

**T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**ŞİZOFRENİ VE BİPOLAR MANİK BOZUKLUĞUN RAMAN
SPEKTROSKOPİK VE KEMOMETRİK ANALİZ YÖNTEMLERİYLE
AYRIŞTIRILMASI**

**YÜKSEKLİSANS TEZİ
HAYRUNNİSA NUR KABUK
1102040003**

Anabilim Dalı: Fizik

Program: Fizik

Tez Danışmanı: Prof. Dr. Gülce ÖĞRÜÇ İLDİZ

OCAK 2021

**T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**ŞİZOFRENİ VE BİPOLAR MANİK BOZUKLUĞUN RAMAN
SPEKTROSKOPİK VE KEMOMETRİK ANALİZ YÖNTEMLERİYLE
AYRIŞTIRILMASI**

**YÜKSEKLİSANS TEZİ
HAYRUNNİSA NUR KABUK
1102040003**

Anabilim Dalı: Fizik

Program: Fizik

**Tez Danışmanı: Prof. Dr. Gülce ÖĞRÜÇ İLDİZ
Jüri: Prof. Dr. Olcay BÖLÜKBAŞI YALÇINKAYA
Doç. Dr. Ayberk YILMAZ**

OCAK 2021

ÖNSÖZ

Bu tez çalışmasında bana türlü fedakarlık gösteren, maddi ve manevi desteğini esirgemeyen, çalışmalarımın yönlendirilmesi, sonuçların değerlendirilmesi ve yazımı aşamasında en büyük katkı sahibi değerli tez hocam Prof. Dr. Gülce ÖĞRÜÇ İLDİZ'a en içten duygularıyla minnettarlığımı sunuyorum.

Çalışma süresi boyunca bana sağladıkları destek ve sonsuz sabırlarından ötürü sevgili babam İbrahim ve annem Hatice Kabuk'a ayrıca canım ablam Ayşe Hümeysra Kabuk ve güzel kardeşim Hüsna Şule Arslan'a, çalışma arkadaşlarıma, özellikle Gizem Yavuz'a ve tüm gerçek dostlarıma teşekkürlerimi sunuyorum.

Ocak, 2021

Hayrunnisa Nur KABUK

İÇİNDEKİLER

İÇİNDEKİLER	II
KISALTMALAR	IV
TABLO LİSTESİ	V
ŞEKİL LİSTESİ.....	VI
SİMGE LİSTESİ.....	VII
ÖZET.....	VIII
ABSTARCT	IX
1. GİRİŞ.....	- 1 -
2. MATERYAL VE METOD.....	- 6 -
2.1. Klinik Aşama	- 6 -
2.1.1. Hasta ve Kontrol Seçimi.....	- 6 -
2.1.2. Numune Hazırlama.....	- 7 -
2.2. Spektroskopik Aşama	- 8 -
2.2.1. Raman Spektroskopisi.....	- 8 -
2.2.1.1. Kuantum Mekaniksel Olarak Raman Spektroskopisinin Açıklanması	- 8 -
2.2.1.2. Klasik Görüş Altında Raman Spektroskopisinin Açıklanması..	- 10 -
2.2.2. Raman Spektrumlarının Kaydedilmesi	- 14 -
2.2.3. Spektrumların Ön İşlenmesi.....	- 14 -
2.3. İstatistiksel Aşama	- 16 -
2.3.1. Temel Bileşenler Analizi (PCA).....	- 17 -
2.3.2. Hiyerarşik Kümeleme Analizi	- 19 -
2.3.3. Kısmi En Küçük Kareler Ayırıştırma Analizi (PLS-DA)	- 21 -

2.3.4. İstatistiksel Model Performansının Değerlendirilmesi.....	- 23 -
2.3.5. Sınıflandırma Modellerinin Geliştirilmesi ve Test Edilmesi.....	- 24 -
3. BULGULAR	- 26 -
3.1. Kan Raman Spektrumlarının Ön Analizi.....	- 26 -
3.2. Raman Spektrum Tabanlı Sınıflandırma Modellerinin Geliştirilmesi	- 29 -
3.3. Modellerin Tahmin Yeteneğinin Test Edilmesi.....	- 34 -
4. TARTIŞMA VE SONUÇ	- 40 -
KAYNAKLAR	- 41 -



KISALTMALAR

SZ	: Şizofreni
BP	: Bipolar
K	: Kontrol Grubu
PCA	: Temel Bileşenler Analizi
PLS-DA	: Kısmi En Küçük Karelere Ayırıştırma
DSM	: Ruhsal Bozuklukların Tanısal ve İstatiksel El Kitabı
WHO	: Dünya Sağlık Organizasyonu
FTIR	: Fourier Dönüşümlü Kızılötesi Spektroskopisi
SST	: Serum Ayırma Tüpü
RMSE	: Hata Karaler Ortalamasının Karekökü
SCI	: Bilim Alıntı Endeksi
PANSS	: Pozitif ve Negatif Semptom Ölçeği
YMRS	: Young Mani Derecelendirme Ölçeği
NIPALS	: Doğrusal Olmayan Yinelemeli En Küçük Kareler Algoritması
EMSC	: Genişletilmiş Çarpımsal Dağılım Düzeltme Yöntemi
DP	: Doğru Pozitif
DN	: Doğru Negatif
YP	: Yanlış Pozitif
YN	: Yanlış Negatif

TABLO LİSTESİ

- Tablo 3.1.** : Kan serumu Raman spektrumundaki temel bantların işaretlemeleri...-27-
- Tablo 3.2.** : Karışıklık (Confusion) Tablosu ve 3S Modeli için istatistiksel model performans göstergeleri.....-35-
- Tablo 3.3.** : Karışıklık (Confusion) Matrisi ve 3S Modeli için istatistiksel model performans göstergeleri.....-35-
- Tablo 3.4.** : Karışıklık (Confusion) Tabloları ve 2 / 2S Modeli için istatistiksel model performans göstergeleri.....-36-

