Visualizing and Predicting Heart Diseases with an Interactive Dash Board

PROPOSED SOLUTION

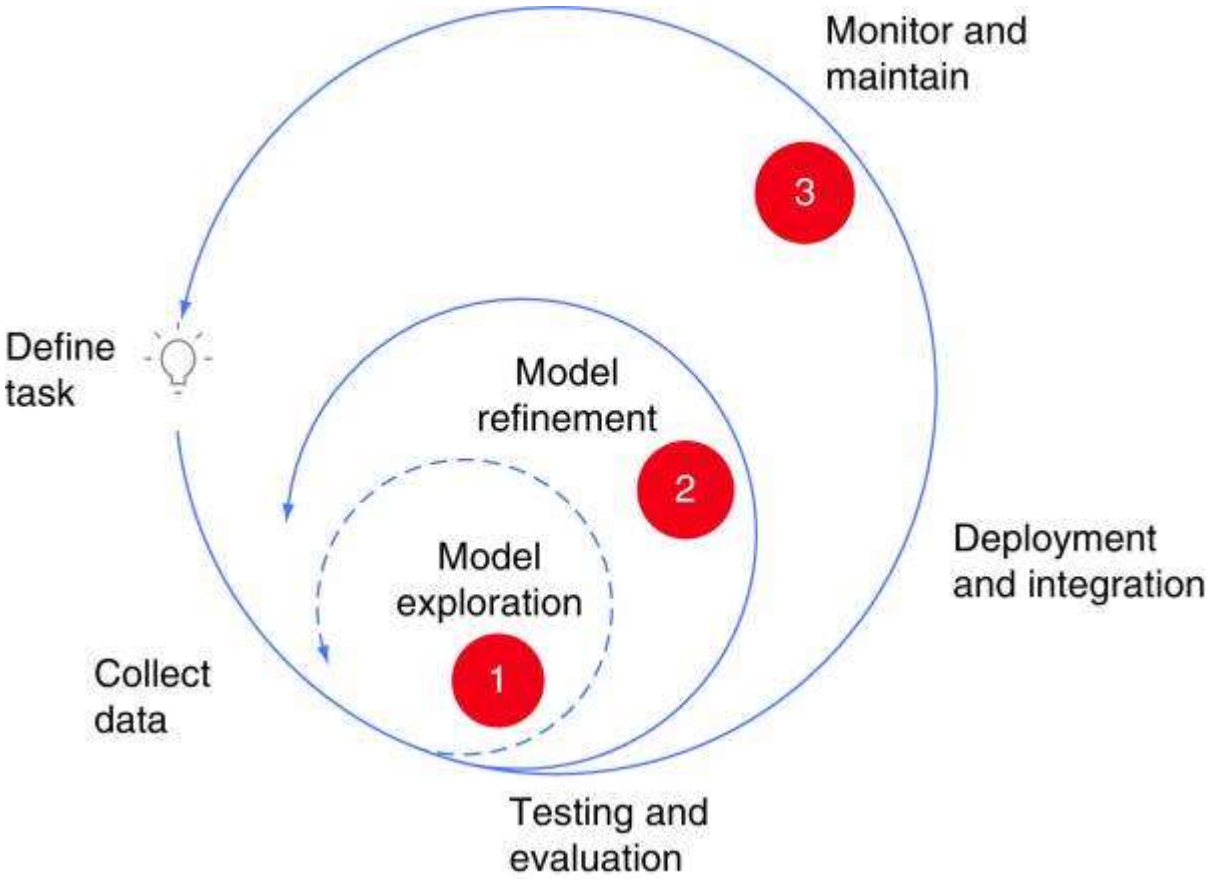
In this study, an effective heart disease prediction system (EHDPS) is developed using neural network for predicting the risk level of heart disease. The system uses 15 medical parameters such as **age, sex, blood pressure, cholesterol, and obesity for prediction.**

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive.

The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease.

Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs.

According to (Wurz & Takala,) the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending. The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still „information rich" but „knowledge poor". There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data for African genres.



