

# Siğil Tedavisinde Kullanılan Immunotherapy Yönteminin Uygunluğunun Bayes Yöntemi ile Tespiti

Sümeyye Çelik<sup>1\*</sup>, Melike Şişeci Çeşmeli<sup>1</sup>, İhsan Pençe<sup>1</sup>, Adnan Kalkan<sup>1</sup>

<sup>1</sup> Burdur Mehmet Akif Ersoy University, \* Corresponding author, sumeyye.celik334@gmail.com

## Özet

Siğil hastalığı tıp alanında sıkça rastlanan vakalar arasındadır. Özelliklerine göre birkaç farklı yöntem ile tedavisi mümkündür. Uygulanacak tedavi yöntemi seçilirken literatürde kabul görmüş özelliklere göre en uygun olduğu düşünülen yöntem seçilmektedir. Bu çalışmada, literatürde belirlenen özelliklere göre; tedavi yöntemlerinden olan immunotherapy yönteminin, hastaya uygulanıp uygulanmaması konusunda veri madenciliği yöntemleri ile ön bir değerlendirme yapılmış ve değerlendirme başarı oranının artırılması sağlanmıştır. Böylece hekime tedavi yöntemini seçerken, immunotherapy yöntemini seçip seçmeme konusunda daha doğru karar vermesi için yardımcı olunabilecektir. Başarı oranının artırabilmek için veri seti üzerinde birçok yöntem denenmiştir. Gözlemlenen sonuçlara göre en yüksek başarı oranı, bayes net ile yapılan sınıflandırmada %85.55 olarak görülmüştür. Çalışmada kullanılan veri seti UCI machine learning repository yayınlanan, İran Ghaem Hastanesinin dermatoloji kliniğinde yapılan bilimsel bir araştırmanın sonuçlarıdır. Bayes net algoritması WEKA (Waikato Environment for Knowledge Analysis) uygulaması ile gerçekleştirilmiştir.

**Anahtar kelimeler:** Siğil tedavisi, bayes net, immunotherapy, veri madenciliği.

## Title

Determination of the Suitability of Immunotherapy Method Used in the Wart Treatment Using Bayes Method

## Abstract

Wart disease is a common occurrence in medicine. According to their characteristics, treatment with several different methods is possible. When the treatment method to be applied is selected, the method which is deemed most appropriate is selected according to the features accepted in the literature. In this study, according to the characteristics determined in the literature, a preliminary evaluation was made with the data mining methods on the immunotherapy method applied to the patient and the evaluation success rate was increased. This will help the doctor to decide whether to choose the immunotherapy method when choosing a treatment method. Various methods have been tested on data sets in order to increase the success rate. According to the observed results, the highest success rate was 85.55% in bayes net classification. The data set used in the study is the results of a scientific research conducted in the dermatology clinic of the Iranian Ghaem Hospital, published in the UCI machine learning repository. The Bayesian net algorithm was implemented using the Waikato Environment for Knowledge Analysis (WEKA).

**Keywords:** Wart treatment, Bayes net, Immunotherapy, Data mining.

**Citation:** Çelik, S., Çeşmeli, M. Ş., Pençe, İ., Kalkan, A. (2018, October) *Siğil Tedavisinde Kullanılan Immunotherapy Yönteminin Uygunluğunun Bayes Yöntemi ile Tespiti*. Paper presented at the Fifth International Management Information Systems Conference.

**Editor:** H. Kemal İltar, Ankara Yıldırım Beyazıt University, Turkey

**Received:** August 19, 2018, **Accepted:** October 18, 2018, **Published:** November 10, 2018

**Copyright:** © 2018 IMISC Çelik et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

# SİĞİL TEDAVİSİNDE KULLANILAN IMMUNOTHERAPY YÖNTEMİNİN UYGUNLUĞUNUN BAYES YÖNTEMİ İLE TESPİTİ

**Özet:** Siğil hastalığı tıp alanında sıkça rastlanan vakalar arasındadır. Özelliklerine göre birkaç farklı yöntem ile tedavisi mümkündür. Uygulanacak tedavi yöntemi seçilirken literatürde kabul görmüş özelliklere göre en uygun olduğu düşünülen yöntem seçilmektedir. Bu çalışmada, literatürde belirlenen özelliklere göre; tedavi yöntemlerinden olan immunotherapy yönteminin, hastaya uygulanıp uygulanmaması konusunda veri madenciliği yöntemleri ile ön bir değerlendirme yapılmış ve değerlendirme başarı oranının artırılması sağlanmıştır. Böylece hekime tedavi yöntemini seçerken, immunotherapy yöntemini seçip seçmeme konusunda daha doğru karar vermesi için yardımcı olunabilecektir. Başarı oranının artırabilmek için veri seti üzerinde birçok yöntem denenmiştir. Gözlemlenen sonuçlara göre en yüksek başarı oranı, bayes net ile yapılan sınıflandırmada %85.55 olarak görülmüştür. Çalışmada kullanılan veri seti UCI machine learning repository yayınlanan, İran Ghaem Hastanesinin dermatoloji kliniğinde yapılan bilimsel bir araştırmanın sonuçlarıdır. Bayes net algoritması WEKA (Waikato Environment for Knowledge Analysis) uygulaması ile gerçekleştirilmiştir.

