Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 6, Issue 8, 2014

Original Article

ENDORSEMENT OF SMALL PATIENTS POPULATION STUDY THROUGH DATA MINING CLASSIFICATION: SIGNIFICANCE TO MANIFEST DRUG INTERACTION STUDY OF CARDIOVASCULAR DOSAGE FORMULATION

RAKESH DAS, SUBHASIS DAN, TAPAN KUMAR PAL*

Bioequivalence Study Center, Dept. of Pharmaceutical Technology, Jadavpur University, Jadavpur, Kolkata, W.B. India
Email: tkpal12@gmail.com

Received: 19 Jun 2014 Revised and Accepted: 22 Jul 2014

ABSTRACT

Objective: A simple, sensitive, precise computational classifiers justifies the positive indication of drug interaction through statistical validation and confirms for further root level investigation.

Methods: The blood pressure (BP) & Lipid profile valued data sheet was prepared from 100 patients those were chronically treating with cardiovascular formulation consisting Atorvastatin 10mg + Olmesartan 20mg. The data sheet contains 100 patients with 10 variables and final decision attributes of working & non-working. Then, with the operation of seven different related classifier the details of % of accuracy by class, correct & incorrect classified instance and stratified cross- validation were estimated. Those statistical results of classifiers were compared, correlate and interpreted to bring a fixed conclusion based on it.

Results: The % of accuracy for all classifiers results commonly 95.9596 %, 93.9394 % and 96.9697 % and inter-depending class attributes denoting by a = NW & b =W Matrix values are 84 | 11, 84 | 9, 87 | 9 respectively . Thus, the accuracy is excellent covering within the limits of (±15%) as a correct classified instant.

Conclusion: Statistical computation on less populated patients through classifiers, evidentially confirms the drug-interaction profile of collected data through data mining process. So that, it can proceeds further upto root level through instrumental bioanalysis.

Keywords: Seven Classifiers, computatIonal statistical analysis, Physio-chemical patients data, Data mining process.

INTRODUCTION

Rapid inter-collaboration between Clinical pharmacy researchers working in Cardiovascular therapy and computer scientists are looking at the application of data mining techniques to the area of individual patients diagnosis, based on the clinical records. An investigation of seven different classification models on cardiovascular data for estimation of patient risk in cardiovascular domains is presented [1-7].

A major Challenge on healthcare sectors is the provision of quality services at affordable cost. Quality service implies diagnosing patients exactly and providing treatments that are cost-effective. Bad clinical decision can lead to negative consequences which are therefore wasteful. They can get onto these results by implicating appropriate computer-based information and /or decision support systems [8]. One in eight women over their lifetime has a risk of developing breast cancer. An analysis of the most recent data has shown that the survival rate is 88% after 5 years of diagnosis and 80% after 10 years of diagnosis [9]. The nature of a population can be seemed to establish the reasons associated with a specific endpoints. Prospective studies, such as statistical learning and data mining, can approach the association of the variables to the outcome, but were not always establish the cause-and-effect relationship of the association. Data holding statistical research is becoming a common breeze to many scientific areas like medicine and biotechnology. This trend commonly observed as in the studies of Houston et al. and Cios et al. [10-11]. A literature survey convey several studies on the survivability prediction problem using statistical approaches and artificial neural networks. However, we observe few studies related to medical diagnosis and survivability using data mining approaches like decision trees [12-14]. Other than the breadth of stored information, which gradually includes long-term outcome and associated biological and genetic data, mining for potentially novel and useful biomedical associations in Clinical Data Repositories (CDRs) is a relatively recent approach [15-18]. Research has dual effect: to develop clinical participated databases of cancer patients, and to conduct data mining and learning studies on collected patient records [19]. Statistics and data mining differ in the use of machine learning methods, the volume of data, and the role of computational complexity. Requirement for analysis is preceding our abilities to handle the complexity. Preprocessing is much vital with large datasets, especially as we reaches the pentabyte level. However, data mining is concentrate on the data mining process itself with little traces on the knowledge actually extracted[20-22]. Cardiovascular decision-making support experiences increasing research interest of scientists. Simultaneous collaborations between clinical pharmacy researchers and computer scientists are focusing at the implication of data mining techniques to the area of individual patient diagnosis, based on clinical datas, Bayesian network [23].