Novelty:

The progress in CVD risk models is ordinarily based on conventional laboratory-based predictors. The primary purpose was to develop a prediction model, which was based on a 10-year risk. To predict 10-year risk, a Cox regression proportional method was applied. A dataset from Framingham Original Cohort and SCORE of men and women aged between 30 and 62 was taken to predicate CVD risk. In this study, the CVD risk was being predicted on a gender basis (men/woman). The risk factors involved are age, sex, BMI, SBP, and diabetes. Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease.A proposed study related to cardiac arrest prediction using the real-time dataset is classified based on gender and age. Our study has used two methods to find the risk classification

probabilities among gender-based and age-based. The two methods used—Kaplan-Meier survival analysis and Cox regression model survival analysis—are considered. These two methods have their importance in analyzing risk probabilities.

Kaplan-Meier Method:

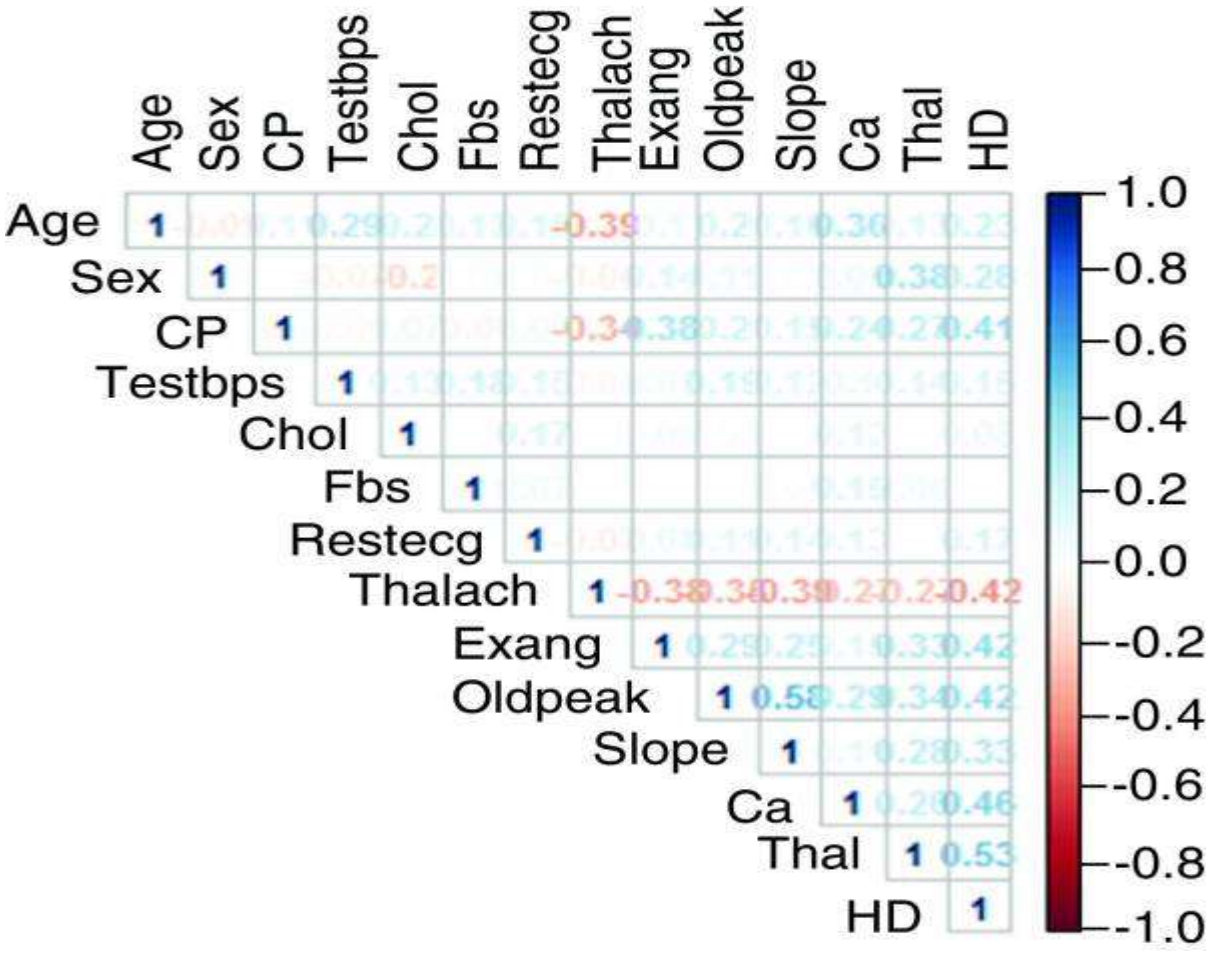
This method is a non-parametric survival function for the estimation of survival probability. Kaplan-Meier method also estimates the survival curve by giving a statistical comparison between two groups: men and women.