**Anahtar kelimeler:** Siğil tedavisi, bayes net, immunotherapy, veri madenciliği

## DETERMINATION OF THE SUITABILITY OF IMMUNOTHERAPY METHOD USED IN THE WART TREATMENT USING BAYES METHOD

**Abstract:** Wart disease is a common occurrence in medicine. According to their characteristics, treatment with several different methods is possible. When the treatment method to be applied is selected, the method which is deemed most appropriate is selected according to the features accepted in the literature. In this study, according to the characteristics determined in the literature, a preliminary evaluation was made with the data mining methods on the immunotherapy method applied to the patient and the evaluation success rate was increased. This will help the doctor to decide whether to choose the immunotherapy method when choosing a treatment method. Various methods have been tested on data sets in order to increase the success rate. According to the observed results, the highest success rate was 85.55% in bayes net classification. The data set used in the study is the results of a scientific research conducted in the dermatology clinic of the Iranian Ghaem Hospital, published in the UCI machine learning repository. The Bayesian net algorithm was implemented using the Waikato Environment for Knowledge Analysis (WEKA).

**Keywords:** Wart treatment, bayes net, immunotherapy, data mining

### 1. Giriş

Bilişim teknolojilerindeki gelişmeler neticesinde bu alandaki ürün ve hizmetlerin daha ulaşılabilir ve ucuz olması beraberinde veri boyutlarının da artmasına sebep olmaktadır. Verinin içinde saklı olan bilginin varlığı da çeşitli yöntemlerin doğmasına sebep olmuştur. Veri madenciliği ham veri içindeki bilgiye ulaşma ihtiyacından ortaya çıkmış, boyutları büyüyen veriler arasından gizli kalmış örüntülerin keşfedilmesini sağlayan yöntemlerdir (Hong ve Wu, 2011). Veri miktarı büyüdükçe aralarındaki ilişkiler çıplak gözle fark edilemeyecek kadar karmaşık bir hale bürünmekte ve özel algoritmalar kullanılmaktadır. Veri madenciliği sağlık, biyoloji, genetik, kimya, sigortacılık, telekomünikasyon, finans, metin madenciliği, internet madenciliği, pazarlama, sosyal bilimler, davranış bilimleri, gibi birçok alanda kullanılmaktadır (Talan, 2016).

Gelişen teknoloji ile birlikte Tıp alanında uygulanan yöntemler teknolojiye bağımsız düşünülemez bir hal almıştır (Tsumoto vd., 2017). Veri madenciliğinin de tıp alanındaki kullanımı oldukça yaygındır (Moskov, 2016). Söz konusu insanın sağlığı olduğu için Tıp alanında yapılan çalışmalar çok fazla öneme sahiptir. Teşhis konulurken hızlı ve doğru kararlar alınması gerekmektedir. Aksi durumda hastalık ilerleyebilmekte ya da tedavi süreci uzayabilmektedir. Bu durum hem hekim hem de hasta için zorlu bir süreçtir ve istenilen bir durum değildir (Coulter vd., 2001). Veri madenciliği tıp alanında, hastalıkların erken teşhis edilmesi ve hastalığı etkileyen önemli faktörlerin ortaya çıkarılması gibi birçok alanda kullanılmaktadır.

Tıp alanında veri madenciliği yöntemleri literatürde sıklıkla kullanılmaktadır (Göreke, 2014). Bu şekilde hekime yardımcı olacak sistemler geliştirilmekte ve daha verimli çalışmalarına yardımcı olunmaktadır (Akdemir, 2016). Veri madenciliğinin tıbbi araştırmalarda her alanda kullanılmakla birlikte, en önemli alanlardan biri de deri hastalıklarıdır (Khozeimeh vd., 2017).

Siğiller, çoğu insanı etkileyen, iyi huylu tümörler olarak tanımlanır (Çokluk, 2014). Özellikle eller ve ayaklarda olmak üzere vücudun her yerinde görülebilmektedir. Bulaşıcı olabilmekte ve bazı durumlarda kanserojen etkilere sahip olabilmektedir. Siğiller yaratabilecekleri sağlık sorunların yanı sıra bireylerin dış görünüşlerinde deformasyonlara sebep olduğundan psikolojik rahatsızlıklara da sebep olabilmektedir. Bu sebeplerden dolayı etkili ve kalıcı bir tedavinin önemi artmaktadır (Erdem, 2012). Siğillerin oluşmasında birçok sebep olmakla birlikte en belirgin sebep literatürde Human Papillomavirus Virus (HPV) mikrobi olarak gösterilmektedir (Çokluk, 2014).

