Heart Disease Prediction Analysis Using Various Machine Learning Techniques - A Survey

By Nishanth Vaidya, Mohit Kandu, and Dr. T. Srinivasa Rao.
Department of Computer Science and Engineering(CSE),
Sambhram Institute of Technology, Bangalore, India.

Abstract:The human services industry is found with gigantic measure of rich data yet absence of information to deal with the huge information so as to get the satisfactory outcomes. As of late loads of scientists have been ventured down to contemplate the information with various data mining frameworks to find out the realities among huge information to give the ongoing issue in medicinal services industry. Analyze of coronary illness in beginning times is one of the difficulties for these expert by adjusting ongoing AI strategies which diminishes the quantity of test to be occur and furthermore decrease the quantity of test to be occur and furthermore lessen the quantity of death from the coronary illness. Information mining gives the ordinary techniques to address these issues and sort out with the productive system to foresee the coronary illness at beginning periods utilizing different AI predication calculation. This paper outline different AI calculation associated with looking for better execution in coronary illness determination, so analysts can pick the ideal system for additional therapeutic critical thinking with high exactness.

***IndexTerms*-Hear disease, Data Mining, Machine learning, Prediction.**

I. INTRODUCTION

The heart is a huge organ of each and every living individual, which expect an essential activity of blood directing to the rest of the organs through the veins of the circulatory structure. In case the progression of blood in the body is wrong the organs like personality suffer and if the heart stops working absolute and passing occurs. Life is absolutely dependent on the right working of the heart. Coronary ailment primarily happens as a result of some life penchants, and wrong sustenance affinities and progressively coronary ailment leaps out at male

Coronary sickness is the fundamental wellspring of reliably passing on the planet. A couple of sorts of illness happens ambush for heart. Sorts of infirmity considered are coronary ailment, angina pectoris, congestive heart disillusionment, cardiomypathy, natural coronary ailment, arrhythmias, Myocarditis, heart attack; heart dangerous development, etc right now particularly unequivocal very hazards contamination to cardiovascular affliction or coronary illness.There are viewed as 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 have accomplished middle age — are twice as at risk to have coronary ailment or experience a stroke as people who don't have diabetes. Adults with diabetes will when all is said in done have heart attacks at a progressively energetic age. Will undoubtedly experience different heart attacks if they have insulin block or high blood glucose levels. The clarification behind this is the association among glucose and vein prosperity. High blood glucose levels that aren't managed can extend the proportion of plaque that structures inside the dividers of the veins. This squares or stops the movement of blood to the heart. If you have diabetes, you can decrease the peril of coronary ailment by managing your glucose carefully. Seek after a diabetes-obliging eating routine that is affluent in fibber and low in sugar, fat, and fundamental starches.

Managing your glucose levels can in like manner help balance cut down your danger for eye disorder and scattering issues. You should moreover keep up a sound weight. Additionally, in case you smoke, by and by's a good time to think about halting. There are various purposes behind coronary sickness, for instance, stress, hypertension, sedate abuse, nonappearance of action, sustenance affinity, cholesterol, etc. Our vein ends up fragile in view of oily sustenance which can provoke coronary ailment. The dividers in heart become thicker when more weight is associated with our stockpile courses. As the dividers become thick, it can ruin the movement of blood and can in like manner make the square which lead to coronary disease as per the Indian Heart Association, half of heart strokes occur under 50 years of age and 25% of all heart strokes occur under 40 years of age in Indians. Urban masses is thrice as feeble against heart ambushes as rural people due to sustenance inclinations and air tainting.

Data mining is data assessment approach used to perceive hid models from colossal proportion of data. It has been successfully used in different domains for learning exposure. Data mining or learning disclosure has transformed out into see, as a champion among the most unique domains in correspondence structuring, information headways and biomedical. This data can be used for the early disclosure of the coronary ailment, which can support to lessen the amount of heart attacks. The therapeutic administrations associations and industry aggregate tremendous level of data that are not expelled to find covered learning for capable decision. The key point of convergence of data recognizing evidence is to discover structures that were in the past not known. At the point when these models are found they can be used to choose incredible decisions We thusly propose to assemble relevant data relates all parts related to our field of study, train the data as per the proposed count of AI and anticipate how strong is there a credibility for a patient to get a coronary ailment. With the ultimate objective of patients entering data, we prescribe to make use Data mining techniques have been commonly used in making decision sincerely steady systems for illnesses conjecture through a ton of therapeutic datasets.