ŞEKİL LİSTESİ

- Şekil 2.1.** : Rayleigh, Stokes ve Anti Stokes Enerji Seviye Diyagramı.....-10-
- Şekil 2.2.** : Atomlar arası uzaklık.....-13-
- Şekil 2.3.** : 2 boyutlu bir veri kümesinin (değişkenler: X, Y) iki ana ekseninin (PC1 ve PC2) yönlerinin tanımı.....-19-
- Şekil 2.4.** :A-E örnekleri için hiyerarşik kümeleme dendrogramının açıklayıcı örneği.....-20-
- Şekil 3.1.** : Kontrol (K) grubunun kan serumunun ortalama Raman Spektrumu, kontrol (K) grubunun kan serumunun ortalama Raman Spektrumunun, SZ (siyah) ve BP(M) (kırmızı) gruplarının normalleştirilmiş ortalama Raman spektrumlarından çıkararak elde edilen fark spektrumları....-26-
- Şekil 3.2.** : BP(M), SZ ve K grubunun ortalama Raman Hiyerarşik kümeleme analizi dendrogramı.....-28-
- Şekil 3.3.** : SZ, BP(M) ve K gruplarının alan normalize ortalama kan serumu Raman spektrumlarının PCA skorları.....-28-
- Şekil 3.4.** : 2/2S Modelinin ikili adımları.....-30-
- Şekil 3.5.** : 3S Modeli için 3D skor grafiği.....-31-
- Şekil 3.6.** : 2/2S Modelin ilk PLS-DA (1) adımı için skor grafiği.....-32-
- Şekil 3.7.** : 2/2S Modelin ilk PLS-DA (1) adımı için 3D skor grafiği.....-33-
- Şekil 3.8.** : 2/ 2S Modelinin ikinci adımı PLS-DA(2) için 3D skor grafiği.....-34-
- Şekil 3.9.** : 3S Modelindeki (BP(M)) örnekler için tahmin edilen Y değerleri..-37-
- Şekil 3.10.** : 3S Modelindeki (K) örnekler için tahmin edilen Y değerleri.....-37-
- Şekil 3.11.** : 3S Modelindeki (SZ) örnekler için tahmin edilen Y değerleri.....-37-
- Şekil 3.12.** : 2/2S Modelinin 1. adımındaki (SZ/BP(M)) örnekler için tahmin edilen Y değerleri.....-38-
- Şekil 3.13.** : 2/2S Modelinin 1. adımındaki (K) örnekler için tahmin edilen Y değerleri.....-38-
- Şekil 3.14.** : 2/2S Modelinin 2. adımındaki (SZ) örnekler için tahmin edilen Y değerleri.....-38-
- Şekil 3.15.** : 2/2S Modelinin 2. adımındaki (BP(M)) örnekler için tahmin edilen Y değerleri.....-39-

SİMGE LİSTESİ

cm	: Santimetre
ml	: Mililitre
μ l	: Mikrolitre
$h\nu_0$: Elastik Saçılma
t	: Zaman
ΔE	: Molekülün İki Enerji Seviyesi Arasındaki Fark
h	: Planck Sabiti
λ	: Dalga Boyu
E	: Elektrik Alan
E_2	: Üst Enerji Seviyesi
E_1	: Alt Enerji Seviyesi
α	: Kutuplanabilirlik Katsayısı
α_0	: Molekülün Kutuplanabilme Yatkınlığı
$\vec{\mu}$: İndüklenmiş Dipol Moment
\vec{E}	: Uygulanan Elektrik Alan Vektörü
α_d	: Molekülün Denge Konumunda Kutuplanabilirliği
Q	: Titreşim Koordinatı
r	: Çekirdekler Arası Uzak
Σ	: Kovaryans Matrisi
δ	: Standart Sapma
ν	: Frekans
ν_0	: Gelen Işığın Frekansı
ν_{tit}	: Molekülün Titreşim Frekansı
I	: Şiddet
$I_{antistokes}$: Anti Stokes Çizgilerinin Şiddeti
I_{stokes}	: Stokes Çizgilerinin Şiddeti

ÖZET

Şizofreni (SZ) ve Bipolar duyu durum bozukluğu (BP), dünyada yaklaşık olarak nüfusun %2'sini etkileyen nöropsikiyatrik sendromları olan ve bu sendromların çok çeşitlilik gösterdiği ve dolayısıyla tanılanmaları oldukça zor olan hastalıklardır. Bipolar bozukluk; manik epizod BP(M), depresif epizod ve ötimik olmak üzere 3 fazda gözlenmektedir. Bipolar bozukluğun manik fazının Şizofreni ile benzer semptomlar göstermesi bu iki bozukluğun ayrıştırılmasını zorlaştırmakta ve tanılarında güçlükler yaşanmasına neden olmaktadır. Şizofren ve Bipolar hastalıkların tanısında, hastalar ile birebir yapılan görüşmelerde ortaya çıkan bazı sübjektif verilere dayanılarak, sonuca gidilmeye çalışılmaktadır. Hâlihazırda bu hastalıkların tanısı için geliştirilmiş, klinik tanı yöntemlerini destekleyecek nitelikte moleküler tabanlı herhangi bir tanı yöntemi bulunmamaktadır.

Bu tez çalışması çerçevesinde SZ ve BP(M) hastalarını birbirinden ve sağlıklı bireylerden ayırabilen, hastaların kan serumunun Raman spektrumları analizine dayanan, SZ ve BP(M)'nin yardımcı ayırıcı tanısı için analitik bir model geliştirildi. Bu model, Raman spektrumlarının X-bağımsız değişkenlerinin olduğu iki sınıflı, iki PLS-DA adımının hiyerarşik bir dizimine dayanmaktadır. Analizler sonucunda 400-3100 cm^{-1} Raman spektroskopik aralığının bu bozukluklar için hassas bir spektroskopik biyobelirteç olarak çalıştığı sonucuna varıldı. Önerilen metodoloji güvenilir, hızlı, ucuz, portatif ve klinik ortamda kolayca uygulanabilir niteliktedir.

ABSTARCT

Schizophrenia (SZ) and Bipolar mood disorder (BP) are neurological diseases that have neuropsychiatric syndromes and together affecting approximately 2% of the world population. Due to diversity of their syndromes, they are difficult to be diagnosed. BP exhibits three distinct phases: manic and depressive episodes and the euthymic phase. The manic phase of bipolar disorder showing similar symptoms with schizophrenia which makes it difficult to differentiate these two disorders and causes difficulties in their diagnosis. Diagnostic approaches are based on constructed interviews and subjective evaluations of clinical symptoms of these disorders. Currently, there is no molecular-based diagnostic method developed for the diagnosis of these diseases that supports clinical diagnostic methods.

Within the framework of this thesis study, an analytical model was developed for the auxiliary differential diagnosis of SZ and BP(M), based on the analysis of patients' blood serum Raman spectra, which can distinguish SZ and BP(M) patients from each other and from healthy individuals. This model is based on a hierarchical sequence of two PLS-DA steps of two classes, where spectra are X -predictor variables. As a result of the analysis, it was concluded that the 400-3100 cm^{-1} Raman spectroscopic range works as a sensitive spectroscopic biomarker for these disorders. The propounded methodology is reliable, fast, inexpensive, portable, and easily applicable in a clinical environment.

1. GİRİŞ

Şizofreni (SZ); kişinin düşüncesini, hareketlerini, duygularını ifade şeklini, gerçeği algılamasını çarpıtan, delüzyon, halüsinasyon, davranış ve konuşma bozukluğu, duygu eksikliği gibi belirtilerle karakterize edilen zihinsel bir rahatsızlıktır. Psikiyatrik tanı sınıflamalarında geçmiş yıllarda yer alan şizofreninin alt tipleri (paranoid, dezorganize, katatonik, ayrışmamış ve rezidüel tipler), sınırlı tanısallıkları, düşük güvenilirlikleri ve zayıf geçerlilikleri nedeniyle yeni sınıflamalarda bulunmamakta ve tek bir hastalık olarak değerlendirilmektedir (DSM-5, 49-51; DSM-IV, 2000; Carpenter vd., 2013). Şizofreninin etiyolojisi hakkında bilinenler oldukça azdır. Günümüze kadar elde edilen bulgular SZ'nin metabolizma değişimi, hormonal regülasyonlar ve bağışıklık durumuyla ilişkili heterojen bir rahatsızlık olduğunu göstermektedir (Tomasik, 2012).

