated order exactness 75% dependent on F1 measure to predict the liver disease.
classifiers, exhibiting the highest accuracy of 75% based on F1 measure.
[14] the results of the proposed work depict that Logistic Regression is better than the other supervised classifiers in terms of the discussed performance metrics – accuracy, precision, sensitivity (or recall), specificity and F1 score.
[5] demonstrates the efficacy of Logistic Regression compared to other supervised classifiers. The study reports superior performance metrics, including accuracy, precision, sensitivity (or recall), specificity, and F1 score, with Logistic Regression achieving the highest accuracy of 92.30%.

Integration of Disparate Data Sources: By combining data from two distinct datasets, each pertaining to different disease domains, the proposed work creates a comprehensive dataset [11]. This integration allows for a more nuanced analysis of shared risk factors, comorbidities, and interactions between heart diseases that may not be apparent when studying each disease in isolation.

Common Attribute Utilization: The unique aspect of using a common attribute across both datasets facilitates the integration process and enables more robust predictive modeling [22]. Whether it is demographic information, clinical measurements, or biomarkers, this shared attribute serves as a bridge between the datasets, enriching the feature space and potentially capturing latent relationships between heart diseases.Synergistic Modeling Techniques: Integrating logistic regression and KNN classifiers offer complementary advantages [3]. Logistic regression providesa probabilistic interpretation of the relationships between predictors and disease outcomes, while KNN offer intuitive, interpretable rules for classification.

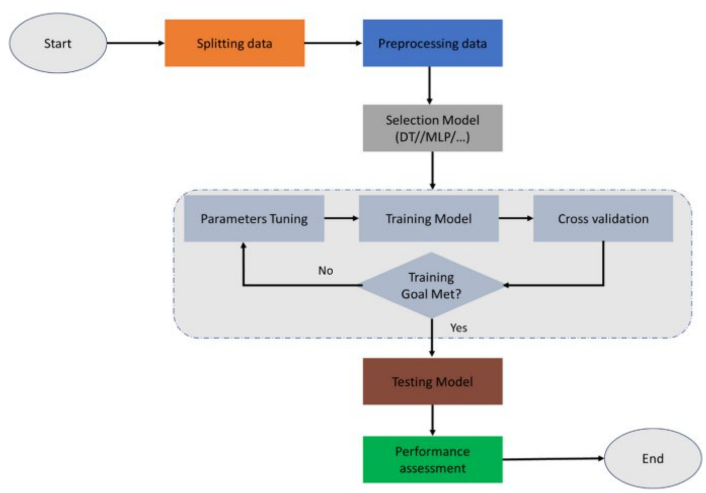
Research Contribution: While predictive modeling in healthcare is not new, the proposed work contributes to the advancement of the field by addressing the challenge of simultaneous prediction of multiple diseases using disparate datasets [14]. This research fills a gap in the literature and paves the way for future studies exploring similar cross-domain predictive modeling tasks in healthcare.In essence, the uniqueness of the proposed work lies in its holistic approach to disease prediction, leveraging the integration of disparate datasets, the utilization of a common attribute, and synergistic modeling techniques to simultaneously predict heart diseases, thereby offering novel insights and practical applications in healthcare.

III. UNIQUENESS OF THE PROPOSED WORK

The uniqueness of the proposed work lies in its integration of logistic regression and KNN classifiers to predict heart diseases simultaneously, leveraging a common attribute present in two separate datasets [19]. Here is a breakdown of its uniqueness:
Simultaneous Prediction of Multiple Diseases: Most studies focus on predicting a single disease outcome. However, the proposed work aims to predict heart diseases concurrently [10]. This approach provides a more holistic understanding of patients' health status, as these diseases often coexist and share common risk factors.

