

Doğrusal, Karesel ve Esnek Ayırma Analizlerinin Türetilmiş ve Gerçek Veriler Üzerinde Karşılaştırılması

Comparison of Linear, Quadratic and Flexible Discriminant Analysis by Using Generated and Real Data

Ahmet Öztürk, PhD.

Department of Biostatistics
Erciyes University, Medical Faculty
ahmetozturk@erciyes.edu.tr

Kazım Özdamar, PhD.

Department of Biostatistics
Eskişehir Osmangazi University, Medical Faculty
ozdamar@ogu.edu.tr

This study was orally presented in the X National Biostatistics Congress September, 5-7th, 2007.

Submitted : December 17, 2007
Revised : January 18, 2008
Accepted : March 14, 2008

Corresponding Author:

Ahmet Öztürk, PhD.
Department of Biostatistics,
Erciyes University, Medical Faculty
Kayseri, Turkey

Telephone : +90- 352 4374937-23476
E-mail : ahmetozturk@erciyes.edu.tr

Özet

Amaç: Bu çalışmada parametrik ayırma analizi yöntemleri olan Doğrusal Ayırma Analizi (DAA) ve Karesel Ayırma Analizi (KAA) ile parametrik olmayan Esnek Ayırma Analizi'nin (EAA) türetilmiş ve gerçek veri setlerinde performansları karşılaştırıldı.

Gereç ve Yöntemler: Araştırmada tıp alanında gerçek ve türetilmiş veri setleri kullanıldı. Türetilmiş veri setleri, grup sayısı $g=2, 3$ ve 4 olacak şekilde, 4 bağımsız değişken içeren farklı ortalama vektörleri ve varyans-kovaryans matrislerine sahip çok değişkenli normal dağılım gösteren ve göstermeyen koşullarda Monte Carlo simülasyon yöntemi kullanılarak 1000 tekrar yapılarak oluşturuldu.

Bulgular: Türetilmiş veri setlerindeki gruplar kendi içinde homojen yapıda ve aralarında heterojen yapıda ise tüm yöntemler daha iyi performans gösterdi. Grup merkezleri birbirine yakın olduğunda; eğer varyanslar küçük ise, yöntemlerin Doğru Sınıflandırma Oranları'nın (DSO) daha yüksek olduğu saptandı. Bununla beraber birimlerin özellikleri birbirlerine karıştığında KAA, DAA'ya göre daha etkin DSO değerleri verdi. EAA yöntemlerinin ise genellikle farklı koşullarda DAA'ya göre yüksek, ancak KAA'dan daha düşük performansı olduğu belirlendi. Tıp alanında 6 sette, EAA ve DAA performansı KAA yöntemine göre daha yüksek iken 4 sette yöntemler benzer sonuçlar verdi.

Sonuç: Normal dağılımayan ve heterojen veri yapılarında EAA yönteminin, normal dağılan v homojen veri yapılarında DAA'nın tercih edilmesi gerektiği, normal dağılan fakat büyük değerli varyanslara sahip heterojen dağılımlarda ise KAA'nın ayırma işlemlerinde kullanılmasının uygun olacağı belirlendi.

Anahtar Kelimeler: **Ayırma Analizi; Çok Değişkenli Analiz; Monte Carlo Yöntemi; Sınıflandırma.**

Abstract

Purpose: Parametric Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and nonparametric Flexible Discriminant Analysis (FDA) were compared by using generated and real data sets.

Material and Methods: The research used generated and six medical data sets. A thousand repeated data were produced by Monte Carlo simulation method under multivariate normal and nonnormal distributed conditions, different mean vectors and variance-covariance matrices of containing four independent variables as in groups $g=2, 3$ and 4.

Results: It was determined that all the methods demonstrated good performances in the data homogeneous within groups and heterogeneous between groups. When the mean vectors of groups were similar and variances were small, it was determined that all methods showed high accurate classification rates. When overlapping in the data, QDA gave more efficient performance than LDA. Methods of FDA gave more performance than LDA, but gave less performance than QDA. That FDA and LDA performance of two data compared in six medical data performance gave high performance according to QDA, while 4 data gave similar results methods.

Conclusion: We concluded that the use of FDA methods in the nonnormal and heterogeneous sets and in the normal and homogeneous sets LDA should be used, but heterogeneous sets QDA may be used.

Keywords: **Classification; Discriminant Analysis; Monte Carlo Method; Multivariate Analysis.**

Giriş

Toplum içinde var olan doğal ya da gizli grupları (alt toplum) ortaya çıkarmak tıbbi sorunların çözümünde büyük önem taşımaktadır. Özellikle belirtileri birbirlerine benzeyen ancak değişkenlerin değişimlerinde küçük farklılıkların olduğu durumlarda tanı ve tedaviye ulaşmak oldukça güçleşmektedir. Hastalık tiplerinin belirlenmesinde yararlanılacak ayırma kriterlerini belirlemek, tanı ve tedavide etkin, tutarlı ve yüksek güvenilirlikte kararlar üretmek için gereklidir (1-4).

Çok değişkenli problemlerin incelenmesinde normal dağılım varsayımını kullanmak alışılmış ve yaygın kullanılan bir yaklaşım olmasına karşın, parametrik istatistiksel yöntemler (Doğrusal Ayırma Analizi, DAA; Karesel Ayırma Analizini, KAA) yanında normal dağılım varsayımı gerektirmeyen parametrik olmayan çok değişkenli istatistiksel yöntemler (Esnek Ayırma Analizi, EAA) de geliştirilmiştir (3, 5-11). Parametrik Ayırma analizleri çok sayıda varsayım gerektirmektedir. Bu koşulların her zaman tıbbi verilerce karşılanması mümkün olmamakta; daha sınırlı, parametrik olmayan ayırma analizlerine gereksinim duyulmaktadır. Bu nedenle EAA'nın karma veri yapıları içeren tıbbi problemlerde uygulanabilirliğinin araştırılması önemlidir (5, 12-14).

Literatürde EAA'nın tıbbi problemlere uygulanışı ile ilgili sınırlı sayıda yayın bulunmaktadır. Ancak bu yayınlarda, veri ve değişken sayısı ile dağılım varsayımlarının yöntemlerin performansına etkileri yeterince incelenmemiştir. DAA ya da KAA yöntemlerine uygun olan veri yapılarında, EAA'nın performansını değerlendirmek ve DAA ya da KAA'nın uygulanmadığı tıbbi problemlerde ayırma ve gruplama işlemlerinin EAA yöntemlerinden hangisi ile yapılmasının uygun olacağını belirlemek gerekli görülmektedir (6,10).

