International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34109-34113

Empirical Study On Heart Disease Prediction System using Particle Swarm Optimization Clustering Mechanism

Sana Bharti

M.tech Scholar Amity,University,Noida
bhartisana90@gmail.com

Dr.Shaliendra Narayan Singh

Associate Professor Amity,University,Noida
snsingh36@amity.edu

Abstract
The medical health care centers has a large collection of health care data that is not organized and mined properly. Nowadays Data mining helps in predicting several medical diseases. In medical science, Bioinformatics, data mining helps in discovering the hidden information and potentially useful information from a large set of database and respirators which help in predicting the disease with greater accuracy [12]. Basically data mining is the process of finding co-relation and pattern among large database by using advanced data mining techniques [19]. A heart disease prediction system based on clinical records of patients is being developed that may help many medical professionals in predicting heart disease. This approach includes four stages. First, I have, selected 11 important features, which are maximum heart rate sex, age, rest electrocardiogram , cholesterol, slope, trestbps , old peak , fasting sugar , chest pain type, thal from the Cleveland dataset which have 76 attributes and 303 records. Then make two clusters of these these important particles according to analytical study one cluster has non patients and another have patients suffering from heart disease. Thirdly, Particle Swarm Algorithm is implemented so as to determine the updated velocities of the particles which is calculated from input values of a new patient. Finally the heart disease is predicted by calculating two times Euclidean Distance, the first one between the updated velocities and average values of each particle in the cluster who are suffering from heart disease(ed1) and the second one between updated velocities and average value of each particle in the cluster who are not suffering from heart disease(ed0) and if ed0<ed1 then Patient is not suffering from heart disease else he is suffering from heart disease. The accuracy of prediction is near 85%.

Index Terms: Clustering, Heart Disease ,Particle Swarm Optimization,Ecludiean Distance.

1. Introduction

Heart disease, which is known as a heart and blood vessel disease, is one of the most life threatening disease and thus becoming an important part of research for many researchers in medical as well as in computer science field. According to statistical data of the British Heart Foundation , Cardiovascular disease causes more than a quarter of all deaths in U.K and around 160,000 deaths each year. Approximately around 7 million people in U.K are suffering from cardiovascular disease. According, to statistical data of the Indian government the frequency of occurring heart failure due to coronary heart disease, high cholesterol, hypertension or obesity ranges from approximately 1.3 million to 4.6 million. About 30% of the urban population and about 15% rural population is suffering from heart attacks and high blood pressure. Another important fact that should be highlighted is that every year in India around 130,000 children are born with heart disease. Heart Disease gives rise to many problems like angina,coronary heart disease , arrhythmia etc and the first step followed by these problems is Atherosclerosis. Atherosclerosis arise in the body when plaque started building up in the walls of the arteries, which will narrow the arteries that create difficulty for the blood to flow. Eventually a clot is formed. The clot blocks the flow of the blood completely from the particular part of the heart muscle, which is abound by that artery and it begins to die and this lead to heart attack [23]. Identify the risk factors of cardiovascular disease and diagnose them at an early stage ,consult to the doctor for its treatment and take preventive measures reduce the cardiac mortality and morbidity. Many predictive models have been proposed by several researchers. These model develop a predictive model by combining different type of data with different type of intelligent algorithms for diagnosis. The proposed system has developed a Prediction System to predict Heart Disease that will assist medical professionals in predicting heart disease status based on the clinical data of patients. This approach includes four stages. First, I have,selected 11 important features, which are maximum heart rate sex, age, rest electrocardiogram, cholesterol, slope, trestbps , old peak, fasting sugar, chest pain type, thal. Then cluster these important particles according to analytical study. Thirdly, Particle Swarm Algorithm is implemented so as to determine the updated velocities of the particles. Finally the heart disease is predicted by calculating Euclidean Distance between the updated velocities and input values and classify the heart disease according to the above mentioned clinical features. Our approaches include four steps. The important particles are clustered from the provided features.The data is

clustered in two groups in patients and non patients. These clinical features will work as parameters for PSO. The choice of PSO parameters can have a large impact on optimizating performance. Selection of PSO parameters can give good performance. Using attributes like age, sex, blood sugar, cholesterol helps in predicting the heart disease. It establishes a significant knowledge or relation between medical attributes to predict disease.