Cox Regression Proportional Model:

Cox regression proportional model does regression analysis with survival data without making any strong assumption. It calculates the hazard ratio (HR) of the covariates used in our data and is based on the equation, and it shows risk probability. Cox regression proportional model is a model that helps to find risk probability or score on more than one covariate. In general, it is a method to identify the effect of variables at some time interval on some event that occurs.

Feasibility of idea:

In the present study, we extract the variables of importance in one model in an effort to demonstrate the feasibility of including a risk assessment component in the ML model. Furthermore, we have openly hosted the computer code that was generated in this study on GitHub, so that we may provide the settings of a workflow model that promotes public contribution toward improved risk detection of CAD using ML, as others may contribute by extending the code to larger cohort groups.The remainder of the manuscript is organized as follows. The Dataset & preprocessing section provides information on the data source, describes the data variables, explains the data preprocessing steps and provides a high-level analysis of the data. The next section provides concise descriptions of the ML models considered in this study and the methodologies of how the model results are evaluated. The results of the modeling efforts are presented and discussed in the Results & discussion section. The concluding section outlines the limitations of the present modeling efforts and provides a workflow model framework for extending our study in future. The quality service implies diagnosing patients properly and administering effective treatments. The available heart disease database consists of both numerical and categorical data. Before further processing, cleaning and filtering are applied on these records in order to filter the irrelevant data from the database



Business model:

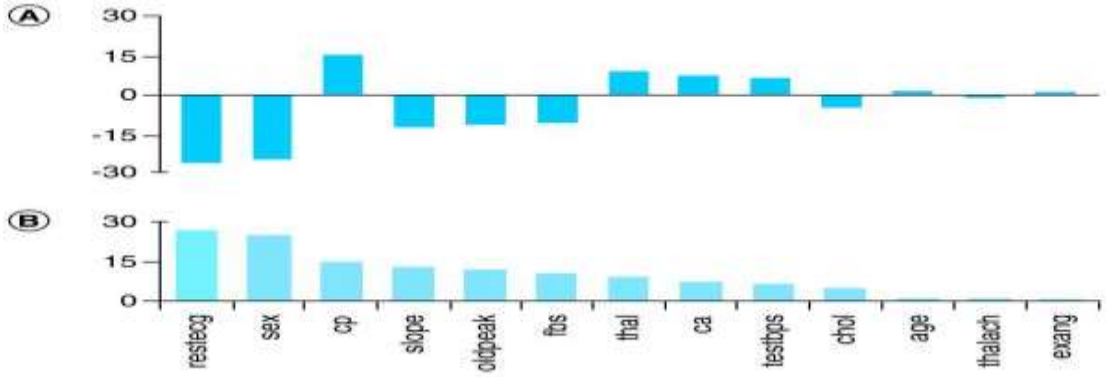
In applying machine learning models, it is generally understood that no single algorithm is superior to the others. In machine learning, if every instance in the dataset is given to the model with known labels (the corresponding correct outputs), like in the Cleveland dataset, then the learning is called 'supervised', in contrast to 'unsupervised' learning, in which instances are unlabeled. Below, we present the general idea on how each of the six supervised machine learning algorithms work on the dataset and any assumptions we make in each case. Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive.Each algorithm is first trained (or fitted) with a fraction of the dataset, usually known as the 'training set' and then tested on the 'test set' that is put aside as 'unseen data' for evaluating the algorithm. For a detailed description of the models we refer the reader to the excellent treatise by James.

Social impact:

- Cardiovascular disease, which includes heart diseases and stroke, accounts for 366 million healthy life years lost across all age groups and genders. Individually, ischemic heart disease and stroke are two of the five leading causes of healthy years lost globally.

- The physiological, social and psychological impact of these cardiovascular diseases vary across populations and individuals. Fortunately, there is an array of treatment options available, but timely diagnosis, appropriate interpretation of investigation results and apt patient selection for the various intervention methods are essential.