II. MACHINE LEARNING METHODS

Recognizing the maladies prior encourages preventive human services mediations, which can prompt improved wellbeing results and cost sparing as far as treatment. Different calculations are produced for anticipating the cardio illness event by considering the present wellbeing information of the patients utilizing AI procedure.

Categories of Machine Learning:

Machine Learning Techniques are divided into four different categories [1]

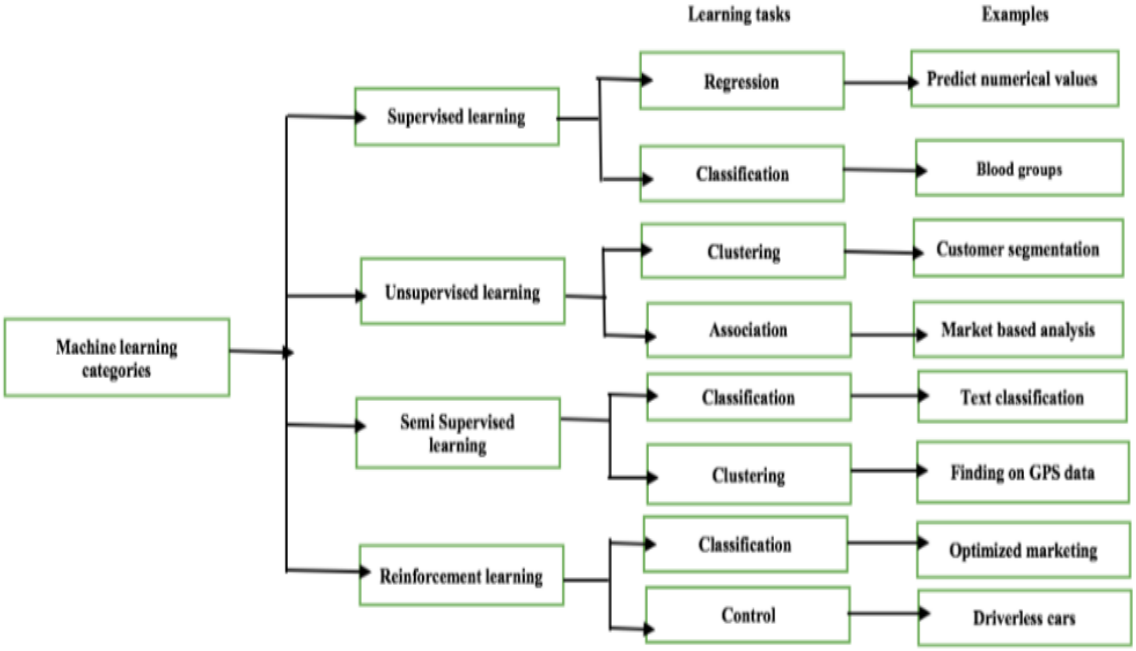

Figure 1: Machine learning categories [1]

2.1. Supervised Learning

Supervised learning is the framework where both the information and wanted the yield are accommodated future information preparing. In this, there are two sorts of learning assignments: Regression[1] and Classification[1]. Some of the most common algorithms are Support Vector Machines (SVM), Genetic algorithms, Decision Trees (DT), k-Nearest Neighbors (k-NN) and Artificial Neural Networks (ANN).

2.2. Unsupervised Learning

UnSupervised learning taking in draw findings from datasets including data without marked responses. Right now, two learning tasks: Association and Clustering. To find the associations in the objects of a database, Association learning was proposed by Rakesh Agarwal [2]. The most regular count used in association rule is Apriori and grouping is to assemble relative kind of datasets [3] and [4]. Without a doubt the most ordinary computations are k-implies bundling and association rule learning figuring..

2.3. Semi-Supervised Learning

Semi-regulated learning is a mix of named and unlabelled data, which falls in directed and solo learning. This adjusting essentially used arranged by site page, innate sequencing and talk affirmation. It is comprehensively characterized into two learning tasks:classification and bunching

2.4. Reinforcement Learning

It is worried about how programming specialists naturally decide the perfect conduct inside a particular setting, so as to amplify its presentation. Fortification sign sends the reward input for the operator to get familiar with its conduct. It comprises of two learning errands:classification and control. Some of the applications are computer played board games, robotic hands, and self-driving cars. Most commonly used algorithms are Q-learning, Temporal difference, Deep Adversarial.

