IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE PREDICTION SYSTEM

[1]Santosh M, [2]Bindu K V, [3]Bhagyalakshmi N, [4]Arolene cynthia, [5]Kausalya R

[1] Assistant Professor, [2] Student, [3] Student, [4] Student, [5] Student

Department of Information Science and Engineering,
Cambridge Institute of Technology, Bangalore, India

Abstract - Cardiovascular diseases, particularly heart disease, remain a leading cause of mortality worldwide, necessitating advanced diagnostic systems that leverage clinical data for early and accurate prediction. Machine Learning integration techniques, particularly ensemble methods, is a way that enhances the precision and reliability of predictive models for heart disease diagnosis. The complex nature of heart diseases demands a comprehensive analysis of clinical data to derive actionable insights. While traditional diagnostic approaches have relied on individual risk factors, the combination of diverse clinical parameters offers a more broad perspective, enabling a more understanding and prediction of cardiovascular outcomes.

Keywords- Data Classification, Learning Algorithms for machines, Data Analysis

1.INTRODUCTION

Heart is a vital organ of the human body. If it fails to work effectively, mind and other organs will stop working and within a few moments the person will pass on. Detection of cardiovascular disease signs early is doctor's most challenging issues today. Each year, cardiovascular disease kills a large number of people throughout the world. Because of the gravity of the problem, the cardiovascular disease needs prompt attention. Heart disease is typically difficult to detect as there is wide range of potential contributing factors, including but not limited to hypertension, hyperlipidemia, arrhythmia, and other health issues. This means Artificial Intelligence (AI) has the potential to aid in the early diagnosis and management of health problems. Machine Learning (ML) provides dynamic calculations without a specific program to build an intelligent machine that can simplify various troublesome issues.

2. OBJECTIVE

The primary objective of this paper is to develop a straightforward model to use in the medical field. The patient's clinical data will be entered, and based on those details, the algorithm will identify the heart disease and classify it.

1. Collection of data and addressing the problems
2. Determining the algorithms that is best fit for the project
3. Create user interface
4. Testing all possible datasets for the interface
5. Determining the possible outcome from the implementation of the project

3. LITERATURE SURVEY

Li Y, Sperrin M [1] In this paper, performance model including calibration, discrimination and consistency of individual risk prediction for same patients among models with comparative model performance. Twelve machine learning family of models (grid searched for optimal models), three Cox proportional hazards models (local fitting, QRISK3, and Framingham), three parametric survival models, and one logistic model were among the 19 prediction methodologies used. Similar population level performance were seen among various models (C statistics about 0.87 and similar calibration).The predictions of individual CVD risks varied widely between and within different models, even when using similar predictors. Logistic models and commonly used ML models were found to be unsuitable for predicting long-term risks without considering censoring in survival analysis. Predictions for patients with higher risks were particularly inconsistent across different models.

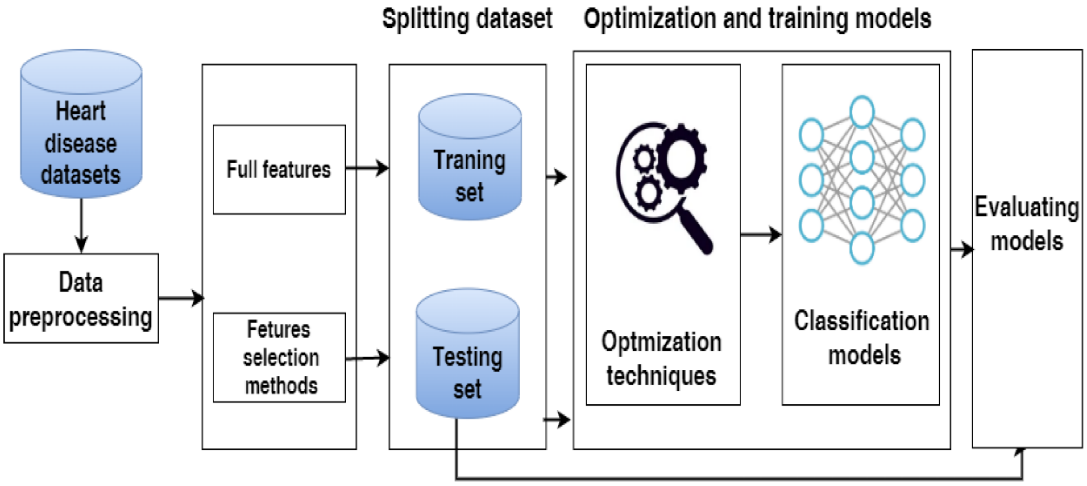
Tougui I, Jilbab A [2] For this project, we have chosen to classify heart disease using six machine learning techniques (Logistic Regression, Support Vector Machine, K Nearest Neighbors, Artificial Neural Network, Naïve Bayes, and Random Forest) and compare them with six popular data mining tools (Orange, Weka, RapidMiner, Knime, Matlab, and Scikit-Learn). The study's dataset included 303 instances, 13 characteristics, one target variable, and 139 cases of cardiovascular disease and 164 cases of healthy patients. The accuracy was one of three performance metrics used to compare how well each tool's techniques performed. The Artificial Neural Network model in MATLAB was determined to have the highest accuracy (95.38%), sensitivity (96.43%), and specificity (94.03%), making it the best-performing tool.

