International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 8, August - 2014

Heart Disease Diagnosis using Genetic and Particle Swarm Optimization

Swati Sharma
Student of M.tech CSE
N.C.C.E , Israna , Panipat
Jind , Haryana , India

Dr. Sukhvir Singh
H.O.D in C.S.E dept.
N.C.C.E , Israna
Jind , Haryana , India

***Abstract* : Heart disease is the leading Causes of death in the world over the past ten years. The world health organization reported that heart disease in the first leading causes of death in high and low countries . Today diagnosing patients correctly and administering effective treatments have become quite a challenge. Poor clinical decisions may end to patient death and which cannot be afforded by the hospital as it loses its reputation.. So to achieve a correct and cost effective treatment computer-based information and/or decision support systems can be developed to do the task. Now several different methods have been developed for designing a heart disease prediction system . In this paper we are designing a computerized method for predicting heart disease with the help of Data mining and Optimization techniques like PSO and Genetic . The reason behind using Genetic Algorithm is the discovery of high level prediction rules that are highly comprehensible, have high predictive accuracy.**

Keywords— Particle swarm optimization , swarm intelligence , particle , heuristic .

I. INTRODUCTION :

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?" This is the main motivation for this paper.

II. DATA MINING :

The information industry has a very large amount of data. Till data is not converted into useful information it is of no use. It is necessary to analyse this huge amount of data and extract useful information from it . The extraction of information is followed by several other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation . After all these processes are completed, we are now in a position to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration etc.

Data mining uses two strategies: supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used And in this paper I am using this data for heart disease diagnosis.

III. ALGORITHMS DISCUSSED :

Now here is a brief overview of the data mining and optimization algorithms used in my this work :-

1. Particle swarm optimization
2. Genetic Algorithm

Particle Swarm Optimization:

Particle Swarm optimization is a heuristic global optimization technique . It was firstly discovered and described by James Kennedy and Russell C. Eberhart in 1995.[1] This technique was proposed from the study of swarm intelligence . Swarm or a group of flocks when search for food , the type of intelligence they use in interacting with their friend swarms is the main principle behind the origin of this technique. When a group of swarms go for searching food , either they go together or in group till they find a place where food can be found. One of the bird among them searches the best food called the optimum search . Now every bird is moving with some velocity to search the food. Now the method which this bird adopts to convey the message of best food to all other birds and the flock comes to that place is used in PSO . It was thought that this interaction among birds can be efficiently utilized in finding the optimum solution. These birds or swarms are said to be particles in particle swarm optimization. Now in PSO every particle is moving with some velocity and when the optimum solution is found by one particle , there is a memory which helps in conveying the message to all other particles.

GENETIC ALGORITHM :

Genetic algorithm is a type of searching algorithm . This algorithm works on a population of people . The collection of candidate solutions is called population that we are considering during the course of algorithm. New members are born into the population over the generations of the algorithm, while others die out of the population. A single solution in the population is referred to as an individual. The fitness of an individual is a measure of how "good" the solution represented by the individual is. The better the solution, the higher the fitness – obviously, this is dependent on the problem to be solved. The algorithm creates a "population" of possible solutions to
the problem and lets them "evolve" over multiple generations to find better and better
solutions.

IV. ADVANTAGES OF PSO OVER OTHER ALGORITHMS :

There are several advantages of PSO over other algorithms . Some are described below:-

- The results with PSO are much accurate as compared to other algorithms.
- PSO takes much less time as compared with other algorithms due to its interaction with other particles.
- It is based on intelligence.
- In PSO the real number code is adopted and it is decided directly by the solution .
- With the help of PSO researching is much fast and it occupies better optimization ability.

V. PROPOSED METHODOLOGY :

Problem Definition

- Poor clinical decisions may end to patient death and which cannot be afforded by the hospital as it loses its reputation. The cost to treat a patient with a heart problem is quite high and not affordable by every patient. To achieve a correct and cost effective treatment computer-based information and/or decision support Systems can be developed to do the task

- Most hospitals today use some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped.