III. MACHINE LEARNING BASED METHODS FOR HEART DISEASE PREDICTION

HD forecast takes less time and puts forth methodology fast with the planned attempt of Machine learning [Ml][34] and data Mining [6] [7] [8] and [9]. Ml is essentially used for data examination explanation behind accurate learning with less misstep rate [10]. Artificial intelligence frameworks improves precision in desire for HD at the outset time of disease and patients can direct remedial specialists for preventive treatment [11] [12] and [13]. Movement in advancement of bioinformatics gives a basic advantage for organizing datasets [14]. The Cleveland database is the only one in coronary ailment that has been used by

researchers till date [15]. Table 1 delineates the 14 characteristics of the dataset.

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]
SerumCholestoral	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 from UCI. 75% of the passages in the informational index will be utilized for preparing and the staying 25% for testing the precision of the calculation. Moreover, a few stages will be taken for enhancing the calculations in this way improving the precision. These means incorporate cleaning the dataset and information pre-preparing. AI calculations are explored for evaluating and foreseeing the seriousness of heart disappointment by Naïve Bayes Algorithm, Support vector machine (SVM), arrangement and regression tree. At long last, out of these calculations the creators guaranteed SVM performs superior to the next three techniques [16]. In the coronary illness demonstrative critical endeavors are made by creators, best methodology is SVM which gives an exactness of 94.60% [17] and furthermore SVM approach is increasingly precise and less blunders in disease prediction [18] and [19].

In HD Prediction 302 events were compared[21] table and investigated using four AI algorithm , for instance, Naïve Bayes, K-Nearest Neighbour, Random forest and SVM. Maker has done investigate various roads with respect to the models and came about SVM method performed well [20]. SVM techniques moreover used for diabetic patients in HD examination

Table 2:Accuracy / Accuracy measured by class

	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

1. NAIVE BAYES NB

This classifier is an amazing probabilistic depiction, and its usage for portrayal has gotten great thought. This classifier gains from getting ready data the prohibitive probability of every trademark Ai given the class mark C. Gathering is then done by applying Bayes standard to process the probability of C given the particular instances of A1… ..An and after that envisioning the class with the most raised back probability. The target of collection is to successfully envision the estimation of an allocated discrete class variable given a vector of pointers or qualities. In particular, the Naive Bayes classifier is a Bayesian framework where the class has no watchmen and every quality has the class as its sole parent. Regardless of the way that the naïve Bayesian (NB) count is clear, it is feasible in various certifiable world datasets considering the way that it can give favored perceptive accuracy over most likely saw without a doubt comprehended systems like C4.5 and BP [22],[23] and is incredibly compelling in that it learns in an immediate way using social affair instruments, for instance, stowing and boosting, to join classifier desires [24]. Regardless, when qualities are tedious and not normally passed on, the prescient exactness is diminished Hence each element is thought to be free

and self-governing contributing separately to the preparation information guide's likelihood of having a place toward a specific class.As per the Bayes theorem [25][33],

$P(c|x) = P(x|c)P(c)\ P(x)$

$(c|X) = (x1|c)(x2|c) \times \ldots \times P(xn|c)P(c)$

Here $(c|x)$is the posterior probability of class given predictor $P(c)$is the prior probability of class $P(x|c)$is the likelihood probability of predictor given class $P(X)$is the prior probability of predictor

2. SUPPORT VECTOR MACHINE SVM

SVM performs characterization by finding the hyper plane that augment the edge between two classes. The vectors that characterize the hyper plane are the help vectors [26][32]

Steps for Calculation of Hyper plane

1. **Set up training data**
2. **Set up SVM parameter**
3. **Train the SVM**
4. **Region classified by the SVM**
5. **Support vector**

Utilization of the SVM for informational collection grouping has its very own preferences and drawbacks. Restorative informational index can be nonlinear of high dimensionality by watching properties. Unmistakably SVM would be one of the most loved decisions for arrangement. A portion of the bit of leeway to choose the SVM for arrangement decision.

1. Right off the bat regularization parameters which maintain a strategic distance from issue of over fitting which one of the significant difficulties is in choice tree.
2. Portion tree is utilized to keep away from the master information through the learning of kern.
3. SVM is a productive strategy since it use curved advancement issue (COP) which mean it has doesn't neighborhood minima.
4. Mistake appraised is tried which give a more prominent help after misclassification of dataset.