Bipolar bozuklukta (BP) şizofreni gibi etiyolojisi tam aydınlatılmamış kronik bir zihinsel bozukluk olup, şizofrenide suçlanan biyolojik yapılar burada da suçlanmaktadır (Tomasik v.d. 79; Bahn v.d. 2011; Torrey; Scola). Bipolar bozukluk, bütün dünyada her 100 kişiden ikisini etkileyen, dolayısıyla da nispeten sık görülen bir bozukluktur. Bipolar bozukluğu olan kişi, sıklıkla duygu durumunda aşırı yükselmelerden çöktürlere, tekrar yükselmelere dönüşen ve çoğu zaman aralarda normal duygu durum dönemleri bulunan dalgalanmalar yaşar (Strakowski, 2003; Mueser, vd., 1998; Yang, vd., 2013; Tas, vd., 2014; Torrey, vd., 1999).

Bipolar bozuklukta alt gruplarının olduğu ve bu alt grupların kliniğinin ve seyrinin farklılık gösterdiği bilinmektedir. Bipolar bozukluk DSM-5 tanı kriterlerine göre, Bipolar Bozukluk tip I, Bipolar Bozukluk tip II ve Siklotimi olmak üzere üç alt gruba ayrılmaktadır (DSM, 2013; Nusslock, vd., 2014; Bellivier, vd., 2013). Bununla birlikte Bipolar bozukluğun her bir alt tipi manik, depresif ve ötimik olmak üzere 3

fazda gözlenmektedir (DSM-IV, 2000; DSM-5, 2013; Nusslock, vd., 2014; Bellivier, vd., 2013).

SZ ve BP dünya nüfusunun %2'sinden fazlasını etkilemektedir. Bu iki hastalık kişilerin yaşam kalitelerini büyük ölçüde etkilemekte ve Dünya Sağlık Organizasyonunun (WHO) yayınladığı rapora göre dünya genelinde kişilerin çalışma hayatından alıkoyan hastalıklar arasında, sırasıyla 6'ncı ve 8'inci sırada yer almaktadır (Pearlson 251; Berrettini 59; WHO). Her iki rahatsızlıkta, düşünce, ruh hali, algı ve davranış bozukluğu gibi benzer belirtilerle karakterize edilmektedir (Correll, vd., 2007).

Bu hastalıkların tedavisinin gecikmeden yapılması gerekmesine karşın, benzer belirtilere sahip olmaları nedeniyle, SZ veya BP teşhisinin konması hastaların uzun yıllar boyunca takibini gerektirmektedir (Torrey, vd., 1999; Tas, vd., 2014; Loobo vd., 2012; Murray vd., 2004). Bipolar bozukluğun ataklar arası dönemde hiçbir belirti göstermeksizin tam remisyon halinde seyreden doğası geçmiş atak dönemleri iyi sorgulanmadığında, sağlıklı bireylerle ayırımın yapılmasını neredeyse imkansız kılmaktadır. Şizofreni, sıklıkla bipolar bozukluğun manik fazıyla benzer semptomlar göstermektedir. Bu durum bu iki hastalığın tanısının karıştırılmasına yol açmaktadır.

Yapılan çalışmalar DSM tanı kriterlerine göre bipolar teşhisi konulan hastaların %12'sinin bir yıl sonunda yanlış teşhis edildiğini ve bu hastalara şizofren tanısı konduğunu göstermektedir. Benzer şekilde bazı hastaların önceden şizofreni tanı kriterlerini karşılarken yıllar içinde yanlış teşhis edildiği görülerek bipolar tanısı almıştır. Tüm bu nedenlerden ötürü iki rahatsızlığın birbirinden ayırt edilmesi yüz yılı aşkın bir süredir psikiyatristlerin çözüme ulaştırmaya çalıştıkları bir sorun olarak karşımıza çıkmaktadır (Strakowski, 2003; Torrey vd., 1999; Nusslock vd., 2014; F Meyer v.d. 174; Gozales-Pinto v.d. 41; Joyce 145).

Bu mevcut durumu iyileştirmenin olası bir yolu, psikolojik ve davranışsal klinik değerlendirmeden bağımsız yardımcı tanı yöntemleri kullanmak olacaktır. Bu yöntemler, analitik ve güvenilir bir teknik kullanılarak izlenebilecek, SZ ve BP ile ilişkili spesifik biyobelirteçlere dayanmalıdır. Bu iki zihinsel rahatsızlığa metabolizma, hormonal ve bağışıklık sistemindeki değişimlerin eşlik ettiği

bilinmektedir (Tomasik v.d. 2012; Scola vd., 2014; Yang vd., 2013; Lobo, 2012; Schwartz, 2010). Günümüze kadar büyük çabalara rağmen SZ ve BP(M) bozukluğunun ayırıştırılmasına yönelik moleküler tabanlı biyobelirteçler geliştirilememiştir.

SZ ve BP temel olarak beyinsel rahatsızlıklar olarak kabul edilmekle birlikte son yıllarda periferal sistemleri etkileyen sistematik bozukluklarla da ilişkilendirilmektedir. Bu bağlamda kan merkezi ve periferal fonksiyonları regüle eden molekülleri içermesi nedeniyle sistemsel değişimlerden etkilenmektedir. Yapılan çalışmalarda bu tip değişimlerin gözlenebileceği, hem teşhis hem de hastalığın seyrinin takibi amaçlı kullanılabilir en umut verici malzemelerden birinin merkezi ve periferal fonksiyonlarla ilişkili molekülleri içeren, kan olduğu sonucuna varılmıştır. (Bahn vd., 2011; Torrey, 1999; Tomasik vd., 2012).

Spektroskopik yöntemlerin, özellikle Raman spektroskopisinin, vücut sıvılarında, özellikle de kandaki biyomoleküler ortamı araştırmak için güçlü araçlar olduğu gösterilmiştir (Kanagathara vd., 2011; Ahmed, 2010). Spektroskopik yöntemler, organik moleküllerin kalitatif ve yapısal analizinde kullanılmaktadır. Bu yöntemler vasıtasıyla bu tip moleküllerdeki yapısal değişimler belirlenebilmektedir. Raman titreşim spektroskopik yöntemi günümüzde plazma, serum, doku, salya ve ürün gibi biyolojik örneklerin kalitatif analizinde güçlü bir araç haline gelmiştir (Kanagathara vd., 2011; Shiek 2010).

Raman spektroskopisi, tek dalga boylu görünür ışığın maddeyle etkileşmesinin incelenmesi prensibine dayanır. Bu teknik moleküler yapının belirlenmesinde parmak izi niteliğindeki bilgiyi sağlar. Son yıllarda Raman spektroskopisinin, hastalıkların tespiti ve teşhisi için klinik teşhis metodlarına tamamlayıcı bir teknik olarak kullanımı giderek yaygınlaşmıştır. Raman spektroskopik teknikler kullanılarak yapılan çalışmalar sonucunda, çok miktarda spektral veri elde edilmektedir. Bu çerçevede fazla sayıdaki verinin analizi için istatistiksel yöntemlerin kullanımına gerek duyulmaktadır. Raman spektroskopisi özellikle kemometri gibi uygun veri analizi yöntemleriyle birleştiğinde, klinik ortamda tamamlayıcı tanı aracı olarak yer kazanmaktadır (Ralbovsk vd., 2019, Ramesh vd, 2019).

