International Journal For Technological Research In Engineering
Volume 3, Issue 10, June-2016 *ISSN (Online): 2347 - 4718*

A NOVEL TECHNIQUE FOR PREDICTIVE ANALYSIS OF CRITICAL DISEASE USING HYBRID FORM OF CLUSTERING, ASSOCIATION AND PRIORITY ALGORITHMS

Vikas Siwach[1], Satish Kumar[2]
[1]Assistant Professor, CSE Department, [2]M.Tech (SE) Department, U.I.E.T, M.D.U, India

Abstract: In recent scenario data mining play a great deal for various application. In medical researches the mining technique are providing great information and understanding level to human intellect. In this thesis, the proposed system is tools that have been developing through the core knowledge of association, priorities, grouping algorithm. Seperate analysis and it integrated form will enhance the earlier work in prediction of disease through history of patient records. Applying the A-Priori,K-Mean and PSO to the Proposed tools will enhance the capability of detection as well as the accuracy. In proposed tools a brain viewer is added in extra for analysis of brain and its peripheral area. These tools suggesting a patient on their basic diagnosis report that he or she has the chances of disease. It has been categorized in low, normal high, and damaged level .This is very helpful for patient who wants to know the possibility of some disease like heart, and lung. This is a predictive analysis just for research work. This technique could be apply for industrial application and may contain the suggestion box for patients also.
Key Word: A-Priori, K-Mean, PSO, MATLAB, Heart, Lung

I. INTRODUCTION

Data discovery in databases is fully defined method consisting of many distinct steps. Data mining is that the core step, which results within the discovery of hidden however helpful Data from large databases. Quality service implies diagnosing patients properly and administering treatments that are effective. Poor clinical decisions will lead to fatal consequences that are so unacceptable. Treatment records of millions of patients is stored and processed and data mining techniques might facilitate in responsive many vital and important queries and with health care.

1.1 Data Mining

A process used by corporations to show Data into helpful data. By using software package to look for patterns in Big batches of Data, businesses can learn a lot of concerning their customers and develop a lot of effective promoting ways likewise as increase sales and reduce prices. Data mining depends on effective Data assortment and storage likewise as pc process. Data mining is sorting through Data to spot patterns and establish relationships. Data mining parameters include:

A. Association - looking for patterns wherever one event is connected to a different event

B. Sequence or path analysis - looking for patterns wherever one event ends up in another later event

C. Classification - looking for new patterns (May end in a modification within the means the Data is organized however that is ok)

D. Clustering - finding and visually documenting teams of facts not antecedently known

E. Forecasting - discovering patterns in Data that will cause cheap predictions concerning the longer term (This space of Data mining is understood as prognostic analytics.)

1.2 History of Data Mining

The formal extraction of patterns from Data has occurred for centuries. Early methods of characteristic patterns in Data embody Bayes' theorem (1700s) and multivariate analysis (1800s). The proliferation, ubiquity and increasing power of pc technology has dramatically magnified Data assortment, storage, and manipulation ability. As data sets have full-grown in size and quality, direct "hands-on" data analysis has more and more been increased with indirect, automated Data process, aided by different discoveries in pc science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s), and support vector machines (1990s). Data mining is that the method of applying these ways with the intention of uncovering hidden patterns in Big Data sets. It bridges the gap from applied statistics and artificial intelligence (which usually offer the mathematical background) to management by exploiting the approach Data is keep and indexed in databases to execute the particular learning and discovery algorithms a lot of with efficiency, allowing such ways to be applied to ever larger Data sets.

1.3 K-Mean

K-Mean grouping is a generally utilized information bunching for unsupervised learning assignments. Here we demonstrate that essential segments are the nonstop answers for the discrete group enrollment pointers for K-Mean bunching. Equally, we demonstrate that the subspace spread over by the group centroids are given by otherworldly development of the information covariance network truncated at K−1 terms. These outcomes demonstrate that unsupervised measurement diminishment is firmly identified with unsupervised learning. K-Mean technique utilizes K models, the centroids of groups, to portray the information. They are dictated by minimizing the entirety of squared mistakes,

$$J_{\mathrm{K}} = \sum_{k=1}^{K} \sum_{i \in C_k} (\mathbf{x}_i - \mathbf{m}_k)^2$$

where (x_1, ・・・,x_n) = X is the data matrix and mk= ∑i*Ckx_i/$n_k$is the centroid of cluster Ckand nkis the number of points in Ck. Standard iterative solution to K-means suffers from a well-known problem: as iteration proceeds, the solutions are trapped in the local minima due to the greedy nature of the update algorithm.