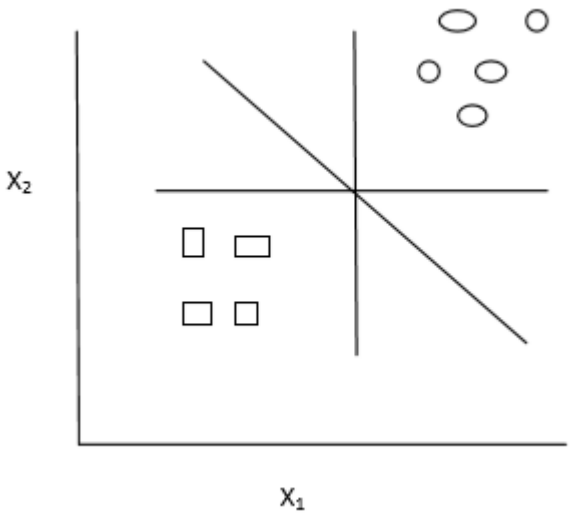

Figure 2 SVM Classifier[26]

Support vectors really allude to the information indicates that are nearest the outlining surface which are subsequently dubious to classify. The metric that implies the optimality of a hyper plane is the edge around the hyper plane. So the issue advances into that of a streamlining one. As set up the most extreme edge classifier took in and got from the preparation information would lead us to ideal hyper plane. This is accomplished by changing the maximal edge classifier as the inward item (aggregate of increase of pair esteems) of two given information focuses as opposed to the information focuses. The general part capacity could then be characterized as pursues [27][33]:

$$(x) = B0 + \Sigma\ (ai \times (x,))\ i$$

Here x is the new information vector and the coefficients $B0$ and ai must be evaluated from preparing information.

All the above highlights could be helpful for restorative analyze dataset which is bringing about structure increasingly productive predication framework for the specialist Support vector machines exist in various structures, straight and non-direct. A help vector machine is a regulated classifier. What is regular in this specific situation, two distinctive datasets are included with SVM, preparing and a test set. In the perfect circumstance the classes are straightly divisible. In such circumstance a line can be discovered, which parts the two classes consummately. Anyway one line parts the dataset flawlessly, yet an entire bundle of lines

do. From these lines the best is chosen as the "isolating line". The best line is found by amplifying the separation to the closest purposes of the two classes in the preparation set. The amplification of this separation can be changed over to a proportionate minimization issue, which is simpler to comprehend. The information focuses on the maximal edge lines are known as the help vectors. Regularly datasets are not pleasantly dispersed with the end goal that the classes can be isolated by a line or higher request work. Genuine datasets contain arbitrary blunders or clamor which makes a less perfect dataset. Despite the fact that it is conceivable to make a model that superbly isolates the information, it isn't attractive, in light of the fact that such models are over-fitting on the preparation information. Over fitting is brought about by fusing the irregular mistakes or commotion in the model. Thusly the model isn't nonexclusive, and makes essentially more blunders on different datasets. Making less difficult models keeps the model from over-fitting. The unpredictability of the model must be adjusted between fitting on the preparation information and being nonexclusive. This can be accomplished by permitting models which can make blunders. A SVM can make a few blunders to maintain a strategic distance from over-fitting. It attempts to limit the quantity of blunders that will be made. Bolster vector machines classifiers are connected in numerous applications. They are exceptionally well known in late research. This notoriety is because of the great generally exact execution. Contrasting the credulous Bayes and the SVM classifier, the SVM has been connected the most [28].

3. RANDOM FOREST

Random Forest Algorithm: Random forest is an ensemble classifier that consists of many decision trees. The output of the classes is represented by individual trees. It is derived from random decision a forest that was proposed by Tin Kam Ho of Bell Labs in 1995 [30]. This method combines with random selection of features to construct a decision trees with controlled variations. The tree is constructed using algorithm as discussed.

Let N be the number of training classes and M be the number of variables in classifier

- The input variable m is used to determine the node of the tree. Note that m<M.
- Choosing n times of training sets with the replacement of all available training cases N by predicting the classes, estimate the error of the tree.
- Choose m variable randomly for each node of the tree and calculate the best split
- At last the tree is fully grown and it is not pruned. The tree is pushed down for predicting a new sample. When the terminal node ends up, the label is assigned the training sample. This procedure is iterated over all trees and it is reported as random forest prediction.

