



Sınıflama Ağaçları Yardımıyla Restless Legs Syndrome (RLS) Hastalarına Tanı Koyma

Gülhan Oreki Temel*, Handan Çamdeviren**, Zeki Akkuş***

*Mersin Üniversitesi, İktisadi İdari Bilimler Fakültesi, İşletme Bölümü, Mersin

**Mersin Üniversitesi, Tıp Fakültesi, Biyoistatistik AD. Mersin

***Dicle Üniversitesi, Tıp Fakültesi, Biyoistatistik AD. Diyarbakır

Sınıflama ağaçları (Classification Trees, CT) parametrik olmayan istatistiksel bir yöntemdir. Bu yöntem bir veya daha fazla risk faktöründen yararlanarak bireylere tanı koyma amacıyla kullanılan bir ağaç algoritmasıdır. Tıbbi araştırmalarda özellikle son yıllarda bu yöntemin kullanılabilirliğinde artış izlenmesine karşın bir çok hastalığın risk faktörlerinin incelenmesinde henüz bu yaklaşım kullanılmamıştır.

Bu çalışmanın amacı; CT hakkında önemli teorik bilgileri özetlemek ve RLS (Restless Legs Syndrome) hastalarının risk faktörlerini farklı bir yaklaşımla incelemektir. Bu amaçla, Mersin Üniversitesi, Tıp Fakültesi, Nöroloji bölümünün 206 denek üzerinde yaptığı anket çalışmasının sonuçları kullanılmış ve deneklerin RLS hastası olup olmama durumunu belirleyen değişkenler sınıflama ağaçları analizi ile tespit edilmiştir.

Analiz sonuçlarına göre, RLS hastalığını belirleyen değişkenler literatürde yer alan pek çok risk faktörüyle paralellik göstermektedir.

Anahtar Kelimeler: Sınıflama ve regresyon ağaçları, Tanı, Karar ağaçları, Hatalı sınıflama, Huzursuz bacak sendromu

Diagnosing Restless Legs Syndrome (RLS) Patients with Help of Classification Tree

Classification trees (CT) are nonparametric statistical methods. This method is a trees algorithm, performed to diagnose the disease with the use of one or more risk factors. Although, in recent years availability of this method especially for medical researches has increased, it is not common for the investigation of numerous risk factors of the disease.

The aim of the study is to summarize theoretical knowledge related to CT and to investigate Restless Legs Syndrome (RLS) risk factors with different approach. For this purpose, 206 patients, with whom a questionnaire form was practiced in the University of Mersin, Faculty of Medicine, Department of Neurology, were included in this study and, CT analysis is used for determination of variables whether these patients have RLS or not.

According to the analysis results, the variables which determine RLS disease was parallel with many risk factors reported in literature.

Key Words: Classification and regression trees (CART), Diagnose, Decision trees, Misclassification error, Restless Legs Syndrome.

Bilimsel çalışmalardan elde edilen verilerin analizinde sınıflama (diskriminant, lojistik regresyon, kümeleme analizleri gibi) ve regresyon modelleri sıkça kullanılmaktadır. Ancak bu tür modellerin gerektirdiği varsayımlar pek çok alanda istatistiksel analiz imkanlarını kısıtlamaktadır. İncelenen veri seti üzerinde hiçbir varsayım gerektirmemesi nedeniyle, Sınıflama ve Regresyon Ağaçları (CART) bu tür istatistiksel sınıflama ve regresyon tekniklerine karşı güçlü bir alternatif olarak ortaya çıkmaktadır. Veri setinin çok karmaşık olduğu durumlarda bile CART, bağımlı değişkeni etkileyen değişkenleri ve bu değişkenlerin modeldeki önemini basit bir ağaç yapısı ile görsel olarak sunabilmektedir. Ele alınan bağımlı değişkenin kategorik yapıda ise yöntem sınıflama ağaçları (Classification Tree, CT), sürekli ise regresyon ağaçları (Regression Tree, RT) olarak adlandırılmaktadır.^{1, 2, 4}