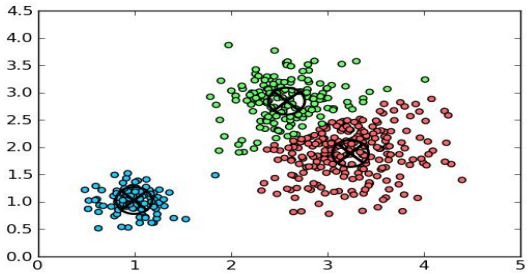

Fig1.1: Showing the K-mean technique

1.4 A-PRIORI

A-priori algorithmic algorithm for association is planned by R.Agarwal., in 1994. It finds out the relationships among item sets victimization a pair of inputs-support and confidence. one in every of the foremost common mining approaches is to hunt out frequent itemsets from a dealings dataset and derive association algorithms. Finding frequent itemsets is not trivial due to its combinate explosion.

1.5 PARTICLE SWARM OPTIMIZATION (PSO)

A big challenge facing in care organizations (hospitals, medical centers) is that the provision of fine quality services at terribly cheap costs. Quality service implies diagnosis patients properly and administering treatments that are effective. Poor clinical choices can cause fateful consequences that are therefore unacceptable. Hospitals should in addition minimize the value of clinical tests. they'll win these results by victimization applicable computer-based Data and/or decision support systems.

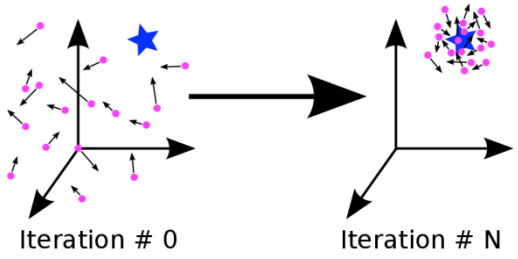

Fig 1.2 : Showing the optimization process under N iteration using PSO

II. LITERATURE REVIEW

[1] QeetharaKadhim Al-Shayea sugeested that the simulated neural system in infection diagnosis.Here two cases are examined. The main clarify that intense nephritis ailment information is the sickness side effects. The second is the coronary illness in which the information is on cardiovascular Single Proton Emission Computed Tomography (SPECT) pictures. Every patient characterized into two classifications: contaminated and non-tainted. Characterization is an essential apparatus in medicinal finding choice backing. Sustain forward back spread neural system is utilized as a classifier to recognize contaminated or non-tainted individual in both cases. The consequences of applying the manufactured neural systems philosophy to intense nephritis conclusion based upon chose indications show capacities of the system to take in the examples comparing to side effects of the individual. [2] Mohd Khalid Awang1 and FadzilahSiraj proposed to survey the utilization of simulated neural system in foreseeing the nearness of coronary illness, for the most part the angina in patients. The expectation and recognition of angina are critical in deciding the most proper type of treatment for these patients. The advancement of the application includes three primary stages. The primary stage is the advancement of Heart Disease Management Data System (HDMIS) for information gathering and patient administration. At that point took after by the second stage, which is the advancement of Neural Network Simulator (NNS) utilizing back engendering neural system for preparing and testing. The last stage is the improvement of Prediction System (PS) for expectation on new patient's information. [3] Irfan Y. Khan, P.H. Zope, S.R. Suralkar proposed to assess manufactured neural system [ANN] in infection conclusion. ANN's are regularly utilized as an effective separating classifier for undertakings in therapeutic analysis for early identification of sicknesses. ANN's are finding numerous utilizations in the medicinal determination application. Two cases are concentrated on. The first is intense nephritis malady; information is the illness indications. The second is the coronary illness. Information is on cardiovascular Single Proton Emission Computed Tomography (SPECT) pictures. Every patient characterized into two classes: contaminated and non-tainted. Grouping is an imperative instrument in restorative determination choice backing. Encourage forward back engendering neural system is utilized as a classifier to recognize contaminated or non-tainted individual in both cases. The analysis is then controlled by taking the entire accessible patients status into the record. At that point contingent upon that, a reasonable treatment is recommended, and the entire procedure may be iterated. In every emphasis, the analysis may be reconfigured, refined, or even rejected. [4] selvakumar.pdr.rajagopalan.s.p proposed the advantages and overhead of different neural system models for coronary illness expectation. The produced Data frameworks ordinarily comprise of vast measure of information. Social insurance associations must have capacity to break down these information. The Health care framework incorporates information, for example, asset administration, tolerant driven and changed information. Information mining procedures are utilized to investigate, break down and remove these information utilizing complex calculations as a part of request to find obscure examples. Numerous information mining procedures have been utilized as a part of the analysis of coronary illness with great exactness. Neural Networks have demonstrated extraordinary potential to be connected in the improvement of expectation framework for different kind of coronary illness. Information digging methods are utilized for Data disclosure as a part of databases by extraction of intriguing Data, for example, non-