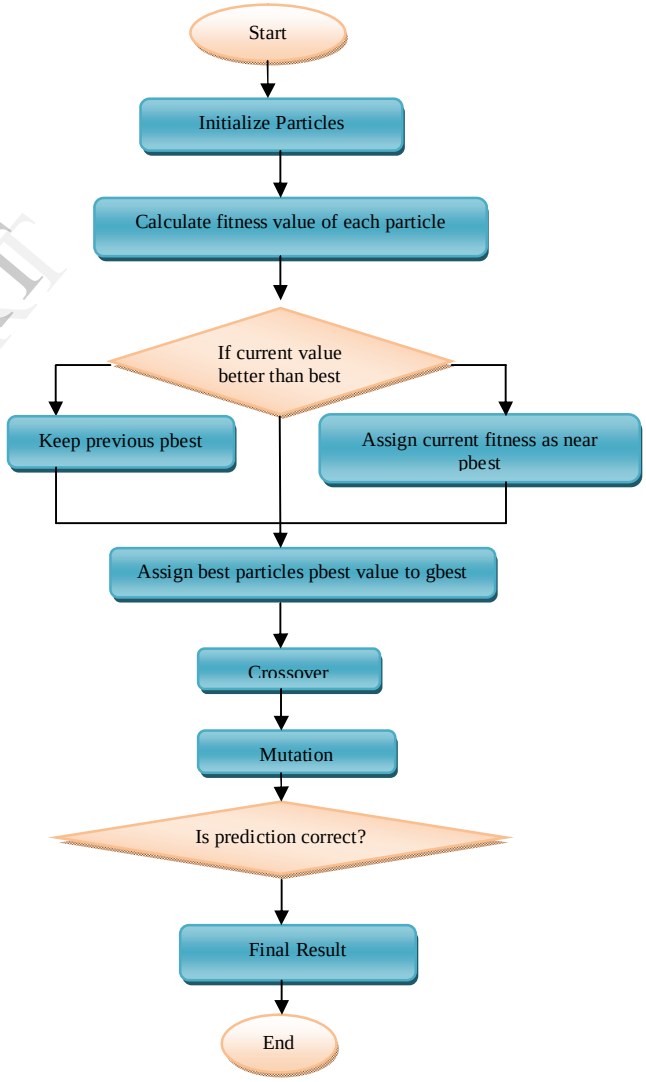
- PSO does not have genetic operators like crossover and mutation.
-
- Compared with GA, all the particles tend to converge to the best solution quickly even in the local version in most cases.

VI. PROPOSED SYSTEM :

To solve the above problems, we proposed a method in which PSO integrate with genetic algorithm to improve the accuracy of result.

Firstly initialize the random particles and searched for optimized solution. Updating of particles take place by fitness value called pbest and a global value gbest or lbest. This gbest value is the best value obtained by particle swarm optimization. In PSO updation of particle take place by its velocity but in GA this process is done by crossover and mutation .So we used GA for updation and sharing information with each other that's why whole flakes of data or particle moves in one bundle towards an optimized area and the resultant accuracy of prediction is more accurate as compare to individuals. Flow chart of proposed method is:

VII. FLOWCHART OF PROPOSED ALGORITHM :



VIII. IMPROVED ALGORITHM :

1. Start the process.
2. Intialize the particles .
3. Calculate fitness value of each paricle.
4. If the current value of particle is better than the pbest assign this the new pbest value.
5. If the current value of particle is less than the pbest value , keep the previous pbest value.
6. Now assign best particle pbest value to the gbest value.
7. Now start the crossover process.
8. Now start mutation.
9. If prediction is correct , then we got the final result.
10. End of algorithm.