MATERIALS AND METHODS

Study Design

On evolving negative therapeutic information for Cardiovascular combined drug formulation of Atorvastatin (10mg) & Olmesartan (20mg) from patient's Clinical history, doctors comments, mortility & morbidity frequencies, clinical study, which was operated on Midinapur Medical college and Hospital under clinical supervisor Dr. Balaram Ghosh. The prospective observational study were conducted among the 100 patients those were under chronic therapy of that formulation. The basic data were collected on the basis of Blood pressure (BP), Lipid profiles (HDL, LDL, VLDL, Triglycerides, Total Cholesterol, Total- Cholest/ HDL, LDL-C/HDL-C levels) and there hypothetical correlative resulting data, Working (W) & Non-working (NW) from patients physiological and biochemistry data chart reports.

Data assemblance

Systolic BP & Diastolic BP of 100 patients were assembled in successively in 1st and 2nd column. The lipid profile for those corresponding 100 patients were spilt to HDL, LDL, VLDL, Triglyceride in chronicle manners in columns, followed by a depending column of Total Cholesterol. The 3rd last and 2nd last

column assists the functional ratio of T-Cholesterol/HDL & LDL-C/HDL-C respectively. Also, in final column attributes are domain with working (W) and Non- working (NW). This patients data sheet were included in **Table- 1.**

Table 1: The medical data sheet exhibiting BP & Lipid profile collected after chronic administration of CVD formulation of 100 patients.

Systolic BP	Diastolic BP	HDL	LDL	VLDL	Triglyceride	T.Chol	T.Chol/HDL	LDL-C/HDL-C	class
172	98	39	143	43	212	198	5.09	3.66	N.W
176	123	36	162	46	231	215	5.97	4.5	N.W
169	136	40	169	42	216	223	5.57	4.22	N.W
173	125	39	155	46	198	234	6	3.97	N.W
168	138	40	159	43	219	197	4.92	3.97	N.W
170	141	36	162	48	223	231	6.41	4.5	N.W
173	139	32	159	46	241	224	7	4.96	N.W
189	141	37	160	42	214	231	6.24	4.32	N.W
172	138	41	163	41	209	214	5.21	3.97	N.W
168	132	35	141	46	213	235	6.71	4.02	N.W
186	134	59	143	46	200	237	4.01	2.42	N.W
180	121	32	157	43	212	224	7	4.9	N.W
167	140	36	146	45	198	232	6.44	4	N.W
182	156	63	144	41	159	214	3.39	2.28	N.W
168	141	36	172	48	172	242	6.72	4.77	N.W
165	100	44	164	42	198	232	5.27	3.72	N.W
177	126	23	139	46	199	219	9.52	6.04	N.W
152	135	69	163	47	222	221	3.2	2.36	N.W
163	132	71	167	42	256	251	3.53	2.35	N.W
178	133	37	143	48	159	236	6.37	3.86	N.W
173	145	43	129	50	168	228	5.3	3	N.W
188	134	41	161	53	183	194	4.73	3.92	N.W
169	123	79	158	42	199	233	2.94	2	N.W
154	119	38	159	51	198	246	6.47	4.18	N.W
183	141	36	165	46	189	213	5.91	4.58	N.W
174	141	69	157	47	231	239	3.46	2.27	N.W
167	134	72	147	43	212	227	3.15	2.04	N.W
124	102	29	112	27	156	195	6.72	3.86	W
183	143	37	134	46	126	249	6.72	3.62	N.W
168	135	35	167	48	147	226	6.45	4.77	N.W
170	142	32	163	44	198	245	7.65	5.09	N.W
189	150	33	158	72	242	243	7.36	4.78	N.W
202	153	33	174	58	271	225	6.81	5.27	N.W
154	96	41	159	54	233	245	5.97	3.87	N.W
164	136	47	169	61	213	236	5.02	3.59	N.W
187	143	40	163	45	243	235	5.87	4.07	N.W
178	126	39	145	43	232	237	6.07	3.71	N.W
167	129	78	160	47	251	231	2.96	2.05	N.W
153	142	29	147	53	202	221	7.62	5.06	N.W
131	123	33	102	30	123	198	6	3.09	N.W
127	89	42	100	27	134	200	4.76	2.38	N.W
112	91	62	95	25	128	187	3.01	1.53	W
190	141	69	170	61	198	236	3.42	2.46	N.W
162	132	60	145	46	231	244	4.06	2.41	N.W
176	132	28	163	49	231	238	8.5	5.82	N.W
168	121	36	156	44	212	237	6.58	4.33	N.W
171	124	38	154	39	189	247	6.5	4.05	N.W
166	112	35	155	38	199	239	6.82	4.42	N.W
182	142	45	170	51	213	235	5.22	3.77	N.W
175	145	42	182	46	215	231	5.5	4.33	N.W
159	120	34	156	57	232	234	6.88	4.58	N.W
169	112	40	149	59	230	241	6.02	3.72	N.W
182	131	72	157	65	219	234	3.25	2.18	N.W
117	82	37	98	32	141	197	5.32	2.64	W
174	147	39	145	63	199	216	5.53	3.71	N.W
165	125	43	134	57	250	234	5.44	3.11	N.W
184	136	47	147	48	262	267	5.68	3.12	N.W
155	114	37	134	52	202	241	6.51	3.62	N.W
162	118	49	139	38	210	242	4.93	2.83	N.W
139	100	68	154	58	223	211	3.1	2.26	N.W
173	152	38	137	49	232	239	6.28	3.6	N.W
120	78	45	98	28	135	187	4.15	2.17	W
181	135	40	126	39	209	242	6.05	3.15	N.W
118	85	32	100	25	129	176	5.5	3.12	W
165	140	38	145	58	231	241	6.34	3.81	N.W
168	131	23	154	55	214	227	9.86	6.69	N.W
174	127	25	171	62	235	238	9.52	6.84	N.W