Yapılan birçok tıbbi arařtırmada hastalıklara tanı koyma iřlemi klasik istatistiksel yöntemlerle belirlenmektedir. Bu yöntemler arasında en sık kullanılanlar ise ki_kare analizi, diskriminant veya lojistik regresyon analizleridir. Bu yöntemlerde ya çok sayıda deęiřkeni bir arada deęerlendirmek ya da bir takım istatistiksel varsayımların yerine gelmesi gerekmektedir ki bu tip kısıtlamalar biyolojik yapıya çoęu zaman uymamaktadır. Ayrıca yine bu yaklařımlarda risk faktörleri arasındaki iliřkileri deęerlendirmek oldukça zor ve zaman alıcıdır.³ Bunlara ilaveten çok sayıda risk faktörünün bir arada deęerlendirildięi büyük veri setlerinde veri kaybı mutlaka ortaya çıkmaktadır. Klasik çok deęiřkenli istatistiksel yöntemlerde kayıp gözlemler veri analizinde bilgi kaybına neden olmaktadır. Bu sıkıntı CART yöntemiyle giderilebilmektedir. Bu avantajlarının yanı sıra henüz yeni bir yöntem olmasından dolayı algoritmasında bazı deęiřiklikler yapılmakta ve teorisi gün geçtikçe gelişmektedir. Ayrıca bir veri seti için ortaya çıkan farklı ağaç yapılarından uygun tanı koyabilecek ağaç yapısının seçiminde de dikkatli olmak gerekir.

Bu çalışmanın amacı, CT hakkında önemli teorik bilgileri özetlemek ve Huzursuz Bacak Sendromu (Restless Legs Syndrome, RLS) hastası olup olmama durumunu belirleyen deęiřkenleri CT analizini kullanarak tespit etmektir.

GEREÇ VE YÖNTEM

Denekler

Mersin Üniversitesi Tıp Fakültesi Hastanesi Nöroloji Bölümünün 206 denek üzerinde yaptıęı anket çalışmasının sonuçları, bu çalışmada veri olarak kullanılmıştır. Bu anket çalışmasının uygulandıęı 206 denegin 103'ü RLS hastası ve geriye kalan 103 denekte ise RLS hastası deęildir. Bu şekilde denekler, hasta ve kontrol grubu olarak ikiye ayrılmış ve denegin ait olduęu grup, sınıflama ağacı uygulamasında iki seviyeli kategorik bağımlı deęiřken olarak kullanılmıştır. RLS ve Kontrol gruplarının ayrımını önemli ölçüde etkileyen bağımsız deęiřkenler, ankette her iki gruba da sorulan ortak sorular kullanılarak tespit edilmiştir. Bu deęiřkenlerden yaş, kilo, boy uzunluęu, öğrenim süresi (yıl) günlük ortalama uyku süresi (saat) ve aylık ortalama gündüz saatlerinde uyuklama miktarı (gün) soruları analizde kullanılan sürekli bağımsız deęiřkenlerdir. Cinsiyet, meslek, yařanılan yer ve bu yerin deniz kıyısına olan uzaklıęı, medeni hal, geçmiş ve bugünkü sigara

kullanma alışkanlıkları, alkol alışkanlıęı, antidepresan ve antiparkinson ilaç kullanımı, beyin ya da omurilik rahatsızlıkları, kansızlık hastalıęı, böbrek yetmezlięi, hipertansiyon, diyabet, migren, depresyon, genel saęlık durumu, son 1 ay içerisindeki moral durumu ve birinci derece akrabalarda gözlenen řikayetler soruları kategorik bağımsız deęiřken olarak ve bir günde içilen sigara adedi, aylık ortalama gece uykusundan uyanma sayısı, aylık ortalama rüya adedi ise kesikli deęiřkenler olarak analize alınmıştır.¹

İSTATİSTİK YÖNTEM

Sınıflama Ağaçları

CT kategorik yapıdaki bağımlı deęiřkenin alacağı deęerleri tahmin etmek üzere geliştirilen parametrik olmayan istatistiksel bir yöntemdir.^{2,4} CT modellerinin kullanımı son yıllarda hızla artış göstermektedir. CT modellerinin kullanım hızındaki artış ařağıdaki nedenlere bağlıdır.

- CT parametrik olmayan bir model olduęu için varsayımları çok kısıtlıdır.
- Modelde deęiřkenlerin türü (sürekli, kategorik, sıralı ya da bunların karışımı) konusunda herhangi bir varsayım gerektirmez ve sınırlama getirmez.³
- Bağımlı ve bağımsız deęiřkenler arasındaki iliřki görsel sunuma sahip olduęundan, ağaç şeklindeki model sonuçları çok fazla istatistik bilgisine gerek duyulmadan kolay bir şekilde yorumlanabilir.
- CT, tanımlanan bağımlı deęiřken için olabilecek bütün bağımsız deęiřkenleri ve onların tüm kombinasyonlarını modele katar ve mümkün olan en doęru sınıflandırmayı yapar. Deęiřken kombinasyonlarına da bakıldıęı için interaksiyonlar da deęerlendirilmiş olur.
- Karmařık veri setlerine kolaylıkla uygulanabilir.
- Hem bağımlı hem de bağımsız deęiřkenler için kayıp veya eksik deęerler ile aşırı uç deęerlerden etkilenmeyen bir metottur.
- Geleneksel birçok istatistik teknięine (çoklu regresyon, varyans analizi, lojistik regresyon, diskriminant analizi, kümeleme analizi) alternatiftir.
- Kesin olmayan ancak saęlam temellere dayanan ağaç metodlarını da hesaba katar.⁵
- Eęer ihtiyaç duyulursa aynı bağımsız deęiřken aynı ağaçta farklı ayırma deęerleriyle (cut off) kullanılabilir.

