

Orijinal Araştırma

Hidrocefali Hastalığının Sınıflandırılması ve İlişkili Faktörlerin Makine Öğrenmesi Yöntemleri ile Belirlenmesi

Classification of Hydrocephalus Disease and Determination of Related Factors by Machine Learning Method

Zeynep Tunç¹, Cemil Çolak¹, Ramazan Özdemir²

¹İnönü Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya, Türkiye

²İnönü Üniversitesi Tıp Fakültesi, Pediatri Anabilim Dalı, Malatya, Türkiye

Özet

Bu araştırmanın birinci amacı çeşitli çekirdek fonksiyonlarıyla oluşturulan destek vektör makineleri (DVM) yardımıyla ve farklı dallanma kriterleriyle oluşturulan Random Forest (RF) yöntemleriyle hidrocefaliye sahip olan ve olmayan hasta bebekleri sınıflandırmaktır.

Bu çalışmada incelenen veriler, İnönü Üniversitesi Turgut Özal Tıp Merkezi Pediatri Anabilim Dalı veri tabanından geriye yönelik (retrospektif) olarak seçilmiştir. Çalışmada yer alan veriler hidrocefalisi olan ve olmayan bebekleri ve bu bebeklere ilişkin diğer özellikleri içermektedir. Hidrocefalinin sınıflandırılması işlemi için DVM ve RF modelleri kullanılmıştır. Performans değerlendirilmesinde doğruluk, sınıflama hatası, kappa istatistiği, ROC eğrisi altında kalan alan (AUC), f ölçümü, duyarlılık, seçicilik ve Youden indeksi ölçütleri kullanılmıştır.

DVM modellerinden en iyi sınıflandırma performansını gösteren ANOVA çekirdeğine ilişkin performans ölçütlerine ait değerler (%95 güven aralığında) sırasıyla; doğruluk için %72.14±8.99, sınıflama hatası için %27.86±8.99, kappa istatistiği için 0.424±0.209, AUC için 0.748±0.120, f ölçümü için %75.78±6.97, duyarlılık için %71.45±8.27, seçicilik için %73.00±22.13, Youden indeksi için 0.444±0.221 olarak elde edilmiştir. RF modellerinden en iyi sınıflandırma performansını gösteren Gini İndeksi ile oluşturulan modele ilişkin performans ölçütlerine ait değerler sırasıyla; doğruluk için %77.47±9.51, sınıflama hatası için %22.53±9.51, kappa istatistiği için 0.554±0.193, AUC için 0.799±0.100, f ölçümü için %78.89±9.43, duyarlılık için %71.90±15.10, seçicilik için %87.00±19.23, Youden indeksi için 0.589±0.197 olarak elde edilmiştir.

Hidrocefaliyi sınıflandırmada en iyi performansı RF modeli vermiştir. İlerleyen çalışmalarda, hidrocefalinin sınıflandırma performansını arttırmak için gelişmiş makine öğrenmesi yöntemleri kullanılacaktır.

Anahtar kelimeler: Destek Vektör Makinesi, Random Forest, Hidrocefali, Makine Öğrenmesi

Abstract

The first objective of this research is to classify babies with and without hydrocephalus by means of support vector machines (SVM) constructed by various kernel functions and Random Forest (RF) methods generated by different branching criteria.

The data analyzed in this study were retrospectively selected from the database of the Department of Pediatric, Turgut Özal Medical Center, Inonu University. The data included in the study include babies with and without hydrocephalus and risk factors related to these babies. SVM and RF models were used for classification of hydrocephalus. As performance evaluation metrics; accuracy, classification error, kappa statistic, area under the ROC curve (AUC), f-measure, sensitivity, specificity and Youden index were used.

For SVM models, the best classification performances were obtained by the SVM with ANOVA kernel model. The performance metrics (with 95% confidence interval) were respectively; 72.14% ± 8.99 for accuracy, 27.86% ± 8.99 for classification error, 0.424 ± 0.209 for kappa statistic, 0.748 ± 0.120 for AUC, 75.78% ± 6.97 for f measurement, 71.45% ± 8.27 for sensitivity, 73.00% ± 22.13 and 0.404 ± 0.221 for Youden index. For RF models, the best classification performances were obtained by RF constructed by the Gini Index. The performance metrics (with 95% confidence interval) were respectively; 77.47% ± 9.51 for accuracy, 22.53% ± 9.51 for classification error, 0.554 ± 0.193 for kappa statistic, 0.799 ± 0.100 for AUC, 78.89% ± 9.43 for f measurement, 71.90% ± 15.10 for sensitivity, 87.00% ± 19.23 and 0.589 ± 0.197 for Youden index.