186	132	39	137	48	223	229	5.87	3.51	N.W
159	89	55	148	67	231	235	4.27	2.69	N.W
165	152	21	187	73	235	223	10.61	8.9	N.W
158	102	70	169	39	227	242	3.45	2.41	N.W
168	115	50	135	54	222	225	4.5	2.7	N.W
174	91	20	165	46	198	237	11.85	8.25	N.W
176	119	42	155	48	199	226	5.38	3.69	N.W
179	138	24	143	54	216	256	10.66	5.95	N.W
142	111	21	138	81	189	249	11.85	6.57	N.W
197	152	47	154	46	217	265	5.63	3.27	N.W
160	129	49	149	39	208	243	4.95	3.04	N.W
123	82	46	101	32	142	179	3.89	2.19	W
115	74	53	98	21	126	181	3.41	1.84	W
127	68	22	96	26	137	189	8.59	4.36	W
176	124	23	156	56	214	265	11.52	6.78	N.W
178	131	22	147	47	207	246	11.18	6.68	N.W
183	115	47	154	54	200	243	5.17	3.27	N.W
155	103	76	147	38	213	233	3.06	1.93	N.W
171	131	65	157	68	215	255	3.92	2.41	N.W
116	74	48	153	54	142	188	3.91	3.18	N.W
121	78	37	98	30	138	185	5	2.64	W
172	116	20	158	45	215	232	11.6	7.9	N.W
189	149	58	154	76	223	265	4.56	2.65	N.W
178	126	72	148	69	200	234	3.37	2.05	N.W
125	80	54	100	23	157	190	3.51	1.85	W
181	152	23	158	46	214	216	9.39	6.86	N.W
125	78	19	92	19	148	194	10.21	4.84	N.W
162	112	24	154	54	231	235	9.79	6.41	N.W
152	100	75	146	48	225	247	3.29	1.94	N.W
187	102	43	148	54	218	243	5.65	3.44	N.W
123	78	43	99	24	132	198	4.6	2.3	W
120	79	56	89	26	136	202	3.6	1.58	W

Foot Notes: Blood Pressure, BP; High Density Lipoprotein, HDL; Low Density Lipoprotein, LDL; Very Low Density Lipoprotein, VLDL; Cholesterol, C; Total Cholesterol, T.Chol; Not working, NW; Working, W; Cardiovascular drug, CVD.

