A REFINED CLASSIFIER MODEL WITH GUI-BASED FRAMEWORK FOR PREDICTION OF CARDIOVASCULAR ILLNESS

Dr. Unnati A. Patel[1], Dr. Jay Nanavati[2], Dr. Kamini Solanki[3], Dr. Arpankumar G. Raval[4] & Dr. Sandeep Gaikwad[5]

[1,2,3,4&5]Assistant Professor, Smt. Chandaben Mohanbhai Patel Institute of Computer Application [CMPICA], Charotar University of Science and Technology (CHARUSAT), Changa, Gujarat, INDIA.

Abstract:

Cardiovascular disorders are one of the most common and complex disorders across the world. According to the sector fitness association, one of the top 10 important reasons for death is heart ailment. Accurate and timely identity is a critical stage in rehabilitation and remedy. A device that is capable of predicting the prevalence, is essential for the detection of cardiovascular illness. To help a health practitioner decide whether or not or not a patient has a coronary heart condition and whether or no longer they may be at a higher danger of developing cardiovascular illness, we're developing a clever clinical system primarily based on system mastering strategies that allow you to perceive an affected person's heart situation. In the publicly-to-be-had dataset, we cope with the troubles of missing records and choppy facts by utilizing a couple of data processing techniques. Furthermore, we have chosen various ML algorithms for predicting cardiovascular illness based on the device getting to know. We must eliminate superfluous and unnecessary information from the facts in order to improve class accuracy. As an end result, by using highlighting the maximum critical trends, function choice techniques may be useful in lowering the cost of analysis. The proposed diagnosing method produced superior results. Distinct metrics, along with sensitivity, accuracy, precision and F-degree, had been utilized to check our framework and confirmed that our proposed technique performs highly better than competing procedures.

Keywords: Cardiovascular Illness, Machine learning algorithms, Pre-processing, Feature selection, Correlation matrices, heart attack

Introduction:

The global healthcare sector has the latest challenge of optimally use information technology for the benefit of mankind. This necessitates prevention, early detection, accurate prediction and treatment with the help of information technology. On the other end, there is a large of health issues and heart-related issues are one of them. When a Cardiovascular illness occurs, the blood supply to all the organs is interrupted. The most common compounds accumulated after a cardiovascular illness are fat, cholesterol, and other compounds. Heart attack, or myocardial infarction, occurs when oxygen in the proper amount is not received by the part of the cardiac muscle. A few causes for cardiovascular illness might be emotional nature or chest pain in the left part of the heart for few minutes or goes away and comes back. Rubbing, uncomfortable strains, pain or fullness can all be unpleasant. Weakness, dizziness, or faintness, you might also break out in a cold sweat. Irritation or pressure on the neck, back, or mask.

Submitted: 21/10/2022 **Accepted**: 24/11/2022

Irritation or Pain in one or both shoulders or arms. Shortening of the breath can also cause a lack of oxygen prior to a Cardiovascular illness but it can also cause heart pain.
According to the Global Burden of Disease, CVDs account for nearly a quarter (24.8 percent) of all deaths in India. An examination of medical certification of cause of death (MCCD) reports reveals an increase in the proportion of deaths caused by CVD. According to WHO, In 2019, due to CVDs, 17.9 million people died, which is near to 32% of all global deaths. From this, the main cause of 85% of deaths was stroke or a heart attack.
Cardiovascular illness statistics: Last year, 735,000 many Americans had a heart attack. In United States, Cardiovascular illness is the leading cause of death. 0.3% of Men and Women under age group 20 to 39 who are suffering from Cardiovascular illness. The average age of the first Cardiovascular illness is 72 years for women and 65 years for men.
[1] Cardiovascular diseases (CVDs) are a group of illnesses of the blood vessels the heart. They include cerebrovascular disease, coronary heart disease, rheumatic heart disease, deep vein thrombosis and pulmonary embolism, peripheral arterial disease and congenital heart disease. Cerebrovascular disease is a disease of the blood vessels which supply blood to the brain. Coronary heart disease is a disease of the blood vessels which supply blood to the heart muscle. Rheumatic heart disease damages to the heart valves and heart muscle from rheumatic fever. This fever is due to streptococcal bacteria. In Deep vein thrombosis and pulmonary embolism, blood clots in the leg veins. Periodically it can dislodge and moves to the lungs and heart. Peripheral arterial disease is a disease of the blood vessels which supply blood to arms and legs. Congenital heart disease is the birth defects that affect the normal functioning and development of the heart caused by malformations of the heart structure from birth.
Heart strokes and attacks are typically sudden events which is caused by a clog that precludes blood from flowing to the brain or heart. The utmost common cause is the gathering of greasy deposits on the internal walls of blood vessels that supply the brain or heart. Strokes can be caused by bleeding from a blood vessel or blood clots in the brain.
The factor of risk: The following are the most important risk factors for Cardiovascular illness control: Alter, alter, alter. Men over 45 and women over 55 are more victims of Cardiovascular illness than younger women and men. Second-hand smoke exposure and smoking were both considered. Their blood pressure is elevated. Over time, the artery gets damaged by high blood pressure. Obesity, high cholesterol, or diabetes all increase the chances of developing high blood pressure. Triglyceride cholesterol or serum levels are higher. LDL i.e. Low-density lipoprotein has been linked to high cholesterol which is known as "bad" cholesterol. As triglycerides (a type of blood fat associated with diet) increases, your Cardiovascular illness risk increases. A high volume of high-density cholesterol lipoprotein i.e. HDL which is known as Good cholesterol would reduce the risk. Obesity is directly connected with high triglyceride levels, high serum cholesterol levels, diabetes and high blood pressure. This risk can be reduced by weight loss. Diabetes - When the body does not produce enough secreted hormones – insulin, blood sugar intensities rise, and in turn, it is going to increase your risk of Cardiovascular illness. The blood sugar that you are getting is not affecting your insulin.

