

# A Unique Method of Finding Heart Attack and Suggesting Preventive Prediction

<sup>1</sup>Fahmitha Noorin S, <sup>2</sup>Dr. S Manimekalai

<sup>1</sup>M.Sc Student, <sup>2</sup>Dean & Associate Professor

<sup>1,2</sup>PG & Research Department of Computer Science,

<sup>1,2</sup>Theivanai Ammal College for Women (A), Villupuram, TamilNadu, India

---

## ABSTRACT

In this modern era cardiovascular diseases are the good number tedious one. This disorder physical attacks a human so in a flicker that it only just gets at anything time to get treated with. So identifying patients truthfully on suitable hypothesis is the most difficult obligation for the therapeutic society. A wrong prediction to find heart disease will be a big problem in patient's life. The medical errors were took place in India and also in other countries. The encouragement behind this paper is to append to an economical knowledge by utilizing the mining improvements for heartening in the sequence base choice sensitively compassionate network. This paper deliberated on the viewpoint which is dismissed in the development.

**Keywords:** K-Means, Heart Disease Prediction, Data Analysis, ClusterAnalysis

---

## 1. INTRODUCTION

The best carry out to evade human mortality origin by life threatening diseases like myocardial infarction is to detect them early prevent their onset. One approach is to device computational methods that capitalize on clinical biomarkers to better screen the patients for their potential risk of experiencing MI. Our objective is to generate huge amount datasets which is taken as input from the patient and generate the output using and clustering algorithm that predict the heart attack with remedies for overcoming it and with contacts of the specialist.

A great deal of elements adds to coronary illness like cholesterol level, circulatory strain, smoking or drinking propensities, beat rate, age, diabetes, and some more. Throughout the most recent years, many AI, information mining and neural organization approaches are embraced to recognize the illness. Different Machine Learning models like K-Nearest Neighbor, Naïve Bayes, Decision Tree, Support Vector Machines are familiar with decide the danger level these illnesses have yet the information from the most recent couple of many years characterize that numerous individual are having these infections at a beginning phase and surprisingly a few new conceived kids are experiencing or passed on a coronary illness. AI viewpoints can assume a significant part in forecast of these infection and the danger level it has.

## 2. REVIEW LITERATURE

There are various works has been done connected with sickness expectation frameworks utilizing various information mining strategies and AI calculations in clinical focuses.

R. Sharmila et al, proposed to involve non- straight grouping calculation for coronary illness expectation. It is proposed to utilize bigdata apparatuses like Hadoop Distributed File System (HDFS), Mapreduce alongside SVM for expectation of coronary illness with enhanced property set. This work made an examination on the utilization of various information digging strategies for foreseeing heart sicknesses. It proposes to involve HDFS for putting away huge information in various hubs and executing the forecast calculation involving SVM in more than one hub at the same time utilizing SVM. SVM is utilized in equal style which yielded preferred calculation time over consecutive SVM.

K. Polaraju et al, proposed Prediction of Heart Disease utilizing Multiple Regression Model and it demonstrates that Multiple Linear Regression is suitable for foreseeing coronary illness possibility. The work is performed utilizing preparing informational index comprises of 3000 examples with 13 distinct traits which has referenced before. The informational collection is partitioned into two sections that is 70% of the information are utilized for preparing and 30% utilized for testing. In light of the outcomes, obviously the grouping exactness of Regression calculation is better contrasted with different calculations.

Marjia et al, created coronary illness forecast utilizing KStar, j48, SMO, and Bayes Net and Multilayer discernment utilizing WEKA programming. In view of execution from various element SMO and Bayes Net accomplish ideal execution

than KStar, Multilayer discernment and J48procedures utilizing k means cross approval. The precision exhibitions accomplished by those calculations are as yet not palatable. In this way, the exactness' presentation is worked on additional to give better choice to determination illness.

