

Analysis and Detection of Heart Related Issues: A Proposed Algorithm

Kumari Nirmala¹, R.M.Singh²

¹Computer Science and Engineering, Amity University
Noida, Uttarpradesh

²Defence Research and Development Organisation
Metcalfe House, New Delhi

Abstract— The change in lifestyle gives invitation to many diseases. Heart disease or cardio vascular diseases, as reported by the WHO are the main cause of death, especially in Asia subcontinent. As there are many symptoms which can lead to heart disease so it becomes the primary requirement to reduce the data set so that it becomes quite easy task for doctors to clinically treat the disease with filtered attributes or values that are responsible of heart disease. Hence, this paper provided a comprehensive study of various kinds of heart diseases and various reasons that are responsible for heart disease. An algorithm based on the patterns has been proposed in this paper in which we have combined MTD(Mega Trend Diffusion Function) that have the capability of classifying the attributes to the maximum extent how much they are overlapping. If the two values are overlapping, then MTD function has the capability to classify it to the maximum extent by using synthetic operators, and hence can classify the data or attributes into two classes. After using MTD function we can apply any classification algorithm like PCA(Principal Component Analysis) which can take reduced the dataset as an input and thus provide better results in classifying the disease or the number of symptoms, which are more relevant in giving decision to the prediction of heart diseases. The mathematical implementation is also included in this review paper

Keywords— Pattern recognition, MTD, machine learning, PCA, WHO

I. INTRODUCTION

Pattern recognition, a new emerging branch in today's scenario is a special branch of artificial intelligence. Different researchers gave many definitions of Pattern recognition, however the dictionary meaning of PATTERN RECOGNITION \Rightarrow Pattern + Recognition

There are a diversity of applications of pattern recognition that are as follows: Diagnosis of various medical diseases like Heart attack or Lung Cancer, Satellite image capturing based on some particular patterns, Image processing, Detection of Radar, Process control, Weather prediction, Sensing of life on remote planets, Behaviour analysis, Management systems of information, Character recognition and there are many more.

From these above applications, I've chosen the broader area of medical diagnosis; this is because the pattern of human beings has no changed since many years as compared to patterns of machines and computer which change continuously as the new technology land in market. Heart Attack diseases that are basically of many types are the

leading cause of death throughout the world according to world health organization (WHO).

1.1 Symptoms of Heart Diseases

After a long research, scientists have learned that Pain in the chest is not always a symptom of heart attack and also the symptoms are not crystal clear. Any sort of heart disease depends upon various parameters like how old are you, what is your cholesterol level, whether you are a woman, man or child , etc. likewise there are many other factors.

Following are the general symptoms that are collected from various sites, these are as follows:

1. Chest pressure
2. Shortness of breath
3. Pain in the chest and further spreading to the neck, arms, jaws and back.
4. Severe anxiety or confusion
5. Gastric or indigestion problems particularly when it doesn't respond to antacids
6. High cholesterol which leads to the blood clotting in the blood vessels and which restricts to the passage of blood to and from the heart.
7. Smoking is another symptom of heart disease, in which the nicotine can deprive the oxygen level in blood vessels
8. High Blood pressure can leads to the damage of arteries which could further stop the passage of blood supply to the body.
9. Nausea or fainting
10. Swelling of eyes or face

These symptoms doesn't at all assures that a heart attack is in progress, but the more symptoms one have, the more chances it is a heart attack; and more oneself educate about these symptoms, the more we can avoid fatal episodes of heart attack, and other heart related diseases.

II. RELATED WORK

A method has been proposed, namely "Heart disease diagnosis using Neural Network" [1] focusing on the work of heart disease classification using neural network with back propagation network .Another technique which is used for heart prediction was presented[2], considering risk factors using two data mining tools i.e. neural network and genetic algorithms this they introduced a hybrid system and compared the results to back propagation method, and the results were found that the risk of disease was obtained with an accuracy of 89%. [2] . An intelligent heart disease Prediction System was designed by [3] ,employing a

method of weighted associative classifiers using Java (front end) and Ms access(backend) as implementation tool, classified whether a patient has heart disease or not. One of the most famous and successful hybrid approaches “Neuro-fuzzy approach” was proposed by J.S.R. Jang, which is the hybridization of two approaches. This approach combined the features of fuzzy system that have the capability of human-like thinking [4] and neural network carrying learning capability.