The best classification performance to predict Hydrocephalic achieved by RF model. As future study, more sophisticated machine learning algorithms will be used to improve classification performance of the Hydrocephalic.

Keywords: Missing Value (S) Analysis; Assignment Methods; Web Based Software

Giriş

Beyinde su birikmesi olarak kabul edilen hidrocefali pediatrik beyin cerrahisi bölümünü en çok meşgul eden halk sağlığı problemleri arasındadır. Beyin omurilik sıvısının (BOS) fizyolojisini ve hidrocefaliyi inceleyen çok

sayıda gözlemsel ve deneysel araştırmalar söz konusudur. Hidrocefali; çok fazla miktarda BOS dilate serebral ventriküller ile subaraknoid boşlukta birikmeye başlamasıyla seyreden bir bozukluktur (1, 2). Klasik olarak bilinen hidrocefaliyi BOS dolanım merkezlerinde tıkanma ya da BOS yapımı ve tekrar emilimi

arasındaki dengenin bozulmasına bağlı olarak ortaya çıkan, ventriküllerde genişleme ve kafa içi basıncının (KİB) artmasıyla seyreden bir klinik tablo olarak tanımlanabilir. Gerçek konjenital hidrosefalinin görülme sıklığı her 1000 canlı doğumda 1-1.5 arasında bildirilmiştir. Buna ek olarak, diğer konjenital nörolojik hastalıklarla birlikte oluşmuş ve edinilmiş birçok intrakraniyal patoloji etkisinde gelişen hidrosefali ile bu oran her 1000 canlı doğumda 3-4'e ulaşmaktadır (3). Çocuklarda oluşan hidrosefali genel olarak artmış kafa içi basıncı ile ilişkili olduğu bildirilmektedir. Çoğu hastada bu durum, BOS dolanımındaki bozukluktan kaynaklı olarak (obstrüktif veya non-komünike hidrosefali) fazla miktardaki BOS'un serebral ventriküllerde birikmesinden kaynaklanır. Daha az durumda ise emilimdeki yetersizlik (komünike hidrosefali) BOS birikimine neden olur ve kafa içi basıncı artar (4).

Veri madenciliği bilgi teknolojilerinin doğal bir süreci olarak görülebilir. Elde edilen fazla miktarda bilgiyi depolamayı, uzun yıllar sürecektir uygulamalar için yorum çıkartabilmeyi ve ileriye dönük tahminlemeler yapabilmeyi sağlayan bir bilgi sürecidir. Bu süreç verinin temizlenmesi, verilerin birleştirilmesi, istenilen verinin seçimi, verileri uygun formlara dönüştürme, veri madenciliği uygulamaları ve yorumlanması şeklinde sıralanır. Veri madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar verme modelleri yaratmaktır (5, 6).

DVM yöntemleri son yılların popüler algoritmaları arasında yerini almıştır. Vapnik - Chervonenkis tarafından geliştirilen DVM, sınıflandırma problemlerine ek olarak regresyon problemlerinde de kullanılan bir makine öğrenmesi modelidir. DVM verileri dönüştürmek için çekirdek çözümü (kernel trick) adında bir teknikten faydalanır. Çekirdek çözümü yöntemleri veri dönüşümü modellerine dayanarak olası sonuçlar arasından en uygun

sınırı belirler. Yani çekirdek çözümü yöntemleri öncelikle karmaşık veri dönüşümleri yapar ve sonra bu verilerin, tanımlanan etiketler veya sonuçlara dayalı olarak ne şekilde ayrılacağını belirler (7-10).

RF; Leo Breiman ve Adele Cutler tarafından geliştirilen ve içerisinde oylama metodunu barındıran bir sınıflama yöntemidir. Birçok karar ağacının bir araya gelmesiyle oluşur ve bireysel ağaçlar tarafından oylanarak kazanan sınıf belirlenir. Ormanda yer alan karar ağaçları, birbirinden bağımsızdır ve veri setinden bootstrap tekniği ile çekilen örneklerden oluşturulur (11). RF yöntemi çok sayıda karar ağacından oluşan orman sınıflayıcısıdır ve bu yöntemle sınıflama veya regresyon ağaçları kurulabilmektedir. Veri setindeki "sınıf değişkeni" kategorik ise sınıflama, sürekli ise regresyon ağaçları kurulmaktadır (12).

Bu araştırmanın birinci amacı çeşitli çekirdek fonksiyonlarıyla oluşturulan DVM'ler yardımıyla ve farklı kriterlerde oluşturulan RF yöntemleriyle hidrosefaliye sahip olan ve olmayan hasta bebekleri sınıflandırmaktır. Araştırma içindeki ikinci amaç ise en iyi sınıflamayı hangi çekirdek fonksiyonu ile oluşturulan DVM ve hangi dallanma kriteriyle oluşturulan RF modelleriyle elde edildiğini belirlemektir.