### 3. PROPOSED METHODOLOGY

#### A. Information Collection

A significant writing survey was performed toward the start of the review to distinguish the hole in the current investigations. At the same time, huge review issues were found through writing audit which helped to gather significant information and variables. This study was directed between July 2018 and September 2019 at Dhaka, Bangladesh. Essential information was gathered through both field overviews and online studies. An online survey (Google structure) was searched the overall set of different ages.

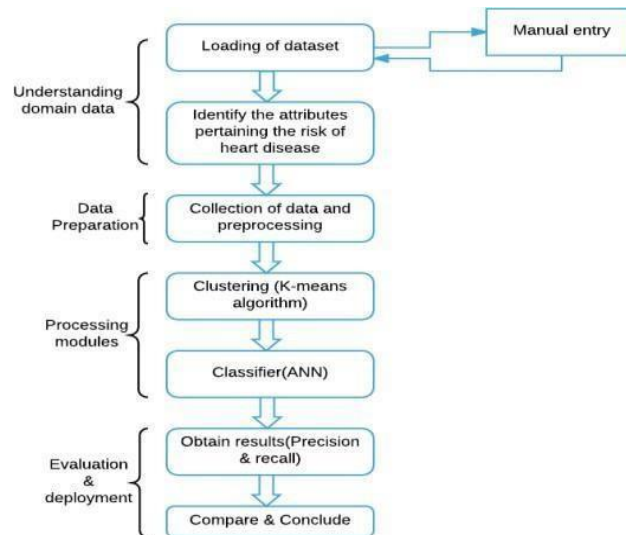
#### B. Members

A sum of 3500 polls were conveyed and 1247 legitimate records were assembled (field study: 952 web study: 295), including unconstrained female (43% and male (57%) members.

#### C. Information Preprocessing

To get the exact worth, the dataset was completely disposed until dispensable, incomplete, conflicting data was taken out. A few python libraries were utilized to preprocess the crude information. Simultaneously, we additionally completed information change from string to mathematical worth to fit the information with the classifiers.

#### D. Keyword Extraction



Keyword extraction is a course of restricting dimensionality by decreasing less suitable components from unrefined data which helps separate huge components to work on insightful precision and control over fitting. Fig. 2 shows additional huge highlights from base to top. From 14 elements, we took the initial 11 elements that assisted us with getting more precision. To extricate the significant elements "sklearn.ensemble.ExtraTrees Classifier" class has been utilized.

#### E. Model Selection

The last informational collection was parted into a preparation set (80%) and a testing set (20%). The main six normal calculations have been chosen to investigate the best-execution AI classifier for foreseeing coronary illness.

a) Decision Tree (DT): A non-parametric directed characterization and relapse learning strategy. The objective is to make a model that predicts the worth of an objective variable by learning basic guidelines of a choice in light of information

attributes.

b) Support Vector Machine (SVM): A Vector Support Machine (SVM) is officially characterized as a one-sided classifier by a specific hyper plane. The assurance capacity of SVMs depends on a portion of the preparation information sub-sets called help vectors.

c) K Means: It is one of the learning calculations that are usually utilized. The NB classifier is a k method rule based probabilistic model.

#### 4. SYSTEM ARCHITECTURE

The system architecture diagram depicts the flow of execution of the proposed system. To achieve the proposed system, the user has to collect the data for preprocessing, clustering and classification has to perform. The experimental results have to perform the evaluation and deployment.

#### 5. ALGORITHM DESCRIPTION

K-means clustering is well known preprocessing calculation broadly utilized in data mining. The fundamental point of this preprocessing is to find the positions  $\mu_i$ ,  $i = 1 \dots k$  inside group to limit amount of squares distance from the centroid. K-means clustering relies upon k groups, and it might have stuck for various arrangements. So, to eliminate such dependence or altered for further developed k-implies was proposed. K means went with Lloyd's calculation to dispose of conditions. Utilizing this technique, the outcomes show the nature of groups isn't compromised. Speculations for K-implies calculation are:

1. Introduce the focal point of the preprocessing from n information focuses  $x_i$ ,  $i=1 \dots n$  that must be parceled in k groups.
2. Characteristic the nearest cluster to every information point utilizing Euclidean distance.
3. Set the place of each group to the mean of all information guides having a place toward that cluster.
4. Repeat stages 2-3 until optimal value receives.