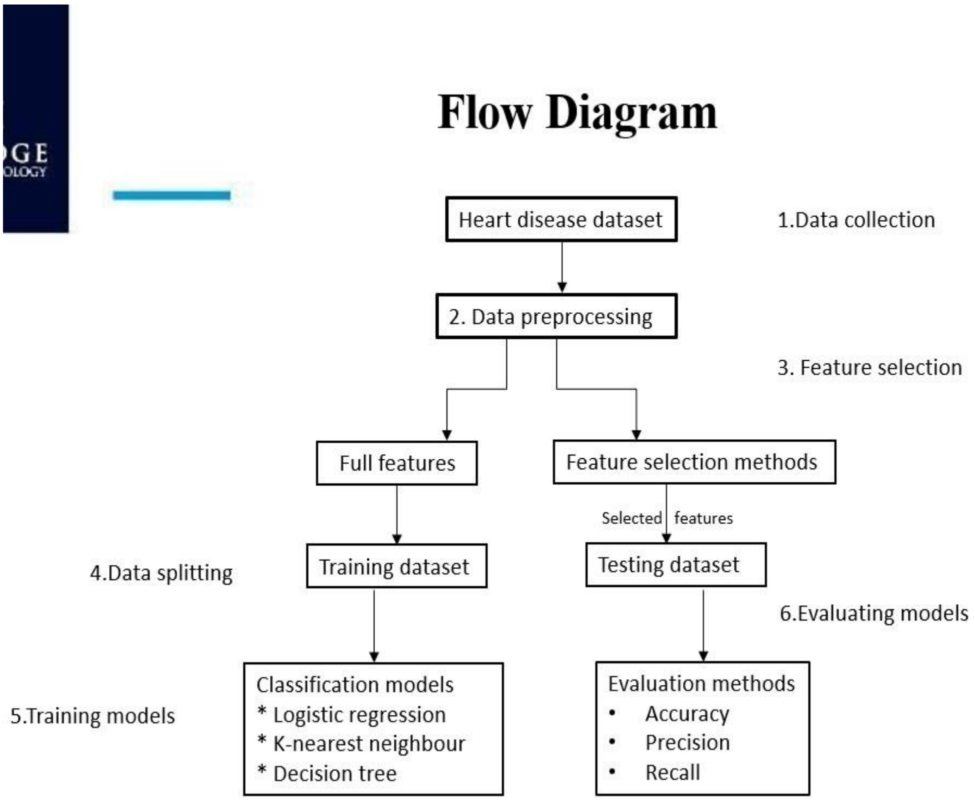
Joo G, Song Y [3] In order to supplement earlier studies, we examined the features of big data and machine learning for CVD risk prediction in this work, as well as the data from the Korean National Health Insurance Service-National Health Sample Cohort (KNHSC). To be more precise, we evaluated how well different machine learning techniques predicted the 2-year and 10-year risk of cardiovascular disease (CVD), which includes heart failure, atrial fibrillation, coronary artery disease, and strokes. We took into account comorbidities, prior medication information from the KNHSC data, results from questionnaire surveys, and typical medical examination data while creating prediction models. Using logistic regression, deep neural networks, random forests, and LightGBM, we created a variety of ML-based prediction models. We then used metrics like receiver operating characteristic curves, precision-recall curves, sensitivity, and specificity to validate our models.

Shah D,Patel S [4] This study outlines several characteristics associated with heart disease and proposes a model based on supervised learning techniques, such as random forest, decision trees, K-nearest neighbor, and Naïve Bayes. Just 14 of these 76 attributes are taken into account during testing, which is crucial to proving the effectiveness of various algorithms. The purpose of this research work is to estimate the patients' risk of acquiring heart disease. The findings show that K-nearest neighbor yields the highest accuracy score.