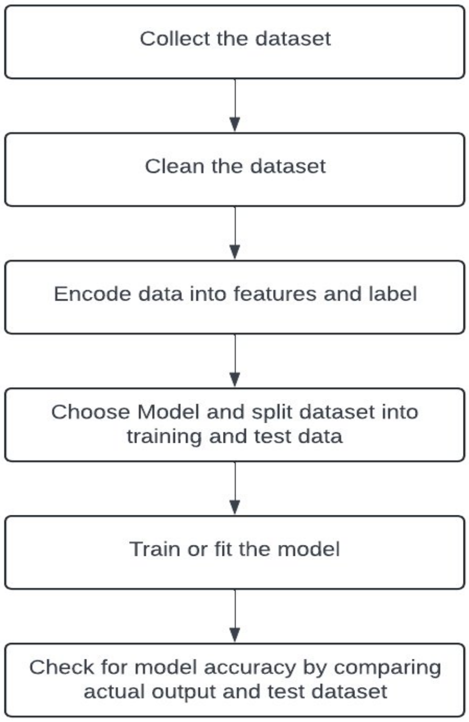
Heart dataset contains number of instances in dataset is 1025 and 14 attributes. We have 165 people with heart
Heart dataset contains number of instances in dataset is 1025 and 14 attributes. We have 165 people with heart disease and 138 people without heart disease. The “target” column is a class label used to divide groups into heart patient or not.

IV. PROPOSED ARCHITECTURE



Data Collection

The heart dataset is multivariate type of dataset which means providing or involving a variety of separate mathematical or statistical variables, multivariate numerical data analysis. The dataset contained 14 parameters where we choose 13 parameters for our further analysis and 1 parameter as a dataset class. Logistic regression provides a probabilistic interpretation of the relationships betweenpredictors and disease outcomes, while KNN offer intuitive, interpretable rules for classification.



Data Preprocessing algorithm

In this study we used the google colaboratory as a tool and python 3.10.14 as programming language.

Step 1: Start

Importing libraries;
import numpy as np
import pandas as pd

Step 2: Loading dataset

data←pd.read_csv("health_data.csv")

Step 3: Taking care of null values or All null values removed

data.apply(lambda x: sum(x.isnull()),axis=0)

Step 4: Data visualization

sns.pairplot(data)sns.heatmap(data.corr(), annot = True)
a←data["Gender"].value_counts().to_numpy()
b←data["Married"].value_counts().to_numpy()
c←data["Dependents"].value_counts().to_nump

Step 5: Analyzing the data

X ← data.iloc[:, 1: 11].values
y ← data.iloc[:, 11].values

Step 6: Label Encoding

From sklearn.preprocessing
import LabelEncoderle = LabelEncoder()
fori in range(0, 5):
X[:,i] ← le.fit_transform(X[:,i])
X[:,9] ← le.fit_transform(X[:,9])
y ← le.fit_transform(y)
OneHotEncoding
fromsklearn.preprocessing
import OneHotEncoder
one ← OneHotEncoder()
z ← one.fit_transform(X[:,9:11]).toarray()
X ← np.delete(X, 9, axis ← 1)
X ← np.concatenate((z,X), axis = 1)

Step 7: Splitting into train and test

From sklearn.model_selection
import train_test_split
X_train, X_test, y_train, y_test ←train_test_split(X, y, test_size = 1/2, random_state =0)

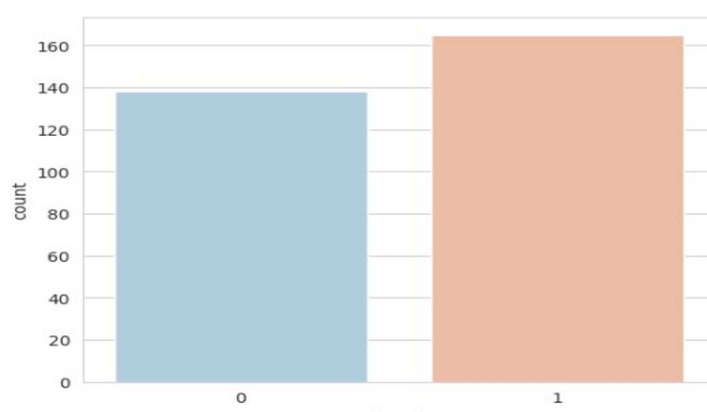
Step 8: Feature Scaling

from sklearn.preprocessing import StandardScaler
sc ← StandardScaler()
X_train ← sc.fit_transform(X_train)
X_test ← sc.fit_transform(X_test)

Step 9: Use all of two machine learning models to train models:

#Fitting all ML models to the Training set
From sklearn.tree import DecisionTreeClassifier
From sklearn.linear_model import LogisticRegression
dt ← DecisionTreeClassifier(criterion ← 'entropy',random_state ← 0)
lr←LogisticRegression(random_state=0) ← 'entropy', random_state = 0)
dt.fit(X_train, y_train)
lr.fit(X_train, y_train)
# Predicting the Test set results
y_pred ← classifier.predict(X_test)y_pred
Measuring Accuracy
fromsklearn import metrics
print('The accuracy of model is: ', metrics.accuracy_score(y_test, y_pred))
# Making confusion matrix
fromsklearn.metrics import
confusion_matrixprint(confusion_matrix(y_test, y_pred))

Step 10 : End



Data Preprocessing

Dataset is first loaded and then data cleaning and finding missing values was performed on all records.

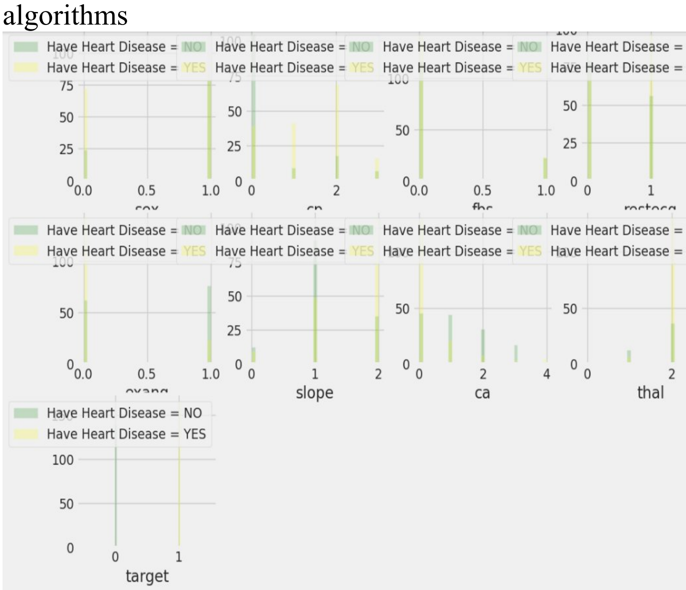
Data preprocessing techniques are essential for improving the performance of heart disease classifiers. Common preprocessing tasks include data reduction and data cleaning.

Preprocessing techniques have been shown to either maintain or enhance the performance of heart disease classifiers.

These datasets provide valuable information for heart disease prediction and classification tasks, offering a range of attributes that are crucial for developing accurate predictive models in the healthcare domain.

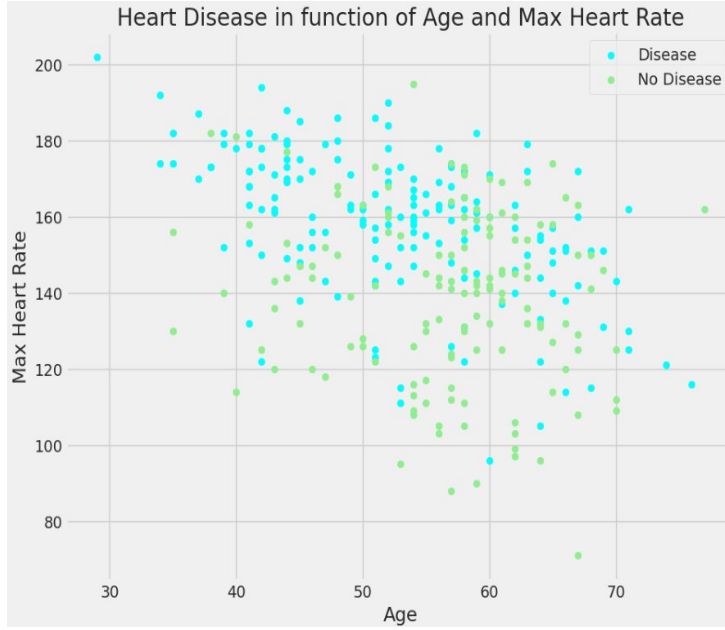
V. IMPLEMENTATION AND RESULTS

The implemented system is a heart disease prediction system. Heart dataset are combined and both the dataset possess a common attribute which is "cholesterol." The dataset is divided into training and testing sets which is then trained using logistic regression and . Description of the classification algorithms