Gereç ve Yöntem

Veri seti

Bu çalışmada kullanılan veriler İnönü Üniversitesi Turgut Özal Tıp Merkezi Pediatri Anabilim Dalı için geliştirilen veri tabanından geriye yönelik (retrospektif) olarak seçilmiştir. Çalışmada yer alan veriler 133 çocuktan oluşup, hidrosefalisi olan 81 ve olmayan 52 çocuklara ilişkin bazı değişkenleri içermektedir. İncelenen verideki tüm değişkenlere ait tanımlayıcı bilgiler Tablo 1'de verilmiştir.

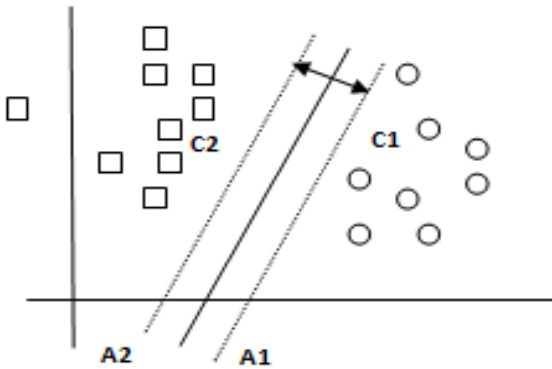
Tablo 1. Çalışmada kullanılan değişkenlerin tanımları

Değişkenler	Değişken Tipi	Değişken Açıklaması	Ölçüm Birimi	Değişken Rolü
Hidrosefali	Kategorik	Var/yok	-	Bağımlı/hedef
Anne yaşı	Nümerik	Doğal sayı	Yıl	Bağımsız/tahminleyici
Doğum haftası	Nümerik	Doğal sayı	Hafta	Bağımsız/tahminleyici
Doğum ağırlığı	Nümerik	Doğal sayı	Gr	Bağımsız/tahminleyici
Baş çevresi	Nümerik	Doğal sayı	Cm	Bağımsız/tahminleyici
Hidrosefali	Kategorik	Var/yok	-	Bağımlı/hedef

Destek vektör makineleri

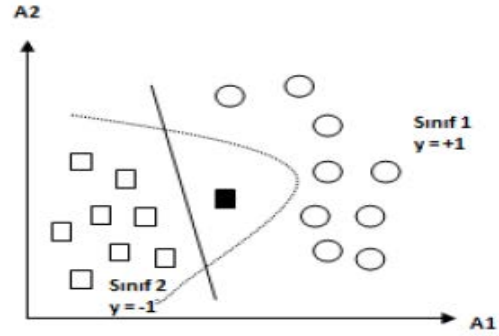
DVM'deki asıl amaç hedef değişkene ait sınıfları birbirinden en uygun şekilde ayıracak bir hiperdüzlemin elde edilmesidir. DVM regresyon metodundaki temel fikir, eldeki eğitim verilerinin özelliklerini mümkün olduğunca gerçeğe çok yakın bir şekilde yansıtan ve istatistiksel öğrenme teorisine uygun olan doğrusal ayırıcı fonksiyonun bulunmasıdır. DVM' de karşılaşılabilecek iki durum vardır. Bunlar verilerin Doğrusal olarak ayrılabilirlikleri bir yapıda olması veya Doğrusal olarak ayrılabilen bir yapıda olmasıdır. DVM ile sınıflandırma da regresyon da yapılacak olsa doğrusal olmayan durumların çözümü için çekirdek fonksiyonları kullanılır (13, 14).

DVM modellerinin temelinde verilerin Doğrusal ayrılabilirlik veya ayrılabilen durumu yer alır. Eğitim veri setinde kullanılacak N elemandan oluşan verinin, $\theta = \{x_i, y_i\}$, $i = 1, 2, \dots, N$ olduğu varsayalım. Burada $y_i \in \{-1, 1\}$ etiket değerleri ve $x_i \in \mathbb{R}^d$ özellikler vektörüdür. Doğrusal olarak ayrılabilirlikte, iki değerden oluşan veriler direkt olarak bir aşırı düzlem ile ayrılabilir. Bu aşırı düzleme ayırıcı aşırı düzlem adı verilir. DVM'nin amacı bu aşırı düzlemin iki ayrı sınıfta bulunan örnek grubuna eş uzaklıkta olmasını sağlamaktır (14). Doğrusal ayrılabilirlik durumu Şekil 1'de verilmiştir (14).