Bu çalışmada "sınıflama ve gruplama" amacıyla kullanılan parametrik olmayan çok değişkenli istatistiksel yöntemlerin performanslarının klasik yöntemler olan Doğrusal ve Karesel ayırma analizleri ile, benzetim çalışmalarından elde edilen veriler ve tıp alanı ile ilgili gerçek veriler üzerinde karşılaştırılması amaçlandı.

Gereç ve Yöntemler

Araştırmada gerçek ve türetilmiş verilerden yararlanıldı.

Veri Türetimi. Veri, bir bağımlı değişken ve 4 bağımsız değişken içeren ortalama vektörleri ve kovaryans matrisleri farklı olan çok değişkenli normal dağılım gösteren ve

göstermeyen koşulları sağlayacak biçimde, Monte Carlo simülasyon yöntemi kullanılarak, 1000 tekrarlı setler olarak türetildi (11).

Analiz edilecek veri setinde grup sayısı (g) 2, 3 ve 4 olacak şekilde X veri matrisleri oluşturuldu. Her bir türetim için veri setlerini elde etmek için grup sayısı dikkate alınarak farklı parametrelili X_i ($i=1, g$) veri matrisleri alta birleştirildi. Veri setlerinin oluşturulmasında yararlanılan farklı grup ortalama vektörleri ($\mu_1=(0,0,0,0)$, $\mu_2=(0.5,0.5,0.5,0.5)$, $\mu_3=(1,1,1,1)$, $\mu_4=(2,2,2,2)$, $\mu_5=(0.5,1,1.5,2)$, $\mu_6=(0.5,1,2,3)$, $\mu_7=(2.5,3,3.5,4)$, $\mu_8=(3,3,3,3)$, $\mu_9=(4,4,4,4)$, $\mu_{10}=(4.5,5,5.5,6)$, $\mu_{11}=(6,6,6,6)$) belirlendi. Türetimlerde farklı varyans ve kovaryans düzeylerinin yöntemlerin sınıflama performansına etkilerini incelemek için farklı kovaryans matrisleri düzenlendi. Bu kovaryans matrislerinde varyansların 1, 3 ve 10 olduğu fakat kovaryansların 0, 0.25, 0.75 ve 0.90 olduğu durumlar (V1 ile V12 arası kovaryans matrisleri) belirlendi. Ortalama vektörler ve grupların farklı kovaryans matrisleri kullanılarak her grupta eşit olmak üzere 50, 100, 250, 500 ve 1000 birimlik veri türetildi. Grup sayısına göre farklı ortalama vektörü ve kovaryans matrislerine göre türetilmiş X_i veri matrisleri birleştirilerek 13 tane çok değişkenli normal dağılım gösteren ve 2 tane normal dağılım göstermeyen koşullarda 15 veri seti (SET) belirlendi.

SET olarak isimlendirilen her analiz veri setinde çok sayıda birime ilişkin homojen yapıda gruplardan oluşmuş veri setleri, kovaryans matrisinin seçimine bağlı olarak ortalamaları yakın ancak kovaryans matrisleri heterojen olan üst üste binmiş (overlap) karmaşık sınıfları içeren veri setleri belirlendi. Ayırma yöntemlerinin homojen ve farklı konumlardaki sınıfları içeren veri setleri ve heterojen yapıda ve üst üste çakışan sınıfları içeren veri setlerinde yöntemlerin ayırma performansları hesaplandı.

Yöntemlerin ayırma performanslarını belirlemek için doğru sınıflama oranları (DSO) hesaplandı. DSO; birimlerin türetildiği gruplar gerçek gruplar varsayılarak, birimlerin gerçekte ait olduğu grup ile model aracılığı ile atandığı grubun benzer olduğu durumlara ilişkin birim sayısının (nDSO) gruplardaki toplam birim sayısına (N) bölünmesi ile elde edildi. Ayrıca çok değişkenli normal dağılım göstermeyen iki veri seti daha türetildi.

Veri Analizi Yöntemleri. Oluşturulan 158 türetilmiş veri setinin her biri DAA, KAA ve EAA'nın yaygın kullanımı olan EAA/MARS degree=1, EAA/MARS degree=2 ve

EAA/BRUTO yöntemlerine göre analiz edildi. Verilerin analizinde DAA ve KAA'yı uygulamak için R 2.3.1 istatistik paket programının (The R Project for Statistical Computing) MASS modülünden yararlanıldı. EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemleri için ise R paket programının MDA (Mixture Discriminant Analysis, Karma Ayırma Analizi) modülünden (packages) yararlanıldı (11). Pakette yer almayan bazı makrolar araştırmacılar tarafından yazıldı ve programa entegre edildi. MedCalc paket programı kullanılarak, ayırma yöntemlerinin gerçek veri setlerindeki performansları ROC (receiver operator curve) eğrisi yöntemi ile karşılaştırıldı.

Doğrusal Ayırma Analizi (DAA) p değişkenli X veri setindeki değişkenlerin iki ve daha fazla gerçek gruba ayrılmasını ve belirlenen ayırma fonksiyonları aracılığı ile yeni gözlenen birimlerin belirlenen sınıflara doğru olarak atanmasını sağlayan bir yöntemdir (1-4,15). Örnek kovaryans matrislerinden g grup sayısı olmak üzere S1 ve S2, ..., Sg'nin birleşimi olan Sp (pooled variance-covariance matrix) hesaplandı. Grupların varyans-kovaryans matrislerinin birbirlerinden farklı olduğu durumda, veri matrislerinin ayırma analizinde KAA (Quadratic Discriminant Analysis) uygulandı (1-4,15).

Karesel Ayırma Analizi (KAA). Verilerin normal dağıldığı ancak, grupların varyans-kovaryans matrislerinin birbirlerinden farklı olduğu durumda, veri matrislerinin ayırma analizinde KAA (Quadratic Discriminant Analysis) uygulanır. KAA'da katsayıların hesaplanmasında ortak kovaryans matrisi yerine (s) grupların kovaryans matrislerinin farkları alındı (1-4,15).

