

DIAGNOSING HYPERLIPIDEMIA USING ASSOCIATION RULES

Sengul Dogan¹ and Ibrahim Turkoglu²

¹ Department of Bioengineering, Firat University, Elazig, Turkey

²Department of Electronics and Computer Science, Firat University, Elazig, Turkey
vsenguldogan@hotmail.com, iturkoglu@firat.edu.tr

Abstract- Data mining methodologies have been developed for exploration and analysis of large quantities of data to discover meaningful patterns and rules. This paper presents a new approach, that employs data mining, to find association rules an effective method for discovering Hyperlipidemia. The propose system has been projected from the biochemistry blood parameters which will be very helpful for and will make everything easier for the physicians in the diagnosis of Hyperlipidemia. The basic characteristic of the lipide parameters that is Total cholesterol, LDL, Triglyceride, HDL and VLDL parameters are used in the process of entering the system and finally Hyperlipidemia (T) and Hyperlipidemia (F) results have been evaluated at the end of this process. Data of 492 patients are evaluated in the projected system. The results of the decision support system have completely matched with those of the physicians decisions.

Keywords- Data Mining, Association Rules, Hyperlipidemia, Expert system.

1. INTRODUCTION

Hyperlipidemia is a term doctors use to describe several conditions in which high concentrations of lipids (fats) exist in the bloodstream and it occurs because of the abnormal increase of the fats in blood [1]. Lipid is the scientific term for fats in the blood and like vitamins and minerals, certain fats are useful to the body as an energy source and to help build cells and hormones. The measurement of plasma lipid levels makes it possible to determine the hyperlipidemia patients and so necessary precautions can be taken easily. Most commonly, changes of the serum levels is taken as the base in hyperlipidemia diagnosis. The parameters taken into consideration in hyperlipidemia diagnosis are Total cholesterol, LDL, Triglyceride, HDL and VLDL [2-5]. These enzymes form great data stacks when used in each person's diagnosis. Also, as these kinds of researches in the biology and medical science world increase, this will cause the increase in different kinds of data related with the real-life cases [6]. It became quite difficult for people to be able to comprehend and change hundreds of characteristics and thousands of images into meaningful knowledge in modern medical science [7, 8], DNA [9] and protein synthesis [10], biological state measurements [11], graphs [12] and enzymes [6] are some of the main types of data which are still being used. As the data increased it became quite a boring and difficult work for the medical science experts to prepare a guide that will be used in the processes of reading, simplifying, classifying the findings and making a decision at the end. Also a lot of findings hidden in these data have remained as the stack in this way. It is necessary to get these data automatic in order to obtain useful information. Getting information from the databases or the data mining is the kind of method mostly used for solving these kinds of

problems [13]. Increasingly, researchers in medical informatics are using natural language processing, information extraction, and data mining methods to utilize existing clinical data for secondary purposes. Researchers have employed such techniques in decision support efforts, to encode medical reports, to provide structured data for data mining and clinical research, in order entry and in information retrieval [14]. In this context data mining has been applied so successfully in many fields of medical science [13,15]. For instance, there are a lot of work that can be seen in the literature related with discovering certain rules in the diagnosis and treatment of acute ailments; applying the determined rules and making complementary data mining for defining the enlargement states of the working procedures of the refined, shared, organized and produced information in the system [16]. In these works, data stores are known as the clinical stores in medical field [6]. Data warehouses in medicine field are called clinical warehouse too. Then, those clinical stores containing biological, clinical and administrating data unit the patients information. Thus, the possibility of the usage of the systems related with the patient is improved [17].

Among the data-mining technologies, finding association rules in transaction databases is most commonly seen. Association rule induction is a powerful data mining method for finding temporal trends in large datasets. The goal of data mining is to automate the process of finding interesting temporal patterns. The output of a data-mining method should be a “summary” of the data sets. Such goal is difficult to achieve due to the vagueness of the term “interesting”. The solution is to define various types of trends (patterns) and to look for only those defined trends in the data sets. One such type of trend is the association rule [18]. Association rules identify the set of items that are most often purchased with another set of items. For example, an association rule may state that “95% of customers who bought items A and B also bought C: and D.” Association rules may be used for catalog design [19], biomedical [20], store layout [21], product placement, target marketing [22], etc [23]. Association rules in biomedical are used the microarray gene expression in applications [24], the ant colony system in applications [25], biomedical data classification [26], visual text mining [27], extracted data in large databases [28].

