DOI: 10.37990/medr.1088631 Med Records 2022;4(3):286-92

MEDICAL RECORDS-International Medical Journal

Research Article

Assessment of Association Rule Mining Using Interest Measures on the Gene Data

Gen Verileri Üzerinde İlginçlik Ölçütleri Kullanılarak Birliktelik Kuralları Madenciliğinin Uygulanması

Kubra Elif Akbas[1], Mehmet Kivrak[2], Ahmet Kadir Arslan[3], Tugce Yakinbas[4], Hasan Korkmaz[5], Ebru Etem Onalan[4], Cemil Colak[3]

[1]Firat University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Elazig, Turkey
[2]Recep Tayyip Erdogan University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Rize, Turkey
[3]Inonu University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey
[4]Firat University, Faculty of Medicine, Department of Medical Biology, Elazig, Turkey
[5]Firat University, Faculty of Medicine, Department of Cardiology, Elazig, Turkey

Abstract

Aim: Data mining is the discovery process of beneficial information, not revealed from large-scale data beforehand. One of the fields in which data mining is widely used is health. With data mining, the diagnosis and treatment of the disease and the risk factors affecting the disease can be determined quickly. Association rules are one of the data mining techniques. The aim of this study is to determine patient profiles by obtaining strong association rules with the apriori algorithm, which is one of the association rule algorithms.

Material and Method: The data set used in the study consists of 205 acute myocardial infarction (AMI) patients. The patients have also carried the genotype of the FNDC5 (rs3480, rs726344, rs16835198) polymorphisms. Support and confidence measures are used to evaluate the rules obtained in the Apriori algorithm. The rules obtained by these measures are correct but not strong. Therefore, interest measures are used, besides two basic measures, with the aim of obtaining stronger rules. In this study For reaching stronger rules, interest measures lift, conviction, certainty factor, cosine, phi and mutual information are applied.

Results: In this study, 108 rules were obtained. The proposed interest measures were implemented to reach stronger rules and as a result 29 of the rules were qualified as strong.

Conclusion: As a result, stronger rules have been obtained with the use of interest measures in the clinical decision making process. Thanks to the strong rules obtained, it will facilitate the patient profile determination and clinical decision-making process of AMI patients.

Keywords: Data mining, association rules, apriori algorithm, interest measures, gene expression data

Öz

Amaç: Veri madenciliği, önceden büyük ölçekli verilerden ortaya çıkarılmayan faydalı bilgilerin keşfedilme sürecidir. Veri madenciliğinin yaygın olarak kullanıldığı alanlardan biri de sağlıktır. Veri madenciliği ile hastalığın tanı ve tedavisi ile hastalığı etkileyen risk faktörleri hızlı bir şekilde belirlenebilmektedir. Birliktelik kuralları, veri madenciliği tekniklerinden biridir. Bu çalışmanın amacı, birliktelik kuralı algoritmalarından biri olan apriori algoritması ile güçlü birliktelik kuralları elde ederek hasta profillerini belirlemektir.

Materyal ve Metot: Çalışmada kullanılan veri seti 205 akut miyokard enfarktüsü (AMI) hastasından oluşmaktadır. Hastalar ayrıca FNDC5 polimorfizmlerinin rs3480, rs726344, rs16835198 genotipini de taşımaktadır. Apriori algoritması ile elde edilen kuralları değerlendirmek için destek ve güven ölçüleri kullanılır. Ancak bu ölçütler ile elde edilen kurallar doğrudur ancak güçlü değildir. Bu nedenle, daha güçlü kurallar elde etmek amacıyla iki temel ölçütün yanı sıra ilginçlik ölçütleri kullanılmaktadır. Bu çalışmada daha güçlü kurallara ulaşmak için ilginçlik ölçütlerinden kaldıraç, kanaat, kesinlik faktörü, cosine, korelasyon katsayısı (phi) ve karşılıklı bilgi ölçütleri uygulanmıştır.

Bulgular: Çalışmada 108 kural elde edilmiştir. Bu kurallara ilginçlik ölçütlerinin de uygulanması ile elde edilen kural sayısı 29 olmuştur ve bu kurallar güçlü kural olarak nitelendirilmiştir.