trifling, covered up, already obscure, potential helpful and at last justifiable Data or examples from information's in substantial databases. Information digging gives diverse systems to basic leadership, critical thinking, investigation, arranging, finding, location, combination, aversion, learning and advancement and determining. [5] Parvathi I, SiddharthRautaray Computer Science, KIIT University proposed information mining as a rule (e.g. Definition, errands of information mining, use of information mining)and gives a brief synopsis of different information digging calculations utilized for characterization, bunching, and affiliation. Discourse is made to empower the malady conclusion and anticipation, and the disclosure of concealed biomedical and social insurance designs from related databases is offered alongside a talk of the utilization of information mining to find such connections as those between wellbeing conditions and a sickness, connections among ailments. It further talks about the device that can be utilized for the mining and grouping of information and the benefits of WEKA. the field of training information mining is massively utilized and is a developing field . As consistently a huge number of understudies are enlisted the nation over with Big number of advanced education applicants, we trust that information mining innovation can help spanning Data crevice in higher instructive frameworks. Information Mining distinguishes shrouded examples, affiliations, and oddities from instructive information and can enhance basic leadership forms in higher instructive frameworks. [6]HianChyeKoh and Gerald Tan proposed an information mining applications in medicinal services. Specifically, it talks about information mining and its applications inside medicinal services in real zones, for example, the assessment of treatment viability, administration of social insurance, client relationship administration, and the discovery of extortion and misuse. It additionally gives an illustrative case of a social insurance information mining application including the distinguishing proof of danger elements and with the onset of diabetes.

2.1 Problem Statement

This research had been implemented based on the basic need that require good and efficient way to detect and diagnose the disease. This will not be based on the assumption rather the data will be detecting the stage at which the disease is by using this particular system. Although it know the implementation of such system has not been taken place but by implementing such kind of patient monitoring system will allow the doctors to easily predict at what stage the disease is now in patient monitoring system. This is to provide an efficient and easy way to monitor.

III. PROPOSED WORK FOR A-PRIORI ,K-MEAN & PSO

K-Mean : It works for categorizing the group from which is could be decided that the where occurance of data lies.It means that if we have patient information through data then it is break into subgroup.The role of K-Mean is to form a cluster of data.The cluster is form of a near by value of derived sets that may be overlsped or non overlapped.

A-Priori: It has to calculate the effort of individual or commulative effort for making a breakthrough result.If on a scaling system it is to be decided that it need a threshold level then a priori will give the information that one or multiple cause will responsible for that level.

PSO: It has separate tool that will just give the optimized value of data input.It will consider both weight and direction of swarm.Weight is the fitness value of the data and direction is the possibilities of occurnace of the data for occurance second time or uptime.

Integrating K-Mean,A-Priori, & PSO for analysis of patient data will give a very accrate prediction.Now the proposed solution is as follows.

Step 1: Clustering Data Sets or categorizing the data set for patient data set and take a prediction for heart attributes.

Step 2: Making transaction chart for attributes and understanding the threshold value as follows in below table.Age will be differentiated in three group like A1,A2,and A3.Chest Pain will in Cp1 and Cp2.Blood pressure is in B1 and B2.Cholestrol level will be in two and last thal level in just a single category.

Step3: Making Attribute transaction status for 20 patients

Step4: Taking Intersection rule between the attributes and find out the maximum transaction sequence of given data.

Step 5: It will provide the lower and higher threshold value to data.

