

The Working of Various Prediction Techniques For Heart Diseases – A Case Study

Nishanth Vaidya

UG Student, Department of Computer Science
Sambhram Institute of Technology, Bengaluru

Mohit Kandu

UG Student, Department of Computer Science
Sambhram Institute of Technology, Bengaluru

Rahul Kumar Yadav

UG Student, Department of Computer Science
Sambhram Institute of Technology, Bengaluru

Nikhil Bharadwaj

PG Student, Department of Computer Science
Syracuse University, New York

Abstract: Heart disease or cardiovascular disease (CVD) is a kind of illness involving heart and/or blood vessels of people. Millions of scientists have been ventured down to contemplate the information with various data mining frameworks to find out the accuracy among huge information to help the medical industry. Analysis of coronary illness at the beginning is one of the more difficult parts for these expert by adjusting ongoing AI strategies which reduces the number of tests to be conducted by a significant margin . This paper gives an outline of different AI calculations associated with looking for more accurate executions in coronary illness determination, so analysts can pick the ideal system for additional therapeutic critical thinking with greater accuracy and success rates.

IndexTerms-Heart diseases, Data Mining and Machine learning, Heart Disease Prediction

1.INTRODUCTION

The heart is a vital organ of each and every living individual, which produces and sends blood to the rest of the organs through the veins . In case the progression of blood in the body is wrong, other organs suffer and if the heart stops working imminent death occurs. Life is absolutely dependent on the proper working of the heart. Coronary ailment primarily happens as a result multiple factors like heredity, different harmful habits like ingestion of alcohol and so on.

Coronary sickness is even to today the greatest factor of mortality on the planet. Few of the diseases that affect the heart are sudden and very potent, some of them being congestive heart disillumination, coronary ailment, angina pectoris, cardiomyopathy, natural coronary ailment, arrhythmias, Myocarditis and finally heart attack. There are views that some significant reasons of coronary illness National Institute of Diabetes and Digestive and Kidney Diseases evaluates that people with sort 2 diabetes — and especially the people who are at least middle aged — are at twice the risk to have coronary ailment or experience a stroke as compared people who don't have any type of diabetes.

Data mining is a data assessment approach used to perceive hidden models from colossal proportion of data. It has been successfully used in different domains for learning different outcomes. Data mining or learning disclosure has transformed into a champion amongst the most unique domains in correspondence structuring, information

hiding and biomedical industry. This data can be used for the early disclosure of the problems like coronary ailment, which can lessen the amount of heart attacks happening across the world. The therapeutic administrations associations and industry aggregate tremendous levels of data that aren't expelled to find covered learning for capable decision making processes. The key point of convergence of data recognizing is to discover structures that were not known in the past. At the point of discovery of these models, they can be used to make decisions that will prove to be greatly significant. We thusly propose to assemble relevant data related to all parts of our field of study, train the data as per the proposed count and anticipate how strong there is a credibility for a patient to get a coronary ailment via these techniques. With the ultimate objective of patients entering their data, we prescribe to make use of Data mining techniques that have been commonly used in making decision systems for illnesses conjecture through large number of datasets.

2.MACHINE LEARNING METHODS

Recognizing the problems in the early stages encourages preventive treatment, which can result in improved wellbeing and cost sparing as far as treatment goes. Different calculations are produced for anticipating the cardiovascular diseases by considering the present wellbeing of the patients utilizing AI procedure.

Categories of Machine Learning:

Machine Learning Techniques are classified into four different categories :