Al'Aref S. J, Anchouche K [5] We provide a concise synopsis of machine learning techniques utilized in the development of inferential and predictive data-driven models in this paper. We highlight many areas of machine learning use, including electrocardiography, echocardiography, and newly developed non-invasive imaging modalities including coronary computed tomography angiography and coronary artery calcium scoring. We wrap up by examining the drawbacks of the modern use of ML algorithms in the study of cardiovascular illness.

4. SYSTEM ARCHITECTURE





5. SECTION OF MODULES

Let's talk about the numerous modules that make up our suggested system and how each one helps us get closer to our objective.

1. Data/Input Gathering:

The acquisition of data from various sources may come from internal or external sources in order to address company needs or issues. Any format could contain data. Here, use CSV, XML, JSON, etc. To ensure that the correct data is in the anticipated format and organization, big data is essential. Data transformation, data cleaning, missing value filling, and feature extraction are the key data mining techniques we use. Prior to using the classification model, we encode the categorical values in the data purification section. Section of Modules

Let's talk about the numerous modules that make up our suggested system and how each one helps us get closer to our objective.

2.Data Processing (EDA):

1. Understanding the given dataset and helping clean up the given dataset.
2. It gives you a better understanding of the features and the relationships between them.
3. Extracting essential variables and leaving behind/removing non-essential variables.
4. Handling Missing values or human error.
5. Identifying outliers.
6. The EDA process would be maximizing insights of a dataset.

3.Feature engineering:

1. Handling missing values in the variables
2. Convert categorical into numerical since most algorithms need numerical features.
3. Need to correct not Gaussian(normal). linear models assume the variables have Gaussian distribution.
4. Finding Outliers are present in the data, so we either truncate the data above a threshold or transform the data using log transformation.
5. Scale the features. This is required to give equal importance to all the features, and not more to the one whose value is larger.
6. The process of feature engineering is costly and time-consuming.
7. Feature engineering can be a manual process, it can be automated.

4.Training and Testing:

1. Cross-validation of data is used to ensure improved accuracy and efficiency of the algorithm used to train the machine, and training data is used to ensure that the machine recognizes patterns in the data.
2. Test data is used to see how well the machine can predict new answers based on its training.
3. The train-test split procedure is used to estimate the ML performance of algorithms when they are used to make predictions on data that is not
used to train the model.

Training

1. Training data is the data set on which you train the model.
2. Train data from which the model has learned the experiences.
3. Training sets are used to fit and tune your models.

Testing

1. The purpose of test data is to determine whether the model has sufficiently learned from the events it has encountered in the train data set.
2. Test sets are “unseen” data to evaluate your models.

Train data: It trains our machine learning algorithm.

Test data: After the training the model, test data is used to test its efficiency and performance of the model.

5.Data Split into Training/Testing Set:

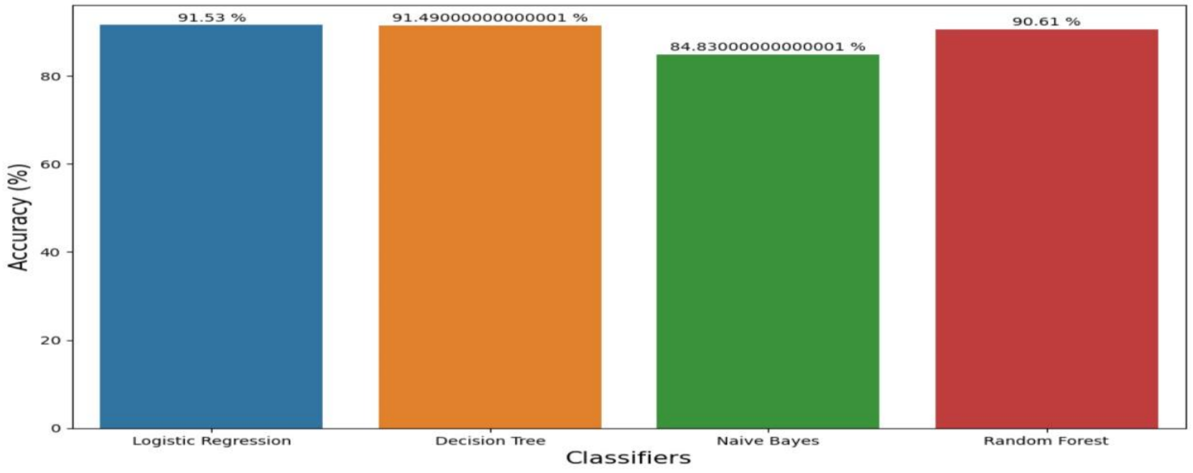
1. We used to split a dataset into training data and test data in the machine learning space.
2. The split range is usually 20%-80% between testing and training stages from the given data set.
3. A major amount of data would be spent on to train your model
4. You can use the remaining funds to assess your test model.
5. However, you are unable to combine or utilize the same data for testing and training.
6. Your model may be excessively overfitted if you test it using the same set of training data. Next, there's the question of how well models anticipate fresh data.
7. Therefore, you should have separate training and test subsets of your dataset.