1.1 Data Mining

Data mining is a process of extracting significant ,previously unknown data which is stored in large database or resporartries Useful information helps to increase revenue, save cost and time .Data mining is a combination of various fields like machine learning, neural network,statistics,pattern recognition.Software for data mining helps for analyzing the data. Data mining basically based on two approaches:

a) Supervised learning:Supervised training is a machine learning task of gaining knowledge from labeled training dataset , i.e. an external person gives neural network input data and its actual output and seeks to build a predictive model that will able to generate correct predictions of new data.
b) Unsupervised learning: Unsupervised is also a machine learning task which predicts the output from dataset consisting of input data without labeled responses. It is an iterative learning process in which input is given to the neural network and weights are associated with inputs are adjusted or updated each time. Cluster analysis is one of the methods of unsupervised training [17].

Data mining is a technique of selecting,exploring and modeling a huge collection of data and then find similarities or the relation between the data which are unknown, so as to get clear and useful information,yet we can say that through data mining we discover new knowledge from data.Thus data mining is popularly known as Knowledge mined from data.

1.2 Particle Swarm Optimization

Particle Swarm Optimization is invented by Kennedy and Eberhart, which imitate the behavior of bird flocking[13]. Particle Swarm is a group of birds or fish schools in which all these particles are searching for food and attain a better position in each iteration. But in each iteration they do not know how much away they are from their food. Then what will be the best strategy to find their food? A best way is to follow the bird which is nearer to the food. In this paper , I am relating Heart Disease prediction with the particle swarm intelligence algorithm. In PSO a group of random particles in an N dimensional space is initialized and each particle has its own velocity and position.In each and every iteration particle is being updated with two numeric values and record them:

a) The first is the best solution, it had achieved . That fitness value will be stored in the memory. This value is known as P_{best}.
b) Second "best" value that is tracked by the PSO optimizer is the best value, obtained by any of the particle in the population. This value is known as global (G_{best}).

After we have found these two numeric values we update the position and velocity of particle by using the below two equations.

$$V_{i+1}^{k}[] = w*V_i^{k}[] + b1* \mu()$$

$$*(P_{best}^{k}[] - pre_i^{k}[]) + b2*\mu()*(G_{best}^{k}[] - pre_i^{k}[]) \quad (1)$$

$$pre^{k}_{i+1}[] = pre^{k}_{i}[] + V^{k}_{i+1}[] \quad (2)$$

V_{i+1} is the velocity of kth particle at i+1 time , pre is the present velocity of particle, w is the inertia factor which have controls on the velocity of the particle through weighing the contribution of particles previous velocity and control the previous particles' flight direction which will influence the new velocity , its value range[0,1] to have good convergent behaviour and normally its value taken as 0.7298, μ is a number between (0,1) ,it is used to have good exposure of the problem space and avoid entrapment in local minima ,b1,b2 are positive learning vectors and usually b1=b2=2. If the V_{i+1}^{k} exceed the given value , then velocity at that position is equal to V_{max} .

1.3 Clustering

Clustering is an unsupervised data mining technique which helps in partitioning of unstructured data into the different clusters based on the degree of similarity. Objects in Each cluster are same with respect to each other and dissimilar to the other cluster. Clustering is one of the popular approach as it automatically finds the concepts, classes or pattern. In heart disease prediction system using clustering we get the cluster of all the patients of same risk factor like in one cluster we have group of patients who have high cholesterol , blood pressure, fasting blood sugar , abnormal ECG, slope, Thal, etc. and in other cluster we have patients having everything normal and are healthy [21]. Basically, there are two types of Clustering namely: Hierarichal and Partioning.