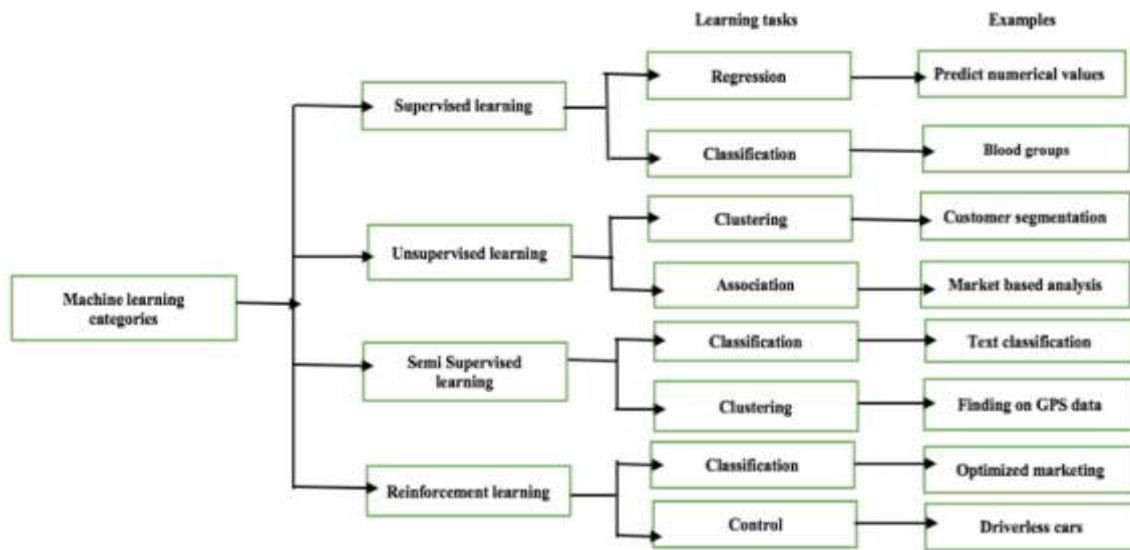


Figure 1: Machine learning categories

2.1. Supervised Learning

Supervised learning is the framework where the information and yield is acquired for future preparation. In this, there are two sorts of learning assignment being regression and classification. Some of the most common algorithms are Support Vector Machines (SVM), k-Nearest Neighbours (k-NN) and Artificial Neural Networks (ANN), Genetic algorithms and Decision Trees (DT).

2.2. Unsupervised Learning

Unsupervised learning involves taking in findings from the datasets including data without marked responses. Right now, there are two learning tasks being Association and Clustering. To find the associations of the objects in a database, Association learning was proposed by Rakesh Agarwal. The most regular count that was used in association rule is Apriori and grouping is used to assemble relative kind of data sets.

2.3. Semi-Supervised Learning

Semi-regulated learning is a mix of labelled and unlabelled data, which falls into the field of both directed as well as solo learning. This adjusting is essentially used to be arranged by site page, innate sequencing and talk affirmation. It is comprehensively characterized into two learning tasks which are known as Classification and Bunching.

2.4. Reinforcement Learning

It is affiliated to how programming specialists naturally decide the perfect conduct for particular setting, so as to amplify it's presentation. Fortification sends the reward input for the operator to get familiar with its working. It comprises of two learning errands being Classification and Control. Some of the applications are computer played board games, self-driving cars and robotic arms,. Most commonly used algorithms are Q-learning and the Temporal difference algorithm.

3.LITERATURE SURVEY

Monika Gandhi et.al,[1] made use of Naive Bayes, Decision tree and neural network algorithms. There is a huge number of features that are involved. So, there is a need to reduce the number of features .this can be done by feature selection

Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte [3] made use of Data mining techniques for appropriate results. They use Neural network to predict the different Heart Diseases. For prediction, the system uses sex, blood pressure like 13 different medical parameters.

Chaitrali S. Dangare, Sulabha S. Apte[4] Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques. They made use of Naive Bayes, Decision tree and Neural Networks and they concluded that Neural Network provides accurate results as compared to Decision tree & Naive Bayes.

M.Marimuthu, S. Deivarani, Gayathri.R[5] Analysis of Heart Disease Prediction Using Various Machine Learning Techniques like k-nearest neighbor, Naive Bayes, SVM and Decision trees are used. They calculated the accuracy of each algorithm and concluded Support Vector Machine as the most accurate.