Computational Analysis

Computational analysis were carried out through various important workable and relating classifiers *i.e, a) NaiveBayes; b) SMO; c) Lazy.KStar Beta-version; d) Meta. adaBoostM1; e) Meta. Bagging; f) rules. PART; and g) Tress. J48*, to understand the percentage (%) of accuracy analysis in those small population of 100 patients. Based on results to these analytical justification through those classifiers, the further permission and requirement of drug interaction study proceedings could be determined.

Accuracy evaluation

The weka software is most supportive to compute data sheet variables. Each variables of all the patients were statistically developed on respect to Mean, Std. dev., weight sum, & precision. So that according to each & every seven classifier, accuracy could be traced out. These accuracy and inexactness ratio of all well reputed classifiers reflects its signified conclusion after correlating. The standard statistical reports would aid on making decision to prepare valid reason to start investigation even on lower populated patients data.

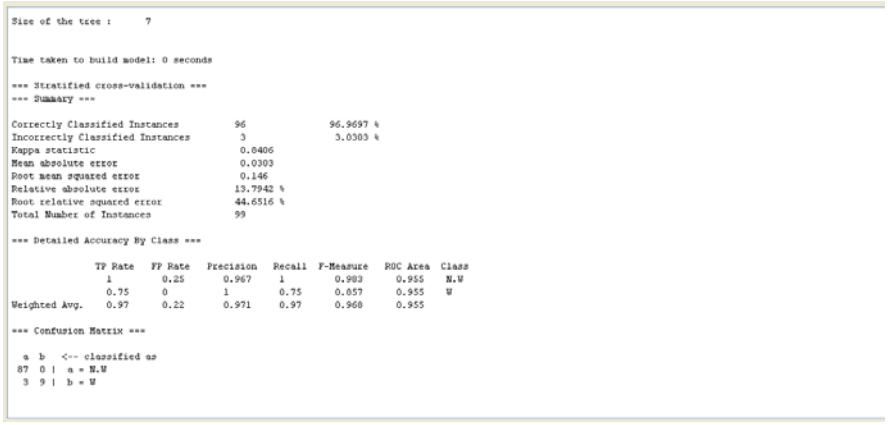
Recruitment of based Analysis

The analysis of bio-analytes (peptides, enzymes & other biochemical product) and analytes inter-related to investigation content is required to access drug-interactions of patients out of 100 enrolled in data sheet.

RESULTS

The biochemical and physiological variation marked after the cardiovascular drug therapy among all 100 patients, were statistically analyse for the accuracy using computational process of data mining, **Figure-1**.

The computational statistical complete analysis were carried out with classifiers- Naive Bayes Classifier, SMO, KStar Beta Verion (0.1b), AdaBoost, bagging, rules. PART; and Trees. J48.



```
Size of the tree :      7


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          96               96.9697 %
Incorrectly Classified Instances         3                3.0303 %
Kappa statistic                          0.8406
Mean absolute error                      0.0303
Root mean squared error                  0.146
Relative absolute error                 13.7942 %
Root relative squared error             44.6516 %
Total Number of Instances               99

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.25       0.967     1         0.983      0.955    N.W
                 0.75      0          1         0.75      0.857      0.955    W
Weighted Avg.    0.97      0.22       0.971     0.97      0.968      0.955

=== Confusion Matrix ===

  a  b   <-- classified as
 87  0 |  a = N.W
  3  9 |  b = W
```

Fig. 1: Classification of accuracy with best data mining tool classifiers.