In this study, a decision support system which will be helpful for the diagnosis of hyperlipidemia is improved to be used for the aim of classifying the association rule construction from the data mining techniques. Lipide parameters from biochemistry data were used for application and 492 patient’ data have been evaluated successfully.

2. THEORETICAL VIEW

In this part, the informations related with the easier conception of the decision support system’s construction in hyperlipidemia diagnosis is introduced in the shape of sub-classifications.

2.1. Lipide Parameters

Hyperlipidemia is related with concentration increase of plasma lipoproteins. More than one lipoprotein class can come together due to the increased construction or release in circulation or because of the decreased clearance or of being far away from the circulation. These changes in the metabolic events are often related with the changes in

the apolipoproteins, receptors, enzymes or cofactors related with the lipoprotein metabolism. Such changes rooted from the genetic changes are classified as primary disorders of the lipid metabolisms. In Hyperlipidemia Total Cholesterol, LDL, HDL, Triglyceride ve VLDL values are taken as the base [2-5,29]. These parameters are shortly explained in the following part:

Total Cholesterol: Cholesterol is a steroid alcohol (sterol) found in animal fats and oils. It is widely distributed throughout the body, especially in the blood, brain, liver, kidneys, and nerve fiber myelin sheaths, and it is an essential component of cell membrane development and production of bile acids, adrenal steroids, and sex hormones. Cholesterol testing evaluates the risk of atherosclerosis, myocardial occlusion, and coronary arterial occlusion. Cholesterol relates to coronary heart disease (CHD) and is an important screening test for risk factors. It is part of the lipid profiles. Elevated cholesterol levels are a major component in the hereditary hyperlipoproteinemias.

Triglycerides: Triglycerides account for >90% of dietary intake and comprise 95% of fat stored in tissues. Because they are insoluble in water, they are the main plasma glycerol ester. Normally stored in adipose tissue as glycerol, fatty acids, and monoglycerides, the liver reconverts these to triglycerides are in VLDL, and 15% are in LDL. This test evaluates suspected atherosclerosis and measures the body's ability to metabolize fat. Elevated triglycerides, together with elevated cholesterol, are atherosclerotic disease risk factors. Because cholesterol and triglycerides can vary independent of each other, measurement of both values is more meaningful.

HDL: HDL-C is a class of lipoproteins produced by the liver and intestines. HDL is comprised of phospholipids and 1 or 2 apolipoproteins. It plays a role in the metabolism of the other lipoproteins and in cholesterol transport from peripheral tissues to the liver. LDL and HDL may combine to maintain cellular cholesterol balance through the mechanism of LDL moving cholesterol into the arteries and HDL removing it the from the arteries. Decreased HDL levels are atherogenic, whereas elevated HDL levels protect against atherosclerosis by removing cholesterol from vessel walls and transporting it to the liver where it is removed from the body.

LDL-VLDL: Most serum cholesterol is present in the LDL. LDLs are the cholesterol-rich remnants of the VLDL lipid transport vehicle. Because LDL has a longer half-life (3-4 days) than its precursor VLDL, LDL is more prevalent in the blood. It is mainly catabolized in the liver and possibly in nonhepatic cells as well. The VLDLs are major carriers of triglycerides. Degradation of VLDL is a major source of LDL. Circulating fatty acids form triglycerides in the liver, and these are packaged with apoprotein and cholesterol to be exported into the blood as VLDLs. Therefore, LDL is the test of choice because of its longer half-life and the fact that VLDLs are extremely hard to measure. This test is specifically done to determine CHD risk. LDLs are closely associated with increased incidence of atherosclerosis and CHD [30].

2.2. Data Mining

Data mining has recently emerged as a growing field of multidisciplinary research. It combines disciplines such as databases, machine learning, artificial intelligence, statistics, automated scientific discovery, data visualization, decision science, and high performance computing. It is our contention that data mining

techniques can be used to provide the knowledge required to assist users in locating relevant information on the web, through automating the analysis of the current users navigation and combining this with datamined knowledge of multiple users' buying behaviour [31]. For instance, some standart softweres used in a medical diagnosis system were problematic because of the unsystematic data, the abscence of control, the usage of too many various kinds, being unable to make a consistent and systematic analysis on databases, comparing the examples and determining the critical differences. Data mining and techniques are also accepted in the medical science world because they make everything easier for the experts and provides a necessary and important help for the practitioners [32]. In fact data mining is evaluated as a part of the information discovery process both in medical science and in the other fields [33]. Data mining stages are presented in Figure 1 [34].