Rajesh N, T Maneesha, Shaika Hafeez, Hari Krishna[6] states that Naïve Bayes are more accurate when compared to Decision Tree Algorithm

4. MACHINE LEARNING BASED METHODS FOR HEART DISEASE PREDICTION

HD forecast takes less time and puts forth the methodology that is faster with the planned attempt of Machine learning and data Mining. ML is essentially used for data examination's explanations behind accurate learning with minimal misstep rate [10]. Artificial intelligence frameworks improves precision in desire for HD at the outset time of diseases and patients can direct remedial specialists for preventive treatment [11]. Movement in advancement of bio informatics gives a basic advantage for organizing the datasets [14]. The Cleveland database is the only one in coronary ailment that has been used by researchers till date [10].

Table 1. Attributes of the Heart disease dataset

Attribute	Representation	Information Attribute	Description
Age	Age	Integer	Age in years (29 to 77)
Sex	Sex	Integer	Gender instance (0 = Female, 1 = Male)
ChestPainType	Cp	Integer	Chest pain type (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)
RestBloodPressure	Trestbps	Integer	Resting blood pressure in mm Hg[94, 200]
SerumCholesterol	Chol	Integer	Serum cholesterol in mg/dl[126, 564]
FastingBloodSugar	Fbs	Integer	Fasting blood sugar > 120 mg/dl (0 = False, 1= True)
ResElectrocardiographic	Restecg	Integer	Resting ECG results (0: normal, 1: ST-T wave abnormality, 2: LV hypertrophy)
MaxHeartRate	Thalach	Integer	Maximum heart rate achieved[71, 202]
ExerciseInduced	Exang	Integer	Exercise induced angina (0: No, 1: Yes)
Oldpeak	Oldpeak	Real	ST depression induced by exercise relative to rest[0.0, 62.0]
Slope	Slope	Integer	Slope of the peak exercise ST segment (1: up-sloping, 2: flat, 3: down-sloping)
MajorVessels	Ca	Integer	Number of major vessels coloured by fluoroscopy (values 0 - 3)
Thal	Thal	Integer	Defect types: value 3: normal, 6: fixed defect, 7: irreversible defect
Class	Class	Integer	Diagnosis of heart disease (1: Unhealthy, 2: Healthy)

Table our system will be implementing the following four algorithms:

- Support Vector Machine (SVM)
- Random Forest
- Naïve Bayes Algorithm
- KNN

The calculations will be prepared utilizing the informational index acquired via UCI. 75% of the passages in the informational index will be utilized for preparing the training data and the remaining 25% for testing the precision of the calculation. Moreover, a few of the stages will be taken for enhancing the calculations which in turn improves the precision. These different means incorporate cleaning the dataset and information pre-preparing. AI calculations are explored for evaluating and foreseeing the impact of heart disease by Naïve Bayes Algorithm, Support vector machine (SVM), arrangement and regression tree. At last, out of these calculations the creators guaranteed that SVM's performance is superior to the next three techniques. In the coronary illness demonstration critical endeavours are made by creators, which concludes with the best methodology being SVM which gives an accuracy of 94.60% and furthermore SVM approach is increasingly precise and less blunders in disease prediction [18] and [19].

In HD Prediction 302 events were compared [21] investigated using four AI algorithm that were Naïve Bayes, K-Nearest Neighbour, Random forest and SVM. The user has investigated various roads with respect to the models and came to the conclusion that SVM method performed better[20]. SVM techniques moreover used for diabetic patients in HD examination

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
K-NN	0.753	0.258	0.785	0.753	0.769	Absence
	0.742	0.247	0.706	0.742	0.724	Presence
SVM	0.867	0.2	0.844	0.867	0.855	Absence
	0.8	0.133	0.828	0.8	0.814	Presence
RF	0.847	0.225	0.825	0.847	0.836	Absence
	0.775	0.153	0.802	0.775	0.788	Presence
NB	0.867	0.2	0.844	0.867	0.855	Absence
	0.8	0.133	0.828	0.8	0.814	Presence

