

# Klinik Veritabanlarında Etkin Veri Madenciliği Amacıyla Çoklu-ilişkisel Veri Madenciliği Yaklaşımı

Pınar YILDIRIM<sup>a</sup>, Ü.Erkan MUMCUOĞLU<sup>b</sup>, Mehmet R. TOLUN<sup>a</sup>,  
Mustafa N. ÖZMEN<sup>c</sup>, Mahmut ULUDAĞ<sup>d</sup>

<sup>a</sup>Bilgisayar Mühendisliği Bölümü, Çankaya Üniversitesi, Ankara

<sup>b</sup>Enformatik Enstitüsü Tıp Bilişimi AD, Orta Doğu Teknik Üniversitesi, Ankara

<sup>c</sup>Hacettepe Üniversitesi Hastaneleri, Ankara

<sup>d</sup>European Bioinformatics Institute, Cambridge, UK

## Multi-Relational Data Mining Approach for Effective Data Mining in Clinical Databases

### **Abstract**

*Today's hospitals and clinics keep large amounts of data about patients and their medical conditions. These data contains hidden knowledge and they are important resources for data mining studies. Despite many successful applications of data mining studies in several domains, medical domain is difficult area for these studies. One of the factors affecting on this problem is the quality of medical data and the other one is the representation of the data. Medical data may include repeated or missing values and also is usually represented in relational structures and stored in multi-relational database systems. In this study, all of these challenges are explained in Percutaneous Cysts Hydatid treatment database used in Interventional Radiology Department in Hacettepe University Hospitals. Multi-relational databases consist of multiple tables and complex relations. Conventional data mining algorithms look for hidden patterns in a single table or flat file. These algorithms can handle single-table data and do not provide an effective way for hospital databases. Usually, manual pre-processing and data transformation are required before analysis of data. Therefore they can not directly work on relational databases. This limitation leads to some problems such as loss of information and redundancy. This paper represents a detailed explanation of how to improve the quality of medical data by using preprocessing techniques and highlights relational data mining approaches for effective data mining that are different from conventional approaches.*

### **Key Words:**

*Clinical Data Mining, Multi-relational Data Mining, Knowledge Discovery in Medical Data*

## Özet

Günümüzde hastaneler ve klinikler, hastalara ve onların tıbbi durumlarına ait yoğun veri birikimine sahiptirler. Bu veriler aynı zamanda birçok gizli bilgiler içerirler ve veri madenciliği çalışmaları için önemli araştırma kaynaklarıdır. Veri madenciliği çalışmaları birçok alanda başarıyla uygulanmalarına karşın tıbbi verilerin analizi için uygulanmaları biraz daha zordur. Buna etki eden en önemli faktörlerden birincisi verilerin kalitesi, diğeri de verilerin yapısıdır. Tıbbi alandaki veriler gürültülü olmalarının yanı sıra, eksik veya tekrarlanmış olabilirler. Ayrıca bu veriler ilişkisel bir yapıdadırlar ve çoklu ilişkisel veritabanlarında saklanır ve yönetilirler. Bu çalışmada, bütün bu zorluklar Hacettepe Üniversitesi Hastaneleri Girişimsel Radyoloji Bölümünde kullanılan Perkütan Kist Hidatik tedavisi veritabanı üzerinde açıklanmıştır. Geleneksel veri madenciliği algoritmaları sadece tek bir tablo yapısındaki verilere uygulanabilmekte ve hastane veritabanları için uygun yöntem değildirler. Genellikle veriler analiz edilmeden önce önışlemlerden geçirmeli ve bazı dönüşümler yapılmalıdır. Bu işlemler bazı verilerin kaybolmasına ve bazı verilerin ise tekrar etmesine neden olur. Bu çalışmada tıbbi verilerin önışlemlerden geçerek kalitesinin artırılması ve geleneksel veri madenciliği algoritmalarından farklı olan çoklu-ilişkisel veri madenciliği yöntemleri açıklanmıştır.