Oldukça güçlü bir analiz tekniği olmasına karşın CT" nin birtakım sınırlılıkları da vardır. CT tekniğinin en önemli sınırlılığı sonuçların bir olasılık modeline dayanmıyor olmasıdır. Veri setine uygun bir CT ağacından alınmış tahmini sınıflandırmaya yardım edebilecek bir olasılık derecesi ya da güven aralığı yoktur. CT tarafından üretilen sonuçların doğruluğuna duyulabilecek güven tamamen geçmiş verilere dayalı doğruluğuyla orantılıdır.^{3,4}

CT bireyleri doğru grubuna sınıflayabilmek için bir başlangıç veri seti kullanır ve bu veri seti Learning Sample olarak adlandırılır. CT, kök düğümden başlayarak devam eden ve her düğümden o düğüme ait deney ünitelerine uygulanan basit sorulardan alınan evet/hayır cevaplarına göre oluşan yollar içerir. Her düğümden uygulanan bu sorulara ayrıca denir. Bu işlem ayırma olarak adlandırılır.⁶

Deney ünitelerinin birden fazla bağımsız değişken içermesi durumunda değişen tek şey ayırma ağaçlarının tüm değişken ve değişken kombinasyonlarını tek tek ele almasıdır. Bu durumda, deney ünitelerinin içerdiği bağımsız değişkenler ve bu değişkenlerin birbirleri ile kombinasyonlarının tanımlı bulunduğu aralıklardaki tüm olası değerler birer ayırma olarak düşünülüp, mümkün olan tüm olası ayırmalar belirlenir. Oluşan ağaçlarda homojen olmayan düğümlere çocuk düğümü homojen düğümlere ise terminal düğüm adı verilir ve hesaplamalar sonucunda terminal düğümler hasta veya kontrol grubu olarak adlandırılır ve bu düğümler yorumlanır.

Herhangi bir düğümün heterojenlik değeri safsızlık (impurity) ölçüsü olarak adlandırılır ve bu değer safsızlık fonksiyonu kullanılarak hesaplanır. Safsızlık ölçüsü sıfır değerini alıyorsa düğüm tamamen homojendir. Sınıflama ağaçlarında kullanılacak birçok alternatif safsızlık ölçüsü (Gini, Twoing, Chi-square, G-square) vardır. Ayırma fonksiyonundan anlaşılacağı gibi, kullanılan safsızlık ölçüsü herhangi bir t düğümü için en iyi ayırmanın seçimini önemli bir şekilde etkilemektedir. Bu nedenle safsızlık ölçüleri literatürde en iyi ayırma kriterleri (ya da ayırma kuralları) olarak da bilinirler. En yaygın olarak kullanılan ayırma kriterleri Gini Diversity Index (Gini) ve Twoing Kuralı'dır.^{1,4}

CT metodunda bir sınıflama ağacı oluşturulurken ön olasılıklar (prior probabilities) kullanır. Ön olasılıklar deney ünitelerinin ait olacağı sınıfın belirlenmesini

etkiler.⁷ j sınıfı için ön olasılık değeri (π_j) ile gösterilir ve bu değerler ya veri setinden hesaplanır ya da araştırmacı tarafından bildirilir.

Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır (Doğruluk Oranı = 1 - Hata Oranı). Verilerin sınıflandırılması için oluşturulan modellerin hata oranlarına karar vermek için risk matrisi kullanılmaktadır. Ayırma sonucunda ortaya çıkan herhangi bir düğüme atanacak olan en uygun sınıf aşağıdaki gibi tahmin edilir;

$C(j/i)$: i sınıfını j sınıfı gibi sınıflamanın maliyeti (risk matrisi katsayıları),