III. PROPOSED WORK

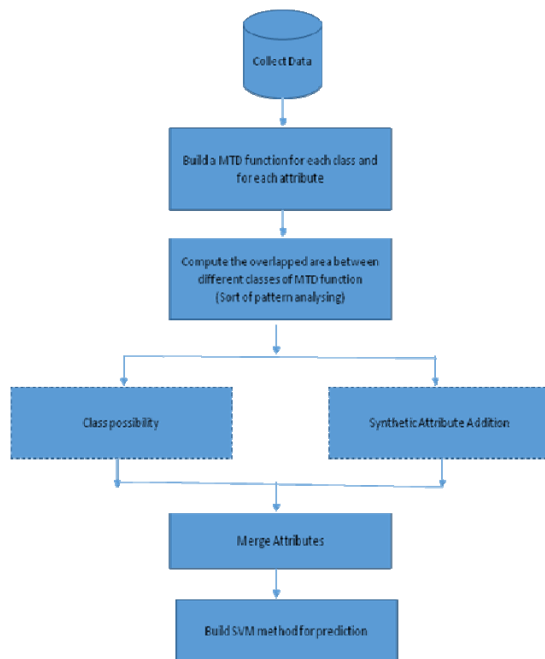


Fig. 1 Algorithm model

Proposed Algorithm

Step 1: Collect the set of inputs X and divide the same in k classes where each x has M attributes and a have a specific target value ‘t’.

Step 2:- Find the decision boundaries of each class and of each attribute

The formula that we are going to use is

$$a = U_{Set} - skew_L X \sqrt{(-2) * S_L^2 N_L * \ln(f(t))}$$

// Setting boundary values a and b

$$b = U_{Set} + skew_U X \sqrt{(-2) * S_U^2 N_U * \ln(f(t))}$$

in and max denote the minimum and maximum values of X, respectively.

Uset= (min+max)/2 and S_L^2 is the variance

NL is the number of data points smaller than Uset.

NU is the number of data points greater than Uset.

skew_L=NL/(NL+NU) and skew_U=NU/(NL+NU)

f(t) show the rates of skewness in the distribution of the data.

Step 3: Perform transformation (MTD) of each attribute as M¹(X) for attribute 1 and M²(X) for

Attribute 2 for each class

Step 4: Compute the overlap area of MTD function by using Geometric mean and find the

Optimal threshold say δ

- i. δ¹ < mean (δ), low overlap
- ii. δⁱ > mean (δ), low overlap i=1,2,.....M.

Step 5: Attributes construction

5. a) Computing Correlation coefficient matrix Σ for data set X

Step 6: for each correlation coefficient value greater than 0.4 ,we have to find synthetic values by applying synthetic operators i.e Correlation attributes of (yi,yj)>0.4 then the operation followed for synthetic operators are :

- i. yi*yj
- ii. yi/yj
- iii. yj/yi

Step 7: the above steps will generate new data sets

Step 8: Use any classification machine learning algorithm like PCA,LDA etc.

Mathematical Explanation of proposed algorithm:

1. Applying MTD(Mega trend diffusion function) on the given set

1. a) given data

Table 1. Assumed Data Set

Data	Attribute 1	Attribute 2	CLASS
X1	0.10	0.50	A
X2	0.60	0.40	A
X3	0.85	0.12	B
X4	.35	0.50	B
X5	0.30	0.80	A

2. Divide the given data set in k classes and set the decision boundary ‘a’ and ‘b’ by the formula

The formula that we are going to use is

$$a = U_{Set} - skew_L X \sqrt{(-2) * S_L^2 N_L * \ln(f(t))}$$

// Setting boundary values a and b

$$b = U_{Set} + skew_U X \sqrt{(-2) * S_U^2 N_U * \ln(f(t))}$$

in and max denote the minimum and maximum values of X, respectively.