Table 2:Accuracy / Accuracy measured by class

5.CONCLUSIONS

Heart ailments are lethal and remove countless lives each year .When the early side effects of heart ailments are disregarded, the patient may wind up with fatal outcomes. Stationary way of life and unreasonable worry in this day and age has intensified the circumstance for this to happen. In the event that the illness is recognized early, at that point it may be monitored. AI Algorithms are widely utilized in therapeutic line so illness location can be made simple, so that with legitimate treatment the enduring patient can be cured. Data mining can be of generally excellent assistance in choosing the line of treatment to be trailed by retrieving information from such reasonable databases. Coronary illness forecast framework is proposed to recognize the danger of coronary illness accurately.

6.REFERENCES

- 1.SunilRay.(2017)<https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machineExample-Code/>.
2. Agrawal R, Imielinski T, Swami A. (1993) ‘Mining Association Rules Between Sets of Items In Large Databases’,Proceedings of The ACM SIGMOD Conference on Management of Data, P. 207–16.
3. Agrawal R, Srikant R. (1994) ‘Fast Algorithms for Mining Association Rules in Large Databases’,Proceedings of the 20th International Conference on Very Large Databases,P. 478–99.
4. Kavakiotis I, Tzani G, Vlahavas I. (2014), ‘Mining Frequent Patterns and Association Rules from Biological Data’,In: Elloumi M, Zomaya AY, Editors. Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing Of Biological Data. Wiley Book Series on Bioinformatics: Computational Techniques and Engineering new Jersey, USA: Wiley-Blackwell, John Wiley & Sons Ltd.
- 5.NikkiCastle.(2017)<https://www.datascience.com/blog/supervised-and-unsupervised-machine-learningalgorithms>
6. Neha Chauhan, Nisha Gautam. (2016) ‘Heart Disease and Data Mining: Collaboration in The Making’, International Journal of Latest Trends in Engineering and Technology (IJLTET). Volume 7,Issue 2,Pages:49-55.

7. Apurva Joshi, Er. Jitendra Dangra and Dr. M. K. Rawat.(2016) _A Critical Review of Various Techniques for Heart Disease Prediction', International Journal of Technology Research and Management, ISSN (Online): 23489006 Vol 3 Issue 9 September 2016 .
8. Rahul C. Deo. (2015), Machine Learning in Medicine,Boi science for clinicians, <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>.
9. Dr.R. Obulakonda Reddy, R. Nagarjuna Reddy, M. Radha, Sree Vani. (2017) _A Review of Machine Learning Approaches in Data Sensitive Real-World Applications',Journal of advanced research in dynamical and control systems, Volume: 9 | Issue: 3,Pages: 165-171.
10. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee. (2017) _Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review', Advances in Computational Sciences and Technology,ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159.
11. Tanvi Sharma, Sahil Verma, Kavita. (2017) _Intelligent Heart Disease Prediction System Using Machine Learning: A Review', International Journal of Recent Research Aspects, ISSN: 2349-7688, Vol. 4, Issue 2, pp. 94-97.
12. Sonam Nikhar, A.M. Karandikar.(2016) _Prediction of Heart Disease Using Machine Learning Algorithms', Interantional journal of advanced engineering ,management and science,Vol-2, Issue-6.
International Journal of Pure and Applied Mathematics Special Issue
1492
13. Himanshu Sharma, M A Rizvi. (2017) _Prediction of Heart Disease Using Machine Learning Algorithms: A Survey', International Journal on Recent and Innovation Trends in Computing and Communication, ISSN:23218169, Volume:5, Issue:8, Pages 99-104.
14. Almas Jabeen, Nadeem Ahmad, Khalid Raza. (2017) _Machine Learning-Based State-Of-The-Art Methods for The Classification Of RNA-Seq Data', doi: <https://doi.org/10.1101/120592>.
15. Evanthia E. Tripoliti, Theofilos G. Papadopoulos, Georgia S. Karanasiou, K. Naka, Dimitrios I. Fotiadis. (2017) _Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques', Computational and Structural Biotechnology Journal, Volume 15, Pages 26-47.
16. Meherwar Fatima,Maruf Pasha.(2017) _Survey of Machine Learning Algorithms for Disease Diagnostic', Journal ofintelligentlearningsystemsandapplications,Vol.09.pages 10.4236/jilsa.2017.91001.
17. Akhila C S,Vidya M.(2017) _A Survey on Machine Learning Approaches for Disease Predicting System',Journal of Emerging Technologies and Innovative Research,Volume 4,Issue 6,pages 203-205.
18. Gagan Kumar,Rohit Kalra.(2016)_A Survey on Machine Learning Techniques in Health Care Industry', International Journal of Recent Research Aspects, ISSN: 2349-7688 Vol. 3 Issue ,pp. 128-132.
19. Seyedamin Pouriye, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez. (2017) _A Comprehensive Investigation and Comparison of Machine Learning Techniques in The Domain of Heart Disease', Published in: Computers and Communications (ISCC), 2017 IEEE Symposium on 3-6 July 2017, DOI: 10.1109/ISCC.2017.8024530.