Spektral verilerin analizinde kullanılan kemometrik yöntemlerin başında PCA (Temel Bileşen Analizi) ve PLS-DA (Kısmi En küçük kareler ayrıştırma analizi) yöntemleri gelmektedir (Sakudo vd., 2006; Dingari vd., 2012; Dreissig vd., 2009; Travo vd., 2014; Dong vd., 2014; Sheng vd., 2013; Wan vd., 2014; Andrade vd., 2007; Zainal vd., 2011). 20 yılı aşkın bir süredir bu iki yöntem etkili biçimde spektroskopik verilerin analiz edilmesinde ve belirli fonksiyonel gruplarla ilişkili spektral değişimlerin belirlenmesinde kullanılmaktadır. Literatürde PCA ve PLS-DA istatistiksel analiz yöntemlerinin titreşimsel spektroskopik yöntemlerle birlikte uygulandığı birçok çalışma yer almaktadır.

Bu çalışmalar öncelikle Onkoloji ile sınırlandırılmış olmakla beraber bunamanın en yaygın ve en hızlı büyüyen tipi olan Alzheimer hastalığı ve Parkinson hastalığı tanısı için potansiyel bir araç olarak da kullanılmaktadır (Shiek, 2010; Sudworth, 2005; Bright, 2011). Parkinson hastalığı, merkezi sinir sisteminin nörolojik olarak dejeneratif bozukluğudur. Beynin hareket ve koordinasyondan sorumlu olan bölgelerine bilgi gönderen nörotransmitter dopamin'in üretiminde azalmaya sebep olan bazı nöronların ölmesi ya da hatalı çalışması sonucu bu hastalık ortaya çıkmaktadır. Amerika'da 52 hastanın (20'si Parkinson belirtileri göstermekte olan) kırmızı kan hücreleri, kan plazması, trombosit ve kan pıhtıları titreşimsel spektroskopik yöntemlerle incelenmiştir. Yapılan çalışmalar sonucunda oksidatif stres ve serbest radikallerden zarar görmüş protein, lipit (yağ) ve nükleik asitlerdeki metabolik değişimlerle ilişkili, soğurma ve şiddet değerlerinde değişim gösteren spektral bölgeler spektroskopik olarak tespit edilmiştir. Böylelikle Parkinson, hastalarına özgü metabolik imza niteliğinde sonuçlar elde edilmiştir. Bu yöntemler, sadece Parkinson hastalığının tanısında değil, aynı zamanda hastalığın seyrinde de takip amaçlı olarakta kullanılabilir (Shiek, 2010; Wikonson, 2007).

Ağızda oluşan tümörlerle ilişkili yapılan başka bir çalışmada dişteki kolajenlerdeki karbonhidratlarla ilişkili C-O titreşimlerinde ve nükleik asitlerle (DNA) ilişkili fosfat grubuna ait titreşimlerdeki değişimlerin bu hastalık tipi için biyoişaret niteliğinde olduğu belirlenmiştir (Conti vd., 2005). Benzer şekilde bir başka çalışma da kan kanseri hastaları için gerçekleştirilerek, yine hastalığın tespitinde biyoişaret olarak kullanılabilir spektral bölgeler tespit edilmiştir (Erukhimovitch vd., 2006).

FTIR spektroskopik ve PCA, PLS-DA istatistik analiz yöntemleri az sayıda kan örneği üzerinden, Şizofren ve Bipolar hastalıkların sınıflandırılmasında kullanılmıştır [Ogruc Ildiz vd., 2015]. Tüm bu çalışmalar çok az miktarda serum ve plazma kan örnekleri üzerinden titreşimsel spektroskopik yöntemler kullanılarak gerçekleştirilmiştir.

Günümüze kadar yapılan bu tip çalışmaların hepsi, Titreşimsel Spektroskopik yöntemlerin ileriye dönük olarak hastalıkların tanı ve takibinde kullanılabileceğini göstermekte ve klinik yöntemlere alternatif bir yöntem olarak kullanılması hususunda umut vaat etmektedir.

Mevcut literatür tarandığında Alzheimer ve Parkinson gibi bazı nörodejeneratif hastalıklarda benzer yöntemlerle yapılmış olan çalışmalar ve ön çalışma niteliğinde FTIR spektroskopisi kullanılarak, az sayıda kan örneği üzerinden SZ ve BP'nin ayrıştırılmasına ilişkin çalışma olmasına karşın bu tez çalışması çerçevesinde incelenecek olan ve birbiriyle sıklıkla karışan Bipolar bozukluğun manik fazı ve Şizofreni hastalığının birbirinden ve sağlıklı kontrollerden ayrıştırılmasına yönelik Raman spektroskopik ve kemometrik analiz yöntemleri kullanılarak moleküler düzeyde yapılmış bir çalışmaya rastlanmamıştır.

Bu tez çalışması çerçevesinde, hastaların kan serumu Raman spektrumlarının analizine dayanan SZ ve BP'nin ayırıcı tanısı için, klinisyenlerin teşhisine tamamlayıcı nitelikte modeller geliştirildi. Geliştirilen nihai analitik model BP'nin manik fazı ve SZ'yi birbirlerinden ve sağlıklı kontrollerden ayrıştırılmasına olanak sağlamaktadır. Oluşturulan model bize $400-3100 \text{ cm}^{-1}$ aralığındaki spektroskopik bölgenin SZ ve BP(M) bozukluklarını ayırtmada etkin bir araç görevi gördüğünü ve spektroskopik biyobelirteç olarak kullanılabileceğini göstermektedir.

Aşağıda bölümlerde ayrıntılı olarak açıklandığı gibi, metodoloji güvenilir, hızlı, ucuz, minimal invazivdir ve klinik ortamda kolaylıkla uygulanabilir. Bu tez çalışması içinde bursiyer olarak yer aldığım yürütücü tez danışmanım Prof. Dr. Gülce Öğrüş Ildız'ın olduğu 116Z294 numaralı TÜBİTAK 1001 çalışma çerçevesinde elde edilen sonuçlar SCI (Science Citation Index) kapsamında taranan 2.79 etki faktörlü Journal of Raman Spectroscopy dergisinde yayınlamıştır. (Ildiz, vd., 2020)

2. MATERYAL VE METOD

2.1. Klinik Aşama

2.1.1. Hasta ve Kontrol Seçimi

Çalışma, Türkiye İstanbul Sağlık Bilimleri Üniversitesi Kanuni Sultan Süleyman Araştırma ve Eğitim Hastanesi ve Marmara Üniversitesi Pendik Eğitim ve Araştırma Hastanesi tedavi görmekte olan, klinik olarak DSM- 5 tanı kriterlerine göre Şizofreni (SZ) veya Bipolar Manik Epizod (BP(M)) olarak tanı konulmuş hastalardan toplanan kan örnekleri üzerinden gerçekleştirilmiştir. Kan örneklerinin alınacağı hastaların seçimi, temini ve serumlarının ayrıştırılması işlemleri İstanbul Sağlık Bilimleri Üniversitesi Kanuni Sultan Süleyman Araştırma ve Eğitim Hastanesi ve Marmara Üniversitesi Pendik Eğitim ve Araştırma hastanesi, Psikiyatri Bölümleri tarafından gerçekleştirilmiştir.