Significant metabolism, this is because you have asthma, high blood sugar and smoke. The metabolic syndrome will increase the chances of heart disease twice.

Developing and underdeveloped countries lack the infrastructure, doctors and technologies, needed to predict illness in the primary stage and avoid mortality and complications. Because many illnesses are linked to the heart, precise analysis and prediction are required. To find a solution, virtual research into this subject is required. Because these illnesses are typically diagnosed at an advanced stage and are the leading cause of cardiac patient death due to a lack of accuracy, effective illness prediction algorithms must be developed. The advancement of telecommunication and information technology has benefited both poor and rich patients by providing immediate information and monitoring patient health and lowering the cost of diagnosing. Dramatically, it has increased the number of patients' detailed health records. Researchers have access to a large number of medical records. It is a challenge to the medical industry, to utilise massive amounts of medical data. Massive amounts of data have been transformed to obtain accurate and valuable information with the help of machines. As a result, ML (machine learning) becomes a significant field. Machine learning, which is based on precise training and testing and is supported by Python and Python libraries, is one of the most capable, capable, and effective technologies. The method learns directly from data and skill; testing on various types of needs should be done in accordance with the necessary algorithms based on this learning. Machine learning is an effective technology for testing and training. It is the result of artificial intelligence (Artificial Intelligence). Machine learning is one of AI's disciplines. Machine learning technology can be used to perform tasks that humans currently perform using their human intelligence. ML technology includes a number of data-use processes that enable human intelligence traits. It absorbs the ML-described typical phenomenon. Machine learning algorithms and Python libraries must be used for this prediction. Biological elements such as cholesterol (chol), sex, blood pressure (bp), and chest pain (cp) are used in this detection. Using these elements, three ML algorithms are used to predict analysis and determine which technique is best based on the results, including SVM with three different kernels, KNN, Gradient Boosting, Random Forest, and Logistic Regression. Cardiac illness can be prophesied with greater accuracy using ML. In the early stages, it helps healthcare diagnose and treat patients, allowing many patients to be diagnosed in a small amount of time. As a result, lives of millions will be saved.

Description of the Dataset

The Public Health Dataset, which is available on Kaggle, was used in this study. The total no. of attributes is 76, which also includes the predicted attribute. However, in our experiment, we used only 14 out of 76 attributes. Most of the published research work to the best of our knowledge has also considered the same 14 attributes.

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

Figure 1: dataset with attributes

The features used in the proposed work are described below:

1) Age → patient's age in years
2) Sex → (male:1; female:0).
3) Cp → chest pain.
4) Trestbps → resting blood pressure (in mm Hg) (At the time of patient admission to the hospital). The normal range = 120/80 (higher than the normal range is risky).
5) Chol → serum cholesterol. (the amount of triglycerides). (Normal = below 170 mg/dL).
6) Fbs → fasting blood sugar (Normal = Less than 100 mg/dL)
7) Restecg → resting electrocardiographic results.
8) Thalach → maximum heart rate.
9) Exang → exercise-induced angina (yes:1).
10) Oldpeak → ST depression induced by exercise
11) Slope → slope of the peak exercise ST segment
12) Ca → the number of major vessels (0–3).
13) Thal → thalassemia (normal:3; fixed defects: 6; reversible defects: 7).