The balanced, ordered, summary based operating representative prototypes to mutually evaluation compassion, specificity and AUC.  $Sen_i$   $Sen_i$  and  $Spc_i$   $Spc_i$  denote the sensitivity and specificity of the  $i^{th}$  study.  $\sigma_{sen}^2$  is the variance of  $\mu_{sen}$  and  $\sigma_{sen}^2$  is the variance of  $\mu_{spc}$ .

$$\begin{pmatrix} \mu_{seni} \\ \mu_{spci} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_{sen} \\ \mu_{spc} \end{pmatrix}, \begin{pmatrix} \sigma_{sen}^2 & \sigma_{SenSpc} \\ \sigma_{SenSpc} & \sigma_{Spc}^2 \end{pmatrix} \right\}$$

#### PROCESS SYSTEM

Usually to predict any heart disease the patient has to undergone the following process like BP checkup, ECHO Cardiograph, Treadmill Test, Angiogram, etc. But these tests may be take time and it may consume some or huge cost from the people's savings. Without any heart disease symptoms, no one is going to the hospital to take tests. But our paper provides a web based facility to find the basic optimal value of having heart disease. In order to find the optimal value one user has to register in this application the user has to enter the various criteria like whether has any heart disease heredity, smoking, cholesterol, BP, hyper tension, Obesity, etc. clustering is performed here, it is an unsupervised machine learning task. It also includes doctor suggestion and the admin saves the normal ranges of clusters for each and every dataset such as age, gender, Bp, etc. The clustered details are saved in a table and information's are compared with the details with the existing clusters. The user has to enter the details to find the disease prediction status and to find the solution in order to become a normal in abnormal data sets.

The characterization of items into different gatherings or the isolating of dataset into subcategories so that the information in each of the subcategory offer a typical article, regularly the vicinity concerning some clear space degree is known as Clustering. The cluster grouping has been surely understood in various circumstances and inclination to being demonstrated and valuable in numerous medicinal information mining applications. Clustering the therapeutic information into little with important information can help in the disclosure of structures by supporting the deliberation of a few suitable components from each of the accumulations consequently bringing together into the information and offering the utilization of standard information some assistance with mining strategies. Proposed methods are introduced underway for clustering and connected

with the famous K- Means grouping calculation in this methodology.

K-means clustering the information as per their individual qualities into k unmistakable accumulations. Information arranged into the indistinguishable cluster has alike element values. K, the positive number speaking to the quantity of accumulations, should be conveyed ahead of time.

## MODULE DESCRIPTION

### Set Normal Values

The admin saves the normal ranges of clusters for each and every data sets such as age, gender, bp, etc.

### View Patient Details

The admin can view the list of patient information's which are stored in the database and the records are displayed in grid view.

### Trace Patient Records

The admin can view and track the patient records based on the patient id and information.

### Transaction ID

While User enters the data to analyze the test results in this application a transaction id is created and the details are saved in that id.

### Clusters

The clustered details are saved in a table and information's are compared with the details with the existing clusters.

### Retrieve Clusters

The user has to enter the details to find the disease prediction status and to find the solution in order to become a normal in abnormal data sets.

## 6. RESULTS AND DISCUSSION

The general course of successful coronary illness forecast framework depends on the accompanying three stages

1. Data collection
2. Data pre-processing and
3. The classification of data.

Thus the proposed system achieved 99.9% of accuracy.

The data are collected from a standard dataset that contains 303 records. The 15 parameters, such as age, gender, chest pain type (CP), and cholesterol (chol), with some domain values associated with them, considered to predict the probability of heart disease are shown in the below table.