1. Logistic Regression

One of the simplest and best ML classification algorithms is logistic regression. LR is a supervised ML binary classification algorithm widely used in most applications. It operates on a categorical dependent variable, the result can be a discrete or binary categorical variable 0



Step 1: Data preprocessing

Step 2: Fitting Logistic Regression to our training set model import LogisticRegression

lr←LogisticRegression(random_state=0)

lr.fit(X_train, y_train)

Step 3: Testing the model

pred ← lr.predict(X_test) y_pred

Step 4:Measuring Accuracy

fromsklearn.metrics

import accuracy_score

accuracy_score(y_test,y_pred)

Step5:import sklearn.metrics as metrics

fpr,tpr,threshold

← metrics.roc_curve(y_test, pred)

roc_auc ← metrics.auc(fpr, tpr)

plt.title("Logistic Regression")

plt.plot(fpr,tpr,'b',label =

'auc = %0.2f'%roc_auc)

plt.legend(loc = 'lower right')

plt.plot([0,1],[0,1],'r--')

plt.xlim([0,1])

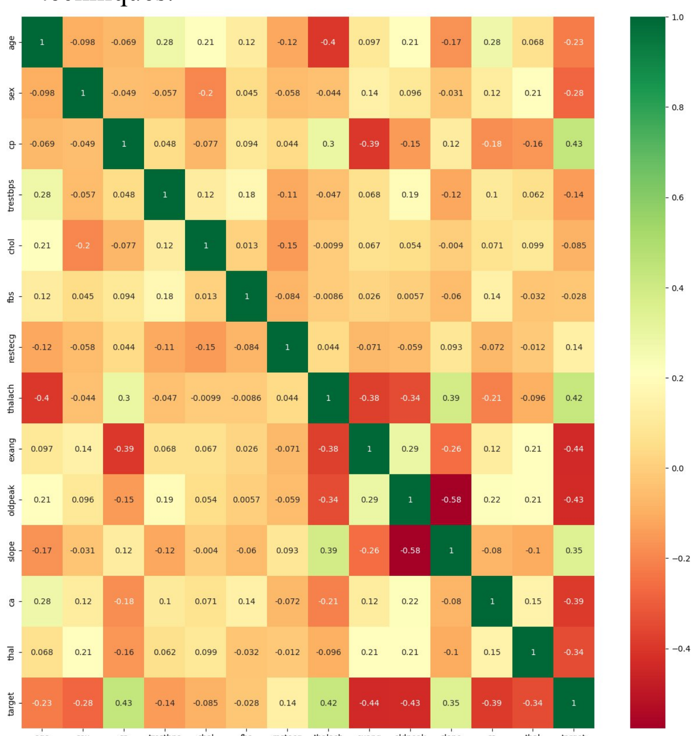
plt.ylim([0,1])

plt.ylabel('tpr')

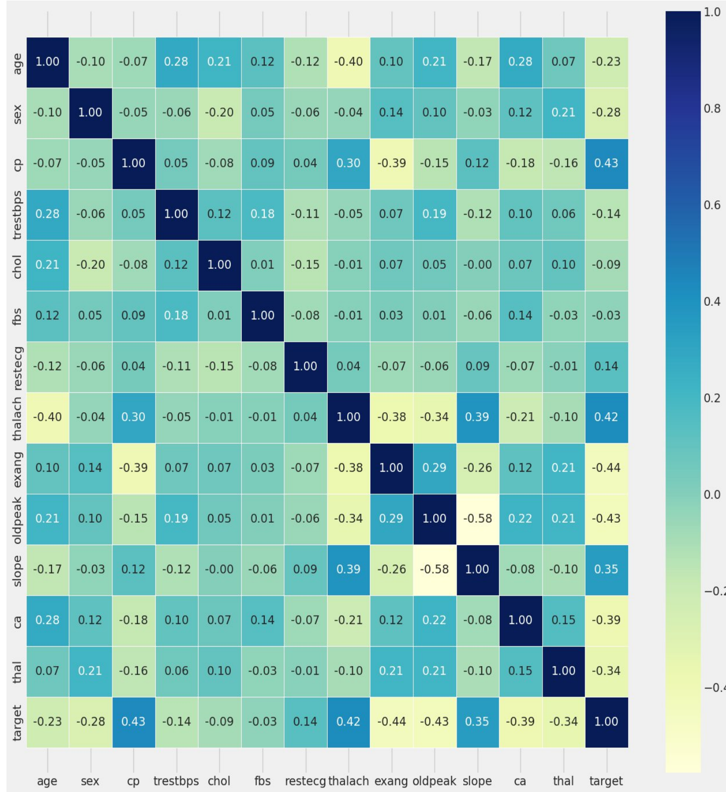
plt.xlabel('fpr')