a) **Hierarchical Clustering:** It has nested set of clusters means each cluster has a sub cluster i.e. data is not put into one specific cluster in first step.In hierarchical clustering sequence of partition takes place in a single cluster so it can be viewed as tree having sub clusters at the nodes. Hierarchical clustering is of two types: bottom up (agglomerative) or bottom up (divisive). In agglomerative each and every object has its own unique cluster and iteratively merges into one large cluster. In divisive there is a single cluster contain all the objects and then splits into a sub cluster or small clusters [5].
b) **Partitional Clustering:** Partional Clustering first determines all the clusters and then partition each the data object into different cluster based upon their degree of similarity. Each and every data object will placed in only one cluster. A Partitional Clustering is suited for clustering of a large dataset [5]. The partition is done on some criteria like minimizing square criteria etc.

2. Proposed Algorithm

Proposed algorithm works like Dynamic Algorithm with Particle Swarm Optimization. Dynamic algorithm is commonly used as an example of the elegance of recursion as a *programming* technique. It is used when the solution can be recursively described in terms of solutions to sub problems. (Optimal substructure). In my algorithm values we enter for a patient like ecg,blood sugar,ecg,cholesterol etc are divided into subproblems i.e we calculate velocities of each particle separately using Particle Swarm Optimization and then recombine them by calculating two times Euclidean Distance,first one between the updated velocities and average values of each particle in the cluster who are suffering from heart disease(ed1) and the second one between updated velocities and average value of each particle in the cluster who are not suffering from heart disease(ed0) and if ed0<ed1 then Patient is not suffering from heart disease else he is suffering from heart disease. to find out whether the person is suffering from heart disease or not.The figure below shows the flow of work.

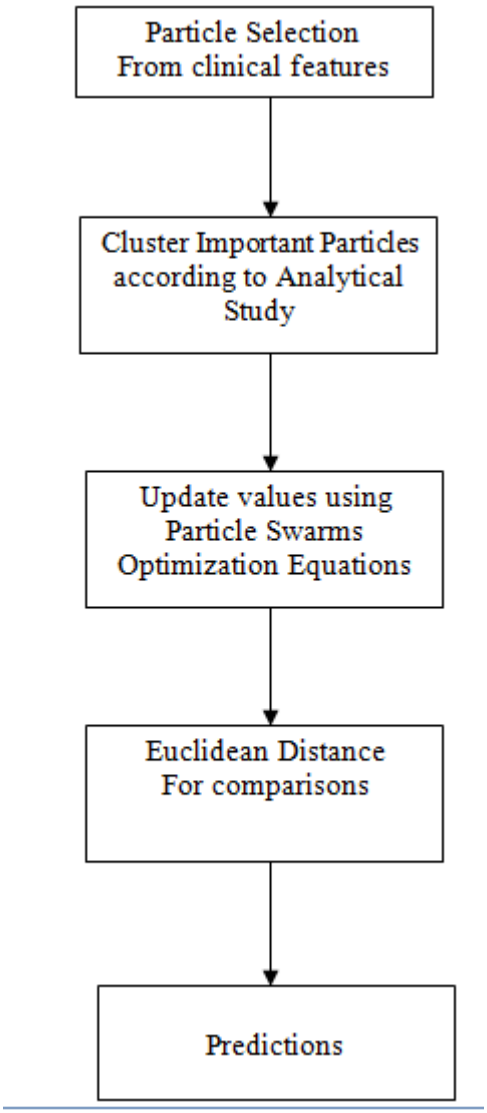

Figure1: Flow of proposed Algorithm.

3. Calculation of Pbest and Gbest Values

For implementation, the proposed work is using following attributes. These attributes work as particles in Particle Swarm Optimization. PSO is initialized with group of these particles and then search for a best solution by updating the generations. In each iteration ,each and every particle In every iteration, each particle is updated by 2 best values

a) First is the best solution or fitness particles had achieved up till now, which is called as pbest.That fitness values is stored one . Suppose, for cholesterol the pbest is 240, for blood pressure pbest is 120(as it is fitness value a normal person to have normal BP) .

b) Second best value is tracked by a particle swarm optimzer is the best value achieved by any particle in a population up till now.This global best value is called as gbest The gbest for cholesterol is 269 as the best value calculated in the cluster as shown in figure 2.

