



**T.C.
ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI**

**SÜREKLİ YAPIDAKİ ÇOKLU TANI TESTLERİ İÇİN
BİRLEŞTİRME YÖNTEMLERİNİN ROC EĞRİSİ ANALİZİ
KULLANARAK KARŞILAŞTIRILMASI**

DOKTORA TEZİ

Muzaffer BİLGİN

**DANIŞMAN
Doç. Dr. Ertuğrul ÇOLAK**

2018



**T.C.
ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI**

**SÜREKLİ YAPIDAKİ ÇOKLU TANI TESTLERİ İÇİN
BİRLEŞTİRME YÖNTEMLERİNİN ROC EĞRİSİ ANALİZİ
KULLANARAK KARŞILAŞTIRILMASI**

DOKTORA TEZİ

Muzaffer BİLGİN

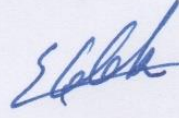
**DANIŞMAN
Doç. Dr. Ertuğrul ÇOLAK**

KABUL VE ONAY SAYFASI

Muzaffer BİLGİN'in Doktora Tezi olarak hazırladığı “**SÜREKLİ YAPIDAKİ ÇOKLU TANI TESTLERİ İÇİN BİRLEŞTİRME YÖNTEMLERİNİN ROC EĞRİSİ ANALİZİ KULLANARAK KARŞILAŞTIRILMASI**” başlıklı bu çalışma Eskişehir Osmangazi Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddesi uyarınca değerlendirerek “**KABUL**” edilmiştir.

02.03.2018

Üye: Doç. Dr. Ertuğrul ÇOLAK (Danışman)



Üye: Prof. Dr. K. Setenay ÖNER



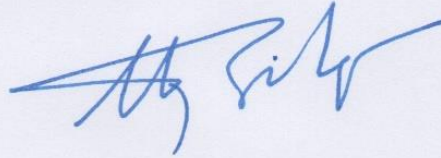
Üye: Doç. Dr. Canan BAYDEMİR



Üye: Doç. Dr. Nurhan DOĞAN



Üye: Doç. Dr. Uğur BİLGE



Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Yönetim Kurulu'nun 16/03/2018 tarih ve ...1163../5722... sayılı kararı ile onaylanmıştır.

Prof Dr. Hasan Veysi GÜNEŞ
Enstitü Müdürü

4

ÖZET

SÜREKLİ YAPIDAKİ ÇOKLU TANI TESTLERİ İÇİN BİRLEŞTİRME YÖNTEMLERİNİN ROC EĞRİSİ ANALİZİ KULLANARAK KARŞILAŞTIRILMASI

Tıp alanında, eksik klinik bilgilere ve klinik sonuçlarla ilgili belirsizliğe rağmen, hastanın tedavisi ve bakımı ile ilgili doğru ve mantıklı kararlar vermek gerekmektedir. Medikal testler tıbbi tanı açısından önemli bir role sahip olmanın yanı sıra doğru tedavinin planlanması ile tedavi maliyetlerinin azaltılması yönünden göz ardı edilemeyecek katkılar sağlamaktadırlar. Hastanın durumu hakkında güvenilir bilgi sağlamak, hasta ile sağlıklı birimleri doğru sınıflamak ve hastanın tedavisi için sağlık personelinin doğru planlama yapmasına olumlu katkılar sağlamak medikal testin temel amaçlarını oluşturmaktadır.

Sınıflandırma kuralının performansını değerlendirmek için yaygın olarak kullanılan istatistiksel tekniklerden biri, Receiver Operating Characteristic (ROC) eğrisidir. Günümüzde, tek bir tanı testi sonucu kullanılarak hastalığa ait tanının belirlenmesi yerine birden fazla test kullanılarak daha kesin bir tanı ya da sınıflandırma yapmak mümkündür. Sağlık alanında yapılan çalışmalarda farklı tanı testleri, hastalığın farklı yönlerine duyarlıdır. Dolayısıyla, bireyin sağlık durumunu değerlendirmek için her zaman tek bir tanı testine güvenilemez. Birden fazla tanı testi kullanılarak gerçeğe daha yakın ve doğru sınıflandırma yapmak mümkündür. Bu amaç doğrultusunda çeşitli modeller öne sürülmüştür. Bu modeller, en iyi doğrusal birleştirme yöntemi, doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizi yöntemleridir.

Bu tez çalışmasının amacı, tanı testlerinin birleştirilmesinde kullanılan yöntemlere alternatif olarak lojistik regresyon modeli ile doğrusal ve karesel ayırma analizlerinin tanı testlerinin birleştirilmesinde kullanılabilirliğini göstermek, lojistik, doğrusal ve karesel ayırma analizlerinin tanı testlerinin birleştirilmesindeki performanslarını, minimax prosedürü ile en iyi doğrusal birleştirme yönteminin performansını parametrik ve parametrik olmayan ROC analizleri ile karşılaştırmak, çok değişkenli normal dağılım varsayımının gerçekleştiği ve gerçekleşmediği durumlarda hangi yöntemin iyi performans gösterdiğini belirlemektir.

Anahtar Kelimeler: En iyi doğrusal birleştirme yöntemi, Doğrusal ayırma analizi, Karesel ayırma analizi, Lojistik ayırma analizi

Summary

COMPARISON OF COMBINING METHODS FOR MULTIPLE CONTINUOUS DIAGNOSTIC TESTS USING ROC CURVE ANALYSIS

In the field of medicine, despite the uncertainty about incomplete clinical information and clinical outcomes, it is necessary to make accurate and logical decisions about the treatment and care of the patient. Medical tests have an important role in terms of medical diagnosis, but they also provide indispensable benefits in terms of planning of correct treatment and reduction of treatment costs. Providing reliable information about the patient's condition, correctly classifying patients and healthy units, and providing positive contributions to healthcare staff's correct planning for the treatment of the patient constitute the main objectives of the medical test.

One of the widely used statistical techniques for evaluating the performance of the classification rule is the Receiver Operating Characteristic (ROC) curve. Nowadays, rather than using a single diagnostic test result to determine the diagnosis of the disease, it is possible to make more precise diagnosis or classification using more than one test. Different diagnostic tests in health field studies are sensitive to different aspects of the disease. Therefore, a single diagnostic test can not always be relied upon to assess an individual's health condition. It is possible to make more accurate and closer classification using more than one diagnostic test. Various models have been proposed for this purpose. These models are best linear combination method, linear discriminant analysis, quadratic discriminant analysis and logistic discriminant analysis methods.

The purpose of this thesis is to demonstrate the utility of the logistic regression model as an alternative to the methods used in the combination of diagnostic tests in combining diagnostic tests of linear and quadratic discriminant analysis. The second objective is to compare the performance of combining diagnostic tests of logistic, linear and quadratic discriminant analyzes with parametric and nonparametric ROC analyzes of the performance of the best linear combination method with the minimax procedure. It is to determine which method performs well in cases where the assumption of the multivariate normal distribution is realized and not realized.

Keywords: Best linear combination method, Linear discriminant analysis, Quadratic discriminant analysis, Logistic discriminant analysis

İçindekiler

ÖZET	iii
Summary.....	iv
Tablo Dizini	vii
Şekil Dizini.....	viii
Simge ve Kısaltmalar Dizini	ix
1- GİRİŞ VE AMAÇ	1
1.1- Temel Kavramlar ve Literatür Bilgisi	1
1.2- Problemin Belirlenmesi ve Çalışma Hipotezi	3
1.3- Tezin Amacı.....	4
2- GENEL BİLGİLER.....	6
2.1- Medikal Testler.....	6
2.1.1- Uygun medikal test için kriterler.....	7
2.2- Binormal Parametrik ROC Eğrisi Analizi	8
2.3- Parametrik Olmayan ROC Eğrisi Analizi	11
2.4- Sınıflandırma	13
2.4.1- Bayes kuralına göre sınıflandırma	14
2.5- Çoklu Tanı Testlerinin Birleştirilmesinde Kullanılan Çok Değişkenli Yöntemler	15
2.5.1- Ayırma analizi.....	15
2.5.1.1- Doğrusal ayırma analizi	16
2.5.1.1.1- Doğrusal ayırma analizi için parametrik ROC modeli... ..	17
2.5.1.1.2- Doğrusal ayırma analizi için parametrik olmayan ROC modeli.....	19
2.5.1.2- Karesel ayırma analizi	20
2.5.1.2.1- Karesel ayırma analizi için parametrik ROC modeli.....	21
2.5.1.2.2- Karesel ayırma analizi için parametrik olmayan ROC modeli.....	23
2.5.2- Lojistik ayırma analizi	25
2.5.2.1- Lojistik ayırma analizi için parametrik ROC modeli.....	27
2.5.2.2- Lojistik ayırma analizi için parametrik olmayan ROC modeli	29
2.5.3- MiniMax prosedürüne göre doğrusal en iyi birleştirme yöntemi	30

3- GEREÇ VE YÖNTEM	33
3.1- Simülasyon Çalışmasında Kullanılan Senaryolar	33
3.2- Simülasyon Algoritması	36
3.3- Uygulama Veri Seti	37
3.4- R Programlama Kodları.....	41
4- BULGULAR	47
4.1- Simülasyon Çalışması Bulguları	47
4.2- Uygulama Veri Seti Bulguları.....	54
5- TARTIŞMA.....	62
6- SONUÇ VE ÖNERİLER	65
KAYNAKLAR DİZİNİ.....	66
EKLER DİZİNİ.....	72
EK – 1	72
ÖZGEÇMİŞ.....	75

Tablo Dizini

Tablo 2.1- Uygun Bir Tanı Testi Geliştirilmesinde Dikkat Edilmesi Gereken Kriterler.....	7
Tablo 3.1- Uygulama Veri Seti.....	38
Tablo 3.2- Shapiro Wilk Normallik Testi.....	39
Tablo 3.3- Doornik-Hansen İkili Normallik Testi.....	40
Tablo 4.1- Senaryo I – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Aynı, Kovaryans Matrislerinin Aynı Olduğu Durum.....	49
Tablo 4.2- Senaryo II – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Aynı, Kovaryans Matrislerinin Farklı Olduğu Durum.....	50
Tablo 4.3- Senaryo III – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Farklı, Kovaryans Matrislerinin Aynı Olduğu Durum.....	51
Tablo 4.4- Senaryo IV – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Farklı, Kovaryans Matrislerinin Farklı Olduğu Durum.....	52
Tablo 4.5- Senaryo V – Hasta Verilerinin Türetiminde Bozulmuş Çok Değişkenli Normal Dağılım Verilerinin Kullanıldığı Durum.....	53
Tablo 4.6- Uygulama Veri Seti Sonuçları.....	57
Tablo 4.7- Uygulama Veri Seti b_0 ve b katsayıları Sonuçları.....	57

Şekil Dizini

Şekil 2.1- Binormal parametrik ROC Eğrisi.....	11
Şekil 2.2- Parametrik Olmayan ROC Eğrisi.....	13
Şekil 2.3- Sınıflandırma Grafiği.....	14
Şekil 2.4- Bayes Kuralına Göre Sınıflandırma Grafiği.....	15
Şekil 2.5- Sigmoid Fonksiyonu.....	26
Şekil 3.1- Uygulama Verisindeki Medikal Tanı Testleri Histogram Grafikleri.....	40
Şekil 4.1- Parametrik ROC Eğrisi.....	54
Şekil 4.2- Parametrik Olmayan ROC Eğrisi.....	55
Şekil 4.3- Skor Değerlerine Göre Birleştirme Yöntemlerine Ait Parametrik Olmayan ve Parametrik ROC Eğrileri.....	56
Şekil 4.4- Simülasyon Senaryosu I'ın AUC Değerleri.....	58
Şekil 4.5- Simülasyon Senaryosu II'nin AUC Değerleri.....	59
Şekil 4.6- Simülasyon Senaryosu III'in AUC Değerleri.....	60
Şekil 4.7- Simülasyon Senaryosu IV'in AUC Değerleri.....	61
Şekil 4.8- Simülasyon Senaryosu V'in AUC Değerleri.....	61

Simge ve Kısaltmalar Dizini

Simgeler

S_n	Duyarlılık
S_p	Özgüllük
π_h	Hastalık tanısı taşıyan toplum
π_s	Sağlıklı toplum
μ_h	Hasta toplumun tanı testi ortalaması
μ_s	Sağlıklı toplumun tanı testi ortalaması
σ_h	Hasta toplumun tanı testi standart sapması
σ_s	Sağlıklı toplumun tanı testi standart sapması
Φ	Kümülatif standart normal dağılım fonksiyonu
n_h	Hasta sınıftaki birim sayısı
n_s	Sağlıklı sınıftaki birim sayısı
$s_h(c)$	Hasta sınıfında tanı testi değeri c 'den büyük olanların sayısı
$s_s(c)$	Sağlıklı sınıfında tanı testi değeri c 'den büyük olanların sayısı
$n_s^{=x}$	Tanı testleri sonuçları x 'e eşit olan gerçek negatif birimlerin sayısı
$n_h^{=x}$	Tanı testi sonuçları x 'e eşit olan gerçek pozitif birimlerin sayısı
$n_s^{<x}$	Tanı testi sonuçları x 'den küçük olan gerçek negatif birimlerin sayısını
$n_h^{>x}$	Tanı testi sonuçları x 'den büyük olan gerçek pozitif birimlerin sayısını
\mathbf{b}	$(px1)$ boyutlu doğrusal modeldeki katsayılar vektörü
\mathbf{b}'	$(1xp)$ boyutlu doğrusal modeldeki katsayılar vektörü
t	$(0, 1)$ aralığında değerler alan sabit
L_M	\mathbf{b} vektöründeki katsayılar kullanılarak tanı testlerinden elde edilen doğrusal skor
$p(\Pi_i \mathbf{x})$	Bayes Kuralın göre sonsal olasılıklar
$f_i(\cdot)$	Hasta ve sağlıklı sınıfı temsil eden alt toplumların olasılık yoğunluk fonksiyonu
$L_{i D,h}$	Hasta grubundaki birimlerin doğrusal ayırma analizinden elde edilen skor değerleri
$L_{j D,s}$	Sağlıklı gruptaki birimlerin doğrusal ayırma analizinden elde edilen skor değerleri
$\bar{L}_{D,h}$	Hasta grubundaki birimlerin doğrusal ayırma analizinden elde edilen skor değerlerine ait ortalama
$\bar{L}_{D,s}$	Sağlıklı gruptaki birimlerin doğrusal ayırma analizinden elde edilen skor değerlerine ait ortalama
$LS_{D,h}$	Hasta grubundaki birimlerin doğrusal ayırma analizinden elde edilen skor değerlerine ait standart sapma

$L_{SD,s}$	Sağlıklı gruptaki birimlerin doğrusal ayırma analizinden elde edilen skor değerlerine ait standart sapma
$L_{iK,h}$	Hasta grubundaki birimlerin karesel ayırma analizinden elde edilen skor değerleri
$L_{jK,s}$	Sağlıklı gruptaki birimlerin karesel ayırma analizinden elde edilen skor değerleri
$\bar{L}_{K,h}$	Hasta grubundaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait ortalama
$\bar{L}_{K,s}$	Sağlıklı gruptaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait ortalama
$L_{SK,h}$	Hasta grubundaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait standart sapma
$L_{SK,s}$	Sağlıklı gruptaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait standart sapma
$L_{iL,h}$	Hasta grubundaki birimlerin lojistik ayırma analizinden elde edilen skor değerleri
$L_{jL,s}$	Sağlıklı gruptaki birimlerin lojistik ayırma analizinden elde edilen skor değerleri
$\bar{L}_{L,h}$	Hasta grubundaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait ortalama
$\bar{L}_{L,s}$	Sağlıklı gruptaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait ortalama
$L_{SL,h}$	Hasta grubundaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait standart sapma
$L_{SL,s}$	Sağlıklı gruptaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait standart sapma
ε_i	Birleştirme olasılıkları
κ	Kappa katsayısı

Kısaltmalar

ROC	Receiver Operating Characteristic
AUC	Eğri Altında Kalan Alan
c	Kesim Noktası
GP	Gerçek Pozitif Oranı
YP	Yalancı Pozitif Oranı
J	Youden İndeksi

1- GİRİŞ VE AMAÇ

1.1- Temel Kavramlar ve Literatür Bilgisi

Tıp alanında, eksik klinik bilgilere ve klinik sonuçlarla ilgili belirsizliğe rağmen, hastanın tedavisi ve bakımı ile ilgili doğru ve mantıklı kararlar vermek gerekmektedir. Hastanın öyküsünden ve fizik muayenesinden elde edilen verilerin rehberliğinde bir tanı koymak çoğunlukla yeterli olmamaktadır. Bu gibi durumlarda, sağlık personeli daha fazla veri elde etmek için medikal testlere başvurur. Medikal testler birçok özelliğe göre sınıflandırılırsalar da genel olarak tarama (screening) ve tanı (diagnostic) testi olarak karşımıza çıkmaktadır. Belli bir hastalık için risk faktörlerini tanımlamak ve asemptomatik kişilerde gizli hastalığı tespit etmek için kullanılan medikal testlere tarama testleri denilmektedir. Aynı zamanda medikal testler tanı amaçlı olarak semptomatik kişilerde hastalığın belirlenmesine yardımcı olmak için kullanılmaktadır. Bu tip testlere ise tanı testleri denilmektedir (Nicoll ve ark., 2012).

Medikal testler tıbbi tanı açısından önemli bir role sahip olmanın yanı sıra doğru tedavinin planlanması ile tedavi maliyetlerinin azaltılması yönünden de göz ardı edilemeyecek katkılar sağlamaktadırlar (Epstein ve ark., 1986). Ancak, tanı testlerinin geliştirilmesinde, uygulanmasında ve sonuçlarının değerlendirilmesinde yapılan hatalar dikkat çekmektedir. Reid, Lachs ve Feinstein (1995), 1978-1993 yılları arasında yayınlanan tanı testleri hakkındaki bilimsel makaleleri incelediklerinde, geliştirilen tanı testlerinin tasarım ve analizinde önemli hataların olduğunu saptamışlardır. Bu hatalar, medikal test çalışmalarının sonuçlarına karşı güvensizliğe, testlerinin seçimi ve yorumlanmasında yanlış anlamalara neden olmuştur. En çok yapılan hatalardan biri, sadece sağlıklı gönüllülerden elde edilen test sonuçlarına göre değerlendirme ya da sınıflandırma kriterleri geliştirmek olmuştur. Örneğin pankreatiti saptamak için geliştirilen bir testte, sağlıklı gönüllülerden elde edilen bir enzim düzeyinin ortalama değerinin üç standart sapma altı kesme noktası olarak belirlenmesini düşünelim. Belirlenen bu kesme noktasının üstünde bir enzim seviyesine sahip birimler sağlam, altında olan birimler ise hasta olarak sınıflandırılırlar. Ancak bu şekilde bir sınıflandırma kriteri belirlemek bazı önemli noktaların gözden kaçmasına neden olur. Bunlar; (1) enzim değerlerinin dağılım şekli, (2) hasta ve sağlıklı birimlerden elde edilen enzim değerlerinin kesişim miktarı (overlapping), (3) yanlış sınıflandırma hatalarının oranı ve bunların klinik önemi, (4) sağlıklı gönüllülerden elde edilen sonuçların hastaları dikkate almadan zayıf bir biçimde genelleştirilmesi (Zhang ve ark., 2002).

Bir medikal testin iki amacı vardır; (1) hastanın durumu hakkında güvenilir bilgi sağlamak, hasta ile sağlıklı birimleri doğru sınıflamak, (2) hastanın tedavisi için sağlık personelinin doğru planlama yapmasına olumlu katkılar sağlamak (Sox ve ark., 1989). Belirlenen bu amaçların elde edilmesi, testin doğru istatistiksel yöntemler kullanılarak değerlendirilmesine ve değerlendirme sonucunda da kullanılabilir olduğuna dair bilimsel kanıtların elde edilmesine bağlıdır. Medikal testlerin değerlendirme süreci ve analizleri sonucunda genel olarak testin tanı doğruluğu ve birimleri doğru ayırabilme özelliği ile ilgili bilgiler elde edilir (Zweig ve Campbell, 1993).

Bir sınıflandırma kuralının performansını değerlendirmek için yaygın olarak kullanılan istatistiksel tekniklerden biri, Receiver Operating Characteristic (ROC) eğrisidir. Basit bir tanı testi probleminde, yeni gözlenen bir birimin hasta ya da sağlıklı sınıfa atanacağı bu tanı testinin kesme noktası değerine göre belirlenir. Sınıflamanın doğruluğu seçilen kesim noktasına bağlıdır. Bir tanı testinin iyi veya doğru sınıflandırmasının değerlendirilmesi altın standarda göre belirlenir ancak altın standardın olmadığı durumlar da söz konusudur (Friedman, Hastie ve ark., 2001).

Kesme noktası belirlendiğinde, gerçek ve yanlış pozitif oranları ile duyarlılık (S_n) ve özgüllük (S_p) değerleri hesaplanır. Her bir kesme değeri için S_n ve S_p değeri elde edilerek, ROC eğrisi yatay ekseninde $(1-S_p)$, dikey ekseninde de S_n olacak şekilde oluşturulur. Belli bir tanı testinin performansını değerlendiren en önemli kriter, 0 ile 1 arasında değerler alan, eğri altında kalan alandır (Area Under Curve, AUC) (Çolak ve ark., 2012). Örneğin hasta toplumda tanı testi değerleri sağlam topluma göre daha yüksek ise; AUC, hasta toplumundan rasgele seçilen bir bireyin tanı testi değerinin sağlıklı toplumdan rasgele seçilen bir bireyin tanı testi değerinden yüksek olma olasılığına eşittir. Bu durumda, AUC değerinin yüksek olması, tanı testinin sınıflandırma performansının daha iyi olduğunu göstermektedir (Hanley ve McNeil, 1982; Metz, 1978; Fawcett, 2006).

Günümüzde, tek bir tanı testi sonucu kullanılarak hastalığa ait tanının belirlenmesi yerine birden fazla test kullanılarak daha kesin bir tanı ya da sınıflandırma yapmak mümkündür. Tanı testleri genelde mükemmel değildir. Ancak aynı hastalık ya da tanı için geliştirilmiş birden fazla tanı testinin birleştirilmiş sonuçlarını kullanmak daha doğru sonuca ulaşmada oldukça önemlidir (Li ve Fine, 2008).

Sağlık alanında yapılan çalışmalarda farklı tanı testleri, hastalığın farklı yönlerine duyarlıdır. Dolayısıyla, bireyin sağlık durumunu değerlendirmek için her zaman tek bir tanı testine güvenilemez. Birden fazla tanı testi kullanılarak gerçeğe daha yakın ve doğru sınıflandırma yapmak mümkündür. Bu amaç doğrultusunda çeşitli modeller öne sürülmüştür. Su ve Liu (1993), en iyi doğrusal birleştirme yöntemini önermişlerdir. Hasta ve sağlıklı toplumlardan elde edilen tanı testlerinin çok değişkenli normal dağılım gösterdiği varsayıldığı bu çalışmada, kovaryans matrislerini iki farklı yönden ele almışlardır. İlk yaklaşımda, kovaryans matrislerinin orantısız olduğunu, diğerinde ise kovaryans matrisleri üzerinde herhangi bir kısıt olmadığını incelemişlerdir. Reiser ve Faraggi (1997), çoklu tanı testinin birleştirilmesinde kullanılan genelleştirilmiş ROC modelinde, AUC'nin güven aralıkları için yöntemler önermişlerdir. Pepe ve Thomson (2000), tanı testlerinin doğrusal birleştirilmesinin belirlenmesinde herhangi bir dağılım varsayımı içermeyen, sıralama puanlarına dayalı bir yaklaşım önermişlerdir. Liu ve ark. (2011), Gao ve ark. (2008) belirlenen sabit bir S_p değerinde, S_n değerini maksimize eden ve en iyi doğrusal birleştirmeyi tahmin eden bir yöntem önermişlerdir. Yu W. ve Park T. (2015), çok değişkenli normal dağılım varsayımı altında, tanı testlerinin doğrusal birleştirilmesi için iki farklı algoritma geliştirmişler. Sameera ve ark. (2016), minimax prosedürünü kullanarak en iyi doğrusal birleştirme yöntemi geliştirmişler ve önceki yaklaşımlarda önerilen yöntemlere göre daha etkili olduğunu ileri sürmüşlerdir. Ayrıca bu yöntemde optimum kesme noktasının elde edilmesi için Youden indeksinden yararlanılmıştır. Birden çok tanı testi kullanılarak elde edilen bu yöntem çok değişkenli ROC (Multivariate ROC, MROC) adını almaktadır (Kang, Liu ve ark, 2016).

1.2- Problemin Belirlenmesi ve Çalışma Hipotezi

Lojistik regresyon ve doğrusal ayırma analizi kategorik bağımlı değişken ile bağımsız değişkenleri içeren verilerin analizi için yaygın olarak kullanılan çok değişkenli istatistiksel yöntemlerdir. Her iki yöntem de doğrusal sınıflandırma modellerinin, yani gruplar arasındaki doğrusal sınırlarla ilişkili modellerin geliştirilmesi için uygundur. İki yöntem arasındaki temel fark içerdikleri varsayımlardır. Lojistik regresyon analizi değişkenlerin dağılımı hakkında bir varsayım içermemekle birlikte doğrusal ayırma analizi çok değişkenli normal dağılım varsayımını içermektedir. Ayrıca doğrusal ayırma analizi, kategorik bağımlı değişkenin belirtmiş olduğu toplumların varyan-kovaryanslarının homojen olduğu varsayımını da içerir. Homojenite varsayımının sağlanmadığı durumlarda da karesel ayırma analizi kullanılmaktadır (Hosmer ve Lemeshow, 2005; Rencher, 2002).

Bu bağlamda, tanı testlerinin birleştirilmesinde kullanılan ve bölüm 1.1’de belirtilen yöntemlere alternatif olarak lojistik regresyon modeli ile doğrusal ve karesel ayırma analizlerinin tanı testlerinin birleştirilmesinde kullanılabilirliğini araştırmak bu tez çalışmasının problemi olarak tanımlanmıştır. Böylelikle çalışmanın ana hipotezi aşağıdaki şekilde belirlenmiştir.

"Lojistik, doğrusal ve karesel ayırma analizlerinin tanı testlerinin birleştirilmesindeki performansları, Sameera ve ark. (2016) tarafından önerilen minimax prosedürü ile en iyi doğrusal birleştirme yönteminin performansı kadar etkilidir."

1.3- Tezin Amacı

Bu tez çalışmasının amacı;

1. Tanı testlerinin birleştirilmesinde kullanılan yöntemlere alternatif olarak lojistik regresyon modeli ile doğrusal ve karesel ayırma analizlerinin tanı testlerinin birleştirilmesinde kullanılabilirliğini göstermek,
2. Lojistik, doğrusal ve karesel ayırma analizlerinin tanı testlerinin birleştirilmesindeki performanslarını, Sameera ve ark. (2016) tarafından önerilen minimax prosedürü ile en iyi doğrusal birleştirme yönteminin performansını parametrik ve parametrik olmayan ROC analizleri ile karşılaştırmak,
3. Çok değişkenli normal dağılım varsayımının gerçekleştiği ve gerçekleşmediği durumlarda hangi yöntemin iyi performans gösterdiğini belirlemek,
4. Elde edilen sonuçlara göre, tanı testlerinin birleştirilmesinde kullanılan bu yöntemleri içeren ve veri yapısına göre uygun yöntemin kullanılacağı bir yazılımı R programında oluşturarak

literatüre katkı sağlamaktır.

Belirlenen hipotezin test edilmesinde ve yöntemlerin karşılaştırılmasında çok deęişkenli normal dağılım varsayımının gerçekleştięi ve gerçekleşmedięi durumlara göre deęişik örnek büyükleri kullanarak belirlenen senaryolar aracılığı ile Monte-Carlo simülasyon çalışması yapıldı. Ayrıca over kanseri tanısında kullanılan testleri içeren gerçek bir veri seti kullanılarak da simülasyon sonuçlarının tutarlılığı denetlendi.



2- GENEL BİLGİLER

2.1- Medikal Testler

Tıbbi bir problemin ya da durumun doğru teşhisi bu problemin tanımlanması ve kontrol altına alınması için ilk adımdır. Uygun tedavi genel olarak durumun ayrıntılı bir şekilde tanımlanması ya da bilinmesine bağlıdır. Yaygın kullanılan tıbbi tanı testleri bakteri kültürünü, radyolojik görüntülemeyi ve biyokimyasal testleri içerir. Streptokoksik enfeksiyon için boğaz sürüntüsü ile elde edilen bakteri kültürü, idrar yolu enfeksiyonu için tam idrar tetkiki, omurilik sıvısı kullanılarak menenjit tanısı, kemik kırıkları için X-ray görüntüleme, beyin travması için bilgisayarlı tomografi, lumbal disk hernisini (fıtıklaşmış disk, bel fıtığı) değerlendirilmesinde manyetik rezonans görüntüleme (MR), karaciğer fonksiyon bozukluklarının belirlenmesinde serum bilirubin düzeyi, böbrek fonksiyon bozuklukları belirlenmesinde serum kreatinin düzeyi, hamilelik testinde Human Chorionic Gonadotropin'inin (hCG) kullanılması medikal testlere örnek olarak verilebilir (Pepe, 2003).

Tıbbi bir problem için belirlenen bir test uygulama ve sonucun elde edilmesi bakımından basit olabilir, örneğin bir enfeksiyon için bakteri kültürü testi. Bazı tıbbi durumlar için ise karmaşık bir süreç içerebilir, örneğin bir dokuya ait gen ekspresyonu profillemesinin mikroarray teknolojisi kullanılarak değerlendirilmesi (Pepe, 2003).

Sağlıklı toplumların gizli (okult) hastalıklar bakımından taranması önemlidir. Çünkü erken teşhis tedavinin başarılı olmasında önemli bir etkidir. Mamografi ile meme kanseri taraması, Pap smear ile serviks kanseri taraması, serum kolesterolü ve kan basıncı izlemi ile kardiyovasküler hastalıkların taraması, yeni doğanların fenilketonüri (PKU), kistik fibroz ve fiziksel bozukluklar için yapılan taramalar, bulaşıcı hastalıkların örneğin AIDS vb. taranması, bu tip hastalıkların yayılmasını kontrol eden ve bu konuda araştırmacılara bilgi sağlayan programlar ya da prosedürler sağlıklı toplum taramalarında kullanılan tarama testlerine örnek olarak verilebilir (Zhou; McClish ve ark; 2009).

Tarama ve tanı testlerinin değerlendirilmesinde kullanılan istatistiksel yöntemler bakımından bir farklılık bulunmamaktadır. Dolayısıyla istatistiksel bakış açısıyla tarama testi ile tanı testleri genel anlamda medikal test olarak kabul edilir. Bu tez çalışmasında tarama ve tanı testleri ayrımı yapılmaksızın tanı testi ifadesi her iki test içinde kullanılmıştır (Pepe, 2003).

2.1.1- Uygun medikal test için kriterler

Bir testin geliştirilmesi ancak hastalığın teşhis edilmesinden sonra elde edilecek bir fayda varsa yararlıdır. Tanı testlerinin geliştirilmesinde dikkat edilmesi gereken kriterler tablo 2.1’de verilmiştir. Bu kriterler Wilson ve Jungner (1968), Cole ve Morrison (1980) ve Obuchowski ve ark. (2004)’nın ortaya atmış olduğu kriterlerin özet bir şeklidir.

Tablo 2.1- Uygun Bir Tanı Testi Geliştirilmesinde Dikkat Edilmesi Gereken Kriterler

1. Hastalık ciddi bir hastalık olmalıdır.
2. Hastalık risk altındaki toplumda nispeten yaygın olmalıdır.
3. Hastalık için tedavi yöntemleri olmalıdır.
4. Tedavi yöntemleri test sonucu pozitif olanlar için uygulanabilir ve ulaşılabilir olmalıdır.
5. Test birimlere zarar vermemelidir.
6. Test hasta ve sağlıklı birimleri yüksek oranda doğru sınıflayabilmelidir.