Uset= (min+max)/2 and S_L^2 is the variance

NL is the number of data points smaller than Uset.

NU is the number of data points greater than Uset.

skew_L=NL/(NL+NU) and skew_U=NU/(NL+NU)

f(t) show the rates of skewness in the distribution of the data.

Table 2. Setting the boundary values ‘a’ and ‘b’

	Class A		Class B	
a	b	a	b	
Attribute 1	-0.46	1.16	-0.13	1.73
Attribute 2		1.27	-0.33	1.17

Step 3: performing MTD function on each attribute by the formula

$$M^k(x) = \begin{cases} \frac{x-a}{uset-a} & a \leq x \leq uset; \\ \frac{b-x}{b-uset} & uset \leq x \leq b \\ 0 & otherwise \end{cases}$$

Table 3. MTD Function on each attribute

	x1	M(x1)	M(x1)	x2	M(x2)	M(x2)	Class
x1	0.10	0.69	0.56	0.50	0.85	0.78	A
x2	0.60	0.32	1.00	0.40	0.70	0.90	A
x3	0.85	0.38	0.78	0.12	0.28	0.78	B
x4	0.35	1.00	0.78	0.50	0.89	0.78	B
x5	0.30	0.94	0.73	0.80	0.70	0.43	A

Step 4: attribute construction

Find the correlation coefficient by using the formula

$$\sum (y_i, y_j) = \frac{cov(y_i, y_j)}{\sqrt{var(y_i) var(y_j)}}$$

$\sum(y_i, y_j)$ always lies between -1 to 1

+ve values of (y_i, y_j) means that it will increase in same direction

-ve values of (y_i, y_j) means that it will increase in reverse direction

Step 5: $(y_i, y_j) > 0.4$ implies overlapping of two classes in greater extent while

$(y_i, y_j) < 0.4 \rightarrow$ minimum overlapping between two classes

Graphically

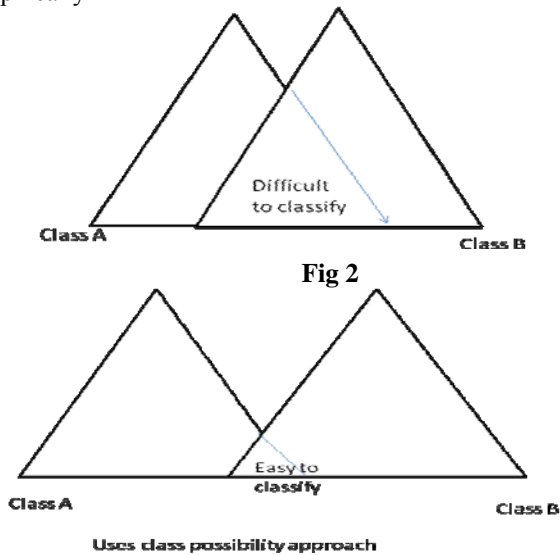


Fig 3. Overlapping of class A and B

Step 6: apply synthetic operators to the attributes having $(y_i, y_j) > 0.4$ by using the formula

- i. $y_i * y_j$
- ii. y_i / y_j
- iii. y_j / y_i

For high co-relation we find synthetic attributes

Table 4. Applying synthetic operators to the attributes

	a1	a2	a1*a2	a1/a2	a2/a1	class
x1	0.1	0.5	0.05	0.2	5	A
x2	0.6	0.4	0.24	1.5	0.67	A
x3	0.85	0.12	0.10	7.08	0.14	B
x4	0.35	0.5	0.175	0.7	1.43	B
x5	0.3	0.8	0.24	0.375	2.67	A

Step 7: merge the synthetic values obtained in step 6 with the values having $(y_i, y_j) < 0.4$ to obtain new dataset

Table 5. Merged values after applying synthetic operators

	a1'	a2'	Class
x1	-2.24	4.44	A
x2	1.01	1.29	A
x3	6.17	3.47	B
x4	-0.07	1.34	B
x5	-0.97	2.46	A

Step 8: Use any classification machine learning algorithm like PCA, LDA etc.