Each particle has clustered data. There are two clusters,one of the persons who are predicted and second of the persons not predicted for heart disease.For each particle, there is a best fitness value called pbest of that particle. The gbest is the group best value in the cluster. The below table shows pbest and gbest values of all 11 particles.

Table 1: Particle Best and Group Best values.

Particle	pbest value	gbest value
Age	53	57
Sex	1	1
Cholesterol	240	269
BP	120	150
Thal	3,6	4.27
Old Peak	2	2
Slope	2	2
Blood Sugar	120	0.18
ECG	0,1	1.3
Chest Pain (CP)	1,2,3	3.75

4. Interface of Heart Disease Prediction System

a) Splash Window comes.

b) A Window comes having three buttons: Analyical Study of Patients,Analytical Study of Non Patients,Intialize particles and optimize..

c) On clicking Analytical Study Of Patients and Non Patients a bar graph comes showing analysis.

d) On clicking on Intialize Particles and optimize a window comes in which user enter the 11 clinical values.

e) Submit Button shows the result, whether predicted or not predicted.

f) View result and Decision button show the graph showing in which range the patient is lying.

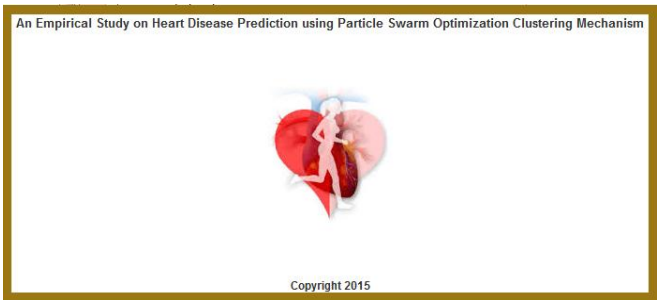

Figure2: Welcome Screen.

