

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462

🕥 6381 907 438

🛛 🖂 ijircce@gmail.com

🙋 www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205294 |

A Novel Approach to Heart Disease Prediction using Xgboost

Punith kumar N, Rajashekara G N, Sahithi M R, Talari Sai Sreenivasulu, Mrs. Pavithra

U.G. Student, Department of Information Science and Engineering, SJB Institute of Technology, Bengaluru,

Kengeri, India

Assistant Professor, Department of Information science and Engineering, SJB Institute of Technology, Bengaluru,

Kengeri, India

ABSTRACT: Heart abnormalities are It is one of the leading causes of loss of life in the world Patients frequently have no symptoms until fatal events occur, requiring skilled personnel to capture cardiac abnormalities even in their care In recent decades electrocardiogram (ECG) [1]. Using a virtual format would provide evidence of increased machine capacity utilization detected to encounter such abnormalities Machine-getting to know (ML) has shown to be effective in predicting profitability on it in how beautiful [2]. This study aims to develop machine learning definitions of cardiovascular disease based on appropriate features. We used the UCI the benchmark data set for cardiovascular prediction in this study, which includes 14 specific parameters associated with heart failure [3] Support vector machines, logistic regression, artificial neural networks, K-nearest on special neighbors, Naïve bays, and decision trees are just a few of the scripts that this tool is based on. Trendy feature selection methods relief minimum redundancy maximum relevance, minimum Absolute shrinkage selection operator and adjacent classes were used said of the article to eliminate improprieties and eliminate inappropriate [4]. With approximately 40 thousand ECGs tagged through cardiologists in specific hospitals and countries, our fashions can cope with 7 unique alerts: normal, AF, heart failure, heart failure, heart failure, heart failure cavity, or other noise prediction models using multiple integrations and multiple statistical characteristic methods [5]. Using an XGBoost rule from the first gadget learning method, we already use the version-derived F1 rating at zero-nuety-three - zero-nine. To our knowledge, this is the first model reporting high performance across hospitals, international locations, and file types [6]

KEYWORDS: Heart disease, Machine Learning, Feature Selection, XGBOOST Algorithm.

I.INTRODUCTION

Heart disease is considered a threat to human life worldwide. According to statistical reports from the World Health Organizations, 17.9 a million individuals will perish from cardiovascular disease (32% of all global deaths) in 2019. [1] Many lives can be saved by detecting HD and appropriate interventions Early diagnosis and treatment of many heart diseases is exceptionally difficult, especially in developing countries, due to the lack of accredited medical centers and. accurate diagnosis of other factors affecting inflammatory heart disease can be reduced Atrial fibrillation (AF) and the typical doses, affecting 1-2% of the total population can lead to stroke, heart attack or myocardial infarction [4]. An issue with the diagnosis of AF is that it is very routinely asymptomatic (discovered incidentally in electrocardiograms of 30–forty-five patients for unrelated reasons [19]) and requires trained personnel them to detect problems from electrocardiograms (ECGs). Doing so could be a deadly opportunity for injury. Similarly, tachycardia (increased heart rate) are not rare heart conditions Although not as harmless as AF, They have the potential to seriously complicate cardiac arrest, if untreated) procedures are increasingly timeconsuming in global health care , according to guidelines for standardized diagnostic methods. In fact, the algorithm can be useful amo in many ways

II. RELATED WORK

The confusion surrounding the new aesthetic comes from its definition, as it includes specific (undefined) details of the heart, making it miles harder for the statue to determine its exact nature. which will be reclassified Figure 2 shows the essential model and its features. The CSI index is highly descriptive in classifying the signal as normal or new, and is actually a popular index used in the literature to capture anomalies in the ECG with R-R intervals greater than 50 seconds is the same as a person with atrial fibrillation has an irregular heartbeat surrounded by non-current devices. i. Less Accuracy

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205294 |

ii. Low Efficiency

III. METHODOLOGY

1) Upload ECG Dataset: This module allows us to upload datasets into the program.