Additionally, a feature named Output is also used; it represents presence (indicated by 1) or absence (indicated by 0) of the cardiac health issue.

Related Work:

In [2], Safa and Pandian applied ML algorithms for cardiac stress prediction using IoT devices. Their system sensed the temperature, blood pressure, pulse, and other relevant parameters of cardiac patients involved in stress. The outcome showed that the quality of KNN classification is greater than that of Decision tree methods and SVM.

D. Bala krishnand et al [3] developed an IoT-based solution to detect arrhythmic heartbeats. A heart rate sensor to be placed on tip of the finger was used in this experiment. The linear regression is employed for the prediction.

Tyagi et al. [4] projected a prototype for classification and probabilistic analysis using map-reduce programming and gave a graphical representation for the same.

In [5], Ali et. Al proposed Fuzzy Association-based rule mining for profile-based assessment of CVD.

Robert et al. [6] worked on heart disease detection and attained an accuracy of 77.1% using logistic regression classification classifier.

Wankhade et al. [7] worked on heart disease diagnosis and obtained an accuracy of 80% using a multi-layer perceptron (MLP) classifier.

Allahverdi et al. [8] worked on a classification model and obtained an accuracy of 82.4% using an artificial neural network.

Awang et al. [9] used Decision Tree and NB for the diagnosis and prediction of heart disease and achieved an accuracy of 80.4% with DT and 82.7% with NB.

Olaniye and Oyedodum [10] proposed an ANN-based model for the prediction of heart disease.

Das and Turkoglu [11] proposed a predictive model for heart disease using ANN ensemble-based.

Likewise, Tomov et al. [12] introduced a model for the prediction of heart disease using deep learning with higher accuracy.

Studies have clearly shown that more than one machine learning algorithm such as Support Vector Machine, Random Forest, Logistic Regression, and Gradient Boosting are utilized successfully as a decision-making tool to predict heart illness based on individual clinical information. Moreover, it can be extended for the prediction of other critical illnesses such as cancer, diabetes, brain tumour etc.

Proposed Model:

Data Pre-processing

Figure 5 depicts our proposed model's sequential chart. Cleaning the collected data may result in noisy or missing values. To obtain an effective and accurate result, these data must be cleaned for noise values and missing values must be completed. To make data more understandable, data needs to be converted into another format.

There are no null values in the dataset. However, the dataset posed two challenges: considerable no. of outliers and skewed distribution of data. Appropriate measure were taken to meet with these challenges.

Checking the Distribution of the Data

This assists the model in identifying patterns in the data that lead to cardiovascular illness. The dataset includes information from 303 people. There are 165 people here who have cardiovascular illness and 138 people who do not. The bar in green colour in the graph represents people who have cardiac issues, whereas the bar in the red colour represents people who do not have cardiac issues.