## Anahtar Kelimeler:

Kliniksel Veri Madenciliği, Çoklu-ilişkisel Veri Madenciliği, Tıbbi Verilerde Bilgi Keşfi

## 1. Giriş

Günümüzde sağlık alanında birçok veri madenciliği çalışması yapılmaktadır. Bunlardan bazıları, örneğin tıbbi tedaviden sonra hastaların durumlarının tahmin edilmesi veya belirli bir hastalığa sahip hastaların ortak özelliklerinin tahmin edilmesi gibi çalışmalardır. Örneğin Harleen Kaur ve arkadaşları şeker hastalarına ait bazı bilgileri, örneğin yaş, cinsiyet ve bazı tetkik sonuçlarını kullanarak, veri madenciliği çalışması yapmışlar ve bu hastalığın sonuçlarını tahmin etmeye çalışmışlardır [1]. Başka bir çalışmada ise kafa yaralanması teşhisi ile tedavi edilen hastaların tedavi öncesi, tedavi zamanı ve sonrası tıbbi durumlarına ait bilgiler kullanılarak veri madenciliği çalışması yapılmış ve tedavinin başarılı olup olmaması konusunda tahmin yapılmaya çalışılmıştır [2].

Bu çalışmada, önerdiğimiz yeni yaklaşımın gösterilmesi amacıyla, Hacettepe Üniversitesi (HÜ) Hastaneleri'nde çalışan radyologlarla görüşülüp Girişimsel Radyoloji bölümünde gerçekleştirilen Perkütan Kist Hidatik Tedavisi (Percutaneous Cyst Hydatid Treatment) veritabanı oluşturulması ve ilgili veri madenciliği çalışması ön hazırlığı sunulmuştur.

### 1.1. Problem Tanımı ve Veriseti

Kist Hidatik(Cyst Hydatid) hastalığı bazı ülkelerde özellikle Ortadoğu, Güney Amerika, Yeni Zelanda ve Türkiye gibi küçükbaş hayvancılığın yaygın olduğu ülkelerde sık görülen, önemli bir sağlık sorunudur. Tedavi sonucunda hastalığın yeniden oluşması ve uzun süre hastanede kalış gibi olaylar gözlemlenir. Son yıllarda Perkütan Kist Hidatik Tedavisinin oldukça başarılı olduğu gösterilmiş ve günümüzün en etkili ve güvenilir tedavi yöntemlerinden birisi haline gelmiştir [3]. Ayrıca bu tedavide tedavi sonrası görülen komplikasyonlar da kliniğin başarısına önemli ölçüde etki eder ve başarıyı artırır veya azaltır. Radyologların bu tedavinin güvenilir, kolay uygulanabilir ve etkili bir metot olduğunu değerlendirebilmeleri ve başarıyı artırabilmeleri için tedavi sonrası komplikasyonları tahmin edebilecekleri bir sisteme ihtiyaçları vardır. Kistin tipi, hacmi ve hastanın yaşı, cinsiyeti gibi özelliklerle tedavi sonrası komplikasyonlar arasında gizli ilişkiler olabilir. Bazen hastalarda kistler yeniden oluşabilir ve tedavi sonrası takip edilmeleri gerekir. Bu tedavinin uygulandığı hastalarda tedavi sonrası sonuçları tahmin eden sınıflandırma yöntemine dayalı veri

madenciliği çalışması tasarlanmıştır. Veriseti olarak HÜ Hastaneleri Girişimsel Radyoloji bölümünde bu tedavinin uygulandığı hastalara ait veritabanı seçilmiştir (Tablo 1,2,3,4,5).

*Tablo-1. Hasta Tablosu*

Hasta_ID (BA)	Yaş	Cinsiyet
1	12	Erkek
9	27	Erkek
11	41	Kadın
13	36	Kadın

*Tablo 2. Perkutan Kist Hidatik Tablosu*

İşlem_ID (BA)	Hasta_ID (YA)	Teknik	Kist Lokasyonu	Kist Tipi	Kist Hacmi (cm <sup>3</sup> )	Komplikasyon
2	1	PAIR	L	I	55	Ateş 39 C
7	9	8F Kat.	R	III	86	Ateş
9	1	6F Kat.	LR	NULL	400	Ürtiker
11	9	PAIR	R	IV	352	NULL
13	9	6F Kat.	L	II	100000	Allerji
14	11	7F Kat.	R	I	350	Rek.