Sonuç: Sonuç olarak, klinik karar verme sürecinde ilginçlik ölçütlerinin kullanılmasıyla daha güçlü kurallar elde edilmiştir. Elde edilen güçlü kurallar sayesinde AMİ hastalarının hasta profili belirleme ve klinik karar verme sürecini kolaylaştıracaktır.

Anahtar Kelimeler: Veri madenciliği, birliktelik kuralları, apriori algoritması, ilginçlik ölçütleri, gen ifadesi verisi

Received: 23.03.2022 **Accepted:** 13.05.2022
Corresponding Author: Kubra Elif Akbas, Firat University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Elazig, Turkey, **E-mail**: keet@firat.edu.tr

INTRODUCTION

Acute myocardial infarction (AMI) is the formation of necrosis in the heart tissue as a result of insufficient blood supply to the heart muscle due to the blockage of the coronary arteries feeding the heart. The heart muscle cannot be fed and damaged in terms of blood and oxygen due to the narrowing or blockage of the coronary arteries. The non-feeding heart muscle is damaged, and the area of the heart muscle cells loses its ability to contract. AMI is an important public health problem because it is a disease that can result in death, it is generally seen in all age groups and then serious complications occur (1).

Data mining, which is also called discovery of data in the systems including data bases, is a process of beneficial information from the large data sets. This new-found information can be used in areas, consisting of information management, inquiry work, and decision-making process control in this aspect (2). Researchers working in the fields of database systems, knowledge base systems, artificial intelligence, machine learning, data collection, statistics, spatial databases and data visualization show great interest in data mining. (3,4). The use of data mining is very important, especially in the field of health. (5,6). For instance, the data mining is used for evaluating the efficiency of medical treatment carried out. Also, the data mining techniques such as classification, association, clustering make a major contribution to detecting epidemics and to preventing various diseases from appearing (6).

Association rules mining is the most important technique of data mining to produce strong association rules in data base (6). The aim of this technique is to reveal relationships between sets of items in transactional databases or other data warehouses (7). By spreading area of usage, the association rules can be applied in field of medical. Thanks to these rules, connection of various variables can be identified in medical information and that situation helps to reach diagnosis (8). Algorithms such as apriori, eclat and FP-growth algorithms are used in association rules mining and apriori algorithm is the most commonly used among these algorithms (6,9). For generating stronger rules by interest measures such as lift, conviction, certainty factor, cosine, coefficient of correlation (phi) and mutual information, this study intends to constitute strong association rules by using the studied data of the patients suffering from acute myocardial infarction. In this study, apriori algorithm is applied in order to constitute association rules from a gene dataset.

MATERIAL AND METHOD

Dataset

The dataset includes 225 patients, suffering from AMI disease in coroner intensive care unit of Firat University Medical Faculty Cardiology Department, Elazig, Turkey. The patients have also carrying the genotype of the FNDC5 (rs3480, rs726344, rs16835198) polymorphisms. In order to generalize the results obtained, the suggested number of observations should be 5 per independent variable, but this number is expected to be above 10, especially between 15 and 20. When these numbers are reached, it is said that the results can be generalized, provided that the sample represents the population. Since the number of independent variables is 15, the minimum sample size was determined as 150 based on the assumption of a multiple of 10 (10). Patients who did not suit the inclusion criteria were excluded from the study, and the study was conducted on 205 patients. The patients included in the study were determined by simple random sampling method. The study received ethics approval from Firat University's Faculty of Medicine Local Ethics Committee (14.10.2014, Decision No: 17/07). AMI is the death of a part of myocardium after coroner arterials supplying to heart are blockaded by blood clot. Criteria for taking part in the experiment are set by assessing AMI diagnosis, clinic symptom and superficial Electrocardiogram (ECG), which was accepted as critic for intensive care unit admission and/or cardiac biomarkers (11). All the association rules analyses were carried out by "arules" package in R programming language (12). Rstudio Version 1.1.463 program language was used for attaining analysis of the data (13). Detailed explanations about the variables used in the current study are given in Table 1.