π_i : i sınıfının önceki olasılığı,

N_i : Learning Sample'da i sınıfında bulunan deney ünitelerinin sayısı,

$N_i^{(t)}$: t düğümünde i sınıfında bulunan deney ünitelerinin sayısı olmak üzere;

$$\frac{C(j/i)\pi_i N_i^{(t)}}{C(i/j)\pi_j N_j^{(t)}} > \frac{N_i}{N_j} \text{ eşitsizliği } j \text{'nin bütün}$$

değerleri ($j = 1, 2, \dots, k$ ve $j \neq i$) için sağlanıyorsa t düğümüne en uygun olarak i sınıfı atanır.⁵

Düğümün yapısına göre bazı durumlarda birden fazla sınıf yukarıda belirtilen eşitsizliği sağlayarak en uygun sınıf konumuna girer ya da hiçbir sınıf bu eşitsizliği sağlayamaz. Böyle bir durumda en uygun sınıfın belirlenmesi için çoğulluk ve minimum risk olmak üzere iki alternatif kural mevcuttur.⁴

Çoğulluk kuralı hatalı sınıflama maliyetini göz önüne almaksızın (eşit varsayarak) düğüm içerisinde en büyük orana sahip olan sınıfı en uygun sınıf olarak atar.⁷

Minimum risk kuralı ise düğüm içerisinde deney ünitelerinin sınıflara dağılımını göz önüne almaksızın (eşit varsayarak) düğüm içerisinde hatalı sınıflama maliyetini minimum yapan sınıfı en uygun sınıf olarak belirler.

CT modellerinde tekrarlı ikili bölünmelerle homojen alt gruplar elde edilir ve ağaç bu şekilde büyümeye devam eder. ^{4, 5, 8, 9}

CT' de meydana gelen büyüme;

1. Her çocuk düğümündeki gözlem sayısı sadece bir gözlem ise veya on gözlem ise,^{4,5}
2. Her düğümde grup içi homojenlik söz konusu ise,
3. Ağacın düzey sayısında analizi yürüten kişi tarafından bir sınırlama yapıldıysa,
4. Yeni oluşacak düğümlerde fazla bir değişiklik yaratmıyorsa durur.

Ağaç inşası sonunda elde edilen ağaç büyük (maksimum) ağaç olarak adlandırılır ve Learning Sample'daki deney ünitelerine en uygun ağaçtır. Ancak maksimum ağaç pratikte iki dezavantaja sahiptir.

1. Maksimum ağaç Learning Sample'ı kusursuz biçimde tanımlar çünkü eklenen her bağımsız değişken hatalı sınıflama oranını düşürür. Bu durumda, maksimum ağaç Learning Sample için olması gerekenden daha iyi bir tahmin modeli (overfitting) sunar. Ancak, Learning Sample'a aşırı uyumlu maksimum ağaçlar farklı bir veri seti (örneğin Test Sample) söz konusu olduğunda iyi bir tahmin sağlayamazlar.

2. Bir sınıflama ağacının karmaşıklık ölçüsü o ağacın terminal düğüm sayısına eşittir. Terminal düğüm sayıları ve dolayısıyla karmaşıklığı yüksek olan maksimum ağacın anlaşılması ve yorumlanması güçtür.

Maksimum ağacın pratikte ortaya çıkardığı bu sorunların çözümü için maksimum ağacın budanması yani maksimum ağaçtan oluşturulan daha küçük bir ağacın seçilmesi gereklidir.¹⁰ Maksimum ağacın budanması daha küçük ağaçlar dizisi oluşturur ve oluşturulan bu dizi içerisinde optimum ağaç seçilir. Optimum ağaç maksimum ağaçtan daha az karmaşıklığa sahiptir ancak, Learning Sample'a maksimum ağaçtan daha az uyumludur ve hatalı sınıflama oranı daha yüksektir.⁴ Maliyet-karmaşıklık budama metoduna göre maksimum ağaç, maliyet-karmaşıklık ölçüsü minimum değerine ulaşıncaya kadar budanır ve optimum ağaç elde edilir.

Sınıflama ağaçlarında üç alternatif doğruluk tahmin yöntemi vardır. Bunlar yeniden yerine koyma tahmini, test örneği tahmini ve çapraz geçerlilik testidir.^{1,4}

Statistica® 6.0 istatistik paket programı kullanılarak yapılan analizlerde ayırma kriteri olarak Gini, budama sonucunda optimum ağacın belirlenmesinde ise sınıflama ağacının doğruluk tahmini olarak 10 katlı

çapraz geçerlilik testi tercih edilmiştir. Hasta ve kontrol gruplarında yer alan denek sayısının eşit olması nedeniyle önsel olasılıklar eşit (%50) alınmıştır. İlk analiz ile elde edilen maksimum ağaç ve bu ağacın budanması sonucu ortaya çıkan diğer ağaçlarla birlikte toplam 8 ağaç oluşturulmuş, ancak gereksiz tekrarlamalardan kaçınmak amacı ile bu çalışmada yalnızca maksimum ağaç ve optimum ağaç açıklanmıştır.