**BA:**Birincil Anahtar

**YA:**Yabancı Anahtar

## 2. Tıbbi Veri Madenciliğinin Zorlukları

Diğer alanlarda yapılan veri madenciliği çalışmaları ile karşılaştırıldığında, tıp alanındaki veri madenciliği çalışmaları birçok farklı özelliğe sahiptir. Bunlardan en önemlisi sağlığın önemli bir unsur olması ve birçok verinin detaylı açıklanmasına ihtiyaç duymasındır. Çoğu kez klinik verilerin veya raporların yeniden üretilmesi için tedavinin veya operasyonun tekrar edilmesi mümkün değildir. Ayrıca veriler birçok belirsizliğin etkisiyle üretilmiş olabilirler. Tıbbi verilerde birçok ölçüm hatası veya kayıp bilgi olabilir. Veri madenciliği çalışması bütün bu problemlerle karşı karşıyadır ve beklenti odur ki, veriler için uygun model yapısı sunabilmelidir [4]. Bütün bu zorluklara rağmen, klinik veritabanları büyük bilgi kaynaklarını oluştururlar. Verilerin kalitesi ve yapısı veri madenciliği çalışmasını etkileyen önemli faktörlerdir [4]. Verilerin öncelikle önışlemeden geçmesi gerekir ve bu aşama veri madenciliği çalışmasının başarımını doğrudan etkiler. Diğer alanlarla karşılaştırıldığında, tıp alanında bu süreç daha çok uzman bilgisi ve zamanı gerektirir. Bu nedenle hatalı verilerin düzeltilmesi ve eksik verilerin tamamlanması oldukça zordur.

Diğer bir önemli fark, hastane ve kliniklerde saklanan veriler, ilişkisel yapıdadırlar ve ilişkisel veritabanı sistemlerinde saklanır ve yönetilirler. Verilerin büyük bir kısmı tek bir tabloda

saklanmadıkları için, geleneksel veri madenciliği algoritmalarının uygulanması zordur. Bu nedenle ilişkisel veri madenciliği yaklaşımı tercih edilebilir.

## 2.1. Veri Kalitesi ile İlgili Problemler

Günümüzde tıp alanındaki veritabanlarının birçoğunda çok sayıda gürültülü, kayıp ve tutarsız veri vardır. Temizlenmiş ve tamamlanmış veriler yapılan araştırmaların hem etkinliğini artırır hem de kolaylaştırır. Genellikle kullanılan veri önleme yöntemleri şunlardır:

- Veri temizleme, gürültülü veriyi ortadan kaldırır ve tutarsız verileri tutarlı ve doğru hale getirir.
- Verilerin entegrasyonu, birçok verinin birleştirilmesini sağlar.
- Veri dönüştürme verilerin farklı formlara dönüştürülmesini veya normalleştirilmesini sağlar.
- Veri azaltma verilerin boyutlarının azaltılmasını veya tekrar eden özelliklerin çıkarılması gibi işlemleri içerir [5].

Tıbbi verilerde en çok görülen problem verilerdeki bilinmeyen değerlerdir. Buna neden olan faktörler şunlardır.

- Veri unutulmuş veya kayıp olabilir
- Veri girişi yapan kişi uzman değildir ve hangi değeri girmesi gerektiğine karar verememiştir. Bu kişi ancak uzmana danışarak değeri girebilir.
- Veri girişi yapan kişi dikkat etmemiş veya önemsememiş olabilir.
- Verinin ne olacağı hakkında belirsizlik olabilir[4].

Bu problemleri açıklamak için Perkütan Kist Hidatik Tedavisi veritabanını incelersek, örneğin Tablo 2’de bazı değerler “NULL” olarak girilmiştir. Tedavi işlemi sonrasında veri girişi yapan personel kistin hacmine ait değerleri veritabanına girmiş olabilir ancak komplikasyon bilgisini girmemiş de olabilir. Çünkü bu kullanıcı uzman doktor olmayabilir ve hastada komplikasyon olup olmadığını ancak uzman doktora danışarak belirleyebilir. Ayrıca kistin tipinin ne olduğu hakkında bilgi unutulmuşsa bu bilgi de ancak uzman doktorun radyolojik görüntüleri yeniden yorumlamasıyla elde edilebilir.

Genelde eksik veya kayıp bilgilerin geçmişe yönelik yeniden elde edilmesi oldukça zordur. Çünkü veri değerlerinin yeniden elde edilebilmesi için yapılan işlem veya tedavinin yenilenmesi gerekir ve bu da hemen hemen mümkün değildir. Ayrıca işlem çok uzun zaman önce uygulanmış ve kullanılan yöntemler zaman içinde değişmiş olabilir [6].

