IJIRIS: International Journal of Innovative Research in Information Security
Volume 09, Issue 03, June 2023
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

A Method to Improve Human Heart Disease Prediction Using Machine Learning Algorithms

Jenifer A,
Assistant Professor, Dept. of. CSE,
Sri Sairam College of Engineering, Bangalore.
jenifera.cse@sairamce.edu.in
Pratyush Priyadarshi,Harini R,Gayathri Devi K
Student, Department of CSE
Sri Sairam College of Engineering,Bangalore
sce20cs088@sairamtap.edu.in, sce20cs082@sairamtap.edu.in
sce20cs097@sairamtap.edu.in

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.09/Issue03/JNIS10104
Research Article | Open Access | Double-blind Peer-reviewed |Article ID: IJIRIS/RS/Vol.09/Issue03/JNIS10104
Received: 02, June 2023 | Accepted: 20, June 2023 | Published: 23, June 2023
Volume 2023 | Article ID JNIS10104 http://www.ijiris.com/volumes/Vol09/iss-03/25.JNIS10104.pdf
ArticleCitation: Jenifer(2023).A Method to Improve Huam Heart Disease. International Journal of Innovative Research in Information Security (IJIRIS), 10, 195-199
doi: https://doi.org/10.26562/ijiris.2023.v0903.25
BibTex key: Jennifer**2023Method**

Copyright: ©2023 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

ABSTRACT:

There is great diversity in the field of medical science due to computational power and technical innovation, especially in identifying human heart disease. Today it is one of the deadliest human heart diseases in the world and have very serious effects on human life. Accurate and timely identification of heart disease in humans can be very helpful in prevent heart failure in its early stages and will improve patient survival. Manual method for determining the heart disease is biased and can vary between researchers. In this regard, efficient and reliable machine learning algorithms resources for detecting and classifying people with heart disease and those who are healthy. According to suggestion in our study, we identified and predicted heart disease in humans using a variety of machine learning algorithms and using heart disease dataset to evaluate its performance using various measures, such as sensitivity, specificity, F-measure, and classifier accuracy. For this purpose, we used nine machine learning classifiers for the final dataset before and after hyper parameter tuning of machine learning classifiers, such as AB, LR, ET, MNB, CART, SVM, LDA, RF, and XGB. In addition, we verify their accuracy on a standard heart disease dataset by performing several standardized, pre-processing procedures of the data set and hyper parameter tuning. In addition, to train and validate machine learning algorithms, we implemented standard K-fold cross-validation technique. Finally, the experimental results show that the accuracy of the predictive classifiers with improved hyper parameter tuning and achieved remarkable results with data normalization and hyper parameter tuning of machine learning classification.

1. INTRODUCTION:

First of all: A person may have heart disease and not feel annihilated. It is indicated for some groups with heart disease. This is when there are physical changes or pain that indicate the presence of disease. Signs of heart disease include:Chest pain, difficulty breathing, palpitations (a tendency for the heart to beat abnormally fast), swelling of the legs and feet, weakness due to insufficient blood supply to the body and brain, cyanosis (skin bluish discoloration). Heart disease has many risk factors.age, sex, smoking, physical inactivity, inappropriate drinking, unfavorable eating habits, body weight, genetic predisposition and family history of cardiovascular disease, elevated heart rate (hypertension), elevated blood sugar (diabetes), Elevated blood cholesterol levels (hyperlipidemia), undetected celiac disease, psychosocial variables, poverty and low educational attainment, air pollution. Individual exposure to each hazard varies by network and ethnic group, but overall exposure to these hazards is very reliable. Some risk factors, such as age, sex, and familial/genetic predisposition, do not change. In any case, many important cardiovascular risk factors can be altered by lifestyle changes, social changes, and drug treatments (such as the development of hypertension, hyperlipidemia, and diabetes). Overweight people are at increased risk of coronary atherosclerosis. Age is a major risk factor for developing cardiovascular disease and heart disease, and the risk increases significantly with each stage of life. Fat streaks in the coronary arteries can spoil youthfulness. A number of explanations have been proposed to clarify why the risk of cardiovascular/heart infections increases with age. One of them corresponds to serum cholesterol levels. Absolute serum cholesterol levels increase with age in many populations step.