Observation: we observe that the performance formed by combining these attributes is greater than the individual class attributes. For any given data set, we have to build a triangular fuzzy membership function say MDT for each class in every attribute. For given set define boundaries say a and b.

If classes overlaps, it means it is difficult to classify, so in this case we apply synthetic operators and for the classes, that don't overlap we use and are easy to classify we use class possibility approach.

IV. CHALLENGES

There are many factors that are responsible for Heart disease cause, and despite of so much progress in the field of early classification and prediction of Heart Disease, there are many roadblocks that come in the growth of this advancement technology, we are going to discuss few yet the most important ones:

- a) To ascertain an unexceptional sensor [6] for the real time and compilation of physiological information from which we can foretell the incident or existence of acute CVD.
- b) To extend fast high-resolution imaging technologies for forecast
- c) To build up precise predictive model [7] due to the accessibility of variety of information sources even contained by a single huge hospital.
- d) Predicting the heart disease on the basis of perception and occurrence rather than facts based on rich data concealed in the database causes:
 - superfluous biases
 - Errors
 - Low quality of examination to patient

V. CONCLUSION

As there are many symptoms which can lead to heart disease so it becomes the primary requirement to reduce the data set so that it becomes quite easy task for doctors to clinically treat the disease with filtered attributes or values that are responsible of heart disease. This paper showcased the various types and symptoms of heart disease and proposed an algorithm with its mathematical implementation. In this proposed algorithm if we have large set of data sets we can reduce the number of data attributes which makes it quite simple for the doctors to diagnose the type of heart disease so that it can be clinically be treated successfully.

In future, we can collect a large amount of data set and create a diagnosing tool in mat lab which can help doctors in early detection of heart disease and proper treatment of the heart disease.

ACKNOWLEDGMENT

The author would like to thank R.M.Singh(senior scientist in DRDO) for his consistent support in writing this paper and also Faculty of artificial intelligence in Amity University of Science and Technology,for her ideas about heart disease that were insightful and helpful .

REFERENCES

- [1] Ankeeta R. Patel and Maulin M. Joshi Heart diseases diagnosis using Neural Network⁴IEEE International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1-5,2013.
- [2] Syed Umar Amin, Kavita Agrawal, Dr. Rizwan Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors, Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT), pp. 1227 – 1231.
- [3] Soni, J., Ansari, U., Dipesh Sharma Intelligent and Effective Heart Disease Prediction System using Weighted Associative classifiers, International Journal on Computer Science and Engineering, vol 3, No. 6, pp.2385- 2392,2011.
- [4] E.P.Ephzibah, Dr. V. Sundarapandian ANeuro fuzzy expert system forHeart disease diagnosis”, Computer Science & Engineering: An International Journal (CSEIJ), Vol.2, No.1, 2012
- [5] Der-Chiang Lia, Chih-SenWua, Tung-I Tsaib, Yao-San Lina (2007)Using mega-trend-diffusion and artificial samples insmall data set learning for early flexible manufacturing system scheduling knowledge,Computers & Operations Research 34 966–982
- [6] Yuan-Ting Zhang, Ya-Li Zheng, Wan-Hua Lin, He-Ye Zhang, and Xiao-Lin Zhou (2013) Challenges and Opportunities in Cardiovascular Health Informatics, IEEE Transactions on Biomedical Engineering, Volume:60 , Issue: 3, pp.633 – 642, ISSN: 0018-9294 .
- [7] Kiyana Zolfaghar, NarenMeadem, AnkurTeredesai, SenjutiBasu Roy, Si-Chi Chin and Brian Muckian (2013) Big Data Solutions for Predicting Risk-of –Readmission for Congestive Heart Failure Patients in conference on Big Data Solution, IEEE, pp. 64-71.
- [8] UCI Machine Learning Repository [homepage on the Internet]. Arlington: The Association; 2006 [updated 1996 Dec 3; cited 2011 Feb 2]. Available from: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>