Parameters	Parameter description	Values
age	Age in years	Continuous
gender	Male or female	1= male
		0= female
the stb ps	Resting blood pressure	Continuous value in mmHg
cp	Chest pain type	1= typical type 1
		2= typical type angina
		3= non-angina pain
		4= asymptomatic

Parameters	Parameter description	Values
chol	Serum cholesterol	Continuous value in mm/dL
fbg	Fasting blood sugar	$1 \geq 120$ mg/dL $0 \leq 120$ mg/dL
restecg	Resting electrographic results	0= normal 1= having ST-T wave abnormal 2= left ventricular hypertrophy
thalach	Maximum heart rate achieved	Continuous value
old peak	ST depression induced by exercise relative to rest	Continuous value
exang	Exercise induced angina	0= no 1= yes
ca	Number of major vessels colored by fluoroscopy	0–3 value
slope	Slope of the peak	1= unsloping
	exercise ST segment	2= flat
		3= downsloping
thal	Defect type	3= normal
		6= fixed
		7= reversible defect
obes	Obesity	1= yes
		0= no
num	Diagnosis of heart disease	$0\% \leq 50\%$

## EXPERIMENTAL RESULTS

## PREDICTING HEART ATTACK FROM ACCESSIBLE PATIENTS MEDICAL DATASETS



[Update Cluster](#)
[Doctors List](#)
[My Records](#)
[Signout](#)

**CREATE CLUSTER**

User ID

Age

Gender

Chest Pain

Blood Pressure

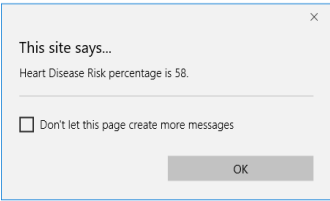
LDL Cholesterol (mg/dl)

Fasting Blood Sugar

Resting Blood Sugar


Resting ECG Results

**Create Cluster**



Generated Percentage

### Heart Attack Signs



Update Cluster

Doctors List

My Records

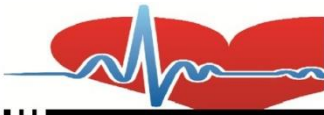
Signout

Signout

VIEW DOCTORS LIST

DoctorName	MobileNo	EMailID	City	AvailableTiming
Hariab	9214891394	hariab@gmail.com	Cuddalore	Morning 9:00-11:10 Evening 5:00-8:00
Renu	8667551679	renu@gmail.com	CHENNAI	morning 9:30 to 12:00

Doctor list




Normal Range

VIEW PATIENT RECORDS

ViewID	Age	Gender	ChestPain	BloodPressure	LDL	Cholesterol	FBS	FBIS	ECG	MaxHR	Rate	Angina	Dysp	Exercise	Slope	Thal	NumChol	Has
rajesh	22	Male	typical angina	200	150	120	160	Normal	75	Yes	Yes	Yes	Yes	Up	Sleeping	Normal	Healthy	10
rajesh	45	Male	typical angina	175	230	120	160	ST-T Wave Abnormality	100	Yes	Yes	Yes	Yes	Down	Fixed	Heart Disease	10	
krumar	45	Male	asymptomatic	160	140	120	140	ST-T Wave Abnormality	85	Yes	No	Yes	Yes	Down	Fixed	Sleeping Defect	Healthy	10
renu	12	45	Male	typical angina	160	220	90	160	Normal	82	Yes	Yes	Yes	Up	Fixed	Sleeping Defect	Healthy	10
renu	12	55	Male	asymptomatic	200	220	120	160	Normal	76	No	No	No	Flat	Normal	Healthy	10	

View the patients records




Normal Range

VIEW PATIENT RECORDS

NEW DOCTOR REGISTRATION

Doctor Name	<input type="text" value="Renu"/>
Mobile No	<input type="text" value="86675551679"/>
EMail ID	<input type="text" value="renu@gmail.com"/>
City	<input type="text" value="CHENNAI"/>
Available Timing	<input type="text" value="morning 9:30 to 12:00"/>
<input type="button" value="Register Now"/>	

Doctor Registration



Home

Sign Up

Login

View Normal Range

Contact Us

Attribute	AType	Description	EPercentage
Alcohol Consumption Continuous Regularly	5		
Alcohol Consumption Continuous Regularly	15		