Şekil 1. DVM'de Doğrusal ayrılabilirlik durumu

Doğrusal ayrılabilirlik söz konusu ise veriler iki sınıfa doğrusal bir düzlem ile ayrılabilir. Fakat bu durum uygulamada her zaman geçerli olmayabilir. Yani doğrusal bir düzlem ile verileri birbirinden ayırmak her zaman mümkün olmayabilir. Doğrusal ayrılabilen durumu Şekil 2'de gösterilmiştir (14).



Şekil 2. DVM'de Doğrusal ayrılabilen durumu

Böyle durumlarda DVM, doğrusal olmayan bir haritalama fonksiyonu aracılığıyla orijinal giriş uzayından kolay bir şekilde doğrusal olarak sınıflandırma yapılabilecek yüksek boyutlu bir nitelik uzayına dönüşüm yapar. Bu şekilde çekirdek fonksiyonları kullanılarak bütün değerlerin tekrar tekrar çarpım değerlerinin hesaplanarak bulunması yerine doğrudan çekirdek fonksiyonunda değer yerine koyularak nitelik uzayındaki değerinin bulunması sağlanır. Bu yolla, oldukça yüksek boyutlu bir nitelik uzayı ile uğraşma sorunu ortadan kalkar. Çekirdek fonksiyonlarının diğer bir avantajı ise eğitim aşamasında herhangi bir eğitim örneği için fonksiyon kurulup değerler bulunduktan sonra diğer örnekler için artık kalıp değerleri eğitim örneği dışında tamamen hazır olduğu için çok daha kolay hesaplanmasıdır (15).

Bu çalışmada farklı çekirdek fonksiyonlarıyla oluşturulan DVM yardımıyla veri setimize ilişkin ilişki, örüntüler çıkarma ve tahminleme işlemleri yapılmıştır. DVM oluşturulurken kullandığımız farklı çekirdek fonksiyonları, onlara ait parametreler ve parametre değerleri Tablo 2'de verilmiştir.

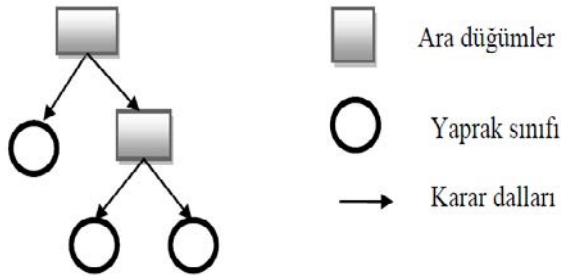
Tablo 2. Çekirdek tipleri, fonksiyonları, parametreler ve değerleri

Çekirdek tipi	Fonksiyon	Parametre	Parametre değer(ler)i
Dot	$k(x, y) = x * y$	-	-
Radial	$exp(-g x - y ^2)$	Çekirdek gamma (g)	1.0
Polinomial	$k(x, y) = (x * y + 1)^d$	Çekirdek derecesi (d)	2.0
Neural	$tanh(a x * y + b)$	Çekirdek a; çekirdek b	1.0; 0.0
Epachnenikov	$(3/4)(1 - u^2)$	Çekirdek sigma, çekirdek derecesi	1.0; 2.0

Random Forest

RF yönteminde dallanma kriterlerinin belirlenmesi ve uygun bir budama yönteminin seçimi çok önemli bir konudur. Rastgele orman sınıflandırıcının dallanma kriterlerinin belirlenmesinde Gini indeksi yöntemi kullanılmaktadır. Gini indeksi sınıf özniteliklerinin zayıflık derecesini ölçmektedir (16). RF yönteminde de başka sınıflandırma metodlarında olduğu gibi uygulayıcı tarafından belirlenmesi gereken parametreler vardır. Bu parametreler, ağaç yapısının kurulmasında gerekli olan her bir düğümde kullanılacak örneklerin sayısı ve oluşturulacak ağaçların sayısıdır. Yani bir sınıflandırma işlemi sırasında karar ormanı, kullanıcı tarafından belirlenen K adet ağaçtan oluşturulur (17, 18). RF'yi oluşturan ağaçların yapısı Şekil 3'de gösterilmiştir (19).

Bu çalışmada kazanç oranı (gain ratio), bilgi kazancı (information gain), gini indeksi ve doğruluk (accuracy) ile elde edilen farklı dallanma yapılarına sahip RF modelleri yardımıyla veri setimize ilişkin ilişki, örüntüler çıkarma ve tahminleme işlemleri yapılmıştır.