Table 1. The detailed explanation of the variables in the dataset

Abbreviation	Explanation
Age	Birth year (year)
Height	Height (cm)
Weight	Weight (kg)
BMI	Body Mass Index (kg/m^2)
Troponin	Troponin (ng/ml)
CK-MB	Creatine Kinase Myocardial Band (ng/ml)
Irisin	Irisin (ng/mL)
LDH	Laktat Dehidrogenaz (mg/dL)
Gender	Gender(1=male, 2=female)
Hypertension	Hypertension (1= presence, 2= absence)
Smoke	Current Smoker (1= presence, 2= absence)
Diabetic	Diabetic (1= presence, 2= absence)
Family history	Family History (1= presence, 2= absence)
Mtype	Myocardial Infarction Type (1=acute inferior, 2=acute anterior, 3= non-ST segment elevation MI)
Rs3480in FNDC5gene	Rs3480 (1=AA wild, 2=AG: heterozygote genotype, 3= GG: polymorphism genotype)
Rs726344in FNDC5 gene	Rs726344 (1=GG wild, 2=GA: heterozygote genotype, 3= AA: polymorphism genotype
Rs16835198in FNDC5 gene	Rs16835198 (1=GG wild, 2=GT: heterozygote genotype 3= TT: polymorphism genotype)

Association Rules Mining

Association rules is one of the data mining methods, which focuses on interesting association among large-scale information. The first studies carried out in the field of data mining aimed to detect similarities between different products in data bases consisting of customer operations and customer purchasing profile. Association rules area of usage is expanded by later studies. One of them is the field of medicine. The use of association rules in the field of medicine has discovered the association of features in the medical data, which has been very useful in the diagnosis of medical conditions (8).

The form of association rules is formulated as X⇒Y. In that formula, X is antecedent of rule and Y is known as consequence (14,15).

Apriori Algorithm

Apriori Algorithm is the most classic and significant algorithm that used for finding frequent data sets in a specific data base (14,16). The apriori algorithm enables multi-transaction on database. During these transactions, an iterative approach is used throughout the research space. In other words, k item set is used for defining (k+1) item set. First, frequent 1- item set is obtained. This set includes an item which provides support threshold and indicated with L1. In every next transition, it begins with the biggest item set of former seed set. This seed set is used to create potentially large item sets, called candidate item sets, and the real support value is taken into account during the process of transferring data to the candidate item set. At the end of the transaction, it is determined which candidate items value is bigger (in frequency) actually and it naturally becomes source for next transaction. By this way, L1 is applied for calculating L2, which is also had to be found for discovering L3 value. The transaction continues until not finding k-item set (14).

Basic Measures

Even if numerous interest measures are developed for assessing association in data mining, two basic measures used for forming strong association rules with frequent item sets are support and confidence (17).

1. Support: Support shows percentage of processes in database. This is calculated as the probability of P(XUY). XUY means, the operation includes X and Y together, in short it means association of X and Y.

It can be defined by support (X⇒Y) = P(XUY) formula (18).

2. Confidence: Other basic measure used for association rules is confidence. Confidence gives certainty degree of association discovered. P(Y|X) means, possibility of including Y in an operation with X together and it is defined as below,

confidence (X⇒Y) = P (Y|X) (18).

Interest Measures

Sometimes conventional confidence measures fall short to explain real interesting associations patterns fully, so various interest measures discussed for association rules. These interest measures are categorized in two groups as subjective and objective. Subjective measures take both data and users into consideration. In this respect, if a pattern discloses unexpected information about data, it is classified as subjective. On the other hand information, which attracts interest of a user, may not get attention of another one. Because of that, objective measures are needed. Objective measures are mostly dependent on contingency, statistic and information theory. Objective measures don't require any information about users and impact area beforehand. It measures interest of an association rule in the way of structure and basic information on the discovery process. Objective measures are generally calculated with frequency counts on contingency table. Some objective measures are mentioned below (19).

1. Lift: Lift is a measure of how many times X and Y are together more than expected if they are statistically independent. It is defined below.

$$Lift(X \Rightarrow Y) = \frac{\text{confidence } (X \Rightarrow Y)}{\text{support } (Y)}$$

Indicates that if the lift value is greater than 1, X and Y appear more often together. If this value is less than 1, it indicates that X and Y appear together less than expected (17).