View Normal Range

## 6. CONCLUSION

The finding of a type of heart illness is a huge and tedious undertaking in drug. The human services industry accumulates of huge procedure of heart illness information that unfortunately, are not "mined" to decided data for viable choice making by social insurance professionals. The term Heart problem incorporates the assorted illnesses that influence the heart. Cardiomyopathy and Cardiovascular sickness are a few classes of heart infections. The decrease of blood and oxygen supply to the heart prompts a heart attack illness. In this paper the information characterization depends on directed machine learning calculations which bring about precision, time taken to manufacture the calculation.

## REFERENCES

- [1] L. R. Martin, S. L. Williams, K. B. Haskard, and M. R. DiMatteo, "The challenge of patient adherence," *Therapeutics Clin. Risk Manage.*, vol. 1, no. 3, pp. 189–199, 2019.
- [2] M. A. Whooley, "To screen or not to screen?: Depression in patients with cardiovascular disease," *J. Amer. Coll. Cardiol.*, vol. 54, no. 10, pp. 891–893, 2020.
- [3] R. J. Hye, A. E. Smith, G. H. Wong, S. S. Vansomphone, R. D. Scott, and M. H. Kanter, "Leveraging the electronic medical record to implement an abdominal aortic aneurysm screening program," *J. Vascular Surg.*, vol. 59, no. 6, pp. 1535–1543, 2021.
- [4] T. Clayton, J. Lubsen, S. Pocock, Z. Vok'o B. Kirwan, K. Fox, and P. Poole- Wilson, "Risk score for predicting death, myocardial infarction, and stroke in patients with stable angina, based on a large randomised trial cohort of patients," *Biomed. J.*, vol. 331, pp. 869–873, 2019.
- [5] D. M. Lloyd-Jones, P. W. Wilson, M. G. Larson, A. Beiser, E. P. Leip, R. B. D'Agostino, and D. Levy, "Framingham risk score and prediction of lifetime risk for coronary heart disease," *Amer. J. Cardiol.*, vol. 94, no. 1, pp. 20–24, 2020.
- [6] W. Levy, D. Mozaffarian, D. Linker, S. Sutradhar, S. Anker, A. Cropp, I. Anand, A. Maggioni, P. Burton, M. Sullivan, B. Pitt, P. Poole-Wilson, D. Mann, and M. Packer, "The seattle heart failure model: Prediction of survival in heart failure," *Circulation*, vol. 113, pp. 1424–1433, 2019.
- [7] S. Alty, N. Angarita-Jaimes, S. Millasseau, and P. Chowieńczyk, "Predicting arterial stiffness from the digital volume pulse waveform," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2268–2275, Dec. 2018.
- [8] S. Chattopadhyay, "Mining the risk of heart attack: A comprehensive study," *Int. J. Biomed. Eng. Technol.*, vol. 11, no. 4, pp. 394–410, 2016.
- [9] X. Song, A. Mitnitski, J. Cox, and K. Rockwood, "Comparison of machine learning techniques with classical statistical models in predicting health outcomes," *Medinfo*, vol. 107, no. Pt 1, pp. 736–740, 2020.
- [10] R. Snyderman and J. Langheier, "Prospective health care: The second transformation of medicine," *Genome Biol.*, vol. 7, pp. 104.1–8, 2021.
- [11] J. Wiens, J. Guttag, and E. J. Horbitz, "Patient risk stratification for hospital- associated C. Diff as a time – series classification task," in *proc. Neural Inf. Process. Syst.*, 2020, pp. 476–484.
- [12] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berphold, M. Schwab, C. R. Antonescu, C. Peterson, and P. S. Meltzer, "classification and diagnostic prediction of cancer using gene expression profiling and artificial neural networks," *Nature Med.*, Vol. 7, No. 6, pp. 673–679, 2020.
- [13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training algorithm for optimal margin classifiers," in *proc. 5<sup>th</sup> Annu. Workshop Comput. Learning Theory*. 2018, pp. 144–152.