Siğil hastalığı için çeşitli tanı ve tedavi yöntemleri vardır. Kullanılacak yöntemler hastaların özelliklerine göre farklılık göstermekte ve tedavide kesin bir sonuç alınamadığı durumlar da olabilmektedir. Bazı durumlarda siğiller yeniden nüksedebilmekte ve tedavi süreci uzamaktadır. Bu sebepten dolayı hasta üzerinde hangi tedavinin daha etkili olacağı belirlenmeli ve tedavi uygulanmalıdır (Khozeimeh vd., 2017).

Bu çalışmanın amacı veri madenciliği yöntemleri kullanılarak siğil tedavisi uygulanacak olan hastaya uygun tedavi yöntemin seçilmesinde hekime yardımcı olacak bilgi çıkarımlarının yapılmasıdır. Siğil tedavisinde kullanılan birçok yöntem bulunmaktadır. Bu yöntemlerden bir tanesi de immunotherapy yöntemidir.

Çalışmada kullanılan veri seti 2017 yılında UCI machine learning repository’de yayınlanan (<https://archive.ics.uci.edu>), İran Ghaem Hastanesinin dermatoloji kliniğinden elde edilmiştir. Güncel olan veri seti üzerinde immunotherapy yönteminin uygulanmasından sonra tedavinin başarılı olup olmayacağının sınıfı tespit edilmeye çalışılmıştır. Böylece immunotherapy tedavisi uygulanmadan önce ortaya çıkabilecek sonuçlar öngörülebilmektedir. Tahminlenen sonuç da hastanın tedaviye vereceği cevap pozitif olarak öngörülür ise hastaya uygulanacak olan tedavi yöntemi olarak immunotherapy yönteminin tercih edilebileceği ortaya konulmuş olacaktır. Eğer hastanın immunotherapy yöntemine karşı vücudunun vereceği tepki negatif olarak öngörülür ise o zaman hastaya uygulanacak olan yöntemin immunotherapy yöntemi olmadığına karar verilmiş olacaktır. Bu durumda hastaya uygulanacak olan tedavinin diğer tedavi yöntemlerinden seçilmesi sonucuna varılacaktır. Dolayısı ile bu çalışma hekime siğil tedavisi konusunda hastaya uygun tedavi yönteminin immunotherapy yöntemi olup olmadığına karar vermesi için yardımcı olacaktır.

Siğil tedavisinde uygulanacak yönteme karar verilirken incelenecek olan öznitelikler literatürde kabul görmüş olan şekli ile kullanılmıştır (Khozeimeh vd., 2017). Kullanılan öznitelikler Hastanın cinsiyeti, yaşı, zaman, siğilin sayısı, tipi, alanı, çap kalınlığı şeklindedir. Bu öznitelikler sonucunda tahmin edilmek istenen sınıf etiketi siğilin immunotherapy tedavi yöntemine karşı verdiği tepkinin olumlu ya da olumsuz olmasıdır.

Veri seti İran'daki Ghaem Hastanesinin dermatoloji kliniğine başvuran ve siğil tedavisi olmak isteyen 180 hastadan alınan bilgilerle oluşturulmuştur. Bu 180 hastanın 90 tanesi immunotherapy yöntemi diğer 90 tanesi de cryotherapy ile yöntemi ile tedavi edilmiştir ve tedaviye verilen tepkiler kaydedilmiştir. Veri seti kullanılan tedavi yöntemlerine göre ikiye ayrılmıştır. Tedavi yöntemine verilen yanıtları öngörmek için bulanık mantık kural tabanlı bir uzman sistem önerilmiş ve uygulanmıştır. Immunotherapy ve cryotherapy yöntemlerinin tahmin doğruluğunun sırasıyla %83.33 ve %80.70 olduğu gözlemlenmiştir (Khozeimeh vd., 2017).

Bu çalışmada daha önceki başarı oranı arttırılmaya çalışılmıştır. Başarı oranının arttırılabilmesi için veriler önışlemden geçirilmiş ve Yeni Zelanda’da ki Waikato Üniversitesi’nde geliştirilmiş, ücretsiz, açık kaynak kodlu WEKA programı kullanılarak birçok farklı sınıflandırma yöntemi denenmiştir. Etkisinin daha fazla olabileceği düşünülen bazı özniteliklerin ağırlıkları arttırılmıştır. Bu şekilde tahminlemenin doğruluk oranı arttırılmaya çalışılmış ve istenen başarı artışı sağlanmıştır.

## 2. Veri Madenciliğinin Tıp Alanında Kullanımı

Tıp alanında ki teknolojik gelişmeler ile daha fazla veri üretilmekte ve saklanmaktadır (Peixoto vd., 2017). Tıp alanındaki verilerin gün geçtikçe artması ve hayati önem taşıyan veriler olması, veri madenciliği yöntemlerinin tıp alanında oldukça fazla kullanılmasına sebep olmaktadır (Moskov, 2016). Özellikle Elektronik hasta kayıtları (EPR) gibi sistemlerde tüm karar verme süreçlerinin saklanması, birçok anlamlı bilginin çıkarılabileceği ortamı oluşturmaktadır (Tsumoto vd., 2017). Tıptaki veriler insan hayatı ile ilgili olduğu için tıp verileri oldukça önemli bir çalışma alanı oluşturmakta ve yapılan çalışmalar birçok fayda sağlamaktadır (Reddy vd., 2012). Bu faydalardan başlıcaları şöyle sıralanabilmektedir; Hastalığın teşhisi ve tedavisinde oluşabilecek hataların önüne geçmek (Bytyçi vd., 2017), maliyetleri azaltmak, süreci kısaltmak (Sağıroğlu, 2016) ve daha doğru kararların verilmesinde hekime yardımcı olacak sistemler geliştirmektir (Tsumoto vd., 2017). Literatürde veri madenciliği yöntemleri kullanılarak yapılan çalışmaların bir kısmı aşağıda listelenmiştir.