Tablo 2.1’de verilen kriterler incelendiğinde, altıncı maddede belirtilen kriter doğrultusunda, tanı testinin doğruluğunun ve sınıflama oranının değerlendirilmesinde istatistiksel yöntemlerden yararlanılır. Bu değerlendirme çeşitli yöntemler kullanılarak yapılmaktadır ancak temelde doğru ve güvenilir bir test, birimleri hastalık durumlarına göre (hasta ve sağlıklı) doğru bir şekilde sınıflayan bir testtir. Hatalı testler hasta bir birimin sağlıklı, sağlıklı bir birimi de hasta olarak sınıflayabilir. Hasta bir bireyin sağlıklı sınıflandırılması yalancı negatiflik, sağlıklı bir bireyinde hasta olarak sınıflandırılması yalancı pozitiflik olarak tanımlanır ve bu iki oran testin yanlış sınıflandırma oranını belirler. Yanlış pozitif sonuçlar, birimlerin gereksiz prosedürlere (genellikle invaziv (girişimsel) ve pahalı tedaviye) maruz kalmasına neden olur. Benzer şekilde tedavi alması gereken bir birimin sağlıklı olarak nitelendirildiği için tedaviden yoksun kalabilir. İyi bir tanı testi için bu tip hataların minimum düzeyde tutulması gerekmektedir. Bundan dolayı bir tanı testinin uygulanmasından önce bu tanı testine ait tanı ya da sınıflama doğruluğu titizlikle değerlendirilmelidir (McNeil ve Hanley, 1983).

Bazı medikal testlerin sonuçları doğrudan ikili olarak gözlemlenir. Örneğin serumdaki spesifik bir metile edilmiş DNA sekansının varlığı ya da yokluğu ikili bir sonuçtur, varsa sonuç pozitif, yoksa sonuç negatiftir. Bazı durumlarda da medikal test sonuçları sıralı ölçekli bir tipte sonuçlanır. Mamogramların değerlendirilmesinde kullanılan BIRADS (Breast Imaging

Reporting And Data System) olarak ifade edilen raporlama yöntemi sıralı ölçekli medikal test örnek olarak verilebilir. Bazı medikal testlerin sonuçları ise sürekli bir değerdir. Doğrudan hasta ya da sağlıklı yerine ölçümsel bir değer verir. Örneğin serumdaki PSA değeri, kreatinin değeri ya da bilirubin değeri hastalığın varlığını ya da yokluğunu doğrudan vermez, elde edilen değerlerin yüksekliğine bağlı olarak yorum yapılır (Pepe, 2003).

Tıp alanında birçok tanı testi sürekli yapıda karşımıza çıkmaktadır. Örneğin kanser için kullanılan PSA, CA-125 vb. testler sürekli değerler içeren sonuçlar verir. Diğer standart biyokimyasal ölçümler örneğin karaciğer hastalıkları için bilirubin, böbrek fonksiyon bozuklukları için kreatinin, kardiyovasküler hastalıklar için serum kolesterol düzeyi, diyastolik/sistolik kan basınçları, diyabet için serum glikoz düzeyi sürekli ölçümler içeren medikal testlerdir.

Sürekli yapıdaki tanı testlerinin performanslarının değerlendirilmesinde en iyi geliştirilmiş istatistiksel yöntem ROC eğrisi yöntemidir ve istatistiksel karar teorisine dayanır. ROC eğrileri 1950 ve 1960'larda geliştirilen sinyal algılama teorisi bağlamında ortaya çıkmıştır (Green ve Swets, 1966; Egan, 1975). 1960'lı yılların başlarında özellikle ROC eğrisi yöntemi ilk olarak tıpta radyolojik görüntüleme karar verme sürecinde kullanılmıştır (Lusted, 1960). Günümüzde ROC eğrisi yöntemini laboratuvar testlerinde, epidemiyolojide, biyoinformatikte, radyolojik çalışmalarda yaygın bir şekilde kullanılmaktadır (Hanley, 1989; Shapiro, 1999).

2.2- Binormal Parametrik ROC Eğrisi Analizi

Bu çalışmada sürekli yapıdaki çoklu tanı testleri için birleştirme yöntemlerinin ROC eğrisi analizi kullanarak karşılaştırılması amaçlandığından dolayı öncelikli olarak binormal tek değişkenli parametrik ROC eğrisi yöntemi açıklanmıştır (Jonson ve Wichern, 1992). Hasta (Π_h) ve sağlıklı (Π_s) toplum için belirlenen tek değişkenli tanı testi (X) binormal dağılım gösterebilir, $X_h \sim \mathcal{N}(\mu_h, \sigma_h^2)$, $X_s \sim \mathcal{N}(\mu_s, \sigma_s^2)$.

Toplumda X 'in büyük değerleri hastalığı gösterebilir. Herhangi bir kesim noktası c için gerçek pozitif oranının (GP),

$$GP(c) = P(X > c | X \in \Pi_h) = \Phi\left(\frac{\mu_h - c}{\sigma_h}\right) \quad (2.01)$$

Yalancı pozitif oranı (YP),

$$YP(c) = P(X > c | X \in \Pi_s) = \Phi\left(\frac{\mu_s - c}{\sigma_s}\right) \quad (2.02)$$

eşitlikleri ile gösterilir.

Burada;

Φ : Kümülatif standart normal dağılım fonksiyonu

Yalancı pozitif oranı $YP(c) = y$ ile gösterilirse, eşitlik 2.01 ve eşitlik 2.02 kullanılarak ROC eğrisi fonksiyonu eşitlik 2.03'de verilmiştir.

$$ROC(y) = \Phi(a + b\Phi^{-1}(y)) \quad (2.03)$$

Burada;

$$a = \frac{\mu_h - \mu_s}{\sigma_h}, b = \frac{\sigma_s}{\sigma_h} \quad (2.04)$$

Yalancı pozitiflik oranı (y) için kesme noktası,

$$c = \mu_s - \sigma_s\Phi^{-1}(y) \quad (2.05)$$

şeklinde tanımlanır.

Eğri altında kalan alan (AUC) eşitlik 2.06'da verilmiştir.

$$AUC = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right) \quad (2.06)$$

a, b ve AUC 'nin en büyük benzerlik tahminicileri $\bar{x}_h, \bar{x}_s, s_h$ ve s_s kullanılarak elde edilir.

$$\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{ih}}{n_h}, s_h = \sqrt{\frac{\sum_{i=1}^{n_h} (x_{ih} - \bar{x}_h)^2}{n_h - 1}} \quad (2.07)$$

$$\bar{x}_s = \frac{\sum_{j=1}^{n_s} x_{js}}{n_s}, s_s = \sqrt{\frac{\sum_{j=1}^{n_s} (x_{js} - \bar{x}_s)^2}{n_s - 1}} \quad (2.08)$$

$$\hat{a} = \frac{\bar{x}_h - \bar{x}_s}{s_h}, \quad V(\hat{a}) = \frac{n_h(\hat{a}^2 + 2) + 2n_s\hat{a}^2}{2n_h n_s} \quad (2.09)$$

$$\hat{b} = \frac{s_s}{s_h}, \quad V(\hat{b}) = \frac{(n_h + n_s)\hat{b}^2}{2n_h n_s} \quad (2.10)$$

$$\widehat{AUC} = \Phi\left(\frac{\hat{a}}{\sqrt{1 + \hat{b}^2}}\right) \quad (2.11)$$

Burada,

n_h : Hasta sınıfındaki birim sayısı

n_s : Sağlıklı sınıfındaki birim sayısı

göstermektedir.

$V(\widehat{AUC})$, delta yöntemi kullanılarak hesaplanmaktadır. Ancak günümüzde bootstrap yöntemi kullanılarak varyans hesaplamaları yapılmaktadır (Sameera, Vardhan ve ark., 2016). Bu tez çalışmasında parametrik ROC için hesaplanan AUC'nin varyansı bootstrap yöntemi kullanılarak hesaplanmıştır.

Optimum kesim noktasını elde etmek için kullanılan yaygın indekslerden biriside Youden indeks'dir (Perkins ve Schisterman,2005).

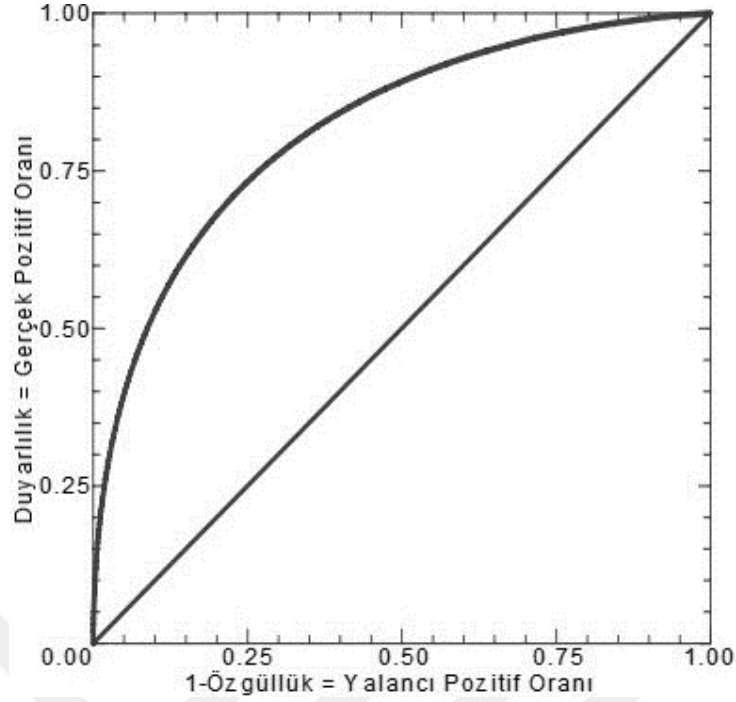
Youden indeks (J) eşitlik 2.12'deki gibi hesaplanır.

$$J = [GP(c) - YP(c)] \quad (2.12)$$

Bu indeks değerleri kullanılarak optimum kesim noktası (c) için tahmin değeri eşitlik 2.13'deki gibi hesaplanır.

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.13)$$

Optimum c noktası aynı zamanda ROC eğrisinin diyagonale olan dik uzaklığının maksimum olduğu noktadır.



Şekil 2.1- Binormal parametrik ROC Eğrisi

2.3- Parametrik Olmayan ROC Eğrisi Analizi

Parametrik olmayan ROC eğrisi analizinde ampirik (empirical) yöntemler kullanılır. Parametrik olmayan ROC eğrisi analizi, tanı testleri hakkında herhangi bir dağılım varsayımı içermez. Bu yaklaşımda gözlenen tüm test sonuçları küçükten büyüğe doğru sıralanır ve her bir değer kesme noktası (c) olarak belirlenerek bu kesme noktasına ait gerçek pozitif ve yalancı pozitif oranları aşağıdaki gibi hesaplanır (Robin ve ark., 2011).

$$GP(c) = P(X > c | X \in \Pi_h) = \frac{s_h(c)}{n_h} \quad (2.14)$$

$$YP(c) = P(X > c | X \in \Pi_s) = \frac{s_s(c)}{n_s} \quad (2.15)$$

Burada,

- $s_h(c)$: Hasta sınıfında tanı testi değeri c 'den büyük olanların sayısı
- n_h : Hasta sınıftaki birim sayısı
- $s_s(c)$: Sağlıklı sınıfında tanı testi değeri c 'den büyük olanların sayısı
- n_s : Sağlıklı sınıftaki birim sayısı

göstermektedir.

Parametrik olmayan ROC eğrisi bu noktaların düz bir çizgi ile birleştirilmesi sonucunda elde edilir. Eğri altında kalan alan trapezoidal kural kullanılarak elde edilir ve tahmin değeri eşitlik 2.16'da verilmiştir.

$$\widehat{AUC} = \frac{1}{n_h n_s} \sum_{i=1}^{n_h} \sum_{j=1}^{n_s} \psi(x_{ih}, x_{js}) \quad (2.16)$$

Burada;

$$\psi(x_{ih}, x_{js}) = \begin{cases} 1 & \text{Eğer } x_{ih} > x_{js} \\ \frac{1}{2} & \text{Eğer } x_{ih} = x_{js} \\ 0 & \text{Eğer } x_{ih} < x_{js} \end{cases} \quad (2.17)$$

x_{ih} ve x_{js} sırası ile hasta ve sağlıklı birimlerin test sonuçlarını göstermektedir.

Bu indeks değerleri kullanılarak optimum kesim noktasının tahmin değeri (c) eşitlik 2.18'deki gibi elde edilir.

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.18)$$

Tahmin edilen AUC'nin varyansı eşitlik 2.19'deki gibi hesaplanır.

$$V(\widehat{AUC}) = \frac{\widehat{AUC}(1 - \widehat{AUC}) + (n_h - 1)(Q_h - \widehat{AUC}^2)}{n_h n_s} + \frac{(n_s - 1)(Q_s - \widehat{AUC}^2)}{n_h n_s} \quad (2.19)$$

Burada;

$$Q_h = \frac{1}{n_s n_h^2} \sum_x n_s^{\bar{x}} \times \left[(n_h^{>x})^2 + n_h^{>x} \times n_h^{\bar{x}} + \frac{(n_h^{\bar{x}})^2}{3} \right] \quad (2.20)$$

$$Q_s = \frac{1}{n_s^2 n_h} \sum_x n_h^{\bar{x}} \times \left[(n_s^{<x})^2 + n_s^{<x} \times n_s^{\bar{x}} + \frac{(n_s^{\bar{x}})^2}{3} \right] \quad (2.21)$$

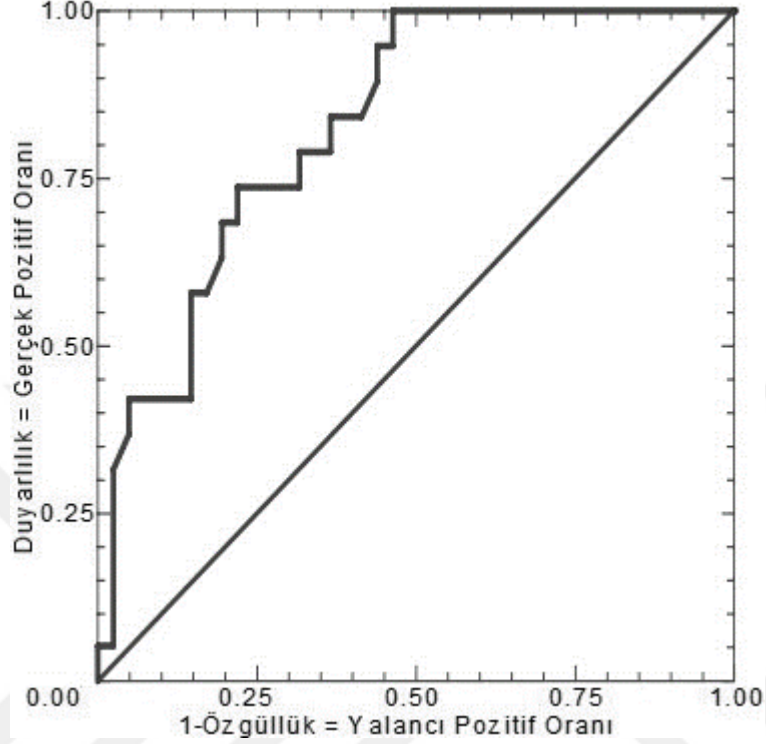
Burada,

$n_s^{\bar{x}}$: Tanı testleri sonuçları x 'e eşit olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{\bar{x}}$: Tanı testi sonuçları x 'e eşit olan gerçek pozitif birimlerin sayısını göstermektedir.

$n_s^{<x}$: Tanı testi sonuçları x 'den küçük olan gerçek negatif birimlerin sayısını göstermektedir.

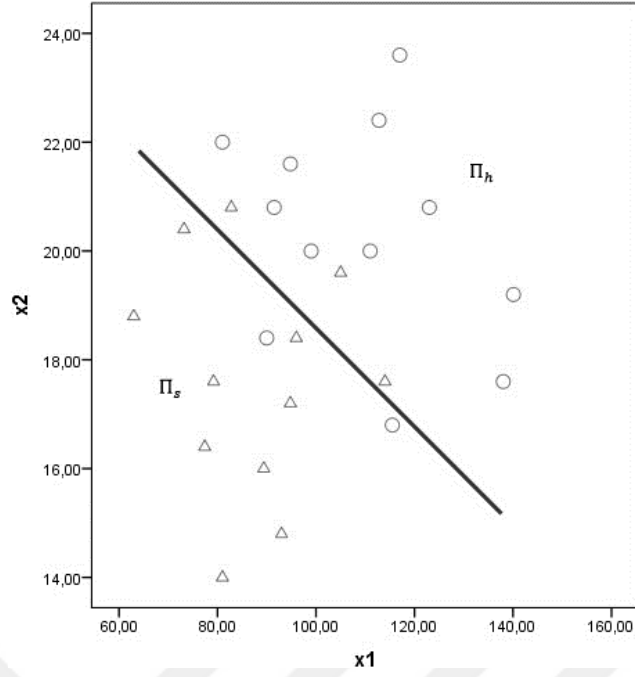
$n_h^{>x}$: Tanı testi sonuçları x 'den büyük olan gerçek pozitif birimlerin sayısını göstermektedir.



Şekil 2.2- Parametrik Olmayan ROC Eğrisi

2.4- Sınıflandırma

Hedef toplum \mathcal{T} ile gösterilsin ve bu toplum, ilgilenilen hastalık tanısı taşıyanlar (Π_h) ile sağlıklı kişiler (Π_s) olmak üzere iki alt sınıftan oluşsun. Bir birime ait p sayıda tanı testine ilişkin rassal vektör $\mathbf{X} = [X_1, X_2, \dots, X_p]$ ile gösterilsin ($\mathbf{X} \in \mathfrak{R}^p$). X_1, X_2, \dots, X_p tanı testleri bireyleri hasta ve sağlıklı sınıflara ayırabilme özellikleri taşımalıdır (Johnson ve Wichern, 2002; Härdle ve Simar, 2007; Hastie ve ark., 1999).



Şekil 2.3- Sınıflandırma Grafiği

2.4.1- Bayes kuralına göre sınıflandırma

Hasta ya da sağlıklı sınıfa ait rasgele seçilen gözlemlerin ($X = x$) önsel olasılıkları,

$$P(X \in \Pi_i) = \pi_i, \quad i = h, s \quad (2.22)$$

olsun.

i . sınıfa ait X 'in koşullu çok değişkenli olasılık yoğunluk fonksiyonu,

$$P(X = x | X \in \Pi_i) = f_i(x), \quad i = h, s \quad (2.23)$$

olsun.

Bu sınıflara ait koşullu çok değişkenli olasılık yoğunluk fonksiyonunu sürekli ya da kesikli yapıda olabilir. 2.22 ve 2.23 eşitlikleri göz önünde bulundurulduğunda Bayes teoremine göre sonsal (posterior) olasılıklar 2.24'deki eşitlikteki gibi elde edilir (Johnson ve Wichern, 2002).

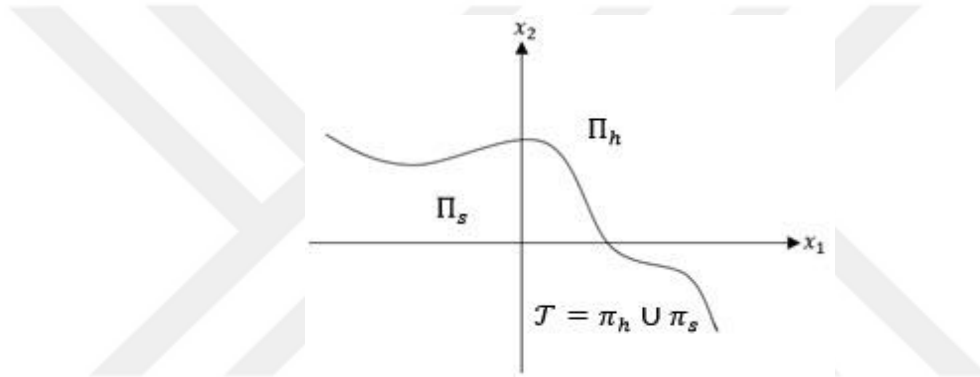
$$p(\Pi_i|x) = P(X \in \Pi_i|X = x) = \frac{f_i(x)\pi_i}{f_h(x)\pi_h + f_s(x)\pi_s}, \quad i = h, s \quad (2.24)$$

Verilen bir x için doğru sınıflandırma stratejisi x 'e ait gözlemi daha yüksek sonsal olasılık değerine sahip olan sınıfa atamaktır. Bu strateji Bayes kuralı sınıflandırıcısı olarak adlandırılır. Bir diğer deyiş ile bir gözlem,

$$\frac{p(\Pi_h|\mathbf{x})}{p(\Pi_s|\mathbf{x})} > 1 \quad (2.25)$$

eşitsizlik 2.25'i sağlıyor ise Π_h sınıfına atanır. Diğer durumlarda Π_s sınıfına atanır. 2.24'e göre eşitlik 2.25 düzenlendiğinde aşağıdaki Bayes kuralı sınıflandırıcısı elde edilir (Johnson ve Wichern, 2002).

$$\frac{f_h(\mathbf{x})}{f_s(\mathbf{x})} > \frac{\pi_s}{\pi_h} \quad (2.26)$$



Şekil 2.4- Bayes Kuralına Göre Sınıflandırma Grafiği

2.5- Çoklu Tanı Testlerinin Birleştirilmesinde Kullanılan Çok Değişkenli Yöntemler

Çoklu tanı testlerinin birleştirilmesinde kullanılan minimax prosedürüne göre en iyi birleştirme yöntemi, ayırma analizi ve lojistik ayırma analizi yöntemleri bu tez çalışmasında kullanılan çok değişkenli yöntemlerdir (Mamtani ve ark., 2006; Ma ve Huang, 2007; Qin ve Zhang, 2010).

2.5.1- Ayırma analizi

Ayırma analizi ve sınıflandırma, önceden sınıflandırılmış iki ya da daha fazla grubu (küme) birbirinden ayıran faktörleri tespit etmeye ve yeni bir gözlemin hangi gruba atanabileceğini gösteren çok değişkenli analiz yöntemleridir. Ayırma analizi açıklayıcı bir yöntem olmasıyla beraber nedensel ilişkilerin belirsiz olduğu durumlarda gözlemsel farklılıkların incelenmesinde kullanılır. Sınıflandırma prosedürleri yeni gözlemleri gruplara atamak için kullanılır (Johnson ve Wichern, 2002).

2.5.1.1- Doğrusal ayırma analizi

Farklı ortalama vektörü ve ortak kovaryans matrisine ($\Sigma_h = \Sigma_s = \Sigma$) sahip çok değişkenli olasılık yoğunluklarını kullanan Fisher varsayımını izleyerek Bayes kuralı sınıflandırıcısı daha özgül hale getirilir (Antonogeorgos ve ark., 2009). Burada hasta ve sağlıklı sınıfı temsil eden alt toplum sırası ile $f_h(\cdot)$, $f_s(\cdot)$ olasılık yoğunluk fonksiyonlarına sahip çok değişkenli normal dağılım gösterebilir. $\Pi_h \sim \mathcal{N}_p(\mu_h, \Sigma)$, $\Pi_s \sim \mathcal{N}_p(\mu_s, \Sigma)$.

İki olasılık yoğunluk fonksiyonunun oranı,

$$\frac{f_h(\mathbf{x})}{f_s(\mathbf{x})} = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_h)' \Sigma^{-1}(\mathbf{x} - \mu_h)\right\}}{\exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_s)' \Sigma^{-1}(\mathbf{x} - \mu_s)\right\}} \quad (2.27)$$

şeklinde tanımlanır. Burada eşit kovaryans matrisinden dolayı $(2\pi)^{-p/2} |\Sigma|^{-1/2}$ normalizasyon faktörü oranda sadeleştirilmiştir. $\log_e(\cdot)$ monoton artan fonksiyonu kullanılarak eşitlik 2.28 aşağıdaki şekilde ifade edilir.

$$\begin{aligned} \log_e \frac{f_h(\mathbf{x})}{f_s(\mathbf{x})} &= (\mu_h - \mu_s)' \Sigma^{-1} \mathbf{x} - \frac{1}{2} (\mu_h - \mu_s)' \Sigma^{-1} (\mu_h + \mu_s) \\ &= (\mu_h - \mu_s)' \Sigma^{-1} (\mathbf{x} - \bar{\mu}) \end{aligned} \quad (2.28)$$

Burada $\bar{\mu} = (\mu_h + \mu_s)/2$. Eşitlik 2.28'nin sağ tarafındaki 2. terim aşağıdaki gibi yazılabilir.

$$(\mu_h - \mu_s)' \Sigma^{-1} (\mu_h + \mu_s) = \mu_h' \Sigma^{-1} \mu_h - \mu_s' \Sigma^{-1} \mu_s \quad (2.29)$$

Böylece aşağıdaki x 'e bağlı doğrusal fonksiyon elde edilmiş olur.

$$L_D(\mathbf{x}) = \log_e \left\{ \frac{f_h(\mathbf{x}) \pi_h}{f_s(\mathbf{x}) \pi_s} \right\} = b_0 + \mathbf{b}' \mathbf{x} \quad (2.30)$$

Bu fonksiyonun katsayıları ise aşağıdaki gibi hesaplanır.

$$\mathbf{b} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_h - \boldsymbol{\mu}_s) \quad (2.31)$$

$$b_0 = -\frac{1}{2}\{\boldsymbol{\mu}'_h \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_h - \boldsymbol{\mu}'_s \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_s\} + \log_e(\pi_s/\pi_h) \quad (2.32)$$

2.5.1.1.1- Doğrusal ayırma analizi için parametrik ROC modeli

Doğrusal ayırma analizinden elde edilen skor değerleri kullanılarak parametrik ROC modeli aşağıdaki şekilde tanımlanır.

Gerçek pozitif ve yalancı pozitif oranı eşitlik 2.33 ve 2.34'de verilmiştir.

$$GP(c) = P(L_D(\mathbf{x}) > c | \mathbf{X} \in \Pi_h) = \Phi\left(\frac{\mu_{D,h} - c}{\sigma_{D,h}}\right) \quad (2.33)$$

$$YP(c) = P(L_D(\mathbf{x}) > c | \mathbf{X} \in \Pi_s) = \Phi\left(\frac{\mu_{D,s} - c}{\sigma_{D,s}}\right) \quad (2.34)$$

Burada,

$\mu_{D,h}, \sigma_{D,h}$ ve $\mu_{D,s}, \sigma_{D,s}$ sırasıyla hasta ve sağlıklı toplum için $L_D(\mathbf{x})$ 'den elde edilen skor değerlerine ait ortalama ve standart sapma değerlerini göstermektedir (Lin, Zhou ve ark., 2011).

Yalancı pozitif oranı $YP(c) = y$ ile gösterilirse, eşitlik 2.36 ve eşitlik 2.37 kullanılarak ROC eğrisi fonksiyonu eşitlik 2.35'de verilmiştir (Johnson ve Wichern, 2002).

$$ROC(y) = \Phi(a + b\Phi^{-1}(y)) \quad (2.35)$$

Burada;

$$a = \frac{\mu_{D,h} - \mu_{D,s}}{\sigma_{D,h}}, b = \frac{\sigma_{D,s}}{\sigma_{D,h}} \quad (2.36)$$

Yalancı pozitiflik oranı (y) için kesme noktası,

$$c = \mu_{D,s} - \sigma_{D,s}\Phi^{-1}(y) \quad (2.37)$$

şeklinde tanımlanır.

Eđri altında kalan alan (AUC) eřitlik 2.38'de verilmiřtir.

$$AUC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \quad (2.38)$$

a, b ve AUC 'nin en byk benzerlik tahmincileri $\bar{L}_{D,h}, \bar{L}_{D,s}, LS_{D,h}$ ve $LS_{D,s}$ kullanılarak elde edilir.

$$\bar{L}_{D,h} = \frac{\sum_{i=1}^{n_h} L_{i D,h}}{n_h}, \quad LS_{D,h} = \sqrt{\frac{\sum_{i=1}^{n_h} (L_{i D,h} - \bar{L}_{D,h})^2}{n_h - 1}} \quad (2.39)$$

$$\bar{L}_{D,s} = \frac{\sum_{j=1}^{n_s} L_{j D,s}}{n_s}, \quad LS_{D,s} = \sqrt{\frac{\sum_{j=1}^{n_s} (L_{j D,s} - \bar{L}_{D,s})^2}{n_s - 1}} \quad (2.40)$$

$$\hat{a} = \frac{\bar{L}_{D,h} - \bar{L}_{D,s}}{LS_{D,h}}, \quad V(\hat{a}) = \frac{n_h(\hat{a}^2 + 2) + 2n_s\hat{b}^2}{2n_hn_s} \quad (2.41)$$

$$\hat{b} = \frac{LS_{D,s}}{LS_{D,h}}, \quad V(\hat{b}) = \frac{(n_h + n_s)\hat{b}^2}{2n_hn_s} \quad (2.42)$$

$$\widehat{AUC} = \Phi\left(\frac{\hat{a}}{\sqrt{1+\hat{b}^2}}\right) \quad (2.43)$$

Burada,

$L_{i D,h}$: Hasta grubundaki birimlerin dođrusal ayırma analizinden elde edilen skor deđerleri

$L_{j D,s}$: Sađlıklı gruptaki birimlerin dođrusal ayırma analizinden elde edilen skor deđerleri

$\bar{L}_{D,h}$: Hasta grubundaki birimlerin dođrusal ayırma analizinden elde edilen skor deđerlerine ait ortalama

$\bar{L}_{D,s}$: Sađlıklı gruptaki birimlerin dođrusal ayırma analizinden elde edilen skor deđerlerine ait ortalama

$LS_{D,h}$: Hasta grubundaki birimlerin dođrusal ayırma analizinden elde edilen skor deđerlerine ait standart sapma

$LS_{D,s}$: Sađlıklı gruptaki birimlerin dođrusal ayırma analizinden elde edilen skor deđerlerine ait standart sapma

n_h : Hasta sınıftaki birim sayısı

n_s : Sađlıklı sınıftaki birim sayısı göstermektedir.

Youden indeks kullanılarak, optimum kesim noktası (c) için tahmin değeri eşitlik 2.44'deki gibi hesaplanır.

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.44)$$

2.5.1.1.2- Doğrusal ayırma analizi için parametrik olmayan ROC modeli

Doğrusal ayırma analizinden elde edilen skor değerleri kullanılarak parametrik olmayan ROC modeli aşağıdaki şekilde tanımlanır (Schulzer, 1994; Segen ve Wade, 2002).

Gerçek pozitif ve yalancı pozitif oranı eşitlik 2.45 ve 2.46'da verilmiştir.

$$GP(c) = P(L_D(x) > c | L_D(x) \in \Pi_h) = \frac{s_h(c)}{n_h} \quad (2.45)$$

$$YP(c) = P(L_D(x) > c | L_D(x) \in \Pi_s) = \frac{s_s(c)}{n_s} \quad (2.46)$$

Burada,

- $s_h(c)$: Hasta sınıfında skor değeri c'den büyük olanların sayısı
- n_h : Hasta sınıftaki birim sayısı
- $s_s(c)$: Sağlıklı sınıfında skor değeri c'den büyük olanların sayısı
- n_s : Sağlıklı sınıftaki birim sayısı

göstermektedir.

Skor değerlerine ait eğri altında kalan alanın tahmin değeri eşitlik 2.47'de verilmiştir.

$$\widehat{AUC} = \frac{1}{n_h n_s} \sum_{i=1}^{n_h} \sum_{j=1}^{n_s} \psi(L_{iD,h}, L_{jD,s}) \quad (2.47)$$

Burada;

$$\psi(L_{iD,h}, L_{jD,s}) = \begin{cases} 1 & \text{Eğer } L_{iD,h} > L_{jD,s} \\ \frac{1}{2} & \text{Eğer } L_{iD,h} = L_{jD,s} \\ 0 & \text{Eğer } L_{iD,h} < L_{jD,s} \end{cases} \quad (2.48)$$

$L_{iD,h}$ ve $L_{jD,s}$ sırası ile hasta ve sağlıklı birimlerin skor değerlerini göstermektedir.

Skor değeri için optimum kesim noktasının tahmin değeri (c) eşitlik 2.49'daki gibi elde edilir.