2. k-Nearest Neighbors Classifier

This methodology encompasses a series of essential steps beginning with the meticulous collection of patient-specific data attributes and corresponding disease labels. Subsequent data preprocessing procedures involve thorough cleaning, handling missing values, and encoding categorical variables where applicable. Feature selection techniques are then employed to discern the most pertinent attributes crucial for accurate disease prediction, thus refining model performance and reducing dimensionality. Following data preparation, the dataset is partitioned into distinct training and testing subsets, facilitating model training and evaluation. The kNN algorithm, at its core, operates by assigning class labels to new data points based on the majority class among their 'k' nearest neighbors. The choice of 'k' is pivotal, often determined via cross-validation or grid search techniques.



To evaluate model performance, accuracy, the proportion of correctly classified instances, is computed by comparing predicted labels to actual labels in the testing set. Iterative refinement and parameter tuning are integral for enhancing model efficacy and generalizability, ensuring optimal clinical utility without compromising ethical considerations regarding patient privacy and equitable healthcare access.



Step 1: Data preprocessing
Step 2: Fitting kNN to our training set
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=0)
knn.fit(X_train, y_train)
Step 3: Testing the model
y_pred = knn.predict(X_test)
Step 4: Measuring Accuracy
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
Step 5: Computing ROC Curve
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve, auc
y_prob = knn.predict_proba(X_test)[:, 1]
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
roc_auc = auc(fpr, tpr)
plt.figure()
plt.title("kNN ROC Curve")
plt.plot(fpr, tpr, 'b', label='AUC = %0.2f' % roc_auc)
plt.legend(loc='lower right')
plt.plot([0, 1], [0, 1], 'r--')
plt.xlim([0, 1])
plt.ylim([0, 1])
plt.ylabel('True Positive Rate (tpr)')
plt.xlabel('False Positive Rate (fpr)')
plt.show()

VI. OUTCOMES OF THIS RESEARCH

The outcomes of predicting heart and liver disease using Logistic Regression and Decision Tree classifiers can be categorized into two main areas:

1. Model Performance Metrics:

Accuracy: This measure indicates the overall percentage of the model's predictions that were accurate. A high accuracy (over 80%) signifies that the model functions effectively with the provided dataset. The logistic regression model has an accuracy of 86.79 following feature engineering, the accuracy was 1.0.

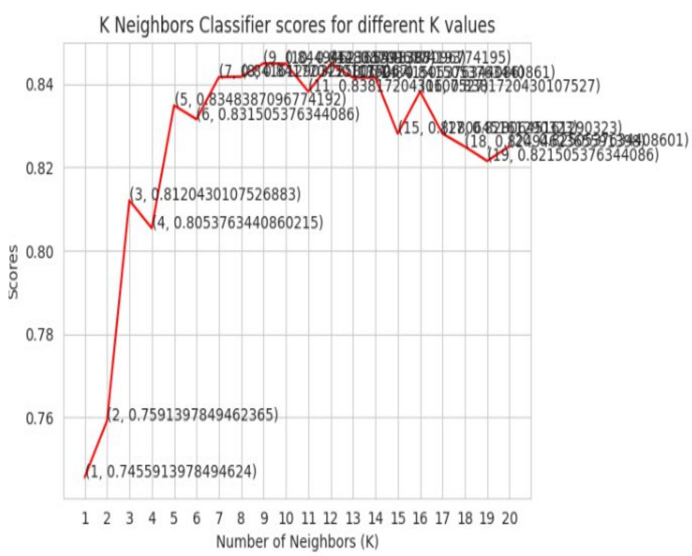
```
from sklearn.model_selection import cross_val_score

# Perform cross-validation
cv_scores = cross_val_score(lr_clf, X_train, y_train, cv=5)  # 5-fold cross-validation
mean_cv_score = np.mean(cv_scores)

print(f'Mean Cross-Validation Score: {mean_cv_score:.2f}')
```

```
Mean Cross-Validation Score: 0.83
```