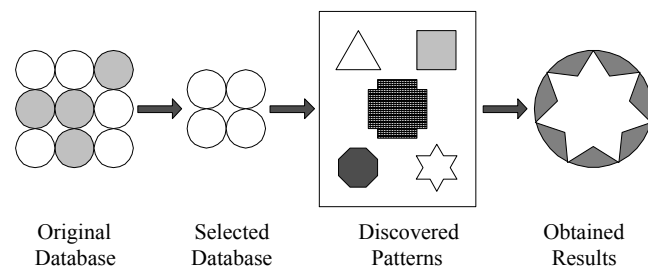


Figure 1. Data mining processes.

Data mining has an interaction with the user and the database. Interesting data patterns are showed to the user [35]. And also, they can be saved in the database if wanted. According to this, data mining goes on till the hidden data patterns are found. Firstly the necessary data are taken, classified and then processed while obtaining meaningful information from the databases. It is an important problem to be able to save the state of a patient and to predict the characteristic of the data such as all the laboratory test results, findings and signals of all the patients. These points are also problems in machine learning and data mining which works in many fields such as classification and problem detection [8]. Classification in data mining is used for an automatic definition of the interesting object in great data and for the information discovery in the applications including the classification of the trend in the market. There are a lot of methods used for classifying these data. In data mining the decision trees, association rules, genetic algorithm etc. among these techniques [36].

2.3. Association Rules

A number of data mining algorithms have been introduced to the community that perform summarization of the data, classification of data with respect to a target attribute, deviation detection, and other forms of data characterization and interpretation. One popular summarization and pattern extraction algorithm is the association rule algorithm, which identifies correlations between items in transactional databases. Given a set of transactions, each described by an unordered set of items, an association rule $X \rightarrow Y$ may be discovered in the data, where X and Y are conjunctions of items. The intuitive meaning of such a rule is that transactions in the database which contain the items in X , tend to also contain the items in Y . An example of such a rule might be that many observed customers who purchase tires and auto accessories also

buy some automotive services. In this case, $X = \{\text{tires, auto accessories}\}$ and $Y = \{\text{automotive services}\}$. Two numbers are associated with each rule, that indicate the support and confidence of the rule. The support of the rule $X \rightarrow Y$ represents the percentage of transactions from the original database that contain both X and Y . The confidence of rule $X \rightarrow Y$ represents the percentage of transactions containing items in X that also contain items in Y . Applications of association rule mining include cross marketing, attached mailing, catalog design and customer segmentation. An association rule discovery algorithm searches the space of all possible patterns for rules that meet the user-specified support and confidence thresholds. The problem of discovering association rules can be divided into two steps:

1. Find all *itemsets* (sets of items appearing together in a transaction) whose support is greater than the specified threshold. Itemsets with minimum support are called *frequent itemsets*.
2. Generate association rules from the frequent itemsets. To do this, consider all partitionings of the itemset into rule left-hand and right-hand sides. Confidence of a candidate rule $X \rightarrow Y$ is calculated as $\text{support}(XY) / \text{support}(X)$. All rules that meet the confidence threshold are reported as discoveries of the algorithm.

In addition to the antecedent (the "if" part) and the consequent (the "then" part), an association rule has two numbers that express the degree of uncertainty about the rule. In association analysis the antecedent and consequent are sets of items (called itemsets) that are disjoint (do not have any items in common). The support is simply the number of transactions that include all items in the antecedent and consequent parts of the rule. (The support is sometimes expressed as a percentage of the total number of records in the database.) Confidence is the ratio of the number of transactions that include all items in the consequent as well as the antecedent (namely, the support) to the number of transactions that include all items in the antecedent [37].