Tanımlar, Şizofreni ve Bipolar bozukluk için sırasıyla Pozitif ve Negatif Semptom Ölçeği (PANSS) (Opler vd., 2017) ve Young Mani Derecelendirme Ölçeği (YMRS) (Young, vd., 1978) kullanılan ayrıntılı tıbbi ve psikiyatrik yapılandırılmış öykü görüşmelerini içermektedir.

Kan serumları toplam 160 numune olmak üzere, BP(M) grubuna ait 40'ar hasta ile SZ grubuna ait 60'ar hastadan temin edilmiştir. Ayrıca İstanbul Kültür Üniversitesi personeli ve öğrencileri arasından kontrol grubu (K) olarak 60 sağlıklı birey seçilmiştir . Kontrol grubunun seçilen üyelerinin hiçbir psikopatoloji öyküsü yoktur. Çalışma İstanbul Koç Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu tarafından, 24/02/2016 tarihli kurul kararıyla onaylanmıştır.

Tüm çalışmaya katılan hastalara ve gönüllülere çalışmayla ilgili detaylı yazılı bilgilendirme ve çalışmanın ayrıntılı bir açıklaması verilmiş ve izinleri alınmıştır.

Alzheimer ve Parkinson gibi önemli tıbbi öyküsü olan hastalar ve sonuçları etkileyebilecek diyabet, kalp krizi, anjin, kanser ve romatoid artrit hastalıklarına sahip kişiler hariç tutulmuştur. Aynı zamanda alkol ve ilaç kullanım geçmişi olan ve Şizofren ve Bipolar dışında zihinsel rahatsızlık tanısı konmuş hastalar da çalışma dışı bırakılmıştır. Kontrol grubu, herhangi bir kronik fiziksel, nörolojik hastalık ve duyu durum bozukluğu olmayan, sağlıklı gönüllülerden seçilmiştir. Farklı gruplara ait katılımcıların cinsiyet ve yaş dağılımları (18-70 aralığında) aynıdır ve aynı grup içindeki denekler aynı tıbbi tedaviye maruz kalan hastalardan seçilmiştir.

2.1.2. Numune Hazırlama

Hasta ve kontrollerin her birinden temin edilen 5 mL kan numuneleri serumlarını ayırtmak amacıyla Vacutainer plastik SST(Serum Seperater Tube-Serum Ayırıştırıcı Tüp) jelli tüplerere kondu. Tüplerdeki jel, santrifüjden sonra serum ve kan hücreleri arasında fiziksel bir bariyer oluşturmakta ve tüp duvarındaki silika parçacıkları sayesinde serum pıhtılaşmasını hızlandırmaktadır. Tüp çalkalanmadan, kanın, tüpün çeperindeki silika partikülleri ile iyice temas etmesi için 4-5 kez yavaşça altüst edildi. Tüp içerisindeki kan kendiliğinden pıhtılaşana kadar (en az 30 dk.) beklenildi.

Kan numunesinin pıhtılaşması beklenildikten sonra, serumu hücresel materyalden ayırmak için 14.000 rpm'de 10 dakika santrifüj edilerek tüp içerisindeki jelin serum ile kan hücreleri arasında bariyer oluşturması sağlandı. Elde edilen serum örnekleri Eppendorf tüplere konarak -80 °C' de donduruldu.

Raman spektroskopik analizi için, daha önceden dondurulan kan serum numuneleri çözdürüldü, 1 µL serum numunesi bir alüminyum tabaka üzerine yerleştirildi ve başka bir işlem yapılmadan kullanıldı.

2.2. Spektroskopik Aşama

2.2.1.Raman Spektroskopisi

1928 yılında Hintli Fizikçi Sir Chandrasekhra Venteka Raman çıktığı bir deniz yolculuğunda hava kirliliğinden uzak ve temiz bir gökyüzünde çıplak gözle Rayleigh'in bahsettiği saçılmayı gözlemledi ve 1931'de kaynak olarak güneş ışığı dedektörü teleskop ve alıcıyı da kendi gözleri olarak belirlediği gaz moleküllerini ışıkla bombardıman ettiği bir deneyle Nobel ödülüne layık görüldü (Smekal,310-311).

Raman Spektroskopisi moleküler yapıları titreşimsel olarak karakterize eden bir yöntemdir. İncelenen sistemin yakın infrared, görünür ve morüstü bölgede tek frekanslı (genel olarak kaynak lazerdir) elektromanyetik dalga ile ışınlayarak saçılan elektromanyetik dalganın incelenmesi esasına dayalı bir tekniktir.

Raman saçılması sırasında saçılan ışığın enerjisindeki değişim molekülün titreşim enerji düzeyleri arasındaki farka karşılık gelmektedir. Bu yüzden Raman saçılma bantları moleküllerin titreşim ve dönü düzeylerini açıklar ayrıca moleküllerin parmak izleri niteliğindedir. Ölçülen frekans değerleri daha sonra incelenen maddenin veya maddelerin yapısını belirlemek için kullanılır. Fonksiyonel gruplar karakteristik titreşimlere sahip olduklarından ötürü, incelenen sistemlerin moleküllerinin gözlenen frekansları ve yapısal elemanları arasındaki doğrudan ilişki kurabilir. Raman bant şiddetleri ve bant genişlikleri, sırasıyla kimyasal analiz için niceliksel bilgi ve örnek moleküllerdeki dinamik özellikler ve etkileşimler hakkında bilgi sağlar (Öztürk, Tuğçe P.,3, C.V. Raman vd., K.S. Krishnan,121,Carey, P.R.).

2.2.1.1. Kuantum Mekaniksel Olarak Raman Spektroskopisinin Açıklanması

Bir molekül topluluğu $h\nu_0$ enerjili fotonlarla bombardıman edildiğinde fotonlar hedef moleküllerle elastik veya inelastik saçılma gerçekleştirebilirler. Eğer gönderilen foton ve saçılan fotonun enerjisi arasında bir değişim olmazsa bu saçılma inelastik bir saçılmadır. Şayet, molekül ile foton arasında bir enerji alışverişi olmuşsa, çarpışma inelastiktir. Bu durumda saçılan fotonlar gelen fotonlardan daha yüksek yahut daha düşük enerjiye sahip olarak saçılır ve enerjisi;

$$\Delta E = h (\nu_0 \pm \nu_{tit}) \quad (2.1)$$

ifadesine eşit olur. Eğer molekül, başlangıçta taban titreşim enerji seviyesindeyse, fotondan belli bir enerji soğurarak uyarılmış duruma geçer. Bu durumda fotonun enerjisi;

$$\Delta E = h (\nu_0 - \nu_{tit}) \quad (2.2)$$

ifadesine eşittir. Molekül başlangıçta uyarılmış enerji seviyesinde ise fotona enerji aktararak taban duruma geçer. Bu durumda ise saçılan fotonun enerjisi;

$$\Delta E = h (\nu_0 + \nu_{tit}) \quad (2.3)$$

eşit olur.