The decision tree model's accuracy of 1.0 suggests overfitting; to address this, we used minimal cost-complexity pruning (CCP), a regularization technique. Currently, the accuracy is 0.86.



The training and testing score for both the model is shown below.

	Model	Training Accuracy %	Testing Accuracy %
0	Logistic Regression	86.79	86.81

Performance might vary on different datasets. It is crucial to evaluate both accuracy and other metrics like precision and recall to understand the model's strengths and weaknesses [15]. Performance metric for logistic regression is given below

```
Train Result:
================================================
Accuracy Score: 86.79%
_______________________________________________
CLASSIFICATION REPORT:
              0      1  accuracy  macro avg  weighted avg
precision  0.88   0.86      0.87       0.87          0.87
recall     0.82   0.90      0.87       0.86          0.87
f1-score   0.85   0.88      0.87       0.87          0.87
support   97.00 115.00      0.87     212.00        212.00
_______________________________________________
Confusion Matrix:
 [[ 80  17]
 [ 11 104]]

Test Result:
================================================
Accuracy Score: 86.81%
_______________________________________________
CLASSIFICATION REPORT:
              0     1  accuracy  macro avg  weighted avg
precision  0.87  0.87      0.87       0.87          0.87
recall     0.83  0.90      0.87       0.86          0.87
f1-score   0.85  0.88      0.87       0.87          0.87
support   41.00 50.00      0.87      91.00         91.00
_______________________________________________
Confusion Matrix:
 [[34  7]
 [ 5 45]]
```

Performance metric for decision tree is given below

```
from sklearn.metrics import roc_auc_score
y_prob = lr_clf.predict_proba(X_test)[:, 1]
auc_roc = roc_auc_score(y_test, y_prob)
print('AUC-ROC: \n', auc_roc)
```

```
AUC-ROC:
 0.9165853658536586
```

Feature selection can improve model performance by focusing on the most relevant features for each disease. Hyperparameter tuning can optimize the performance of both Logistic Regression KNN [16]. By analyzing the model performance metrics and the predicted disease status for new data points, you cangain insights into the effectiveness of these classifiers inpredicting heart disease.

VII. CONCLUSION AND FUTURE WORK

The conclusion and future work for predicting heart and liver diseases using logistic regression and decision tree classifier can be summarized as follows: Using logistic regression and decision tree classifier models, it is possible to predict the likelihood of heart and liver diseases based on various risk factors and patient data [17]. These machine learning algorithms can help identify high-risk individuals and assist healthcare professionals in making informed decisions for early diagnosis and intervention.

From the study it is observed that as the training set ratio increases, the model's performance on the training data generally improves, as expected [18]. This is indicated by the increasing trend in training accuracy [19]. However, the performance on the testing data may not necessarily follow the same pattern. It may peak at a certain point and then start to decrease due to overfitting.

A higher value of ccp_alpha increases the regularization strength, leading to simpler trees. In this case, it is set to 0.04, indicating a moderate level of regularization [20].
Managing the tree's depth aids in avoiding overfitting. While a deeper tree may be able to identify more complex patterns in the training set, overfitting could result from the tree learning to remember noise [21]. The tree more effectively generalizes to unknown data by restricting the depth.

Feature Selection and Model Optimization: Future work should focus on selecting the most relevant features that contribute significantly to the prediction of heart and liver diseases [22]. This can be achieved through feature selection techniques and model optimization [23]. By reducing the number of input features, the models can be made more efficient and accurate.