Singh ve Sivabalakrishnan (2015), farklı öznitelik seçimi algoritmaları ile makine öğrenimi tabanlı gen ifadesi üzerine çalışmışlardır. Yapılan sınıflandırmanın yanında baskın olarak kullanılan öznitelik seçimlerinin üstünde durulmuştur.

Vanitha ve arkadaşları (2015), Destek Vektör Makinesi (SVM) kullanılarak gen sınıflandırması için etkili bir yöntem sunmaktadır. Genler ve sınıf etiketi arasındaki karşılıklı bilgi (MI), bilgilendirici genleri tanımlamak için kullanılmıştır. Seçilen genler, SVM sınıflandırıcısıyı eğitmek için kullanılmıştır. Diğer yaklaşımlarla karşılaştırıldığında sınıflandırma doğruluğunun arttığı ve girdi özelliklerinin boyutunu azalttığı belirtilmiştir.

El-Bialy arkadaşları (2015), Koroner arter hastalığını (KAH) hedef alan farklı veri setlerine uygulanan makine öğrenme analizinin sonuçlarının bir entegrasyonu gerçekleştirmişlerdir. Veri setine karar ağacı uygulanmış ve budamalar gerçekleştirilmişlerdir. Yapılan çalışmalar soncunda sınıflandırma doğruluğunun daha yüksek olduğunu belirtmişlerdir.

Khalilabad ve arkadaşları (2016), meme kanseri tipinin saptanması için bir çalışma yapmıştır. Yapılan çalışmada genlerin kesin yerini belirlemek için görüntü işleme teknikleri kullanılmıştır. Elde edilen veriler karar ağacı kullanılarak analiz edilmiştir ve meme kanseri tipi %95.23'lük bir doğrulukta teşhis edilmiştir.

Tarek ve arkadaşları (2017), gen ekspresyon profilleri için etkili bir analize ulaşmak için veri madenciliği yöntemleri, makine öğrenme algoritmaları ve istatistiksel yöntemler kullanarak kanser sınıflandırması yapmışlardır. Oluşturulan bu yeni sistemin öncekilerden daha hızlı ve daha fazla kanser türünü kapsadığını belirtmişlerdir.

Peixoto ve arkadaşları (2017), Yoğun Bakım Ünitelerinin (YBÜ'lerin) performansını iyileştirmek ve sağlık profesyonellerine, hastaları hakkında karar vermede yardımcı olmak amacıyla hastaneden temin edilen veriler üzerinde veri madenciliği tekniklerini uyguladıklarını ve başarı oranını %90 bulduklarını belirtmişlerdir.

Mohebian ve arkadaşları (2017), çalışmalarında 5 yıllık meme kanseri nüks tahmini için Parçacık Sürü Optimizasyonu (PSO), karar ağacı, yapay sinir ağı yöntemlerini kullanarak başarı oranlarını karşılaştırmışlardır.

Wang ve arkadaşları (2018), Çin'in sağlık durumunu değerlendirmek için anket yoluyla topladıkları verileri, veri madenciliği yöntemlerinden lojistik regresyon modeli ile analiz etmişlerdir. Veri analizi sonuçları, 2017 yılında Çinli programcıların sağlıksızlık sorununun temel olarak metabolik bozukluklar, depresyon, baskı, memnuniyet ve cinsel yaşamlarının kalitesinden kaynaklandığını göstermiştir.

Wu ve arkadaşları (2018), iki tip şeker hastalığı (T2DM) tahmini için veri madenciliği tekniklerine dayalı yeni bir model önermiştir. Uygulama kısmında ön işleme adımları uygulandıktan sonra k-means ve lojistik regresyon algoritması kullanılmıştır. Daha önceki çalışmalara göre yüzde 3.04 başarı artışı sağlamışlardır.

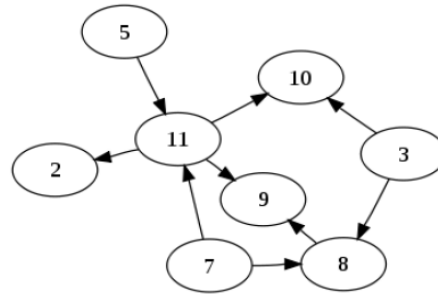
Peng ve arkadaşları (2018), Hastalık-10'uncu revizyon kodlarının (ICD-10) idari sağlık verilerine ilişkin durumunu ortaya koyan bir çalışma yapmıştır. Sunulan kodlama ve ilişkilendirme kuralları ile, veri kalitesi değerlendirmesi için kural madenciliğinin kullanımıyla ilgili gelecekteki araştırmalar için bir referans olabileceği vurgusu yapılmıştır.