For men, this increase plateaus around age 45 to 50. In women, there is a marked increase between the ages of 60 and 65.Diagnosing heart disease is perhaps the most basic and challenging task in the health services sector. Saving lives requires fast, productive and accurate analysis. Patients must undergo numerous tests, and the results must be carefully reviewed by medical professionals. For this reason, scientists attach great importance to heart disease prediction and have developed various heart disease prediction frameworks using different AI calculations. Some of them gave better results than others. Many companies used the well-known UCI cardiology dataset to prepare and test their classifiers, while others freely used information from various clinics. Therefore, prevention of heart disease has become more important than necessary. A better information-driven framework for predicting heart disease will improve overall research and action, enabling more people to live healthier lives. AI can help predict heart disease, and the predictions are very accurate. This includes researching cardiac patient records and preparing information accordingly. At this point, various models were prepared and expectations were built using various computations such as ANN, SVM, Decision Trees, Naive Bayes, Random Forest, and Logistic Regression.

2. RELATED WORKS:

Related Works classifier to organize information into her five classifications. Four class classifications were performed using this method.None, normal, low, high, very high. This method uses information to predict the likelihood of coronary artery abnormalities. Table 1 shows the dataset used for testing. The proposed method showed an accuracy of 88.96%. Das et al. 1 implemented a framework that uses ANNs with ensemble strategies. For this we used the dataset shown in Table 1. Group models have led to increased guesswork by integrating different models developed for similar projects. The device used to conduct the tests was SAS Project Excavator

5.2. Results showed that the model predicted coronary artery abnormalities with an accuracy of 89.01. Chen et al. [13] created a framework for Heart Disease Day Prediction (HDPS) using ANNs. In this study, learning vector quantization (LVQ) was used. The ANN used in the model has 13 neurons in the information layer, 6 neurons in the secret layer, and 2 neurons in the profit layer. The decision tree J48 calculation used decision tree calculations for cardiovascular disease selection and cardiovascular disease prediction. Thirteen clinical credits/highlights were used in the database and 240 clinical records were collected during test preparation and rest periods. When using all highlights, the achieved accuracy was 75.8333%. On the other hand, highlight selection improves accuracy to 76.67%. In addition, accuracy improves to 85% as more unwanted highlights are removed. The article argues that the J48 computation allows for minimal highlight selection and improves the expected accuracy. Patel et al. [16] In 8, several choice tree computations were analyzed using his WEKA device based on the dataset to determine the presence or absence of heart disease. Various calculations were attempted, including J48, strategy model tree, and irregular outback, but J48 surpassed them all with 56.76

°Curacy. implemented a K nearest neighbor (ANN) to predict coronary artery anomalies using the Cleveland dataset. Voting divides the information into modules and applies a classifier to each module. The evaluation was completed with 10 superimposed mutual recognitions.

Sabarinathan et al Decision tree J48 calculations used decision tree calculations for cardiovascular disease selection and cardiovascular disease prediction. 13 clinical credits/highlights were used in the database and 240 clinical records were collected for test preparation and rest periods. When using all highlights, the achieved accuracy was 75.8333%. On the other hand, highlight selection improves accuracy to 76.67%. In addition, accuracy improves to 85% as more unwanted highlights are removed. The paper claims that the J48 computation allows the selection of minimal highlights, improving the expected accuracy. Patel et al. [16]8, several selection tree computations using the WEKA apparatus were analyzed on the dataset to determine the presence or absence of cardiac disease. Various computations tried were J48, strategy model tree, irregular outback, etc. J48 beats them all with 56.76% accuracy. Showman et al. implemented K nearest neighbors (KNN) to predict coronary anomalies using the Cleveland dataset. Voting is the technique of dividing information into modules and applying a classifier to each module. The evaluation was completed using 10 overlay mutual recognitions. The results clearly showed that the unvoted accuracy increased from 94% to 97.4% with different case numbers for