2) Dataset Preprocessing:Using in this module, we'll learn all the dataset values then remove the missing values then normalize and shuffle the values

3) Train and Test Split: The processed values can be split into train and test with 80% used for training and 20% used for testing the data set.

4) Run XGBoost Algorithm: In order to determine the prediction accuracy, we will use this module to train a set of XGBOOST rules on 80% of school data and then test them on the remaining 20% of the data.

5) Comparison Graph: We will use this module to create equality and equality graph

6) Predict Heart Disease from Test Data: This module will be used to upload the test records after which the utility will check the facts and assign an expected label as regular or coronary heart attack. To run the mission can be found below by double clicking on the run.Bat file

IV. EXPERIMENTAL RESULT

To initiate the project, double-click on run.Bat file to access the screen below In above display click on 'Upload ECG Dataset' button to upload dataset and get under outpu

	ADODE FIEDGIANT.			- s b .
	Machine Learn	ting for Real Time He	art Disease Prediction	
Upload ECG Dataset	Dataset Preprocessing	Trais & Test Split	Run XGBoost Algorithm	
Comparison Graph	Predict Heart Disease from	Test Data		
				Activate Windows Gato Settingene activate Windows
Inchese Lawreng fai Real Time Heart D				
				- 0
Open		×	art Disease Prediction	
Open + + landd + ECOC Stylence + New folder	eN + Detanat ~ 6 Search Da	× 101 A	rt Disease Prediction.	
Open + lan21 + ECOD Organize + New Folder + Guick scores	energy Production eN + Determine → Date model	x toot p III • III • Ind Type	art Disease Frediction. Run XGBeost Algorithm	
Open + in - land) + ECOCH Trgance + Face folder Calck access Creditive Calck access Creditive	eN + Detent C Search Date Date result data core Boy 31-07-302 Boy 31-07-302	Land p III - III O Tod Type Table Macand Land I ISO Macand Land I	rt Disease Prediction.	
Open + in and + ECOD Ingence + New Folder Conclose Conclose Conclose The PC The PC	PL + Default	X Int J Int - D Ind Type TRUK Microsoft Incel (Int D Microsoft Incel (et Disease Production. Run XGBsost Algorithm	
Open + is - land + ECCC Tiperare + Plane Todae Clark access Clark ac	PA + Default	taset p III + III • Tod Type III.20 Microsoft East 1 III.20 Microsoft East 1	art Disease Production. Ran XGBsost Algorithm	
Open	 Portunati w (b) Search Date Date recent Date recent Date recent Maximum 21-07-2021 Second Date 	teet p III + III O Tota Type III + Meccard Level (III + Meccard Level (art Disease Prodiction.	
Open	N + Detent w C Search Da Data result Infance 31-01-302 Index 99 11-01-302	Kant p iii - Di o hol 798 Table Meccent Lond L Meccent Lond L	art Disease Production.	
Copen	Ne - Detent Ditection	Anati p iii - Iii Iii - Iii Iiii - Iiii Iiii - Iiii Iiii - Iiii Moorent Even I Moorent Even I	er Disease Production	
F Open T open T open T open to the total T open total	PL + Default	Kand p III - III p III - III p III - III p III - III - III - III III - III - III - III - III Microsoft Ford (rt Disease Production. Run XGBsort Algorithm	
F Open	PD + Default up (), Search Da Date recent Ideator Date recent Date	Examination of the second seco	rt Disease Production.	
F Open	PAI + Default	Acad p III - III o III O IIII O IIII O III O III O III O III O I	rt Disease Production.	
Popen Popen Popence Parent Marco Popence Parent Marco Popence Parent Marco Popence Parent Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Popence Pope	NP + Default	Kandi pi III - III P III P IIII P III P IIII P III	rt Disease Production. Ren XGBsort Algorithm	
Popen	MP + Default	Konti pi Illi - Illi P Inter Type Inter Moreauth Level 1 Norreuch Level 1 Center	rt Disease Production. Ren XGBsort Algorithm	
Open *********************************	Mar Paul Lan Mar Control Cont	Konti pi Illi - Illi P Into Type Into: Microsoft Ecolor Into: Microsoft Ecolor Cencit	art Disease Prediction.	