Burada ν_{tit} , molekülün titreşim frekansıdır. İyi bir yaklaşıklık olarak fotonun kinetik enerjisi ve molekülün kinetik enerjisinin çarpışmadan önce ve sonra aynı kaldığını söyleyebiliriz. Enerjinin korunumu yasası

$$h\nu + E = h\nu' + E' \quad (2.4)$$

olduğunu ifade eder. Burada 'E' molekülün dönme, titreşim ve elektronik enerjisini temsil etmektedir. 'E'' ise aynı enerjilerin çarpışma sonundaki durumunu göstermektedir. Denklem

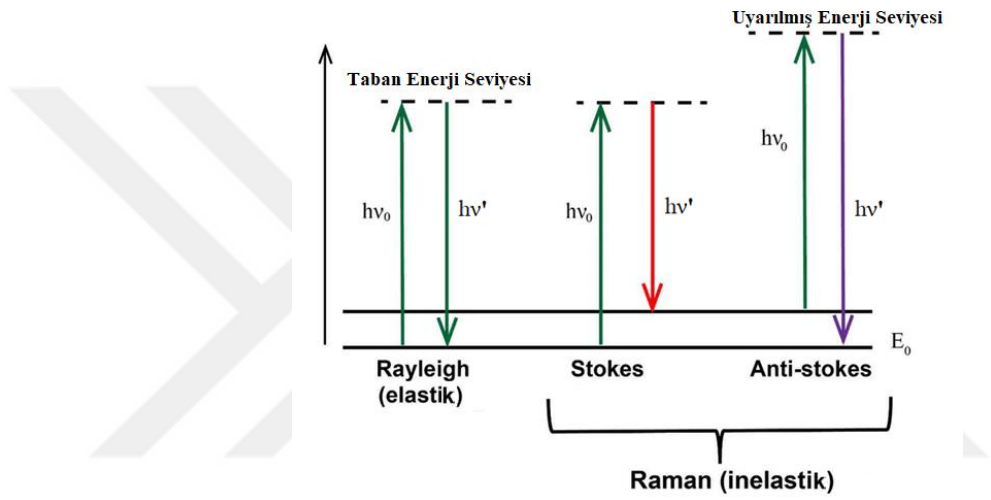
$$\frac{E - E'}{h} = \nu' - \nu \quad (2.5)$$

olur.

Saçılmalarda çıkan enerji,

- $h\nu_0$ (Elastik saçılma) : Rayleigh saçılması, $E = E'$ ($\nu = \nu'$) } Rayleigh
 $h(\nu_0 - \nu_{tit})$: Stokes saçılması, $E > E'$ ise ($\nu < \nu'$) } Raman
 $h(\nu_0 + \nu_{tit})$: Anti-Stokes saçılması $E < E'$ ise ($\nu > \nu'$) } Saçılması

Stokes ve anti-Stokes saçılmaları Raman saçılması olarak bilinir. Bu saçılmalar şematik olarak Şekil 2.1.'de verilmiştir.



Şekil 2.1.: Rayleigh, Stokes ve Anti Stokes Enerji Seviye Diyagramı
(Köksal, F. ve Köseoğlu, 209-216, Yurdakul, Ş.,77-81)

2.2.1.2. Klasik Görüş Altında Raman Spektroskopisinin Açıklanması

Bir molekül topluluğu üzerine ν frekanslı bir elektromanyetik dalga gönderildiğinde, moleküller elektrik alan içerisinde elektron ve çekirdeklerine etkiyen kuvvet sebebiyle değişime uğrar. Bu nedenle eğer molekülün başlangıçta bir elektrik dipol momenti yoksa dipol moment oluşur. Eğer başlangıçta dipol moment varsa, bu dipol momente ilave olarak dipol moment yüklenir ve değişir.

$$\vec{E} = \vec{E}_0 \sin(2\pi\nu t) \quad (2.6)$$

Bu oluşan veya değişime uğrayan (denklem 2.6.) dipol momenti $\vec{\mu}$, elektromanyetik dalganın, elektrik alanının şiddetiyle orantılıdır. İndüklenmiş dipol momenti,

$$\vec{\mu} = \alpha \vec{E} \quad (2.7)$$

eşitliği ile ifade edilir.

Burada;

α : Molekülün kutuplanabilme yatkınlığı

$\vec{\mu}$: İndüklenmiş dipol moment

\vec{E} : Uygulanan elektrik alan vektörünü ifade etmektedir.

Burada α bir tensördür ve $\vec{\mu}$ aşağıdaki şekilde ifade edilir.

$$\begin{matrix} \vec{\mu}_x \\ \vec{\mu}_y \\ \vec{\mu}_z \end{matrix} = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} \begin{bmatrix} \vec{E}_x \\ \vec{E}_y \\ \vec{E}_z \end{bmatrix} \quad (2.8)$$

Yukarıdaki terimler daha açık yazılırsa,

$$\vec{\mu}_x = \alpha_{xx}\vec{E}_x + \alpha_{xy}\vec{E}_y + \alpha_{xz}\vec{E}_z$$

$$\vec{\mu}_y = \alpha_{yx}\vec{E}_x + \alpha_{yy}\vec{E}_y + \alpha_{yz}\vec{E}_z$$

$$\vec{\mu}_z = \alpha_{zx}\vec{E}_x + \alpha_{zy}\vec{E}_y + \alpha_{zz}\vec{E}_z$$

$$E^2 = E_x^2 + E_y^2 + E_z^2$$

(2.9)

eşitlikleri elde edilir.

Denklem (2.8) ve denklem (2.9)'da α_{ij} elemanlarının fiziksel anlamı elektrik alanı herhangi bir doğrultuda ise, dipol momentinin o doğrultularda bileşenleri olabileceğidir. Yani, i doğrultusunda dipol momentini ve j doğrultusunda elektrik alan bileşeni vardır. Kutuplanma tensörü simetriktir, $\alpha_{ij} = \alpha_{ji}$ dir. Öyleyse, bütün $i \neq j$ ler için α_{ij} sıfır olduğu esas eksen sistemine geçebiliriz. Bu sistemde indüklenmiş dipol momentleri aşağıdaki ifadelerle verilir.

$$\begin{aligned}
\vec{\mu}_{x'} &= \alpha_{x'x'} \vec{E}_{x'} \\
\vec{\mu}_{y'} &= \alpha_{y'y'} \vec{E}_{y'} \\
\vec{\mu}_{z'} &= \alpha_{z'z'} \vec{E}_{z'}
\end{aligned}
\tag{2.10}$$

Burada üstler yeni koordinat sistemini göstermektedir, $\alpha_{x'x'} = \alpha_{y'y'} = \alpha_{z'z'}$ olması durumunda molekül izotropiktir.