RLS hekimler tarafından da yaygın olarak atlanabilen ancak oldukça sık rastlanan ve en önemli uykusuzluk nedenlerinden biri olup, özellikle de bacaklarda istirahat sırasında veya yatarken ortaya çıkan nahoş duygular sonucunda bacakları sürekli hareket ettirme ihtiyacı duyma ve bu nedenle uykuya dalmama ile karakterize bir rahatsızlıktır.¹¹

Şekil 1'de sunulan maksimum ağaç, değişkenler arasındaki ilişkileri en ayrıntılı biçimde gösteren sınıflama ağacıdır. Bu ağaçta terminal düğümleri kırmızı, çocuk düğümleri ise mavi renkte kareler olarak şekillendirilmiş ve her bir düğüm içerisinde o düğümün hangi sınıfa (RLS yada Kontrol) ait olduğu belirtilmiştir. Maksimum ağaçta 13 terminal düğüm, 12 çocuk düğüm vardır. Düğümlerin içerisinde sağ üst köşede N harfi ile o düğümde toplam kaç deneğin bulunduğu, sol üst köşede ise ID harfi ile düğüm numarası gösterilmektedir. Düğümler içerisinde ayrıca, o düğümde yer alan deneklerin ait oldukları sınıflar bar grafiği ile sunulmuş ve çoğulluk kuralına göre o düğüme atanan sınıf belirtilmiştir.

Şekil 1'de sunulan sınıflama ağacı başlangıçta 206 deneğin tümünü aynı grupta kabul ederek analize başlamıştır. Aile düğümünü iki çocuk düğümüne ayıran ilk ayırıcı 1. derece akrabalarınızda bacaklarda uyuşma, karıncalanma ve hareket ettirdikçe geçen bu tür şikayetlerin olup olmadığı sorusudur. Bu soruya verilen evet, hayır ya da bilmiyorum cevapları ile aile düğümü iki çocuk düğümüne ayrılmıştır. Toplam 206 denek içerisinde, bu soruya cevabı evet olan 51 denek sol çocuk düğümüne (Düğüm 2), cevabı hayır ya da bilmiyorum olan 155 denek ise sağ çocuk düğümüne (Düğüm 3) ayrılmaktadır. Diğer çocuk düğümleri de aynı yöntemle (sadece sorular değiştirilerek) tekrarlı olarak ayrılarak ayırma işlemi terminal düğümlere kadar devam etmektedir. 206 deneğin tümü Şekil 1'de sunulan sınıflama ağacına tabi tutulduğunda, gerçekte RLS grubuna ait olan 103 deneğin 83'ünün yine RLS grubunda (doğru sınıflandırma) 20'sinin ise Kontrol