Esnek Ayırma Analizleri (EAA) normal dağılım göstermeyen, çok değişkenli toplumlardan alınan X veri matrislerinin sınıflandırılmasında yararlanılan parametrik olmayan bir yöntemdir. Ayırmada yararlanılan yaklaşımlara bağlı olarak farklı isimlerle anılan EAA yöntemleri vardır. Bu yöntemlerden MARS (Multivariate Adaptive Regression Splines) ve BRUTO (adaptive additive modeling) sunulan çalışmada kullanılmıştır. EAA/MARS yöntemi uygulamalarında iki farklı algoritma (degree=1 ve degree=2) kullanılmıştır. BRUTO modeli, budama (backfitting) olarak da bilinen yinelemeli Gauss-Seidel algoritması aracılığı ile tahmin edildi (6-10, 16)

Gerçek veri uygulaması: GEBE veri seti. Toplam 20 soru içeren 0-5 arasında puanlandırma yapılan likert tipi

Quebek ölçeği ile toplam 10 soru içeren 0-5 arasında puanlandırma yapılan likert tipi Oswestry ölçeklerinin 1047 gebeye uygulanması ile elde edilmiş bir veri setidir. Quebec ölçeği bel ağrısının günlük yaşam aktivitelerini nasıl etkilediğini gösterir. Oswestry ölçeği ise bel ağrısının günlük aktivitelerin üstesinden gelme becerilerini belirler. Bağımlı değişken gebelerde bel ağrısı olup olmaması şeklinde iki düzeylidir.

BKİ veri seti. 6-17 yaş arası 5358 okul çocuğunun bel çevresi, ortakol çevresi, triseps deri kıvrım kalınlığı, bel boy oranı, kol yağ alanı, beden kitle indeksinin (BKİ, kilo/boy²) bulunduğu veri setidir. Bağımlı değişken kategorileri BKİ'nin 85'inci percentilin yukarısında olanlar ve olmayanlar olarak ifade edilmiştir.

YAĞ1 veri seti. 6-17 yaş arası 5358 okul çocuğunun kol yağ alanının, 3 antropometrik ölçümlerinin (bel çevresi, bel boy oranı, BKİ) bulunduğu veri setidir. Bağımlı değişken (kol yağ alanı) kategorileri 5'inci percentilin altı olanlar, 5-85'inci percentil arasında olanlar, 85-95'inci percentil arasında olanlar, 95'inci percentil yukarısında olanlar olarak dört düzeyde ifade edildi. Ayrıca bağımlı değişken ikinci bir analiz için (YAĞ2) 85'inci percentilin altı olanlar 85-95'inci percentil arasında olanlar, 95'inci percentil yukarısında olanlar olarak üç düzeyde ifade edilmiştir.

DDİMER veri seti. Acil cerrahi girişim gereken hasta ve kontrollerden oluşan toplam 77 hastanın D-dimer kan değerinin ölçüldüğü veri setidir. Bağımlı değişken 4 alt gruptur (1-Mezenter damar hastalığı nedeniyle acil cerrahi girişim gerektiren hastalar, 2-Mezenter damar hastalığı olmayan ama acil cerrahi girişim gerektiren başka patolojisi olan hastalar, 3-Acil cerrahi girişim gerekmeyen cerrahi hastalığı olan hastalar ve 4- Sağlıklı bireyler).

İNS veri seti. 6-20 yaş arası 279 kişinin HOMA indeksi değerinin ölçüldüğü veri setidir. Bağımlı değişken kan insülin değeridir ve 300'ün altında olanlar ve olmayanlar olarak iki alt gruptur.

Bulgular

Ayırma Analizi Yöntemleri'nin (DAA, KAA, EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO), değişik koşullara göre çokdeğişkenli normal dağılımdan türetilen setlerin DSO değerleri Tablo I-IV arasında verilmiştir. Bu tablolarda yer alan her bir set oluşturulan koşullara göre yöntemlerin verdikleri doğru sınıflandırma

oranlarına (DSO) bakılarak performans değerlendirmeleri yapılmıştır. Tablolarda X matrisinin birim sayısına göre ve kovaryans matrislerinin içeriklerine göre belirlenen DSO değerleri irdelenmiştir.

Tablo I'de 5 ayırma analizi yönteminin kovaryanslarının sıfır olduğu durumda belirlenen sonuçlar verildi. Beş ayırma yönteminin değişik koşullara göre SET'lerde hesaplanan DSO değerleri varyanslara göre hesaplandı. Varyans 1 iken SET 1'de alt grupta N=100 değerinin üzerine çıktığında bütün yöntemlerin DSO değerlerinin arttığı, diğer setlerde ise herhangi bir değişiklik olmadığı saptandı. Bu 5 veri setinde, varyansın 1 olduğu koşulda bütün yöntemler birbirlerine yakın DSO değerleri gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük DSO değerleri gösterdiği bulundu.

Beş ayırma analizi yönteminin kovaryansı 0.25 ve varyansı 1 olduğu koşulda bütün yöntemlerin DSO değerleri birbirlerine yakınken, varyansın 3 ve 10 olduğu koşulda DAA ile diğer yöntemlerden daha düşük DSO değerleri elde edildi (Tablo II).

Beş veri setinde kovaryansı 0.75 ve varyansın 1 olduğu koşulda bütün yöntemler birbirlerine yakın DSO değerleri elde edilirken, varyansın 3 ve 10 olduğu koşulda DAA ile diğer yöntemlerden daha düşük DSO değerleri elde edildi. KAA'nın ise en iyi sınıflandırma performansını gösterdiği belirlendi (Tablo III).

Tablo IV'de 5 ayırma analizi yönteminin kovaryansın 0.90 olduğu durumda elde edilen sonuçlar görülmektedir. Bağımlı değişken 2 gruplu ve kovaryans 0.90 olduğunda, bu 5 veri setinde varyansın 1 olduğu koşulda bütün yöntemler ile birbirlerine yakın DSO değerleri elde edilirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği ve KAA'nın ise en iyi sınıflandırma performansını gösterdiği belirlendi.

Beş ayırma analizi yönteminin 3 gruplu ve kovaryansı 0 ve 0.25 olan 4 veri setinde, varyansın 1 olduğu koşulda tüm yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA diğer yöntemlerden daha düşük bir performans gösterdi, KAA'nın hemen hemen tüm şartlarda performansı diğer yöntemlerden daha yüksek bulundu. Kovaryansın 0.75 ve 0.90 olduğu durumlarda, grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük ise yöntemlerin sınıfları yüksek DSO değerleriyle birbirlerinden ayırdığı belirlendi.

Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) ve varyans arttığında KAA, DAA'ya göre daha etkin; EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük DSO değerleri verdi.

Beş ayırma analizi yönteminin 4 gruplu ve kovaryansı 0 olan 4 veri setinde, KAA ve EAA, DAA'ya göre daha yüksek DSO değerleri verirken EAA, KAA'dan daha düşük DSO değerleri verdi. Kovaryansı 0.25 olan koşulda, yine varyansların küçük olduğu durumlarda yöntemlerin sınıfları yüksek DSO değerleriyle ayırdığı belirlendi. Varyans arttığında KAA, DAA'ya göre daha yüksek DSO değerleri verdi. EAA yöntemleri ile DAA benzer sonuçlar verse de KAA'dan daha düşük DSO değerlerine sahipti. Kovaryansı 0.75 ve 0.90 olan koşulda, yine varyansların küçük olduğu durumlarda yöntemlerin sınıfları yüksek doğruluk oranında ayırdığı belirlendi. Ancak birimlerin özellikleri birbirlerine karıştığında ve varyans arttığında KAA, DAA'ya göre daha doğru sınıflandırma sonuçları verdi. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük DSO değerleri verdi.