K. At K=7, the accuracy was the highest at 97.4%. In any case, using democratic procedures did not improve accuracy. The results showed that at K=7 the precision dropped to 92.7%. Wihart et al. Using a dataset (UCI) to analyze coronary artery anomalies, he examined the accuracy of SVM calculations. The investigation includes various SVM types such as Scaler preprocessed Decision Direct Asymmetric Graph (DDAG), Binary Tree Support Vector Machine (BTSVM), One-Against-All (OAA), Exhaustive Output Error Correction Code (ECOC) was included. (Minimum Maximum). The next step is to perform the calculations completed using the SVM calculation above. In a trade show evaluation, BTSVM performed well compared to other computations, with an overall accuracy of 61.86%.

Pourier et al conducted extensive research on different characterization strategies within the Cleveland database to find out which classifiers perform better than others. Classifiers included were Single Bond Rule Learner (SCRL), Decision Tree (DT), Multilayer Perceptron (MLP), Naive Bayes (NB), Radial Basis Function 9 (RBF), and others. The authors used his K-fold cross-validation. A procedure for evaluating the best classifier. The evaluation measures exhibited in each classifier included accuracy, authenticity, validation, and ROC curvature. After trying different estimates of K for each individual ANN classifier, we found K= 9 to be optimal. ANN tested several neurons, 13, 2, and 7, to show the best combination. The investigation he carried out in two modules.The first step was the comparison of the various classifiers mentioned above, and the subsequent step was the application of the detection-method.

3. METHODOLOGY:

DATAPRE-PROCESSING:

The representation and accuracy of predictive models are influenced not only by the calculations used, but also by the type of data set and its strategy. This preprocessing phase is important as it prepares the data set and conforms it to the structure that will be processed in the computation. Another important point is the magnitude of change in the dataset. Some datasets have a large number of attributes, making it more difficult to analyze them by computation, find examples, or make accurate predictions. Such issues can be resolved by examining datasets and using appropriate information preprocessing techniques. Information pre-processing steps include:Information cleansing, information modification, missing quality attribution, information standardization, data determination, and other advances based on dataset thinking. 5 Datasets may contain erroneous information, missing information, redundancies, inconsistencies, and numerous other problems, and it is not acceptable to use them for AI calculations as they are due to erroneous information.

PERFORMANCE EVALUATION MATRICES:

The following measurements are used by scientists to evaluate the expectations model and present its presentation results. Without delving too deeply into complex mechanics and numerical constraints, we briefly define each technique.

ACCURACY = (TP+TN) / (TP+TN+FP+FN)PRESICION=TP/(TP+FP)
SENSITIVITY=TP/(TP+FN)
F-MEASURE=(2*PRECISION*RECALL)/(PRESICION+RECALL) ROCCURVE,Y=P /(P+N)

Accuracy is the most widely used evaluation metric in presentations and is used in all exam questions covered in this article. And the focus of this review article is to rank, compare, and review past works base done their accuracy.

HEART DISEASE DATASET:

The dataset used in most studies is the Cardiology dataset obtained from the UCI.(University of California, Irvine) Center for AI and Intelligent Frameworks. Contains 4 Information database of four clinics. Each dataset has a similar number of highlights, 14 so far A set of different records. The Cleveland dataset is the most commonly used dataset by AI scientists. Because it has more records with fewer attributes missing than others. of The "Number" field indicates the presence of heart disease in the patient. Number estimated from 0 (No) The Cleveland dataset contains 303 instances. various record attributes Their values are described in Table 3.3.1.