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.49)$$

Tahmin edilen AUC'nin varyansı eşitlik 2.50'deki gibi hesaplanır.

$$V(\widehat{AUC}) = \frac{\widehat{AUC}(1 - \widehat{AUC}) + (n_h - 1)(Q_h - \widehat{AUC}^2)}{n_h n_s} + \frac{(n_s - 1)(Q_s - \widehat{AUC}^2)}{n_h n_s} \quad (2.50)$$

Burada;

$$Q_h = \frac{1}{n_s n_h^2} \sum_{L_D(x)} n_s^{=L_D(x)} \times \left[(n_h^{>L_D(x)})^2 + n_h^{>L_D(x)} \times n_h^{=L_D(x)} + \frac{(n_h^{=L_D(x)})^2}{3} \right] \quad (2.51)$$

$$Q_s = \frac{1}{n_s^2 n_h} \sum_{L_D(x)} n_h^{=L_D(x)} \times \left[(n_s^{<L_D(x)})^2 + n_s^{<L_D(x)} \times n_s^{=L_D(x)} + \frac{(n_s^{=L_D(x)})^2}{3} \right] \quad (2.52)$$

Burada,

$n_s^{=L_D(x)}$: skor değerleri $L_D(x)$ 'e eşit olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{=L_D(x)}$: skor değerleri $L_D(x)$ 'e eşit olan gerçek pozitif birimlerin sayısını göstermektedir.

$n_s^{<L_D(x)}$: skor değerleri $L_D(x)$ 'den küçük olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{>L_D(x)}$: skor değerleri $L_D(x)$ 'den büyük olan gerçek pozitif birimlerin sayısını göstermektedir.

2.5.1.2- Karesel ayırma analizi

Kovaryans matrisleri birbirine eşit olmadığı durumlarda sınıflandırma kuralı olumsuz yönde etkilenir. Bu durumda eşitlik 2.28 aşağıdaki şekilde gösterilir.

$$\log_e \frac{f_h(x)}{f_s(x)} = \frac{1}{2} \left\{ \log_e \frac{|\Sigma_h|}{|\Sigma_s|} + \mu'_h \Sigma_h^{-1} \mu_h - \mu'_s \Sigma_s^{-1} \mu_s \right\} - \frac{1}{2} x' (\Sigma_h^{-1} - \Sigma_s^{-1}) x + (\mu'_h \Sigma_h^{-1} - \mu'_s \Sigma_s^{-1}) x \quad (2.53)$$

Bu durumda eşitlik 2.53'deki x 'in karesel bir fonksiyonu haline dönüşür. Aşağıdaki eşitlik 2.54 'e dönüştürülür (Johnson ve Wichern, 2002).

$$L_K(x) = b_0 + b'x + x'\Omega x \quad (2.54)$$

Burada,

$$\Omega = -\frac{1}{2} (\Sigma_h^{-1} - \Sigma_s^{-1}) \quad (2.55)$$

$$b = \Sigma_h^{-1} \mu_h - \Sigma_s^{-1} \mu_s \quad (2.56)$$

$$b_0 = -\frac{1}{2} \left\{ \log_e \frac{|\Sigma_h|}{|\Sigma_s|} + \mu'_h \Sigma_h^{-1} \mu_h - \mu'_s \Sigma_s^{-1} \mu_s \right\} - \log_e (\pi_s / \pi_h) \quad (2.57)$$

şeklinde hesaplanır (Johnson ve Wichern, 2002).

2.5.1.2.1- Karesel ayırma analizi için parametrik ROC modeli

Gerçek pozitif ve yalancı pozitif oranı eşitlik 2.58 ve 2.59'da verilmiştir.

$$GP(c) = P(L_K(x) > c | X \in \Pi_h) = \Phi \left(\frac{\mu_{K,h} - c}{\sigma_{K,h}} \right) \quad (2.58)$$

$$YP(c) = P(L_K(x) > c | X \in \Pi_s) = \Phi \left(\frac{\mu_{K,s} - c}{\sigma_{K,s}} \right) \quad (2.59)$$

Burada,

$\mu_{K,h}$, $\sigma_{K,h}$ ve $\mu_{K,s}$, $\sigma_{K,s}$ sırasıyla hasta ve sağlıklı toplum için $L_K(x)$ 'den elde edilen skor değerlerine ait ortalama ve standart sapma değerlerini göstermektedir.

Yalancı pozitif oranı $YP(c) = y$ ile gösterilirse, eşitlik 2.58 ve eşitlik 2.59 kullanılarak ROC eğrisi fonksiyonu eşitlik 2.60'da verilmiştir.

$$ROC(y) = \Phi(a + b\Phi^{-1}(y)) \quad (2.60)$$

Burada;

$$a = \frac{\mu_{K,h} - \mu_{K,s}}{\sigma_{K,h}}, b = \frac{\sigma_{K,s}}{\sigma_{K,h}} \quad (2.61)$$

Yalancı pozitiflik oranı (y) için kesme noktası,

$$c = \mu_{K,s} - \sigma_{K,s} \Phi^{-1}(y) \quad (2.62)$$

şeklinde tanımlanır.

Eğri altında kalan alan (AUC) eşitlik 2.63'de verilmiştir.

$$AUC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \quad (2.63)$$

a, b ve AUC 'nin en büyük benzerlik tahmincileri $\bar{L}_{K,h}, \bar{L}_{K,s}, LS_{K,h}$ ve $LS_{K,s}$ kullanılarak elde edilir.

$$\bar{L}_{K,h} = \frac{\sum_{i=1}^{n_h} L_{iK,h}}{n_h}, \quad LS_{K,h} = \sqrt{\frac{\sum_{i=1}^{n_h} (L_{iK,h} - \bar{L}_{K,h})^2}{n_h - 1}} \quad (2.64)$$

$$\bar{L}_{K,s} = \frac{\sum_{j=1}^{n_s} L_{jK,s}}{n_s}, \quad LS_{K,s} = \sqrt{\frac{\sum_{j=1}^{n_s} (L_{jK,s} - \bar{L}_{K,s})^2}{n_s - 1}} \quad (2.65)$$

$$\hat{a} = \frac{\bar{L}_{K,h} - \bar{L}_{K,s}}{LS_{K,h}}, \quad V(\hat{a}) = \frac{n_h(\hat{a}^2 + 2) + 2n_s\hat{b}^2}{2n_h n_s} \quad (2.66)$$

$$\hat{b} = \frac{LS_{K,s}}{LS_{K,h}}, \quad V(\hat{b}) = \frac{(n_h + n_s)\hat{b}^2}{2n_h n_s} \quad (2.67)$$

$$\widehat{AUC} = \Phi\left(\frac{\hat{a}}{\sqrt{1+\hat{b}^2}}\right) \quad (2.68)$$

Burada,

$L_{iK,h}$: Hasta grubundaki birimlerin karesel ayırma analizinden elde edilen skor değerleri

$L_{jK,s}$: Sağlıklı gruptaki birimlerin karesel ayırma analizinden elde edilen skor değerleri

$\bar{L}_{K,h}$: Hasta grubundaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait ortalama

$\bar{L}_{K,s}$: Sağlıklı gruptaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait ortalama

$LS_{K,h}$: Hasta grubundaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait standart sapma

$LS_{K,s}$: Sağlıklı gruptaki birimlerin karesel ayırma analizinden elde edilen skor değerlerine ait standart sapma

n_h : hasta sınıftaki birim sayısı

n_s : sağlıklı sınıftaki birim sayısı

göstermektedir.

Youden indeks kullanılarak, optimum kesim noktası (c) için tahmin değeri eşitlik 2.69 gibi hesaplanır (Wu ve ark., 1996).

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.69)$$

2.5.1.2.2- Karesel ayırma analizi için parametrik olmayan ROC modeli

Karesel ayırma analizinden elde edilen skor değerleri kullanılarak parametrik olmayan ROC modeli aşağıdaki şekilde tanımlanır (Colak ve ark., 2012).

Gerçek pozitif ve yalancı pozitif oranı eşitlik 2.70 ve 2.71'de verilmiştir.

$$GP(c) = P(L_K(x) > c | L_K(x) \in \Pi_h) = \frac{s_h(c)}{n_h} \quad (2.70)$$

$$YP(c) = P(L_K(x) > c | L_K(x) \in \Pi_s) = \frac{s_s(c)}{n_s} \quad (2.71)$$

Burada,

$s_h(c)$: Hasta sınıfında skor değeri c'den büyük olanların sayısı

n_h : Hasta sınıftaki birim sayısı

$s_s(c)$: Sağlıklı sınıfında skor değeri c'den büyük olanların sayısı

n_s : Sağlıklı sınıftaki birim sayısı

göstermektedir.

Skor değerlerine ait eğri altında kalan alanın tahmin değeri eşitlik 2.72'de verilmiştir.

$$\widehat{AUC} = \frac{1}{n_h n_s} \sum_{i=1}^{n_h} \sum_{j=1}^{n_s} \psi(L_{iK,h}, L_{jK,s}) \quad (2.72)$$

Burada;

$$\psi(L_{iK,h}, L_{jK,s}) = \begin{cases} 1 & \text{Eğer } L_{iK,h} > L_{jK,s} \\ \frac{1}{2} & \text{Eğer } L_{iK,h} = L_{jK,s} \\ 0 & \text{Eğer } L_{iK,h} < L_{jK,s} \end{cases} \quad (2.73)$$

$L_{iK,h}$ ve $L_{jK,s}$ sırası ile hasta ve sağlıklı birimlerin skor değerlerini göstermektedir.

Skor değeri için optimum kesim noktasının tahmin değeri (c) eşitlik 2.74'deki gibi elde edilir.

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.74)$$

Tahmin edilen AUC'nin varyansı eşitlik 2.75'deki gibi hesaplanır.

$$V(\widehat{AUC}) = \frac{\widehat{AUC}(1 - \widehat{AUC}) + (n_h - 1)(Q_h - \widehat{AUC}^2)}{n_h n_s} + \frac{(n_s - 1)(Q_s - \widehat{AUC}^2)}{n_h n_s} \quad (2.75)$$

Burada;

$$Q_h = \frac{1}{n_s n_h^2} \sum_{L_K(x)} n_s^{=L_K(x)} \times \left[\left(n_h^{>L_K(x)} \right)^2 + n_h^{>L_K(x)} \times n_h^{=L_K(x)} + \frac{\left(n_h^{=L_K(x)} \right)^2}{3} \right] \quad (2.76)$$

$$Q_s = \frac{1}{n_s^2 n_h} \sum_{L_K(x)} n_h^{=L_K(x)} \times \left[\left(n_s^{<L_K(x)} \right)^2 + n_s^{<L_K(x)} \times n_s^{=L_K(x)} + \frac{\left(n_s^{=L_K(x)} \right)^2}{3} \right] \quad (2.77)$$

Burada,

$n_s^{=L_K(x)}$: Skor değerleri $L_K(x)$ 'e eşit olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{=L_K(x)}$: Skor değerleri $L_K(x)$ 'e eşit olan gerçek pozitif birimlerin sayını göstermektedir.

$n_s^{<L_K(x)}$: Skor değerleri $L_K(x)$ 'den küçük olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{>L_K(x)}$: Skor değerleri $L_K(x)$ 'den büyük olan gerçek pozitif birimlerin sayısını göstermektedir.

2.5.2- Lojistik ayırma analizi

Bayes kuralında belirtilen sonsal olasılıklar kullanılarak lojistik regresyon eşit 2.78'deki gibi elde edilir.

$$\text{logit} [p(\Pi_h|\mathbf{x})] = L_L(\mathbf{x}) = \log_e \left(\frac{p(\Pi_h|\mathbf{x})}{1 - p(\Pi_h|\mathbf{x})} \right) = b_0 + \mathbf{b}'\mathbf{x} \quad (2.78)$$

Ayrırma analizine lojistik yaklaşım uygulandığında eşitlik 2.22 aşağıdaki şekilde \mathbf{x} 'in doğrusal bir fonksiyonu olarak modellenebilir (Friedman, 1991; King ve Rayn, 2002; Komarek ve Moore, 2003; Moawed ve Osman, 2017; Wilson ve Langenberg, 1999).

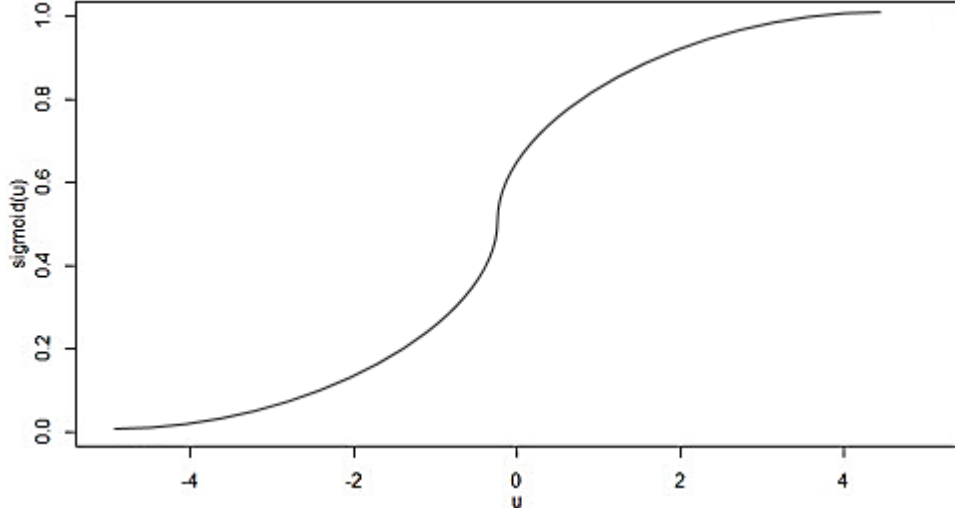
$$p(\Pi_h|\mathbf{x}) = \frac{e^{L_L(\mathbf{x})}}{1 + e^{L_L(\mathbf{x})}} \quad (2.79)$$

$$p(\Pi_s|\mathbf{x}) = \frac{e^{L_L(\mathbf{x})}}{1 + e^{L_L(\mathbf{x})}} \quad (2.80)$$

Eşitlik 2.22, eşitlik 2.81 haline dönüştürülebilir.

$$p(\Pi_h|\mathbf{x}) = \frac{1}{1 + e^{-L_L(\mathbf{x})}} = \sigma(L_L(\mathbf{x})) \quad (2.81)$$

Burada $u \in \mathfrak{R} (0,1)$ aralığında değerler alan $\sigma(u) = 1/(1 + e^{-u})$ S-şeklinde sigmoid fonksiyondur.



Şekil 2.5- Sigmoid Fonksiyonu

En çok olabilirlik tahmini

X 'in ait olduğu sınıfları tanımlayan W cevap değişkenini,

$$W = \begin{cases} 1 & X \in \Pi_h \text{ ise} \\ 0 & \text{Diğer Durumlarda} \end{cases} \quad (2.82)$$

şeklinde tanımlansın (Johnson ve Wichern, 2002).

Y Bernoulli dağılımı gösterir. Verilen n gözlem için, (X_j, W_j) $j = 1, 2, \dots, n$ için koşullu olabilirlik fonksiyonu, eşitlik 2.78 göz önünde bulundurulur $p(\Pi_h | \mathbf{x})$ için $p_h(\mathbf{x}, b_0, \mathbf{b})$ ve benzer şekilde $p(\Pi_s | \mathbf{x})$ için $p_s(\mathbf{x}, b_0, \mathbf{b})$ şeklinde ifade edilirse eşitlik 2.83'deki gibi gösterilir.

$$\ell(b_0, \mathbf{b}) = \prod_{j=1}^n \left(p_h(\mathbf{x}_j, b_0, \mathbf{b}) \right)^{w_j} \left(1 - p_h(\mathbf{x}_j, b_0, \mathbf{b}) \right)^{1-w_j} \quad (2.83)$$

Buradan koşullu log-olabilirlik fonksiyonu ise eşitlik 2.84 ve 2.85'deki gibi elde edilir.

$$\ell(b_0, \mathbf{b}) = \sum_{j=1}^n \left\{ w_j \log_e p_h(\mathbf{x}_j, b_0, \mathbf{b}) + (1 - w_j) \log_e (1 - p_h(\mathbf{x}_j, b_0, \mathbf{b})) \right\} \quad (2.84)$$

$$\ell(b_0, \mathbf{b}) = \sum_{j=1}^n \left\{ w_j (b_0 + \mathbf{b}'\mathbf{x}_j) - \log_e (1 - e^{b_0 + \mathbf{b}'\mathbf{x}_j}) \right\} \quad (2.85)$$

(b_0, \mathbf{b}) 'nın en çok olabilirlik tahminleri, $\ell(b_0, \mathbf{b})$ fonksiyonu kullanılarak iteratif tekrar ağırlıklandırılmış en küçük kareler algoritması ile elde edilir.

2.5.2.1- Lojistik ayırma analizi için parametrik ROC modeli

Gerçek pozitif ve yalancı pozitif oranı eşitlik 2.86 ve 2.87'de verilmiştir.

$$GP(c) = P(L_L(\mathbf{x}) > c | \mathbf{X} \in \Pi_h) = \Phi\left(\frac{\mu_{L,h} - c}{\sigma_{L,h}}\right) \quad (2.86)$$

$$YP(c) = P(L_L(\mathbf{x}) > c | \mathbf{X} \in \Pi_s) = \Phi\left(\frac{\mu_{L,s} - c}{\sigma_{L,s}}\right) \quad (2.87)$$

Burada,

$\mu_{L,h}, \sigma_{L,h}$ ve $\mu_{L,s}, \sigma_{L,s}$ sırasıyla hasta ve sağlıklı toplum için $L_L(\mathbf{x})$ 'den elde edilen skor değerlerine ait ortalama ve standart sapma değerlerini göstermektedir (Shultz, 1987; Sharma, 2014; Swets ve ark, 2000; Yu ve Park, 2015).

Yalancı pozitif oranı $YP(c) = y$ ile gösterilirse, eşitlik 2.89 ve eşitlik 2.90 kullanılarak ROC eğrisi fonksiyonu eşitlik 2.88'de verilmiştir.

$$ROC(y) = \Phi(a + b\Phi^{-1}(y)) \quad (2.88)$$

Burada;

$$a = \frac{\mu_{L,h} - \mu_{L,s}}{\sigma_{L,h}}, b = \frac{\sigma_{L,s}}{\sigma_{L,h}} \quad (2.89)$$

Yalancı pozitiflik oranı (y) için kesme noktası,

$$c = \mu_{L,s} - \sigma_{L,s}\Phi^{-1}(y) \quad (2.90)$$

şeklinde tanımlanır.

Eğri altında kalan alan (AUC) eşitlik 2.91'da verilmiştir.

$$AUC = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right) \quad (2.91)$$

a, b ve AUC 'nin en büyük benzerlik tahmincileri $\bar{L}_{L,h}, \bar{L}_{L,s}, LS_{L,h}$ ve $LS_{L,s}$ kullanılarak elde edilir.

$$\bar{L}_{L,h} = \frac{\sum_{i=1}^{n_h} L_{iL,h}}{n_h}, \quad LS_{L,h} = \sqrt{\frac{\sum_{i=1}^{n_h} (L_{iL,h} - \bar{L}_{L,h})^2}{n_h - 1}} \quad (2.92)$$

$$\bar{L}_{L,s} = \frac{\sum_{j=1}^{n_s} L_{jL,s}}{n_s}, \quad LS_{L,s} = \sqrt{\frac{\sum_{j=1}^{n_s} (L_{jL,s} - \bar{L}_{L,s})^2}{n_s - 1}} \quad (2.93)$$

$$\hat{a} = \frac{\bar{L}_{L,h} - \bar{L}_{L,s}}{LS_{L,h}}, \quad V(\hat{a}) = \frac{n_h(\hat{a}^2 + 2) + 2n_s\hat{b}^2}{2n_h n_s} \quad (2.94)$$

$$\hat{b} = \frac{LS_{L,s}}{LS_{L,h}}, \quad V(\hat{b}) = \frac{(n_h + n_s)\hat{b}^2}{2n_h n_s} \quad (2.95)$$

$$\widehat{AUC} = \Phi\left(\frac{\hat{a}}{\sqrt{1 + \hat{b}^2}}\right) \quad (2.96)$$

Burada,

$L_{iL,h}$: Hasta grubundaki birimlerin lojistik ayırma analizinden elde edilen skor değerleri

$L_{jL,s}$: Sağlıklı gruptaki birimlerin lojistik ayırma analizinden elde edilen skor değerleri

$\bar{L}_{L,h}$: Hasta grubundaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait ortalama

$\bar{L}_{L,s}$: Sağlıklı gruptaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait ortalama

$LS_{L,h}$: Hasta grubundaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait standart sapma

$LS_{L,s}$: Sağlıklı gruptaki birimlerin lojistik ayırma analizinden elde edilen skor değerlerine ait standart sapma

n_h : Hasta sınıftaki birim sayısı

n_s : Sağlıklı sınıftaki birim sayısı

göstermektedir.

Youden indeks kullanılarak, optimum kesim noktası (c) için tahmin değeri eşitlik 2.97'deki gibi hesaplanır (Copas, 1988; DeLong ve Clarke-Pearson, 1988).

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.97)$$

2.5.2.2- Lojistik ayırma analizi için parametrik olmayan ROC modeli

Lojistik ayırma analizinden elde edilen skor değerleri kullanılarak parametrik olmayan ROC modeli aşağıdaki şekilde tanımlanır (Swets, 1988).

Gerçek pozitif ve yalancı pozitif oranı eşitlik 2.98 ve 2.99'da verilmiştir.

$$GP(c) = P(L_L(x) > c | L_L(x) \in \Pi_h) = \frac{s_h(c)}{n_h} \quad (2.98)$$

$$YP(c) = P(L_L(x) > c | L_L(x) \in \Pi_s) = \frac{s_s(c)}{n_s} \quad (2.99)$$

Burada,

- $s_h(c)$: Hasta sınıfında skor değeri c 'den büyük olanların sayısı
- n_h : Hasta sınıftaki birim sayısı
- $s_s(c)$: Sağlıklı sınıfında skor değeri c 'den büyük olanların sayısı
- n_s : Sağlıklı sınıftaki birim sayısı

göstermektedir.

Skor değerlerine ait eğri altında kalan alanın tahmin değeri eşitlik 2.100'da verilmiştir.

$$\widehat{AUC} = \frac{1}{n_h n_s} \sum_{i=1}^{n_h} \sum_{j=1}^{n_s} \psi(L_{iL,h}, L_{jL,s}) \quad (2.100)$$

Burada;

$$\psi(L_{iL,h}, L_{jL,s}) = \begin{cases} 1 & \text{Eğer } L_{iL,h} > L_{jL,s} \\ \frac{1}{2} & \text{Eğer } L_{iL,h} = L_{jL,s} \\ 0 & \text{Eğer } L_{iL,h} < L_{jL,s} \end{cases} \quad (2.101)$$

$L_{iL,h}$ ve $L_{jL,s}$ sırası ile hasta ve sağlıklı birimlerin skor değerlerini göstermektedir.

Skor değeri için optimum kesim noktasının tahmin değeri (c) eşitlik 2.102'deki gibi elde edilir.

$$\hat{c} = \max(\hat{J}) = \max[\widehat{GP}(c) - \widehat{YP}(c)] \quad (2.102)$$

Tahmin edilen AUC'nin varyansı eşitlik 2.103'deki gibi hesaplanır.

$$V(\widehat{AUC}) = \frac{\widehat{AUC}(1 - \widehat{AUC}) + (n_h - 1)(Q_h - \widehat{AUC}^2)}{n_h n_s} + \frac{(n_s - 1)(Q_s - \widehat{AUC}^2)}{n_h n_s} \quad (2.103)$$

Burada;

$$Q_h = \frac{1}{n_s n_h^2} \sum_{L_L(x)} n_s^{=L_L(x)} \times \left[\left(n_h^{>L_L(x)} \right)^2 + n_h^{>L_L(x)} \times n_h^{=L_L(x)} + \frac{\left(n_h^{=L_L(x)} \right)^2}{3} \right] \quad (2.104)$$

$$Q_s = \frac{1}{n_s^2 n_h} \sum_{L_L(x)} n_h^{=L_L(x)} \times \left[\left(n_s^{<L_L(x)} \right)^2 + n_s^{<L_L(x)} \times n_s^{=L_L(x)} + \frac{\left(n_s^{=L_L(x)} \right)^2}{3} \right] \quad (2.105)$$

Burada,

$n_s^{=L_L(x)}$: Skor değerleri $L_L(x)$ 'e eşit olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{=L_L(x)}$: Skor değerleri $L_L(x)$ 'e eşit olan gerçek pozitif birimlerin sayısını göstermektedir.

$n_s^{<L_L(x)}$: Skor değerleri $L_L(x)$ 'den küçük olan gerçek negatif birimlerin sayısını göstermektedir.

$n_h^{>L_L(x)}$: Skor değerleri $L_L(x)$ 'den büyük olan gerçek pozitif birimlerin sayısını göstermektedir.

2.5.3- MiniMax prosedürüne göre doğrusal en iyi birleştirme yöntemi

Minimax prosedürünü (Anderson ve Bahadur, 1962) kullanılarak Sameera ve ark. (2016) 'larının önermiş olduğu tanı testlerinin doğrusal kombinasyonları aşağıdaki eşitlik 2.106'daki gibi verilmiştir.

$$\mathbf{b} = [t\boldsymbol{\Sigma}_h + (1 - t)\boldsymbol{\Sigma}_s]^{-1}(\boldsymbol{\mu}_h - \boldsymbol{\mu}_s) \quad (2.106)$$

Burada;

t : (0, 1) aralığında değerler alan sabiti göstermektedir. t 'nin değeri deneme yanılma yöntemi ile belirlenir.

\mathbf{b} : ($px1$) boyutlu doğrusal modeldeki katsayılar vektörü

L_M : \mathbf{b} vektöründeki katsayılar kullanılarak tanı testlerinden elde edilen doğrusal skor değeridir. Aşağıdaki eşitlikte 2.107'deki gibi hesaplanır.

$$L_M(\mathbf{x}) = \mathbf{b}'\mathbf{x} \quad (2.107)$$

$$c = \mathbf{b}'\boldsymbol{\mu}_s + \sqrt{\mathbf{b}'\boldsymbol{\Sigma}_s\mathbf{b}} \Phi^{-1}(1 - x) \quad (2.108)$$

Eşitlik 2.107 ve 2.108 kullanılarak gerçek pozitif ve yalancı pozitif oranı eşitlik 2.109 ve 2.110'de verilmiştir.

$$GP(c) = P(L_M(\mathbf{x}) > c | \mathbf{X} \in \Pi_h) = \Phi\left(\frac{\mathbf{b}'\boldsymbol{\mu}_h - c}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_h\mathbf{b}}}\right) \quad (2.109)$$

$$YP(c) = P(L_M(\mathbf{x}) > c | \mathbf{X} \in \Pi_s) = \Phi\left(\frac{\mathbf{b}'\boldsymbol{\mu}_s - c}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_s\mathbf{b}}}\right) \quad (2.110)$$

Eşitlik 2.108 ve 2.110 kullanılarak gerçek pozitiflik oranı eşitlik 2.111'de verilmiştir.

$$GP = \Phi\left(\frac{\mathbf{b}'(\boldsymbol{\mu}_h - \boldsymbol{\mu}_s) - \sqrt{\mathbf{b}'\boldsymbol{\Sigma}_s\mathbf{b}} \Phi^{-1}(1 - y)}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_h\mathbf{b}}}\right) \quad (2.111)$$

Burada,

y yalancı pozitiflik oranını göstermektedir.

Minimax prosedürü GP ile $1 - YP$ oranların maksimize ederken, yanlış sınıflandırma oranını minimize etme durumudur. (Sameera, Vardhan ve ark., 2016; Anderson and Bahadur, 1961).

Minimax prosedürüne göre kesim noktası olan c aşağıdaki eşitlikte 2.112'deki gibi elde edilir.

$$c = \frac{\mathbf{b}'\boldsymbol{\mu}_h\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_s\mathbf{b}} + \mathbf{b}'\boldsymbol{\mu}_s\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_h\mathbf{b}}}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_h\mathbf{b}} + \sqrt{\mathbf{b}'\boldsymbol{\Sigma}_s\mathbf{b}}} \quad (2.112)$$

Çok deęişkenli ROC analizi için eęri altında kalan alan eřitlik 2.113'deki gibidir.

$$AUC = \Phi \left\{ \frac{\mathbf{b}'(\boldsymbol{\mu}_h - \boldsymbol{\mu}_s)}{\sqrt{\mathbf{b}'(\boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_h)\mathbf{b}}} \right\} \quad (2.113)$$

Çok deęişkenli ROC modeli eęrisi monoton artan bir eęridir. Bu eęrinin eęimi olabilirlik oranına eřittir. Tanı testlerine uygulanacak dönüşümlerden etkilenmez (Sameera ve ark., 2016).

Çok deęişkenli ROC modeli için Youden indeksi eřitlik 2.114'deki gibi verilmiştir.

$$J = \max[GP - YP] = \max \left[\Phi \left(\frac{\mathbf{b}'\boldsymbol{\mu}_h - c}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_h\mathbf{b}}} \right) - \Phi \left(\frac{\mathbf{b}'\boldsymbol{\mu}_s - c}{\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_s\mathbf{b}}} \right) \right] \quad (2.114)$$

3- GEREÇ VE YÖNTEM

Sürekli yapıdaki çoklu tanı testleri için birleştirme yöntemlerini karşılaştırmak amacı ile 1000 tekrarlı Monte Carlo simülasyon tekniği kullanıldı. Doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizi ve en iyi doğrusal birleştirme yöntemleri farklı senaryolar altında karşılaştırıldı. Doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizinden elde edilen skor değerleri hem parametrik hemde parametrik olmayan ROC eğrisi yöntemine yerleştirilerek farklı sonuçlar elde edildi.

Çalışmada kullanılan örneklem büyüklükleri hasta ve sağlıklı gruplar için eşit sayıda ve $n=50, 100, 250$ ve 500 olarak belirlendi. Farklı senaryolar altında veri türetimi gerçekleştirmek için çok değişkenli normal dağılımdan faydalanıldı. Yöntemlerin karşılaştırılması ROC eğrileri altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeksler göz önünde bulundurularak gerçekleştirildi.

Simülasyon çalışmaları dışında yöntemler sağlık alanına ait bir uygulama veri seti üzerinde de karşılaştırılmıştır. Veriler "Sağlık Bilimleri Üniversitesi Tepecik Eğitim ve Araştırma Hastanesi" 'nden elde edildi.

Simülasyon çalışmaları ve uygulama veri setinin analizi R Studio (v1.1.383) paket programında gerçekleştirildi. *mvtnorm*, *plotROC*, *ggplot2*, *pROC*, *dplyr*, *haven* ve *sortinghat* paketleri kullanılarak sonuçlar elde edildi.

3.1- Simülasyon Çalışmasında Kullanılan Senaryolar

Yöntemlerin karşılaştırılmasında veri türetimi için 5 farklı senaryo kullanılmıştır. Her bir senaryoda 4 farklı tanı testi olduğu varsayılmıştır. İlk 4 senaryo çok değişkenli normal dağılım kullanılarak, 5. senaryoda ise bozulmuş (kontamine olmuş) çok değişkenli normal dağılım modeli kullanılarak veri türetimleri yapılmıştır. Kullanılan senaryolara ait detaylar aşağıdaki gibidir:

Senaryo I: Ortalama vektörleri aynı, kovaryans matrisleri aynı. Bu senaryo kullanılarak yapılacak karşılaştırmalarda hasta ve sağlıklı gruplara ait kullanılan ortalama vektörleri ve kovaryans matrisleri birbirine eşit aşağıdaki gibi seçilmiştir.

$$\mu_h = \mu_s = \begin{bmatrix} 0.5 \\ 0.8 \\ 1 \\ 1.5 \end{bmatrix}$$

$$\Sigma_h = \Sigma_s = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

Senaryo II: Ortalama vektörleri aynı, kovaryans matrisleri farklı. Bu senaryo kullanılarak yapılacak karşılaştırmalarda hasta ve sağlıklı gruplara ait kullanılan ortalama vektörleri birbirine eşit ve kovaryans matrisleri birbirinden farklı olacak şekilde aşağıdaki gibi seçilmiştir.

$$\mu_h = \mu_s = \begin{bmatrix} 0.5 \\ 0.8 \\ 1 \\ 1.5 \end{bmatrix}$$

$$\Sigma_h = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 \end{bmatrix} \quad \Sigma_s = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix}$$

Senaryo III: Ortalama vektörleri farklı, kovaryans matrisleri aynı. Bu senaryo kullanılarak yapılacak karşılaştırmalarda hasta ve sağlıklı gruplara ait kullanılan ortalama vektörleri birbirinden farklı ve kovaryans matrisleri aynı olacak şekilde aşağıdaki gibi seçilmiştir.

$$\mu_h = \begin{bmatrix} 0.5 \\ 0.8 \\ 1 \\ 1.5 \end{bmatrix} \quad \mu_s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Sigma_h = \Sigma_s = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

Senaryo IV: Ortalama vektörleri farklı, kovaryans matrisleri farklı. Bu senaryo kullanılarak yapılacak karşılaştırmalarda hasta ve sağlıklı gruplara ait kullanılan ortalama vektörleri birbirinden farklı ve kovaryans matrisleri de birbirinden farklı olacak şekilde aşağıdaki gibi seçilmiştir.

$$\mu_h = \begin{bmatrix} 0.5 \\ 0.8 \\ 1 \\ 1.5 \end{bmatrix} \quad \mu_s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Sigma_h = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 \end{bmatrix} \quad \Sigma_s = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix}$$

Senaryo V: Çok deęişkenli normal daęılım kullanılarak bozulmuş (kontamine olmuş) çok deęişkenli daęılım modeli.

$$(1 - \varepsilon_1)N_p(\mu_1, \Sigma_1) + \varepsilon_1N_p(\mu_2, \kappa \Sigma_2) + (1 - \varepsilon_2)N_p(\mu_2, \Sigma_2) + \varepsilon_2N_p(\mu_1, \kappa \Sigma_1) \quad (3.1)$$

Bu senaryoda hasta gruba ait verilerin türetimi bozulmuş çok deęişkenli normal daęılım modeli eşitlik 3.1 kullanılarak gerçekleştirildi. Modelde de görüldüğü gibi iki farklı çok deęişkenli normal daęılımdan türetilen veriler belirli ε_i olasılıkları oranında birleştirilmiştir. İkinci çok deęişkenli normal daęılım kovaryans matrisi κ (kappa) katsayısı kadar büyütülerek verilerin deęişkenliği arttırıldı. Böylelikle, sonuçta elde edilecek verilerin çok deęişkenli normal daęılıma uygunluk göstermemesi sağlandı.