Select and upload the 'HeartData.csv' file to the screen above and after that, press "Open" to load the dataset.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205294 |

and get the output below

Upload ECG Dataset	Dataset Preprocessing Train & Test Split	Run XGBeest	Algorithm							
Comparison Graph	Predict Heart Disease from Test Data	1	🛞 Figure 1							×
ee/VeukatJan23/ECGCNN/Dat	aset HeartData.csv loaded		-		Data	set La	bels Found	in Dat	aset	
V0 V1 V2 V3 28.650031 3.874866 4.063665	V4 V76 V77 V78 V79 Label -1.037575-0.108496 0.410211 -0.465336 0.400407 -1.	725821 Normal	200 -							
9.474361 -5.042980 -2.972559 - 11.151052 -5.831333 -1.944207 162.099300 -6.037885 -8.439421	L368801-2.7354171.138248 2.076897 0.142733 -0.6 -1.808658 -1.3531091.755086 -0.518044 -0.385620 1 -2.570125 -0.358789 5.679461 3.399706 -0.183589 4	683443 Normal .025638 Normal .227241 Normal	₹ ^{150 -}							
-12.088207 -3.678261 -3.648417	-2.322801 -1.208475 0.912765 1.538851 1.125998 2/	002777 Normal	8 100 -							
4 -40.265626 -5.658991 -1.43713 4 -39.822603 -5.387832 -1.31920	9-3.395375-0.659091 3.060949 0.389245 -3.639728 - 7-2.864167-0.545985 35.501952 -22.565685 27.672386	-3.961224 Noisy 5 33.516999 Nois	50 -							
-40.565438 -5.112076 -2.70454	a -39/4050 0.005209 = -0.444542 0.014301 -1.500515 9-2.653160 0.072387 0.845234 1.351859 0.088758 / 2.3693050 0.701652 1.581385 0.755186 0.055145	0.182930 Noisy 0.182930 Noisy	0	2	4		2	7	5	
S rows x SI columns]	0-2-072070-0.791002 - 1.001555 -0.275100 0.007145 -	-0.477704 (1009		9452	hythm	dycard	Not	Narm	8	hucard
els : ['AF' 'Arrhythmia' 'Bradyca	rdia' 'Noisy' 'Normal' 'Other' 'Tackycardia']				Her.	ga	Labels			Tac

Above display screen dataset is loaded and It is visible to us dataset of numeric and non-numeric data but gadget earning algorithm will take numeric data most efficient so we have to first process the facts to convert into numeric and x-axis in the above graph represents the name of coronary heart failure and the Y-axis represents the dependence on the statistics for that disorder observed within the dataset. Click the upper graph now, and then click the 'Dataset Preprocessing' button to execute the dataset technique and get the output

Machine Learning for Real-Time Heart Dise	ase Prediction					
	Machine Leas	uing for Real-Time Ho	art Disease Prediction			
Upload ECG Dataset	Dataset Preprocessing	Train & Test Split	Run XGBoost Algorithm			
Comparison Graph	Predict Heart Disease from	a Test Data				
Dataset processing completed. Nora	nalized dataset values					
[[1.67220456 -0.06732135 -0.22024 0.42943089]	4210.2412331 -0.47867198					
[-0.52651126 -0.05790945 -0.04957: 0.13813453]	5520.42153413 -0.40566843					
[-1.2469493 -0.09360907 -0.031073 -2.47195493]	96 3.81964784 2.20161991					
	6690.21978892 0.04797235					
[-0.11033704 -0.10068871 -0.095825	5090.01332763 0.03984782					
[2.64723545 -0.03570871 -0.303414 -0.7289063]]	453 0.20224672 0.75147949					
				Activate Windo		