3. DEVELOPED METHOD

The most common and main tests that help the physicians decide on a diagnosis are biochemistry tests which are mostly successful in diagnosing. An hyperlipidemia diagnosis can be obtained by controlling the lipide parameters which were taken into consideration for the hyperlipidemia diagnosis in biochemistry test results. The construction of the improved decision support system is showed in Figure 2. The functions of system are constituted of these following steps:

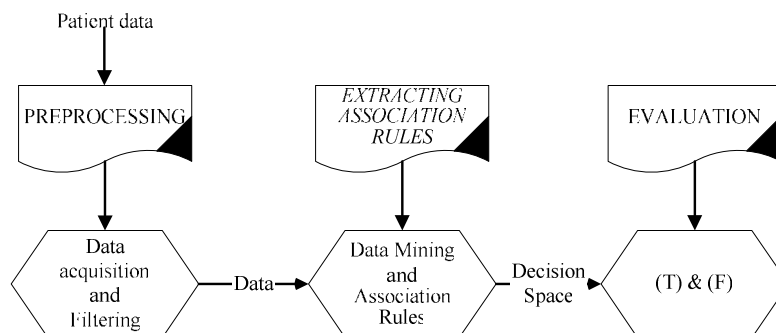


Figure 2. The algorithm of decision support system.

Step – 1: Pre-processing

Collecting the data is the step in which the ones those are proper for the goal are choosen. The necessary data were obtained by collecting the biochemistry test results of the patients applied for the internal medicine and by analyzing the ones related with the lipide parameters with the direction of the physicians. The database was created by choosing the test results of the Hyperlipidemia suspected-ill from the data collected.

Step – 2: Feature Extraction and Classification

This is the most important step of association rule based decision support system which was developed for processing the data in the collected database. The data collected are being processed in this stage. In this paper, enzymes are classified normal and reference values A, B, C and are showed table 1. For this, the rules underlined in table2 and determined according to the physician's knowledge. The possibility of rules were given with percentage rates in table2. This values over 50 percent, it shows that the disease is positive. As for the others can say that negative.

Table 1. Lipide refence value.

Lipides	Normal values	Reference					
T. Cholesterol	0-200 mg/dL	<i>TCHOL>200</i>	<i>A</i>	<i>0<TCHOL<200</i>	<i>B</i>	<i>TCHOL<0</i>	<i>C</i>
LDL	80-180 mg/dL	<i>LDL>180</i>	<i>A</i>	<i>80<LDL<180</i>	<i>B</i>	<i>LDL<80</i>	<i>C</i>
Triglyceride	60-165 mg/dL	<i>TG>165</i>	<i>A</i>	<i>60<TG<165</i>	<i>B</i>	<i>TG<60</i>	<i>C</i>
HDL	40-60 mg/dL	<i>HDL>60</i>	<i>C</i>	<i>40<HDL<60</i>	<i>B</i>	<i>HDL<40</i>	<i>A</i>
VLDL	20-50 mg/dL	<i>VLDL>50</i>	<i>A</i>	<i>20<VLDL<50</i>	<i>B</i>	<i>VLDL<20</i>	<i>C</i>

Table 2. The rule base of proposed system.

Condition & Result		Confidence
1	If TCHOL = A Then LDL=A & HYPER=T	% 52.00
2	If TCHOL = A Then TG=A & HYPER=T	% 72.00
3	If TCHOL = A Then HDL=A & HYPER=T	% 58.00
4	If LDL = A Then TCHOL =A & HYPER=T	% 92.00
5	If LDL = A Then TG =A & HYPER=T	% 71.00
6	If LDL = A Then HDL =A & HYPER=T	% 63.00
7	If LDL = B Then TCHOL =A & HYPER=T	% 78.00
8	If LDL = B Then TG =A & HYPER=T	% 66.00
9	If LDL = B Then HDL =A & HYPER=T	% 59.00
10	If TG = A Then TCHOL =A & HYPER =T	% 91.00
11	If TG = A Then HDL =A & HYPER =T	% 60.00
12	If TG = A Then VLDL =A & HYPER =T	% 60.00
13	If HDL = A Then TCHOL =A & HYPER =T	% 81.00
14	If HDL = A Then TG =A & HYPER =T	% 66.00
15	If HDL = B Then TCHOL =A & HYPER =T	% 93.00
16	If HDL = B Then TG =A & HYPER =T	% 71.00
17	If LDL = A Then TCHOL =A & TG =A & HYPER =T	% 71.00
18	If LDL = A Then TCHOL =A & HDL =A & HYPER =T	% 63.00
19	If LDL = B Then TCHOL =A & TG =A & HYPER =T	% 57.00
20	If TG = A Then TCHOL =A & HDL =A & HYPER =T	% 53.00
21	If TG = A Then TCHOL =A & VLDL =A & HYPER =T	% 58.00
22	If HDL = A Then TCHOL =A & TG=A & HYPER =T	% 58.00
23	If TCHOL = A & LDL=A Then HYPER =T	% 98.00
24	If TCHOL = A & LDL=B Then HYPER =T	% 97.00
25	If TCHOL = A & TG=A Then HYPER =T	% 98.00
26	If TCHOL = A & HDL=A Then HYPER =T	% 98.00
27	If TCHOL = A & HDL=B Then HYPER =T	% 96.00

