VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI-590018, KARNATAKA.

A PROJECT SYNOPSIS

On

"Heart Disease Prediction Using Novel Quine McCluskey Binary Classifier"

Submitted in Partial Fulfillment for the requirement of VII Semester

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By:

Preetham S G	**1SG20CS082**
Santhosh D	**1SG20CS094**
Shravan L	**1SG20CS100**
Vivan Kushal Heneger	**1SG20CS114**

Under the Guidance of

Prof.Lavanya K

Assistant Professor

Department of Computer Science and Engineering

(Accredited by NBA)

SAPTHAGIRI COLLEGE OF ENGINEERING

(Affiliated to Visvesvaraya Technological University, Belagavi& Approved by AICTE, New Delhi.)

ISO 9001-2015 & 14001-2015 Certified, Accredited by NAAC with 'A' Grade

14/5,Chikkasandra, Hesarghatta Main Road

Bengaluru–560057.

2023-2024

TITLE OF THE PROJECT:

Heart Disease Prediction Using Novel Quine McCluskey Binary Classifier.

ABSTRACT:

Cardiovascular disease is the primary reason for mortality worldwide, responsible for around a third of all deaths. To assist medical professionals in quickly identifying and diagnosing patients, numerous machine learning and data mining techniques are utilized to predict the disease. s. In this study, we introduce a new ensemble Quine McCluskey Binary Classifier (QMBC) technique for identifying patients diagnosed with some form of heart disease and those who are not diagnosed. The QMBC model utilizes an ensemble of seven models, including logistic regression, decision tree, random forest, K-nearest neighbour, naive bayes, support vector machine, and multilayer perceptron, and performs exceptionally well on binary class datasets. We employ feature selection and feature extraction techniques to accelerate the prediction process.

INTRODUCTION

- The term ‘‘Heart Disease’’ (HD) is used to refer to a variety of pathological disorders that have an impact on the heart and blood vessels. It encompasses a variety of heart-related conditions, including but not limited to vascular diseases and disturbances in heart rhythm.
- The major contributions of this project are discussed below:
  - The first step of this work involves preprocessing the dataset, followed by applying FS and FE techniques to optimize computation time.
  - Specifically, we utilized the Chi-Square and ANOVA techniques to eliminate irrelevant and redundant attributes and create subsets of features.
  - We then applied the PCA FE technique to extract prime components from these subsets, further improving the efficiency of our model.
  - We introduced a novel ensemble technique called Quine McCluskey Binary Classifier (QMBC) that combines predictions from multiple models to predict patients diagnosed and not diagnosed with some form of HD.

STATEMENT OF THE PROBLEM

Heart disease is one of the most common causes of death around the world nowadays. Often, the enormous amount of information is gathered to detect diseases in medical science. All of the information is not useful but vital in taking the correct decision. Thus, it is not always easy to detect the heart disease because it requires skilled knowledge or experiences about heart failure symptoms for an early prediction. Most of the medical dataset are dispersed, widespread and assorted.

WHY IS THE PARTICULAR TOPIC CHOSEN?

As per the World Health Organization (WHO), it is the deadliest and most devastating disease, taking over 18 million in lives a year. To diagnose it, healthcare professionals rely on a patient's medical history and various tests, such as blood pressure, blood sugar, and cholesterol tests. Additionally, modern medical procedures like electrocardiograms, exercise stress tests, X-rays, echocardiography, coronary angiography, radionuclide tests, MRI scans, and CT scans can aid in the identification of cardiac condition. Heart failure is the result of chronic issues that damage or weaken the heart muscles, leading to reduced ejection fraction. It is a condition that can affect both adults and children and cause severe damage to other vital organs in the body. To detect heart disease early we have chosen this topic.

OBJECTIVE AND SCOPE OF THE PROJECT

Objective:

- To build an efficient model to predict heart disease.
- To reduce the execution time .
- To improve accuracy of heart disease prediction.
- To compare the various machine learning algorithms to detect heart disease.

Scope:

Using machine learning to classify cardiovascular disease occurrence can help diagnosticians reduce misdiagnosis. This research develops a model that can correctly predict cardiovascular diseases to reduce the fatality caused by cardiovascular diseases.