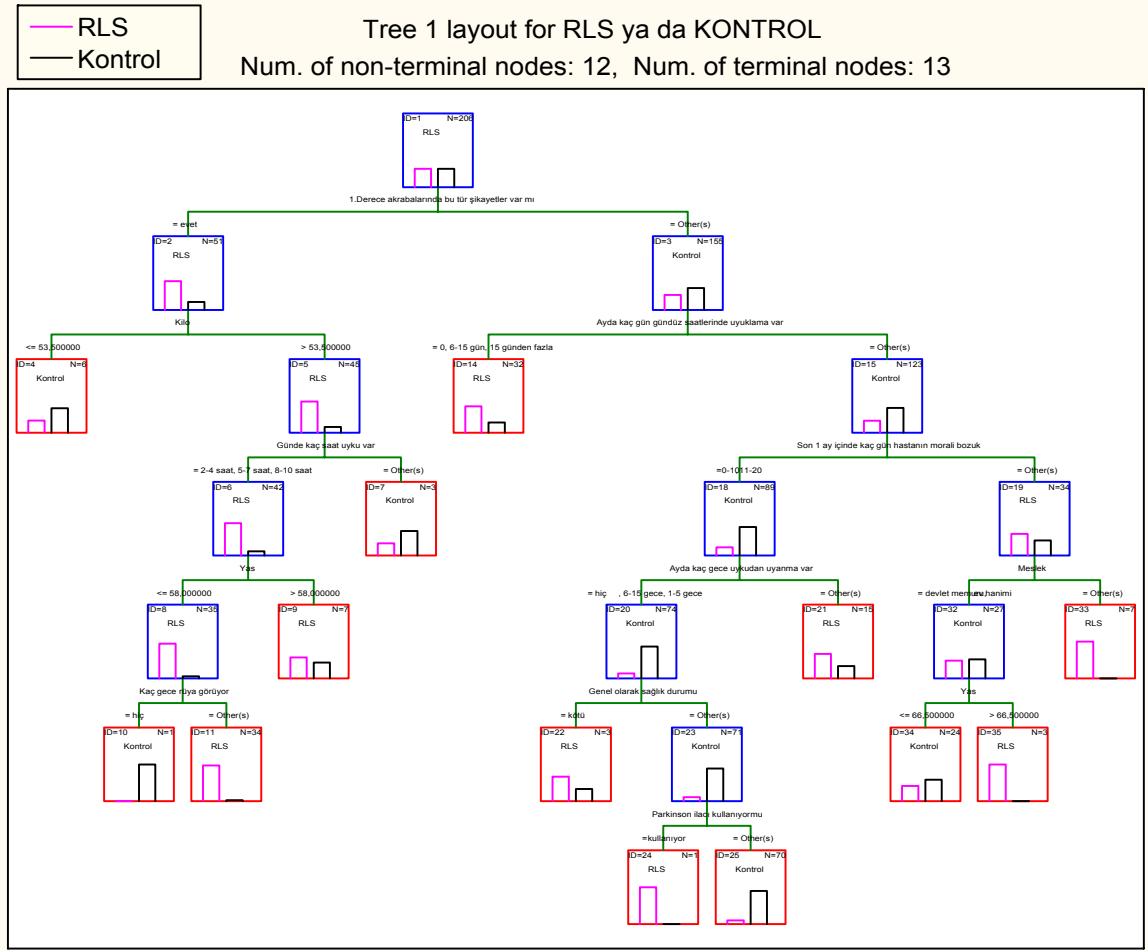
Sınıflama Ağaçları Yardımıyla Restless Legs Syndrome (RLS) Hastalarına Tanı Koyma

grubunda (yanlış sınıflandırma) olduğu görülmüştür. Gerçekte Kontrol grubuna dahil olan 103 denegün 84'ü ise doğru sınıflandırılarak yine Kontrol grubunda, 19'u ise yanlış sınıflandırılarak RLS grubunda yer almıştır. Bu durumda, maksimum ağaç deneklerin %81'ini doğru sınıflandırmıştır.

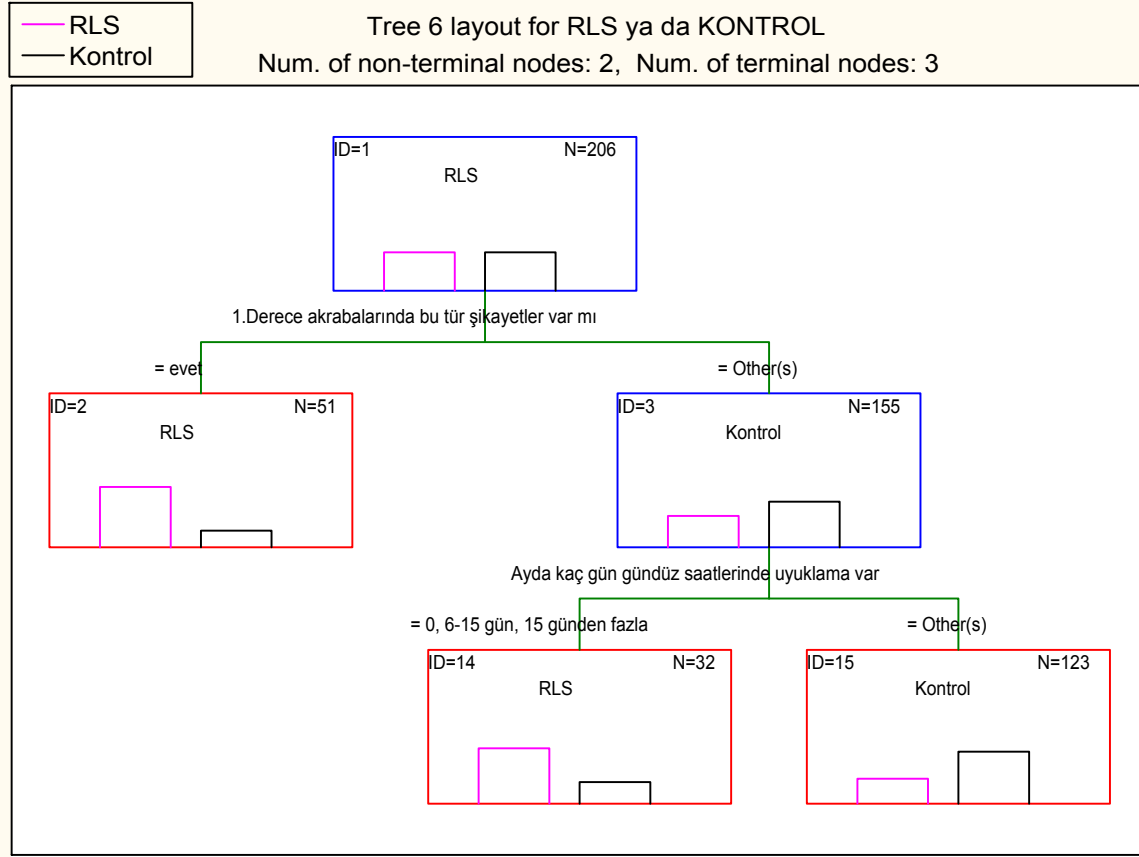
Maksimum ağacın en alt çocuk düğümünden başlayarak aşamalı olarak yukarıya doğru budanması ile her aşamada, bir önceki ağaçtan daha az düğüme sahip ancak hatalı sınıflama oranı daha yüksek olan, yeni bir ağaç üretilmiştir. Budama sonucunda elde edilen tüm ağaçlar içerisinde, düğüm sayısındaki azalış ile 10 katlı