Dönme ve titreşim hareketleri molekülün elektronik dağılımını devamlı değiştirdiğinden kutuplanma, molekülün denge konumu etrafında Taylor serisine açtığımızda,

$$\alpha = \alpha_d + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q + \frac{1}{2} \left(\frac{\partial^2 \alpha}{\partial Q^2}\right)_0 Q^2 + \dots
\tag{2.11}$$

denklemini elde edilir.

Burada;

α_d : molekülün denge konumunda kutuplanabilirliği

$Q \equiv r - r_d$: titreşim koordinatı

r : herhangi bir anda çekirdekler arası uzaklığı ifade eder.

Eğer molekül ν_0 frekansı ile titreşirse, Q 'da zamanın fonksiyonudur.

$$Q = Q_0 \sin 2\pi \nu_0 t
\tag{2.12}$$

Denklemin 2.11'de ilk iki terimin dışındakileri yok sayarsak, basit harmonik olması durumunda,

$$\alpha = \alpha_d + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q_0 \sin 2\pi \nu_0 t
\tag{2.13}$$

şeklinde yazabiliriz.

Bu durumda indüklenen dipol momentini;

$$\vec{\mu} = \alpha_e \vec{E}_0 \sin 2\pi \nu t + \vec{E}_0 Q_0 \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q_0 \sin 2\pi \nu_0 t \sin 2\pi \nu t
\tag{2.14}$$

ifade ederiz.

Burada, Denklem (2.16)'i elde ederken trigonometride sinüs çarpımı kuralı (2.15),

$$\sin A \sin B = \frac{1}{2} \cos(A-B) - \frac{1}{2} \cos(A+B) \quad (2.15)$$

kullanılarak yeniden düzenlenecek olursa μ indüklenmiş dipol momenti için aşağıdaki ifade elde edilir.

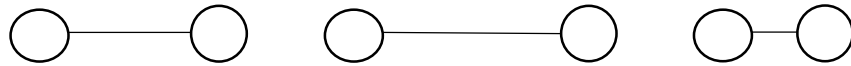
$$= \alpha_e E_0 \sin \underbrace{2\pi\nu t}_{\text{Rayleigh Terimi}} + \frac{1}{2} E_0 Q_0 \left(\frac{\partial \alpha}{\partial Q} \right)_0 [\underbrace{\cos 2\pi(\nu - \nu_0)t}_{\text{Stokes Terimi}} - \underbrace{\cos 2\pi(\nu + \nu_0)t}_{\text{Anti-Stokes Terimi}}] \quad (2.16)$$

Rayleigh Terimi Stokes Terimi Anti-Stokes Terimi

Burada ilk terim Rayleigh saçılmasını, ikinci terim Stokes saçılmasını ve son terim ise anti-Stokes saçılmasını verir. Eğer molekülün titreşimi sırasında α değişiyorsa, o zaman bu titreşim Raman spektroskopisinde gözlenebilir. Buna Raman aktiflik gözlem şartı denir.

$$\text{Raman} \left(\frac{\partial \alpha}{\partial Q} \right) \neq 0 \quad (2.17)$$

ise o titreşim Ramanda gözlenebilir, yani Raman aktiftir. İzotropik iki atomlu moleküllerin moleküler dönü spektrumuyla dönü enerji seviyeleri gözlenemez. Ramanda gözlenebilir. Çünkü şekilde verildiği gibi α , bağ uzunluğuna bağlı olduğundan değişen α 'sı vardır.



Şekil 2.2. : Atomlar arası uzaklık

Atomlar arası uzaklık ne kadar fazla ise kutuplanma yatkınlığı o kadar fazladır.

$\vec{\mu}$, elektromanyetik ışınının titreşen elektrik alanının etkisi ile görüşe göre, gönderilen ışınım ile aynı frekansta dalga yayar. Bu Rayleigh saçılmasının klasik açıklamasıdır. (Çağlayan, Çağlar, 44; Köksal, F. ve Köseoğlu, 210).

2.2.2. Raman Spektrumlarının Kaydedilmesi

Raman spektroskopik analizi için, eppendorf (1 µL) tüplerde daha önceden dondurulan kan serum örnekleri oda sıcaklığında çözündürülerek, mikroskop lamel cam üzerine alüminyum kaplanmış, yüzey üzerine yerleştirilip daha başka bir işleme tabi tutulmadan kullanıldı. Çözünen her numune için, daha önceden oluşturulan karanlık oda içinde bulunan, lazer dalgaboyu 785 nm olan, taşınabilir, B & W-Tek i-Raman Plus-785 model mikro-Raman sistemi (50x objektif)'ni kullanıldı. Sonraki aşamadaki spektral ve istatistiksel analizde kullanılmak için lazer gücü 280 mW olarak ayarlanan Raman Spektrometresiyle, 400-3100 cm⁻¹ arasındaki spektral aralıkta, (30 sn ve 32 tarama entegrasyon süresi kullanılarak) numunenin farklı yerlerinden toplam 5 spektrum kaydedildi.

2.2.3. Spektrumların Ön İşlenmesi

Veri analizinden önce, aşağıda bahsedilen prosedürler kan serumu Raman spektrumlarına uygulanmıştır.

(1) Spektrumlara, baseline correction (taban çizgisi düzeltmesi) - spektrumların sıfır şiddet seviyesini düzeltmek için; bu, özellikle floresans gözlemlendiğimiz Raman spektrumları için etkin bir işlemdir;

(2) Smoothing (Düzeltilme) - Savitzky-Golay veya medyan filtresi gibi algoritmalar kullanarak sinyal-gürültü oranını artırmak için;

(3) Normalleştirme -; spektrumların toplam şiddetini (tüm spektral aralıkta) alan normalizasyonuyla bire normalize ederek birbirleriyle doğrudan karşılaştırılabilir hale getirmek için;

(4) Aykırı değerlerin (Outliers) çıkarılması - farklı nedenlerle yanlış spektrumlara neden olan örnekleri elemek için yapılmıştır. Aykırı değerlerin çıkarılması için, her bir örnekle ilişkili replika spektrumların ortalaması alınarak herbir örnek sınıfı (K, SZ, BP(M)) için, Doğrusal Olmayan Yinelemeli Kısmi En Küçük Kareler (NIPALS) Algoritması (Wold vd., 1975) kullanılarak Temel Bileşen Analizi (PCA) yapıldı. Aykırı değerler, farklı örneklerin PCA skor grafiğindeki konumuna ve

ayrıca uygunluk rezidüellerine ve Hotellings parametrelerine (belirli bir örneğin PCA eksenlerinin belirlenmesine olan etkisinin bir ölçüsünü sağlayan) bakılarak belirlendi.

PCA analizi; çok sayıda, birbiri ile ilişkili değişkenler içeren veri setinin boyutlarını veri içerisinde var olan değişimlerin mümkün olduğunca korunarak daha az boyuta indirgenmesini sağlayan bir dönüşüm tekniğidir. Çok değişkenli verilerle ilgili en önemli sorunlardan biri büyük hacimli verilerin tamamının desen ve ilişkilerinin görülmesini engellemesidir. Temel bileşenler analizinin temel hedefi, aralarında, korelasyon bulunan orijinal değişkenlerin doğrusal kombinasyonları olan esas bileşenleri bularak verilerin boyutunu küçültmektir. Dönüşümler sonrasında elde edilen değişkenler, ilk değişkenlerin temel bileşenleri olarak adlandırılmaktadır. İlk temel bileşen varyans değeri en büyük olandır ve diğer temel bileşenler varyans değerleri azalacak şekilde sıralanır (Uyanık, 2012; Esbensen vd., 2005).