Şekil 3. RF yapısına ilişkin grafiksel gösterim

Değerlendirme ve yorumlama:

Sınıflandırma işleminde daha önceden belirtilen çeşitli çekirdek fonksiyonları yardımıyla oluşturulan DVM ve farklı dallanma kriterleriyle oluşturulan RF modellerinin tahmin performansları, 10-katlı çapraz geçerlilik tekniği ile değerlendirilmiştir. Çapraz geçerlilik tekniğinde veriler öncelikle n parçaya ayrılır ve

kullanılan model n parça için uygulanır. n parçadan bir tanesi test için kullanılırken, diğer n-1 parça modelin eğitimi için kullanılır. Elde edilen değerlerin ortalaması, çapraz geçerlilik yöntemi için değerlendirilir (20). RF modellerinde 10 ağaç kullanılarak sonuçlar ele alınmıştır. Ayrıca, performans ölçütlerinden doğruluk, sınıflama hatası, kappa istatistiği, ROC eğrisi altında kalan alan (AUC), f ölçümü, duyarlılık, seçicilik ve Youden indeksi kullanılarak, modellerin performans sonuçları yorumlanmıştır. Modelleme ve analizler Rapidminer Studio Free 8.1.000 versiyonu ile yapılmıştır.

Biyoistatistiksel analiz

Veriler uygun olduğu yerlerde ortalama \pm standart sapma ve medyan (min-maks) ile özetlendi. Verilerin normal dağılıma uygunluğunu Kolmogorov Smirnov testi ve varyansların homojenlik kontrolü ise Levene testi ile değerlendirildi. Verilerin analizinde bağımsız örneklerde t-testi ve Mann-Whitney U testi uygun yerlerde kullanıldı. Analizlerde IBM SPSS Statistics version 24.0 for Windows paket programı kullanıldı. $p < 0.05$ değeri istatistiksel olarak önemli kabul edildi.

Sonuçlar

Bu çalışmada incelenen hidrosefalinin varlığı ve yokluğu incelendiğinde, hidrosefalisi olan bireylerin sayısı 81 (%60,9) ve olmayanların sayısı 52 (%39,1) olarak gözlemlendi. Bağımsız değişkenlerimizin bağımlı değişken olan hidrosefalinin varlığı ve yokluğu durumuna göre dağılımları Tablo 3'de verilmiştir. Bu bulgulara göre baş çevresi değişkeni açısından hidrosefalinin varlığı ve yokluğu arasında istatistiksel olarak anlamlı bir fark vardır ($p < 0,05$). Fakat doğum ağırlığı, anne yaşı ve doğum haftası değişkenleri açısından bağımlı değişkenimiz olan hidrosefalinin varlığı ve yokluğu arasında istatistiksel olarak anlamlı bir fark yoktur ($p > 0,05$).

Tablo 3. Değişkenlerimizin hidrosefali varlığına ve yokluğuna göre dağılımı

Değişkenler	Hidrosefali		P
	Yok (n=52)	Var (n=81)	
Anne yaşı (yıl)	29.83±6.51	29.83±6.239	<1
Doğum ağırlığı (gr)	2997.69±545.357	3147.78±633.040	0.162
Doğum haftası	38.5(28-40) ^a	38(31-40) ^b	0.865
Baş çevresi (cm)	34(30-40)	37(30-50)	<0.001

Veriler ortalama \pm standart sapma ya da ortanca (min-maks) olarak özetlendi. APA stiline göre her satırda farklı harfleri taşıyan değerler, istatistiksel olarak birbirinden farklıdır ($p < 0.05$).

Farklı çekirdek fonksiyonları ile oluşturulan DVM modellerinin çeşitli performans ölçütlerine göre sınıflandırma performansları ve farklı dallanma kriterleri baz alınarak oluşturulan RF modelinin çeşitli performans ölçütlerine göre sınıflandırma performansları (% 95 güven aralığı ile) Tablo 4’de verilmiştir.

Bu tabloya göre en iyi sonuçları ANOVA çekirdeği ile elde edilen DVM ve Gini indeksi kriterine göre oluşturulan RF modeli vermiştir.

Hidrosefalinin sınıflandırılmasında en iyi sonuçları veren ANOVA çekirdeği ile elde edilen DVM ve Gini indeksi kriterine göre oluşturulan RF modellerine göre hangi bağımsız değişkenin bağımlı değişken üzerinde daha etkili olduğunu belirlemek amacıyla kullanılabilen ağırlıklar Tablo 5 ile verilmiştir.

Bu sonuçlara göre hidrosefali üzerinde en etkili değişken her iki modelden de baş çevresi olarak tahmin edilmiştir.