Tıbbi verilerde tutarsız değerler olması da ayrı bir problemdir. Örneğin Tablo 2’de postoperatif komplikasyonlar alanında bazı hastalar için “ateş 39” bazıları için de sadece “ateş” değerleri girilmiştir. İkinci veri sadece hastanın ateşi olduğunu belirtirken birinci veri de ayrıca ölçüm değeri de girilmiştir. Bu iki açıklama aynı anlamdadır ve aynı şekilde ifade edilmelidir. Bazen tıbbi verilerde çok büyük değerlere de rastlanmaktadır. Örneğin Tablo 2’de bir hastanın kist hacmi “1 000 000” girilmiştir. Bu değer hiç şüphesiz yanlıştır ve gerçek kist değerini tanımlayamayacak kadar büyüktür. Bu tutarsızlıklar verilerin kalitesini düşürmekte ve veri madenciliği çalışmasının güvenilirliğini azaltmaktadır [6].

Veri Madenciliği sembolik ve kategorize edilmiş verilerde iyi sonuç verebilmektedir. Bazı sürekli verilerin de kategorize edilerek önlemeden geçmeleri gerekir [4]. Örneğin Perkütan

Kist Hidatik Tedavisi veritabanındaki yaş, kist hacmi gibi nitelikler kategorize edilirse bilgi keşfi için daha iyi sonuçlar alınabilir.

## 2.2. Perkütan Kist Hidatik Tedavi Verilerinin Önişlenmesi

Hacettepe Üniversitesi Hastaneleri Girişimsel Radyoloji Bölümünden elde edilen verilerle oluşturduğumuz veritabanı 1991-2008 yılları arasında takip edilen hastaların demografik bilgileri ve tedaviye ait bilgilerden oluşmaktadır. Veritabanı 686 hastaya ait kimlik ve tedavi bilgisinden oluşmaktadır. Tedavi ve komplikasyonlara ait veriler, örneğin kistin tipi Gharbi sınıflandırılmasına göre girilmiştir ve herhangi bir dönüşüm işlemine ihtiyaç yoktur. Gharbi, kistleri ultrasonografi görüntülerine göre farklı sınıflara ayıran bir sınıflandırma yöntemidir [7]. Bazı alanlar ise yaş ve kist hacmi gibi kategorize edilmeleri gerekmektedir. Sınıflandırma işlemi için literatür taraması yapılmış ve radyologlara danışılmıştır. Yaş bilgisi hastanın çocuk veya yetişkin olması durumuna göre, kist hacmi de 50-3550 arasında belirlenen 10 değere kategorize edilmiştir.

Tablo-3. Gharbi Sınıflandırmasına Göre Kist Tipleri [7].

Kist Tipi	Gharbi Sınıflandırması
Tip I	İçeriği sadece sıvıdan oluşan kist
Tip II	Germinatif membranın perikistten ayrıldığı kist görünümü (nilüfer çiçeği görünümü)
Tip III	Kist içerisinde kız kistlere bağlı septa görünümleri (balpeteği görünüm)
Tip IV	Sıvı içeriğini kaybetmiş yün yumağına benzer heterojen içerikli dejenere kist
Tip V	Kalsifiye duvar

Tablo-4. Yaş Aralıkları

< 16
>=16

Tabl- 5. Komplikasyonlar

Rekürens
Abse
Enfeksiyon
Ürtiker
Ağrı
Ateş
Kız vezikül
Allerji

### 3. Geleneksel Veri Madenciliği Algoritmalarının Dezavantajları

Geleneksel veri madenciliği yöntemlerindeki ana yapı sabit nitelik setidir. Nesnelere tek bir nitelik-değer çifti (tek bir tablo) olarak tanımlanıp, matris yapısında gösterilirler. Verilerin bu şekilde veri-ambarına dönüştürülmesinin birçok dezavantajı vardır. Bazı durumlara nesnelere tek bir tablo ile ifade edilemezler [8]. Çünkü nesnelere farklı bölümlerde olabilir ve bu bölümler farklı sayıda ve hacimde olabilirler. Bu nedenle nesnelere tek bir nitelik-değer şekline dönüştürülmeleri orijinal yapısına zarar verebilir.