- 20.G.Parthiban,S.K.Srivatsa.(2012) ‘Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients’, International Journal of Applied Information Systems (IJ AIS), ISSN : 2249-0868 , Volume 3–No.7.
- 21.Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization Youness Khourdifi1* Mohamed Bahaj1 International Journal of Intelligent Engineering and Systems, Vol.12, No.1, 2019 DOI: 10.22266/ijies2019.0228.24
22. Domingos P and Pazzani M. “Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier”, in Proceedings of the 13th Conference on Machine Learning, Bari, Italy, pp105-112, 1996.
23. Elkan C. “Naive Bayesian Learning, Technical Report CS97-557”, Department of Computer Science and Engineering, University of California, San Diego, USA, 1997.
- 24.B.L Deekshatulua Priti Chandra “Reader, PG Dept. Of Computer Application North Orissa University, Baripada, Odisha – India. “Empirical Evaluation of Classifiers’ Performance Using Data Mining Algorithm” International Journal of Computer Trends and Technology (IJCTT) – Volume 21 Number 3 – Mar 2015 ISSN: 2231-2803 <http://www.ijcttjournal.org> Page 146
- 25.D.S. Medhekar, M.P. Bote, and S. D. Deshmukh, “Heart Disease Prediction System using Naive Bayes,” International Journal of Enhanced Research in Science Technology and Engineering, vol. 2, no. 3, Elsevier 2013
- 26 .D. K. Srivastava and L. Bhambhu, “Data classification using support vector machine,” J. Theor. Appl. Inf. Technol., 2009
27. Suykens, Johan AK, and JoosVandewalle. "Least squares support vector machine classifiers." Neural processing letters vol. 9, 3, 1999, pp. 293-300
28. Witten I and Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java, Morgan Kauffman Publishers, California, USA, 1999
- 29.P. V. Ankur Makwana, “Identify the patients at high risk of re-admissionin hospital in the next year,” International Journal of Science andResearch, vol. 4, pp. 2431–2434, 2015
- 30.Y. E. Shao, C.-D. Hou, and C.-C. Chiu, “Hybrid intelligent modelling, schemes for heart disease classification,” Applied Soft Computing,vol. 14, pp. 47–52, 2014.
31. Bhatia and C. Author, “Survey of Nearest Neighbor Techniques,” IJCSIS) Int. J. Comput. Sci. Inf. Secur., vol. 8, no. 2, pp. 302–305, 2010
- 32.Prediction of Heart Disease using Machine Learning Algorithms: A Survey Himanshu Sharma, M A RizviIJRITCC | August 2017, Available @ <http://www.ijritcc.org>
- 33.Intelligent Heart Disease Prediction System using Machine Learning: A Review Tanvi Sharma, Sahil Verma, Kavita© 2017 IJRRRA All Rights Reserved page-9
- 34.Machine Learning and Deep Learning Methods in Heart Disease (HD) Research Kusuma.S1, Divya Udayan.J2International Journal of Pure