### **3. Materyal ve Metot (Bayes net)**

Bayes ağları olasılık hesaplamalarına dayalı veri modelleme yöntemlerinden bir tanesidir (Cihan ve Kalıpsız, 2015). Bayes ağları bayes teoreminden ortaya çıkmış olan bir yöntemdir. Bayes'in olasılık kuralını kullanan, grafiksel modellemeye dayalı bir yöntemdir (Kartal, 2012; Çığışar, 2017). Öznitelikler arasındaki ilişkileri ortaya çıkarmak için kullanılmaktadır. Oluşturulan model nodes (düğüm) ve edge (kol) den oluşmaktadır (Sigurdsson vd., 2001). Bayes ağlarında her bir düğüm öznitelikleri temsil etmektedir. Kollar ise öznitelikler arasındaki olasılıkları göstermektedir (Cihan ve Kalıpsız, 2015). Böylece belirlenen parametrelerin sisteme olan etkileri tek tek ölçülebilmektedir (Kurum, 2013). Bayes ağları oluşturulurken bazı kabuller yapılarak başlanmaktadır. Öncelikle çıkış parametresine etkisi olan öznitelikler belirlenmekte, sonra bu özniteliklerin her birinin sonuca etki değeri istatistiksel yöntemlerle hesaplanarak ağırlıklandırılmaktadır (Cihan ve Kalıpsız, 2015; Bulut, 2016). Böylece oluşturulan model ile, sonuca etki değeri fazla olan özniteliklerin ağırlık değerleri

daha fazla, sonuca etki değeri daha az olan özneliliklerin ise ağırlık değerleri daha az olmaktadır (Kartal, 2012). Belirlenen bu ağırlıklara göre sınıflandırma işlemi gerçekleştirilmektedir (Ediş ve Ulaş, 2017).

Bayes ağları yönlü dönüşsüz ağlardır (directed acyclic network) (Efe, 2018). Bayes ağlarında ilişkili olan değişkenler birbirlerine oklarla bağlıdır ve oklar tek yönlüdür (Bulut, 2016). Okların yönü aynı zamanda ilişkininde yönünü göstermektedir. Düğümler arasındaki kolların geçişleri istatistiksel bir takım kabuller ile belirlenmektedir. Okların çıktığı kaynak düğümlere Parent Nodes (ebeveyn düğüm), okların gösterdiği hedef düğümlere Child Nodes (çocuk düğüm), kendilerine hiçbir ok gelmeyen düğümlere ise Root Nodes (kök düğüm) denilmektedir (Trucco vd., 2008).



Şekil 1: Bayes ağları diyagramının örnek bir görüntüsü (Dereli, 2012).

Şekil 1’de de görüldüğü gibi bayes net yöntemi ile birçok ebeveyn ve birçok çok düğümden oluşan diyagramlar oluşturmak mümkündür (Dereli, 2012). Bayes ağı niteliksel ve niceliksel olarak iki ana yapıdan meydana gelmektedir. Niteliksel kısımda değişkenler arası olası ilişkilerin diyagramı oluşturulmaktadır. Niceliksel tarafta ise niteliksel tarafta oluşturulan diyagramlardan yararlanılarak olasılıklar hesaplanmaktadır (Lee vd., 2009).

Bayes net modeli oluşturulurken öznelilikler arası ve öznelilikler ile etiketler arasındaki ilişkiler, koşullu olasılık yardımı ile keşfedilmektedir. Kullanılan koşullu olasılık formülü şekil 2 de gösterilmiştir. Öğrenmenin gerçekleştirileceği veri kümesindeki her bir örneğin n boyutlu bir uzayda tanımlı olduğu düşünüldüğünde;  $X=(x_1,x_2,x_3,...,x_n)$  ve veri seti içerisinde toplamda m adet sınıf var ise  $C_1,C_2,C_3,...,C_m$  gibi bu kümeler içerisindeki en büyük  $P(C_i|X)$  oranı bulunmaya çalışılmaktadır. Eşitlik (1)’deki gibi hesaplanan  $P(C_i|X)$  değerinde, tüm sınıflar için  $P(X)$  olasılığı sabittir ve olasılığı en büyük olan değer aranırken  $P(X)$  olasılığı kullanılır (Kartal, 2012).

$$P(C_j | x) = \frac{p(x | C_j)P(C_j)}{p(x)} = \frac{p(x | C_j)P(C_j)}{\sum_k p(x | C_k)P(C_k)} \quad (1)$$

Eşitlik (1)'de  $P(Cj)$ :  $j$  sınıfındaki bir durumun ilk olasılığı,  $P(X|Cj)$ :  $j$  sınıfındaki bir durumun  $X$  olma olasılığı,  $P(X)$ : herhangi bir durumun  $X$  olma olasılığı,  $P(Cj|X)$  ise  $X$  olan bir durumun  $j$  sınıfından olma olasılığıdır (Kartal, 2012).