Örneğin Perkütan Kist Hidatik Tedavisi veritabanında geleneksel veri madenciliği çalışması için iki tabloyu birleştirdiğimizi varsayalım (Tablo 6). Burada bir hasta birden çok tedavi olmuş olabilir ve her seferinde hastanın adı, yaşı ve cinsiyeti gibi demografik bilgileri tekrar eder. Bu dönüşüm veri dosyasının hacmini artırır ve hem diskte hem de bellekte fazla yer kaplamasına neden olur. Perkütan Kist Hidatik Tedavisi veritabanı MySQL DBMS de 44 KB hacminde yer tutarken, tek bir tablo haline dönüştürüldüğü zaman 123 KB olmuştur. Ayrıca Provost F. ve Kolluri V yaptıkları çalışmada müşteri ve ürün bilgilerinden oluşan bir veritabanı 100 MB hacme sahipken, düz dosya haline dönüştürüldüğünde 2.5 GB olduğunu belirtmişlerdir. Ayrıca, düz dosya yapısında, bazı veriler tekrar ettiği için güncelleme ve silme işlemleri de risklidir.

Tablo- 6. Hasta ve Perkütan Kist Hidatik Tablolarının Birleşmiş Görüntüsü

Hasta_ID	Yaş	Cins.	Islem_ID	Tek.	Kist Lok.	Kist Tipi	Kist Hac.(cc)	Komp.
1	12	Erkek	2	PAIR	L	I	55	Ateş 39 C
1	12	Erkek	9	6F Kat.	LR	NUL L	400	Ürtiker
9	27	Erkek	7	8F Kat.	R	III	86	Ateş
9	27	Erkek	11	PAIR	R	IV	352	NULL
9	27	Erkek	13	6F Kat.	L	II	1000 00	Allerji
11	41	Kadın	14	7F Kat.	R	I	350	Rek.
13	36	Kadın	NULL	NULL	NULL	NULL	NULL	NULL

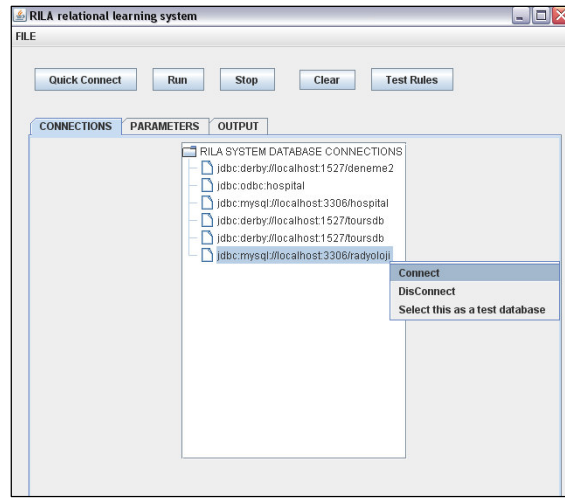
Verilerin dönüşüm işlemine farklı bir yöntem uygulanarak tekrarlar belki önlenir. Örneğin hastaların demografik bilgilerinin yanına kaç kez tedavi olduklarına ait özet bilgi yazılabilir [Tablo 7]. Bu şekilde tekrarlar ortadan kaldırılabilir ancak bu durumda tedaviye ait bilgiler yok olur.

Tablo-7. Özet Bilgi İle Hasta Tablosu

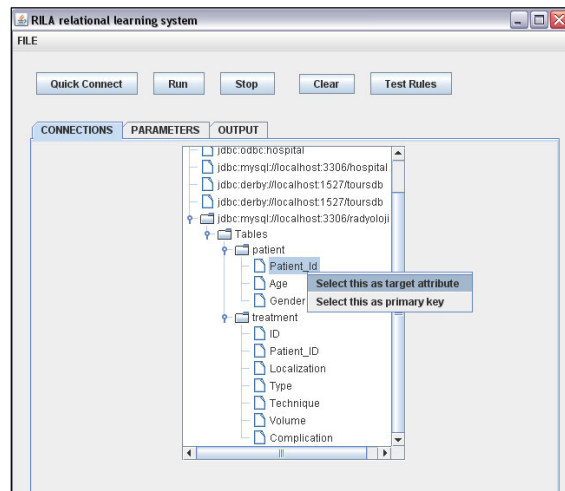
Hasta_ID	Yas	Cinsiyet	Tedavi Sayısı
1	12	Erkek	2
9	27	Erkek	3
11	41	Kadın	1
13	36	Kadın	0