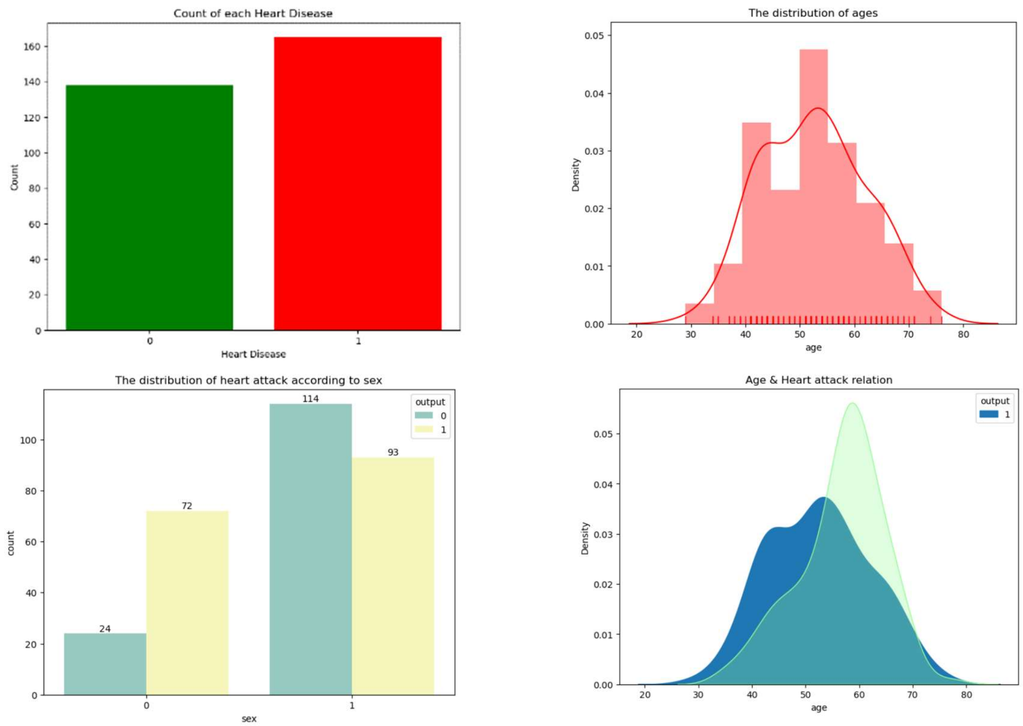

Figure 2: Data Distribution

Analysis of Data

The following figure (Fig. 3) represents analysis of quantitative features from the dataset:

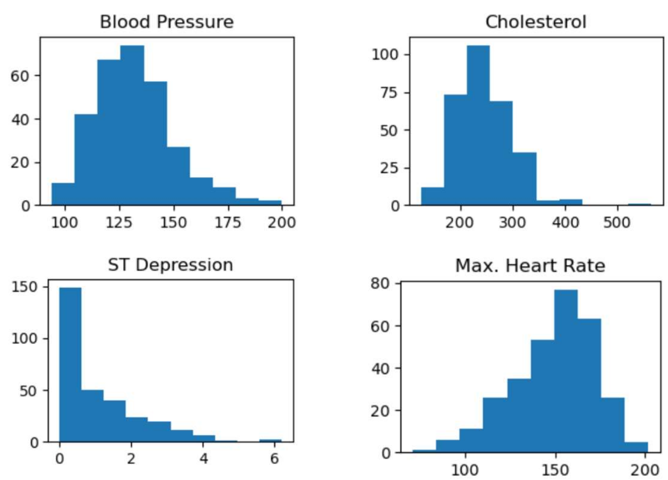

Figure 3: Quantitative Data

Followings are the some of the important observations regarding quantitative features:

1. trestbps : resting blood pressure, if it is more than 130, raises worries.
2. chol above 200 is alarming.
3. The ST depression vs. rest stares at heart stress.
4. thalach above 140 increases chances of the heart attack.

The following figure (Fig. 4) represents analysis of qualitative features from the dataset:

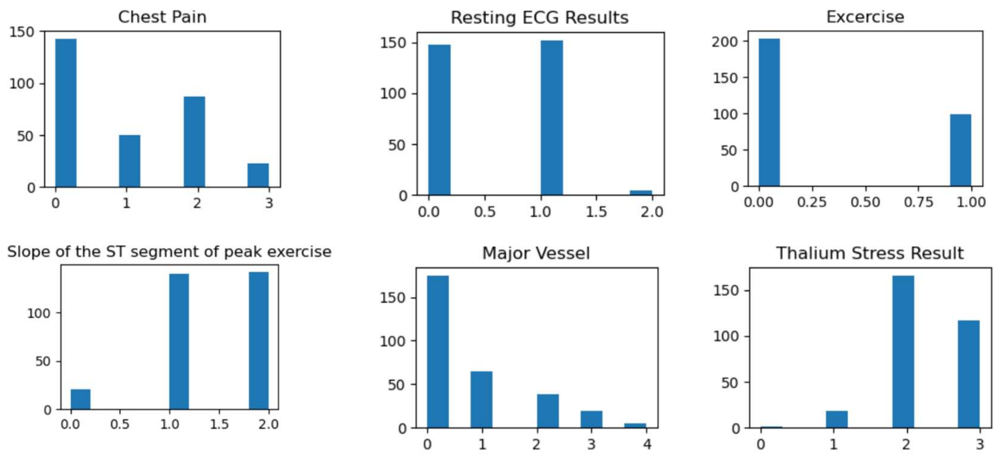

Figure 4: Plots for categorical data

Followings are the some of the important observations regarding qualitative features:

1. Chest Pain: 0 means absence of risk whereas a greater value of indicates a risk of a heart attack.
2. Resting ECG results: 1 means having an abnormal heart rate, and it indicates chances of having a cardiovascular illness.
3. Exercise: represents exercise - encouraged angina, 0 = chances of cardiovascular illness, 1 = it absence of cardiovascular illness.
4. Slope of the ST segment of peak exercise: 2 = an unhealthy heart and additional chances of having a cardiovascular illness. 0 = best heart rate with exercise. 1 = strong heart.
5. Major Vessel: No. of major vessels stained by fluoroscopy (0 - 3). It is a measure of movement of blood. 0 - chances of cardiovascular illness.
6. Thalium stress result : 2 - increased chances of cardiovascular illness.

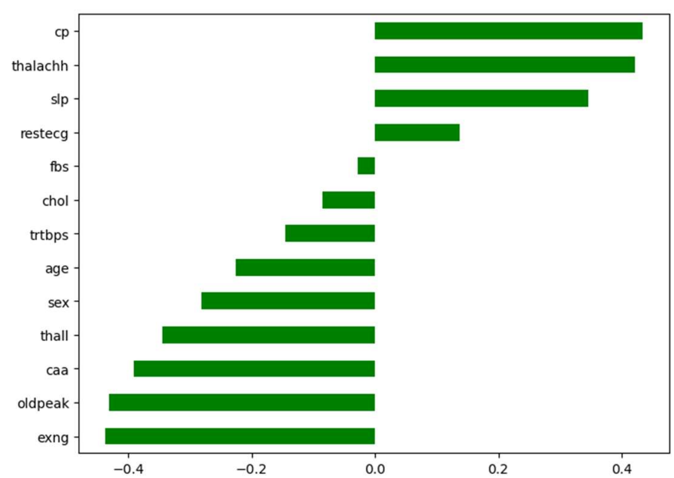

Figure 5: Correlation Matrix

From the Correlation Matrix which is shown in Figure 5, it can be concluded that fbs and chol have relatively low impact on the output variable. However, rest of the variables have a substantial impact on the output variable.

Proposed Methodology

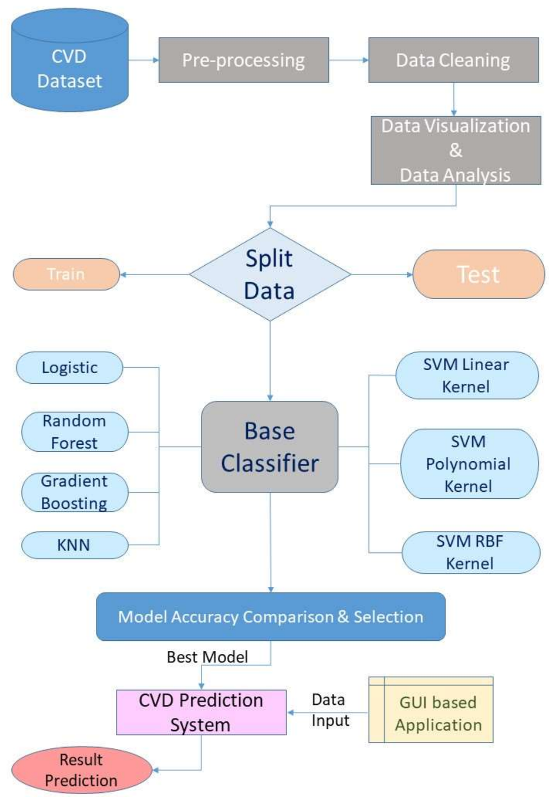

Figure 6: Proposed Research step-by-step process

Figure 6 depicts the proposed model in a step-by-step process. In this study, we used a dataset of 303 records with 13 observable attributes, which we then converted into floating data. We pre-processed the data after conversion, taking care of missing values, cleaning the data, and normalization. We analyzed the data after pre-processing it, checking the dependency of a few variables on others. Following data analysis, data is divided into training and testing datasets. (KNN), Logistic Regression, Random Forest Classifiers, Gradient Boosting, SVM Linear Kernel, SVM Polynomial Kernel, and SVM RBF Kernel are the classifiers used in the proposed model. The models were then evaluated for accuracy and performance using various performance metrics. Figure 7 depicts the assessment of accuracy for all classifiers.