IX. DATABASE USED :

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	male	typ_angina	145	233	t	left_vent_hy...	150	no	2.3	down	0	fixed_defect	<50
67	male	asympt	160	286	f	left_vent_hy...	108	yes	1.5	flat	3	normal	>50_1
67	male	asympt	120	229	f	left_vent_hy...	129	yes	2.6	flat	2	reversable...	>50_1
37	male	non_anginal	130	250	f	normal	187	no	3.5	down	0	normal	<50
41	female	atyp_angina	130	204	f	left_vent_hy...	172	no	1.4	up	0	normal	<50
56	male	atyp_angina	120	236	f	normal	178	no	0.8	up	0	normal	<50
62	female	asympt	140	268	f	left_vent_hy...	160	no	3.6	down	2	normal	>50_1
57	female	asympt	120	354	f	normal	163	yes	0.6	up	0	normal	<50
63	male	asympt	130	254	f	left_vent_hy...	147	no	1.4	flat	1	reversable...	>50_1
53	male	asympt	140	203	t	left_vent_hy...	155	yes	3.1	down	0	reversable...	>50_1
57	male	asympt	140	192	f	normal	148	no	0.4	flat	0	fixed_defect	<50
56	female	atyp_angina	140	294	f	left_vent_hy...	153	no	1.3	flat	0	normal	<50
56	male	non_anginal	130	256	t	left_vent_hy...	142	yes	0.6	flat	1	fixed_defect	>50_1
44	male	atyp_angina	120	263	f	normal	173	no	0	up	0	reversable...	<50
52	male	non_anginal	172	199	t	normal	162	no	0.5	up	0	reversable...	<50
57	male	non_anginal	150	168	f	normal	174	no	1.6	up	0	normal	<50
48	male	atyp_angina	110	229	f	normal	168	no	1	down	0	reversable...	>50_1
54	male	asympt	140	239	f	normal	160	no	1.2	up	0	normal	<50
48	female	non_anginal	130	275	f	normal	139	no	0.2	up	0	normal	<50
49	male	atyp_angina	130	266	f	normal	171	no	0.6	up	0	normal	<50
64	male	typ_angina	110	211	f	left_vent_hy...	144	yes	1.8	flat	0	normal	<50
58	female	typ_angina	150	283	t	left_vent_hy...	162	no	1	up	0	normal	<50
58	male	atyp_angina	120	284	f	left_vent_hy...	160	no	1.8	flat	0	normal	>50_1
58	male	non_anginal	132	224	f	left_vent_hy...	173	no	3.2	up	2	reversable...	>50_1
60	male	asympt	130	206	f	left_vent_hy...	132	yes	2.4	flat	2	reversable...	>50_1
50	female	non_anginal	120	219	f	normal	158	no	1.6	flat	0	normal	<50
58	female	non_anginal	120	340	f	normal	172	no	0	up	0	normal	<50
66	female	typ_angina	150	226	f	normal	114	no	2.6	down	0	normal	<50

PRINCIPAL INVESTIGATOR OF THIS DATASET :-

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D. 2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Explanation of database entities :-
This consists of 76 attributes while 14 of them are actually used.
1 id: patient identification number
2 ccf: social security number (I replaced this with a dummy value of 0)
3 age: age in years
4 sex: sex (1 = male; 0 = female)
5 painloc: chest pain location (1 = substernal; 0 = otherwise)
6 painexer (1 = provoked by exertion; 0 = otherwise)
7 relrest (1 = relieved after rest; 0 = otherwise)
8 pncaden (sum of 5, 6, and 7)
9 cp: chest pain type
-- Value 1: typical angina
-- Value 2: atypical angina
-- Value 3: non-anginal pain
-- Value 4: asymptomatic
10 trestbps: resting blood pressure (in mm Hg on admission to the hospital)
12 chol: serum cholestoral in mg/dl
13 smoke: I believe this is 1 = yes; 0 = no (is or is not a smoker)
14 cigs (cigarettes per day)
15 years (number of years as a smoker)
16 fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
17 dm (1 = history of diabetes; 0 = no such history)
18 famhist: family history of coronary artery disease (1 = yes; 0 = no)
19 restecg: resting electrocardiographic results
-- Value 0: normal
-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
-- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
20 ekgmo (month of exercise ECG reading)
21 ekgday(day of exercise ECG reading)
22 ekgyr (year of exercise ECG reading)
23 dig (digitalis used furing exercise ECG: 1 = yes; 0 = no)
24 prop (Beta blocker used during exercise ECG: 1 = yes; 0 = no)
25 nitr (nitrates used during exercise ECG: 1 = yes; 0 = no)
26 pro (calcium channel blocker used during exercise ECG: 1 = yes; 0 = no)
27 diuretic (diuretic used used during exercise ECG: 1 = yes; 0 = no)
28 proto: exercise protocol
1 = Bruce
2 = Kottus
3 = McHenry
4 = fast Balke
5 = Balke
6 = Noughton
7 = bike 150 kpa min/min (Not sure if "kpa min/min" is what was written!)
8 = bike 125 kpa min/min
9 = bike 100 kpa min/min
10 = bike 75 kpa min/min
11 = bike 50 kpa min/min
12 = arm ergometer
29 thaldur: duration of exercise test in minutes
30 thaltime: time when ST measure depression was noted
31 met: mets achieved
32 thalach: maximum heart rate achieved
33 thalrest: resting heart rate
34 tpeakbps: peak exercise blood pressure (first of 2 parts)
35 tpeakbpd: peak exercise blood pressure (second of 2 parts)
36 dummy
37 trestbpd: resting blood pressure
38 exang: exercise induced angina (1 = yes; 0 = no)
39 xhypo: (1 = yes; 0 = no)
40 oldpeak = ST depression induced by exercise relative to rest
41 slope: the slope of the peak exercise ST segment
-- Value 1: upsloping
-- Value 2: flat