Tablo V'de görüldüğü gibi 5 ayırma analizi yönteminin iki ve üç gruplu veri matrisi birleştirmeli sette, varyansın 1 olduğu koşulda tüm yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği belirlendi. EAA 2 alt set haricinde KAA'dan daha yüksek DSO değerleri bulundu. Ancak veri matrisinin normal dağılmadığını göz önüne almak gerekmektedir. EAA modelleri olarak ele alınan yöntemler hem $g=2$ ve $g=3$ için hem de varyans kovaryans yapılarına göre birbirlerine benzer sonuçlar verdi. Ancak BRUTO, her koşulda MARS modellerinden biraz daha düşük DSO değerlerine sahip olduğu bulundu.

Tablo VI'da görüldüğü gibi, 4 ayırma analizi yönteminin tıp alanında 6 veri seti üzerinde performanslarının karşılaştırılmasında, İNS veri setinde, EAA ve DAA'nın DSO değerleri KAA yöntemine göre istatistiksel olarak anlamlı düzeyde daha yüksek ($p<0.05$) iken GEBE, BKİ, YAĞ1, DDİMER ve YAĞ2 veri setlerinde yöntemlerin benzer sonuçlar verdiği belirlendi ($p>0.05$).

Tablo I. Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları sıfır olan (V1,V2,V3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki (n=50, 100, 250, 500 ve 1000) gözlem doğru sınıflandırma oranı değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3
SET 1	100	0.603	0.671	0.661	0.622	0.818	0.946	0.615	0.806	0.924	0.617	0.809	0.928	0.586	0.785	0.918
	200	0.698	0.662	0.645	0.709	0.805	0.940	0.765	0.805	0.924	0.735	0.802	0.930	0.745	0.921	0.921
	500	0.696	0.656	0.635	0.700	0.798	0.939	0.702	0.796	0.920	0.702	0.796	0.928	0.696	0.796	0.922
	1000	0.693	0.654	0.630	0.695	0.796	0.938	0.696	0.792	0.926	0.698	0.792	0.922	0.692	0.794	0.919
	2000	0.692	0.653	0.629	0.694	0.795	0.938	0.695	0.789	0.915	0.696	0.790	0.923	0.693	0.793	0.918
SET 2	100	0.851	0.789	0.746	0.858	0.867	0.952	0.860	0.859	0.930	0.869	0.867	0.938	0.852	0.842	0.925
	200	0.845	0.784	0.738	0.848	0.858	0.947	0.853	0.853	0.930	0.858	0.860	0.938	0.846	0.846	0.928
	500	0.842	0.781	0.735	0.843	0.851	0.944	0.846	0.846	0.926	0.848	0.852	0.934	0.844	0.846	0.926
	1000	0.842	0.780	0.734	0.843	0.850	0.943	0.841	0.842	0.921	0.846	0.847	0.931	0.842	0.844	0.920
	2000	0.842	0.780	0.733	0.843	0.849	0.944	0.843	0.842	0.921	0.843	0.845	0.931	0.843	0.844	0.923
SET 3	100	0.979	0.930	0.862	0.981	0.955	0.967	0.982	0.947	0.950	0.984	0.961	0.961	0.981	0.940	0.944
	200	0.978	0.929	0.858	0.979	0.953	0.965	0.980	0.944	0.947	0.984	0.955	0.958	0.979	0.940	0.945
	500	0.978	0.928	0.858	0.978	0.950	0.963	0.976	0.940	0.942	0.978	0.950	0.954	0.980	0.940	0.942
	1000	0.977	0.926	0.857	0.978	0.948	0.962	0.975	0.935	0.939	0.976	0.946	0.952	0.976	0.936	0.940
	2000	0.977	0.926	0.857	0.977	0.948	0.962	0.975	0.935	0.939	0.975	0.946	0.953	0.976	0.936	0.940
SET 4	100	0.999	0.862	0.918	0.999	0.967	0.983	0.999	0.950	0.969	0.999	0.961	0.980	0.999	0.943	0.963
	200	0.999	0.979	0.916	0.999	0.990	0.982	0.999	0.986	0.967	0.999	0.991	0.977	0.999	0.997	0.965
	500	0.999	0.979	0.914	0.999	0.990	0.981	0.999	0.960	0.960	0.999	0.999	0.974	0.998	0.984	0.960
	1000	0.999	0.978	0.914	0.999	0.989	0.980	0.998	0.984	0.959	0.998	0.989	0.972	0.998	0.984	0.961
	2000	0.999	0.979	0.914	0.999	0.990	0.981	0.999	0.983	0.959	0.998	0.988	0.971	0.998	0.984	0.961
SET 5	100	0.972	0.921	0.852	0.975	0.947	0.965	0.968	0.936	0.970	0.978	0.949	0.958	0.965	0.929	0.928
	200	0.972	0.919	0.850	0.972	0.945	0.963	0.965	0.935	0.946	0.975	0.945	0.956	0.964	0.930	0.943
	500	0.971	0.918	0.848	0.972	0.943	0.961	0.962	0.928	0.942	0.970	0.940	0.952	0.960	0.928	0.940
	1000	0.971	0.916	0.848	0.971	0.941	0.960	0.959	0.927	0.938	0.968	0.937	0.950	0.960	0.925	0.939
	2000	0.971	0.917	0.848	0.971	0.941	0.960	0.960	0.926	0.936	0.967	0.936	0.949	0.958	0.925	0.938