2. Conviction: That measure is an alternative of confidence measure. If there is a relationship between the actual occurrences of X without Y, conviction compares the probability that X will do so. In this respect, it is similar to lift, but unlike lift, it also uses knowledge of the absence of the consequence (17,20). It is defined in the form of;

$$Conviction\ (X \Rightarrow Y) = \frac{1 - support\ (Y)}{1 - confidence\ (X \Rightarrow Y)}$$

3. Certainty Factor: Certainty factor is interpreted as a measure of the probability that Y is involved in a transaction when only X is considered and it is described below (15).

$$CF(X \Rightarrow Y) = \begin{cases} \frac{confidence\ (X \Rightarrow Y) - support\ (Y)}{1 - support(Y)} & \text{confidence } (X \Rightarrow Y) > support(Y) \\ \frac{confidence\ (X \Rightarrow Y) - support\ (Y)}{support\ (Y)} & \text{confidence } (X \Rightarrow Y) < support(Y) \\ 0 & \text{in other conditions} \end{cases}$$

4. Cosine: It is similarity measure used widespread for vector space model and it is described below (21).

$$Cosine = \frac{P(X,Y)}{\sqrt{(P(X)P(Y)}}$$

5. Correlation Coefficient (Phi): It is a measure assessing association between X and Y by calculating coefficient of correlation for continuous variables and it can be stated as below (18,21).

$$\varphi(X \Rightarrow Y) = \frac{P(X,Y) - P(X)P(Y)}{\sqrt{P(X)P(Y)(1-P(X))(1-P(Y))}}$$

6. Mutual Information: Mutual information is an entropy-based measure that used to assess dependence between variables. When the value of a second variable is known, it represents the amount of decrease of a variable's entropy (21). Mutual information is employed to boost performance of data mining process (22).

$$\varphi(X \Rightarrow Y) = \frac{P(X,Y) - P(X)P(Y)}{\sqrt{P(X)P(Y)(1-P(X))(1-P(Y))}}$$

RESULTS

Descriptive statistics according to the demographic information and diagnostic test results of the patients are given in Table 2. For continuous variables, mean, standard deviation was given as descriptive stastistic, whereas numbers and percentages was given for categorical variables.

Age, height, weight, BMI, troponin, irisin, LDH and CK-MB continuous variables, they were transformed into categorical variables in order to the generation of association rules. For experimental results, minimum support value was set as 12 % and confidence value was set as 94 %. At the result of analysis, there were 108 rules, including triple and quadruple association rules, which were the most-observed with confidence and support values. It is possible to conclude that the rules, constituted by taking support and confidence measures into consideration, are right; but not so strong. Many interest measures, used in literature, are recommended to obtain stronger rules. In this study, lift, conviction, certainty factor, cosine, coefficient of correlation (phi) and mutual information measures are used in addition to support and confidence. Lift value is referred to how many times X and Y are seen together (23). Lift value, which is over 1, indicates that X and Y are seen more frequently (17). The present study included lift values over 1. Cosine is geometric mean among interest factor and support values. As cosine (X=>Y) approaches to 1, operations with the X item includes Y item, too (17). So values close to 1 are taken into consideration. The correlation coefficient (phi) close to 0 when X and Y are independent (24). Conviction value (X=>Y) gives X's conditional possibility if Y is absentee. It can be commented similarly with lift. As a distinguishing difference, conviction value measures Y' presence possibility when X is present, but the lift take how often X and Y are seen together into consideration. As a result of that, conviction values are used over 1 (17). Certainty factor value approaching to 1 is reckoned that rules, which are obtained during the study, are indicators of fairly high accuracy (15).

Considering support and confidence measures, 108 rules were constituted in the analyses. Those association rules were right; but not strong so that new rules are obtained by using six interest measures. In this respect, 29 of 108 rules are qualified as strong. The rules obtained are given in Table 3.

Generally, in associative classification methods, the rule with the highest confidence is used for classification (25). In this context three examples of interpretation of the rules given in table 3 are given below.

Rule 20: Patients, who with a weight of 80 to 105, a troponin value between 0.01 and 0.48 and male are 100% of the rs726344 polymorphism genotype type is GG wild.

Rule 24: Patients, who height between 170 and 185 with myocardial infarction type acute inferior and rs726344 polymorphism genotype type GG wild are 100% of male.

Rule 25: Patients, who height between 170 and 185, smoke, rs726344 polymorphism genotype type GG wild, and don't have hypertension are 100% of male.

The other rules given in Table 3 will be interpreted similarly.