Table 2: Statistical results of all correctly classified (N.W) and incorrectly classified (N) attribute for BP and Lipid profile analysis as per model classifiers.

Class	Attribute	Attribute
	N.W	W
	(0.87)	(0.13)
Systolic BP		
mean	169.6552	120.45
std. dev.	15.0777	4.1866
weight sum	87	12
precision	1.8	1.8
Diastolic BP		
mean	127.0179	81.6028
std. dev.	17.6302	8.0336
weight sum	87	12
precision	1.8723	1.8723
HDL		
mean	43.0135	42.9348
std. dev.	15.5519	11.5989
weight sum	87	12
precision	1.3043	1.3043
LDL		
mean	152.2943	98.5444
std. dev.	15.4126	4.715
weight sum	87	12
precision	2.1778	2.1778
VLDL		
mean	49.4679	26.7154
std. dev.	10.1452	3.2276
weight sum	87	12
precision	1.5122	1.5122
Triglyceride		
mean	209.4661	137.761
std. dev.	28.169	9.5607
weight sum	87	12
precision	2.7925	2.7925
T.Cholesterol		
mean	232.4631	188.9643
std. dev.	15.9614	7.8472
weight sum	87	12
precision	1.8571	1.8571
T.Chol/HDL		
mean	6.1322	4.772
std. dev.	2.2604	1.5425
weight sum	87	12
precision	0.1001	0.1001
LDL-C/HDL-C		
mean	3.996	2.5091
std. dev.	1.5216	0.8554
weight sum	87	12
precision	0.0899	0.0899

Results of the cross-validation were justified statistically under Naive Bayes, SMO, AdaBoost, Bagging Classifiers. The classified instant which carried 95.95% of correctly and 4.04% of incorrect accuracies. Other than that Naive Bayes has kappa statistic, Mean absolute error, root mean squared error, relative absolute error and Root relative squared error are 0.8231, 0.0398, 0.1983, 18.13%, 60.62% respectively. And the population of instances on patients investigation was performed out-off 100. Also, the not working (NW) & working (W) classes were expressed in 84 | 3 | and 1 | 11 in matrix calculation respectively for Naïve Bayes classifier. According to SMO classifier, number of kernel evaluations are 322 (69.594% cached). Statistical calculations of SMO; AdaBoost; Bagging classifiers upon kappa statistic, Mean absolute error, root mean squared error, relative absolute error and Root relative squared error are exhibits 0.8242, 0.0404, 0.201, 18.39%, 61.46%; 0.8231, 0.043, 0.2015, 19.57%, 61.61%; 0.8231, 0.0813, 0.1925, 36.9984%, 58.8542% respectively. Not working (NW) & working (W) classes were expressed in 84 | 3 | and 0 | 12 in matrix calculation respectively for SMO classifier. And not working (NW) & working (W) classes were expressed in |84| and 1 | 11 in matrix calculation respectively for AdaBoost & Bagging classifiers.

The cross-validation were justified statistically under Kenstar beta version classifier, were The classified instant which carried 93.93% of correctly and 6.06% of incorrect accuracies. Statistical evaluation of Kenstar beta version classifier lodges kappa statistic, Mean absolute error, root mean squared error, relative absolute error and Root relative squared error are represents- 0.7155, 0.0584, 0.2339, 26.599%, 71.51%. And not working (NW) & working (W) classes were expressed in 84 | 3 | and 3 | 9 in matrix calculation respectively for KenSTAR Beta version classifier.

The cross-validation were justified statistically under Rules PART & Trees J48 classifier, were The classified instant which carried 96.9697% of correctly and 3.0303% of incorrect accuracies.

Statistical evaluation of Rules PART & Trees J48 classifiers represents kappa statistic, Mean absolute error, root mean squared error, relative absolute error and Root relative squared error are represents- 0.8406, 0.0303, 0.146, 13.7942%, 44.65% & 0.8406, 0.0303, 0.146, 13.7942%, 44.65% respectively. And not working (NW) & working (W) classes were expressed in 87 | 0 | and 3 | 9 in matrix calculation respectively for Rules PART & Trees J48 classifier.