#### 4. Çoklu İlişkisel Veri Madenciliği Yaklaşımı

Birçok tablo ve ilişkinin analizi sonucu yapılan veri madenciliği yöntemi Çoklu İlişkisel Veri Madenciliği olarak tanımlanır. Bu algoritmalar doğrudan ilişkisel veritabanı üzerinde çalışır ve veriler için herhangi bir dönüşüme ihtiyaç duymazlar [11]. Genellikle Veritabanı Yönetim Sistemlerinin sorgulama, SQL fonksiyonları ve Meta-Data yapısını kullanırlar. RILA (Relational Inductive Learning Algorithm), bir çoklu ilişkisel veri madenciliği algoritmasıdır [12]. Şekil 1 ve 2 RILA kullanıcı arabirimini göstermektedir. Veritabanı bağlantısı ve tablo seçimi sihirbazı kullanıcının kolaylıkla ilgili tabloları ve hedef niteliği(class attribute) seçerek öğrenme işlemi başlatmasını sağlar. Öğrenme aşaması veritabanına Meta-Data sorguları gönderilip nitelik(attribute) ve birincil/yabancı anahtar(Primary/Foreign Key) bilgilerinin alınmasıyla oluşur. Bu aşamadan sonra SQL sorgularıyla hipotezler oluşturularak, bu hipotezlerin olasılıklarına göre kurallar elde edilir [12].



Şekil-1. RILA Veritabanı Bağlantısı [11]



Şekil-2. RILA Hedef Nitelik(target attribute) Seçimi [11]

Çoklu İlişkisel Veri Madenciliği yaklaşımının diğer bir avantajı sonuç bilgisinin ilişkisel yapıda gösterilebilmesidir [13-14]. Perkütan Kist Hidatik Tedavisi veritabanını ele alırsak, hem geleneksel hem de çoklu ilişkisel veri madenciliği yaklaşımıyla verileri analiz etmiş olsaydık. Örneğin

Geleneksel yaklaşım( nitelik-değer) kuralı:

IF Yaş=27 AND Cinsiyet="Erkek" AND Kist Tipi ="II" THEN Komplikasyon= "Ateş"

Çoklu-ilişkisel yaklaşım kuralı:

IF Hasta.Yaş=27 AND Hasta.Cinsiyet="Erkek" AND PCH-Tedavi.Kist Tipi ="II" THEN PCH-Tedavi.Komplikasyon= "Ateş"

gibi kurallar elde edebiliriz. Burada görüldüğü gibi, kistin hangi hastalık veya tedaviye ait olduğu net bir şekilde anlaşılmamaktadır.

İlişkisel yaklaşım büyük veritabanları için efektif ve ölçeklenebilir çözümler sunar. Örneğin Provost ve Kolluri [9] öğrenme algoritmalarının ölçeklenebilirliği üzerine çalışmalar yapmışlar ve şu üç faktörü önermişlerdir:

- Hızlı algoritma
- Verinin küçük bölümlere ayrılarak analiz edilmesi
- İlişkisel yaklaşımın kullanılması

Ayrıca, birçok geleneksel veri madenciliği algoritması tek bir tablodan oluşan dosyanın tamamını belleğe yükler ve onun üzerinde çalışır. Oysaki ilişkisel veri madenciliği algoritmaları veritabanı yönetim sisteminin etkili veri sorgulama ve getirme yöntemlerini kullandığı için bütün veriyi bellekte tutmasına gerek yoktur [9].