Bu çalışmada hastalara ait türetilecek veriler için kullanılan bozulmuş çok deęişkenli normal daęılım modeli aşğıdaki ortalama vektörleri ve kovaryans matrisleri ile oluşturuldu. μ_s ve Σ_s kullanılarak türetilen çok deęişkenli normal daęılım verilerinin hasta verilerinin %30'unu oluşturmasına ve bu daęılımın dięer daęılım tarafından bozulma yüzdesinin $\varepsilon_1 = 0.05$ olmasına karar verildi. μ_h ve Σ_h kullanılarak türetilen çok deęişkenli normal daęılım verilerinin ise hasta verilerinin %70'ini oluşturmasına ve bu daęılımın dięer daęılım tarafından bozulma yüzdesinin de $\varepsilon_2 = 0.10$ olmasına karar verildi. Kovaryans matrisinin deęişkenleri büyütecek katsayı $\kappa = 10$ olarak belirlendi.

$$\mu_s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mu_h = \begin{bmatrix} 0.5 \\ 0.8 \\ 1 \\ 1.5 \end{bmatrix}$$

$$\Sigma_s = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix} \quad \Sigma_h = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 \end{bmatrix}$$

Sağlıklı bireylere ait verilerin türetimi ise çok değişkenli normal dağılımdan aşağıdaki ortalama vektörü ile kovaryans matrisi kullanılarak gerçekleştirildi.

$$\mu_s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Sigma_s = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix}$$

3.2- Simülasyon Algoritması

Bu tez çalışmasında birleştirme yöntemlerini karşılaştırmak amacı ile simülasyon algoritması aşağıdaki şekilde oluşturuldu.

1. Seçilen senaryoya göre hasta ve sağlıklı gruplar için ayrı ayrı tanı testlerine ait veriler türetildi.
2. En iyi birleştirme yöntemi doğrusal ayırma, karesel ayırma ve lojistik ayırma analizi yöntemleri kullanılarak modeller kuruldu. Bu modellerden tek değişkenli skor değerler elde edildi.
3. Doğrusal ayırma, karesel ayırma ve lojistik ayırma analizi ile elde edilen skorlar parametrik ve parametrik olmayan ROC eğrisi yöntemlerine ayrı ayrı yerleştirilerek ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeks değerleri elde edildi. Optimum kesim noktası duyarlılık ve özgüllük seçimi Youden indeksin en büyük olduğu durum göz önünde bulundurularak yapıldı.
4. En iyi birleştirme yöntemi için ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeks değerleri yöntemin kendi özgün algoritması kullanılarak hesaplandı. Bu hesaplama sırasında modelin katsayıları b elde edilirken kullanılan t değeri için $t = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$ vektörü kullanıldı.
5. İlk 4 adım farklı örneklem büyüklüklerinde birbirinden bağımsız olarak 1000 defa tekrar edilerek sonuçlar kaydedildi.

6. Elde edilen ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeks değerlerin ortalama ve standart sapmaları her bir yöntem için ayrı ayrı hesaplandı.

7. Bu değerler göz önünde bulundurularak yöntemler karşılaştırılıp değerlendirildi.

3.3- Uygulama Veri Seti

Simülasyon çalışmalarına ek olarak, yöntemlerin karşılaştırılması gerçek uygulama veri seti kullanılarak yapılmıştır. Bu veri seti over kanserli 60 hasta ile 70 sağlıklı kadın bireylere ait CA-125, CA-19.9, CA-15.3 ve CEA olmak üzere 4 farklı medikal tanı test değerlerini içermektedir. Veriler Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Girişimsel Olmayan Klinik Araştırma Etik Kurul'a başvuru gerçekleştirilmiş olup, 09.08.2017 gün ve 805587-21/G-217 sayı numaralı etik kurul onayı alındıktan sonra, "Sağlık Bilimleri Üniversitesi Tepecik Eğitim ve Araştırma Hastanesi" İzmir'den elde edilmiştir. Etik kurul onayına ek-1'de yer verilmiştir.

Bu tez çalışmasında uygulama verisinde kullanılan medikal tanı testleri CA-125, CA-19.9, CA-15.3 ve CEA'dır. Medikal tanı testlerinin açıklaması aşağıdaki gibidir.

CA-125: CA 125 testi, kanda bulunan CA-125 proteininin (kansere antijen 125) miktarını ölçmeye yarar. Yumurtalıklarından ve rahim kanseri hastalığının tedavisini gören hastalarda tümör olup olmadığını tespit etme ve ona göre tedavi uygulama amacı ile kullanılmakta olan bir yöntemdir.

CA-15.3: Kansere antijeni 15-3 normal meme hücreleri tarafından üretilen bir proteindir. Hastanın meme kanseri tedavisine verdiği yanıtı izlemek ve meme kanseri nüksünü izlemektir.

CA-19.9: Pankreas, hepatobiliyer sistem, mide, kolon ve rektum kanserlerinin teşhisi, tedaviye alınan cevabın izlenmesi ve tedavisi tamamlanmış hastalarda meydana gelebilecek nüksün erkenden belirlenmesine yönelik takip çalışmaları sırasında kullanılabilir.

CEA: Özellikle kalın bağırsak kanseri başta olmak üzere bazı kanserlerde seviyesi artan bir moleküldür.

Tablo 3.1- Uygulama Veri Seti

Grup	CA-15.3	CA-19.9	CEA	CA-125	Grup	CA-15.3	CA-19.9	CEA	CA-125	Grup	CA-15.3	CA-19.9	CEA	CA-125
H	110	20	0.7	5963	H	22	25	1.7	477	K	55	25	0.5	14
H	168	20	0.8	5844	H	60	1000	5.8	414	K	12	8	0.4	61
H	137	21	1.4	5172	H	61	24	2.4	373	K	14	5	0.4	84
H	267	2.5	0.9	5000	H	41	12	1.1	357	K	14	1.4	2	42
H	300	7	1.8	4460	H	65	10	2.5	320	K	12	1	0.4	40
H	267	2.5	0.9	4400	H	19	10	0.3	318	K	26	2.5	0.9	33
H	103	26	1.1	4221	H	11	1.1	2.5	309	K	15	13	1	33
H	120	2	1	4250	H	101	2.5	2	307	K	18	23	4.2	33
H	113	80	1.8	4046	H	211	0.83	1.75	290	K	31	15	0.9	32
H	112	52	2.1	2901	H	57	123	1.6	289	K	42	25	2.7	31
H	45	25	2.5	2319	H	27.9	4.02	1.69	282	K	51	24	5.3	27
H	121	5.5	0.2	2241	H	300	7.8	0.6	281	K	24	10	2.1	26
H	106	11	0.5	2163	H	17	5.4	1.96	141	K	25	2	0.4	25
H	106	11	0.5	2163	H	13	25	2.1	133	K	13	2.5	0.7	23
H	113	32	2.5	2136	H	182	24	1.05	126	K	10	2	0.8	22
H	300	1	4	2129	H	27	1000	28	122	K	8	5	7	22
H	76	12	0.8	2110	K	65	9.6	2.63	359	K	5	27	13	20
H	76	12	0.8	2110	K	24	16	1.8	344	K	5	14	0.6	19
H	21	20	2	1962	K	26	54	0.5	335	K	18	17	0.9	19
H	59	1000	4	1815	K	13	16	1	310	K	21	21	0.5	18
H	54	87	4.8	1391	K	15	14	0.4	309	K	22	42	4.6	16
H	10	27	2	1344	K	24	16	0.3	299	K	23	3	1.3	15
H	12	25	0.5	1336	K	28	18	1.43	287	K	26	18	3	14
H	10	5	0.6	1325	K	204	5.9	2.2	286	K	24	5	5.4	12
H	257	4	0.5	1286	K	26	5.2	2.1	284	K	15	10	3.2	11
H	20	48	5.8	1251	K	27	3.2	0.8	200	K	2	22	3.5	11
H	52	45	5.1	1203	K	55	41	8	191	K	14	24	3.1	9
H	20	56	2.3	1122	K	23	1.59	5	188	K	21	5.7	0.7	9
H	43	65	1.9	1047	K	15	26	3.8	177	K	11	48	0.3	9
H	56	121	1.4	995	K	14	25	3.2	163	K	9	27	2.7	9
H	287	437	17.4	893	K	26	9	2.5	38	K	11	55	3	9
H	300	161	1.06	771	K	29	5	2.4	21	K	28.85	23.47	2.49	8.8
H	56	152	2.4	744	K	14	30	4.6	339	K	43	25	3.1	8
H	60	29	0.4	718	K	15	9.3	0.8	9	K	29	31	8.3	7.9
H	409	54	0.3	693	K	33	18	2.05	31	K	31	69	5.2	7.1
H	24	1.5	0.75	691	K	31	12	2.4	50	K	25	23	2.5	5.6
H	45	2.5	1.1	658	K	33	0.6	2.8	97	K	41	45	2.2	5.2
H	24	55	2.5	607	K	17	6	0.3	248	K	60	124	2.5	2.5
H	54	24	1.1	598	K	34	29	2.5	53	K	27	27	0.3	5.1
H	66	59	2	582	K	25	5.9	0.7	43	K	54	63	0.5	2.5
H	156	23	1	567	K	24	5	1.6	37	K	55	58	2.4	3.5
H	27	12	3.7	541	K	59	23	2.9	54	K	40	54	2.3	6
H	26	2.5	3.3	500						K	46	51	1.1	25

Hasta ve sağlıklı kadın bireylere ait ortalama vektörleri ve kovaryans matrisleri aşağıdaki elde edilmiştir.

$$\mu_h = \begin{bmatrix} 100.43 \\ 85.91 \\ 2.54 \\ 1554.92 \end{bmatrix} \quad \mu_s = \begin{bmatrix} 29.04 \\ 21.74 \\ 2.40 \\ 80.40 \end{bmatrix}$$

$$\Sigma_h = \begin{bmatrix} 9361.68 & -1291.26 & -18.43 & 48171.26 \\ -1291.26 & 48748.23 & 605.84 & -49911.2 \\ -18.43 & 605.84 & 16.96 & -1238.45 \\ 48171.26 & -49911.2 & -1238.44 & 2440254 \end{bmatrix}$$

$$\Sigma_s = \begin{bmatrix} 673.61 & 90.1 & -0.61 & 509.27 \\ 90.1 & 437.02 & 7.58 & -358 \\ -0.61 & 7.58 & 4.96 & -19.71 \\ 509.27 & -358 & -19.71 & 12049.04 \end{bmatrix}$$

Çok değişkenli varyans analizi (MANOVA) kullanıldığında, hasta ve sağlıklı grupların ortalama vektörleri arasında anlamlı fark bulundu (Wilks' Lambda=0.5893; $p<0.001$). Box-M testi sonucu ile hasta ve sağlıklı bireylerin kovaryans matrisleri homojen olmadığı gözlemlenmiştir (Ki-Kare=574.05; $p<0.001$).

Medikal tanı testlerine ait verilerin gruplara göre çok değişkenli normal dağılıma uygunluğu Doornik-Hansen testi ile araştırıldı. Hasta grubun (Ki-Kare=921.484; $p<0.001$) ve sağlıklı grubun (Ki-Kare=484.974; $p<0.001$) çok değişkenli normal dağılıma uymadığı belirlendi.

Medikal tanı testlerinin her birinin tek değişkenli olarak gruplar bazında ayrı ayrı normal dağılıma uygunluğu Shapiro Wilk testi ile incelendi. Hasta ve sağlıklı gruplarda tüm medikal tanı testlerinin normal dağılıma uygun olmadığı belirlendi ($p<0.001$).

Tablo 3.2- Shapiro Wilk Normallik Testi

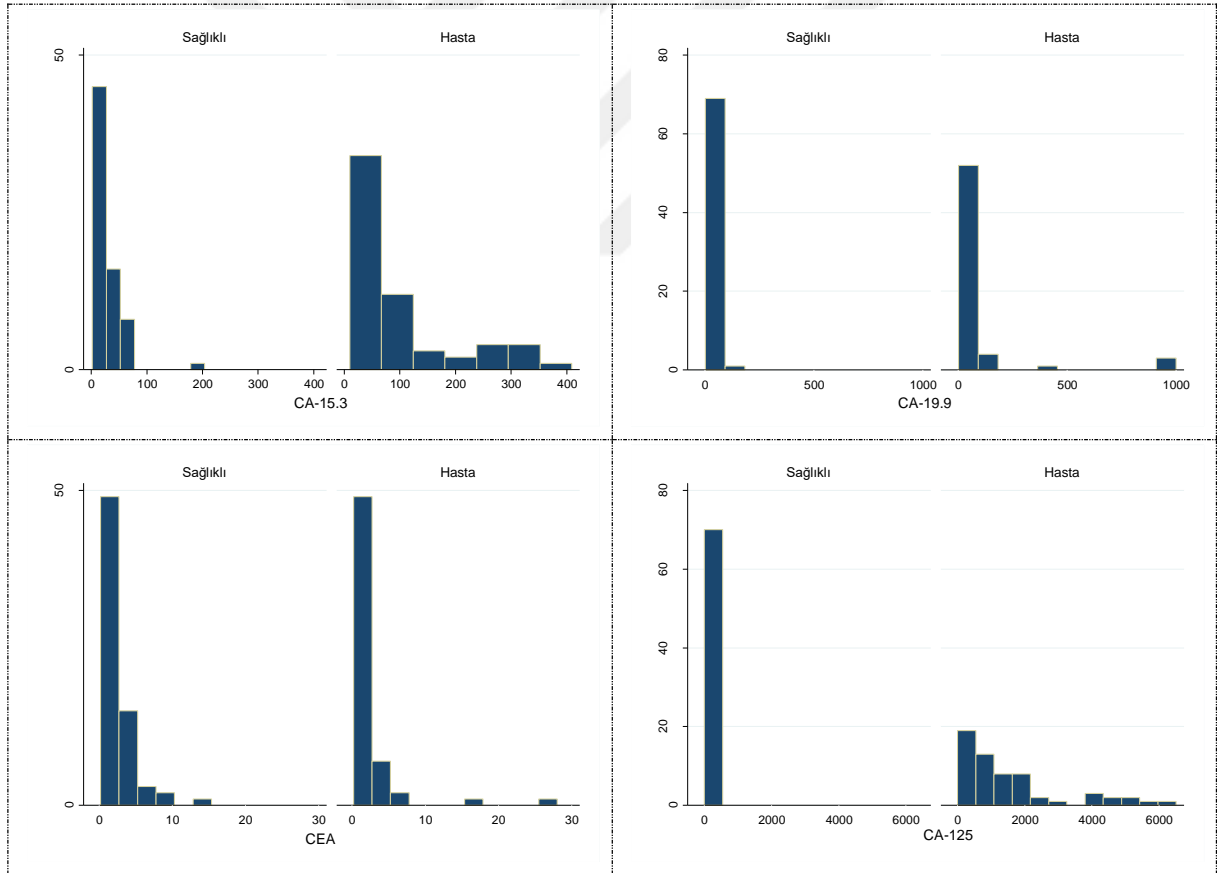
Tanı Testleri	Grup			
	Hasta		Sağlıklı	
	W İstatistiği	p	W İstatistiği	p
CA-15.3	0.603	<0.001	0.810	<0.001
CA-19.9	0.804	<0.001	0.381	<0.001
CEA	0.793	<0.001	0.438	<0.001
CA-125	0.680	<0.001	0.796	<0.001

Tanı testlerinin ikili olarak çok değişkenli normal dağılıma uygunluğunun test sonuçları tablo 3.3' de belirtilmiştir. Bu sonuçlara göre tanı testleri ikili olarak çok değişkenli normal dağılıma uygunluk göstermemektedir ($p<0.001$).

Tablo 3.3- Doornik-Hansen İkili Çok Değişkenli Normallik Testi

Tanı Testleri	Grup	Grup			
		Sağlıklı		Hasta	
		Ki-Kare İstatistiği	p	Ki-Kare İstatistiği	p
CA-15.3	CA-19.9	348.94	<0.001	599.36	<0.001
	CEA	309.45	<0.001	600.79	<0.001
	CA-125	337.72	<0.001	100.32	<0.001
CA-19.9	CEA	102.77	<0.001	847.17	<0.001
	CA-125	167.78	<0.001	596.18	<0.001
CEA	CA-125	171.53	<0.001	587.88	<0.001

Hasta ve sağlıklı grupların her bir medikal tanı testi için histogram grafikleri şekil 3.1’de verildi. Dağılımları incelendiğinde medikal tanı testlerinin sağa çarpık bir yapıda olduğu görüldü.



Şekil 3.1- Uygulama Verisindeki Medikal Tanı Testleri Histogram Grafikleri

3.4- R Programlama Kodları

```
1. #Veri Türetim - Multivariate Normal Dağılım İçin Fonksiyon
2. veritüretim = function(n, muh, mus, sh, ss)
3. {
4.   g1 = t(t(rep(1, n)))
5.   g0 = t(t(rep(0, n)))
6.   g = rbind(g1, g0)
7.   veri_h = rmvnorm(n, mean = muh, sigma = sh)
8.   veri_s = rmvnorm(n, mean = mus, sigma = ss)
9.   x = rbind(veri_h, veri_s)
10. veri = data.frame(g = g, x1 = x[, 1], x2 = x[, 2], x3 = x[, 3], x4 = x[, 4])
11. }
12.
13. #Doğrusal Ayırma Analizi Fonksiyonu - Skor Değerleri Hesaplama
14. d = function(veri, n)
15. {
16.   muh = as.matrix(colMeans(veri[1: n, 2: 5]))
17.   mus = as.matrix(colMeans(veri[(n + 1): (2 * n), 2: 5]))
18.   sh = cov(veri[1: n, 2: 5])
19.   ss = cov(veri[(n + 1): (2 * n), 2: 5])
20.   s = ((n - 1) * sh + (n - 1) * ss) / (2 * n - 2)
21.   b = solve(s) % * % (muh - mus)
22.   b0 = -0.5 * (t(muh) % * % solve(s) % * % muh - t(mus) % * % solve(s) % * % mus)
23.   b00 = rep(b0, 2 * n)
24.   l = as.numeric(b00 + as.matrix(veri[, 2: 5]) % * % b)
25. }
26.
27. #Karesel Ayırma Analizi Fonksiyonu - Skor Değerleri Hesaplama
28. k = function(veri, n)
29. {
30.   muh = as.matrix(colMeans(veri[1: n, 2: 5]))
31.   mus = as.matrix(colMeans(veri[(n + 1): (2 * n), 2: 5]))
32.   sh = cov(veri[1: n, 2: 5])
33.   ss = cov(veri[(n + 1): (2 * n), 2: 5])
34.   omega = -0.5 * (solve(sh) - solve(ss))
35.   b = solve(sh) % * % muh - solve(ss) % * % mus
36.   b0 = -
37.     0.5 * (log(det(sh) / det(ss)) + t(muh) % * % solve(sh) % * % muh - t(mus) % * % solve(ss) %
38.       * % mus)
39.   q = 0
40.   for (i in 1: (2 * n)) {
41.     x = as.matrix(veri[i, 2: 5]) q[i] = b0 + x % * % b + x % * % omega % * % t(x)
42.   }
43.   q
44. }
45. #Lojistik Regresyon - Skor Değerleri Hesaplama
46. l = function(veri)
47. {
48.   lr = glm(g~x1 + x2 + x3 + x4, data = veri, family = "binomial")
49.   logit = predict(lr, type = c("link"))
50. }
51. #Doğrusal, Karesel ve Lojistik Yöntemler İle Elde Edilen Skor Değerleri ile Parametrik ROC
52. parROC = function(score)
53. {
54.   a = (mean(score[1: n]) - mean(score[(n + 1): (2 * n)])) / sd(score[1: n])
55.   b = sd(score[(n + 1): (2 * n)]) / sd(score[1: n])
56.   AUC = pnorm(a / sqrt(1 + b ^ 2))
57.   muh = mean(score[1: n])
58.   mus = mean(score[(n + 1): (2 * n)])
59.   sh = sd(score[1: n])
60.   ss = sd(score[(n + 1): (2 * n)])
61.   c = seq(min(mus - 3 * ss, muh - 3 * sh), max(mus + 3 * ss, muh + 3 * sh), length.out = 100)
```

```

61. sp = 1 - pnorm((mus - c) / ss)
62. sn = pnorm((muh - c) / sh)
63. j = (sn + sp - 1)
64. kriter = data.frame(AUC, c, sn, sp, j)
65. max_kriter = kriter[which.max(kriter$j), ]
66. }
67.
68. #Doğrusal, Karesel ve Lojistik Yöntemler İle Elde Edilen Skor Değerleri ile Parametrik Olma
yan ROC
69.
70. nonparROC = function(score) {
71.   grup = veri$g uygulama = data.frame(grup, score)
72.
73. # Burada c 'ler skorlarımızdır.
74. a = calculate_roc(score, grup, ci = FALSE, alpha = 0.05)
75. j = (a$TPF - a$FPF)
76. AUC = calc_auc(ggplot(uygulama, aes(m = score, d = grup)) + geom_roc()) $AUC
77. se_AUC = (sqrt(var(roc(grup, score))))
78. non_kriter = data.frame(AUC, a, j, se_AUC)
79. maxnon_kriter = non_kriter[which.max(non_kriter$j), ]
80. }
81. #En İyi Doğrusal Birleştirme Modeli
82. mm = function(veri)
83. {
84. muh = as.matrix(colMeans(veri[1: n, 2: 5]))
85. mus = as.matrix(colMeans(veri[(n + 1): (2 * n), 2: 5]))
86. sh = cov(veri[1: n, 2: 5])
87. ss = cov(veri[(n + 1): (2 * n), 2: 5])
88. t = seq(0.1, 0.9, 0.1)
89. c = 0;
90. AUC = 0;
91. sn = 0;
92. sp = 0;
93. j = 0
94.
95. for (i in 1: length(t))
96. {
97. b = solve(t[i] * sh + (1 - t[i]) * ss) % * % (muh - mus)
98. cpay = t(b) % * % muh * sqrt(t(b) % * % ss % * % b) + t(b) % * % mus * sqrt(t(b) % * % sh %
* % b)
99. cpayda = sqrt(t(b) % * % sh % * % b) + sqrt(t(b) % * % ss % * % b)
100.   c[i] = cpay / cpayda
101.   AUC[i] = pnorm((t(b) % * % (muh - mus)) / sqrt(t(b) % * % (ss + sh) % * % b))
102.   sn[i] = pnorm((t(b) % * % muh - c[i]) / sqrt(t(b) % * % sh % * % b))
103.   sp[i] = pnorm((c[i] - t(b) % * % mus) / sqrt(t(b) % * % ss % * % b))
104.   j[i] = (sn[i] + sp[i] - 1)
105. }
106.   mm_kriter = data.frame(AUC, c, sn, sp, j) maxmm_kriter = mm_kriter[which.max(mm_krit
er$j), ]
107. }
108. #Simülasyon Fonksiyonu Multivariate Normal İçin
109. sim = function(n, mus, muh, ss, sh, s)
110. {
111.   parKriter_d = matrix(c(0), s, 5)
112.   parKriter_k = matrix(c(0), s, 5)
113.   parKriter_l = matrix(c(0), s, 5)
114.   nonparKriter_d = matrix(c(0), s, 6)
115.   nonparKriter_k = matrix(c(0), s, 6)
116.   nonparKriter_l = matrix(c(0), s, 6)
117.   mm_kriter = matrix(c(0), s, 5)
118.
119.   for (i in 1: s)
120.   {
121.     veri = verituret(n, muh, mus, sh, ss)
122.     dscore = d(veri, n)
123.     kscore = k(veri, n)

```

```

124.     lscore = l(veri)
125.     parKriter_d[i, ] = data.matrix(parROC(dscore), rownames.force = NA)
126.     parKriter_k[i, ] = data.matrix(parROC(kscore), rownames.force = NA)
127.     parKriter_l[i, ] = data.matrix(parROC(lscore), rownames.force = NA)
128.     nonparKriter_d[i, ] = data.matrix(nonparROC(dscore), rownames.force = NA)
129.     nonparKriter_k[i, ] = data.matrix(nonparROC(kscore), rownames.force = NA)
130.     nonparKriter_l[i, ] = data.matrix(nonparROC(lscore), rownames.force = NA)
131.     mm_kriter[i, ] = data.matrix(mm(veri), rownames.force = NA)
132.     }
133.     sonuc = list(parKriter_d, parKriter_k, parKriter_l, nonparKriter_d, nonparKriter_k,
nonparKriter_l, mm_kriter)
134.     }
135.
136.     library("mvtnorm")
137.     library("plotROC")
138.     library("ggplot2")
139.     library("pROC")
140.     library("dplyr")
141.     library("haven")
142. # Örnek Hacimleri 50, 100, 250, 500 Olacak Şekilde Simülasyonlar Yapıldı
143.     n = 500
144.     s = 1000
145.     muh = c(0.5, 0.8, 1.0, 1.5)
146.     mus = c(0.5, 0.8, 1.0, 1.5)
147.     mus = c(0, 0, 0, 0)
148.     sh = matrix(c(1, 0.5, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 0.5, 1
), nrow = 4)
149.     ss = matrix(c(1, 0.5, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 0.5, 1
), nrow = 4)
150.     sh = matrix(c(1, 0.8, 0.8, 0.8, 0.8, 1, 0.8, 0.8, 0.8, 0.8, 1, 0.8, 0.8, 0.8, 0.8, 1
), nrow = 4)
151.     ss = matrix(c(1, 0.2, 0.2, 0.2, 0.2, 1, 0.2, 0.2, 0.2, 0.2, 1, 0.2, 0.2, 0.2, 0.2, 1
), nrow = 4)
152.     veri = veritretim(n, muh, mus, sh, ss)
153.     dscore = d(veri, n)
154.     kscore = k(veri, n)
155.     lscore = l(veri)
156.     parKriter_d = parROC(dscore);
157.     parKriter_k = parROC(kscore);
158.     parKriter_l = parROC(lscore)
159.     nonparKriter_d = nonparROC(dscore);
160.     nonparKriter_k = nonparROC(kscore);
161.     nonparKriter_l = nonparROC(lscore)
162.     mm_kriter = mm(veri)
163.     sonuc = sim(n, mus, muh, ss, sh, s)
164.     pardogrusal=as.data.frame(sonuc[1])colnames(pardogrusal)=c("AUC", "c", "Sn", "Sp", "j")
165.     parkaresel=as.data.frame(sonuc[2])colnames(parkaresel)=c("AUC", "c", "Sn", "Sp", "j")
166.     parlojistik=as.data.frame(sonuc[3])colnames(parlojistik)=c("AUC", "c", "Sn", "Sp", "j")
167.     nonpardogrusal=as.data.frame(sonuc[4])colnames(nonpardogrusal)=c("AUC", "1-
Sp", "Sn", "c", "j", "se")
168.     nonparkaresel=as.data.frame(sonuc[5])colnames(nonparkaresel)=c("AUC", "1-
Sp", "Sn", "c", "j", "se")
169.     nonparlojistik=as.data.frame(sonuc[6])colnames(nonparlojistik)=c("AUC", "1-
Sp", "Sn", "c", "j", "se")
170.     mm = as.data.frame(sonuc[7]) colnames(mm) = c("AUC", "c", "Sn", "Sp", "j")
171.     ort_pardogrusal = data.matrix(colMeans(pardogrusal))
172.     ort_parkaresel = data.matrix(colMeans(parkaresel))
173.     ort_parlojistik = data.matrix(colMeans(parlojistik))
174.     ort_nonpardogrusal = data.matrix(colMeans(nonpardogrusal))
175.     ort_nonparkaresel = data.matrix(colMeans(nonparkaresel))
176.     ort_nonparlojistik = data.matrix(colMeans(nonparlojistik))
177.     ort_mm = data.matrix(colMeans(mm))
178.
179.     sd_pardogrusal = data.matrix(apply(pardogrusal, 2, sd))
180.     sd_parkaresel = data.matrix(apply(parkaresel, 2, sd))
181.     sd_parlojistik = data.matrix(apply(parlojistik, 2, sd))

```

```

182. sd_nonpardogrusal = data.matrix(apply(nonpardogrusal, 2, sd))
183. sd_nonparkaresel = data.matrix(apply(nonparkaresel, 2, sd))
184. sd_nonparlojistik = data.matrix(apply(nonparlojistik, 2, sd))
185. sd_mm = data.matrix(apply(mm, 2, sd))
186.
187. ci_pardogrusal = t(data.matrix(apply(pardogrusal, 2, function(x) {
188.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
189. ci_parkaresel = t(data.matrix(apply(parkaresel, 2, function(x) {
190.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
191. ci_parlojistik = t(data.matrix(apply(parlojistik, 2, function(x) {
192.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
193. ci_nonpardogrusal = t(data.matrix(apply(nonpardogrusal, 2, function(x) {
194.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
195. ci_nonparkaresel = t(data.matrix(apply(nonparkaresel, 2, function(x) {
196.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
197. ci_nonparlojistik = t(data.matrix(apply(nonparlojistik, 2, function(x) {
198.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
199. ci_mm = t(data.matrix(apply(mm, 2, function(x) {
200.     mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
201.
202. veri <- read_dta("C:/veri.dta")
203. veri = as.data.frame(veri)
204. n1 = 60
205. n2 = 70
206. b = 1000
207. #Doğrusal Ayırma Analizi Fonksiyonu - Skor Değerleri Hesaplama
208. d_b = function(veri, n1, n2)
209. {
210.     muh = as.matrix(colMeans(veri[1: n1, 2: 5]))
211.     mus = as.matrix(colMeans(veri[(n1 + 1): (n1 + n2), 2: 5]))
212.     sh = cov(veri[1: n1, 2: 5])
213.     ss = cov(veri[(n1 + 1): (n1 + n2), 2: 5])
214.     s = ((n1 - 1) * sh + (n2 - 1) * ss) / ((n1 + n2) - 2)
215.     b = solve(s) % * % (muh - mus)
216.     b0 = -0.5 * (t(muh) % * % solve(s) % * % muh - t(mus) % * % solve(s) % * % mus)
217.     b00 = rep(b0, (n1 + n2)) l = as.numeric(b00 + as.matrix(veri[, 2: 5]) % * % b)
218. }
219.
220. #Karesel Ayırma Analizi Fonksiyonu - Skor Değerleri Hesaplama
221. k_b = function(veri, n1, n2)
222. {
223.     muh = as.matrix(colMeans(veri[1: n1, 2: 5]))
224.     mus = as.matrix(colMeans(veri[(n1 + 1): (n1 + n2), 2: 5]))
225.     sh = cov(veri[1: n1, 2: 5]) ss = cov(veri[(n1 + 1): (n1 + n2), 2: 5])
226.     omega = -
227.     0.5 * (solve(sh) - solve(ss)) b = solve(sh) % * % muh - solve(ss) % * % mus
228.     b0 = -
229.     0.5 * (log(det(sh) / det(ss)) + t(muh) % * % solve(sh) % * % muh - t(mus) % * % solve(ss) %
230.     * % mus)
231.     q = 0
232.     for (i in 1: (n1 + n2))
233.     {
234.         x = as.matrix(veri[i, 2: 5]) q[i] = b0 + x % * % b + x % * % omega % * % t(x)
235.     }
236.
237. #Lojistik Regresyon - Skor Değerleri Hesaplama
238. l_b = function(veri)
239. {
240.     lr = glm(g~x1 + x2 + x3 + x4, data = veri, family = "binomial")
241.     logit = predict(lr, type = c("link"))
242. }
243.