Set 1 Inter pair

$$\begin{cases} A_1 \cap A_2 = \emptyset, & A_1 \cap A_3 = \emptyset, \ A_2 \cap A_3 = \emptyset \\ Cp_1 \cap Cp_2 = \emptyset, & B_1 \cap B_2 = \emptyset, \ Ch_1 \cap Ch_2 = \emptyset \end{cases}$$

Set 2 Inter pairing

$$\begin{cases} A_1 \cap Cp_1 = \{T_1,T_5\}, & A_1 \cap Cp_2 = \{T_3,T_7\} \\ A_1 \cap B_1 = \{T_1,T_3,T_7\}, & A_1 \cap B_2 = \{T_5 \\ A_1 \cap Ch_1 = \{T_1,T_7\}, & A_1 \cap Ch_2 = \{T_3,T_5\} \end{cases}$$

$A_1 \cap Th_1 = \{T_7$

‘

‘

‘

‘

‘

$A_2 \cap Cp_1 = \emptyset, \quad A_2 \cap Cp_2 = \{T_2$

‘

‘

‘

‘

‘

$Ch_2 \cap Th_1 = \emptyset,$

Set 3 inter pairing of Attributes

$A_1 \cap A_2 \cap A_3 = \emptyset$

Set 4 Inter pairing

$A_1 \cap Cp_1 \cap B_1 = T_1$

$A_1 \cap Cp_1 \cap B_2 = T_5$

Increase the pairing until the last inter pairing. We get the transaction status of each patient.

Table 3.1: Showing the Transaction status

	A_1 A_2 A_3	$Cp1$ Cp_2	$B1$ B_2	Ch_1 Ch_2	Th_1
T_1	x	x	x	x	
T_2	x	x	x	x	x
T_3	x	x	x	x	
T_4	x	x	x	x	x
T_5	x	x	x	x	
T_6	x		x	x	
T_7	x	x	x	x	x
T_8	x	x	x	x	x
T_9	x	x	x	x	x
T_{10}	x	x	x	x	x

IV. RESULT & DISCUSSION

4.1 K-means, A-priori and PSO

With the proposed technique which is based on finding the value of critical or threshold level lead to develop a GUI through MATLAB. Which has two different blocks, One is representing a parameter inserting of new patient and second is independent for brain viewer and PSO optimization.

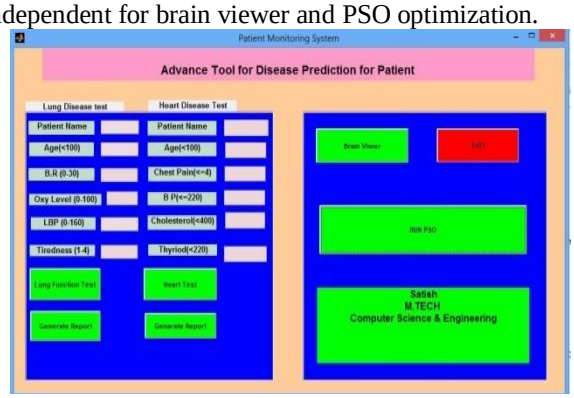

Fig 4.1: GUI of Proposed model

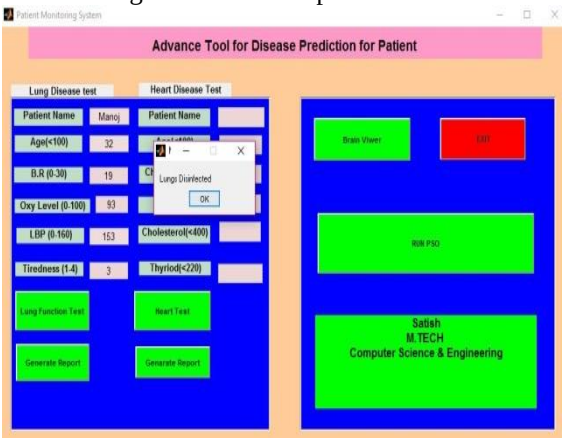

Fig 4.2: Testfiying the result after inserting the patient data(Lung-Disinfected)