Tablo II. Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.25 olan (V4,V5,V6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki (n=50, 100, 250, 500 ve 1000) gözlem doğru sınıflandırma oranı değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO			
2 GRUPLU	VERİ SETLERİ	$N=n_1+n_2$	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6
SET 1	100	0.651	0.650	0.650	0.680	0.830	0.950	0.675	0.820	0.930	0.687	0.810	0.930	0.643	0.780	0.910	
	200	0.660	0.635	0.627	0.670	0.816	0.948	0.660	0.797	0.922	0.670	0.811	0.929	0.650	0.790	0.919	
	500	0.650	0.630	0.612	0.656	0.806	0.944	0.656	0.786	0.914	0.660	0.800	0.924	0.650	0.786	0.916	
	1000	0.648	0.627	0.610	0.652	0.806	0.944	0.653	0.783	0.911	0.656	0.796	0.923	0.648	0.785	0.914	
	2000	0.648	0.626	0.606	0.649	0.803	0.943	0.650	0.779	0.909	0.651	0.790	0.919	0.648	0.782	0.912	
SET2	100	0.780	0.750	0.720	0.790	0.860	0.960	0.790	0.830	0.930	0.810	0.850	0.930	0.780	0.820	0.920	
	200	0.708	0.742	0.708	0.952	0.850	0.952	0.926	0.832	0.925	0.934	0.848	0.934	0.923	0.823	0.923	
	500	0.776	0.738	0.702	0.780	0.846	0.950	0.782	0.826	0.918	0.786	0.840	0.932	0.780	0.824	0.922	
	1000	0.776	0.739	0.700	0.778	0.846	0.948	0.779	0.823	0.916	0.780	0.836	0.928	0.778	0.824	0.919	
	2000	0.775	0.737	0.699	0.776	0.844	0.948	0.776	0.821	0.914	0.777	0.832	0.926	0.776	0.823	0.917	
SET3	100	0.940	0.900	0.840	0.940	0.940	0.970	0.950	0.930	0.940	0.960	0.940	0.960	0.940	0.920	0.940	
	200	0.937	0.895	0.836	0.938	0.933	0.966	0.941	0.918	0.941	0.947	0.933	0.952	0.939	0.913	0.939	
	500	0.776	0.738	0.702	0.780	0.846	0.950	0.782	0.826	0.918	0.786	0.840	0.932	0.780	0.824	0.922	
	1000	0.935	0.892	0.832	0.936	0.928	0.963	0.935	0.910	0.932	0.937	0.923	0.946	0.934	0.908	0.934	
	2000	0.935	0.892	0.832	0.935	0.929	0.963	0.934	0.908	0.931	0.935	0.921	0.945	0.934	0.908	0.933	
SET4	100	1.000	0.970	0.910	1.000	0.980	0.980	1.000	0.970	0.970	1.000	0.980	0.980	1.000	0.970	0.960	
	200	0.989	0.965	0.907	0.989	0.980	0.981	0.990	0.974	0.961	0.993	0.981	0.976	0.990	0.945	0.959	
	500	0.988	0.964	0.904	0.988	0.978	0.982	0.988	0.970	0.956	0.988	0.978	0.966	0.988	0.970	0.956	
	1000	0.988	0.964	0.905	0.988	0.979	0.980	0.988	0.968	0.954	0.989	0.987	0.965	0.988	0.988	0.955	
	2000	0.988	0.964	0.904	0.988	0.978	0.979	0.987	0.967	0.953	0.988	0.975	0.964	0.987	0.982	0.954	
SET5	100	0.960	0.890	0.840	0.960	0.940	0.960	0.960	0.930	0.940	0.960	0.940	0.950	0.960	0.920	0.940	
	200	0.954	0.901	0.834	0.954	0.935	0.966	0.950	0.920	0.941	0.958	0.933	0.953	0.949	0.914	0.939	
	500	0.952	0.898	0.834	0.952	0.932	0.962	0.946	0.914	0.936	0.952	0.928	0.950	0.948	0.914	0.938	
	1000	0.952	0.897	0.833	0.952	0.932	0.963	0.946	0.912	0.933	0.952	0.926	0.946	0.945	0.910	0.935	
	2000	0.832	0.897	0.832	0.962	0.931	0.962	0.931	0.911	0.931	0.944	0.924	0.944	0.933	0.910	0.933	

Tablo III. Ayırma Analizi Yöntemleri'nin, iki alt grulu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.75 olan (V7,V8,V9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki (n=50, 100, 250, 500 ve 1000) gözlem doğru sınıflandırma oranı değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9
SET1	100	0.644	0.632	0.643	0.666	0.911	0.982	0.636	0.762	0.897	0.638	0.861	0.933	0.636	0.810	0.894
	200	0.630	0.615	0.615	0.644	0.905	0.980	0.650	0.766	0.899	0.630	0.858	0.929	0.628	0.753	0.897
	500	0.614	0.605	0.595	0.624	0.901	0.978	0.618	0.758	0.890	0.622	0.846	0.919	0.614	0.752	0.895
	1000	0.612	0.601	0.590	0.616	0.900	0.978	0.615	0.752	0.885	0.617	0.840	0.913	0.612	0.750	0.893
	2000	0.610	0.599	0.586	0.612	0.899	0.978	0.612	0.749	0.883	0.614	0.834	0.910	0.610	0.751	0.893
SET2	100	0.722	0.706	0.693	0.743	0.921	0.983	0.731	0.795	0.905	0.740	0.882	0.939	0.720	0.777	0.903
	200	0.717	0.698	0.674	0.727	0.916	0.982	0.722	0.797	0.905	0.729	0.876	0.933	0.717	0.784	0.902
	500	0.713	0.692	0.667	0.717	0.913	0.980	0.716	0.789	0.895	0.721	0.866	0.923	0.713	0.783	0.900
	1000	0.712	0.692	0.665	0.714	0.911	0.980	0.714	0.783	0.892	0.716	0.860	0.917	0.712	0.782	0.899
	2000	0.712	0.691	0.664	0.713	0.911	0.980	0.713	0.779	0.891	0.714	0.856	0.913	0.712	0.783	0.898
SET3	100	0.875	0.845	0.803	0.880	0.952	0.987	0.882	0.888	0.929	0.896	0.930	0.955	0.877	0.870	0.923
	200	0.872	0.840	0.798	0.875	0.948	0.986	0.877	0.880	0.924	0.886	0.921	0.948	0.874	0.870	0.921
	500	0.868	0.839	0.794	0.869	0.946	0.985	0.870	0.875	0.915	0.874	0.915	0.939	0.869	0.869	0.917
	1000	0.868	0.839	0.793	0.868	0.945	0.984	0.869	0.869	0.913	0.871	0.911	0.933	0.868	0.867	0.916
	2000	0.867	0.838	0.792	0.867	0.944	0.985	0.868	0.866	0.910	0.869	0.909	0.929	0.867	0.867	0.914
SET4	100	0.886	0.934	0.886	0.991	0.978	0.991	0.953	0.952	0.953	0.973	0.973	0.973	0.946	0.934	0.946
	200	0.955	0.931	0.882	0.956	0.975	0.991	0.958	0.945	0.946	0.964	0.967	0.965	0.957	0.963	0.945
	500	0.952	0.929	0.880	0.953	0.974	0.990	0.954	0.940	0.939	0.957	0.960	0.957	0.954	0.939	0.940
	1000	0.952	0.929	0.879	0.952	0.974	0.990	0.953	0.938	0.936	0.954	0.958	0.954	0.954	0.939	0.938
	2000	0.952	0.929	0.879	0.953	0.974	0.990	0.953	0.937	0.930	0.953	0.956	0.952	0.953	0.937	0.937
SET5	100	0.986	0.900	0.823	0.986	0.965	0.988	0.979	0.922	0.928	0.987	0.952	0.957	0.981	0.915	0.922
	200	0.984	0.895	0.819	0.984	0.962	0.987	0.975	0.916	0.927	0.984	0.945	0.952	0.976	0.910	0.923
	500	0.983	0.893	0.816	0.984	0.959	0.985	0.972	0.910	0.917	0.980	0.937	0.944	0.972	0.907	0.919
	1000	0.984	0.893	0.815	0.984	0.959	0.985	0.971	0.909	0.915	0.978	0.934	0.921	0.974	0.906	0.918
	2000	0.983	0.893	0.814	0.983	0.959	0.985	0.970	0.907	0.912	0.978	0.932	0.937	0.968	0.906	0.917