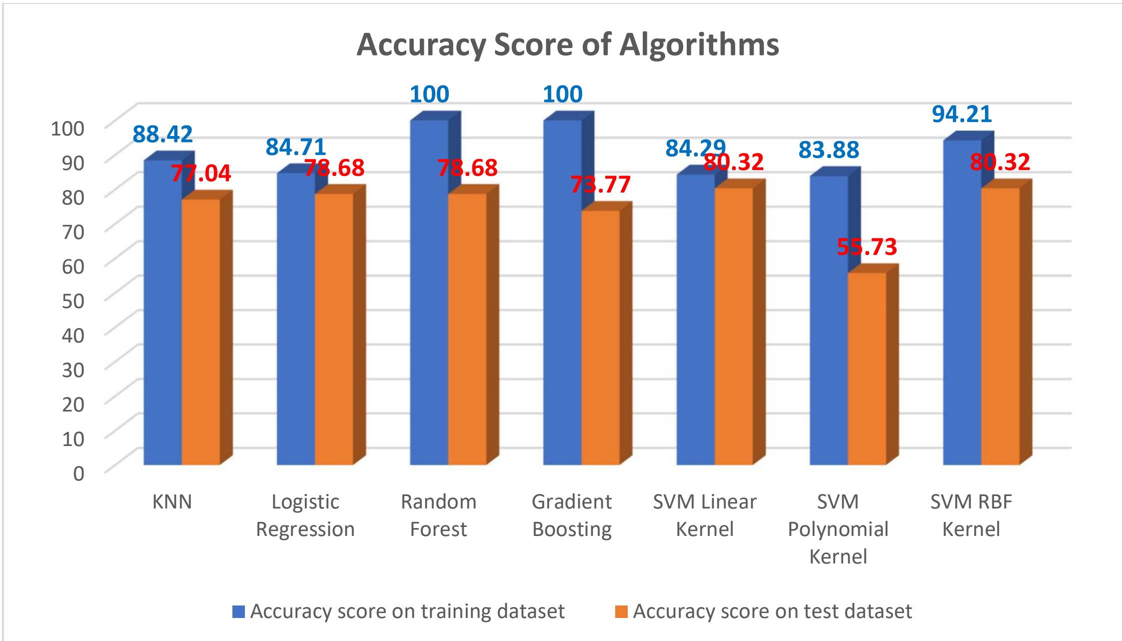

Figure 7: Accuracy Assessment of ML Algorithms

A cardiovascular prediction system with a graphical user interface (GUI) is created. The GUI allows the user to provide various health parameters as input to the system. Finally, cardiovascular defects can be predicted using the input provided, and the best classifier (selected) model.

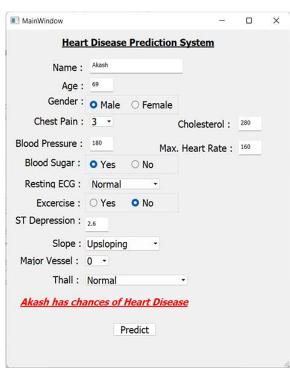

Figure 8 (a)

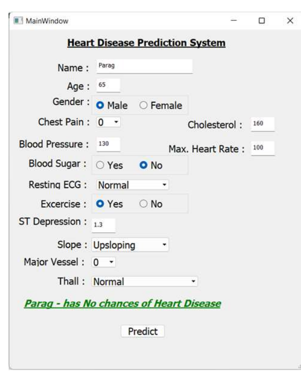

Figure 8 (b)

The Cardiovascular prediction system is depicted in Figures 8 (a) and (b). Figure 8 (a) shows that the system predicted that there is no risk of cardiovascular illness based on the user input. Figure 8 (b) depicts the system's prediction of a cardiovascular illness problem based on user input.

Result Analysis:

Table 1 Accuracy Score of classifiers

Sr. No.	Classifier	Accuracy score on training dataset (%)	Accuracy score the on test dataset (%)
1	KNN	88.42	77.04
2	Logistic Regression	84.71	78.68
3	Random Forest	100	78.68
4	Gradient Boosting	100	73.77
5	SVM Linear Kernel	84.29	80.32
6	SVM Polynomial Kernel	83.88	55.73
7	**SVM RBF Kernel**	**94.21**	**80.32**

The accuracy score of all classifiers is shown in Table 1. By comparison of the results, we come to know that SVM with RBF Kernel Classifier provides the highest accuracy. The algorithms used in this experiment offer higher accuracy, require less computational power and generate quick results. Table 2 in this study represents the confusion matrix of all classifiers used. The SVM RBF kernel yields the best results.