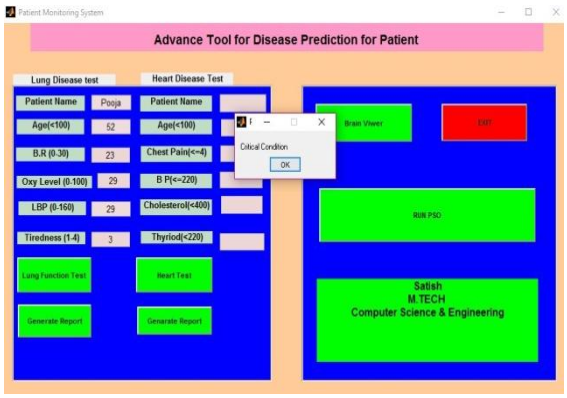

Fig 4.3: Testfiying the result after inserting the patient data(Lung-Critical)

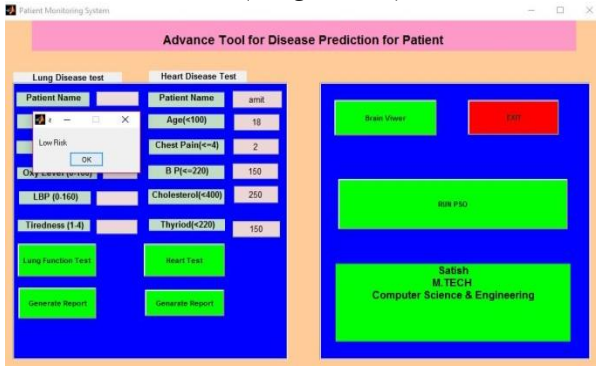

Fig4.4: Testfiying the result after inserting the patient data(Heart-Low Risk)

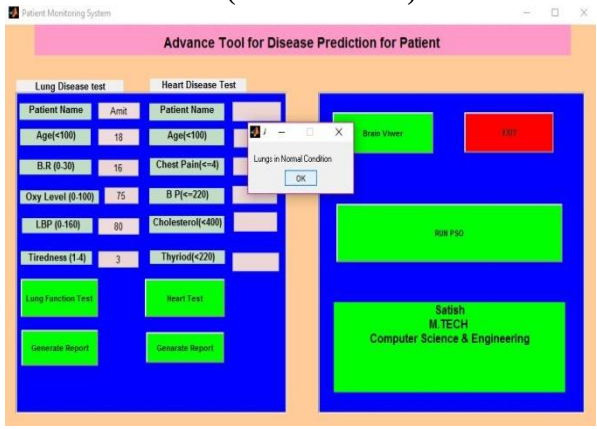

Fig 4.5 : Testfiying the result after inserting the patient data(Lung-Normal)

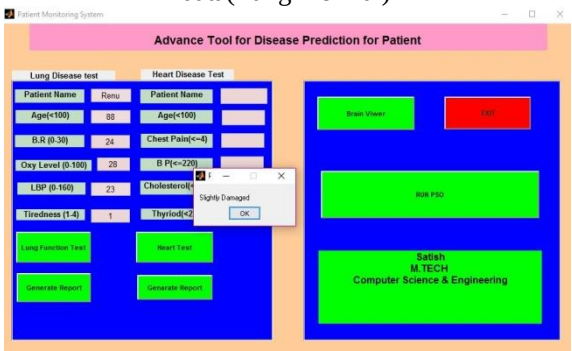

Fig 4.6: Testfiying the result after inserting the patient data(Lung-Slightly Damage)

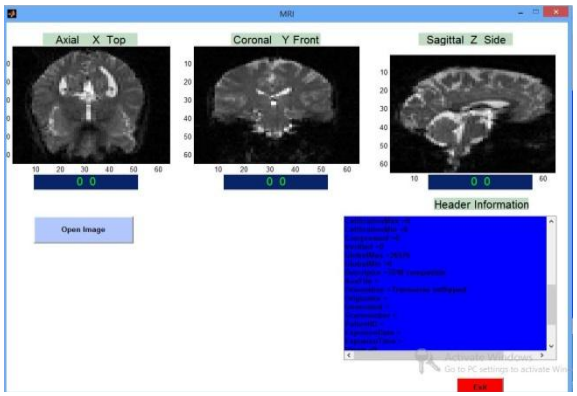

Fig4.7: Prefetching the image of brain for analysis of affected area

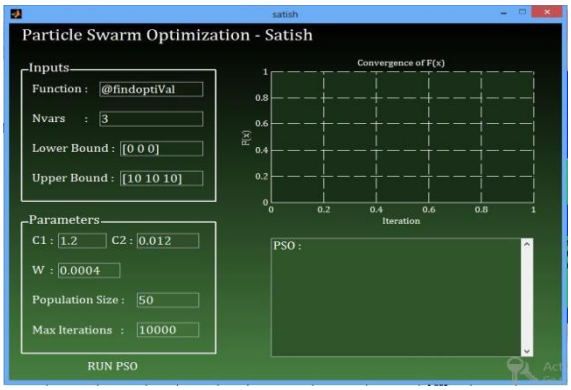

Fig 4.8 : Implementaing the data set for Optimazation (PSO)