Table 2. Descriptive statistics of patients according to demographic information and diagnostic test results

Variables		AMI (n=225) (Mean±SD)
(Mean ± SD)		
Age (years)		64.95±12.45
Height		164.94±17.95
Weight		73.87±13.79
BMI (kg/m^2)		26.37±4.52
Troponin (ng/mL)		13.08±18.29
CK-MB (ng/mL)		70.49±66.8
Irisin (ng/ml)		3.12±2.06
LDH (mg/dL)		347.13±154.22
Gender (%)	Male	148 (72.2)
	Female	57 (27.8)
Hypertension (%)		107 (52.2)
Smoking (%)		89 (43.4)
Diabetes (%)		85 (41.5)
Family history (%)		135 (65.9)
Mtype (%)	A. inferior	109 (53.2)
	A. anterior	62 (30.2)
	NSTEMI	34 (16.6)
Rs3480 in FNDC5 gene	AA	59 (28.8)
	AG	111 (54.1)
	GG	35 (17.1)
Rs726344 in FNDC5 gene	GG	180 (87.8)
	GA	15 (7.3)
	AA	10 (4.9)
Rs16835198in FNDC5 gene	GG	108 (52.6)
	GT	86 (42.0)
	TT	11 (5.4)

Table 3. The generated association rules

ID	Rule No	Association Rules (X =>Y)	Confidence	Count	Lift	Support	Certainty Factor	Conviction	Cosine	Mutual Information	Phi
1	1	familyhistory=1. rs3480=3=>rs16835198=1	1	26	1.898	0.127	1	NA	0.491	NA	0.361
2	2	rs3480=3, rs726344=1=>rs16835198=1	1	25	1.898	0.122	1	NA	0.481	NA	0.353
3	3	diabetic=1, rs3480=1=> rs726344=1	1	28	1.139	0.137	1	NA	0.394	NA	0.148
4	4	rs3480=1, rs16835198=2=>rs726344=1	1	31	1.139	0.151	1	NA	0.415	NA	0.157
5	5	hypertension=1, rs3480=1=>rs726344=1	1	36	1.139	0.176	1	NA	0.447	NA	0.172
6	6	mtype=1, rs3480=1=>rs726344=1	1	27	1.139	0.132	1	NA	0.387	NA	0.145
7	7	smoke=2, rs3480=1=>rs726344=1	1	35	1.139	0.171	1	NA	0.441	NA	0.169
8	8	diabetic=2, rs3480=1=>rs726344=1	1	31	1.139	0.151	1	NA	0.415	NA	0.157
9	9	familyhistory=1, rs3480=1=>rs726344=1	1	42	1.139	0.205	1	NA	0.483	NA	0.189
10	10	gender=1, rs3480=1=> rs726344=1	1	43	1.139	0.210	1	NA	0.489	NA	0.192
11	11	age=[27,59), troponin=[0.01,0.48)=>rs726344=1	1	25	1.139	0.122	1	NA	0.373	NA	0.139
12	12	weight=[80,105], troponin=[0.01,0.48)=>rs726344=1	1	29	1.139	0.141	1	NA	0.401	NA	0.151
13	13	age=[27,59), familyhistory=2=>rs726344=1	1	26	1.139	0.127	1	NA	0.380	NA	0.142
14	14	ldh=[389,919], familyhistory=2=>rs726344=1	1	26	1.139	0.127	1	NA	0.380	NA	0.142
15	15	familyhistory=2, rs16835198=2=>rs726344=1	1	27	1.139	0.132	1	NA	0.387	NA	0.145
16	16	height=[170,185], mtype=1=>gender=1	1	34	1.385	0.166	1	NA	0.479	NA	0.277
17	17	hypertension=1, smoke=2, rs3480=1=>rs726344=1	1	27	1.139	0.132	1	NA	0.387	NA	0.145
18	18	hypertension=1, familyhistory=1, rs3480=1=>rs726344=1	1	26	1.139	0.127	1	NA	0.380	NA	0.142
19	19	gender=1, familyhistory=1, rs3480=1=>rs726344=1	1	28	1.139	0.137	1	NA	0.394	NA	0.148
20	20	weight=[80,105], troponin=[0.01,0.48), gender=1=>rs726344=1	1	25	1.139	0.122	1	NA	0.373	NA	0.139
21	21	weight=[80,105], gender=1, rs16835198=2=>rs726344=1	1	25	1.139	0.122	1	NA	0.373	NA	0.139
22	22	height=[170,185], hypertension=2, smoke=1=>gender=1	1	28	1.385	0.137	1	NA	0.435	NA	0.247
23	23	height=[170,185], smoke=1, familyhistory=1=>gender=1	1	28	1.385	0.137	1	NA	0.435	NA	0.247
24	24	height=[170,185], mtype=1, rs726344=1=>gender=1	1	30	1.385	0.146	1	NA	0.450	NA	0.257
25	25	height=[170,185],hypertension=2,smoke=1,rs726344=1=>gender=1	1	26	1.385	0.127	1	NA	0.419	NA	0.237
26	26	height=[170,185], hypertension=2 => gender=1	0.974	38	1.350	0.185	0.908	10.844	0.500	0.104	0.273
27	27	height=[170,185], smoke=1, rs726344=1=>gender=1	0.974	37	1.349	0.180	0.905	10.566	0.493	0.102	0.268
28	30	height=[170,185], hypertension=2, rs726344=1=>gender=1	0.973	36	1.348	0.176	0.903	10.288	0.486	0.100	0.263
29	98	weight=[0,70), gender=1=> bmi=[0,25.1)	0.944	34	2.847	0.166	0.917	12.029	0.687	0.396	0.601