Tablo 4. DVM’nin tüm çekirdek fonksiyonları ve RF’nin farklı dallanma kriterleri için oluşturulan modellerin çeşitli performans metriklerine göre sınıflandırma performansları (%95 güven aralığı ile)

Model	Çekirdek Tipi/Kriter	Doğruluk	Sınıflama Hatası	Kappa İstatistiği	AUC	F ölçümü	Duyarlılık	Özgüllük	Youden indeksi
DVM	Dot	%69.07±6.47	%30.93±6.47	0.372±0.139	0.775±0.096	%72.56±6.36	%67.98±9.18	%71.33±17.90	0.393±0.161
	Radial	%70.77±5.71	%29.23±5.71	0.360±0.142	0.724±0.121	%76.67±5.08	%80.20±12.78	%55.33±23.72	0.355±0.158
	Polinomial	%60.88±4.37	%39.12±4.37	0.018±0.107	0.611±0.175	%75.26±2.61	%97.78±4.44	%4.00±12.00	0.018±0.096
	Neural	%60.82±10.73	%39.18±10.73	0.191±0.211	0.612±0.157	%64.57±13.99	%62.34±17.69	%57.33±19.43	0.197±0.216
	ANOVA	%72.14±8.99	%27.86±8.99	0.424±0.209	0.748±0.120	%75.78±6.97	%71.45±8.27	%73.00±22.13	0.444±0.221
	Epachnenikov	%66.32±7.66	%33.68±7.66	0.213±0.202	0.670±0.134	%76.19±5.35	%89.17±11.48	%31.00±22.31	0.202±0.194
RF	Kazanç oranı	%75.82±10.93	%24.18±10.93	0.515±0.228	0.776±0.077	%77.96±9.63	%70.34±9.83	%84.67±20.88	0.550±0.248
	Bilgi kazanımı	%75.27±8.70	%24.73±8.70	0.502±0.197	0.802±0.104	%77.3±7.53	%70.65±14.25	%82.67±23.89	0.533±0.205
	Gini indeksi	%77.47±9.51	%22.53±9.51	0.554±0.193	0.799±0.100	%78.89±9.43	%71.90±15.10	%87.00±19.23	0.589±0.197
	Doğruluk	%74.51±8.17	%25.49±8.17	0.495±0.161	0.859±0.083	%76.33±8.25	%69.23±12.13	%83.67±18.28	0.529±0.173

RF: Random Forest; DVM: Destek Vektör Makinesi; AUC: ROC eğrisi altında kalan alan

Tablo 5. Hidrosefaliyi etkileyen bağımsız değişkenlerin ağırlık tablosu

Değişkenler	RF (Gini İndeksi)	DVM (ANOVA)
Baş çevresi (cm)	0.407	0.909
Doğum ağırlığı (gr)	0.308	0.104
Anne yaşı (yıl)	0.259	0.023
Doğum haftası	0.211	0.248

RF: Random Forest; DVM: Destek Vektör Makinesi.

Tartışma

Pediyatrik hidrosefali özellikle yeni doğan bebeklerde sık görülen ve birçok yönden hala tam olarak anlaşılmamış bir hastalıktır. Bu hastalığın en önemli özelliği, çocukluktan erişkinliğe kadar süregelen kronik bir rahatsızlık olması ve hayatın her döneminde gelişimsel sorunlara yol açabilmesidir (21).

Hidrosefali etiyolojilerinin ve hidrosefaliye eşlik eden merkezi sinir sistemi malformasyonlarının çok çeşitli olması, tedavi seçeneklerinin, tedavide kullanılan şant tiplerinin ve basınçlarının farklılığı, uzun süreli izlem gerekliliği gibi nedenlerle hasta seçiminin standardizasyonunun neredeyse mümkün olmaması, prognoz ve izlem konusundaki yayınların yeterli sonuçlara ulaşmasını engellemektedir. İlgili literatür incelendiğinde, endoskopik 3. ventrikülostomi, intrauterin tedavi gibi yöntemlerin kullanıma girmesinin, şantlarla ve ameliyathanelerle ilgili teknik gelişmelerin, cerrahi sonuçların iyileşmesine neden olduğu,

bunların çocukların duygusal, bilişsel, zeka ve sosyal sağlık durumuna pozitif etki gösterdiği görülmekte, ancak bu etkinin tam olarak neden kaynaklandığı ve ne kadar olduğu bilinmemektedir. Bununla birlikte, görüntüleme yöntemleri, nörofizyoloji ve moleküler biyolojideki son dönem gelişmeler, daha iyi sonuçların yakın bir gelecekte olduğuna işaret etmektedir (22).