çapraz geçerlilik testi kullanılarak hesaplanan hatalı sınıflama oranındaki artış arasındaki dengeyi en iyi şekilde sağlayan optimum ağaç belirlenerek Şekil 2'de sunulmuştur. Şekil 2'den kolayca anlaşılacağı gibi optimum ağaç maksimum ağaçtan çok daha az sayıda düğüm (aynı zamanda daha az sayıda soru) içermektedir. Bununla birlikte 206 denegün tümü optimum ağaca tabi tutularak sınıflandırıldığında, gerçekte RLS grubuna ait olan 103 denegün 63'ünün, gerçekte Kontrol grubuna ait olan 103 denegün ise 84'ünün doğru sınıflandırıldığı görülmüştür. Budamanın bir sonucu olarak, maksimum ağaçta %81 olan doğru sınıflama oranı optimum ağaçta %71'e düşmüştür.

Şekil 1. Maksimum Ağaç



Şekil 2. Optimum Ağaç



Maksimum ağaç değerlendirildiğinde, bu çalışmada dikkate alınan risk faktörleri içerisinde;

- Birinci derece akrabalarınızda şikayet var mı?
- Son 1 ay içerisinde hastanın morali kaç gün bozuk?
- Kilo?
- Ayda kaç kez gündüz saatlerinde uyuklama var?
- Günde kaç saat uyku var?
- Yaş?
- Bir ayda ortalama kaç gece rüya görüyor?
- Genel olarak sağlık durumu
- Parkinson ilacı kullanıyor musunuz?
- Mesleğiniz nedir?

Olmak üzere toplam 10 soru RLS hastası olan ve olmayan bireyleri ayırmada anlamlı risk faktörleri olarak belirlenmiştir. Ayrıca anlamlı etkiye sahip olan bu 10

risk faktörünün birbirleriyle olan ilişkileri de şekil 1 üzerinde görünmektedir.

Klasik sınıflama modellerine alternatif olarak ortaya çıkan CT, modelde yer alan değişkenler üzerinde hiçbir varsayım gerektirmemesi nedeniyle, özellikle tıbbi araştırmalarda, elde edilen verilerin fazla sayıda ve kompleks bir yapıda olduğu durumlarda sınıflama analizlerinde kolaylıkla uygulanabilen güçlü bir tekniktir. CT analiziyle oluşturulan sınıflama ağacı incelendiğinde bağımlı değişkenleri hangi bağımsız değişkenlerin etkilediği, bağımsız değişkenlerin risk durumunu belirleyen cut-off değerleri ve bağımsız değişkenler arasındaki etkileşim kolaylıkla görülebilmektedir. Bu özelliği ile CT bir hekime, daha önce hastalık üzerinde etkisi olmadığını düşündüğü bir bağımsız değişkeni

tespit ederek, tekrar düşünme ya da üzerinde çalışma şansı verebilmektedir.