All the values on the aforementioned screen have been transformed into calculations and you can now click on the 'Train & Test Split' button to split the dataset into trains and test to get the output below

|| Volume 12, Issue 5, May 2024 ||

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

| DOI: 10.15680/IJIRCCE.2024.1205294 |

	M	chine Learn	ing for	Real-T	ime He	art Dive	ese Pro	liction			
Upload ECG Dataset	Dataset Prepre	® figure 1								- = ×	
Comparison Graph	Predict Heart I			XG	Boost (Confusi	on mal	rix		- 10	
Boost Accuracy : 99.0		chycardia -							40	-	
Boost Procision : 99.0 Boost Recall : 98.0 Boost FScore : 97.0		Other -					٥	43	-	- 40	
		Normal -							0	- 30	
		Noisy -				37	0		0		
		adycardia -		o	50	0			ø	- 20	
		rrhythmia -	0	42	1				0	- 10	
		AF -	45	0					(0)		
			N	dimia -	ardia -	- Voisy -	ormal -	other -	ardia -	- 0	
		# + >	+	Q₽	8		3		-		
						_	_	_	_	Activ Gene	rate Windows defings to activate the

In the screen above where XGBOOST we achieved 99% accuracy as well as in the confusion matrix graph the x-axis represents Predicted Labels and the y-axis represents True Labels and all colored boxes represent correct prediction estimates and blue boxes have incorrect prediction too conservative approximations.Now close the top graph and then click on the 'Comparison Graph' button to get the bottom graph

In above screen selecting and uploading 'testData.csv' file and then click on 'Open' button to load dataset and get below output

	Machine Learning for Real-TL	me Heart Disease Prediction	
Upload ECG Dataset	Dataset Preprocessing Train & Test Spi	it Ran XGBoost Algorithm	
Comparison Graph	Predict Heart Disease from Test Data		
t Data = 75.44193439 1.715340	27 -6.86944371 0.70147015 0.44350455		
16936085 1.10200342 0.973296 69476687 1.46242979 2.914837	33 -0.19007141 -1.24271575 74 -1.06890934 -0.45505124		
31924174 0.50601996 1.001871 26634878 2.59423241 -0.594583	49 -3.35111716 0.99005798 82 -0.86293525 -0.68615121		
75636874 -3.0957992 3.1161274	4 -0.9571896 -1.45955831		
.67375734 5.83186999 -1.614810	52 0.36113281 -6.02425395		
66870583 11.22404399 5.332209	95 2.49573898 -2.25146065		
33342932 2.33242909 -1.00/340	01 2.12204/48 3.982192/4 19 3.18286626 0.2032702		
16930506 15.70601491 -16.64922	447 4.60385598 -27.42750218		
.76521052 -24.50836456 -0.16715	015-23.33685838 8.8324224		
.95328224 -13.19806634 27.10181	265 -3.1954281 2.17907715		
.75523316 0.32984874 19.440513	06 -7.87354958 -9.1584596		
35999817 -12.6348126 -6.805011	12 -2.54748976 -9.10691919		
10403324 -3.000/9291 -2.418321	03 2.29131087 -0.85663384j	Normal	
st Data = [1.97607711e+02 -1.8946	64662e+00-1.30536539e+01 4.53802488e+00		
66560426e-01 9.64003977e-01 6.9	03089137e+00 -1.06810045e+00		
87167357e-01 2.22257825e+00-1	26364907e+01 -1.75566193e+00		
69526185+00 -6.67286562+00 -5	13070184e+01 1.33046617e+02		
39886632e+01 2.17567882e+01 3	00273189e+01 -1.38503443e+01		
and a second sec			



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/LJIRCCE.2024.1205294 |

In the square cell of the top screen we can see the ECH signal test data and after the arrow symbol = we can see the predicted output and scroll down the output screen to view the new prediction output