Table 3: Detailed accuracy by class of all 100 patients according to their respective attributes and weighted averages of each status respect to Classifiers.

Classifiers	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
Naive Bayes	0.966	0.083	0.988	0.966	0.977	0.993	N.W
	0.917	0.034	0.786	0.917	0.846	0.993	W
	0.96	0.077	0.964	0.96	0.961	0.993	wt. avg.
SMO	0.954	0	1	0.954	0.976	0.977	N.W
	1	0.046	0.75	1	0.857	0.977	W
	0.96	0.006	0.97	0.96	0.962	0.977	wt.avg.
KenSTAR Beta version	0.966	0.25	0.966	0.966	0.966	0.971	N.W
	0.75	0.034	0.75	0.75	0.75	0.971	W
	0.939	0.224	0.939	0.939	0.939	0.971	Wt.avg.
Meta AdaBoost	0.966	0.083	0.988	0.966	0.977	0.969	N.W
	0.917	0.034	0.786	0.917	0.846	0.969	W
	0.96	0.077	0.964	0.96	0.961	0.969	Wt.avg.
Meta Bagging	0.966	0.083	0.988	0.966	0.977	0.97	N.W
	0.917	0.034	0.786	0.917	0.846	0.97	W
	0.96	0.077	0.964	0.96	0.961	0.97	Wt.avg.
Rules PART	1	0.25	0.967	1	0.983	0.955	N.W
	0.75	0	1	0.75	0.857	0.955	W
	0.97	0.22	0.971	0.97	0.968	0.955	Wt.avg.
Trees J48	1	0.25	0.967	1	0.983	0.955	N.W
	0.75	0	1	0.75	0.857	0.955	W
	0.97	0.22	0.971	0.97	0.968	0.955	Wt.avg.

DISCUSSION

On stratified cross-validation after 0.02 sec build model of classifiers - Naive Bayes Classifier, SMO, KStar Beta Verion (0.1b), AdaBoost, bagging, rules. PART; and Trees. J48 represents % of accuracy of 95.9596 %, 95.9596 %, 93.9394 %, 95.9596 %, 95.9596 %, 96.9697 %, 96.9697 % respectively. Also, the incorrect classified instants/ inexactness evolved 4.0404 %, 4.0404 %, 6.0606 %, 4.0404 %, 4.0404 %, 3.0303 %, 3.0303 % respectively. Thus, the accuracy is excellent covering within the limits of (±15%) as a correct classified instant. And depending on the correct classified instant 95, 93, 96 the class attributes denoting by a = NW & b =W Matrix values are 84 | 11, 84 | 9, 87 | 9 respectively. On briefly undergoing through the Kappa Statistics, Mean absolute errors, Root mean squared error, Relative absolute error, root relative square error; and also understanding the correct & incorrect classified instant, it was discovered that classifiers statistical results and values are nearly same.

ABBREVIATIONS

ROC area- Receiver Operating Characteristics; TP-rate- True Positive rate; FP-rate- False Positive rate; F-measure- F-score in F-test measurement; W-working; NW-Not working.

CONCLUSION

The accuracy details, stratified cross-validations and (a,b) matrix of class attributes represents that the statistical results of all the seven classifiers are moreover equal. And its conclude, that data mining application could evaluate and confirms the perfectness of drug interaction even from small populated patients with detail medical history data. And thus, it signifies the positive indication to start investigation further study to get the concrete explanation.

CONFLICT OF INTEREST

We authors are declaring no conflict of interest.

ACKNOWLEDMENT

I myself Mr. Rakesh Das, as a main author of research project funded by All India Council of Technical Education (A.I.C.T.E.) and author Shubhasis Dan is acknowledging University Grants Commission (UGC) for financial support through UGC-BSR fellowship scheme. We authors were extremely thankful to Dr. Balaram Ghosh, MD, Clinical Pharmacologist, Midinapore Medical college and Hospital, West Bengal, to supervising & cooperating to collect medical data reports. And also, I am thankful to patients and patient's party to allow me for performing uninterrupted research study. I am extremely thankful to faculties of computer Sc. & Engg. for technical support.