```



```

244.      #Doğrusal, Karesel ve Lojistik Yöntemler İle Elde Edilen Skor Değerleri ile Parametrik ROC
      ik ROC
245.      parROC_b = function(score, n1, n2)
246.      {
247.      a = (mean(score[1: n1]) - mean(score[(n1 + 1): (n1 + n2)])) / sd(score[1: n1])
248.      b = sd(score[(n1 + 1): (n1 + n2)]) / sd(score[1: n1])
249.      AUC = pnorm(a / sqrt(1 + b ^ 2))
250.      muh = mean(score[1: n1])
251.      mus = mean(score[(n1 + 1): (n1 + n2)])
252.      sh = sd(score[1: n1])
253.      ss = sd(score[(n1 + 1): (n1 + n2)])
254.      c = seq(min(mus - 3 * ss, muh - 3 * sh), max(mus + 3 * ss, muh + 3 * sh), length.out
      = 1000)
255.      sp = 1 - pnorm((mus - c) / ss)
256.      sn = pnorm((muh - c) / sh) j = (sn + sp - 1)
257.      kriter = data.frame(AUC, c, sn, sp, j) max_kriter = kriter[which.max(kriter$j), ]
258.      }
259.
260. #Doğrusal, Karesel ve Lojistik Yöntemler İle Elde Edilen Skor Değerleri ile Parametrik Olmayan ROC
261.      nonparROC_b = function(score, n1, n2)
262.      {
263.      grup = veri$g uygulama = data.frame(grup, score) # Burada c 'ler skorlarımızdır.
264.      a = calculate_roc(score, grup, ci = FALSE, alpha = 0.05)
265.      j = (a$TPF - a$FPF)
266.      AUC = calc_auc(ggplot(uygulama, aes(m = score, d = grup)) + geom_roc()) $AUC
267.      se_AUC = (sqrt(var(roc(grup, score))))
268.      non_kriter = data.frame(AUC, a, j, se_AUC)
269.      maxnon_kriter = non_kriter[which.max(non_kriter$j), ]
270.      }
271. #En İyi Doğrusal Birleştirme Modeli
272.      mm_b = function(veri, n1, n2)
273.      {
274.      muh = as.matrix(colMeans(veri[1: n1, 2: 5]))
275.      mus = as.matrix(colMeans(veri[(n1 + 1): (n1 + n2), 2: 5]))
276.      sh = cov(veri[1: n1, 2: 5])
277.      ss = cov(veri[(n1 + 1): (n1 + n2), 2: 5])
278.      t = seq(0.1, 0.9, 0.1)
279.      c = 0;
280.      AUC = 0;
281.      sn = 0;
282.      sp = 0;
283.      j = 0
284.
285.      for (i in 1: length(t))
286.      {
287.      b = solve(t[i] * sh + (1 - t[i]) * ss) % * % (muh - mus)
288.      cpay = t(b) % * % muh * sqrt(t(b) % * % ss % * % b) + t(b) % * % mus * sqrt(t(b) % *
      % sh % * % b)
289.      cpayda = sqrt(t(b) % * % sh % * % b) + sqrt(t(b) % * % ss % * % b)
290.      c[i] = cpay / cpayda AUC[i] = pnorm((t(b) % * % (muh - mus)) / sqrt(t(b) % * % (ss +
      sh) % * % b))
291.      sn[i] = pnorm((t(b) % * % muh - c[i]) / sqrt(t(b) % * % sh % * % b))
292.      sp[i] = pnorm((c[i] - t(b) % * % mus) / sqrt(t(b) % * % ss % * % b))
293.      j[i] = (sn[i] + sp[i] - 1)
294.      }
295.      mm_kriter = data.frame(AUC, c, sn, sp, j) maxmm_kriter = mm_kriter[which.max(mm_kriter$j), ]
296.      }
297.      sonuc_boot <- function(veri, b)
298.      {
299.      n1 = 60 n2 = 70
300.      vh <- veri[1: n1, -1]
301.      vs <- veri[(n1 + 1): (n1 + n2), -1]
302.      parKriter_d = matrix(c(0), b, 5)
303.      parKriter_k = matrix(c(0), b, 5)

```

```

304.     parKriter_l = matrix(c(0), b, 5)
305.     nonparKriter_d = matrix(c(0), b, 6)
306.     nonparKriter_k = matrix(c(0), b, 6)
307.     nonparKriter_l = matrix(c(0), b, 6)
308.     mm_kriter = matrix(c(0), b, 5)
309.     for (i in 1: b) {
310.         vhboot <- data.frame(g = rep(1, n1), sample_n(vh, size = n1, replace = TRUE))
311.         vsboot <- data.frame(g = rep(0, n2), sample_n(vs, size = n2, replace = TRUE))
312.         veriboot = rbind(vhboot, vsboot)
313.         dscore = d_b(veriboot, n1, n2)
314.         ksore = k_b(veriboot, n1, n2)
315.         lsore = l_b(veriboot)
316.         parKriter_d[i, ] = data.matrix(parROC_b(dscore, n1, n2), rownames.force = NA)
317.         parKriter_k[i, ] = data.matrix(parROC_b(ksore, n1, n2), rownames.force = NA)
318.         parKriter_l[i, ] = data.matrix(parROC_b(lsore, n1, n2), rownames.force = NA)
319.         nonparKriter_d[i, ] = data.matrix(nonparROC_b(dscore, n1, n2), rownames.force = NA)
320.         nonparKriter_k[i, ] = data.matrix(nonparROC_b(ksore, n1, n2), rownames.force = NA)
321.         nonparKriter_l[i, ] = data.matrix(nonparROC_b(lsore, n1, n2), rownames.force = NA)
322.         mm_kriter[i, ] = data.matrix(mm_b(veriboot, n1, n2), rownames.force = NA)
323.     }
324.     sonuc = list(parKriter_d, parKriter_k, parKriter_l, nonparKriter_d, nonparKriter_k,
nonparKriter_l, mm_kriter)
325.     }
326.     sonuc = sonuc_boot(veri, b)
327.     pardogrusal=as.data.frame(sonuc[1])colnames(pardogrusal)=c("AUC", "c", "Sn", "Sp", "j")
328.     parkaresel=as.data.frame(sonuc[2])colnames(parkaresel)=c("AUC", "c", "Sn", "Sp", "j")
329.     parlojistik=as.data.frame(sonuc[3])colnames(parlojistik)=c("AUC", "c", "Sn", "Sp", "j")
330.     nonpardogrusal=as.data.frame(sonuc[4])colnames(nonpardogrusal)=c("AUC", "1-
Sp", "Sn", "c", "j", "se")
331.     nonparkaresel=as.data.frame(sonuc[5])colnames(nonparkaresel)=c("AUC", "1-
Sp", "Sn", "c", "j", "se")
332.     nonparlojistik=as.data.frame(sonuc[6])colnames(nonparlojistik)=c("AUC", "1-
Sp", "Sn", "c", "j", "se")
333.     mm = as.data.frame(sonuc[7])
334.     colnames(mm) = c("AUC", "c", "Sn", "Sp", "j")
335.
336.     ort_pardogrusal = data.matrix(colMeans(pardogrusal))
337.     ort_parkaresel = data.matrix(colMeans(parkaresel))
338.     ort_parlojistik = data.matrix(colMeans(parlojistik))
339.     ort_nonpardogrusal = data.matrix(colMeans(nonpardogrusal))
340.     ort_nonparkaresel = data.matrix(colMeans(nonparkaresel))
341.     ort_nonparlojistik = data.matrix(colMeans(nonparlojistik))
342.     ort_mm = data.matrix(colMeans(mm))
343.     sd_pardogrusal = data.matrix(apply(pardogrusal, 2, sd))
344.     sd_parkaresel = data.matrix(apply(parkaresel, 2, sd))
345.     sd_parlojistik = data.matrix(apply(parlojistik, 2, sd))
346.     sd_nonpardogrusal = data.matrix(apply(nonpardogrusal, 2, sd))
347.     sd_nonparkaresel = data.matrix(apply(nonparkaresel, 2, sd))
348.     sd_nonparlojistik = data.matrix(apply(nonparlojistik, 2, sd))
349.     sd_mm = data.matrix(apply(mm, 2, sd))
350.     ci_pardogrusal = t(data.matrix(apply(pardogrusal, 2, function(x) {
351.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
352.     ci_parkaresel = t(data.matrix(apply(parkaresel, 2, function(x) {
353.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
354.     ci_parlojistik = t(data.matrix(apply(parlojistik, 2, function(x) {
355.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
356.     ci_nonpardogrusal = t(data.matrix(apply(nonpardogrusal, 2, function(x) {
357.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
358.     ci_nonparkaresel = t(data.matrix(apply(nonparkaresel, 2, function(x) {
359.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
360.     ci_nonparlojistik = t(data.matrix(apply(nonparlojistik, 2, function(x) {
361.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))
362.     ci_mm = t(data.matrix(apply(mm, 2, function(x) {
363.         mean(x) + c(-1.96, 1.96) * sd(x) / sqrt(length(x))})))

```

4- BULGULAR

Yapılan simülasyonlar sonucunda her bir yöntemden elde edilen ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeks değerleri kullanılarak değerlendirilmeler yapıldı. Elde edilen sonuçlar her bir senaryo için ayrı tablolarda verildi. Her bir tabloda farklı örnek hacimleri için ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeks değerlerinin ortalamaları ve standart hataları sırası ile gösterildi. Daha sonra uygulama veri seti kullanılarak yöntemlerin sonuçları değerlendirildi.

4.1- Simülasyon Çalışması Bulguları

Ortalama vektörlerinin ve kovaryans matrislerinin ortak olduğu senaryo I kullanılarak türetilen veri setlerinden elde edilen simülasyon sonuçları Tablo 4.1'de verildi.

Tüm yöntemlerde ROC eğrisi altında kalan alan, özgüllük ve duyarlılık değerlerinin oldukça düşük olduğu gözlemlendi. Hatta örneklem büyüklüğündeki artış sonucunda değerlerin daha da düştüğü gözlemlendi. Parametrik ve parametrik olmayan ROC yöntemlerinden elde edilen sonuçların birbirlerine oldukça yakın sonuçlar verdiği görüldü.

Tablo 4.2'de kovaryans matrislerinin heterojen ancak ortalama vektörlerinin eşit olduğu senaryo II'ye ait sonuçlar yer almaktadır. Tüm sonuçlar incelendiğinde parametrik ve parametrik olmayan ROC eğrisi kullanılarak karesel ayırma analizi yöntemi ile elde edilen ROC eğrisi altında kalan alan, duyarlılık ve özgüllük değerleri diğer yöntemlere göre daha yüksek sonuçlar verdiği görüldü. Doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizlerinin parametrik ve parametrik olmayan değerleri birbirine yakın sonuçlar vermektedir. En iyi birleştirme yöntemiyle parametrik ROC eğrisi ile elde edilen yöntemlerin sonuçlarına yakın sonuçlar verdiği görüldü.

Senaryo III kullanılarak veri türetimi yapılan simülasyon çalışması sonuçları tablo 4.3'de verildi. Kovaryans matrislerinin ortak olduğu ancak ortalama vektörlerinin birbirinden farklı olduğu durumda tüm yöntemlerin eğri altında kalan alan, duyarlılık ve özgüllük değerleri tablo 4.1 ve tablo 4.2'ye göre daha yüksek olduğu bulundu. Doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizlerinin parametrik ve parametrik olmayan ROC eğrisi yöntemi ile elde edilen sonuçlar birbirine oldukça yakın

olduđu bulundu. En iyi dođrusal birleřtirme yntemi de bu yntemler ile olduka yakın sonular verdiđi bulundu. rneklem byklklerdeki artıř, ROC eđrisi altında kalan alan, duyarlılık ve zgllk deđerlerinde deđiřim yaratmadıđı gzlendi.

Hem ortalama vektrlerinin hem de kovaryans matrislerinin birbirine eřit olmadıđı senaryo IV'e gre tretilen veri setlerine iliřkin sonular tablo 4.4'de verildi. Tm yntemler yksek ROC eđrisi altında kalan alan, duyarlılık ve zgllk deđerleri sergilemektedir ancak bunlar ierisinde karesel ayırma analizinden elde edilen skorların parametrik olmayan ROC eđrisi yntemine yerleřtirilmesi ile oluřturulan ynteme ait sonuların her bir rneklem byklđnde daha stn olduđu grld.

Hasta verilerinin tretiminde bozulmuř ok deđerkenli normal dađılım verilerinin kullanıldıđı senaryo V'e gre elde edilen sonular tablo 4.5'de verilmektedir. Parametrik olmayan ROC eđrisi yntemi kullanılarak dođrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizinden elde edilen ROC eđrisi altında kalan alan, duyarlılık ve zgllk deđerleri parametrik ROC eđrisi ynteminden elde edilen sonulara daha yksek olduđu bulundu. En iyi dođrusal birleřtirme yntemi de parametrik ROC eđrisi ynteminden elde edilen sonulara yakın sonular verdiđi grld. Tm yntemler gz nnde bulundurulduđunda her bir rneklem byklđ iin parametrik olmayan ROC eđrisi yntemi ile elde edilen karesel ayırma analizi ynteminin sonuları belirgin bir Őekilde diđer yntemlerden daha yksek ROC eđrisi altında kalan alan, duyarlılık ve zgllk sonuları verdiđi gzlendi.

Tablo 4.1- Senaryo I – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Aynı, Kovaryans Matrislerinin Aynı Olduğu Durum

			<i>AUC</i>	<i>SH(AUC)</i>	<i>c</i>	<i>SH(c)</i>	<i>S_n</i>	<i>SH(S_n)</i>	<i>S_p</i>	<i>SH(S_p)</i>	<i>j max</i>	<i>SH(j max)</i>
n=50	P	Doğrusal	0.6096	0.0400	-0.0014	0.1054	0.5870	0.1265	0.5793	0.1286	0.1664	0.0562
		Karesel	0.6942	0.0336	-0.0141	0.2786	0.6635	0.1351	0.6505	0.1353	0.3140	0.0604
		Lojistik	0.6096	0.0400	-0.0015	0.1087	0.5871	0.1267	0.5793	0.1288	0.1664	0.0563
		MM	0.6096	0.0400	-0.0038	0.3597	0.5784	0.0291	0.5784	0.0291	0.1569	0.0582
	NP	Doğrusal	0.6104	0.0424	0.9804	0.1901	0.6143	0.1801	0.6280	0.1818	0.7577	0.0653
		Karesel	0.6924	0.0405	0.0354	0.2480	0.6587	0.1527	0.6825	0.1501	0.3412	0.0663
		Lojistik	0.6105	0.0424	0.0190	0.1937	0.6147	0.1802	0.6275	0.1814	0.2422	0.0654
		MM	0.6105	0.0424	0.0190	0.1937	0.6147	0.1802	0.6275	0.1814	0.2422	0.0654
n=100	P	Doğrusal	0.5780	0.0280	-0.0009	0.0732	0.5607	0.1186	0.5560	0.1195	0.1167	0.0393
		Karesel	0.6426	0.0254	-0.0053	0.1961	0.6171	0.1402	0.6096	0.1397	0.2267	0.0449
		Lojistik	0.5780	0.0280	-0.0009	0.0741	0.5607	0.1186	0.5560	0.1196	0.1167	0.0393
		MM	0.5780	0.0280	-0.0022	0.2427	0.5554	0.0200	0.5554	0.0200	0.1109	0.0401
	NP	Doğrusal	0.5784	0.0290	0.9847	0.1343	0.5711	0.1770	0.6022	0.1762	0.8266	0.0457
		Karesel	0.6369	0.0299	0.0180	0.1810	0.6132	0.1604	0.6299	0.1579	0.2432	0.0478
		Lojistik	0.5784	0.0290	0.0149	0.1352	0.5719	0.1763	0.6015	0.1757	0.1734	0.0455
		MM	0.5784	0.0290	0.0149	0.1352	0.5719	0.1763	0.6015	0.1757	0.1734	0.0455
n=250	P	Doğrusal	0.5475	0.0174	-0.0001	0.0487	0.5372	0.1269	0.5343	0.1275	0.0715	0.0242
		Karesel	0.5909	0.0173	0.0015	0.1253	0.5701	0.1427	0.5731	0.1435	0.1432	0.0304
		Lojistik	0.5475	0.0174	-0.0001	0.0489	0.5372	0.1269	0.5343	0.1275	0.0715	0.0242
		MM	0.5475	0.0174	-0.0064	0.1438	0.5336	0.0124	0.5336	0.0124	0.0673	0.0247
	NP	Doğrusal	0.5475	0.0184	0.9986	0.0838	0.5536	0.1832	0.5542	0.1835	0.8922	0.0281
		Karesel	0.5861	0.0195	0.0067	0.1096	0.5676	0.1660	0.5861	0.1644	0.1537	0.0313
		Lojistik	0.5475	0.0184	0.0013	0.0846	0.5537	0.1837	0.5541	0.1839	0.1078	0.0281
		MM	0.5475	0.0184	0.0013	0.0846	0.5537	0.1837	0.5541	0.1839	0.1078	0.0281
n=500	P	Doğrusal	0.5337	0.0121	0.0024	0.0352	0.5166	0.1276	0.5341	0.1276	0.0507	0.0167
		Karesel	0.5654	0.0123	0.0030	0.0906	0.5454	0.1462	0.5574	0.1457	0.1027	0.0215
		Lojistik	0.5337	0.0121	0.0024	0.0353	0.5166	0.1276	0.5341	0.1276	0.0507	0.0167
		MM	0.5337	0.0121	-0.0005	0.1037	0.5238	0.0086	0.5238	0.0086	0.0477	0.0171
	NP	Doğrusal	0.5337	0.0126	0.9936	0.0611	0.5222	0.1884	0.5546	0.1887	0.9232	0.0198
		Karesel	0.5616	0.0137	0.0001	0.0755	0.5512	0.1607	0.5594	0.1614	0.1106	0.0216
		Lojistik	0.5337	0.0126	0.0062	0.0614	0.5227	0.1886	0.5541	0.1889	0.0768	0.0198
		MM	0.5337	0.0126	0.0062	0.0614	0.5227	0.1886	0.5541	0.1889	0.0768	0.0198

AUC: Eğri Altında Kalan Alan Sn: Duyarlılık Sp: Özgüllük c: Optimum Kesim Noktası J: Youden İndeksi p: Parametrik ROC Eğrisi Analizi Yöntemi
NP: Parametrik Olmayan ROC Eğrisi Yöntemi mm: MiniMax Prosedürü

Tablo 4.2- Senaryo II – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Aynı, Kovaryans Matrislerinin Farklı Olduğu Durum

			<i>AUC</i>	<i>SH(AUC)</i>	<i>c</i>	<i>SH(c)</i>	<i>S_n</i>	<i>SH(S_n)</i>	<i>S_p</i>	<i>SH(S_p)</i>	<i>j max</i>	<i>SH(j max)</i>
n=50	P	Doğrusal	0.6058	0.0386	-0.1411	0.1186	0.7725	0.1535	0.4426	0.1319	0.2151	0.0651
		Karesel	0.8322	0.0320	-0.8185	0.2436	0.9416	0.0340	0.7024	0.0404	0.6440	0.0575
		Lojistik	0.6058	0.0386	-0.1474	0.1245	0.7732	0.1538	0.4424	0.1321	0.2157	0.0654
		MM	0.6056	0.0385	0.0117	0.3760	0.5776	0.0288	0.5776	0.0288	0.1553	0.0575
	NP	Doğrusal	0.6055	0.0412	1.1108	0.1653	0.7747	0.1722	0.5006	0.1612	0.7247	0.0672
		Karesel	0.8893	0.0322	0.1051	0.4502	0.8827	0.0713	0.7871	0.0795	0.6698	0.0679
		Lojistik	0.6055	0.0412	-0.1168	0.1713	0.7754	0.1728	0.5004	0.1615	0.2758	0.0669
		MM	0.6055	0.0412	-0.1168	0.1713	0.7754	0.1728	0.5004	0.1615	0.2758	0.0669
n=100	P	Doğrusal	0.5746	0.0265	-0.1164	0.1034	0.7734	0.1720	0.4040	0.1472	0.1773	0.0535
		Karesel	0.8301	0.0229	-0.7969	0.1516	0.9416	0.0215	0.6929	0.0302	0.6345	0.0423
		Lojistik	0.5746	0.0265	-0.1186	0.1056	0.7737	0.1722	0.4039	0.1474	0.1776	0.0536
		MM	0.5744	0.0265	-0.0076	0.2507	0.5544	0.0195	0.5544	0.0195	0.1088	0.0389
	NP	Doğrusal	0.5740	0.0288	1.1019	0.1302	0.7720	0.1882	0.5545	0.1711	0.7825	0.0550
		Karesel	0.8823	0.0241	0.0477	0.3552	0.8754	0.0588	0.7655	0.0631	0.6410	0.0522
		Lojistik	0.5740	0.0288	-0.1040	0.1322	0.7722	0.1880	0.5544	0.1712	0.2177	0.0550
		MM	0.5740	0.0288	-0.1040	0.1322	0.7722	0.1880	0.5544	0.1712	0.2177	0.0550
n=250	P	Doğrusal	0.5470	0.0172	-0.0884	0.0741	0.7882	0.1772	0.3584	0.1509	0.1466	0.0512
		Karesel	0.8288	0.0149	-0.7783	0.0871	0.9421	0.0128	0.6884	0.0194	0.6305	0.0268
		Lojistik	0.5470	0.0172	-0.0891	0.0747	0.7883	0.1773	0.3584	0.1510	0.1467	0.0512
		MM	0.5469	0.0172	-0.0025	0.1582	0.5342	0.0126	0.5342	0.0126	0.0685	0.0252
	NP	Doğrusal	0.5472	0.0185	1.0812	0.0848	0.7825	0.1890	0.6103	0.1656	0.8278	0.0502
		Karesel	0.8768	0.0152	0.0191	0.2846	0.8698	0.0461	0.7471	0.0489	0.6169	0.0339
		Lojistik	0.5472	0.0185	-0.0818	0.0861	0.7827	0.1895	0.6103	0.1662	0.1724	0.0503
		MM	0.5472	0.0185	-0.0818	0.0861	0.7827	0.1895	0.6103	0.1662	0.1724	0.0503
n=500	P	Doğrusal	0.5333	0.0121	-0.0689	0.0553	0.7928	0.1821	0.3369	0.1541	0.1297	0.0509
		Karesel	0.8280	0.0102	-0.7686	0.0604	0.9416	0.0088	0.6858	0.0133	0.6274	0.0184
		Lojistik	0.5333	0.0121	-0.0691	0.0555	0.7929	0.1820	0.3369	0.1541	0.1298	0.0509
		MM	0.5333	0.0121	0.0014	0.1093	0.5242	0.0088	0.5242	0.0088	0.0484	0.0177
	NP	Doğrusal	0.5333	0.0132	1.0651	0.0608	0.7878	0.1879	0.6417	0.1632	0.8539	0.0487
		Karesel	0.8745	0.0107	0.0006	0.2260	0.8673	0.0371	0.7378	0.0388	0.6051	0.0235
		Lojistik	0.5333	0.0132	-0.0652	0.0607	0.7879	0.1874	0.6417	0.1627	0.1461	0.0487
		MM	0.5333	0.0132	-0.0652	0.0607	0.7879	0.1874	0.6417	0.1627	0.1461	0.0487

AUC: Eğri Altında Kalan Alan S_n : Duyarlılık S_p : Özgüllük c : Optimum Kesim Noktası J : Youden İndeksi p : Parametrik ROC Eğrisi Analizi Yöntemi
 NP: Parametrik Olmayan ROC Eğrisi Yöntemi mm: MiniMax Prosedürü

Tablo 4.3- Senaryo III – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Farklı, Kovaryans Matrislerinin Aynı Olduğu Durum

			<i>AUC</i>	<i>SH(AUC)</i>	<i>c</i>	<i>SH(c)</i>	<i>S_n</i>	<i>SH(S_n)</i>	<i>S_p</i>	<i>SH(S_p)</i>	<i>j max</i>	<i>SH(j max)</i>
n=50	P	Doğrusal	0.8761	0.0320	0.0003	0.0463	0.7956	0.0392	0.7964	0.0412	0.5920	0.0632
		Karesel	0.8776	0.0307	0.0036	0.2953	0.8053	0.0627	0.8069	0.0654	0.6122	0.0595
		Lojistik	0.8756	0.0318	0.0008	0.1041	0.7951	0.0394	0.7960	0.0413	0.5911	0.0628
		MM	0.8761	0.0320	0.3205	0.3282	0.7951	0.0318	0.7951	0.0318	0.5903	0.0635
	NP	Doğrusal	0.8774	0.0325	0.9277	0.4424	0.8135	0.0849	0.8257	0.0813	0.3608	0.0701
		Karesel	0.8910	0.0303	0.1035	0.4453	0.8190	0.0812	0.8421	0.0758	0.6611	0.0674
		Lojistik	0.8780	0.0327	0.0779	0.4317	0.8134	0.0835	0.8278	0.0805	0.6412	0.0707
		MM	0.8730	0.0238	-0.0002	0.0339	0.7916	0.0292	0.7915	0.0295	0.5831	0.0468
n=100	P	Karesel	0.8738	0.0228	0.0012	0.2109	0.7968	0.0505	0.7975	0.0500	0.5943	0.0445
		Lojistik	0.8728	0.0237	0.0009	0.0717	0.7914	0.0291	0.7913	0.0295	0.5827	0.0467
		MM	0.8730	0.0238	1.2971	0.2249	0.7911	0.0235	0.7911	0.0235	0.5823	0.0470
		Doğrusal	0.8739	0.0242	0.9423	0.3548	0.7999	0.0697	0.8161	0.0662	0.3840	0.0523
	NP	Karesel	0.8808	0.0228	0.0550	0.3635	0.8064	0.0687	0.8201	0.0666	0.6264	0.0513
		Lojistik	0.8741	0.0242	0.0557	0.3617	0.8009	0.0707	0.8157	0.0680	0.6166	0.0525
		Doğrusal	0.8703	0.0147	-0.0002	0.0232	0.7882	0.0187	0.7879	0.0184	0.5761	0.0286
		Karesel	0.8705	0.0144	-0.0020	0.1469	0.7909	0.0362	0.7902	0.0353	0.5811	0.0284
n=250	P	Lojistik	0.8703	0.0147	-0.0008	0.0467	0.7881	0.0187	0.7878	0.0184	0.5759	0.0286
		MM	0.8703	0.0147	1.2671	0.1351	0.7879	0.0143	0.7879	0.0143	0.5757	0.0286
		Doğrusal	0.8705	0.0150	0.9726	0.2585	0.7938	0.0507	0.8012	0.0478	0.4051	0.0331
		Karesel	0.8733	0.0148	0.0316	0.2657	0.7950	0.0516	0.8044	0.0496	0.5994	0.0332
	NP	Lojistik	0.8706	0.0150	0.0261	0.2582	0.7942	0.0504	0.8009	0.0485	0.5951	0.0330
		Doğrusal	0.8693	0.0111	0.0000	0.0162	0.7867	0.0133	0.7869	0.0134	0.5736	0.0215
		Karesel	0.8694	0.0110	0.0021	0.0997	0.7875	0.0251	0.7883	0.0252	0.5758	0.0213
		Lojistik	0.8693	0.0111	0.0008	0.0308	0.7866	0.0133	0.7869	0.0134	0.5735	0.0215
n=500	P	MM	0.8693	0.0111	1.2568	0.0951	0.7867	0.0107	0.7867	0.0107	0.5734	0.0215
		Doğrusal	0.8694	0.0111	0.9824	0.2089	0.7904	0.0401	0.7964	0.0386	0.4132	0.0244
		Karesel	0.8708	0.0110	0.0122	0.2136	0.7924	0.0405	0.7964	0.0397	0.5889	0.0241
		Lojistik	0.8695	0.0111	0.0209	0.2080	0.7899	0.0407	0.7970	0.0388	0.5870	0.0244

AUC: Eğri Altında Kalan Alan Sn: Duyarlılık Sp: Özgüllük c: Optimum Kesim Noktası J: Youden İndeksi p: Parametrik ROC Eğrisi Analizi Yöntemi
NP: Parametrik Olmayan ROC Eğrisi Yöntemi mm: MiniMax Prosedürü

Tablo 4.4- Senaryo IV – Hasta ve Sağlıklı Toplumun Medikal Tanı Testi Değerleri Ortalama Vektörleri Farklı, Kovaryans Matrislerinin Farklı Olduğu Durum

			<i>AUC</i>	<i>SH(AUC)</i>	<i>c</i>	<i>SH(c)</i>	<i>S_n</i>	<i>SH(S_n)</i>	<i>S_p</i>	<i>SH(S_p)</i>	<i>j max</i>	<i>SH(j max)</i>
n=50	P	Doğrusal	0.8773	0.0326	-0.0154	0.0500	0.8084	0.0399	0.7873	0.0439	0.5957	0.0645
		Karesel	0.8946	0.0272	-0.9073	0.3741	0.9207	0.0373	0.7750	0.0396	0.6956	0.0497
		Lojistik	0.8767	0.0324	-0.0477	0.1237	0.8099	0.0421	0.7856	0.0454	0.5955	0.0640
		MM	0.8770	0.0327	1.3713	0.3207	0.7968	0.0325	0.7968	0.0325	0.5936	0.0650
	NP	Doğrusal	0.8788	0.0334	0.9175	0.4327	0.8208	0.0830	0.8217	0.0809	0.3574	0.0702
		Karesel	0.9533	0.0199	0.1526	0.5287	0.9151	0.0535	0.8840	0.0600	0.7991	0.0583
		Lojistik	0.8794	0.0338	0.0677	0.4342	0.8203	0.0831	0.8239	0.0813	0.6443	0.0716
		MM	0.8733	0.0233	-0.0172	0.0401	0.8020	0.0299	0.7826	0.0317	0.5846	0.0454
n=100	P	Karesel	0.8949	0.0193	-0.8701	0.2343	0.9162	0.0280	0.7689	0.0278	0.6852	0.0363
		Lojistik	0.8730	0.0232	-0.0419	0.0897	0.8034	0.0318	0.7814	0.0332	0.5848	0.0453
		MM	0.8731	0.0233	1.3357	0.2259	0.7917	0.0229	0.7917	0.0229	0.5833	0.0457
		Doğrusal	0.8738	0.0238	0.9573	0.3647	0.8097	0.0713	0.8063	0.0695	0.3841	0.0519
	NP	Karesel	0.9487	0.0141	0.0878	0.4142	0.9053	0.0436	0.8653	0.0484	0.7706	0.0415
		Lojistik	0.8741	0.0239	0.0272	0.3624	0.8098	0.0709	0.8070	0.0690	0.6168	0.0522
		Doğrusal	0.8698	0.0155	-0.0179	0.0269	0.7972	0.0195	0.7786	0.0209	0.5758	0.0301
		Karesel	0.8944	0.0125	-0.8611	0.1401	0.9146	0.0180	0.7641	0.0184	0.6786	0.0240
n=250	P	Lojistik	0.8697	0.0155	-0.0392	0.0589	0.7985	0.0207	0.7774	0.0221	0.5759	0.0301
		MM	0.8697	0.0155	1.2967	0.1361	0.7875	0.0151	0.7875	0.0151	0.5750	0.0302
		Doğrusal	0.8700	0.0156	0.9971	0.2588	0.8029	0.0515	0.7915	0.0499	0.4057	0.0342
		Karesel	0.9454	0.0091	0.0506	0.3185	0.8977	0.0357	0.8529	0.0372	0.7506	0.0280
	NP	Lojistik	0.8701	0.0156	-0.0040	0.2606	0.8018	0.0516	0.7928	0.0499	0.5945	0.0342
		Doğrusal	0.8687	0.0110	-0.0174	0.0197	0.7952	0.0140	0.7776	0.0148	0.5728	0.0213
		Karesel	0.8947	0.0086	-0.8552	0.0990	0.9135	0.0127	0.7629	0.0126	0.6764	0.0171
		Lojistik	0.8687	0.0110	-0.0362	0.0423	0.7962	0.0150	0.7767	0.0157	0.5729	0.0213
n=500	P	MM	0.8687	0.0110	1.2864	0.0946	0.7862	0.0107	0.7862	0.0107	0.5724	0.0213
		Doğrusal	0.8689	0.0111	0.9971	0.2109	0.7982	0.0407	0.7873	0.0409	0.4145	0.0243
		Karesel	0.9444	0.0063	0.0408	0.2547	0.8946	0.0291	0.8471	0.0300	0.7417	0.0200
		Lojistik	0.8690	0.0111	-0.0181	0.2098	0.7995	0.0406	0.7860	0.0407	0.5855	0.0242
	NP	Doğrusal	0.8687	0.0110	-0.0174	0.0197	0.7952	0.0140	0.7776	0.0148	0.5728	0.0213
		Karesel	0.8947	0.0086	-0.8552	0.0990	0.9135	0.0127	0.7629	0.0126	0.6764	0.0171
		Lojistik	0.8687	0.0110	-0.0362	0.0423	0.7962	0.0150	0.7767	0.0157	0.5729	0.0213
		MM	0.8687	0.0110	1.2864	0.0946	0.7862	0.0107	0.7862	0.0107	0.5724	0.0213