28	If TCHOL = A & VLDL=A Then HYPER =T	% 97.00
29	If TCHOL = A & VLDL=B Then HYPER =T	% 96.00
30	If LDL = A & TG=A Then HYPER =T	% 98.00
31	If LDL = A & HDL=A Then HYPER =T	% 98.00
32	If LDL = B & TG=A Then HYPER =T	% 98.00
33	If LDL = B & HDL=A Then HYPER =T	% 92.00
34	If TG = A & HDL =A Then HYPER =T	% 98.00
35	If TG = A & HDL =B Then HYPER =T	% 97.00
36	If TG = A & VLDL =A Then HYPER =T	% 98.00
37	If TCHOL = A & LDL =A Then TG=A & HYPER =T	% 71.00
38	If TCHOL = A & LDL =A Then HDL=A & HYPER =T	% 63.00
39	If TCHOL = A & LDL =B Then TG=A & HYPER =T	% 72.00
40	If TCHOL = A & TG =A Then HDL=A & HYPER =T	% 58.00
41	If TCHOL = A & TG =A Then VLDL=A & HYPER =T	% 64.00
42	If TCHOL = A & HDL =A Then LDL=A & HYPER =T	% 55.00
43	If TCHOL = A & HDL =A Then TG=A & HYPER =T	% 72.00
44	If TCHOL = A & VLDL =A Then TG=A & HYPER =T	% 97.00
45	If LDL = A & TG=A Then TCHOL =A & HYPER =T	% 98.00
46	If LDL = A & HDL=A Then TCHOL =A & HYPER =T	% 98.00
47	If LDL = B & TG=A Then TCHOL =A & HYPER =T	% 85.00
48	If TG = A & HDL=A Then TCHOL =A & HYPER =T	% 88.00
49	If TG = A & VLDL=A Then TCHOL =A & HYPER =T	% 97.00
50	If TCHOL = A & LDL =A & TG=A Then HYPER =T	% 100.00
51	If TCHOL = A & LDL =A & HDL=A Then HYPER =T	% 98.00
52	If TCHOL = A & LDL =B & TG=A Then HYPER =T	% 100.00
53	If TCHOL = A & TG =B & HDL=A Then HYPER =T	% 100.00
54	If TCHOL = A & TG =B & VLDL=A Then HYPER =T	% 100.00
55	If TCHOL = A & TG =B & VLDL=A & HDL=A Then HYPER =T	% 100.00

Step – 3: Evaluation

In this step, rules which are chosen support and confidence values evaluated. Support value 25 and confidence values are choosen 50. Because this values is accepted for reliability to increase in the developed decision support system and the data obtained after the feature extraction and classification process are presented after the evaluation. Therefore, rules have evaluated level A and B in the system. Support and confidence are calculated by the following formula:

$$\text{Support}(XY) = \text{Support count of } XY / \text{Total number of transaction in Database}$$

$$\text{Confidence}(XY) = \text{Support}(XY) / \text{Support}(X)$$

4. RESULTS

The developed decision support system is experienced on the biochemistry test results of the 492 patients. In the system Hyperlipidemia (T) diagnosis was certain in 311 (224 test data and 87 training) of the 492 patients and the other 81 (68 test data and 13 training) were healthy. And we saw that these results have successfully matched with the physicians'. Results are showed table3.