REFERENCES

1. Baesens B, Egmont-Petersen M, Castelo R, Vanthienen J. Learning Bayesian network classifier for credit scoring using Markov chain Monte Carlo serach Proceedings. Int Congress on Pattern Recognition 2002;3:49-52.
2. Brent M. Instance-Based learning:Nearest neighbor with generalization. Master's thesis at the university of Waikato, New Zealand;1995. p. 1-76.
3. Davis DN, Nguyen TT. Generating and verifying risk prediction models using data mining:A case studyfrom Cardiovascular medicine. Chapter of data mining and medical knowledge management:Cases and applications, ISBN10:1605662186. J IGI Global Inc 2009.
4. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. J Al Magazine 1996;17(3):37-45.
5. Garofalakis M, Hyun D., Rastogi R, Shim K. Building decision trees with constraints. J Data Mining and Knowledge Discovery 2003;7(2):187-214.
6. Mitchell T M. Machine learning. Mc Graw-Hill Companies. In USA 1997;414.
7. Nilson NJ. Introduction to machine learning. Unpublished draft;In Standford University, USA, 1996.
8. Palaniappan S, Awang R. Intelligent heart Disease prediction system using data mining Techniques. Int J of Computer Sc and Network Security 2008;8(8):343-50.
9. American Cancer Society. Breast Cancer Facts & Figures 2005-2006. Atlanta:American Cancer Society. J Inc
10. Houston, Andrea L. and Chen, *et al*. Medical Data Mining on the Internet:Research on a Cancer Information System. J Artificial Intelligence Rev 1999;13:437-66.
11. Cios KJ, Moore GW. Uniqueness of medical data mining. J Artificial Intelligence in Medicine 2002;26:1-24.
12. Zhou ZH, Jiang Y. Medical diagnosis with C4.5Rule preceded by artificial neural network ensemble. J IEEE Trans Inf Technol Biomed 2003;7(1):37-42.
13. Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. J Oncology 1999;57: 281-6.
14. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. J Artificial Intelligence in Medicine 2005;34(2):113-27.
15. Holmes JH, Durbin DR, Winston FK. Discovery of predictive models in an injury surveillance database: An application of

data mining in clinical research. J Proc AMIA Symp 2000;359-63.

16. Downs SM, Wallace MY. Mining Association rules from a pediatric primary care decision support system. J Proc AMIA Symp 2000;200-04.
17. Brossette SE, Sprague AP, Hardin JM, Waites KB, Jones WT, Moser SA. Association rules and data mining in hospital infection control and public health surveillance. J Am Med Inform Assoc 1998;5,:373-81.
18. Prather JC, Lobach DF, Goodwin LK, Hales LK, Hage ML, Hammond WE. Medical data mining: Knowledge discovery in a clinical data warehouse. Proc AMIA Symp 1997;101-05.
19. John Hayward. Mining Oncology Data: Knowledge Discovery in Clinical Performance of Cancer Patients. A Thesis submitted to Worcester Polytechnical Institute, Aug. 2006, MA 01609, United States.
20. Hosking JR, Pednault EP, Sudan M. Statistical perspective on data mining. J Future Generaltion Computer System. 1997;13(3):117-34.
21. Keim DA, Mansmann F, Schneidewind J, Ziegler H. Challenges in visual data analysis.
22. *Information Visualization.* DOI:10.1109/IV.2006.31, 10th Int Conference 2006;9-16.
23. Mannila H. Data mining: machine learning, statistics and databases. Paper presented at:8th J Int Conference on Scientific and Statistical Database Systems 1996.
24. Bohacik J, Darryl ND. Estimation of cardiovascular patient risk with a Bayesian network. J Transcom 2011;27:129-32.