## 5. Çoklu İlişkisel Veri Madenciliğindeki Güçlükler

Organizasyonlar ve kurumlar kişilere, işlemlere ve bilimsel çalışmalara ait verileri veritabanı sistemlerinde saklar ve yönetirler. Günümüzde bu verilerin hacimleri oldukça büyüktür. Veri madenciliği araştırmacıları bu yapılara uygun çoklu-ilişkisel veri madenciliği algoritmalarını uygulamak ve bilgi keşfi yapmak için çalışırlar. Bu algoritmaların çok büyük veritabanları üzerinde efektif bir şekilde çalışabilmeleri araştırma konusudur [15]. Örneğin, bir çoklu ilişkisel veri madenciliği algoritması veritabanına bir takım aday sorgular gönderip, sonra da bunların her biri için frekansı hesaplar. Tek bir sorgu birçok ilişki içerebilir ve sorgunun çalışması geleneksel yöntemlere göre daha karmaşıktır. Bu işlem çok fazla CPU zamanı harcayıp, çok daha fazla bellek ihtiyacı gerektirebilir.

## 6. Sonuç

Klinik verileri genellikle ilişkisel yapılarda olup, ilişkisel veritabanlarında saklandıkları için geleneksel yaklaşım uygun bir çözüm olmayabilir. Eğer çoklu-ilişkisel veri madenciliği yaklaşımı kullanılırsa, verilerin tek bir tablo haline dönüşümüne gerek kalmadan, doğrudan veritabanı üzerinde çalıştırılabilmeleri mümkündür. Böylece, geleneksel yaklaşımdaki verilerin tekrar etmesi veya kayıp olması gibi negatif etkiler önlenmiş olur. Ancak, bu yeni yaklaşımın çok büyük veritabanları üzerinde efektif bir şekilde çalışabilmesi araştırma konusudur. Bundan sonraki çalışmalarımızda, hastane veri-tabanından ve hasta dosyalarından



oluşturacağımız daha çok veri seti üzerinde geleneksel ve çoklu-iliskisel veri madenciliği yöntemleri uygulanacak ve detaylı karşılaştırmalar yapılacaktır.

## 7. Kaynakça

- [1] Kaur H , Wasan S., Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science 2(2) 2006.
- [2] Pilih A. I., Mladenic D., Lavrac H., Tine S. Prevec, Using Machine Learning for Outcome Prediction of Patients with Severe Head Injury, Tenth IEEE Symposium on Computer Based Medical Systems 1997.
- [3] Akhan O., Özmen MN., Percutaneous treatment of liver hydatid cysts, European Journal of Radiology , Volume 32 , Issue 1 , SS 76 – 85, 1999.
- [4], Bellazi R., Zupan B., Predictive Data Mining in Clinical Medicine:Current issues and guidelines, International Journal of Medical Informatics 2006.
- [5] Han J., Kamber M., Data Mining Concepts and Techniques, 2001 Academic Press, 105.
- [6] Zizka j., Hudik T., Machine Learning-Based Knowledge Extraction from Complex Clinical Oncological Data, SIS 2003.
- [7] Chen Y., Yeh T., Tseng J., Huang S., Ynlin D., Hepatic Hydatid Cysts with Superinfection in a Non-Endemic Area in Taiwan, Am. J. Trop. Med. Hyg., 67(5), 2002, ss. 524–527
- [8] Knobbe A.J., Multi Relational Data Mining, Dissertation, Utrecht University, 2004.
- [9] Provost F. and Kolluri V., A Survey of Methods for scaling Up Inductive Algorithms, Kluwer academic Publishers, Boston 1999.
- [10] Atramentov A., Leiva H., Honavar V., A Multi-Relational Decision Tree Learning Algorithm -Implementation and Experiments, In: Proceedings of the ACM/SIGKDD Workshop on Multi-Relational Decision Tree Learning. KDD-2002.
- [11] Toprak S., Data Mining for Rule Discovery in Relational Databases, MS Thesis, Middle East Technical University, 2004.
- [12] Mahmut Uludağ, Mehmet R. Tolun, A New Relational Learning System Using Novel Rule Selection Strategies, Knowledge-Based Systems, Volume 19 , Issue 8 , 2006, S Yass: 765-771.
- [13] Struyf J., Fast Relational Data Mining, Industry-ready Innovative Research: 1st Flanders Engineering PhD Symposium, Brussels, Belgium, December 11, 2003.
- [14] Kovalerchuk B., Vityaev E., Yusupov H., Symbolic Methodology in Numeric Data Mining:Relational Techniques for Financial Applications, Computational Engineering, Finance, and Science, cs.CE/0208022, 20 s. 2002.
- [15] Domingos R., Prospects and Challenges for Multi-Relational Data Mining, SIGKDD Explorations, vol. 5 (ss. 80-83), 2003.