Bu çalışmada hidrosefali olan ve olmayan bebeklerden oluşan bir örneklem için hidrosefaliye neden olabileceği düşünülen değişkenlerin değişik çekirdek fonksiyonları ile geliştirilen DVM modelleri ve farklı dallanma kriterleriyle oluşturulan RF modellerinin tahmin performanslarının değerlendirilmesi amaçlanmıştır. Bu bağlamda 8 farklı çekirdek fonksiyonu ile oluşturulan DVM modelleri elde edilmiştir. Farklı çekirdek fonksiyonları ile oluşturulan DVM modelleri arasında hidrosefaliyi sınıflamada belirlenen performans

ölçütleri (doğruluk, sınıflama hatası, kappa istatistiği, AUC, f ölçümü, duyarlılık, özgüllük, Youden indeksi) dikkate alındığında, en iyi sınıflandırma performansı ANOVA çekirdek fonksiyonu ile oluşturulan DVM modeline aitti. Performans ölçütlerine ilişkin değerler (%95 güven aralığında) sırasıyla; doğruluk için %72.14±8.99, sınıflama hatası için %27.86±8.99, kappa istatistiği için 0.424±0.209, AUC için 0.748±0.120, f ölçümü için %75.78±6.97, duyarlılık için %71.45±8.27, seçicilik için %73.00±22.13, Youden indeksi için 0.444±0.221 olarak elde edilmiştir.

DVM modeline ayrı olarak, 4 farklı dallanma kriteri kullanılarak oluşturulan RF modeli için belirlenen performans ölçütleri incelendiğinde, en iyi sınıflandırma performansı Gini İndeksi kullanılarak oluşturulan RF modelinden elde edilmiştir. Performans ölçütlerine ilişkin değerler (%95 güven aralığında) sırasıyla; doğruluk için %77.47±9.51, sınıflama hatası için %22.53±9.51, kappa istatistiği için 0.554±0.193, AUC için 0.799±0.100, f ölçümü için %78.89±9.43, duyarlılık için %71.90±15.10, seçicilik için %87.00±19.23, Youden indeksi için 0.589±0.197 olarak elde edilmiştir.

İki sınıflandırma modeline ait bulgular birlikte ele alındığında, Gini İndeksi ile oluşturulan RF modelinin, ANOVA çekirdeği ile oluşturulan DVM'e göre daha iyi sınıflandırma performansına sahip olduğu belirlenmiştir. Ayrıca, hidrosefali ile ilişkili faktörler, makine öğrenmesi yöntemlerinden DVM ve RF algoritmaları ile belirlenerek, bu faktörlerin söz konusu hastalık üzerindeki etkileri tahmin edilmiştir.

En iyi performansı gösteren RF modeli sonuçlarına göre hidrosefaliye neden olabilecek en önemli faktör baş çevresi olarak elde edilmiştir. Klinik olarak, hidrosefalinin artmış beyin ventrikül büyüklüğüne eşlik ettiği ve eğer kranial dikişler henüz kapanmamışsa, hastanın baş çevresinin artması sonucunda ortaya çıkan bu durumun her 1000 canlı doğumdan 0.8 vakada gerçekleştiği öngörülmektedir (23).

Hidrosefalinin sınıflandırılmasında, bu hastalığı etkileyebilecek en önemli ikinci faktörün doğum ağırlığı olduğu belirlenmiştir. Prematüre doğum ve düşük doğum ağırlığının bir sonucu olarak ortaya çıkabilen intraventriküler kanamalar, hidrosefali için en sık görülen perinatal nedenlerdendir. Araştırmacıların büyük çoğunluğu, bu tip kanamaların, doğum ağırlığı 1500 gr olan yeni doğan bebeklerin yaklaşık

%40'ında meydana geldiğini öngörmektedir (24).

Anne yaşının hidrosefaliye neden olabilecek faktörler arasında olması sonucu, bu çalışmada oluşturulan RF modeli çıktılarına dayanarak çıkarılmıştır. Bilindiği kadarıyla, ileri anne yaşı ile hidrosefali arasında klinik olarak doğrudan bir ilişki olduğu henüz gösterilmemiştir.

Bu çalışmada hidrosefalinin sınıflandırılmasında diğer önemli bir faktörün de doğum haftası olduğu belirlenmiştir. Erken doğan bebeklerin nörogelişimsel bozukluklara maruz kaldığı ve bunun sonucu olarak da ağır nörolojik sekeller arasında yer alan hidrosefalinin ortaya çıkması söz konusu olmaktadır (25).

İlerleyen çalışmalarda, daha büyük klinik verilerde değişik çekirdek fonksiyonlardan elde edilen DVM ve farklı dallanma kriterleri kullanılarak oluşturulan RF modellerine ek olarak diğer makine öğrenmesi algoritmalarının kullanılması hastalıkların sınıflandırma başarısını daha da artırabilir.