AUC: Eğri Altında Kalan Alan Sn: Duyarlılık Sp: Özgüllük c: Optimum Kesim Noktası J: Youden İndeksi p: Parametrik ROC Eğrisi Analizi Yöntemi NP: Parametrik Olmayan ROC Eğrisi Yöntemi mm: MiniMax Prosedürü

Tablo 4.5- Senaryo V – Hasta Verilerinin Türetiminde Bozulmuş Çok Değişkenli Normal Dağılım Verilerinin Kullanıldığı Durum

			<i>AUC</i>	<i>SH(AUC)</i>	<i>c</i>	<i>SH(c)</i>	<i>S_n</i>	<i>SH(S_n)</i>	<i>S_p</i>	<i>SH(S_p)</i>	<i>j max</i>	<i>SH(j max)</i>
n=50	P	Doğrusal	0.7541	0.0446	0.1414	0.1022	0.6205	0.0719	0.7767	0.0683	0.3972	0.0708
		Karesel	0.7493	0.0560	0.9129	0.6360	0.6133	0.0729	0.8919	0.0862	0.5052	0.0575
		Lojistik	0.7534	0.0443	0.1995	0.1500	0.6146	0.0716	0.7861	0.0715	0.4007	0.0715
		MM	0.7536	0.0446	0.3762	0.1614	0.6901	0.0360	0.6901	0.0360	0.3802	0.0719
	NP	Doğrusal	0.7628	0.0441	0.8358	0.2944	0.6691	0.1140	0.8013	0.1083	0.5297	0.0765
		Karesel	0.8232	0.0398	0.0132	0.4105	0.7206	0.1047	0.8241	0.0956	0.5447	0.0734
		Lojistik	0.7633	0.0444	0.2028	0.3079	0.6648	0.1122	0.8056	0.1090	0.4704	0.0760
		MM	0.7633	0.0444	0.2028	0.3079	0.6648	0.1122	0.8056	0.1090	0.4704	0.0760
n=100	P	Doğrusal	0.7417	0.0328	0.1874	0.0727	0.5878	0.0538	0.7931	0.0508	0.3810	0.0504
		Karesel	0.7240	0.0493	0.9968	0.3362	0.5751	0.0408	0.9267	0.0512	0.5018	0.0447
		Lojistik	0.7412	0.0328	0.2394	0.1027	0.5822	0.0535	0.8029	0.0523	0.3852	0.0507
		MM	0.7413	0.0328	0.3383	0.1016	0.6800	0.0259	0.6800	0.0259	0.3601	0.0518
	NP	Doğrusal	0.7540	0.0307	0.8158	0.2050	0.6369	0.0873	0.8027	0.0834	0.5603	0.0530
		Karesel	0.7996	0.0295	0.0098	0.3170	0.6788	0.0942	0.8141	0.0891	0.4928	0.0556
		Lojistik	0.7543	0.0308	0.2148	0.2233	0.6328	0.0889	0.8069	0.0846	0.4398	0.0527
		MM	0.7543	0.0308	0.2148	0.2233	0.6328	0.0889	0.8069	0.0846	0.4398	0.0527
n=250	P	Doğrusal	0.7349	0.0213	0.2171	0.0420	0.5678	0.0348	0.8056	0.0322	0.3734	0.0322
		Karesel	0.7010	0.0346	0.9924	0.1934	0.5563	0.0216	0.9502	0.0238	0.5065	0.0288
		Lojistik	0.7345	0.0212	0.2668	0.0572	0.5626	0.0338	0.8153	0.0320	0.3778	0.0324
		MM	0.7345	0.0213	0.3290	0.0641	0.6746	0.0167	0.6746	0.0167	0.3493	0.0333
	NP	Doğrusal	0.7486	0.0192	0.8222	0.1548	0.6180	0.0691	0.7980	0.0654	0.5840	0.0335
		Karesel	0.7844	0.0184	-0.0489	0.2371	0.6574	0.0796	0.7976	0.0744	0.4550	0.0366
		Lojistik	0.7484	0.0192	0.2211	0.1666	0.6097	0.0682	0.8064	0.0651	0.4161	0.0334
		MM	0.7484	0.0192	0.2211	0.1666	0.6097	0.0682	0.8064	0.0651	0.4161	0.0334
n=500	P	Doğrusal	0.7324	0.0152	0.2284	0.0272	0.5604	0.0247	0.8102	0.0216	0.3706	0.0225
		Karesel	0.6911	0.0266	0.9737	0.1314	0.5519	0.0157	0.9574	0.0153	0.5094	0.0208
		Lojistik	0.7321	0.0152	0.2769	0.0384	0.5553	0.0240	0.8200	0.0213	0.3753	0.0227
		MM	0.7320	0.0152	0.3242	0.0438	0.6727	0.0119	0.6727	0.0119	0.3454	0.0237
	NP	Doğrusal	0.7473	0.0140	0.8153	0.1226	0.6086	0.0545	0.7981	0.0524	0.5932	0.0252
		Karesel	0.7788	0.0134	-0.0546	0.1897	0.6435	0.0675	0.7954	0.0628	0.4389	0.0281
		Lojistik	0.7469	0.0140	0.2290	0.1347	0.5989	0.0546	0.8077	0.0528	0.4066	0.0250
		MM	0.7469	0.0140	0.2290	0.1347	0.5989	0.0546	0.8077	0.0528	0.4066	0.0250

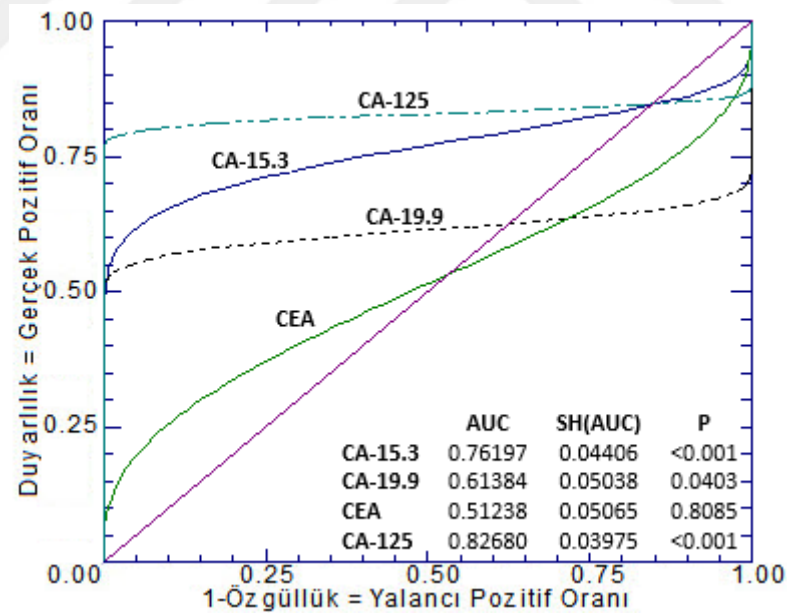
AUC: Eğri Altında Kalan Alan Sn: Duyarlılık Sp: Özgüllük c: Optimum Kesim Noktası J: Youden İndeksi p: Parametrik ROC Eğrisi Analizi Yöntemi
NP: Parametrik Olmayan ROC Eğrisi Yöntemi mm: MiniMax Prosedürü

4.2- Uygulama Veri Seti Bulguları

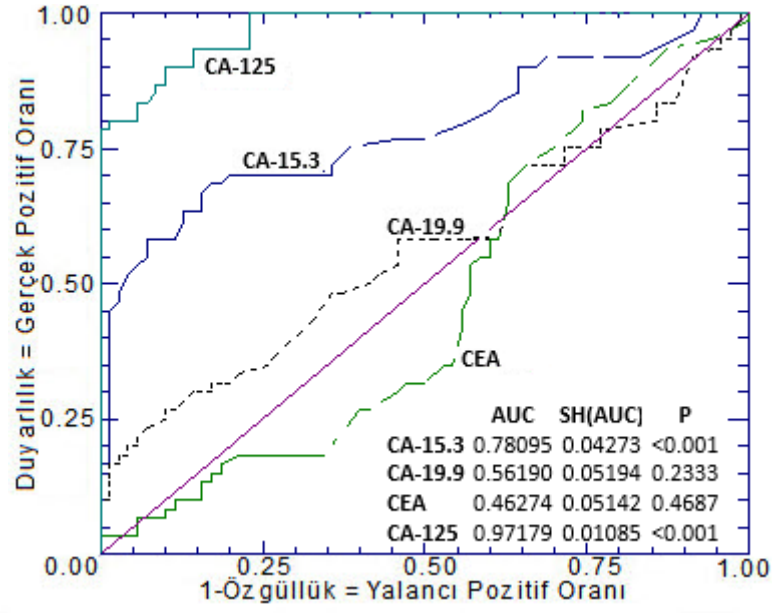
Over hastalığı tanısı koyma kullanılan 4 farklı medikal tanı testi (CA-15.3, CA-19.9, CEA, CA-125) simülasyon çalışmalarında da kullanılan birleştirme yöntemleri ile modellenerek her biri için ayrı ayrı ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeksi hesaplaması yapılmıştır. Bu değerlere ait standart hatalar 1000 tekrar olacak şekilde bootstrap (yeniden örnekleme tekniği) kullanılarak elde edildi. Yöntemlere ait sonuçlar tablo 4.6'da verilmiştir.

Skorların elde edilmesinde kullanılan doğrusal ayırma analizi, karesel ayırma analizi, lojistik ayırma analizi ve en iyi doğrusal birleştirme yöntemlerine ait modellerin katsayıları tablo 4.7'de verildi.

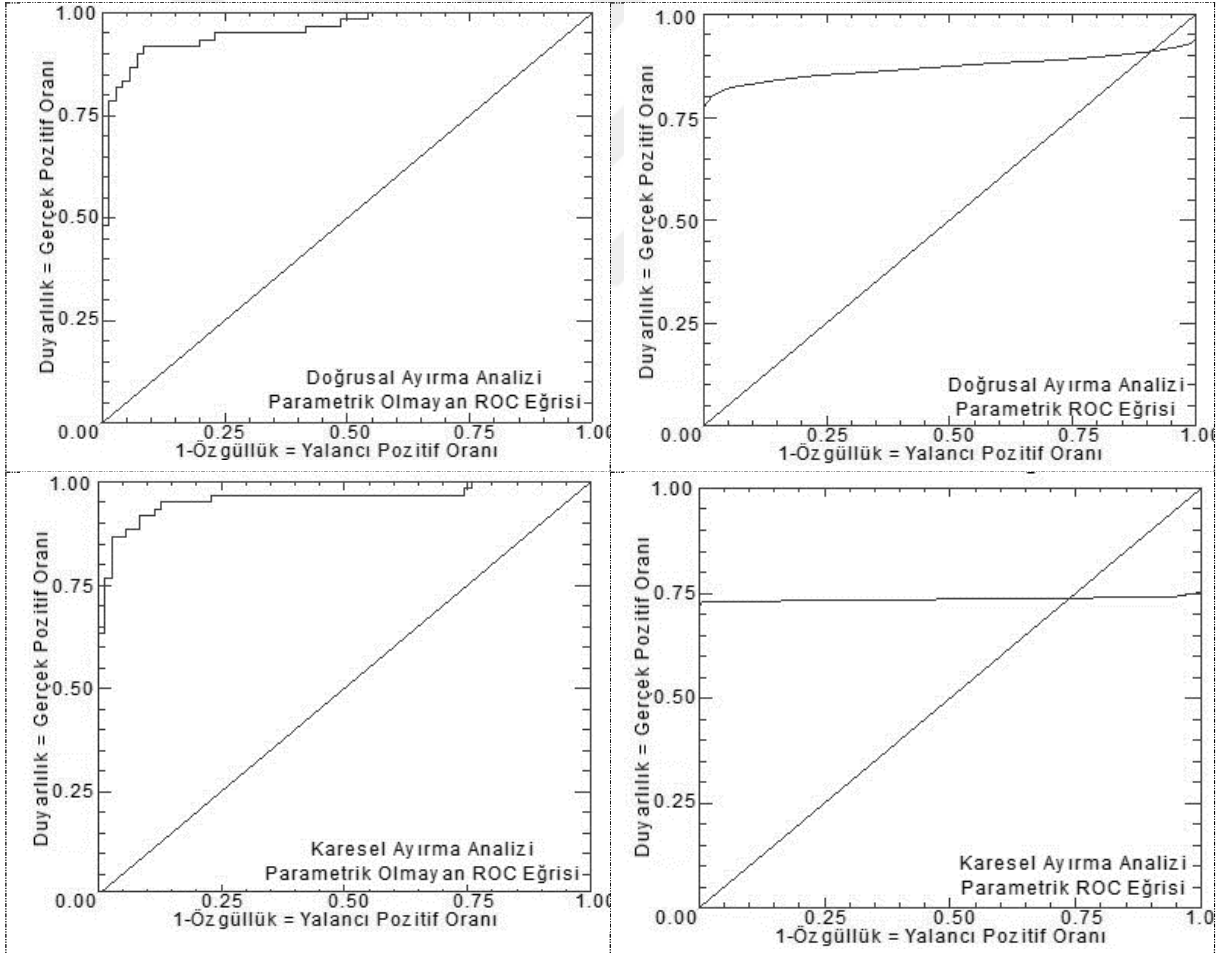
Birleştirme yöntemlerine ait sonuçlar incelendiğinde en yüksek eğri altında kalan alan, duyarlılık ve özgüllük değerlerinin parametrik olmayan ROC eğrisi yöntemi ile elde edildiği gözlemlendi. Bu sonuçlarda özellikle lojistik ayırma analizinin ne yüksek eğri altında kalan alan, duyarlılık ve özgüllük değerleri verdiği görüldü.

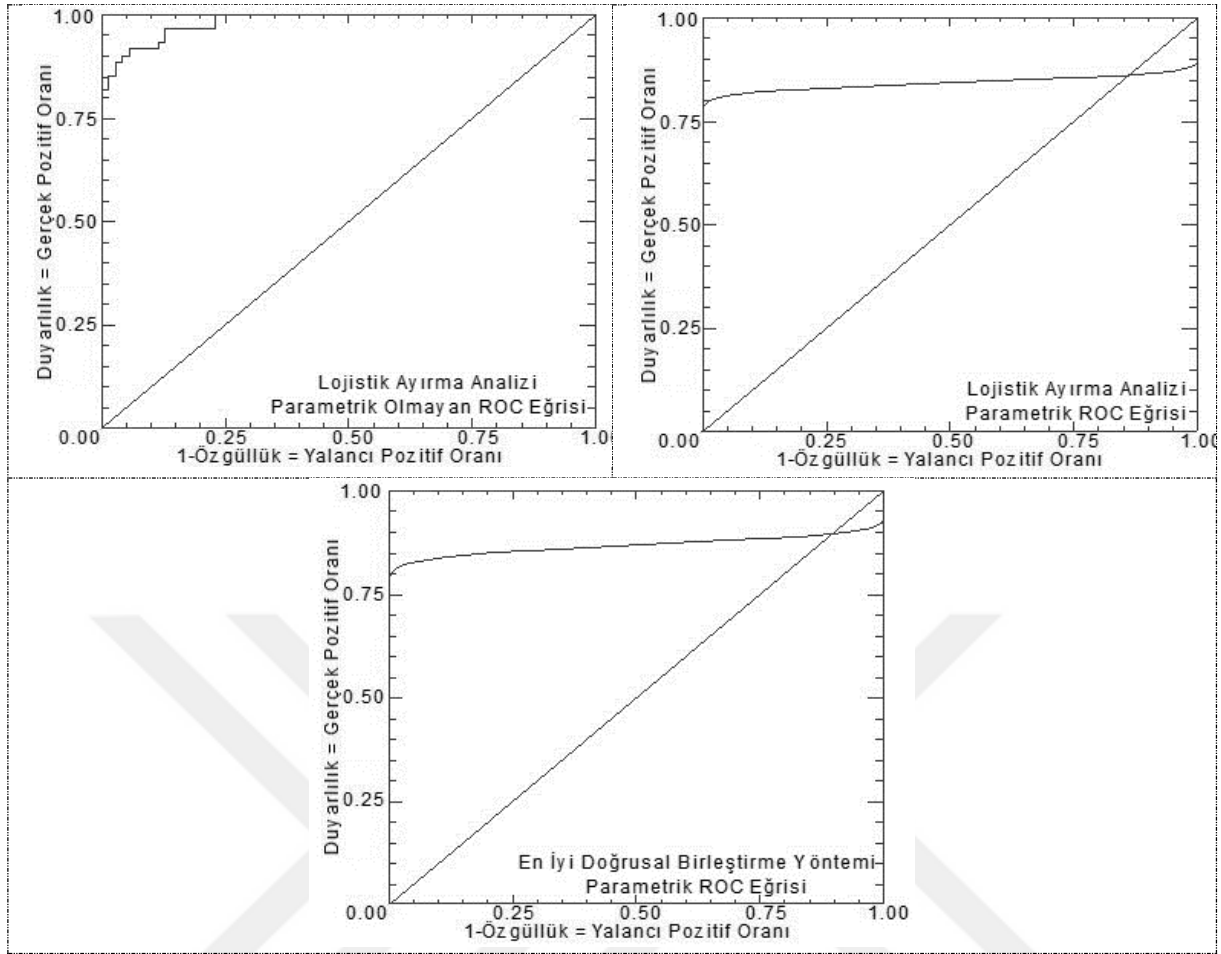


Şekil 4.1- Parametrik ROC Eğrisi



Şekil 4.2- Parametrik Olmayan ROC Eğrisi





Şekil 4.3- Skor Değerlerine Göre Birleştirme Yöntemlerine Ait Parametrik Olmayan ve Parametrik ROC Eğrileri

Tablo 4.6- Uygulama Veri Seti Sonuçları

		<i>AUC</i>	<i>SH_b(AUC)</i>	<i>c</i>	<i>SH_b(c)</i>	<i>S_n</i>	<i>SH_b(S_n)</i>	<i>S_p</i>	<i>SH_b(S_p)</i>	<i>j max</i>	<i>SH_b(j max)</i>
P	Doğrusal	0.8708	0.0196	-0.6459	0.2464	0.7989	0.0313	0.9830	0.0089	0.7819	0.0380
	Karesel	0.7361	0.0247	6.1572	4.4446	0.7269	0.0257	0.9999	0.0007	0.7268	0.0259
	Lojistik	0.8419	0.0199	0.4649	12.7805	0.7955	0.0235	0.9913	0.0019	0.7869	0.0245
	MM	0.8679	0.0201	2.6811	0.6075	0.8443	0.0205	0.8443	0.0205	0.6887	0.0410
NP	Doğrusal	0.9571	0.0188	-1.0459	0.0393	0.9167	0.0516	0.9143	0.2330	0.8310	0.0524
	Karesel	0.9586	0.0190	-1.3712	0.0392	0.8667	0.0446	0.9714	2.0713	0.8381	0.0459
	Lojistik	0.9831	0.0079	-0.1777	0.0356	0.9167	0.0398	0.9429	1.4215	0.8595	0.0414

AUC: Eğri Altında Kalan Alan Sn:Duyarlılık Sp: Özgüllük c:Optimum Kesim Noktası J: Youden İndeksi p:Parametrik ROC Eğrisi Analizi Yöntemi NP: Parametrik Olmayan ROC Eğrisi Yöntemi

Tablo 4.7- Uygulama Veri Seti b_0 ve b katsayıları Sonuçları

	Katsayılar				
	b_0	b			
		CA-15.3	CA-19.9	CEA	CA-125
Doğrusal Ayırma Analizi	-1.771875	0.0100710320	0.004796214	-0.041812562	0.001180130
Karesel Ayırma Analizi	-6.000855	-0.024373082	-0.041003315	-0.256387982	-0.006682174
Lojistik Ayırma Analizi	-4.732216	0.013830	0.009664	-0.109353	0.013295
En İyi Doğrusal Birleştirme Analizi (t=0.1)	-	0.02772900	0.01857967	-0.04080667	0.00557608

Uygulama veri setinin analizi sonucunda elde edilen b_0 ve b katsayıları ile kurulan ayırma fonksiyonları aşağıdaki gibidir.

Doğrusal ayırma analizi için ayırma fonksiyonu

$$-1.772 + 0.010 \times CA15.3 + 0.004 \times CA19.9 - 0.042 \times CEA + 0.001 \times CA125$$

Karesel ayırma analizi için ayırma fonksiyonu

$$-6.000 - 0.024 \times CA15.3 - 0.041 \times CA19.9 - 0.256 \times CEA - 0.007 \times CA125$$

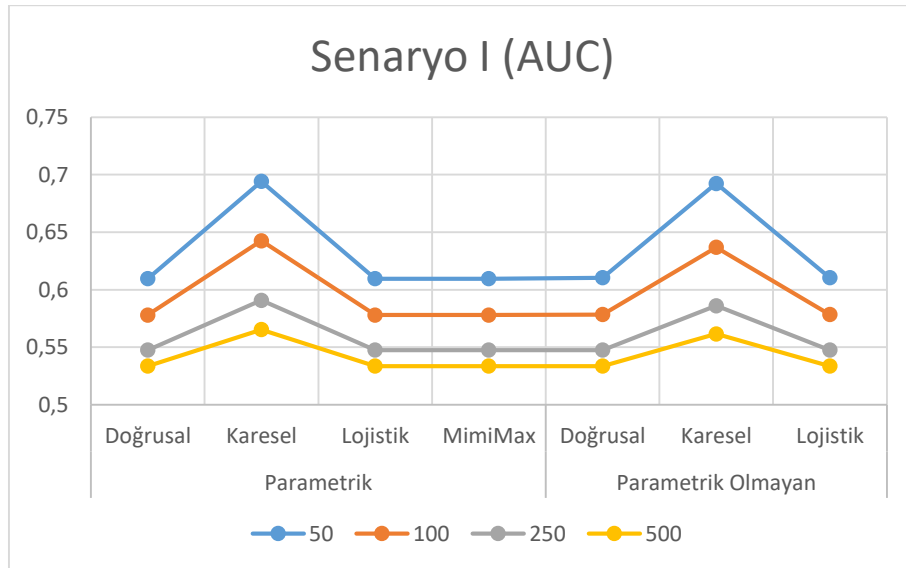
Lojistik ayırma analizi için ayırma fonksiyonu

$$-4.732 + 0.014 \times CA15.3 + 0.010 \times CA19.9 - 0.109 \times CEA + 0.013 \times CA125$$

En iyi doğrusal bireştirme analizi için ayırma fonksiyonu

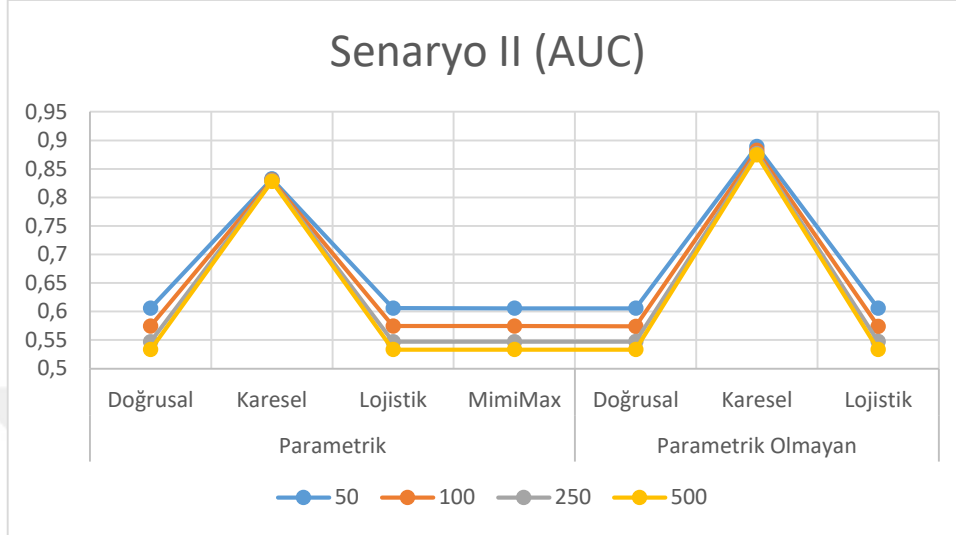
$$0.028 \times CA15.3 + 0.019 \times CA19.9 - 0.041 \times CEA + 0.006 \times CA125$$

Şekil 4.1’de medikal tanı testlerinin hasta ve sağlıklı grup için ortalama vektörlerinin aynı, kovaryans matrislerinin aynı olduğu durumun grafiği incelenmiştir. Medikal tanı testlerinin gruplara göre ortalama değerleri aynı olduğu için, medikal tanı testlerinin grup sınıflamasını yapamaması beklenmektedir. Diğer bir ifade ile eğri altında kalan alanın 0.5’e yakınsaması beklenmektedir. Şekil 4.1’de de görüldüğü gibi örneklem birim sayısı arttığında ROC eğrisi altında kalan alanın 0.5 değerine yakınsadığı görülmüştür. Kovaryans matrislerinin eşit olduğundan dolayı doğrusal ayırma analizinin karesel göre daha iyi sonuç verdiği bulundu. Doğrusal ayırma analizi ile lojistik ayırma analizi sonuçlarının birbirine benzer olduğu bulundu. Ayrıca parametrik ROC modeli ve parametrik olmayan ROC modeli incelendiğinde, her yöntemin her iki model içinde birbirine çok yakın sonuçlar verdiği bulundu.



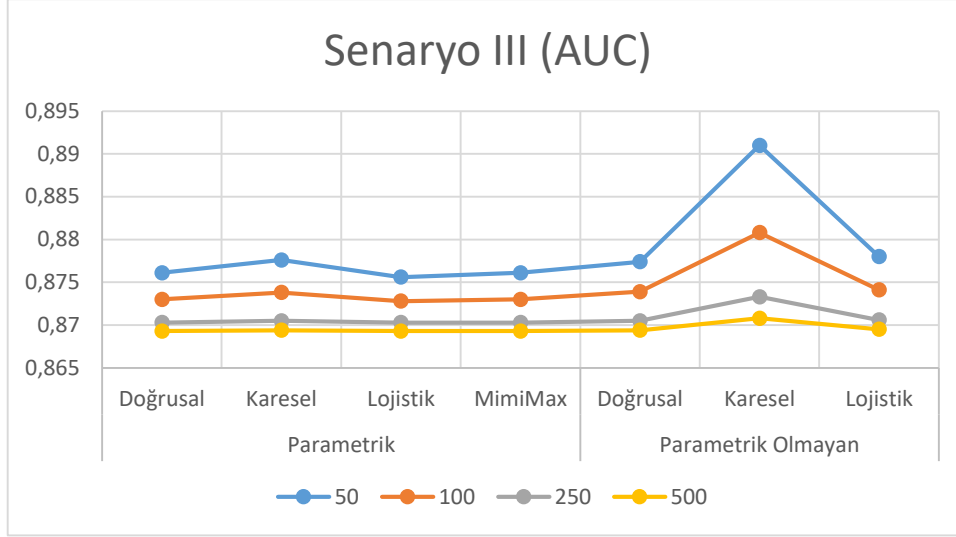
Şekil 4.4- Simülasyon Senaryosu I’in AUC Değerleri

Şekil 4.2’de medikal tanı testlerinin hasta ve sağlıklı grup için ortalama vektörlerinin aynı, kovaryans matrislerinin farklı olduğu durumun grafiği incelenmiştir. Kovaryans matrisleri farklı olduğu için karesel ayırma analizinin eğri altında kalan alanı daha yüksek tahminlediği bulundu.



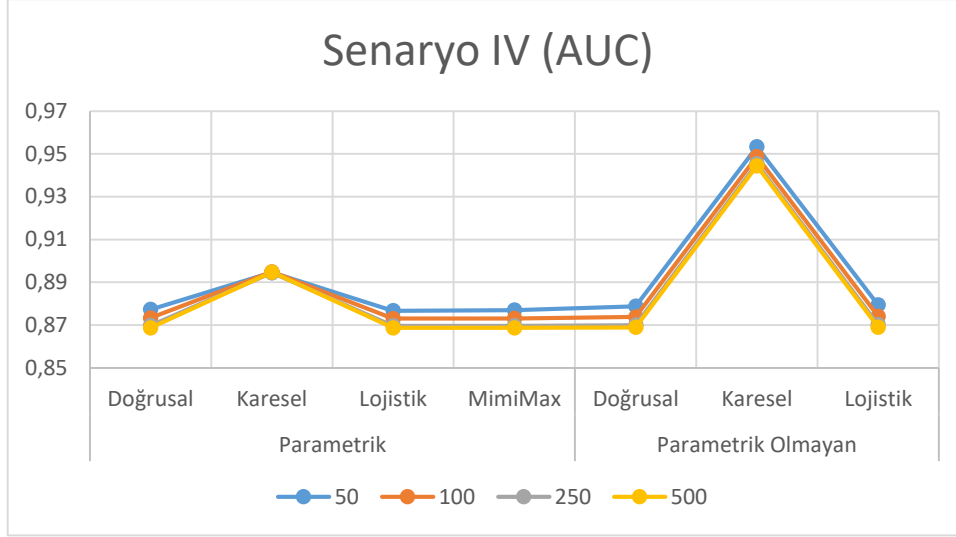
Şekil 4.5- Simülasyon Senaryosu II’nin AUC Değerleri

Şekil 4.3’de medikal tanı testlerinin hasta ve sağlıklı grup için ortalama vektörlerinin farklı, kovaryans matrislerinin aynı olduğu durumun grafiği incelenmiştir. Hasta ve sağlıklı grubun medikal tanı testi ortalamaları birbirinden farklı olduğu için ROC eğrisinin altında kalan alanın 1’e ne kadar yakın olursa o yöntemin daha iyi olacağı varsayılır. Şekil 4.3 incelendiğinde küçük örnek hacminde parametrik olmayan karesel ayırma analizinin çok yüksek AUC değeri verdiği bulunmuştur. Bunun nedeni örnek hacminin yetersizliğinden dolayı olduğu düşünülmektedir. Örnek birim sayısı arttıkça parametrik ve parametrik olmayan her yöntemin eğri altında kalan alanların birbiri ile benzer sonuçlar verdiği bulunmuştur. Kovaryans matrisleri aynı olduğu için doğrusal ayırma analizinin daha yüksek sonuçlar vereceği düşünülmüştür fakat bu durum ile karşılaşılmamıştır.