Table:3.3.1:DATASETATTRIBUTES

ATTRIBUTE	VALUESANDMEANINGS
Age1	Age in year
Gender1	Value1and0 formale and female
Cp1	Pain in chest yes/no
Testbps1	Blood pressure during resting
Chol1	Cholesterol of serum in mg/dl
Fbs1	Blood sugar during fasting
Restecg1	Resting ECG results
Oldpeak1	ST depression induced byexercise relative totest
Slope1	The slope of peak exercise of ST segmentValue 1: unslopingValue2:flat Value3:downsloping
Ca1	Number major vessels (0-3) coloured by of fluoroscopy
Thal1	3=normal 6=fixed defect
Num1	Diagonal of heartdiseaseValue0:noriskVal ue 1: low risk Value 2: risk Value 3:highrisk Value4:higher risk

APPLYING VARIOUS MACHINE LEARNING ALGORITHMS

LOGISTIC REGRESSION

This algorithm is used for learning grouping calculation used to fore see the similarity of an objective variable. The main idea of ward variable is dichotomous, which implieseither1or0.Numerically, a logistic regression model predicts P(Y=1) as a component of X. It is one of the least complex ML calculations that can be utilized for different grouping issues, for example, spam identification, Diabetes expectation, disease location ,and so forth.

KNN ALGORITHM

This algorithm performs computations that are simple, governed AI computations. Used to address both placement and fallback issues. But it'snot hard to check by running Due to its large size, it has the majordrawback of being very small The information we use evolves. Useinformation from several classes to predict: Characterization of news ample points.

SVM

Backing Vector Machines (SVMs) are amazing yet adaptive AI-driven computations Used for both treatment and relapse. In any case, they are generally used characterization problem. Unlike other AIs, SVM shave their own way of executing calculation.

NAIVE-BAYES

This algorithm performs computations that rely primarily on Bayesian computations. Sentence. this is nothing else A single calculation is performed, but a group of calculations where each calculation performs a common calculation. In a typical standard,for example,each pair of placed highlights arefree of each other. Gullible Bayesisa probabilistic AI computation that can be used in a variety of commissioned projects. Common uses include filtering spam, group in greports,composing messages,etc.

DECISION TREE

Decision tree algorithms belong to the group of regulated learning computations. In contrast, Other regulated learning computations canuse decision tree algorithms to deal with recurrence. And characterization issues. The purpose of using decision trees is to provide: A model that can be used to predict the class or guess of the target variable by inclusion Simple selection guidelines derived from historical data (training information).

RANDOM FOREST

Random forests are an effective group learning strategy for sequencing, recurrence, and other errands. Through the development of a large numberof selection trees in preparation and the method of classes(characterization) or moderate/normal expectations. It is also probably the most commonly used one Calculation due to its simplicity and versatility.

4. CONCLUSIONS

This task describes different machine learning grouping strategies for classifying heart disease. A number of exemplary works on the use of AI techniques were viewed and categorized. The accuracy of the proposed model depends on the device and dataset used. The quantity of quality is recorded in the dataset. Improved information gathering. Understand datasets. I prepared various machine learning calculation records and tried them out. Among all computations ,ANN and random forest computations showed the best performance. Accuracy is 88.52%.We believe that creating an accurate heart state requires a dataset with enough examples and laws. Take advantage of the information. The corresponding calculations should also be performed during construction. Finally, using AI to classify heart disease is very important. This can help both healthcare professionals and patients. The field is still developing, and although clinics and medical centers have easy access to patient information, little information is widely available.

5. REFERENCE

[1]. Kononenko, "Machine Learning for Medical Diagnosis: History, State-of-the-Art, Perspectives," Artificial Intelligence in Medicine, vol. 23, no. 1,pp. 89–109, 2001.

[2]. J. Soni et al., "An intelligent and effective System for predicting heart disease using weights," "Associative Classifiers,"International Journal on Computer Science and Engineering, vol. 3, No. 6 ,pp. 2385–2392, 2011.