NA*: Not available

DISCUSSION

In conventional association rules mining, rules are constituted if rules meet the measures of support and confidence and their threshold values. However, these rules may be sometimes misleading. So, confidence and support values are not enough while constituting association rules. There are many alternative interest measures for attaining stronger rules in association rules mining. In association rules mining, if it does not take interest measures into consideration it can cause misleading assessments (15). Assessing association rules by using only confidence and support measures might create a lot of disadvantages; so it will be more accurate to assess it with certainty factor, asserted by Shortliffe and Buchanan (26). It demonstrates similar characteristics with the result of study carried out by Berzal and his friends (15). Ilayaraja and Meyyappan in their study aimed at identify frequency of diseases in particular geographical area at given time period with the aid of association rule based Apriori data mining

technique (27). Nahar et al. presented a rule inference on heart disease data using three different association rule mining algorithms: Apriori, Predictive Apriori and Tertius. In their study, they made rule mining analysis by classifying data based on gender. They also found significant risk factors for heart disease for both men and women (28). Sathyavani and Sharmila show that in addition to support and confidence measures, using mutual information measure we can obtain more effective rules (22). Besides, Marimaran et al. suggested that in addition to support and confidence measures, and can discover more interesting rules by applying eight interest measures as lift, chi-square, hyper-lift, hyper-confidence, conviction, coverage, leverage, and cosine (17).

Rules are created when the threshold support and confidence value is provided in the apriori algorithm used in association rules mining. However, these rules are not strong enough. There are many interest measures proposed in the literature to obtain stronger rules. In the experiment of the current study, 108 rules were constituted in order to eliminate misleading rules and obtain stronger rules. It was determined that if lift, conviction, certainty factor, cosine, coefficient of correlation (phi), mutual information measure values were added to those rules 29 strong association rules were generated as a result of more comprehensive analysis.

Similar to this study, Karolis et al. used the different algorithms of rule extraction and evaluation like the C4.5 decision trees and the apriori algorithms for coronary heart events. They found that there were extracted rules with risk factor like sex (male), smoking, high density lipoprotein, glucose, family history, and history of hypertension (29). Safdari et al. employed the data mining techniques for predicting the risk of myocardial infarction. The seven predictive algorithms and one association rule algorithm were applied on the database. The result of the study indicated that variables such as gender, age, family history of coronary artery disease, history of smoking, history of hypertension history of diabetes, history of high blood fat, and body mass index were the most important variables that can be used for predicting the risk of myocardial infarction (30). Licastro et al. assessed a new risk approach for AMI by an innovative algorithm, the predictive capacity of Artificial Neural Networks (ANNs) in consistently distinguishing the two different conditions (AMI or control) and the identification of variables with the maximal relevance for AMI. Their findings showed that ANNs were useful in distinguishing risk factors selectively associated with the disease (31). Known risk factors such as hypertension, smoking, gender, BMI and family history in terms of AMI were found to be among the most important risk factors similar to the literature in terms of 100 % association rules in the present study. In addition, the 20th, 24th and 25th rule associations of polymorphisms of the FNDC5 gene with other physical properties have been shown to have a 100 % risk for the disease. In particular, these data indicated that the association of rs726344 polymorphism wild types in the FNDC5 gene might be a risk factor.