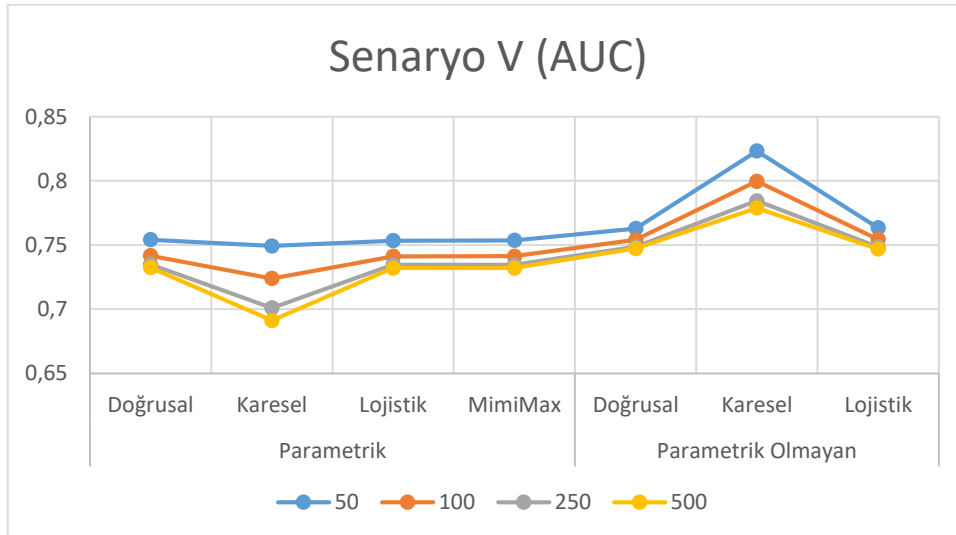
Şekil 4.6- Simülasyon Senaryosu III'in AUC Değerleri

Şekil 4.4'de medikal tanı testlerinin hasta ve sağlıklı grup için ortalama vektörlerinin farklı, kovaryans matrislerinin farklı olduğu durumun grafiği incelenmiştir. Bu senaryo uygulama verimizin ortalama vektörleri ve kovaryans matrisleri ile benzerlik göstermektedir. Uygulama alanında en sık karşılaşılan veri dağılımı bu şekilde olduğu görülmüştür. Şekil 4.4 incelendiğinde karesel ayırma analizi yönteminin hem parametrik hemde parametrik olmayan modellerde daha yüksek eğri altında kalan alan hesapladığı bulunmuştur. Bunun nedeni kovaryans matrislerinin farklı olmasından dolayıdır. Karesel ayırma analizi kendi başına incelendiğinde ise parametrik olmayan karesel ayırma analizinin performansının parametrik olan ayırma analizinden daha yüksek olduğu bulunmuştur. Her zaman medikal tanı testlerinin dağılımı normal dağılıma uygunluk gösteremeyeceğinden dolayı senaryo IV'deki veri yapıları ile karşılaşıldığında parametrik olmayan karesel ayırma analizi kullanılması önerilmektedir.



Şekil 4.7- Simülasyon Senaryosu IV'in AUC Değerleri

Şekil 4.5'de hasta verilerinin türetiminde bozulmuş çok değişkenli normal dağılım verilerinin kullanıldığı durum incelenmiştir. Parametrik olmayan her yöntemin parametrik yöntemlerin tamamından çok daha yüksek eğri altında kalan alan hesapladığı bulunmuştur. Bunu sebebi veri yapımızı bozarak normallik varsayımlarını bozduğumuz içindir. Normal dağılım göstermediği için parametrik olmayan yöntemler daha başarılı sonuçlar vermiştir. Karesel ayırma analizi diğer parametrik olmayan yöntemlerden daha başarılı sonuçlar vermiştir. Parametrik yöntemlerden en iyi birleştirme yöntemi, lojistik ayırma yöntemi ve doğrusal ayırma yöntemi birbirine benzer sonuçlar verdiği bulunmuştur.



Şekil 4.8- Simülasyon Senaryosu V'in AUC Değerleri

5- TARTIŞMA

Tıp alanında tedaviye karar verme açısından medikal tanı testleri önemli bir role sahip olmanın yanı sıra tedavinin planlanmasında kullanılmaktadır. Medikal tanı testleri kullanılarak, hasta ile sağlıklı birimleri doğru sınıflamak ve hastanın tedavisi için sağlık personelinin doğru planlama yapmasına olumlu katkılar sağlamak amacı ile kullanılmaktadır. Tek bir tanı testi sonucu kullanılarak hastalığa ait tanının belirlenmesi yerine günümüzde birden fazla tanı testi kullanılarak bu belirlemenin daha başarılı olması sağlanmaktadır.

Medikal tanı testleri ile hasta ve sağlıklı grup olarak sınıflamada birden fazla yöntemin kullanıldığı bilinmektedir. Yapılan literatür taramaları sonucu, hasta ve sağlıklı sınıflama yapabilmek için en sık kullanılan yöntemler arasında doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizi yöntemleri olduğu belirlendi. Bu üç yöntem ek olarak Su ve Liu (1993) tarafından önerilen en iyi doğrusal birleştirme yöntemi kullanılmaktadır. Qin ve Zhang (2009) yapmış oldukları çalışmada medikal testlerin birleştirilmesi sonucu eğri sınıflama başarılarının arttığını bulmuşlardır. Bu çalışmada; sağlık çalışanlarına yol göstermek ve kullanacakları yöntemin seçiminde kılavuzluk etmesi amacı ile bu yöntemlerin performans karşılaştırmaları yapıldı.

Yöntemlerin karşılaştırmalarında Monte Carlo simülasyon yöntemi kullanılarak farklı örneklem büyüklüğü, ortalama vektörleri ve kovaryans matrislerine göre veri setleri oluşturuldu ve her bir yöntemde bu veri setleri kullanıldı. Yöntemlerin performanslarının değerlendirilmesinde, ROC eğrisi altında kalan alan, optimum kesim noktası, duyarlılık, özgüllük ve Youden indeks değerleri kullanıldı. Yapılan simülasyon çalışmaları sonucunda elde edilen sonuçlar doğrultusunda, sağlık alanında kullanılabilirlik durumu değerlendirmek için gerçek bir uygulama verisi üzerinde çalışma yapıldı. Uygulama verisi CA-15.3, CA-19.9, CEA, CA-125 tanı testlerini birleştirilerek analizler yapıldı.

Simülasyon çalışmasına konu olan 4 yöntemin temelinde birleştirme yöntemleri yer almaktadır. Burada temel amaç farklı tanı testlerini doğrusal birleştirilerek sınıflama başarısını arttırmaktır. Bu tez çalışmasında, 4 yöntem ile tanı testlerinin birleştirilmesi amacı ile sürekli yapıda olması koşulu yer almaktadır. Medikal tanı testlerinin çok değişkenli normal dağılıma uygunluk gösterme durumuna göre parametrik veya parametrik olmayan ROC eğrisi modeli kullanıldı. Simülasyon çalışmalarında 5 farklı senaryo yapılması

planlandı. Bu senaryolar; ortalamalar aynı/farklı, kovaryans matrisleri benzer/farklı ve bozulmuş veri yapısı olacak şekildedir.

Medikal tanı testleri birleştirilmesindeki amaç belirtildiği gibi karşılaştırma kriteri olan ve 0 ile 1 arasında değer alan eğri altında kalan alanın 1'e yaklaşmasıdır. Tanı testlerinin çok değişkenli normal dağılıma uygunluk gösterdiğinde elde edilen eğri altında kalan alanların parametrik olmayan yöntemlere göre daha iyi eğri altında kalan alan tahmin ettiği bulunmuştur. En iyi doğrusal ayırma analizi yönteminde parametrik ROC eğrisi yöntemi ile elde edilen sonuçlara benzer sonuçlar vermiştir. Bu yönü ile en iyi doğrusal birleştirme yöntemi güçlü bir yöntem olarak ifade edilebilir. Sameera, Vardhan ve ark.'nın (2016) çalışmalarında en iyi doğrusal birleştirme yönteminin sonuçları ile benzerlik gösterdiği bulundu. Medikal tanı testlerinin normale uygunluk dağılımları incelendiğinde hasta grubundaki dağılımların çarpık bir dağılıma sahip olduğu bulundu. Bu durumda parametrik olmayan yöntemlerin parametrik yöntemlere göre daha yüksek eğri altında kalan alan hesapladığı bulundu. Yu ve Park (2015), belirli bir özgüllük aralığında, eğri altında kalan alanı optimize eden çoklu biyolojik belirteçlerin en iyi doğrusal kombinasyonunu bulmak için kolay uygulanabilen iki algoritma önermişlerdir. Algoritmalarından biri parametrik olmayan ROC eğrisi yöntemidir. Bu tez çalışması sonucu ile Yu ve Park'ın bulguları örtüşmektedir.

Ortalama vektörlerinin aynı olduğu durumda eğri altında kalan alanın diyagonal üzerinde çıkması beklenmektedir. Senaryo 1 incelendiğinde tüm yöntemlerin 0.5 değerine yakınsadığı bulunmuştur. Karesel ayırma analizi kovaryans matrisinden etkilendiği için 0.5 değerinden sapmalar gözlenmiştir. Fakat senaryo 2'de kovaryans matrisleri farklı olduğu için karesel ayırma analizi diğer yöntemlerden daha yüksek eğri altında kalan alan tahminlediği bulunmuştur. Doğrusal ve lojistik ayırma analizlerinin sonuçlarının birbirine benzer olduğu bulundu.

Ortalama vektörlerinin farklı olduğu durumda ise eğri altına kalan alanın 1'e ne kadar yakın olursa medikal tanı testlerinin sınıflama başarısının arttığı bilinmektedir. Senaryo 3 ve 4'de ortalama vektörlerinin farklı olduğu durum incelendi. Medikal tanı testlerinin çok çarpık bir dağılıma sahip olduklarından dolayı, normale uygunluk göstermediği bulundu. Bu durumdan dolayı parametrik olmayan yöntemlerin eğri altında kalan değerlerinin daha yüksek olduğu bulundu. Senaryo 3 ve 4'de de benzer şekilde olduğu gibi, kovaryansların ortaklık varsayımına göre karesel ayırma analizinin performansı değişiklik göstermektedir.

Kovaryans matrislerinin homojenliđinin bozulduđu durumlarda da yöntemler iyi sonuçlar vermiştir. Bu da doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizinden elde edilen skorların medikal tanı testlerinin normallik varsayımında sapmasından etkilenmediđini göstermektedir. Özellikle karesel ayırma analizi ön plana çıkmıştır. Çünkü karesel ayırma analizi yöntemi daha çok bu amaçta kullanılmaktadır. Kovaryans matrislerinin heterojenliđi durumunda parametrik olmayan ROC eğrisi yöntemleri ile elde edilen sonuçlar, parametrik ROC eğrisine göre elde edilen sonuçlara göre daha yüksek duyarlılık ve özgüllük değerleri verdiđi görüldü.

Dođrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizinin kovaryans matrislerinin benzerliđi ve tanı testlerinin çok deđişkenli normal dağılıma benzerliđinden etkilendiđi bulundu. Fakat en iyi doğrusal birleştirme yönteminde kovaryans matrislerin eşitliđi ve ya normallik varsayımın sağlanamaması durumundan etkilenmediđi görüldü.

Uygulama verisindeki sonuçlar incelendiđinde parametrik olmayan ROC analizi ile kullanılarak yöntemlerden elde edilen sonuçların daha yüksek eğri altında kalan alan, duyarlılık ve özgüllük değerleri verdiđi görülmüştür. Bunun nedeni medikal tanı testlerinin çok deđişkenli normal dağılıma uygunluk göstermemesinden kaynaklanmaktadır. Uygulama veri setinden elde edilen sonuçlar ile senaryo V'e göre elde edilen verilerin simülasyon sonuçları birbirine benzerlik göstermektedir. Medikal tanı testleri çok deđişkenli normal dağılıma uygunluk göstermediđinde parametrik olmayan ROC eğrisi yöntemi ile birleştirme yapmak daha avantajlı olacaktır.

6- SONUÇ VE ÖNERİLER

Bu tez çalışmasında, sürekli yapıdaki çoklu tanı testleri için birleştirme yöntemlerinin ROC eğrisi analizi kullanarak karşılaştırılmıştır. Ortalama vektörlerinin ve kovaryans matrisilerinin benzer ya da farklı olduğu tüm senaryo durumları incelendiğinde en iyi doğrusal birleştirme yönteminin medikal tanı testlerinin birleştirilmesinde en etkin yöntem olduğu belirlenmiştir. Diğer yöntemlerden olan karesel ayırma analizi ve lojistik ayırma analizi bu alanda yaygın kullanılan bir yöntem olmasına rağmen kovaryans matrislerinin ortaklığından etkilenmektedir.

Medikal tanı testlerinin hasta ve sağlam sınıflamasını yaparken yaygın olarak karşılaşılan durum, hasta grubundaki birimlerin değerlerinin çok farklı aralıkta yer almasıdır. Bu durumdan dolayı hasta grubunun normal dağılım varsayımından etkilenmemektedir. Medikal tanı testlerinin birleştirilmesinde parametrik olmayan yöntemler ile bu tez çalışmasında ele alınan en iyi doğrusal birleştirme yönteminin kullanılması önerilmektedir.

Yapılan simülasyon çalışmaları sonucunda geleneksel olan doğrusal ayırma analizi, karesel ayırma analizi ve lojistik ayırma analizini paket programlar aracılığı ile yapabilmek mümkündür. En iyi doğrusal birleştirme yöntemi ise programlama dili kullanılarak uygulanabilir. Gelecek çalışmalarda ise en iyi doğrusal birleştirme yönteminin paket program haline getirilmesi önerilmektedir.

KAYNAKLAR DİZİNİ

- Anderson, T. W., & Bahadur, R. R. (1962). Classification into two multivariate normal distributions with different covariance matrices. *The annals of mathematical statistics*, 33(2), 420-431.
- Antonogeorgos, G., Panagiotakos, D. B., Priftis, K. N., & Tzonou, A. (2009). Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10-to 12-years-old children: Divergence and similarity of the two statistical methods. *International journal of pediatrics*, 2009.
- Colak, E., Mutlu, F., Bal, C., Oner, S., Ozdamar, K., Gok, B., & Cavusoglu, Y. (2012). Comparison of semiparametric, parametric, and nonparametric ROC analysis for continuous diagnostic tests using a simulation study and acute coronary syndrome data. *Computational and mathematical methods in medicine*, 2012.
- Cole, P., & Morrison, A. S. (1980). Basic issues in population screening for cancer. *Journal of the National Cancer Institute*, 64(5), 1263-1272.
- Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 225-265.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837-845.
- Egan, J. P. (1975). Signal detection theory and {ROC} analysis.
- Epstein, A. M., Begg, C. B., & McNeil, B. J. (1986). The use of ambulatory testing in prepaid and fee-for-service group practices. *New England Journal of Medicine*, 314(17), 1089-1094.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. doi:<http://dx.doi.org/10.1016/j.patrec.2005.10.010>
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1-67.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1): Springer series in statistics New York.
- Gao, F., Xiong, C., Yan, Y., Yu, K., & Zhang, Z. (2008). Estimating optimum linear combination of multiple correlated diagnostic tests at a fixed specificity with receiver operating characteristic curves. *Journal of Data Science*, 6(1), 105-123.
- Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging*, 29(3), 307-335.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.
- Härdle, W., & Simar, L. (2007). *Applied multivariate statistical analysis*: Springer Science & Business Media.
- Hastie, T., Tibshirani, R., & Andreas, B. (1999). Flexible discriminant and mixture models. *Statistics and neural networks: advances at the interface*, 1-23.
- Hosmer, D. W., & Lemeshow, S. (2005). *Logistic Regression for Matched Case-Control Studies*. *Applied Logistic Regression*, Second Edition, 223-259.
- Johnson, R. A., & Wichern, D. (2002). *Multivariate analysis*. John Wiley & Sons, Ltd.
- Jonson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis*. In: Prentice Hall Inc.
- Kang, L., Liu, A., & Tian, L. (2016). Linear combination methods to improve diagnostic/prognostic accuracy on future observations. *Stat Methods Med Res*, 25(4), 1359-1380. doi:10.1177/0962280213481053
- King, E. N., & Ryan, T. P. (2002). A preliminary investigation of maximum likelihood logistic regression versus exact logistic regression. *The American Statistician*, 56(3), 163-170.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Komarek, P., & Moore, A. W. (2003). Fast Robust Logistic Regression for Large Sparse Datasets with Binary Outputs. Paper presented at the AISTATS.
- Li, J., & Fine, J. P. (2008). ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics*, 9(3), 566-576. doi:10.1093/biostatistics/kxm050
- Lin, H., Zhou, L., Peng, H., & Zhou, X. H. (2011). Selection and combination of biomarkers using ROC method for disease classification and prediction. *Canadian Journal of Statistics*, 39(2), 324-343.
- Liu, C., Liu, A., & Halabi, S. (2011). A min-max combination of biomarkers to improve diagnostic accuracy. *Statistics in Medicine*, 30(16), 2005-2014.
- Ma, S., & Huang, J. (2007). Combining multiple markers for classification using ROC. *Biometrics*, 63(3), 751-757. doi:10.1111/j.1541-0420.2006.00731.x
- Mamtani, M. R., Thakre, T. P., Kalkonde, M. Y., Amin, M. A., Kalkonde, Y. V., Amin, A. P., & Kulkarni, H. (2006). A simple method to combine multiple molecular biomarkers for dichotomous diagnostic classification. *BMC bioinformatics*, 7(1), 442.
- McNeil, B. J., & Hanley, J. A. (1983). Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Medical decision making: an international journal of the Society for Medical Decision Making*, 4(2), 137-150.
- Metz, C. E. (1978). Basic principles of ROC analysis. Paper presented at the Seminars in nuclear medicine.
- Moawed, S. A., & Osman, M. M. (2017). The Robustness of Binary Logistic Regression and Linear Discriminant Analysis for the Classification and Differentiation between Dairy Cows and Buffaloes. *International Journal of Statistics and Applications*, 7(6), 304-310.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Nevin, J. A. (1969). SIGNAL DETECTION THEORY AND OPERANT BEHAVIOR: A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. *Journal of the Experimental Analysis of Behavior*, 12(3), 475-480.
- Nicoll, D., McPhee, S. J., & Pignone, M. (2012). *Pocket guide to diagnostic tests*. McGraw-Hill Medical.
- Obuchowski, N. A., Lieber, M. L., & Wians, F. H. (2004). ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clinical chemistry*, 50(7), 1118-1125.
- Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Medicine.
- Pepe, M. S., & Thompson, M. L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics*, 1(2), 123-140.
- Perkins, N. J., & Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American journal of epidemiology*, 163(7), 670-675.
- Qin, J., & Zhang, B. (2010). Best combination of multiple diagnostic tests for screening purposes. *Statistics in Medicine*, 29(28), 2905-2919.
- Reid, M. C., Lachs, M. S., & Feinstein, A. R. (1995). Use of methodological standards in diagnostic test research: getting better but still not good. *Jama*, 274(8), 645-651.
- Reiser, B., & Faraggi, D. (1997). Confidence intervals for the generalized ROC criterion. *Biometrics*, 644-652.
- Rencher, A. C. (2002). Principal component analysis. *Methods of Multivariate Analysis*, Second Edition, 380-407.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, 12(1), 1.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Sameera, G., Vardhan, R. V., & Sarma, K. V. S. (2016). Binary classification using multivariate receiver operating characteristic curve for continuous data. *Journal of biopharmaceutical statistics*, 26(3), 421-431.
- Schisterman, E. F., Perkins, N. J., Liu, A., & Bondell, H. (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*, 16(1), 73-81.
- Schulzer, M. (1994). Diagnostic tests: a statistical review. *Muscle & nerve*, 17(7), 815-819.
- Segen, J. C., & Wade, J. (2002). *The patient's guide to medical tests: everything you need to know about the tests your doctor orders*. Infobase Publishing.
- Shapiro, D. E. (1999). The interpretation of diagnostic tests. *Statistical methods in medical research*, 8(2), 113-134.
- Sharma, B. (2014). Right choice of a method for determination of cut-off values: A statistical tool for a diagnostic test.
- Shultz, E. K. (1987). An extension to ROC curve analysis: MultiROC. Paper presented at the Proceedings of the Annual Symposium on Computer Application in Medical Care.
- Sox Jr, H. C., Koran, L. M., Sox, C. H., Marton, K. I., Dugger, F., & Smith, T. (1989). A medical algorithm for detecting physical disease in psychiatric patients. *Psychiatric Services*, 40(12), 1270-1276.
- Su, J. Q., & Liu, J. S. (1993). Linear-Combinations of Multiple Diagnostic Markers. *Journal of the American Statistical Association*, 88(424), 1350-1355.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285-1293.


KAYNAKLAR DİZİNİ (Devam Ediyor)

- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological science in the public interest*, 1(1), 1-26.
- Wilson, J. M. G., Jungner, G., & World Health Organization. (1968). *Principles and practice of screening for disease*.
- Wilson, P. D., & Langenberg, P. (1999). Usual and shortest confidence intervals on odds ratios from logistic regression. *The American Statistician*, 53(4), 332-335.
- Wu, W., Mallet, Y., Walczak, B., Penninckx, W., Massart, D. L., Heuerding, S., & Erni, F. (1996). Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica Chimica Acta*, 329(3), 257-265.
- Yu, W., & Park, T. (2015). Two simple algorithms on linear combination of multiple biomarkers to maximize partial area under the ROC curve. *Computational Statistics & Data Analysis*, 88, 15-27.
- Yu, W., Kwon, M. S., & Park, T. (2015). Multivariate quantitative multifactor dimensionality reduction for detecting gene-gene interactions. *Human heredity*, 79(3-4), 168-181.
- Zhang, D. D., Zhou, X. H., Freeman, D. H., & Freeman, J. L. (2002). A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. *Statistics in medicine*, 21(5), 701-715.
- Zhou, X. H., McClish, D. K., & Obuchowski, N. A. (2009). *Statistical methods in diagnostic medicine* (Vol. 569). John Wiley & Sons.
- Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, 39(4), 561-577.

EKLER DİZİNİ

EK – 1

Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Girişimsel Olmayan Klinik Araştırma Etik Kurul 09.08.2017 gün ve 805587-21/G-217 sayılı numaralı etik kurul kararı



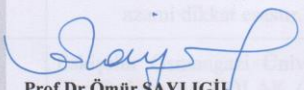
ESKİŞEHİR OSMANGAZI ÜNİVERSİTESİ
GİRİŞİMSEL OLMAYAN KLİNİK ARAŞTIRMALAR ETİK KURULU
BAŞKANLIĞI

09 Ağustos 2017

Sayı: 80558721/G – 217
Konu: Karar – Arş.Gör.Muzaffer BİLGİN

Sayın; Doç.Dr.Ertuğrul ÇOLAK
Eskişehir Osmangazi Üniversitesi Tıp Fakültesi
Biyostatistik Anabilim Dalı

Sorumlu Araştırmacısı olduğunuz **“Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması”** başlıklı çalışma hakkında alınan karar ilâşikte gönderilmiştir.
Bilgilerinizi ve gereğini saygı ile rica ederim.



Prof.Dr.Ömür ŞAYLIGİL
Etik Kurul Başkanı
Eskişehir Osmangazi Üniversitesi
Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu

Araştırmacıların beşerler dileği.

Prof.Dr.Ömür ŞAYLIGİL
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Aile Hekimliği Anabilim Dalı

Doç. Dr. Uğur BİLGE
(Başkan Yardımcısı)
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Aile Hekimliği Anabilim Dalı

Doç. Dr. Özlem ÖRSAL
(Raportör)
Eskişehir Osmangazi Üniversitesi
Eskişehir Sağlık Bilimleri Fakültesi
Halk Sağlığı Hemşireliği Anabilim Dalı

Prof. Dr. Setenay DİNÇER
ÖNER
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Biyostatistik Anabilim Dalı

Prof. Dr. Hilmi ÖZDEN
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Anatomi Anabilim Dalı

Prof. Dr. Varol ŞAHİNTÜRK
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Histoloji ve Embriyoloji Anabilim Dalı

Doç.Dr.Timuçin KAŞIFOĞLU
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
İç Hastalıkları Anabilim Dalı /
Romatoloji Bilim Dalı

Doç. Dr. Altan EŞİZOĞLU
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Ruh Sağlığı ve Hastalıkları
Anabilim Dalı

Doç.Dr.Ömer KILIÇ
Eskişehir Osmangazi Üniversitesi
Tıp Fakültesi
Çocuk Sağlığı ve Hastalıkları
Anabilim Dalı /
Çocuk Enfeksiyon Hast. Bilim Dalı

Doç. Dr. Batu Can YAMAN
Eskişehir Osmangazi Üniversitesi
Diş Hekimliği Fakültesi
Restoratif Diş Tedavisi
Anabilim Dalı

Prof. Dr. Bekir YAŞAR
Genel Cerrahi Uzmanı

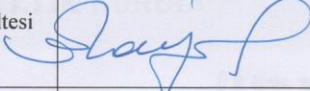
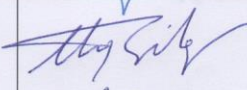

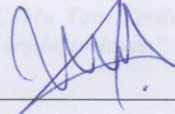
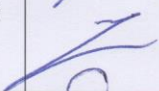
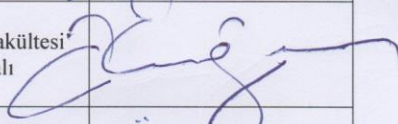
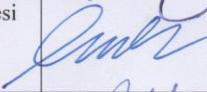
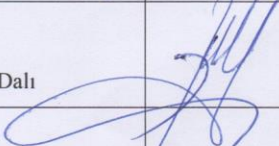
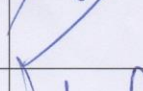
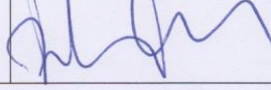
Av. Önder CAN
Avukat

Etik Kurul Sekreterliği
Aysun SERTTAŞ
Makbule SARIÇİÇEK
Tel: 0 222 239 29 79 / 4690

ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
GİRİŞİMSEL OLMAYAN KLİNİK ARAŞTIRMALAR ETİK KURULU BAŞKANLIĞI
KARAR FORMU

Başvuru Tarihi: 17.07.2017	Çalışmanın Başlığı: “Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması” Çalışmacılar: Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı - Doç.Dr.Ertuğrul ÇOLAK (Doktora Tez Danışmanı), Arş.Gör.Muzaffer BİLGİN (Doktora Tez Sahibi), Sağlık Bilimleri Üniversitesi Tepecik Eğitim ve Araştırma Hastanesi Kadın Hastalıkları ve Doğum Bölümü - Op.Dr.Aşkın DOĞAN
Çalışmanın değerlendirildiği ilk toplantı tarihi: 31.07.2017	Sonuç: <ol style="list-style-type: none">1. Kişisel bilgi kişi hak ve özgürlüklerini yakından ilgilendirmesi sebebiyle özellikle korunması gereken bir kavramdır. Kişisel bilgi başlı başına bir kişi hakkı olarak değerlendirildiği gibi aynı zamanda kişinin özel hayatının gizliliği kavramı ile de yakından bağlantılıdır.2. Tıbbi kayıtlar kanun ve yönetmeliklerde belirtilen şekillerde gizliliği ve güvenilirliği korunarak gereksiz ayrıntılar içinde kaybolmadan özenli bir şekilde dikkat gösterilerek doğruluğundan emin olunarak kaydedilmelidir.3. Kişisel veri özelliğinde olduğu için gizliliği yetkili olan birimlerce kişinin özel hayatına saygı olarak etkin bir biçimde korunmalı, bu verilerin yetkisiz kişilerin eline geçmesi engellenmelidir.4. Araştırmacılar İyi Klinik Uygulamalar Kılavuzu (2015), Dünya Tıp Birliği Helsinki Bildirgesi (2013)’ni imzalamış bu imza ile sorumluluğu kabul ettiklerini beyan etmişlerdir.5. Ulusal ve uluslararası mevzuatta dile getirilen “Tıbbi verilerin gizliliğine azami dikkat esastır.”
Karar Tarihi: 31.07.2017 Karar No: 14	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Doç.Dr.Ertuğrul ÇOLAK (Doktora Tez Danışmanı) sorumluluğunda yürütülen “Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması” başlıklı çalışmanın yapılmasının etik açıdan uygun olduğuna oy birliğiyle karar verilmiştir. Araştırmacılara başarılar dileriz.

ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
GİRİŞİMSEL OLMAYAN KLİNİK ARAŞTIRMALAR ETİK KURULU

	Unvanı/Adı/Soyadı	Kurumu	İmza
1	Prof. Dr.Ömür ŞAYLIGİL (Başkan)	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Tıp Tarihi ve Etik Anabilim Dalı	
2	Doç. Dr. Uğur BİLGE (Başkan Yardımcısı)	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Aile Hekimliği Anabilim Dalı	
3	Doç. Dr. Özlem ÖRSAL (Raportör)	Eskişehir Osmangazi Üniversitesi Eskişehir Sağlık Bilimleri Fakültesi Halk Sağlığı Hemşireliği Anabilim Dalı	
4	Prof. Dr. Setenay DİNÇER ÖNER	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Biyostatistik Anabilim Dalı	
5	Prof. Dr. Hilmi ÖZDEN	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Anatomi Anabilim Dalı	
6	Prof. Dr. Varol ŞAHİNTÜRK	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Histoloji ve Embriyoloji Anabilim Dalı	
7	Prof. Dr. Timuçin KAŞİFOĞLU	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi İç Hastalıkları Anabilim Dalı / Romatoloji Bilim Dalı	
8	Doç. Dr. Altan EŞSİZOĞLU	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Ruh Sağlığı ve Hastalıkları Anabilim Dalı	
9	Doç. Dr. Ömer KILIÇ	Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı/ Çocuk Enfeksiyon Hastalıkları Bilim Dalı	
10	Doç. Dr. Batu Can YAMAN	Eskişehir Osmangazi Üniversitesi Diş Hekimliği Fakültesi Restoratif Diş Tedavisi Anabilim Dalı	
11	Prof. Dr. Bekir YAŞAR	Genel Cerrahi Uzmanı	
12	Av. Önder CAN	Hukuk	

ÖZGEÇMİŞ

MUZAFFER BİLGİN

ARAŞTIRMA GÖREVLİSİ

2013-2018 yılları arasında özgeçmiş



E-Posta Adresi : mbilgin@ogu.edu.tr
Telefon (İş) : 2222392979-4617
Telefon (Cep) : 5556321597
Faks :
Adres : Eskişehir Osmangazi Üniversitesi Tıp
Fakültesi Temel Tıp Bilimleri Biyoistatistik
Anabilim Dalı

Projelerde Yaptığı Görevler:

1. Kahve İçmenin Obez ve Olmayanlarda Glukoz ve Metabolizmasını Etkileyen Hormonlar Üzerine Akut Etkisi, Yükseköğretim Kurumları tarafından destekli bilimsel araştırma projesi, Araştırmacı, , 07/04/2015 - 19/02/2016 (ULUSAL)
2. Öğrencilerin Fonksiyonel Olmayan İnanç ve Uygulamalara İlişkin Görüşleri, Yükseköğretim Kurumları tarafından destekli bilimsel araştırma projesi, Araştırmacı: AÇIKGÖZ AYFER, Yürütücü: KAYA YELİZ, Araştırmacı: ÇAKŞAK AYLİN, Araştırmacı: BİLGİN MUZAFFER, , 28/06/2016 - 02/06/2017 (ULUSAL)
3. Türkiye Kronik Hastalık Bakımını ve Yaşam Kalitesi ile İlişkinisi ve İlaç Uyumunu Değerlendirme Çalışması, Yükseköğretim Kurumları tarafından destekli bilimsel araştırma projesi, Araştırmacı, , 08/01/2015 (Devam Ediyor) (ULUSAL)

Eserler

Uluslararası hakemli dergilerde yayımlanan makaleler:

1. ERENOĞLU SON NAZAN, TAŞKIRAN BENGÜR, YORULMAZ GÖKNUR, SON OSMAN, KÖKOĞLU BURCU, BİLGİN MUZAFFER, BİLGE UĞUR (2017). RELATIONSHIP BETWEEN THE 6-ITEM MORISKY MEDICATION ADHERENCE SKALA (MMAS-6) SCORE AND GLYCEMIC CONTROL IN A TURKISH POPULATION WITH TYPE 2 DIABETES MELLITUS. Acta Medica Mediterranea, 33(863), 921-928., Doi: 10.19193/0393-6384_2017_6_146 (Yayın No: 3554734)
2. GÖKALP HİLAL, BULUR IŞIL, KAYA ERDOĞAN HİLAL, KARAPINAR TEK DEN, BİLGİN MUZAFFER (2017). Evaluation of the effects of acne vulgaris on quality of life in Turkey by using TAQLI. Journal of Cosmetic Dermatology, 16(4), 485-490., Doi: 10.1111/jocd.12294 (Yayın No: 3483714)
3. BULUR IŞIL, KAYA ERDOĞAN HİLAL, ÇİFTÇİ EVRİM, CANAZ FUNDA, YİĞİTASLAN SEMRA, YILDIZ PINAR, SARAÇOĞLU ZEYNEP NURHAN, BİLGİN MUZAFFER (2017). Evaluating the Role of Neurotrophins in the Psoriasis and Metabolic Syndrome Relationship. Turkish Journal of Dermatology / Türk Dermatoloji Dergisi, 11(4), 168-173., Doi: 10.4274/tdd.3320 (Yayın No: 3855997)

4. ÖZTAŞ DİLEK,BİLGE UĞUR,KESKİN AHMET,YILDIZ PINAR,KILIÇ SEDA,SELÇUK ENGİN BURAK,BİLGİN MUZAFFER (2017). Sleep disorders in type 2 diabetes mellitus. Does vitamin D have roles on sleep disorders and metabolic parameters?. Biomedical Research, 615-620. (Yayın No: 3858096)
5. YILDIZ PINAR,AYDIN YUNUS,BİLGİN MUZAFFER,HASSA HİKMET (2017). Evaluation of in vitro Fertilization Planned Patients in terms of Internist Perspective. Erciyes Tıp Dergisi/Erciyes Medical Journal, 39(2), 54-58., Doi: 10.5152/etd.2017.16101 (Yayın No: 3569327)
6. BALCIOĞLU HÜSEYİN,ÜNLÜOĞLU İLHAMİ,BİLGİN MUZAFFER (2017). Evaluation of the Treatment Methods of Patients with Mechanical Low Back Pain Complaints Admitted to Family Physician's Clinic. Ankara Medical Journal, 17(2), 119-126. (Yayın No: 3542211)
7. BALCIOĞLU HÜSEYİN,ÜNLÜOĞLU İLHAMİ,BİLGİN MUZAFFER (2017). Aile Hekimliği Polikliniğine Mekanik Bel Ağrısı Nedeni ile Başvuran Hastaların Tercih Ettikleri Tedavi Yöntemlerinin Değerlendirilmesi. Ankara Med J(2), 119-126., Doi: 10.17098/amj.302448 (Yayın No: 3681135)

Uluslararası hakemli dergilerde yayımlanan makaleler:

8. BİLGE UĞUR,ÜNLÜOĞLU İLHAMİ,GÜLTEKİN SARI GÖZDE,MENGÜLLÜOĞLU NEHİR,YAŞAR BİLGE NAZİFE ŞULE,BİLGİN MUZAFFER (2017). The Acute Effect of Coffee Intake on Hormones that Affect Glucose Metabolism in Healthy Individuals. Biomedical Research, 513-515. (Yayın No: 3681088)
9. ERENOĞLU SON NAZAN,TAŞKIRAN BENGÜR,YORULMAZ GÖKNUR,OSMAN SON,KÖKOĞLU BURCU,BİLGİN MUZAFFER (2017). Relationship Between The 6-Item Morisky Medication Adherence Scale (MMAS-6) Score And Glycemic Control In A Turkish Population With Type 2 Diabetes Mellitus.. Acta Medica Mediterranea, 33, 921-928., Doi: DOI: 10.19193 (Yayın No: 3680963)
10. YAŞAR BİLGE NAZİFE ŞULE,BİLGE UĞUR,KILIÇ SEDA,BİLGİN MUZAFFER (2017). The Assesment of Sleep Quality in Rheumatoid Arthritis Patients. Biomedical Research, 410-413. (Yayın No: 4149400)
11. YAŞAR BİLGE NAZİFE ŞULE,BİLGE UĞUR,kılıç seda,BİLGİN MUZAFFER (2017). The assesment of sleep quality in rheumatoid arthritis patients. Biomedical Research (Yayın No: 3791758)
12. YAŞAR BİLGE NAZİFE ŞULE,KILIÇ SEDA,BİLGE UĞUR,BİLGİN MUZAFFER (2016). Romatoid Artrit Hastalarında Hastalık Aktivitesi ile Alerjik Hastalık Sıklığı İlişkisinin Değerlendirilmesi. Ankara Medical Journal, 16(4), 332-338., Doi: 10.17098/amj.48109 (Yayın No: 3409877)
13. YAŞAR BİLGE NAZİFE ŞULE,GÜLER SEDA,BİLGİN MUZAFFER,BİLGE UĞUR (2016). Comparison of the Frequency of Sleep Disorders Among Patients with Rheumatoid Arthritis and Type 2 Diabetic Patients. Konuralp Tıp Dergisi, 8(2), 100-103. (Yayın No: 3410079)
14. Gökalp Hilal,BULUR IŞIL,KAYA ERDOĞAN HİLAL,Karapınar Tekden,BİLGİN MUZAFFER (2016). Evaluation of The Effects of Acne Vulgaris on Quality of Life in Turkey by Using TAQLI. Journal of Cosmetic Dermatology, 1-6., Doi: 10.1111/jocd.12294 (Yayın No: 3409981)
15. Doğan Aşkın,Gülhan İbrahim,Uyar İbrahim,Ekin Atalay,Gezer Cenk,BİLGİN MUZAFFER,Taner Cüneyt E,Ertay İbrahim E (2016). Methotrexate Treatment in Progressive Tubal Ectopic Pregnancies and hCG-Related Clinicosurgical Implications. The Kaohsiung Journal of Medical Sciences, 32(6), 317-322., Doi: 10.1016/j.kjms.2016.05.004 (Yayın No: 3409847)

16. DUR ŞADİYE,GÖZEN DUYGU,BİLGİN MUZAFFER (2016). Devlet ve Özel Hastanedeki Hemşirelerin Aile Merkezli Bakıma İlişkin Tutum ve Davranışları. *Güncel Pediatri Dergisi*, 14(1), 1-9., Doi: 10.4274/jcp.67699 (Yayın No: 3409915)
17. Çilekar Murat,YORULMAZ GÖKNUR,BALCIOĞLU HÜSEYİN,BİLGİN MUZAFFER,BİLGE UĞUR (2016). Renal Function Improvement in the Patients Who Had Sleeve Gastrectomy. *BIOMEDICAL RESEARCH-INDIA*, 27(2), 526-529. (Yayın No: 3410029)
18. KESKİN AHMET,BİLGE UĞUR,ÜNALACAK MURAT,Güler Seda,YILDIZ PINAR,SELÇUK ENGİN BURAK,BİLGİN MUZAFFER (2016). Effects of Sleep Disorders and 25-OH Vitamin D Levels on HbA1c Levels in Geriatric Type 2 Diabetic Patients. *BIOMEDICAL RESEARCH-INDIA*, 27(2), 537-542. (Yayın No: 3410087)
19. ÖZŞEKER BURAK,YAŞAR NECDET FATİH,BİLGİN MUZAFFER,Kurt Yasemin,BALCIOĞLU HÜSEYİN,BİLGE UĞUR (2016). Turkish Validation of National Institutes of Health (NIH) Patient-Reported Outcomes Measurement Information System (PROMIS®) Gastroesophageal Reflux Disease (GERD) Scale. *BIOMEDICAL RESEARCH-INDIA*, 27(2), 577-581. (Yayın No: 3410047)
20. BİLGE UĞUR,ÜNLÜOĞLU İLHAMİ,Kurt Yasemin,Yüce Ayşe,BİLGİN MUZAFFER (2016). Eskişehir Osmangazi Üniversitesi Aile Hekimliği Polikliniğine Başvuran Hastaların Boy ve Kilo Farkındalıklarının ve Fiziksel Egzersiz Durumlarının Değerlendirilmesi. *Ankara Medical Journal*, 16(3), 285-292., Doi: 10.17098/amj.01682 (Yayın No: 3409931)
21. YORULMAZ GÖKNUR,GÖNÜLLÜ EMEL,AKALIN AYSEN,Şenol Murat,YILDIZ PINAR,Tomus Narguler,Akcan Esra,BİLGİN MUZAFFER (2015). The Effect of Non-Autoimmune Thyrotoxicosis to Vitamin D Status During Pregnancy. *BIOMEDICAL RESEARCH-INDIA*, 27(1), 225-228. (Yayın No: 3410135)
22. SON OSMAN,SON NAZANERENOĞLU,BİLGE UĞUR,SOĞUKPINAR CENK,YORULMAZ GÖKNUR,ÇİLEKER MURAT,BİLGİN MUZAFFER (2015). Evaluation of Patients Who Went Under Sleeve Gastrectomy Retrospective Evaluation of 3 Groups of Patients. *Biomedical Research*, 27(1), 872-877. (Yayın No: 1755916)
23. KESKİN AHMET,ÜNALACAK MURAT,BİLGE UĞUR,YILDIZ PINAR,GÜLER SEDA,SELÇUK ENGİN BURAK,BİLGİN MUZAFFER (2015). Effects of Sleep Disorders on Hemoglobin A1c Levels in Type 2 Diabetic Patients. *Chinese Medical Journal*, 128(24), 3292-3297., Doi: 10.4103/0366-6999.171415 (Yayın No: 1756518)
24. BOYRAZ İLKER,BİLGE UĞUR,ÜNALACAK MURAT,BİLGİN MUZAFFER (2015). A Comparison of 25 OH Vitamin D Levels in Patients with Type 2 Diabetes on Oral Hypoglycemic Agents and Insulin Treatment. *Biomedical Research*, 27(1), 806-811. (Yayın No: 1755534)
25. DOĞAN NAZİLE BİLGİN,ÖZPELİT EBRU,AKDENİZ SELMA,BİLGİN MUZAFFER,BARIŞ NEZİHİ (2015). Simple Clinical Risk Score for No-Reflow Prediction in Patients Undergoing Primary Percutaneous Coronary Intervention with Acute STEMI. *Pakistan Journal of Medical Sciences*, 31(3), 576-581.,

Uluslararası hakemli dergilerde yayımlanan makaleler:

26. Kibar Melike,AKTAN ŞEBNEM,BİLGİN MUZAFFER (2015). Dermoscopic findings in scalp psoriasis and seborrheic dermatitis Two new signs Signet ring vessel and hidden hair. *Indian Journal of Dermatology*, 60(1), 41-45., Doi: 10.4103/0019-5154.147786 (Yayın No: 1808636)

27. BİLGE UĞUR,GÜLTEKİN SARI GÖZDE,BİLGİN MUZAFFER,ÜNLÜOĞLU İLHAMİ (2015). Homeostatic Model Assessment Insulin Resistance HOMA IR Değerleri ile Glomerüler Filtrasyon Hızları Arasındaki İlişki Retrospektif Bir Çalışma. Ankara Medical Journal, 15(4), 220-225., Doi: 10.17098/amj.18395 (Yayın No: 1759522)
28. ACAR NURDAN,ÖZÇELİK HAMİT,ÇEVİK ARİF ALPER,ÖZAKIN ENGİN,YORULMAZ GÖKNUR,KEBAPÇI MEDİNE NUR,BİLGE UĞUR,BİLGİN MUZAFFER (2014). Low perfusion index affects the difference in glucose level between capillary and venous blood. Therapeutics and Clinical Risk Management, 10, 985-991., Doi: 10.2147/TCRM.S73359 (Yayın No: 1758613)
29. GEZER CENK,EKİN ATALAY,ÖZEREN MEHMET,TANER CÜNEYT EFTAL,ÖZER ÖZGE,KOÇ ALTUĞ,BİLGİN MUZAFFER,GEZER NACİYE SİNEM (2014). Chromosome abnormality incidence in fetuses with cerebral ventriculomegaly. Journal of Obstetrics & Gynaecology, 34(5), 387-391., Doi: 10.3109/01443615.2014.896885 (Yayın No: 1758926)
30. KİBAR MELİKE,AKTAN ŞEBNEM,LEBE BANU,BİLGİN MUZAFFER (2014). Current understanding of the process of tooth formation transfer from the laboratory to the clinic. Australian Dental Journal, 59, 48-54., Doi: 10.1111/adj.12102 (Yayın No: 1759331)
31. KİBAR MELİKE,AKTAN ŞEBNEM,BİLGİN MUZAFFER (2014). Scalp Dermatoscopic Findings in Androgenetic Alopecia and Their Relations with Disease Severity. Annals of Dermatology, 26(4), 478-484., Doi: 10.5021/ad.2014.26.4.478 (Yayın No: 1759273)
32. Ertas Ibrahim Egemen, Taskin Salih, Goklu Rifat, Bilgin Muzaffer, Goc Goksu, Yildirim Yusuf, Ortac Firat (2014). Long term oncological and reproductive outcomes of fertility sparing cytoreductive surgery in females aged 25 years and younger with malignant ovarian germ cell tumors. Journal of Obstetrics and Gynaecology Research, 40(3), 797-805., Doi: 10.1111/jog.12253 (Yayın No: 580413)
33. Kibar Melike, Aktan Şebnem, Lebe Banu, Bilgin Muzaffer (2013). Trichoscopic findings in alopecia areata and their relation to disease activity severity and clinical subtype in Turkish patients. Australasian Journal of Dermatology, Doi: 10.1111/ajd.12102 (Yayın No: 580290)

B. Uluslararası bilimsel toplantılarda sunulan ve bildiri kitaplarında (proceedings) basılan bildiriler :

1. BİLGİN MUZAFFER,ÇOLAK ERTUĞRUL (2017). Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması. XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi, 74 (Özet Bildiri/Sözlü Sunum)(Yayın No:3808128)
2. BİLGİN MUZAFFER,ÇOLAK ERTUĞRUL (2017). The position of Turkey in the European Union and Candidate Countries according to Mothers' Index. 1st International Women Research Symposium Eskisehir, 112-113. (Özet Bildiri/Sözlü Sunum)(Yayın No:3796918)
3. BULUR İŞİL,BÜLBÜL BAŞKAN EMEL,ÖZDEMİR MUSTAFA,BALEVİ ALİ,KOCATÜRK GÖNCÜ EMEK,ALTUNAY İLKNUR,GÖNÜL MÜZEYYEN,ERGİN CAN,ERTAM İLGEN,SOLAK BERNA,KAYA ERDOĞAN HİLAL,BİLGİN MUZAFFER (2017). Bulur I, Bülbül Başkan E, Özdemir M, Balevi A, Kocatürk Göncü E, Altunay I, Gonul M, Ergin C, Ertam I, Solak B, Kaya Erdogan H, Bilgin M. The efficacy of omalizumab in refractory chronic spontaneous urticaria: The real

life experience from Turkey.. 26 th EADV Congress, 13-17 September 2017, Geneva, Switzerland. (Özet Bildiri/Poster)(Yayın No:4055629)

4. ayan sümeyye,ÜSKÜDAR CANSU DÖNDÜ,ÜSKÜDAR TEKE HAVA,KAŞİFOĞLU TİMUÇİN,AKAY OLGA MELTEM,BİLGİN MUZAFFER,KORKMAZ CENGİZ (2017). CAN WE PREDICT THROMBOTIC TENDENCY IN RHEUMATOID ARTHRITIS: A THROMBOELASTOGRAPHIC ANALYSIS. EULAR 2017 (Tam Metin Bildiri/Poster)(Yayın No:3469108)
5. KAYHAN MEHMET,MAMUR AHMET,BİLGE UĞUR,SUNGUR SERKAN,BALCIOĞLU HÜSEYİN,BİLGİN MUZAFFER,ÜNLÜOĞLU İLHAMİ (2017). NIH (Ulusal Sağlık Enstitüleri), PROMIS (Hasta Bildirimli Sonuç Ölçüm Sistemi) Gastrointestinal Semptom Ölçeği- Bulantı/Kusma Türkçe Güvenirlik Skalası. 16th International Eastern Mediterranean Family Medicine Congress, 128-129. (Özet Bildiri/Poster)(Yayın No:3518834)
6. MAMUR AHMET,KAYHAN MEHMET,BİLGE UĞUR,SUNGUR SERKAN,BALCIOĞLU HÜSEYİN,BİLGİN MUZAFFER,ÜNLÜOĞLU İLHAMİ (2017). NIH (Ulusal Sağlık Enstitüleri), PROMIS (Hasta Bildirimli Sonuç Ölçüm Sistemi) Gastrointestinal Semptom Ölçeği-Diyare Türkçe Güvenirlik Skalası. 16th International Eastern Mediterranean Family Medicine Congress, 124-125. (Özet Bildiri/Poster)(Yayın No:3518827)
7. ÖZÜTÜRK ÖZGE,ZALOĞLU AKİF,BİLGİN MUZAFFER,YILDIZER KERİŞ ELİF (2017). Maksiller Posterior Diş Patolojileri ile Maksiller Sinüs Patolojileri Arasındaki İlişkinin Konik IşınlıBilgisayarlı Tomografi ile İncelenmesi. Oral Diagnoz veMaksillofasiyal Radyoloji Derneği2. Uluslararası Kongresi”7. Bilimsel Toplantısı” (Özet Bildiri/Poster)(Yayın No:3485288)
8. BALCIOĞLU HÜSEYİN,ÜNLÜOĞLU İLHAMİ,BİLGİN MUZAFFER (2017). AİLE HEKİMLİĞİ POLİKLİNİĞİNE MEKANİK BEL AĞRISI NEDENİ İLE BAŞVURAN HASTALARIN TERCİH ETTİKLERİ TEDAVİ YÖNTEMLERİNİN DEĞERLENDİRİLMESİ. 6. INTERNATIONAL TRAKYA FAMILY MEDICINE CONGRESS, 86 (Özet Bildiri/Poster)(Yayın No:3463099)
9. KAYA ERDOĞAN HİLAL,BULUR İŞİL,KOCATÜRK EVİN,SARAÇOĞLU ZEYNEP NURHAN,ALATAŞ İBRAHİM ÖZKAN,BİLGİN MUZAFFER (2017). Oxidative stress in rosacea: A case control study. 2nd International Dermatology and Cosmetology Congress (Özet Bildiri/Sözlü Sunum)(Yayın No:4114042)
10. BİLGE UĞUR,ÜNALACAK MURAT,KESKİN AHMET,YILDIZ PINAR,GÜLER SEDA,BİLGİN MUZAFFER (2015). Effect of 25 OH Vitamin D Levels on HbA1C Levels in an Elderly Population with Type 2 Diabetes Mellitus. 20th Wonca Europe Conference 2015, İstanbul, Turkey (/)(Yayın No:2374291)
11. BİLGİN MUZAFFER,ÇOLAK ERTUĞRUL,Özdamar Kazım (2015). Comparison of Best Linear Combination and Logistic Regression Methods for Multiple Continuous Diagnostic Tests Using ROC Curve Analysis. 36th Annual Conference of the International Society for Clinical Biostatistics (ISCB 2015) (/)(Yayın No:1809315)
12. GÜLTEKİN SARI GÖZDE,BİLGE UĞUR,ŞİMŞEK NURGÜL,BİLGİN MUZAFFER,ÜNLÜOĞLU İLHAMİ (2015). Relationship Between Homeostatic Model Assessment Insulin Resistance HOMA IR values and Glomerular Filtration Rates GFR A Retrospective Study. 4th Conference AGPFMSEE, 51 (/)(Yayın No:1809453)
13. BİLGE UĞUR,GÜLTEKİN SARI GÖZDE,BİLGİN MUZAFFER,ÜNLÜOĞLU İLHAMİ (2015). Insulin Resistance Rates in a Population Under 40 Years Old. 4th Conference AGPFMSEE, 73 (/)(Yayın No:1809506)

14. Yorulmaz Gökür,AKALIN AYSEN,Tomus Narguler,Şenol Murat,Yıldırım Pınar,GÖNÜLLÜ EMEL,BİLGİN MUZAFFER (2015). The Effect of Thyrotoxicosis to Vitamin D Status During Pregnancy. The Endocrine Society's 97th Annual Meeting & Expo (Özet Bildiri/)(Yayın No:1808843)
15. ÖZAKIN ENGİN,ÖZDEMİR ATİLLA ÖZCAN,Şahin Deniz Gören,AKAY OLGA MELTEM,Tehli Yasemin,ÇEVİK ARİF ALPER,ACAR NURDAN,BALOĞLU KAYA FİLİZ,BİLGİN MUZAFFER (2014). Thrombelastography Can It Predict Intracerebral Bleeding After Tpa Treatment. 4th Eurasian Congress on Emergency Medicine (Özet Bildiri/)(Yayın No:1809161)
16. BOZDEMİR ÖZEL CEMİLE,DEMİRTAŞ RAZİYE NESRİN,ARIKAN HÜLYA,ÇELER ÖZGEN,AKALIN AYSEN,BİLGİN MUZAFFER (2014). Metabolik Sendromlu Kadınlarda Egzersiz Kapasitesi ile Santral Obezitenin İlişkisi. Uluslararası Kardiyometabolik Sendrom Doğu Akdeniz Kongresi ve Doğu Akdeniz Sağlık Bilimleri Öğrenci Sempozyumu (/)(Yayın No:2398730)
17. Yorulmaz Gökür,AKALIN AYSEN,Tomus Narguler,Şenol Murat,Yıldırım Pınar,BİLGİN MUZAFFER (2014). Vitamin D Levels in Pregnants with Gestational Thyrotoxicosis. The Endocrine Society's 96th Annual Meeting & Expo (Özet Bildiri/)(Yayın No:1808982)

D. Ulusal hakemli dergilerde yayımlanan makaleler :

1. GÜRSOY HALUK HÜSEYİN,BİLGEÇ MUSTAFA DEĞER,BİLGİN MUZAFFER (2016). Okul Çağı Çocuklarında Sikloplejinin Optik Biyometriye Etkisi. Türkiye Klinikleri Journal of Ophthalmology, 25(4), 245-250., Doi: 10.5336/ophthal.2016-50289 (Kontrol No: 3409949)
2. GEZER NACİYE SİNEM,GÜLERYÜZ HANDAN,GEZER CENK,KOÇYİĞİT ALİ,YEŞİLİRMAK DİDEM,EKİN ATALAY,BİLGİN MUZAFFER,ERTAŞ İBRAHİM EGEMEN (2015). The Prognostic Role of Prenatal MRI Volumetric Assessment in Fetuses with Isolated Ventriculomegaly. The Turkish Journal of Pediatrics, 57(3), 266-271. (Kontrol No: 1756374)

E. Ulusal bilimsel toplantılarda sunulan ve bildiri kitaplarında basılan bildiriler:

1. CANSU GÜVEN BARIŞ,ÜSKÜDAR CANSU DÖNDÜ,TAŞKIRAN BENGÜR,YAŞAR BİLGE NAZİFE ŞULE,BİLGİN MUZAFFER,KORKMAZ CENGİZ (2017). Romatoloji pratiğinde orta-yüksek doz steroide bağlı hiperglisemiyi saptayabilmek için en uygun kan şeker ölçüm zamanı sabah mıdır?. XVIII.Ulusal Romatoloji Kongresi (Tam Metin Bildiri/Poster)(Yayın No:3629013)
2. SARI YE,BİLGE UĞUR,BALCIOĞLU HÜSEYİN,BİLGİN MUZAFFER,KAYHAN M,ÖZTAŞ DİLEK,ÜNLÜOĞLU İ (2017). Sağlık Çalışanlarının Kardiyovasküler Risk Düzeylerinin Sağlık Çalışanı Olmayan Diğer Meslek Gruplarına Göre Karşılaştırılması. Aile Hekimliği Güz Okulu,, 131-131. (Özet Bildiri/Poster)(Yayın No:3832744)
3. SARI YUNUS EMRE,BİLGE UĞUR,BALCIOĞLU HÜSEYİN,BİLGİN MUZAFFER,KAYHAN MEHMET,ÖZTAŞ DİLEK,ÜNLÜOĞLU İLHAMİ (2017). Sağlık çalışanlarının kardiyovasküler risk düzeylerinin sağlık çalışanı olmayan diğer meslek gruplarına göre karşılaştırılması. 11. Aile Hekimliği Güz Okulu, 131 (Özet Bildiri/Poster)(Yayın No:3597541)
4. ACER ERSOY,KAYA ERDOĞAN HİLAL,SARAÇOĞLU ZEYNEP NURHAN,İĞREK ALİ,PARLAK HATİCE,BİLGİN MUZAFFER (2017). Erken saç beyazlaması ile ilişkili risk faktörlerinin değerlendirilmesi. V. DOD Dermatoloji Gündemi (Özet Bildiri/Sözlü Sunum)(Yayın No:3857367)

5. ACER ERSOY,KAYA ERDOĞAN HİLAL,İĞREK ALİ,PARLAK HATİCE,SARAÇOĞLU ZEYNEP NURHAN,BİLGİN MUZAFFER (2017). Erken saç beyazlaması ile ilişkili risk faktörlerinin değerlendirilmesi.. V. DOD Dermatoloji Gündemi (Özet Bildiri/Sözlü Sunum)(Yayın No:4114140)
6. ACER ERSOY,KAYA ERDOĞAN HİLAL,İĞREK ALİ,PARLAK HATİCE,SARAÇOĞLU ZEYNEP NURHAN,BİLGİN MUZAFFER (2017). Erken saç beyazlaması ile ilişkili risk faktörlerinin değerlendirilmesi.. V. DOD Dermatoloji Gündemi (Özet Bildiri/Sözlü Sunum)(Yayın No:4148165)
7. Gökalp Hilal,BULUR IŞIL,KAYA ERDOĞAN HİLAL,KARAPINAR TEKDEN,BİLGİN MUZAFFER (2015). Akne Vulgariste Yaşam Kalitesi 960 Olgunun Değerlendirilmesi. XXII. Prof. Dr. A. Lütfü Tat Sempozyumu (Tam Metin Bildiri/)(Yayın No:1811916)
8. Bilge N Şule Yaşar,BİLGE UĞUR,Güler Seda,BİLGİN MUZAFFER (2015). Romatoid Artrit Hastalarının Uyku Bozukluğu ve Tedavi Uyum Açısından Değerlendirilmesi. XVI. Ulusal Romatoloji Kongresi, 7(2), 163 (/)(Yayın No:1812104)
9. ACAR NURDAN,DOLGUN HAKAN,ÖZAKIN ENGİN,Çanakçı Mustafa Emin,Kaya Filiz Baloğlu,BİLGİN MUZAFFER,TEPE MURAT (2015). Hastayı Tedavi Et Amilaz Testini Değil Kendi Acil Laboratuvarımızdaki Amilaz Düzeyinin Kesim Değerinin Pankreatit Tanısı Koymak İçin Yeniden Düzenlenmesi. 11. Türkiye Acil Tıp Kongresi, 90 (Tam Metin Bildiri/)(Yayın No:1811952)
10. ÖZAKIN ENGİN,DÖNER EGEMEN,SOLMAZ ELİF REYHAN,Keşcioğlu Ferhat,ACAR NURDAN,BALOĞLU KAYA FİLİZ,SİVRİKOZ MUAMMER CUMHUR,BİLGİN MUZAFFER,ÇEVLİK RIDVAN (2015). Acil Servise Başvuran Spontan Pnömotoraks Hastalarında Pnömotoraks Yüzdesi ve Vital Bulgular Arası İlişkinin Değerlendirilmesi. 11. Türkiye Acil Tıp Kongresi, 93 (Tam Metin Bildiri/)(Yayın No:1811985)
11. BULUR IŞIL,KAYA ERDOĞAN HİLAL,Kocatürk Evin,SARAÇOĞLU ZEYNEP NURHAN,ALATAŞ İBRAHİM ÖZKAN,YILDIZ PINAR,BİLGİN MUZAFFER (2015). Plak Tip Psoriasis Vulgaris Hastalarında Serum İrisin Düzeyi ve İnsülin Direnci İlişkinin Değerlendirilmesi. IV. DOD Dermatoloji GÜndemi (Tam Metin Bildiri/)(Yayın No:1812011)
12. KESKİN AHMET,ÜNALACAK MURAT,BİLGE UĞUR,YILDIZ PINAR,Güler Seda,SELÇUK ENGİN BURAK,BİLGİN MUZAFFER (2015). Geriatrik Tip 2 Diabetik Hastalarda Uyku Bozukluğu Sıklığı Ve HbA1C Düzeylerine Etkisi. 9. Aile Hekimliği Güz Okulu, 143 (/)(Yayın No:1812141)
13. BİLGE UĞUR,ÜNALACAK MURAT,KESKİN AHMET,YILDIZ PINAR,Güler Seda,SELÇUK ENGİN BURAK,BİLGİN MUZAFFER (2015). Diabetik Hastalarda 25OH Vitamin D nin Uyku Bozuklukları Ve Metabolik Parametreler Üzerine Etkisi. 9. Aile Hekimliği Güz Okulu, 45 (/)(Yayın No:1812143)
14. KESKİN AHMET,ÜNALACAK MURAT,BİLGE UĞUR,YILDIZ PINAR,Güler Seda,SELÇUK ENGİN BURAK,BİLGİN MUZAFFER (2015). Tip 2 Diabetik Hastalarda Görülen Uyku Bozukluklarının Metabolik Etkileri. 9. Aile Hekimliği Güz Okulu, 85 (/)(Yayın No:1812164)
15. DUR ŞADİYE,GÖZEN DUYGU,BİLGİN MUZAFFER (2014). Devlet ve Özel Hastanedeki Hemşirelerin Aile Merkezli Bakıma İlişkin Tutum ve Davranışları. 8. Ege Pediatri ve 4. Ege Pediatri Hemşireliği Kongresi (/)(Yayın No:2398800)
16. Giray Sevgi,BİLGİN MUZAFFER,Serçe İmran,Yıldız Çiğdem Nahide,Kızrak Sümeyya Betül,Haneci Güzide,Karabatak Zehra,Sevinçer Şeyma (2014). Eskişehir Sağlık Yüksek Okulu Ebelik Bölümü Öğrencilerinin Premenstrüel

Sendrom Düzeyleri. I. Uluslararası - 5. Ulusal Ebelik Öğrenci Kongresi (Tam Metin Bildiri/)(Yayın No:1812071)

17. Giray Sevgi,BİLGİN MUZAFFER,GÜNEŞ GİZEM,Cesur Güler,Dönertaş Özlem,Oleroğlu Burcu,Uzun Fatma (2014). Adolesan Öğrencilerin Cinsel Sağlık ve Üreme Sağlığı Hakkındaki Bilgi Düzeyleri ve Bunları Etkileyen Faktörler. I. Uluslararası - 5. Ulusal Ebelik Öğrenci Kongresi (Tam Metin Bildiri/)(Yayın No:1812044)
18. YEŞİLOVA ESRA,YAŞAR FÜSUN,BİLGİN MUZAFFER (2014). Total Dişsiz Hastaların Klinik Profili 274 Hasta Üzerinde Yapılan Bir Çalışma. Oral Diagnoz ve Maksillofasial Radyoloji Derneği 6. Ulusal Sempozyumu ve 1. Uluslararası Katılımlı Kongresi (/)(Yayın No:2398503)
19. Mustafa Çağatay Büyükuysal, Muzaffer Bilgin, Fürüzan Köktürk (2013). Farklı Örneklem genişlikleri ve Dağılımlarda Normal Dağılım Testlerinin Monte Carlo Simülasyonu İle Karşılaştırılması. Uluslararası Katılımlı 15. Ulusal Biyoistatistik Kongresi (/)(Yayın No:577506)
20. BÜYÜKUYSAL MUSTAFA ÇAĞATAY,BİLGİN MUZAFFER,KÖKTÜRK FÜRÜZAN (2013). Farklı Örneklem Genişlikleri ve Dağılımlarda Normal Dağılım Testlerinin Monte Carlo Simülasyon ile Karşılaştırılması. Uluslararası Katılımlı 15. Ulusal Biyoistatistik Kongresi, 109-109. (Tam Metin Bildiri/)(Yayın No:2398968)

Kurs

1. Bilgisayar Uygulamalı Yapısal Eşitlik Modellemesi Kursu, Eğitim Kursu, Eskişehir, Kurs, 18.02.2015 -20.02.2015 (Ulusal)
2. Bilgisayar Uygulamalı Yapısal Eşitlik Modelleri (Structural Equation Modelling (SEM)), Eğitim Kursu, Eskişehir, Kurs, 26.08.2014 -28.08.2014 (Ulusal)
3. Başarılı Sunum Yapma Teknikleri, Eğitim Kursu, Eskişehir, Kurs, 25.03.2013 - 26.03.2013 (Ulusal)

Çalıştay

1. Mixed Models with Medical Applications, Erasmus Teaching Activity, İzmir, Çalıştay, 27.05.2013 -28.05.2013 (Uluslararası)