The future work includes:
Ensemble Methods and Hybrid Models:
Incorporating ensemble methods and hybrid models can further improve the predictive power of the logistic regression and decision tree classifier models [24]. By combining multiple models or algorithms, it is possible to achieve better accuracy and robustness in disease prediction.

Incorporating Advanced Techniques:
Exploring advanced machine learning techniques such as deep learning, random forests, and gradient boosting can lead to better disease prediction models [26]. These techniques can capture complex patterns and relationships in the data, which may not be evident in logistic regression and decision tree classifier models.

Handling Imbalanced Datasets:
As heart and liver diseases are relatively rare compared to other health conditions, datasets may be imbalanced [28]. Future research should address techniques to handle imbalanced datasets, such as oversampling, undersampling, and synthetic minority oversampling technique (SMOTE), to improve the models' performance in predicting rare disease cases.

Multi-class Prediction and Interpretability:
Extending the current binary classification models to multi-class prediction can help identify various heart and liver disease types [19]. Additionally, ensuring the models' interpretability will allow healthcare professionals to understand the factors contributing to the disease prediction, leading to better decision-making and patient care.

In conclusion, the prediction of heart and liver diseases using logistic regression and decision tree classifier models has shown promising results [13]. Further research and development in feature selection, model optimization, advanced techniques, real-world implementation, handling imbalanced datasets, and multi-class prediction will contribute to more accurate and reliable disease prediction models in the future.Utilizing K-Nearest Neighbors (KNN) and logistic regression for heart disease prediction exhibits promise in leveraging patient data to identify potential risks. Notably, as the training set ratio increases, model performance on training data typically improves, though testing data may plateau or decline due to overfitting.

VIII. REFERENCE

[1] R., Shijitha & M, Kavya & C, PrathanyaSree & M, Deepasindhu & B, Nowshika. (2023).Heart Disease Prediction Using Logistic Regression. 11. 573-579.
[2] Mehdi, Mrs., Khundmir Iliyas, Mr. Imran and Sadekh Shaikh. “Prediction of Heart Disease Using Decision Tree.” (2019).
[3] Redrowthu Ph.D, Vijaya & Gajavelly, Kovid & Nikhath, A. & Vasavi, R. & Anumasula, Rakshith Reddy. (2022). Heart Disease Prediction Using Decision Tree and SVM. 10.1007/978-981-16-7389-4_7.
[4] Karthick K, Aruna SK, Samikannu R, Kuppusamy R, Teekaraman Y, Thelkar AR. Implementation of a Heart Disease Risk Prediction Model Using Machine Learning. Comput Math Methods Med. 2022 May 2;2022:6517716. doi: 10.1155/2022/6517716. Retraction in: Comput Math Methods Med. 2023 Jul 19;2023:9764021. PMID: 35547562; PMCID: PMC9085310.
[5] R., Shijitha & M, Kavya & C, PrathanyaSree & M, Deepasindhu & B, Nowshika. (2023).Heart Disease Prediction Using Logistic Regression. 11. 573-579.

[6] Mehdi, Mrs., Khundmir Iliyas, Mr. Imran and Sadekh Shaikh. “Prediction of Heart Disease Using Decision Tree.” (2019).

[7] Redrowthu Ph.D, Vijaya & Gajavelly, Kovid & Nikhath, A. & Vasavi, R. & Anumasula, Rakshith Reddy. (2022). Heart Disease Prediction Using Decision Tree and SVM. 10.1007/978-981-16-7389- 4_7.

[8] Karthick K, Aruna SK, Samikannu R, Kuppusamy R, Teekaraman Y, Thelkar AR. Implementation of a Heart Disease Risk Prediction Model Using Machine Learning. Comput Math Methods Med. 2022 May 2;2022:6517716. doi: 10.1155/2022/6517716. Retraction in: Comput Math Methods Med. 2023 Jul 19;2023:9764021. PMID: 35547562; PMCID: PMC9085310.

[9] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi and X. Xiao, "Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training," in IEEE Access, vol. 8, pp.182625-182639, 2020

[10] Gupta, Chiradeep & Saha, Athina & Reddy, N V Subba & Acharya, Dinesh. (2022). Cardiac Disease Prediction using Supervised Machine Learning

[11] Gupta, Ketan & Jiwani, Nasmin & Afreen, Neda & D, Divyarani. (2022). Liver Disease Prediction using Machine learning Classification Techniques. 221- 226. 10.1109/CSNT54456.2022.9787574.