The rs726344 polymorphism does not affect the amino acid sequence of the protein formed because it is located in the intron 5 region of the FNDC5 gene. As with the rs726344 polymorphism, intronic genomic variants can affect gene expression and thus phenotype by altering mRNA stability, alternative mRNA splicing, or binding of transcription factors (32). The minor A-allele of rs726344 is characterized by higher luciferase activity compared to the wild-type G-allele and has recently been associated with decreased insulin sensitivity in vivo (33). The A poliymorphic allele of rs726344 has also been associated with cardiometabolic risk factors (34). However, Staiger et al. (2013) found that there was no relationship between the A allele of the rs726344 SNP and FNDC5 mRNA expression in human myotubes (35). All of the studies summarized above are data obtained from control group comparisons and are case control studies. However, as in the current study, there are no studies in which polymorphisms are associated with other risk factors only within the patient group. In the current study, generally significant rules were found between wild genotypes and other parameters, since the wild allele was found in high density in the patient group. The relationships generally show the inheritance patterns of polymorphic variants or how familial transmission occurs. Rules 1,2,4 and 9 are typical examples of this transmission. As in the 1st rule, rs3480 polymorphic variant and rs16835198 wild variant are transmitted together in our society in individuals with positive family history. In rules 12 and 20, the rs726344 wild genotype was associated with high weight [80,105] and low troponin [0.01,0.48] levels. In rule 11, low age [27,59] and low troponin [0.01,0.48] resulted in rs726344 wild genotype. These data indicate that this genotype poses a risk for AMI at an early age. Similarly, rs726344 wild genotype was associated with hypertension positivity in rule 5. In general, both diabetes and hypertension are known important risk factors for AMI (36). In terms of hypertension, hypertension and non-smoking in rule 17 and hypertension and family history positivity in rule 18 support rule 5.

The dataset used in this article includes AMI patients with FNDC5 polymorphism and their risk factors. To obtain association rules, the apriori algorithm, which is one of the data mining methods, was applied to this data set. With these obtained rules, the interest measures such as lift, conviction, certainty factor, cosine, coefficient of correlation (phi) and mutual information were used, and stronger rules were obtained.. Thanks to the strong rules obtained, the patient profiles of AMI patients were revealed. By using these profiles, it has become easier to determine the clinical profiles of AMI patients in clinical implications.

CONCLUSION

In conclusion, the use of interest measures in the clinical decision making process is very important in terms of eliminating misleading rules and avoiding wrong decisions.

Financial disclosures: *The authors declared that this study hasn't*

received no financial support.

Conflict of Interest: *The authors declare that they have no competing interest.*

Ethical approval: *The study received ethics approval from Firat University's Faculty of Medicine Local Ethics Committee (14.10.2014, Decision No: 17/07).*