Tablo IV. Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.90 olan (V10,V11,V12 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki (n=50, 100, 250, 500 ve 1000) gözlem doğru sınıflandırma oranı değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUT O		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12
SET1	100	0.635	0.623	0.646	0.687	0.953	0.994	0.652	0.745	0.893	0.637	0.882	0.939	0.640	0.718	0.889
	200	0.596	0.611	0.612	0.627	0.958	0.993	0.605	0.758	0.893	0.603	0.879	0.930	0.588	0.742	0.890
	500	0.608	0.600	0.624	0.616	0.956	0.996	0.612	0.750	0.932	0.614	0.864	0.964	0.606	0.736	0.944
	1000	0.606	0.596	0.585	0.610	0.956	0.993	0.608	0.746	0.879	0.610	0.856	0.911	0.606	0.739	0.890
	2000	0.604	0.595	0.733	0.606	0.955	0.992	0.605	0.743	0.876	0.607	0.852	0.908	0.604	0.740	0.890
SET2	100	0.732	0.699	0.682	0.744	0.966	0.993	0.729	0.796	0.900	0.738	0.900	0.942	0.733	0.774	0.895
	200	0.706	0.689	0.669	0.716	0.962	0.994	0.711	0.789	0.899	0.716	0.891	0.934	0.706	0.774	0.895
	500	0.700	0.684	0.662	0.704	0.960	0.992	0.702	0.780	0.890	0.706	0.878	0.922	0.700	0.770	0.892
	1000	0.699	0.683	0.658	0.701	0.960	0.993	0.701	0.774	0.885	0.703	0.874	0.916	0.700	0.770	0.894
	2000	0.700	0.682	0.657	0.701	0.959	0.993	0.701	0.770	0.883	0.703	0.870	0.911	0.701	0.772	0.895
SET3	100	0.858	0.838	0.797	0.864	0.978	0.995	0.865	0.887	0.925	0.878	0.941	0.958	0.859	0.868	0.918
	200	0.854	0.830	0.790	0.858	0.975	0.995	0.859	0.876	0.918	0.868	0.930	0.949	0.855	0.861	0.916
	500	0.852	0.826	0.784	0.852	0.972	0.996	0.854	0.864	0.910	0.858	0.918	0.938	0.852	0.858	0.912
	1000	0.852	0.826	0.783	0.852	0.972	0.994	0.853	0.860	0.905	0.855	0.913	0.932	0.852	0.858	0.910
	2000	0.852	0.826	0.783	0.852	0.972	0.994	0.852	0.858	0.903	0.853	0.910	0.928	0.852	0.858	0.909
SET4	100	0.946	0.930	0.880	0.947	0.989	0.996	0.949	0.950	0.950	0.962	0.977	0.974	0.948	0.942	0.942
	200	0.874	0.922	0.874	0.997	0.987	0.997	0.942	0.940	0.942	0.965	0.968	0.965	0.940	0.947	0.940
	500	0.944	0.918	0.872	0.944	0.986	0.996	0.944	0.934	0.940	0.946	0.960	0.956	0.944	0.932	0.936
	1000	0.941	0.919	0.871	0.942	0.985	0.996	0.942	0.931	0.930	0.943	0.957	0.952	0.941	0.931	0.933
	2000	0.871	0.918	0.871	0.996	0.985	0.996	0.929	0.929	0.929	0.950	0.955	0.950	0.932	0.930	0.932
SET5	100	0.998	0.908	0.824	0.998	0.987	0.995	0.998	0.931	0.926	0.998	0.966	0.963	0.998	0.924	0.918
	200	0.999	0.903	0.815	0.999	0.984	0.995	0.998	0.923	0.922	0.999	0.957	0.953	0.999	0.917	0.917
	500	1.000	0.902	0.902	1.000	0.982	0.982	0.996	0.920	0.920	0.996	0.950	0.950	1.000	0.916	0.916
	1000	1.000	0.899	0.811	1.000	0.983	0.994	0.997	0.915	0.908	0.997	0.944	0.940	0.998	0.913	0.913
	2000	0.999	0.899	0.811	0.999	0.982	0.995	0.997	0.914	0.908	0.997	0.943	0.937	0.998	0.914	0.914

Tablo V. Ayırma Analizi Yöntemleri'nin, iki ve üç gruplu veri matrisi birleştirmeli ve her biri 3 alt gruplu normal dağılıma uygunluk göstermeyen her alt sette doğru sınıflandırma oranı değerleri

Veri Setleri	Grup	N	Varyans	DAA	KA A	EAA/MARS Degree=1	EAA/MARS Degree= 2	EAA/BRUTO
SET	g=2	200	V1	0.712	0.722	0.730	0.735	0.715
			V2	0.676	0.698	0.705	0.710	0.677
			V3	0.655	0.681	0.682	0.692	0.655
SET	g=3	300	V7	0.888	0.895	0.893	0.903	0.893
			V8	0.603	0.654	0.627	0.647	0.607
			V9	0.815	0.904	0.850	0.863	0.853

Tablo VI. Ayırma Analizi Yöntemleri'nin, Gerçek Verilerdeki doğru sınıflandırma oranı değerleri

Gerçek Veri Setleri	D A A	K A A	EAA/MARS Degree=1	EAA/MARS Degree=2
GEBE veri seti iki gruplu	0.698	0.685	0.711	0.712
BK< veri seti iki gruplu	0.915	0.900	0.917	0.918
YA=1 veri seti üç gruplu	0.803	0.801	0.804	0.802
YA=2 veri seti dört gruplu	0.800	0.794	0.800	0.797
DD<MER veri seti dört gruplu	0.714	0.675	0.676	0.715
<NS veri seti iki gruplu	0.692	0.645	0.717	0.718

Tartışma

Sınıflama, gruplama ve atama yaklaşımları tıbbi problemlerin çözümünde önemli sorunlar yaratmaktadır. Veri yapısı, tipi ve birim sayıları sınıflandırma işlemlerinde farklı yaklaşımların ve modellerin kullanılmasını gerektirmektedir. Verilere uyan geçerli, güvenilir ve doğruluk derecesi yüksek sonuçların elde edilmesinde uygun yöntemin ve uygun modelin seçimi büyük önem taşımaktadır (12, 13,17).