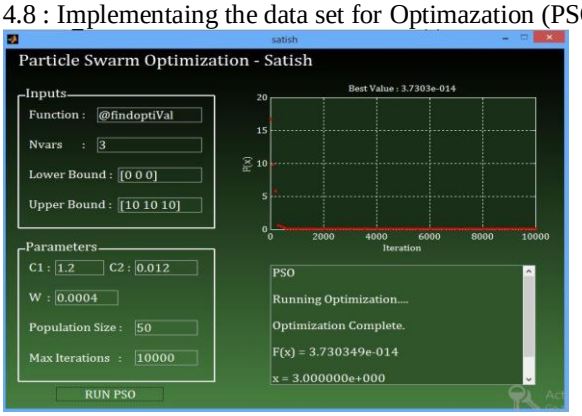

Fig 4.9 : Implementaing the data set for Optimazation (PSO)-Next Phase

Table 4.1 : Heart

Serial No.	Patient Name	Age	Chest Pain Level	Blood Pressure	Cholestrol Level	Thyriod Level	Actual Report	Our Model Test	Match
1	Amit	18	2	150	250	150	Low Risk	Low Risk	yes
2	Manoj	32	2	190	320	150	Low Risk	High Risk	No
3	Pooja	52	4	200	150	150	Medium Risk	Medium Risk	Yes
4	Ankita	88	4	200	300	180	Highest	Highest	yes
5	Stuti	24	2	100	280	160	Medium Risk	Medium Risk	yes
6	Anita	42	2	130	180	125	Low Risk	Low Risk	yes
7	Ankur	14	1	72	70	97	Low Risk	Low Risk	yes
8	Khushi	10	3	87	320	82	Medium Risk	Medium Risk	yes
9	Jaya	49	3	180	380	197	High Risk	High Risk	yes
10	Mohit	78	1	120	200	110	Medium Risk	Medium Risk	yes
11	Deepak	26	4	175	320	164	High Risk	High Risk	yes
12	Vimal	34	2	167	256	156	Medium Risk	Medium Risk	yes
13	Preeti	57	2	173	312	162	High Risk	High Risk	yes
14	Renu	88	3	152	297	141	Medium Risk	Medium Risk	yes
15	Monika	19	1	86	184	75	Low Risk	Low Risk	yes

Table 4.2 : Lung

Serial No.	Patient Name	Age	Breath Rate	Oxygen Level	L.B.P Level	Tiredness Level	Actual Report	Our Model Test	Match
1	Amit	18	16	75	80	3	Normal	Normal	yes
2	Manoj	32	19	93	153	3	Disinfected	Disinfected	Yes
3	Pooja	52	23	29	29	3	Critical	Critical	Yes
4	Ankita	88	20	75	56	1	Normal	Normal	yes
5	Stuti	24	29	62	120	2	Normal	Normal	yes
6	Anita	42	25	54	87	1	Normal	Normal	yes
7	Ankur	14	10	36	151	2	Normal	Normal	yes
8	Khushi	10	12	84	154	3	Disinfected	Disinfected	yes
9	Jaya	49	15	79	53	1	Disinfected	Disinfected	yes
10	Mohit	78	18	57	73	4	Normal	Normal	yes
11	Deepak	26	12	88	130	3	Disinfected	Disinfected	yes
12	Vimal	34	18	95	142	4	Disinfected	Disinfected	yes
13	Preeti	57	19	27	56	2	Normal	Normal	yes
14	Renu	88	24	28	23	1	Slightly Damaged	Slightly Damaged	yes
15	Monika	19	26	28	25	3	Critical	Critical	yes

4.2 Result derives from Proposed Work

As the above result shown by chart it is very clear that the outcome is very accurate and if this will go for a larger data set. This proposed technique will give quite appreciable results. Even a brain viewer will provide 3D model of HDR image which has shown a predictive analysis over the human brain area. And the PSO is different and optimized value of dataset.