-- Value 3: downsloping
42 rldv5: height at rest
43 rldv5e: height at peak exercise
44 ca: number of major vessels (0-3) colored by flourosopy
45 restckm: irrelevant
46 exerckm: irrelevant
47 restef: rest raidonuclid (sp?) ejection fraction
48 restwm: rest wall (sp?) motion abnormality
0 = none
1 = mild or moderate
2 = moderate or severe
3 = akinesis or dyskmem (sp?)
49 exeref: exercise radinalid (sp?) ejection fraction
50 exerwm: exercise wall (sp?) motion
51 thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
52 thalsev: not used
53 thalpul: not used
54 earlobe: not used
55 cmo: month of cardiac cath (sp?) (perhaps "call")
56 cday: day of cardiac cath (sp?)
57 cyr: year of cardiac cath (sp?)
58 num: diagnosis of heart disease (angiographic disease status)
-- Value 0: < 50% diameter narrowing
-- Value 1: > 50% diameter narrowing

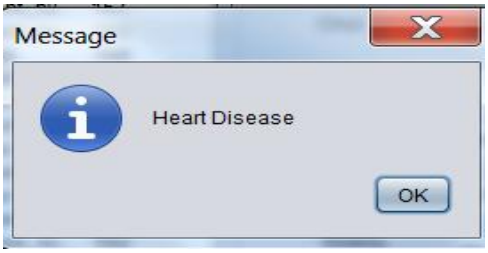
X. WORKING AND RESULTS OF THIS RESEARCH

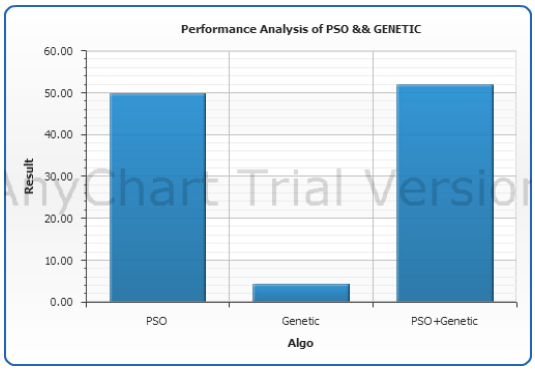
ACCURACY OF THE DATA SET =

$$\frac{\text{Number of objects correctly Classified} * 100\ \%}{\text{Total No. of objects in the test set}}$$

RESULT 1 :

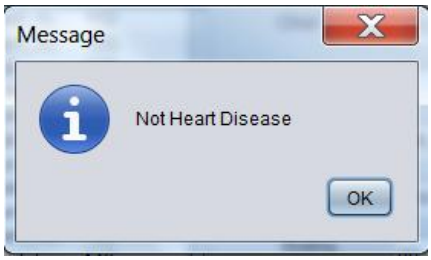
Age = 63
Sex = male
CP = typ_angina
TrestBPS = 145
Chol = 233
FBS = t
RestECG = left_vent_hyper
Thalach = 150
Exang = no
OldPeak = 2.3
Slope = down
Ca = 0
Thal = fixed_defect