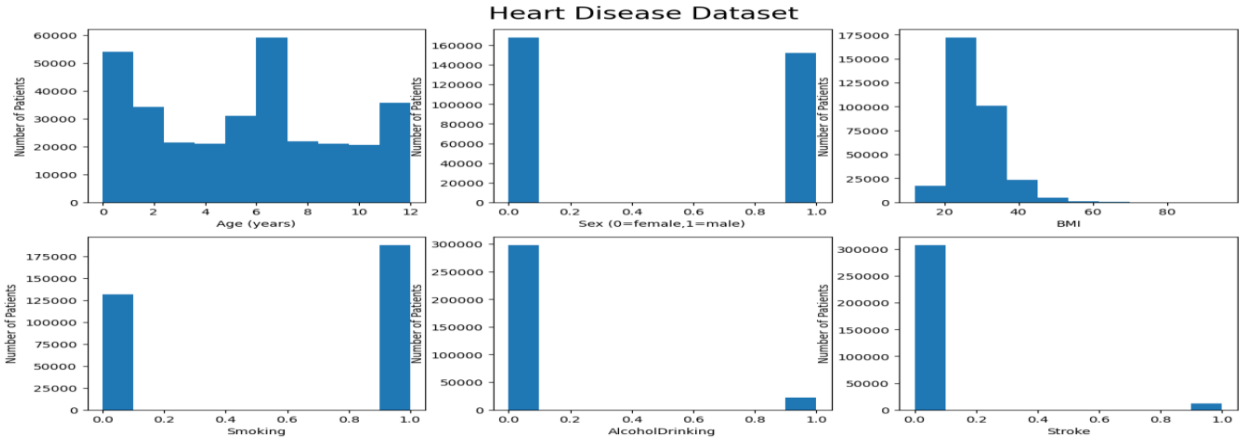
We implement the Logistic regression ,Decision tree ,Random forest, Naïve Bayes classification for dataset as mentioned and evaluate our model.

6.Module evaluation

We evaluate the classification models using confusion matrix and the accuracy score.

6 EXPERIMENTAL RESULTS





7 CONCLUSION

In conclusion, the exploration of "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method" holds substantial promise for advancing cardiovascular healthcare. The proposed approach, grounded in clinical data analysis and leveraging ensemble methods, reflects a thoughtful integration of real-world patient information and sophisticated machine learning techniques. By emphasizing early disease detection, personalized risk assessment, and clinical decision support, the diagnostic system offers the potential to significantly impact patient outcomes. However, addressing challenges related to data quality, model interpretability, and seamless integration into clinical workflows is crucial for the success of such systems. The continuous validation and adaptation of the proposed method, coupled with ethical considerations, will be essential to ensure its long-term efficacy and acceptance within the healthcare ecosystem. As technology evolves, this research contributes to the ongoing quest for effective tools in cardiovascular disease prediction, underscoring the intersection of clinical expertise and innovative machine learning methodologies in shaping the future of cardiac healthcare.

8 REFERENCES

1. Dey, D., Slomka, P. J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P. P., & Marwick, T. H. (2019). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. Journal of the American College of Cardiology, 73(11), 1317-1335.
2. Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., ... & Noseworthy, P. A. (2019). using an electrocardiogram with artificial intelligence to screen for cardiac contractile dysfunction.25(1), 70-74.
3. Madani, A., & Arnaout, R. (2018). Manna from heaven or a poisoned chalice: A systematic review of the clinical implications of convolutional neural networks for echocardiography. European Heart Journal-Cardiovascular Imaging, 19(5), 545-554.
4. Motwani, M., Dey, D., Berman, D. S., Germano, G., Achenbach, S., Al-Mallah, M. H., ... & Slomka, P. J. (2018). A 5-year multi center prospective registry analysis using machine learning to predict all-cause death in patients suspected of having coronary artery disease. European Heart Journal, 39(47), 4245-4253.
5. Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., ... & Friedman, P. A. (2019). A retrospective examination of outcome prediction utilizing an artificial intelligence-enabled ECG method for the detection of atrial fibrillation in sinus rhythm patients.The Lancet, 394(10201), 861-867.
6. Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. American College of Cardiology Journal, 71(23), 2668–2679.
7. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 69(21), 2657-2664.