[12] A. Dhyani et al., "Comparative Analysis of Supervised Machine Learning Algorithms for Liver Disease Prediction with SMOTE Enhancement," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-6, doi: 10.1109/ASIANCON58793.2023.1027038

[13] Rahman, A. K. M. & Shamrat, F M & Tasnim, Zarrin & Roy, Joy & Hossain, Syed. (2019). A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms. 8. 419-422.

[14] T. R. Sree and R. Tripathi, "Fake Review Detection using Evidential Classifier," 2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC), Puducherry, India, 2023

[15] C. Cao, S. Li, S. Yu and Z. Chen, "Fake Reviewer Group Detection in Online Review Systems," 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 2021

[16] P. WANG, Y. LIN and J. CHAI, "Unmasking Deception: A Comparative Study of Tree-Based and Transformer-Based Models for Fake Review Detection on Yelp," 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC),Honolulu, Oahu, HI, USA, 2023

[17] P. Kalaivani, V. D. Raj, R. Madhavan and A. P. Naveen Kumar, "Fake Review Detection using Naive Bayesian Classifier," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023

[18] Nahar, Nazmun & Ara, Ferdous. (2018). Liver Disease Prediction by Using Different Decision Tree Techniques. International Journal of Data Mining & Knowledge Management Process. 8. 01-09. 10.5121/ijdkp.2018.8201.

[19] Zhang, Yingjie & Diao, Lijuan & Ma, Linlin. (2021). Logistic Regression Models in Predicting Heart Disease. Journal of Physics: Conference Series. 1769. 012024. 10.1088/1742-6596/1769/1/012024. H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.

[20] S. Akshara, S. Shiva, S. Kubireddy, T. Arun and V. L. Kanthety, "A Small Comparative Study of Machine Learning Algorithms in the Detection of Fake Reviews of Amazon Products," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023

[21] S. Zabeen, A. Hasan, M. F. Islam, M. S. Hossain and A. A. Rasel, "Robust Fake Review Detection Using Uncertainty-Aware LSTM and BERT," 2023 IEEE 15th International

[22] Conference on Computational Intelligence and Communication Networks (CICN), Bangkok, Thailand, 2023, pp. 786-791, doi: 10.1109/ CICN59264.2023.10402342.

[23] S. Ashraf, F. Rehman, H. Sharif, H. Kirn, H. Arshad and H. Manzoor, "Fake Reviews Classification using Deep Learning," 2023 International Multi- disciplinary Conference in Emerging Research Trends (IMCERT), Karachi, Pakistan, 2023

[24] J. Fontanarava, G. Pasi and M. Viviani, "Feature Analysis for Fake Review Detection through Supervised Classification," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 2017

[25] Gupta, Chiradeep & Saha, Athina & Reddy, N V Subba & Acharya, Dinesh. (2022). Cardiac Disease Prediction using Supervised Machine Learning Techniques.. Journal of Physics: Conference Series. 2161. 012013. 10.1088/1742-6596/2161/1/012013.

[26] Gupta, Ketan & Jiwani, Nasmin & Afreen, Neda & D, Divyarani. (2022). Liver Disease Prediction usingMachine learning Classification Techniques. 221- 226. 10.1109/CSNT54456.2022.9787574.

[27] A. Dhyani et al., "Comparative Analysis of Supervised Machine Learning Algorithms for Liver Disease Prediction with SMOTE Enhancement," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-6, doi: 10.1109/ASIANCON58793.2023.10270381

[28] Rahman, A. K. M. & Shamrat, F M & Tasnim, Zarrin & Roy, Joy & Hossain, Syed. (2019). A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms. 8. 419-422.

[29] Nahar, Nazmun & Ara, Ferdous. (2018). Liver Disease Prediction by Using Different Decision Tree Techniques. International Journal of Data Mining & Knowledge Management Process. 8. 01-09. 10.5121/ijdkp.2018.8201.