Multi-classifiers are the after effect of joining a few individual classifiers. Troupes of classifiers towards expanding the execution have been presented. [29].

Random Forest (RF) is one of the instance of such strategies. RF as a multi classifier framed by decision trees where each tree ht had been made from the arrangement of data getting ready and a vector t of subjective numbers indistinctly spread and liberated from the vectors. Vectors 1, 2... t-1 used to make the classifiers h1; h2; ::; ht-1 . Each choice tree is made from arbitrary subset of the planning dataset. It used an irregular vector that is delivered from some modified probability dispersal, where the probability flow is moved to focus tests that are hard to orchestrate. A Random vector can be joined into the tree-turning out to be process from different points of view. The leaf centres of every one tree are named by assessments of the back spread over the data class names. Each inside centre contains a test that best parts the space of information to be orchestrated. Another, disguised event is requested by sending it down each tree and conglomerating the landed at leaf allocations.

4. K- NEAREST NEIGHBOUR ALGORITHM (KNN)

KNN is slow supervised learning algorithm, it take more time to get trained classification like other algorithm is divided into two step training from data and testing it on new instance . The K Nearest Neighbour working principle is based on assignment of weight to the each data point which is called as neighbour. in K Nearest Neighbour distance is calculate for training dataset for each of the K Nearest data points now classification is done on basis of majority of votes there are three types of distances need to be measured in KNN Euclidian, Manhattan, Minkowski distance in which Euclidian will be consider most one the following formula is used to calculate their distance [31][32]:
$EucledianDistance = Dx, = (xi - yi)\ 2ki = 1$ (1)

K=number of cluster

x, y=co-ordinate sample spaces $Manhattandistance = (xi - yi)ni=1$ (2)
x&y are co-ordinates
Minkowski distances are generally Euclidian distance

$Min = (-yip)\ 1$(3)
Grouping of sample is based on super class in the KNN reduction of sample is the result of proper grouping which is used for further training. Selection of k value plays a pivotal role, if the k value is large then it precise and less noisy
The algorithm for KNN is defined in the steps given below
1. D represents the samples used in the training and k denotes the number of nearest neighbour
2. Create super class for each sample class.
3. Compute Euclidian distance for every training sample 4. Based on majority of class in neighbour, classify the sample

IV ANALYSIS OF AVAILABLE LEARNING ALGORITHM

When it comes to comparing two or more machine learning algorithm, it is most difficult because two algorithms is differ in many ways. Reason for difficulty in comparison because algorithm are highly depended on dataset , it is not wise to decide properly which algorithm is perform for the particular dataset , there is only one way to know about the efficiency of algorithm for the particular dataset is implement them. Analytical comparison is require to properly decide the difference between different machine learning algorithm this type of work could be useful for researchers who want to work in this fieldComparison will highlight the key difference on different background this paper has tried to reflect majority of comparison between different algorithms so that beginner and new .

Table 3. Compares major Machine learning Algorithm based on different parameter [32]

Techniques	Outlier	Online Learning	Over fitting and under fitting	Parametric	Accuracy
SVM	It can handle outlier properly	Online training require less time than ANN	Perform better than over fitting and under fitting	Non parametric model	Higher than other parametric model
Naïve Bays	It is less pruned to outline	It can perform on online testing	It does not suffer over fitting and under fitting	It is parametric	High with limited dataset
Random Forest	Outlier does not play critical role in interoperation of dataset by decision tree	It does not supported online learning	It suffer over fitting and under fitting	Non parametric model	Accuracy depend on the dataset, ensemble technique used decision tree have higher accuracy than SVM
KNN	It is pruned to outlier	Online learning g can take in KNN but more time than SVM	It is more prune d to over fitting than SVM	It is parametric	Higher than all other paramet ric model

IV CONCLUSION

Heart Disease when bothered winding path out of hand. Heart ailments are confused and remove heaps of lives each year .When the early side effects of heart ailments are disregarded, the patient may wind up with intense outcomes in a limited capacity to focus time. Stationary way of life and unreasonable worry in this day and age has intensified the circumstance. In the event that the illness is recognized early, at that point it very well may be monitored. AI Algorithms are widely utilized in therapeutic line so illness location can be made simple and at beginning times, 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 extricating information from such reasonable databases. Coronary illness forecast framework is proposed to recognize the danger of coronary illness precisely.

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, Tzanis 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 Pouriyeh, 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 (IJAIS), 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 C omputer 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 IJRRA 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