Table 2 Classifier Confusion Matrix Analysis

Classifier	True Positive	False Positive	False Negative	True Negative
KNN	22	6	8	25
Logistic Regression	22	6	7	26
Random Forest	22	6	7	26
Gradient Boosting	21	7	9	24
SVM Linear Kernel	20	8	4	29
SVM Polynomial Kernel	16	12	15	18
SVM RBF Kernel	22	6	6	27

Table 3 displays classification results. The precision, recall, and f measures are examined in this section.

Precision: The fraction of records obtained that are significant to the query.

Recall: A subset of the connected records that are efficiently gathered while gathering information.
F measure: An amount of a test's accurateness.

Table 3 Result Scrutiny

Classifier	Class	Recall (%)	Precision (%)	F Measure (%)
KNN	0	79	73	76
	1	76	81	78
Logistic Regression	0	79	76	77
	1	79	81	80
Random Forest	0	79	76	77
	1	79	81	80
Gradient Boosting	0	75	70	72
	1	73	77	75
SVM Linear Kernel	0	71	83	77
	1	88	78	83
SVM Polynomial Kernel	0	57	52	54
	1	55	60	57
SVM RBF Kernel	0	79	79	79
	1	82	82	82

Conclusion:
With the rising number of causalities from cardiovascular ailments, it has become imperative to develop a working model which is capable of offering accurate and timely predictions related to cardiovascular illness. The study's motivation was to find the most efficient ML algorithm for predicting cardiovascular illness. The accuracy score, Confusion matrix, precision, and recall of (KNN), Logistic Regression, Random Forest Classifiers, Gradient Boosting, SVM Linear Kernel, SVM Polynomial Kernel, and SVM RBF Kernel algorithms for predicting heart conditions using the dataset are described briefly in this study. According to the findings of this study, the SVM RBF Kernel algorithm is the most efficient algorithm for predicting cardiovascular illness, with an accuracy score of more than 90%. In his study, a GUI-based application is built to predict cardiovascular illness based on user-supplied inputs. The SVM RBF kernel is used to predict cardiovascular illness, and the prediction result is displayed to the user. In the future, reading can be taken using IoT-based measuring instruments to increase the reliability of the predictions.

References:
[1] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Safa M, Pandian A, Applying machine learning algorithm to sensor coupled IoT devices in prediction of cardiac stress – An integrated approach, Materials Today: Proceedings,2021, ISSN 2214-7853
[3] Balakrishnand D, Rajkumart T. Dhiliphan, Dhanasekaran S, An intelligent and secured heart rate monitoring system using IOT,Materials Today: Proceedings,2020,ISSN 2214-7853
[4] Tyagi A., Mehra R. Intellectual heartbeats classification model for diagnosis of heart disease from ECG signal using hybrid convolutional neural network with Goa. SN Applied Sciences. 2021;3(2): p. 265.
[5] Ali Y, Amir R, Fardin A-M. Profile-based assessment of diseases affective factors using fuzzy association rule mining approach: A case study in heart diseases, Journal of Biomedical Informatics, Volume 116,2021,103695, ISSN 1532-0464
[6] Detrano, R. et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. Am. J. Cardiol. 64, 304–310 (2009).
[7] Gudadhe, M., Wankhade, K. & Dongre, S. In 2010 International Conference on Computer and Communication Technology (ICCCT), 741–745 (IEEE, New York).
[8] Kahramanli, H. & Allahverdi, N. Design of a hybrid system for the diabetes and heart diseases. Expert Syst. Appl. 35, 82–89 (2013).
[9] Palaniappan, S. & Awang, R. In 2012 IEEE/ACS International Conference on Computer Systems and Applications 108–115 (IEEE, New York).
[10] Olaniyi, E. O., Oyedotun, O. K. & Adnan, K. Heart diseases diagnosis using neural networks arbitration. Int. J. Intel. Syst. Appl. 7, 72 (2015).
[11] Das, R., Turkoglu, I. and Sengur, A. (2009) Effective Diagnosis of Heart Disease through Neural Networks Ensembles. Expert Systems with Applications, 36, 7675-7680.
[12] Tomov, N.-S. & Tomov, S. On deep neural networks for detecting heart disease. arXiv:1808.07168 (2018).