RLS hastalığı üzerinde yapılan klinik çalışmalar, genetik,¹² kilo, stres, parkinson ilacı kullanma,¹¹ uyku düzeni¹³ ve meslek¹⁴ gibi faktörlerin önemli olduğunu vurgulamaktadır. Bu çalışma sonucunda elde edilen maksimum ve optimum ağaçlara bakıldığında da, analiz sonuçlarının klinik çalışmalar ile yakın bir paralellik gösterdiği görülmektedir. Elde edilen maksimum ya da optimum ağaç gelecekte ortaya çıkacak vakalar için RLS hastalığına tanı koymada, kabul edilebilir bir hata oranı ile, hekimler tarafından kolaylıkla kullanılabilir. Bunun yanında, elde edilen ağaçlar yardımı ile RLS hastalığını başka hangi faktörlerin etkilediğini ve bu faktörlerin ayrıca değerlerini (hangi sınırdan sonra riskin arttığını) görmek hekimler için son derece önemlidir. Ayrıca, diğer sınıflama ve tahmin problemlerinde sorun yaratan bağımsız değişkenler arasındaki etkileşim, CT için bir sorun teşkil etmemekte aksine bu durum modeli daha güvenilir kılmaktadır.

Çalışmada elde edilen optimum ve maksimum ağaç yapıları değerlendirildiğinde maksimum ağaçta ayrıca olarak 10 risk faktörü kullanılmış ve oluşan ağacın toplam doğru tanı koyma başarısı %81 olarak bulunmuştur buna karşılık optimum ağaçta bu 10 risk faktöründen sadece 2 tanesi ile toplam %71' lik bir doğru tanı koyma başarısı elde edilmiştir. Bu iki ağaç yapısına ait doğru tanı koyma başarıları istatistik olarak anlamlı bulunmuş ve bu sonuca göre söz konusu 2 ağaç yapısının da tanı koyma amacıyla kullanılabilceği

söylenbilir. Ancak unutulmamalıdır ki maksimum ağaç yeni gelen bireylerin bu iki gruba atanmasında daha fazla hata payına sahiptir.

KAYNAKLAR

1. Örekici G. Sınıflama ve Regresyon Ağaçları, Yüksek Lisans Tezi, Mersin Üniversitesi Sağlık Bilimleri Enstitüsü, 2004, 85 s.
2. Fu C. (2003), Combining loglinear model with classification and regression tree (CART): An application to birth data. Computational Statistics & Data Analysis, 2003; (Article In Press).
3. Yohannes Y, Hodinott J. Classification and Regression Trees: An introduction. 2003, Erişim: <http://www.ifpri.org/themes/mp18/techquid/tg03.pdf>, Erişim Tarihi: 10.06.2003.
4. Breiman L, Friedman JH, Olshen RA, and Stone CJ. Classification and Regression Trees, Boca Raton, Florida: Chapman & Hall. 2003.
5. Lewis R. An introduction to classification and regression tree (CART) analysis, Academic Emergency Medicine, 2000. Erişim: <http://www.saem.org/download/lewis1.pdf>.
6. De'ath G, Fabricius K. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology, 2000; 81: 3178-92.
7. STATSOFT. Classification and Regression Trees. 2003. Erişim: <http://www.statsoft.com/textbook/scart.html>, Erişim Tarihi: 18.03.2003
8. Chipman HA, George EI and McCulloch RE. Hierarchical priors for bayesian CART Shrinkage. Statistics and Computing, 2000; 10: 17-24.
9. Bevilacqua M, Braglia M and Montanari R. The classification and regression tree approach to pump failure rate analysis. Reliability Engineering and System Safety. 2003; 79: 59-67.
10. Put R, Questier F, Coomans D et al. Classification and regression tree analysis for molecular descriptor selection and retention prediction in chromatographic quantitative structure-retention relationship studies. Journal of Chromatography A. 2003; 988: 261-76.
11. Atay T. Huzursuz bacak sendromu, 2003. Erişim: <http://www.internationalhospital.com.tr/>, Erişim Tarihi: 18.09.2003
12. Erdal S. Huzursuz bacak sendromu, 2003. Erişim: <http://www.genetikbilimi.com/tip/huzursuz.html>, Erişim Tarihi: 01.12.2003
13. Kaynak H. Uykusuzluk, 2003. Erişim: <http://www.ntvmsnbc.com/news/>, Erişim Tarihi: 27.10.2003
14. Aksu M. Huzursuz bacak sendromu yanlış teşhise yol açabiliyor, 2003. Erişim: <http://www.martiyazilim.com.tr/>, Erişim Tarihi: 18.09.2003

Yazışma Adresi

Yrd.Doç.Dr. Zeki AKKUŞ
Dicle Üniversitesi Tıp Fakültesi Biyoistatistik
Anabilim Dalı, Diyarbakır
Tel : 532 574 4841
E-Posta : zakkus@dicle.edu.tr