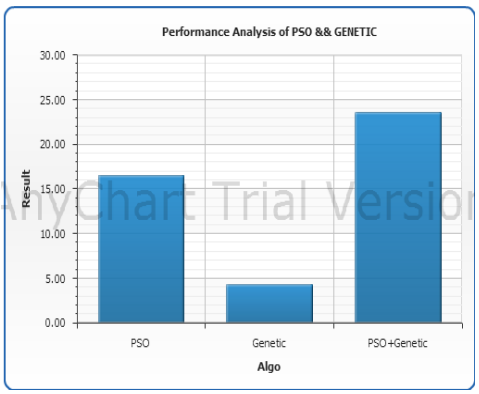




RESULT 2 :

Age = 41
Sex = male
CP = asympt
TrestBPS = 110
Chol = 172
FBS = f
RestECG = left_vent_hyper
Thalach = 158
Exang = no
OldPeak = 0
Slope = up
Ca = 0
Thal = reversable

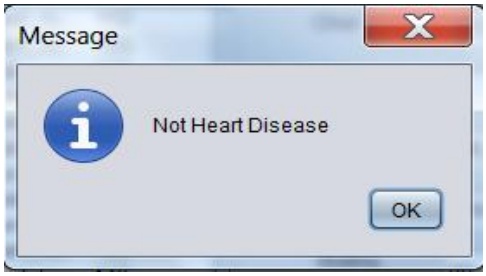


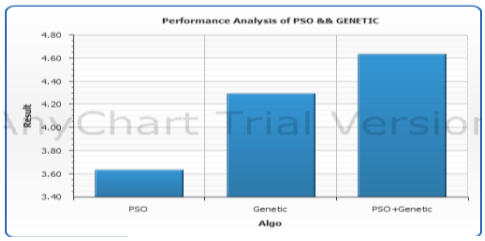


RESULT 3 :

Age = 54
Sex = male
CP = non_anginal
TrestBPS = 125
Chol = 273
FBS = f

RestECG = left_vent_hyper
Thalach = 152
Exang = no
OldPeak = 0.5
Slope = down
Ca = 1
Thal = normal





XI. RESULTS AND DISCUSSION :

These results show that after combining the two techniques the results which come are more accurate than the other techniques used so far.

The graphs shown has algorithms on x- axis and

% results of accuracy on y – axis.

The graphs clearly show that PSO and Genetic

graphs are smaller than the hybrid graphs.

It means PSO and Genetic are less accurate and slow but when we combine the two the accuracy increases.

XII. CONCLUSION :

Around 18 million people, 7 % Indians are affected by heart disease. Heart disease is mostly affected the person under the age of 65. This thesis is based on the heart disease diagnosis of patients. Heart disease is a prevailing disease nowadays. Now due to increasing expenses of heart disease , there was a need to develop a new system which can predict heart diseases in an easy and cheaper way. Various methods had developed previously which had given methods to predict heart disease. The diagnosis is based on data mining processes. Data mining is a process of extraction of knowledge from the data in the database. In the database irrelevant data is present. Till now several data mining techniques namely classification, clustering, fuzzy system and association rules are applied to the health data sets for predicting heart diseases. In current study PSO and Genetic are used for the heart disease prediction. PSO has some benefits over Genetic and Genetic have some other benefits over PSO. So, in this current study all these two techniques are combined to give rise to a good prediction system. This system is more fast and easy to implement. The performance analysis graphs show that PSO and Genetic combined is much better than the individual techniques. In future, new algorithms and techniques are to be developed which overcome the drawbacks of the existing system.

IX. REFERENCES

[1] Analysis of Particle Swarm Optimization Algorithms- Qinghai Bai
[2] Particle Swarm Optimization – A survey – Keisuke Kameyma
[3] Intelligent Heart Disease Prediction System using Data Mining Techniques
[4] A novel hybrid classification method with particle swarm optimization and k-nearest neighbor algorithm for diagnosis of coronary artery disease using exercise stress test data .
[5] Performance analysis of classification data mining techniques over heart disease data base .
[6] Data Mining Techniques For Heart Disease Prediction
[7] Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques
[8] Using data mining techniques in heart disease diagnosis and Treatment.
[9] A comparison of particle swarm Optimization and the genetic algorithm.
[10] Constricted Particle Swarm Optimization based Algorithm for Global Optimization
[11] Introduction to data mining.
[12] Optimizing with genetic algorithms
[13] An Introduction to Genetic algorithms
[14] Heart Disease Prediction System using Associative Classification and Genetic Algorithm
[15] A Study of Heart Disease Prediction in Data Mining