Machine Learning for Real-Time Heart Dise	All and the second s	ration for Bard Time II	and Disease Readinging	- 0
	Stating Liv	Thorg for Acht Thee Is	ALL DESCRIPTION OF A	
Upload ECG Dataset	Dataset Preprocessing	Traia & Test Split	Ran XGBoost Algorithm	
Comparison Graph	Predict Heart Disease fro	n Test Data		
-2.13078235e+00 -5.14964285e-01	4.40274890e+00-2.86059096e+00	Predicted Ax :	achycardia	
Test Data = [-5.27190964e+01 -5.35 -5 18515484e-01 -2.65622600e+00	286131#+00 -1.31251950#+00 -2.3 1.49620365#+00 -1.30471617#+00	6348522e+00		
-4.55899696e-02-1.95238932e+00 1.04108548e+00-1.95593020e+00	1.16413542e+00 3.24101917e-01 9.92565875e-01 9.36739870e-01			
-1.04160401++00 2.49752660+-01 -9.82417249+-01 1.84895332++00	2.40675800e+00 -2.01236764e+00 9.67748456e-01 4.64753114e+00			
-1.43407097e+01-2.69886471e+01 1.12417640e+00-2.37251299e+00	3.70277241++00 2.08779946+-01 3.05598314++00 9.64493484+-01			
6.40222106e+00 6.23577572e-02 - -4.70071812e-01 4.08190797e-01 1	3.95573995e+00 2.39811313e+00 1.96044250e+00 -3.03817417e+00			
1.83478188e-01 -2.72505422e+00 -	1.39787796e+00 -1.83196972e+00			
8.60552183e-01-4.76455255e-01-3	8.71029897e-01 -2.90235545e+00			
-1.75626621e-02 -6.61198548e-01 -	1.35364449e+00 1.35997806e-01 1.23240574e+00 -3.12827477e+00			
-3.73414842e-02 1.28819161e-01 2 1.51004001e+00 1.10475694e+00	5 15541232e 01 -8 57753019e 01			
5.74875735e-01 7.77901222e-01-1	1.52560391e+00 1.78602305e+00			
-5.39212936e-01 -3.27339028e-01 - -5.12924986e-01 -1.03569160e+00	7.73304463e-01 -1.61977651e+00 6.41700041e-01 -1.24633385e+00]	Predicted As : B	radycardia	
Test Data = [-1.78614242e+00 5.03	588703e+02 3.51539137e+00 2.8	2349990e+00		Activate Windows

V. CONCLUSION

In this paper, we suggest a unique approach to discover cardiac abnormalities from freshly recorded ECGs. The prediction techniques can be summed up like this:

signal preprocessing, feature filtering, version schooling, size and evaluation. We formulate a function extraction pipeline of 110 capabilities, with which we educate 5 unmarried fashions on 3 collected information sets. Our models show overall overall performance predictions which might be very strong with admire to unobservable facts, however can be generalized to facts sets of ECGs recorded in precise contexts, and to populations with precise inherent characteristics. Bradycardia, different (non-exceptional), arrhythmia and noise. To correspondingly boom the reliability of our models, we use Temperature Scaling to lessen the anticipated size errors in our put on measurements. Our fashions confirm that on the spot search for symptoms of the QRS complicated gives excessive prediction accuracy. It needs to be expansive functionality effect on the lives of people tormented by heart illnesses. In truth, we expected our paintings to be implemented in a real time setting, with a wearable tool that can continuously show the coronary heart beat of the sufferers at danger.

REFERENCES

- 1. Chapman University and shaoxing people's hospital. https://figshare.com/bibliography/ChapmanECG/4560497/1,
- 2. Tianchihefei High technology cup ecg man-machine intelligence competition. http://hangzhou.oss-cn-hangzhou. aliyuncs.com/231754/round2/hf_route2_train.zip, .
- 3. Alivecor, Inc. https://www.alivecor.com/#, https://www.alivecor.com/#.
- 4. T.S., eds. Akiba, S. Sano, T.S. Yanase, T. Ohta, M. Koyama. Optuna: A next generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,
- 5. Z. D. G. Ari L. Goldberger A. Schwilkin Editor-in- Chief. Goldberger's Treatment in Electrocardiology. https://www.sciencedirect.com/topics/ medicine-and- dentistry/grs-complex, 2017.
- 6. ZI Attia, PA Nakworthy, F Lopez-Jimenez, SJ Asirwatham, AJ Deshmukh, BJ Gersh, RE Carter, X Yao



INTERNATIONAL STANDARD SERIAL NUMBER INDIA







INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

🚺 9940 572 462 应 6381 907 438 🖂 ijircce@gmail.com



www.ijircce.com