REFERENCES

1. Bulduk B, Aktas MC, Bulduk M. Mental disorders developed following acute myocardial infarction. JAREN. 2017;3:24-7.
2. Arslan AK, Colak C, Sarihan ME. Different medical data mining approaches based prediction of ischemic stroke. Comput Methods Programs Biomed. 2016;130:87-92.
3. Chen M-S, Han J, Yu PS. Data mining: an overview from a database perspective. IEEE Transactions Knowledge Data Engineering. 1996;8:866-83.
4. Piateski G, Frawley W. Knowledge discovery in databases: MIT press; 1991;540.
5. Crockett D, Eliason B. What is data mining in healthcare. Insights: Health Catalyst. 2016
6. Jain D, Gautam S. Implementation of apriori algorithm in health care sector: a survey. Int J Computer Science Communication Engineering. 2013;2:22-8.
7. Abdullah, U, Ahmad, J, Ahmed, A. Analysis of effectiveness of apriori algorithm in medical billing data mining. 4th International Conference on Emerging Technologies IEEE. 2008;October:327-31.
8. Zhang, W. J, Ma, D. L, Dong, B. The automatic diagnosis system of breast cancer based on the improved Apriori algorithm. International Conference on Machine Learning and Cybernetics IEEE. 2012;July:63-6.
9. Akbaş KE, Kivrak M, Arslan AK, et al. Assessment of Association Rules based on Certainty Factor: an Application on Heart Data Set. International Artificial Intelligence and Data Processing Symposium (IDAP). 2019;21-2
10. ALPAR R. Applied multivariate statistical methods. 5th Edition. Detay Publishing House, Ankara, 2017, p. 400.
11. Dis O. The investigation of serum irisin levels and genetic variants in patients with acute myocardial infarction. Master's thesis, Firat University, Elazig, 2017.
12. Hornik K, Grün B, Hahsler M. arules-A computational environment for mining association rules and frequent item sets. J Statistical Software. 2005;14:1-25.
13. Team R. RStudio: Integrated development for R (version 1.1. 463)[Computer software]. Boston: RStudio, Inc. http://www.rstudio.com; 2018.
14. Rao S, Gupta P. Implementing improved algorithm over apriori data mining association rule algorithm. IJCST. 2012;3:489-93.
15. Berzal F, Blanco I, Sánchez D, Vila M-A. Measuring the accuracy and interest of association rules: A new framework. Intelligent Data Analysis. 2002;6:221-35.
16. Agrawal, R, Imielinski, T, Swami, A. Mining association rules between sets of items in large databases. ACM SIGMOD international conference on Management of data. 1993;June:207-16.
17. Manimaran J, Velmurugan T. Analysing the quality of association rules by computing an interestingness measures. Indian J Science Technology. 2015;8:1-12.
18. Han J, Pei J, Kamber M. Data mining: concepts and techniques. 3nd edition. Elsevier, 2011.
19. Kiran A, Reddy K. Selecting a right interestingness measure for rare association rules. Management Data. 2010;115:115-24.
20. Brin S, Motwani R, Ullman JD, Tsur S. Dynamic itemset counting and implication rules for market basket data. Acm Sigmod Record. 1997;26:255-64.
21. Tan P-N, Kumar V, Srivastava J. Selecting the right objective measure for association analysis. Information Systems. 2004;29:293-313.
22. Sathyavani D, Sharmila D. Optimized Weighted Association Rule Mining using Mutual information on Fuzzy Data. Int J Scientific Research Management (IJSRM). 2014;2:501-5.
23. Zhang C, Zhang S. Association rule mining: models and algorithms: Springer-Verlag; 2002
24. Hahsler M. A probabilistic comparison of commonly used interest measures for association rules. United States Southern Methodist University. 2015
25. Li, W, Han, J, Pei, J. CMAR: Accurate and efficient classification based on multiple class-association rules. International conference on data mining IEEE, 2001; November:369-76.
26. Shortliffe EH, Buchanan BG. A model of inexact reasoning in medicine. Rule-based expert systems. 1984:233-62.
27. Ilayaraja M, Meyyappan T. Mining medical data to identify frequent diseases using Apriori algorithm. International Conference on Pattern Recognition, Informatics and Mobile Engineering IEEE, 2013;February:194-9.
28. Nahar J, Imam T, Tickle KS, Chen Y-PP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems Applications. 2013;40:1086-93.
29. Karaolis M, Moutiris JA, Papaconstantinou L, et al. Association rule analysis for the assessment of the risk of coronary heart events. Annu Int Conf IEEE Eng Med Biol Soc.2009;2009:6238-41.
30. Safdari R, Saeedi MG, Arji G, et al. A model for predicting myocardial infarction using data mining techniques. Iranian J Med Informat. 2013;2:1-6.
31. Licastro F, Ianni M, Ferrari R, et al. A New Risk Chart for Acute Myocardial Infarction by a Innovative Algoritm. HEALTHINF; 2015
32. Sasabe T, Furukawa A, Matsusita S, et al. Association analysis of the dopamine receptor D2 (DRD2) SNP rs1076560 in alcoholic patients. Neuroscience letters. 2007;412:139-42.
33. Böhm A, Weigert C, Staiger H, Häring H-U. Exercise and diabetes: relevance and causes for response variability. Endocrine. 2016;51:390-401.
34. Badr EA, Mostafa RG, Awad SM, et al. A pilot study on the relation between irisin single-nucleotide polymorphism and risk of myocardial infarction. Biochemistry and biophysics reports. 2020;22:100742.
35. Staiger H, Böhm A, Scheler M, et al. Common genetic variation in the human FNDC5 locus, encoding the novel muscle-derived 'browning'factor irisin, determines insulin sensitivity. PloS one. 2013;8:e61903.
36. Dülek H, Tuzcular Vural E, Gönenç I. Kardiyovasküler hastalıklarda risk faktörleri. J Turkish Family Physician. 2018;9:53-8.