Kaynaklar

1. Fishman MA. Hydrocephalus. In: Eliasson SG, Prensky AL, Hardin WB (editors). Neurological Pathophysiology. New York, NY: Oxford University Press; 1978.
2. Carey CM, Tullous MW, Walker ML. Hydrocephalus: Etiology, pathologic effects, diagnosis and natural history. *Pediatr Neurosurg* 1994; 3: 185-201.
3. Bilginer B, Çataltepe O. Hidrosefali: Sınıflama, Patofizyoloji ve Tedavisi. Korfalı E, Zileli M (editors). TND Temel Nöroşirürji, 2. baskı, Ankara: TNDER, 2010: 1899-910.
4. Canaz H, Alataş İ, Batçık OE, Akdemir AO, Baydın S. Erken Çocuklukta Hidrosefali. *Kafkas Tıp Bilimleri Dergisi* 2013; 2: 88-95.
5. Ertuğrul İ, Organ A, Şavlı A. Veri madenciliği uygulamasına ilişkin PAÜ hastanesinde hasta profilinin belirlenmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi* 2013; 19(2): 97-103.
6. Kahramanlı H. Hibrit bulanık sinir ağını kullanarak bir sınıflandırma ve kural çıkartma sisteminin geliştirilmesi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Doktora tezi, 2008.
7. Vapnik V. The nature of statistical learning theory, Springer science, business media, 2013.
8. Güldoğan E, Arslan AK, Yağmur J. Çeşitli Çekirdek Fonksiyonları ile Oluşturulan Destek Vektör Makinesi Modellerinin Performanslarının İncelenmesi: Bir Klinik Uygulama. *Fırat Tıp Dergisi* 2017; 22(3): 136-42.
9. İplikci S. Controlling the experimental three-tank system via support vector machines.

- In International Conference on Adaptive and Natural Computing Algorithms Springer, Berlin, Heidelberg: 2009; 391-400.
10. Kaban Z, Diri B. Genre and author detection in Turkish texts using artificial immune recognition systems. In Signal Processing, Communication and Applications Conference 2008; 2008. SIU 2008. IEEE 16th:1-4.
 11. Breiman L. Random forests, Machine learning 2001; 45(1): 5-32.
 12. Akman M, Genç Y, Ankaral H. Random forests yöntemi ve sağlık alanında bir uygulama. Türkiye Klinikleri J Biostatı 2011; 3(1): 36-48.
 13. Çoma E. Destek vektör makinelerinin etkin eğitimi için yeni yaklaşımlar. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Doktora tezi 2008.
 14. Yakut E, Elmas B, Yavuz S. Yapay sinir ağları ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 2014; 19(1): 139-57.
 15. Kecman, V. Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. Cambridge: MIT press, 2001.
 16. Mather PM, Koch M. Computer processing of remotely-sensed images: An introduction. 4th edition, John Wiley, Sons, 2011.
 17. Çölkesen İ. Uzaktan Algılamada İleri Sınıflandırma Tekniklerinin Karşılaştırılması ve Analizi. Gebze Yüksek Teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi: 2009.
 18. Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens 2005; 26(1): 217-22.
 19. Korkem E. Mikroarray Gen Ekspresyon Veri Setlerinde Random Forest Ve Naive Bayes Sınıflama Yöntemleri Yaklaşımı, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Yüksek Lisans Tezi 2013.
 20. Bishop C, Bishop CM. Neural networks for pattern recognition. Oxford University Press, 1995.
 21. Zahl SM, Egge A, Helseth E, Wester K. Benign external hydrocephalus: a review, with emphasis on management. Neurosurg Rev 2011; 34(4): 417-32.
 22. Özışık PA. Hidrosefali Prognoz ve İzlem. Türkiye Klinikleri J Neurosurg Special Topics 2015; 5(1): 83-6.
 23. Zielińska D, Rajtar-Zembaty A, Starowicz-Filip A. Cognitive disorders in children's hydrocephal Neurologia Neurochir Polska 2017; 51(3): 234-9.
 24. Liechty EA, Bull MJ, Bryson CQ, et all. Developmental outcome of very low birth weight infants requiring a ventriculo-peritoneal shunt. Pediatr Neurosurg 1983; 10(5): 340-9.
 25. Göçer C, Kavuncuoğlu S, Arslan G, et. all. Çok düşük doğum ağırlıklı erken doğmuş bebeklerin nörogelişimsel sorunları ve nörolojik hastalığa etki eden etmenlerin araştırılması. Türk Pediatri Arşivi 2011; 46(3): 207-14.

Sorumlu Yazar:

Arş. Gör. Zeynep TUNÇ

İnönü Üniversitesi, Tıp Fakültesi,
Biyostatistik ve Tıp Bilişimi Anabilim Dalı,
Malatya, Türkiye.

E mail: zeynep.tunc@inonu.edu.tr