Bu araştırmada kullanılan veriler İran'daki Ghaem Hastanesinin dermatoloji kliniğinde yapılan bilimsel bir araştırmanın sonuçlarıdır. Veri seti Ocak 2013'ten Şubat 2015'e kadar Meşhed'deki Ghaem hastanesinin dermatoloji kliniğinde toplanmıştır. Belirlenen öznitelikler, tedavi için önemli faktörler olarak hekimlerin görüşüne göre seçilmiştir (Khozeimeh vd., 2017). Veri setinin özellikleri tablo 1 de detaylı olarak gösterilmiştir.

Tablo 1: Kullanılan öznitelikler ve özellikleri (<https://archive.ics.uci.edu>, (21.03.2018))

| Öznitelik numarası | Öznitelik adı       | Min Değer | Max Değer | Ortalama |
|--------------------|---------------------|-----------|-----------|----------|
| 1 (input)          | Sex                 | 1         | 2         | 01.544   |
| 2 (input)          | Age                 | 15        | 56        | 31.044   |
| 3 (input)          | Time                | 1         | 12        | 07.231   |
| 4 (input)          | Number_of_Warts     | 1         | 19        | 06.144   |
| 5 (input)          | Type                | 1         | 3         | 01.711   |
| 6 (input)          | Area                | 6         | 900       | 95.007   |
| 7 (input)          | Induration_Diameter | 2         | 70        | 14.333   |
| 8 (output)         | Result_of_Treatment | 0         | 1         | 00.788   |

#### 4. Bulgular ve Değerlendirme

Bu çalışmada veri madenciliği yöntemleri kullanılarak, siğil tedavisi uygulanacak olan 90 hastaya immunoherapy yönteminin uygulanıp uygulanmaması gerektiğinin sınıf niteliği tahmin edilmeye çalışılmıştır. Sınıflandırma amacıyla kullanılan yöntemlerin başarı oranları değerlendirilerek, en yüksek başarı oranının bayes net yönteminde elde edildiği tespit edilmiştir. Modelin başarı oranı hesaplanırken Cross Validation yöntemi kullanılmıştır. Test verisindeki başarı oranı hesaplanırken Cross Validation oranı %10 şeklinde ayarlanmıştır.

Sonuç ile alakalı olan her bir öznitelik sonuç etiketi ile ilişkilendirilmiştir. Bu ilişkilendirme aracılığı ile ebeveyn düğüm ile her bir çocuk düğüm için aralarındaki olasılık değeri ayrı ayrı hesaplanmıştır. Yapılan bu hesaplama sonuçlarına göre özniteliklerin ağırlıkları hesaplanmıştır.

Bayes net ile yapılan sınıflandırmanın sonuçları aşağıdaki gibidir.

|                                  |    |           |
|----------------------------------|----|-----------|
| Correctly Classified Instances   | 77 | 85.5556 % |
| Incorrectly Classified Instances | 13 | 14.4444 % |

|                            |              |
|----------------------------|--------------|
| Kappa statistic            | 0.5578       |
| Mean absolute error        | 0.226        |
| Root mean squared error    | 0.3433       |
| Total Number of Instances  | 90           |
| Time taken to build model: | 0.04 seconds |

=== Detailed Accuracy By Class ===

|               | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|----------|-------|
|               | 0.632   | 0.085   | 0.667     | 0.632  | 0.649     | 0.706    | 0     |
|               | 0.915   | 0.368   | 0.903     | 0.915  | 0.909     | 0.706    | 1     |
| Weighted Avg. | 0.856   | 0.308   | 0.853     | 0.856  | 0.854     | 0.706    |       |

=== Confusion Matrix ===

a b <-- classified as

12 7 | a = 0

6 65 | b = 1

Sonuçlarda da görüldüğü gibi Total Number of Instances değeri 90'dır. Yani bayes net algoritması ile toplamda 90 hastanın bulunduğu veri seti üzerinde çalıştırılmıştır. Cross Validation oranı %10 şeklinde ayarlandıktan sonra test verisi üzerinde yapılan doğru sınıflandırmanın oranı Correctly Classified Instances değeri 77 bulunarak, %85.5556 şeklinde görülmüştür. Yanlış sınıflandırılan veri sayısı ise, Incorrectly Classified Instances değeri 13 bulunup, %14.4444'tür. Verilerin sınıflandırması ile alakalı olarak detaylı bilgiye Confusion Matrix'den bakıldığı zaman; sınıf etiketi 0 olup doğru sınıflandırılan 12 tane veri, sınıf etiketi 1 olup ve doğru sınıflandırılan 65 tane veri olduğu görülmektedir. Yanlış sınıflandırılan verilerin detaylarına bakıldığı zaman 0 etiketli sınıfa ait olduğu halde 1 etiketi olan sınıfa dahil edilen 7 tane veri, sınıf etiketi 1 olduğu halde sınıf etiketi 0 olan sınıfa dahil edilen 6 tane veri olduğu görülmektedir. Modelin oluşturulma ve sınıflandırma işleminin gerçekleşmesi 0.04 saniye sürmüştür. Bu süre yeterince kısa ve ideal bir süre olarak değerlendirilebilir.