V. CONCLUSION

Integrating Clustering technique,Association rule and optimization technique based graphical user interface will provide very easy handling user interface. The patient itself insert the their diagnosis report and the prediction come out. The result will be tested on more than 20 actual patients for verifying the proposed technique. This find a very good prediction. Almost 19 out of 20 accurate prediction provides sufficient accuracy graph of the proposed technique.So integrating the A-priori, PSO, and K-Mean Technique enhancing the earlier work. In this thesis, the proposed system is tools that have been developing through the core knowledge of association, priorities, grouping Algorithm. Seperate analysis and it integrated form will enhance the earlier work in prediction of disease through history of patient records. Applying the A-Priori, K-Mean and PSO to the Proposed tools will enhance the capability of detection as well as the accuracy.

REFERENCE

[1] Qeetharakadhim al-shayea mis department, al zaytoonah university of jordan amman, jordan "artificial neural networks in medical diagnosis" ijcsi international journal of computer science issues, vol. 8, issue 2, march 2011 issn (online): 1694-0814.

[2] Mohd khalid awang1 and fadzilahsirajfaculty of informatics, university of sultan zainalabidin, gong badak campus, 21300 kuala terengganu, terengganu, malaysia "utilization of an artificial neural network in the prediction of heart disease" international journal of bio-science and bio-technology vol. 5, no. 4, august, 2013

[3] Irfan y. Khan, p.h. zope, s.r. suralkar dept. Of ele. & tele. Ssbt's college of engg. & tech, bambhori, jalgaon "importance of artificial neural network in medical diagnosis disease like acute nephritis

disease and heart disease" issn: 2319-5967 iso 9001:2008 certified international journal of engineering science and innovative technology (ijesit) volume 2, issue 2, march 2013

[4] Selvakumar.p, dr.rajagopalan.s.p assistant professor, arm college of engineering and technology, chennai tamil nadu, professor, gkm college of engineering and technology "a survey on neural network models for heart disease prediction" journal of theoretical and applied Data technology 20th september 2014. Vol. 67 no.2 © 2005 - 2014 jatit & lls. All rights reserved.issn: 1992-8645 www.jatit.org e-issn: 1817-3195

[5] Parvathi i, siddharthrautaraycomputer science, kiit university bhubaneswar,odisha,india "survey on data mining techniques for the diagnosis of diseases in medical domain" parvathi i, et al, / (ijcsit) international journal of computer science and Data technologies, vol. 5 (1) , 2014, 838-846.

[6] Hianchyekoh and gerald tan "data mining applications in healthcare" journal of healthcare Data management,vol. 19, no. 2, 2008.

[7] Neelamadhabpadhy, asst. Professor, gandhi institute of engineering and technology, giet, gunupur dept. Of computer science, dr. Pragnyaban mishra cmj university, meghalaya (silong) a research scholar an professor, gandhi institute of engineering and technology, giet, gunupurrasmitapanigrahi lecturer, gandhi institute of engineering and technology, giet, gunupur "the survey of data mining applications and feature scope" international journal of computer science, engineering and Data technology (ijcseit), vol.2, no.3, june 2012

[8] B.venkatalakshmi, m.v shivsankar tifac-core, pervasive computing technologies, velammal engineering college, chennai "heart disease diagnosis using predictive data mining" k.l.n. college of engineering, madurai, tamil nadu, indiaissn (online) : 2319 – 8753 issn (print) : 2347 – 6710 international journal of innovative research in science, engineering and technology volume 3, special issue 3, march 2014 international conference on innovations in engineering and technology

[9] Thirumal p. C. And nagarajan n. Department of it, coimbatore institute of engineering and technology, coimbatore, tamilnadu "utilization of data mining techniques for diagnosis of diabetes mellitus - a case study" vol. 10, no. 1, january 2015 issn 1819-6608 arpn journal of engineering and applied sciences ©2006-2015 asian research publishing network (arpn). All rights reserved.

[10] M.durairaj, v. Ranjani "data mining applications in healthcare sector: a study" international journal of scientific & technology research volume 2, issue 10, october 2013 ISSN 2277-8616 29 ijstr©2013.