Ancak bugüne değin çalışılan yöntem ve modellerden sınıflamayı %100'e yakın oranda belirleyen bir yaklaşım henüz bulunmamaktadır. Yöntemlerin verilerin ortalama vektörlerine ve kovaryans matrislerine bağlı olarak çok değişik yelpazede yapısal özellik göstermeleri, yöntem etkinliklerini olumsuz etkiler. Her yöntemin uygulanabilmesi için ideal koşullar belirlenmiş olmasına karşın gözlemsel veri setlerinde bu koşulların tümü gerçekleşemez. Bu durum yöntemler içinden veri setinin analizi için en uygun olanını seçme sorunlarını gündeme getirir. Tıbbi problemlerde bu durumların gözlenmesi, tıbbi problemlerin çözümünde uygun yöntem/model seçimlerinin daha titiz yapılmasını gerektirir. Benzer problemlerin çözümü için geliştirilen çok sayıda yöntemden problem tiplerine, verilerin taşıdığı koşullara en uygun

olanını seçmek gerekir. Bu amaçla yapılan çok sayıda araştırmada yöntemin değişik koşullara sahip veri setlerinde performansları araştırılmıştır (5-10, 12, 17,18).

Sunulan çalışmada ortalama vektörleri birbirlerine yakın, varyansları bir olan, değişkenler arası kovaryansları sıfır olan veri setlerinde DAA, KAA, EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemleri %89 ve daha yukarı DSO değeri belirlendi. Çok değişkenli normal dağılımdan türetilmiş ve ortalama vektörleri önemli farklılık gösteren iki gruplu verilerde yapılan araştırmalar ile sunulan araştırma bulguları paralellik göstermektedir (6-10,12,16,19).

Sunulan çalışmada veri setlerinde ortalama vektörleri ve birim sayısı değişmemesine karşın kovaryans matrislerinin heterojen yapı gösterdiği durumlarda, yöntemlerin hemen hemen tümünün performansları önemli düzeyde düşme gösterdi. Kovaryans düzeylerinin 0.75 ve 0.90 olduğu durumlarda DAA yöntemi performansını kaybetmekte ancak KAA yöntemi yüksek performans değerlerine ulaşmaktaydı Bu durum KAA'nın heterojen kovaryans matrislerine sahip çok değişkenli veri setlerinde DAA'ya göre güçlü bir yöntem olduğu görüşünü doğrulamaktadır (6-10).

Sunulan çalışmada grup sayısının iki olduğu ve kovaryans matrislerinin homojen ve düşük kovaryans düzeyleri içerdiği durumlarda EAA yöntemlerinin DAA ve KAA yöntemleri ile benzer sonuçlar vermesi, parametrik olmayan ayırma yöntemlerinin güçlerinin parametrik yöntemlerin güçlerine eriştiğini göstermektedir. Ancak heterojen veri yapılarında KAA'ya yakın DAA'dan daha yüksek performans gösterdiklerinden, güçleri DAA'dan daha yüksek bulundu. Grup sayısının düşük olduğu heterojen kovaryans matrislerine sahip verilerin analizinde, EAA yöntemlerinin etkin biçimde kullanılabilir olduğu belirlendi Friedman, 1991 yılındaki yapmış olduğu çalışmada çok değişkenli normal dağılım göstermeyen veri yapılarında MARS degree=1 ve MARS degree=2 modellerinin etkin sınıflama sonuçları verdiğini gösterdi (12). Akdağ ve arkadaşları MARS degree=1 yönteminin parametrik yöntemlere göre daha yüksek performans sergilediğini belirtti (5). Sunulan çalışma ile Friedman ve Akdağ'ın araştırma bulguları paralellik göstermektedir.

Sunulan çalışmada tıp alanı ile ilgili 6 farklı veri seti üzerinde yöntemlerin performansları karşılaştırıldığında, İNS veri setinde, EAA ve DAA performansı KAA yöntemine göre daha yüksek ($p<0.05$) iken DDİMER, GEBE, BKİ, YAĞ1 ve YAĞ2 veri setlerinde yöntemler benzer sonuçlar verdi ($p>0.05$).

Langaas (20) 1995 yılında gerçek ve türetilmiş veri setlerinde yaptığı çalışmasında geleneksel ayırma analizleri (DAA ve KAA) ile halen gelişmekte olan Yapay Sinir Ağları ve EAA analizlerinin performanslarını karşılaştırmıştır. Gerçek bir veri seti üzerinde iki farklı koşula göre uygulama yapmıştır. İlk koşulda EAA DAA'dan anlamlı olarak daha iyi bir performans göstermiş, EAA/BRUTO metodu çalışmamıştır. İkinci koşulda EAA ile DAA yöntemleri arasında anlamlı bir fark bulunmamıştır.

Anne ve Solberg (21) 2004 yılında yaptıkları çalışmada sınıflandırma yöntemlerinin performanslarını karşılaştırmışlardır. Karşılaştırma sonucunda EAA'nın DAA ve KAA'dan anlamlı olarak daha yüksek doğru sınıflandırma oranına sahip olduğunu bulmuşlardır. Ayrıca EAA'nın sınıflandırma amaçları için iyi bir alternatif olarak görülebileceğini ve normal dağılmayan veri setlerinde tutarlı, güvenilir ve geçerli sonuçlar verdiğini göstermişlerdir.

Türe ve arkadaşları (13) hipertansiyonla ilgili bir veri seti üzerinde yaptıkları çalışmalarında farklı sınıflandırma

tekniklerinin performanslarını karşılaştırmışlardır. Bu amaçla hipertansiyon ve kontrol gruplarını tahmin etmek için lojistik regresyon analizi (LR), EAA, EAA/MARS, CHAID, QUEST, CART, Radyal Tabanlı Fonksiyon ve Çok Tabakalı Algılayıcı yöntemlerini kullanmışlardır ve sınıflandırma sonucunda elde edilen duyarlılık, özgüllük ve doğruluk oranlarına dayanarak yöntemleri performanslarına göre Aşamalı Kümeleme Analizi ile kümelendirmişlerdir. Bunun sonucunda ilgili veri seti için Radyal Tabanlı Fonksiyon ve Çok Tabakalı Algılayıcı yöntemlerinin LR, FDA, FDA/MARS, CHAID, QUEST ve CART'dan daha iyi performans gösterdiğini belirlemişlerdir.