## 5. Sonuç ve Öneriler

Bu çalışmada veri madenciliği yöntemleri kullanılarak siğil tedavisi uygulanacak olan hastaya uygun tedavi yöntemin seçilmesinde hekime yardımcı olacak tahminlemeler de bulunulmuştur. Yapılan çalışma ile en iyi tedavi yöntemini seçmede hekimlere yardımcı olmanın yanı sıra hastalara zaman kazandırmak, tedavi maliyetini düşürmek ve tedavi kalitesini iyileştirmek gibi birçok fayda sağlanacaktır.

Bu tahminleme, siğil tedavisi uygulanacak olan hastanın özelliklerine göre immunotherapy tedavi yöntemi uygulandığında tedavinin başarılı olup olmayacağı şeklindedir. Kullanılan veri seti daha önceden yapılan bir çalışmada kullanılmıştır ve %83.33 başarı oranında bir tahminlemede bulunulmuştur (Khozeimeh vd., 2017). Bu çalışmadaki temel hedef başarı oranı arttırmak olmuştur. Başarı oranını arttırmak için birçok yöntem denenerek veri ön işleme adımı üzerinde yoğunlaşmıştır. Ön işleme süreci tamamlandıktan sonra literatürde kabul görmüş sınıflandırma tekniklerinin birçoğu denenerek sınıflandırma işlemi gerçekleştirilmiştir. Sonuçlar incelenmiş ve bayes net tekniği ile yapılan sınıflandırmada başarı oranının %85.55 olarak arttığı görülmüştür. Böylece uygulanacak tedavi yöntemi seçilirken daha doğru karar alınmasına yardımcı olunacaktır.

Yapılan çalışmada veri setindeki kısıtlardan dolayı, tedavi yöntemlerinden sadece bir tanesi (immunotherapy yöntemi) ele alınabilmektedir. Sınıf etiketinin immunotherapy için uygun olmadığı tespit edildiğinde hastaya uygulanacak olan tedavinin diğer tedavi yöntemlerinden hangisi olduğu belirlenemeyecektir. Çünkü veri seti içerisinde gözlemlenen hastaların vücudunun diğer yöntemler için verdiği tepkiler mevcut değildir. Diğer tedavi yöntemleri için de bir veri seti oluşturulabilir. Böylece siğil tedavisi uygulanacak olan hasta için en iyi tedavi yönteminin hangisi olacağına karar vermek mümkün olacaktır.

Bunun yanı sıra yapılacak çalışmalar için önerilen diğer bir çalışma alanı ise veri setleri oluşturulurken görüntü işleme tekniklerinin kullanılması olacaktır. Böylece siğillerin büyüklüğü, kalınlığı gibi değerler elde edilirken daha doğru, kolay ve hızlı ölçümler sağlanabilir. Aynı zamanda siğillerin özellikleri belirlenirken insan gözüyle değil de bilgisayarlı sistemlerle gerçekleştirileceği için fark edilmeden ortaya çıkabilecek problemlerin de önüne geçilmiş olacaktır.

## Kaynakça

- Erdem, Y. (2012). Genital siğilli hastalarda hasta ve aile yaşam kalitesinin değerlendirilmesi. İstanbul: İstanbul Üniversitesi Tıp Fakültesi.
- Wang, B., Yang, S., Huang, Z., Li, D., He, J., & Wang, X. (2018). Data mining-based subhealth analysis of Chinese software programmers in 2017. *Informatics in Medicine Unlocked*, 10, 134-142
- Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, 10, 100-107.
- Peng, M., Sundararajan, V., Williamson, T., Minty, E. P., Smith, T. C., Doktorchik, C. T., & Quan, H. (2018). Data on coding association rules from an inpatient administrative health data coded by International classification of disease-10th revision (ICD-10) codes. *Data in Brief*, 18, 710-712.
- Khalilabad, N. D., Hassanpour, H., & Abbaszadegan, M. R. (2016). Fully automatic classification of breast cancer microarray images. *Journal of Electrical Systems and Information Technology*, 3(2), 348-359.

- Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. *Computational and structural biotechnology journal*, 15, 75-85.
- Singh, R. K., & Sivabalakrishnan, M. (2015). Feature selection of gene expression data for cancer classification: a review. *Procedia Computer Science*, 50, 52-57.
- Tarek, S., Elwahab, R. A., & Shoman, M. (2017). Gene expression based cancer classification. *Egyptian Informatics Journal*, 18(3), 151-159.
- Vanitha, C. D. A., Devaraj, D., & Venkatesulu, M. (2015). Gene expression data classification using support vector machine and mutual information-based gene selection. *procedia computer science*, 47, 13-21.
- El-Bialy, R., Salama, M. A., Karam, O. H., & Khalifa, M. E. (2015). Feature analysis of coronary artery heart disease data sets. *Procedia Computer Science*, 65, 459-468.
- Peixoto, R., Ribeiro, L., Portela, F., Santos, M. F., & Rua, F. (2017). Predicting Resurgery in Intensive Care-A data Mining Approach. *Procedia Computer Science*, 113, 577-584.
- Bytyçi, E., Ahmedi, L., & Lisi, F. A. (2017). Enrichment of Association Rules through Exploitation of Ontology Properties–Healthcare Case Study. *Procedia Computer Science*, 113, 360-367.
- Tsumoto, S., Kimura, T., Iwata, H., & Hirano, S. (2017). Mining Text for Disease Diagnosis. *Procedia Computer Science*, 122, 1133-1140.
- Reddy, S. K. & Kodali, S. R. & Gundabathina, J. L. (2012). Classification of vertebral column using naïve bayes technique. *International Journal of Computer Applications*, 58 (7), 38-42.
- Hong, T. P., ve Wu, C. W. (2011). Mining rules from an incomplete dataset with a high missing rate. *Expert Systems with Applications*, 38(4), 3931-3936.
- Akdemir, Ç. (2016). Hilenin veri madenciliği ile ortaya çıkartılması ve perakende sektöründe bir uygulama. İstanbul: Marmara üniversitesi sosyal bilimler enstitüsü.
- Moskov, D. (2016). Knowledge extraction from published papers in literature for the catalytic methanol production from synthesis gas using data mining tools. İstanbul: Boğaziçi üniversitesi.
- Sağiroğlu, Ç. (2016). Türkiye'nin iller bazında enerji ihtiyacının veri madenciliği ile analizi ve lisanssız yenilenebilir enerji yatırımları için karar destek önermesi. Adana: Çukurova üniversitesi fen bilimleri enstitüsü.
- Coulter, D. M., Bate, A., Meyboom, R. H., Lindquist, M., ve Edwards, I. R. (2001). Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study. *Bmj*, 322(7296), 1207-1209
- Göreke, V., Uzunhisarcıklı, E., ve Güven, A. (2014). Gri Seviyeli Eşoluşum Matrisleri Kullanılarak Sayısal Mamogram Görüntüsünden Doku Öz niteliklerinin Çıkarılması ve Yapay Sinir Ağı ile Kitle Tespiti. *Tıp Teknolojileri Ulusal Kongresi-TıpTekno'14*.
- Khozeimeh, F., Alizadehsani, R., Roshanzamir, M., Khosravi, A., Layegh, P., & Nahavandi, S. (2017). An expert system for selecting wart treatment method. *Computers in biology and medicine*, 81, 167-175.
- Trucco, P., Cagno, E., Ruggeri, F., & Grande, O. (2008). A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Reliability Engineering & System Safety*, 93(6), 845-856.
- Lee, E., Park, Y., & Shin, J. G. (2009). Large engineering project risk management using a Bayesian belief network. *Expert Systems with Applications*, 36(3), 5880-5887.

- Cihan, P., & Kalıpsız, O. (2015). Öğrenci Proje Anketlerini Sınıflandırmada En Başarılı Algoritmanın Belirlenmesi Determining The Most Successful Classification Algorithm For The Student Project Questionnaire.
- Dereli, B. (2012). Bayes ağları ile gelecek odaklı konumlandırma analizi: oyun konsolları pazarında bir uygulama, istanbul teknik üniversitesi, fen bilimleri enstitüsü, İstanbul.
- Kartal, H.B. (2012). Envanter sınıflandırmada yapay öğrenme yöntemlerinin kullanımı ve destek vektör makineleri ile bir uygulama, istanbul teknik üniversitesi, fen bilimleri enstitüsü, istanbul.
- Kurum, F. M. (2013). A study of a group of healthcare datasets in data mining domain, boğaziçi university, istanbul.
- Çığışar, B. (2017). Kredi risklerinde veri madenciliği sınıflandırma algoritmaları, çukurova üniversitesi, fen bilimleri enstitüsü, adana.
- Efe, B. (2018). İş güvenliğinde risk değerlendirme için karar destek modeli önerileri, gazi üniversitesi, fen bilimleri enstitüsü, ankara.
- Bulut, K. (2016). Trijenerasyon sistemlerinin modellenmesine yönelik yeni bir yaklaşım: bir üniversite uygulaması, istanbul teknik üniversitesi, fen bilimleri enstitüsü, istanbul.
- Ediş, S., & Ulaş, E. (2017). Çankırı Acıçay-Tatlıçay Havzalarında arazi kullanım türlerinin Bayes Ağları yöntemiyle tahmin edilmesi. Turkish Journal of Forestry, 18(3), 212-218.
- Sigurdsson, J. H., Walls, L. A., & Quigley, J. L. (2001). Bayesian belief nets for managing expert judgement and modelling reliability. Quality and Reliability Engineering International, 17(3), 181-190.