METHODOLOGY

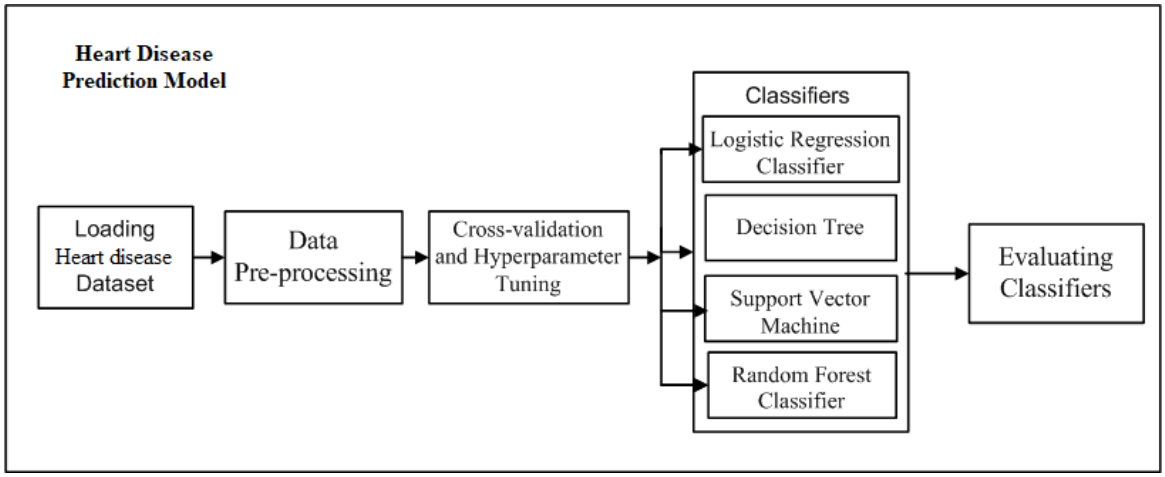

Fig 1.1 Workflow of Project

- The system architecture shows the Heart disease prediction.

- First user should load the UCI heart disease dataset, System will preprocess the dataset tune the feature parameters and pass to the machine learning algorithm models and predict heart disease and shows the performance evaluation between various ML algorithms.

POSSIBLE OUTCOMES

The present study investigates the performance of seven standalone ML models and a voting classifier using the Cleveland dataset, cardiovascular dataset, and HD dataset (Comprehensive). All datasets are preprocessed to ensure the suitability of ML models. To minimize dimensions, computational speed, and remove irrelevant and duplicate features from the dataset, the study utilizes Chi-Square and Anova techniques with PCA FE strategy. Furthermore, a novel Quine McCluskey Binary Classifier (QMBC) is proposed, which ensembles seven standalone ML models to predict the presence of HD in patients. The QMBC model with the fusion of Anova with PCA FE techniques outperformed all state-of-the-art models and existing methodologies, achieving remarkable accuracy, precision, recall, and f1-score for the Cleveland dataset, cardiovascular dataset, and HD dataset (Comprehensive) that are available in an open repository. Specifically, the QMBC model with the fusion of Anova with PCA FE techniques achieved an accuracy of 98.36%, precision of 100%, recall of 97.22%, specificity of 100%, and f1-score of 98.59% for the Cleveland dataset.

REFERENCES

[1] C. G. D. S. E. Silva, G. C. Buginga, E. A. D. S. E. Silva, R. Arena, C. R. Rouleau, S. Aggarwal, S. B. Wilton, L. Austford, T. Hauer, and J. Myers, ‘‘Prediction of mortality in coronary artery disease: Role of machine learning and maximal exercise capacity,’’ Mayo Clinic Proc., vol. 97, no. 8, pp. 1472–1482, Aug. 2022.

[2] World Health Organization. (2009). Cardiovascular Diseases (CVDS). [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/ index.html

[3] M. Ozcan and S. Peker, ‘‘A classification and regression tree algorithm for heart disease modeling and prediction,’’ Healthcare Anal., vol. 3, Nov. 2023, Art. no. 100130.

[4] M. M. Nishat, F. Faisal, I. J. Ratul, A. Al-Monsur, A. M. Ar-Rafi, S. M. Nasrullah, M. T. Reza, and M. R. H. Khan, ‘‘A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset,’’ Sci. Program., vol. 2022, pp. 1–17, Mar. 2022.

[5] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer, ‘‘Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques,’’ IEEE Access, vol. 9, pp. 19304–19326, 2021.

[6] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, ‘‘Classification models for heart disease prediction using feature selection and PCA,’’ Informat. Med. Unlocked, vol. 19, 2020, Art. no. 100330.
[7] S. Bashir, Z. S. Khan, F. H. Khan, A. Anjum, and K. Bashir, ‘‘Improving heart disease prediction using feature selection approaches,’’ in Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2019, pp. 619–623.
[8] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, ‘‘Heart disease identification method using machine learning classification in ehealthcare,’’ IEEE Access, vol. 8, pp. 107562–107582, 2020.
[9] M. Ayar, A. Isazadeh, F. S. Gharehchopogh, and M. Seyedi, ‘‘Chaoticbased divide-and-conquer feature selection method and its application in cardiac arrhythmia classification,’’ J. Supercomput., vol. 78, pp. 5856–5882, Mar. 2022.
[10] S. Shilaskar and A. Ghatol, ‘‘Feature selection for medical diagnosis: Evaluation for cardiovascular diseases,’’ Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug. 2013.
[11] N. C. Long, P. Meesad, and H. Unger, ‘‘A highly accurate firefly based algorithm for heart disease prediction,’’ Expert Syst. Appl., vol. 42, no. 21, pp. 8221–8231, Nov. 2015.