Türe ve arkadaşları (14) bir başka çalışmalarında primer hipertansiyonun tahminlenmesinde LR, EAA ve sinir ağları yöntemlerinin performanslarını karşılaştırmışlardır. Kontrol ve hipertansiyonlu hasta gruplarını ayırmada sinir ağlarının performansının lojistik regresyon ile EAA'dan istatistiksel olarak daha iyi olduğunu belirlemişlerdir.

Sunulan çalışmada, normal dağılım göstermeyen iki ve üç grulu ($p=4$) verilerin ayırımında varyansın 1 olduğu koşulda tüm yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA diğer yöntemlerden daha düşük bir performans göstermiştir. EAA, 2 alt set haricinde KAA'dan genelde daha büyük DSO değerlerine sahiptir. KAA ve DAA'yı bu tür verilerde varsayımları uymadığı için kullanmak doğru olmadığından EAA'nın üç yönteminin de her iki veri setinde ($g=2$, $g=3$) benzer performans vermiş olması yöntemlerin düşük ve yüksek ilişkili verilerde performansları değişmeden etkinliklerini korudukları görülmüştür. Bu yöntemlerin birbirlerine alternatif yöntemler olduklarını düşünmekteyiz. Ancak Friedman, Hastie ve arkadaşları (7,8,12) yaptıkları çalışmalarda EAA/MARS degree=2 modelinin EAA/MARS degree=1 ve EAA/BRUTO modellerine göre daha etkin sonuçlar verdiğini belirtmişlerdir.

Sonuç olarak, veri setlerindeki gruplar kendi içinde homojen yapıda ve aralarında heterojen yapıda ise DAA, KAA, EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemlerinin yüksek performans gösterdiği belirlenmiştir. Grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük ise, yöntemler sınıfları yüksek DSO değerleriyle birbirlerinden ayırmaktadırlar. Ancak birimlerin özellikleri birbirlerine karıştığında yine KAA, DAA'ya göre daha etkin DSO değerleri verdiği ve böylece sınıfları ayırma performansının daha iyi olduğu sonucuna varılmıştır. EAA

yöntemlerinin ise genellikle farklı koşullarda DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri verdiği gözlenmiştir.

DAA yöntemi çok değişkenli normal dağılım gösteren homojen veri setlerinde kullanıldığından, veri setlerindeki küçük değişimlerden kolay etkilenen bir yöntemdir. Sağlık alanında çok değişkenli normal dağılım gösteren homojen veri setlerinin elde edilmesi oldukça zor bir durumdur. Bu şartlar altında, araştırmacıların çok değişkenli normal dağılım gösteren heterojen veri setlerinde KAA analizini, normal dağılım göstermeyen heterojen veri setlerinde ise EAA yöntemini kullanmalarının uygun olacağı düşüncesindeyiz.

Kaynaklar

- 1.Özdamar K. *Paket Programlarla İstatistiksel Veri Analizi 2, Beşinci Baskı. Eskişehir: Kaan Kitabevi; 2005.*
- 2.Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis, Third Ed. Texas: Prentice-Hall Inc; 1992.*
- 3.Press SJ. *Applied Multivariate Analysis. New York: Holt Rinehart and Winston Inc; 1972.*
- 4.Stevens J. *Applied Multivariate Statistics For The Social Sciences. New Jersey (USA): Hillsdale; 1986.*
- 5.Akdağ B, Özkurt S, Ankaralı H, Kavas M, Evyapan F, Kiter G. Tanı koymada MARS ve lojistik regresyon modellerinin kullanımı: Karşılaştırmalı bir uygulama. IX. Ulusal Biyoistatistik Kongresi 5-9 Eylül Zonguldak Bildiri Özetleri Kitabı. 2006. p.25.
- 6.Hastie T, Buja A, Tibshirani R: *Penalized discriminant analysis, Ann Stat 1995; 23:73-102.*
- 7.Hastie T, Tibshirani R, Buja A. *Flexible discriminant and mixture models, Proceeding of "Neural Networks and Statistics" conference. Edinburg: Oxford University Press; 1995. p.1-23.*
- 8.Hastie T, Tibshirani R, Buja A. *Flexible discriminant analysis by optimal scoring, J Am Stat Assoc 1994; 89:1255-1270.*
- 9.Hastie T, Tibshirani R, Friedman J. *Elements of statistical learning: Data mining, prediction, and inference, Stanford: Springer-Verlag; 2001.*
- 10.Hastie T, Tibshirani R: *Discriminant analysis by Gaussian mixtures, J Roy Stat Soc B 1996; 58: 155-76.*
- 11.Maindonald JH, Braun WJ: *Data Analysis and Graphics Using R. Cambridge: Cambridge University Pres; 2003.*
- 12.Friedman JH. *Multivariate Adaptive Regression Splines, Ann Stat 1991; 19(1):1-141.*
- 13.Türe M, Kurt I, Kürüm AT, Ozdamar K. *Comparing classification techniques for predicting essential hypertension, Expert Syst Appl 2005; 29(3): 583-588.*
- 14.Türe M, Kurt I, Yavuz E, Kürüm AT: *Hipertansiyonun tahmini için çoklu tahmin modellerinin karşılaştırılması (Sinir ağları, lojistik regresyon ve esnek ayırma analizleri). Anadolu Kardiyol Derg 2005; 5: 24-28.*
- 15.Hardle W, Simar L. *Applied Multivariate Statistical Analysis, First Ed. Heidelberg: Springer-Verlag; 2003.*
- 16.Eubank RL. *Spline Smoothing and Nonparametric Regression. New York: Marcel Dekker Inc; 1988.*
- 17.McCullagh P, Nelder JA. *Generalized Linear Models, Second Ed. London: Chapman and Hall; 1989.*
- 18.Reynes C, Sabatier R, Molinari N. *Choice of B-splines with free parameters in the flexible discriminant analysis context. Comput Stat Data An 2005; 51: 17651778.*
- 19.Edler L, Suhai S. *Role and result of statistical methods in protein fold class prediction. Math Comput Model 2001; 33(12): 1401-1417.*
- 20.Langaas M. *Discrimination and Classification, Technical Report Statistics 1/95, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim 1995. <www.math.ntnu.no/preprint/statistics/1995/S1-1995.ps>*
- 21.Anne H, Solberg S. *Flexible nonlinear contextual classification. Pattern Recogn Lett 2004; 25(13): 1501-1508.*