- Artificial intelligence (AI) has been a disruptive innovation in the world of health and medicine. Not only has it been applied for medical research, but AI can also provide algorithmic solutions in clinical settings to aid in the diagnosis, prognosis, treatment and visual pattern recognition software in fields such as radiology to aid in the interpretation of imaging. Significant attention is now turning to the potential of AI in the medical field. According to a 2019 biometric study, the number of studies on AI applications in medicine has tripled in the past three years, with heart diseases and stroke as two of the top three topics of interest

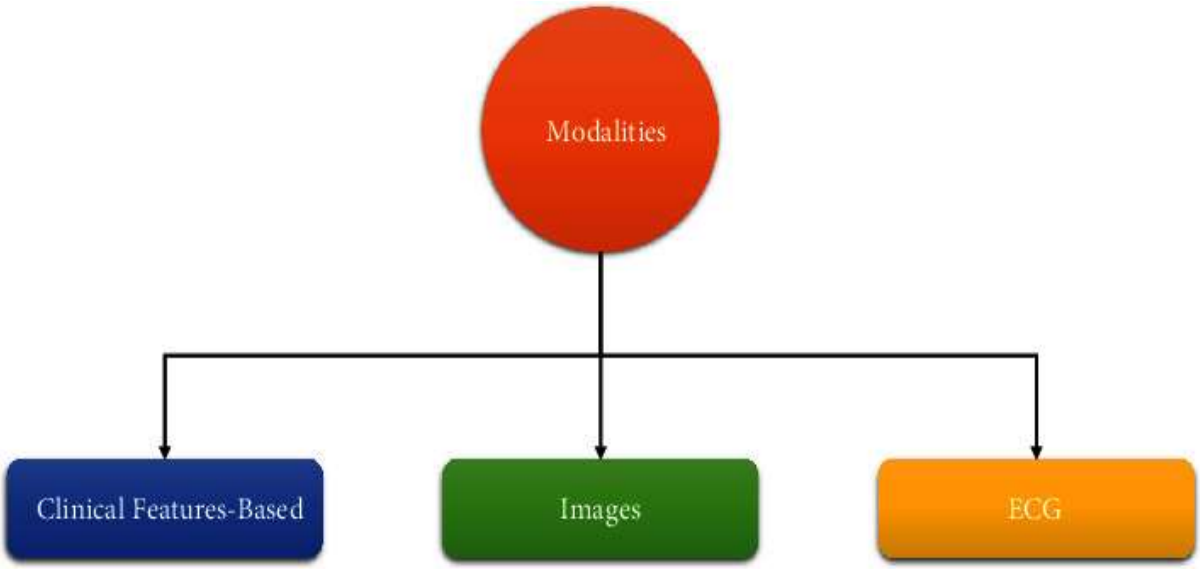


Scalability of solution:

we explore the potential issues in the diagnostic systems based on the images, ECG, and clinical feature-based data modality for heart disease detection and propose solutions. To meet this objective, data is gathered from various databases and sources like ScienceDirect, PubMed, IEEE Xplore Digital Library, Springer,

Hindawi, Plos, and Google Scholar based on the keywords: automated heart disease prediction or detection, ML-based detection of CHF, prediction of heart failure, coronary disease detection, data mining, and CVD. The literature used in this study was selected on the basis of a particular criteria as given:

- ✓ Only CAD, HF, CVD, and CHF are targeted in this study
- ✓ The articles published from 1995 to 2021
- ✓ Those papers were considered that employed ML techniques for the diagnosis of the heart diseases
- ✓ The articles published in the English language are targeted in this study
- ✓ Articles that used different types of data modalities like ECG, images, and clinical features for automated detection of heart diseases were considered
- ✓ The research articles that made use of publicly available datasets and electronic health records



Conclusion

This study contributes classification techniques for detecting and comparing gender-based and age-based probability for cardiac arrest survival. The main objective of our work is to recognize cardiac arrest in a patient as early as possible using an ML model. Apart from that, our main contribution shows survival probability in an individual with gender-based and age-based using two effective methods: Kaplan-Meier and Cox regression methods. From the result, in detecting cardiac arrest risk in gender-based survival probability, female patients

of our collected record show higher chances of survival than male patients. While in detecting cardiac arrest risk in age-based survival probability, patients with age 30 and more may have chances of cardiac risk as per our proposed model. From our classification model, the predicting results of the DT outperform the other ML classifiers.