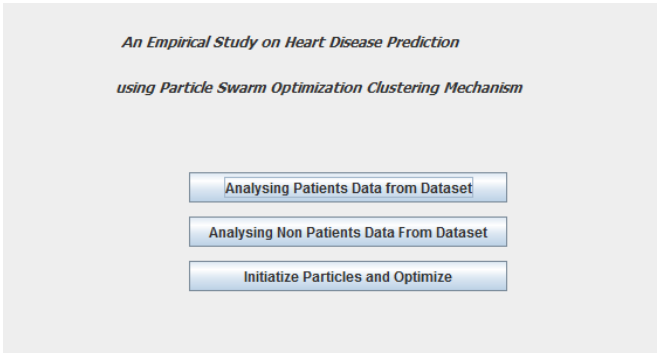

Figure 3: Home Screen

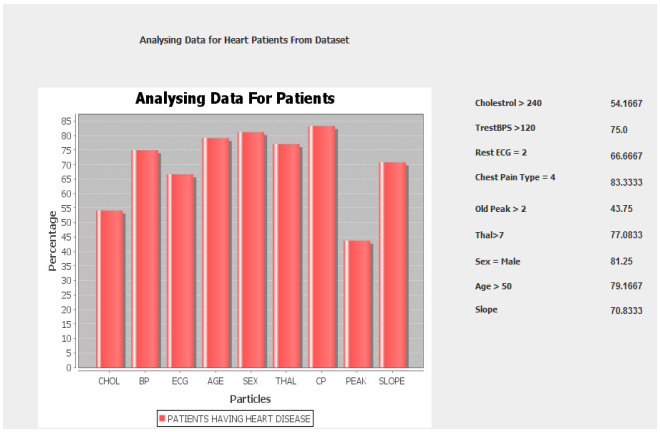

Figure 4: Analyzing of data for patients

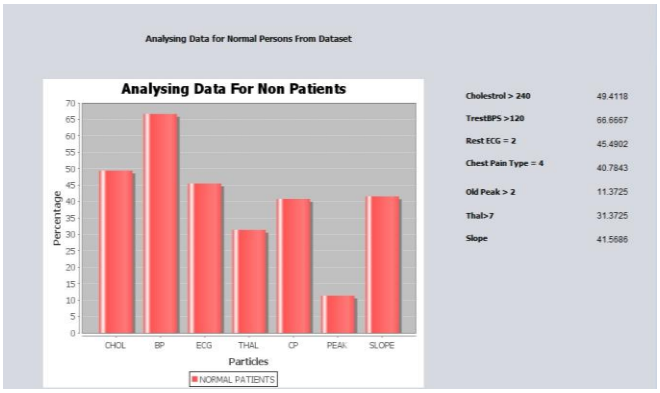

Figure 5:Analyzing data for Non-Patients

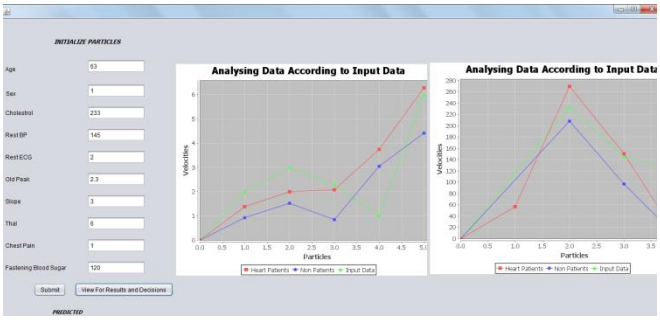

Figure 6: Showing result.

Figure 4 shows the analyzes that person suffering from heart disease i.e

- Patents having Cholesterol>240 is 54.16%
- Patients having Abnormal BP>120 is 75.0%
- They have Chest Pain of type 4 with an enlarged heart=83%
- 77% of the patients have a thal = 7.
- 79% Patients have the age greater than 50.

Figure 5 shows analyzes of non _patients i.e

- Patents having Cholesterol>240 is 49.4%
- Patients having Abnormal BP>120 is 66%
- They have Chest Pain of type 4 =40.7%.
- 31.3% of the patients have a thal = 7.

In figure 6 first graph is plotted by calculating rest ecg,old peak,age,sex,thal,chest pain and the second graph is plotted by calculating cholesterol and resting blood pressure and fasting blood sugar.

5. Conclusion

In this approach, heart disease prediction system was developed using PSO and clustering mechanism in Net beans.Using this approach, the accuracy rate is 85% when applied to my test data set. The importance of PSO is that it has the capability to handle difficult problems.PSO has availability of solutions.PSO is more robust and has a property of inherent parallelism. PSO deals with difficult parameter optimization, but provides the significant results with clustering of particles. Further, using the Euclidean distance between a updated particles velocity and input values. The purpose of a measure of similarity is to compare two lists of data (i.e. vectors of updated velocities values and input values), and compute a single number which evaluates their similarity to predict heart disease.This technique is supposed to be successful in predicting the heart disease using Particle Swarm optimization.

REFERENCES

[1] M. Akhil Jabbar, B.L Deekshatulu & Priti Chandra, **“** Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection”,

Global Journal of Computer Science and Technology Neural & Artificial Intelligence, Volume 13 Issue 3 Version 1.0 Year 2013.

[2] Ms. Preeti Gupta ,Ms. Punam Bajaj, " Heart Disease Diagnosis System Based On Data Mining And Neural Network", INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, June, 2014.

[3] T.John Peter, K. Somasundaram, " AN EMPIRICAL STUDY ON PREDICTION OF HEART DISEASE USING CLASSIFICATION DATA MINING TECHNIQUES", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012.