[3]. N Khateeb and M. Usman, "Efficient Systems forCardiac Disease Prediction Using k-Nearest Neighbors." Proceedings of theInternational Conference on Neighborhood Classification Techniques, Big Data and the Internet of Things(BDIOT),NewYork,NY,USA:ACM,2017,pp.21-26.

[4]. Amer, "Research on Accuracy of Medical Data Mining Techniques,"International Journal of Computer Applications, vol. 168, No. 3, pp. 12-17,June 2017.

[5]. M. Fatima and M. Pasha, Investigations on Machine Learning Algorithms for Disease Diagnosis, Intelligent Learning Systems and Applications, al, vol. 9, no. 01, p. From January 1 to 16, 2017.

[6]. S.Pouriyeh et al., "Comprehensive review and comparison of machine learning technologies in cardiology," IEEE Symposium on Computers Proceedings and Communications (ISCC). Heraklion, Greece: IEEE, July 2017, pp. 204-207.

[7]. RDas, I. Turkoglu, A. Sengur, "Effective neurodiagnosis of heart disease, "Networks Ensembles, Expert Systems with Applications, Vol. 36, No. 4, pp.7675- 7680,2009.N.Waghulde and N.Patil,"Genetic Neural Approach esto Heart Disease Prediction," International Journal of Advanced Computer Research, vol.4,no.3,778pages,2014

[8]. S.Garciaetal.,"Pre processing of Big Data: Methods and Prospects," Big Data Analysis, vol. 1,no. 1,S. November 9, 2016.A. Janosi et al.,Cardiology Dataset,July1988 [Online].Tuganglich:http://archive.ics.uci.edu/ml/datasets/heartDisease.

[9]. K.Venbandasamy, R. Sashipriya, E. Deepa, "Detecting Heart Disease Using a Naive'Bayesian Algorithm'," International Innovative Science , Engineering and Technical Journal,Band2,Nr.9, S.441–444,2015
[10]. D Medhekar,MBote,S Deshmukh, "A system to predict heart disease using naivety." InternationalJournal of Technology and Science Enhanced Research, Bayes. MechanicalEngineering, Volume2,Issue3,p.Vom1.Bis5.January2013
[11]. A.Chenetal.,"HDPS: Herzvorhersagesysteme, Cardiology Computing, Hangzhou, China:IEEE,2011,S.557- 560.
[12]. C Dungare and S.Apte:"A Data Mining Approach "Neural Networks" for Predicting Cardiac Disease, International Journal of Computer Engineering & Technology, Bd. 3, Nr.
[13]. V. Sabarinath anand V Sugumaran, "Diagnosis of heart disease using decision trees,"International Journal of Computer Applications and Research Information Technology, Bd. 3, S. 30-40, 2012. 2, Nr. 6, S. 74-79, 2014.
[14]. J. Patel etal.Prediction of heart disease using machine learning and data miningTechnology,"Cardiology,vol.7,no.1,pp.129-137,2015
[15]. M.Shouman,T.Turner, R. Stocker, "Applying k-Nearest Neighbors in Diagnostics." CardiacPatients,"International Journal of Information Education and Technology,vol. 2, no. 3, p. 220, 2012.
[16]. W. Wiharto, H. Kusnanto,H. Herianto,"Performance Analysis of Multiclass Support:Vector Machine Classification for the Diagnosis of Coronary Artery Disease," International Journal.Computational Science and Applications, vol. 5, nine.
[17]. S. Amin, K.Agarwal, R. Beg, "Genetic Neural Network-Based Data Mining in Prediction"."Studying Cardiac Disease Using Risk Factors" (IEEE Information and Communication Conference) Technology. Indian, Thackeray, Tamil Nadu, April2013,S. 1227-1231.
[18]. N. Amma, "Cardiovascular Disease Prediction Systems Using Genetic Algorithms and Neurals"Presented "Network" at an international conference oncomputing, communications and applications.Dindiguru, Tamil Nadu, India:IEEE,February2012, pp.1-5.March 2008,p. 108～115。