This test data results contain high reliability because it wasn't incorrect classification. Specially, the system proves easiness of the induced confirmation rules, both for the rules generated in this work by the association approach and those already presented in. According to physicians, large amounts of data are generated in modern medicine, but there is a widening gap between data collection and data comprehension. It is clear that a single human can not process all of the data available and make a

rational decision of basic trends. Thus, there is a growing pressure for intelligent data analysis techniques to facilitate the creation of knowledge to support clinicians in making decisions

Table 3. Performance of the decision support system

	Hyperlipidemia			
	(T) Test Data	(T) Training Data	(F) Test Data	(F) Training Data
<i>Total number of samples</i>	224	87	68	13
<i>Correct classification</i>	224	87	68	13
<i>Incorrect classification</i>	---		---	

5. DISCUSSION AND CONCLUSION

The increase in the knowledge obtained under the light of biomedical searches, the processes of reading, simplifying, classifying these findings and making a decision are being so complex day by day. Understanding the major risk factors of a disease is an important factor for clinicians in prevention strategy. The attending physician plays an important role providing information to reduce those risk factors. It is up to the physician whether to warn patients at risk about the major causes of a particular disease and the degree of risk that they are facing. These processes are automated after the data mining took a part in this field. It has considerably helped the medical experts and made it easier to prepare a guide. Also, a lot of findings hidden between the data mining and the data stacks obtained in these kinds of fields have turned to be useful information in this way.

In this paper we dealt with association rule. We restricted ourselves to the “classic” association rule problem, that is the generation of all association rules that medical data with respect to minimal thresholds for support and confidence. The advised system is based on association rules on which so many clever diagnosis systems are constituted. There, the association rules, which take an important part in data mining for the feature extracting and classification stage, have been used. For this disease, decision tree technique, a kind of data mining techniques, has been applied [38]. On the other hand, since association rules find results with percentage rates, it’s better than decision tree technique for those diseases which can’t be separated by definite rules.

The developed decision support system will considerably be helpful for the expert physicians and practitioners for the interpretation of the illnesses. This system construction can also be used in the diagnosis of every illnesses in which the criteria having certain parameters can be controlled.

Acknowledgments - We want to thank, Department of Biochemistry, Ufuk Medical Center, Elazig, TURKIYE for providing the Biochemistry tests to us and Dr. Osman Metin IPCIOGLU.

6. REFERENCES

1. A. Azezli, A. Kubat, Lipide’s place of the hyperlipidemia treatment, *Turkey Clinics J Int Med Sci* **1(20)**, 42-48, 2005.

2. S. Koba, R. Pakala, Hyperlipemic-very low density lipoprotein, intermediate density lipoprotein and low density lipoprotein act synergistically with serotonin on vascular smooth muscle cell proliferation, *Atherosclerosis* **149**, 61–67, 2000.
3. G. Lupattelli, S. Marchesi, Mechanisms of High-Density Lipoprotein Cholesterol Effects on the Endothelial Function in Hyperlipemia, *Metabolism* **52(9)**, 1191–1195, 2003.
4. T.D. Wang, W.J. Chen, Efficacy of fenofibrate and simvastatin on endothelial function and inflammatory markers in patients with combined hyperlipidemia: relations with baseline lipid profiles, *Atherosclerosis* **170**, 315–323, 2003.
5. H. S. Lee, H. C. Ahn, Hypolipemic effect of water extracts of *Picrorrhiza rhizoma* in PX-407 induced hyperlipemic ICR mouse model with hepatoprotective effects: A prevention study, *Journal of Ethnopharmacology* **105**, 380–386, 2006.
6. J. Barrera, R.M. Cesar-Jr, An Environment For Knowledge Discovery in Biology, *Computers in Biology and Medicine* **34**, 427–447, 2003.
7. V. Podgorelec, P. Kokol, Knowledge Discovery with Classification Rules in a Cardiovascular Dataset, *Computer Methods and Programs in Biomedicine* **80**, 39–49, 2005.
8. C. C. Bojarczuk, H. S. Lopes, A Constrained-Syntax Genetic Programming System for Discovering Classification Rules: Application to Medical Data Sets, *Artificial Intelligence in Medicine* **30**, 27–48, 2004.
9. M. M. Yin, J. T. L. Wang, GeneScout: a Data Mining System For Predicting Vertebrate Genes in Genomic DNA Sequences, *Information Sciences* **163**, 201–218, 2003.
10. J. M. Ayub, C. R. Smulski, Protein–Protein Interaction Map of the Trypanosoma Cruzi Ribosomal P Protein Complex, *Gene* **357**, 129 – 136, 2005.
11. R. J. Shebuski, Utility of Point-of-Care Diagnostic Testing in Patients with Chest Pain and Suspected Acute Myocardial Infarction, *Current Opinion in Pharmacology* **2**, 160–164, 2002.
12. I. Turkoglu, A. Arslan, An Intelligent System for Diagnosis of Heart Valve Diseases with Wavelet Packet Neural Networks, *Computer in Biology and Medicine* **33(4)**, 319–331, 2003.
13. K.C. Tan, Q. Yu, Evolutionary Computing for Knowledge Discovery in Medical Diagnosis, *Artificial Intelligence in Medicine* **27**, 129–154, 2002.
14. L. D’Avolio, G. Rees, *Improving the Secondary Utilization of Clinical Data by Incorporating Context*, Los Angeles CA, 2000.
15. T. Unger, Z. Korade, True and False Discovery in DNA Microarray Experiments: Transcriptome Changes in the Hippocampus of Presenilin 1 Mutant Mice, *Methods* **37**, 261–273, 2005.
16. M. J. Huang, M. Y. Chen, Integrating Data Mining with Case-Based Reasoning for Chronic Diseases Prognosis and Diagnosis, *Expert Systems with Applications* **32**, 856–867, 2007.
17. I. M. Mullins, M. S. Siadaty, Data Mining and Clinical Data Repositories: Insights from a 667,000 Patient Data Set, *Computers in Biology and Medicine* **36**, 1357–1377, 2005.
18. L. Ma, C. Tsui, A Framework for Infection Control Surveillance Using Association Rules, *AMIA Annu Symp Processing* **9**, 410–414, 2003.