[4] Qiang Niu, Xinjian Huang, " An Improved Fuzzy C-means Clustering Algorithm based on PSO", JOURNAL OF SOFTWARE, VOL. 6, NO. 5, MAY 2011.

[5] Sunita Sarkar,Arindam Roy,Bipul Shyam Purkayastha, " Application of Particle Swarm Optimization in Data Clustering: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 65– No.25, March 2013.

[6] Rania Hassan Babak Cohanim Olivier de Weck, "A COPMARISON OF PARTICLE SWARM OPTIMIZATION AND THE GENETIC ALGORITHM",46 AIAA Structue,Strctural Dynamics and and Material Confrence ,18-21 April 2005,Austin ,Texas

[7] Dr. G. M. Nasira, S. Radhimeenakshi," Particle Swarm Optimization- A Review on Algorithmic Rule, Application and Scope", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[8] Azhar W. Hammad, Dr. Ban N. Thannoon, " GENETIC ALGORITHM VERSUS PARTICLE SWARM OPTIMIZATION IN N-QUEEN PROBLEM", Journal of Al-Nahrain University Vol.10(2), December, 2007, pp.172-177.

[9] Russell C. Eberhart and Yuhui Shi, " Comparison between Genetic Algorithms and Particle Swarm Optimization", EP'98 Proceeding of the 7th International Conference on Evolutionary VII Pages 611-616.

[10] AH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin, "HDPS: Heart Disease Prediction System"10 Department of Medical Informatics, Tzu Chi University, Hualien City, Taiwan.

[11] Bhuvaneswari Amma N.G"Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network",IEEE

[12] Sheik Abdullah, " A Data Mining Model to Predict and Analyze the Events Related to Coronary Heart Disease using Decision Trees with Particle Swarm Optimization for Feature Selection", International Journal of Computer Applications Volume 55– No.8, October 2012.

[13] Sapna Katiyar," A Comparative Study of Genetic Algorithm and the Particle Swarm Optimization ".

[14] Sivagowry.S , Dr. Durairaj.M, " PSO - An Intellectual Technique for Feature Reduction on Heart Malady Anticipation Data",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 9, September 2014.

[15] S. Muthukaruppan , M.J. Er, "A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease",Expert Sysytem with Applications, 2012,page no: 11657-11665.

[16] G. Komarasamy1 and Amitabh Wahi "IMPROVING THE CLUSTER PERFORMANCE BY COMBINING PSO AND K-MEANS ALGORITHM", ICTACT Journal on Soft Computing, vol.1, issue.4, pp.206- 208, 2011.

[17] N. Aditya Sundar, P. Pushpa Latha , M.Rama Chandra , "PERFORMANCE ANALYSIS OF CLASSIFICATION DATA MINING TECHNIQUES OVER HEART DISEASE DATA BASE , INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY, Volume-2, Issue-3, page no: .470 – 478

[18] Wei Hong Lim, Nor Ashidi Mat Isa, "Particle swarm optimization with increasing topology connectivity, Engineering Applications of Artificial Intelligence, 2014, page no: 80-102, link www.elsevier.com/locate/engappai.

[19] Atul Kumar Pandey,Prabhat Pandey ,K.L. Jaiswal,Ashish Kumar Sen , " A Heart Disease Prediction Model using Decision Tree" , IOSR Journal of Computer Engineering (IOSR-JCE) ,Volume 12, Issue 6 (Jul. - Aug. 2013), PP 83-86.

[20] Beant Kaur , Williamjeet Singh, " Review on Heart Disease Prediction System using Data Mining Techniques", International Journal on Recent and Innovation Trends in Computing and Communication , Volume: 2 Issue: 10 , page no: 3003 – 3008.

[21] Online Avilable,Website Link https://www.heart.org/Conditions

[22] Pritesh Vora, Bhavesh Oza, "A Survey on K-mean Clustering and Particle Swarm Optimization", International Journal of Science and Modern Engineering (IJISME) ISSN: 2319-6386, Volume-1, Issue-3, February 2013