19. R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, *Very Large Databases* **153(8)**, 487–499, 1994.
20. H. Kwasnicka, K. Switalski, Discovery of association rules from medical data - classical and evolutionary approaches, *ISSN* **5**, 204–217, 2006.
21. A. Metwally, D. Agrawal, Using Association Rules for Fraud Detection in Web Advertising Networks. *Very Large Databases* **13**, 169–180, 2005.
22. D. Malerba, F. Esposito, Mining Spatial Association Rules in Census Data, *Research in Official Statistics* **1**, 19–45, 2002.
23. A. Savasere, E. Omiecinsky, An efficient algorithm for mining association rules in large databases, *Very Large Databases* **8(2)**, 134–145, 1995.
24. X. R. Jiang and L. Gruenwald, Microarray gene expression data association rules mining based on BSC-tree and FIS-tree, *Data & Knowledge Engineering* **53**, 3–29, 2005.
25. R.J. Kuo and C.W. Shih, Association rule mining through the ant colony system for National Health Insurance Research Database in Taiwan, *Computers and Mathematics with Applications* **54**, 1303–1318, 2006.
26. Y. He and Y. Tang, Adaptive Fuzzy Association Rule mining for effective decision support in biomedical applications, *Data Mining and Bioinformatics* **1(1)**, 3–18, 2006.
27. A.A. Lopes and R. Pinho, Visual text mining using association rules, *Computers & Graphics* **31**, 316–326, 2007.
28. A. Nanopoulos and A. N. Papadopoulos, Mining association rules in very large clustered domains, *Information Systems* **32**, 649–669, 2007.
29. H. Karpuz, Current approach of hyperlipidemia, *Cardiac Symposium Series* **40**, 69–74, June, 2004.
30. F. Fischbach, *A Manual of Laboratory & Diagnostic Tests*, Lippincott, New York, 2000.
31. X. Zhang and J. Edwards, Personalised online sales using web usage data mining, *Computers in Industry* **58**, 772–782, 2007.
32. F. Alonso and J. P. Caraca-Valente, Combining Expert Knowledge and Data Mining in a Medical Diagnosis Domain, *Expert Systems with Applications* **23**, 367–375, 2002.
33. N. Allahverdi, A artificial intelligence application with expert systems, *Satin Publication Distribution*, Istanbul, 2002.
34. G. S. Oger, The use of data mining for the diagnosis of osteoporos disease, Firat University, Master Thesis, 2003.
35. S. Dogan, An analyse of a diseases with biochemical data by using data mining, Firat University, Master Thesis, 2007.
36. R. Agrawal and M. Mehta, *The Quest Data Mining System*, IBM Almaden Research Center, California, U.S.A., 1996.
37. J. R. Nayak and D. J. Cook, Approximate Association Rule Mining, *University of Texas at Arlington* **1**, 259–263, 2001.
38. S.Dogan and I. Turkoglu, Hyperlipidemia detection from lipide paramets by using decision trees, *e-journal of New World Sciences Academy* **1(2)**, 50–61, 2006.