Bu tez çerçevesinde incelenen veriler, Raman spektral verilerdir. PCA analizi vasıtasıyla analiz edilen her bir örneğe ait spektrumun titreşim frekansları (Raman kayması) eksenleri ortagonal temel bileşenlere karşılık gelen PCA skor grafiklerinde, tek bir noktaya indirgenmektedir. Temel bileşenler, verilerin varyansını azalan sırada açıklar: birinci temel bileşen, PC1, verilerdeki varyansın çoğunu ifade ederken, PC2, verilerdeki kalan varyansın en büyük miktarını açıklar ve böyle devam eder. Sonuç olarak, örneklerin benzerliği (veya farklılığı) hakkında bilgi, benzer örneklerin skor grafiğinde kümelenmesiyle elde edilir. Bir skor grafiğindeki iki nokta arasındaki uzamsal ayırmadaki bir artış, bu iki örnek arasındaki farklılıktaki bir artışa, yani Raman spektrumlarındaki farklılığa karşılık gelir. İki PC bileşenini birbirine göre çizilmesiyle elde edilen skor grafiği (Score plot) genellikle örnekler arasındaki ilişkiyi yorumlamak için kullanılır. Ancak eksenlerin 3 ana bileşene karşılık geldiği 3 boyutlu grafikler de kullanılabilir.

Daha önce açıklandığı gibi ön veri analizi çerçevesinde, her örnek türü için PCA kullanılarak, aykırı değerler çıkarıldı ve her grup için ortalama spektrumlar elde edildi. Spektrumlar, 5 noktalı medyan algoritması (Tukey vd.,1977) kullanılarak düzeltildi ve elde edilen ortalama spektrumlar, taban çizgisi düzeltmesine tabi tutuldu ve alan normalize edildi. Her bir sınıfın ortalama spektrumları, daha sonra genişletilmiş çarpımsal dağılım düzeltme (EMSC) yöntemi kullanılarak, karşılık gelen

sınıfın spektrumlarının taban çizgilerini düzeltmek için, referans spektrumları olarak kullanıldı. (Miller, 1990) Veri setindeki tüm spektrumlar nihayet yeniden normalizasyona tabi tutuldu. OMNIC yazılımı (Sürüm 8.3.103) (Omic) ile yapılan farklı sınıflara ait ortalama Raman Spektrumlarının taban düzeltmesi hariç hem PCA analizleri ve veri ön işleme prosedürleri Unscrambler™ CAMO yazılımı (Sürüm 5.5) ile gerçekleştirildi. (Omic; Unscrambler, 1995)

2.3. İstatistiksel Aşama (Çok Değişkenli Veri İstatistik Kümeleme ve Sınıflandırma Kemometri Yöntemleri)

Birçok alanda olduğu gibi, spektroskopik verilerin analizinde çok değişkenli istatistiksel kümeleme ve sınıflandırma yöntemleri çok başarılı bir şekilde uygulanmaktadır. Çok değişkenli istatistiksel yöntemler, ölçülen birden fazla değişkeni içeren her tür gözlem için kullanılmaya uygun olmakla birlikte çok değişkenli çok sayıda örnek aynı anda uğraşırken en kullanışlı yöntemdir. Bir örnek numune için sıklıkla orta kırmızı altı bölgede ($400-4000\text{ cm}^{-1}$) ve genellikle 1 cm^{-1} çözünürlükle kaydedilen, 3600 tane değişken (frekans) ve her bir frekansa karşılık ölçülen kırmızı-altı sinyallerinin şiddetlerini içeren kırmızı altı spektrumu yukarıda bahsi geçen çok değişkenli veriler için uygun bir örnektir. Numune sayısı arttıkça veri seti matrisinin boyutu içerdiği bilgi artmaktadır. Örneğin 200 spektrum (numuler) içeren numune seti dikkate alındığında, bu örnekteki veri seti matrisi 200×3600 boyutunda olmakta, dolayısıyla 720000 bilgi içermektedir.

Çok değişkenli istatistiksel yöntemler, örneklerin n boyutlu dağılımını (örneğin, standart sapmaları, varyansları, kovaryansları ve olasılık dağılımları aracılığıyla) tanımlamak, değişkenler arasındaki ilişkiyi araştırmak, önemli bilgileri kaybetmeden sayılarını azaltmak ve numuneler arasındaki benzerliği (veya farklılığı) belirlemek için kullanılmaktadır. Çok değişkenli veri analizinin ana hedefi, örneklerin gruplandırılması ve sınıflandırılmasıdır. Günümüze kadar çeşitli istatistiksel yöntemler geliştirerek spektroskopik etkili bir şekilde uygulamaktadır. (İldiz, G.O., Fausto, R, vd, 11)

Tezin bu bölümünde, veri inceleme, kümeleme ve sınıflandırma için en sık kullanılan çok değişkenli istatistik tabanlı yöntemlerden olan PCA ve PLS-DA'dan bahsedilecektir.

2.3.1. Temel Bileşenler Analizi (PCA)

PCA çok sayıda, birbiri ile ilişkili değişkenler içeren veri setinin boyutlarını veri içerisindeki varyansı mümkün olduğunca koruyarak daha az boyuta indirgeyen dönüşüm tekniğidir. Temel bileşenler analizinin temel hedefi, aralarında, korelasyon bulunan orijinal değişkenlerin doğrusal kombinasyonları olan temel bileşenleri (ortogonal özvektörler) bularak verilerin boyutunu küçültmektir.

PCA ayrıca kümeleme sınıflandırma için de kullanılabilir. PCA takipsiz bir metottur, bu da incelenen numunelerin doğası hakkında herhangi bir bilgi ihtiyacı duymaması anlamına gelmektedir. Yöntem, 1901'de İngiliz matematikçi ve biyoistatistikçi Karl Pearson (K. Pearson, v.d, 559,572) tarafından icat edilmiş ve ekonomi alanında da çok etkili olan Amerikalı matematikçi Harold Hotelling (H. Hotelling, 498–520, H. Hotelling, 417–441, H. Hotelling, 321–377) tarafından da geliştirilmiştir.

PCA için ilgili temel istatistiksel büyüklükler olan, ortalama, varyans (σ ' nin standart sapmanın karesi veya ortalamadan kare farklılıklarının ortalaması) ve kovaryans sırasıyla denklem 2.18, 2.19 ve 2.20' de gösterilmektedir:

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} \quad (2.18)$$

$$var(X) = \sigma^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{(N-1)} \quad (2.19)$$

$$kov(X^{(1)}, X^{(2)}) = \frac{\sum_{i=1}^N (X_i^{(1)} - \bar{X}^{(1)})(X_i^{(2)} - \bar{X}^{(2)})}{(N-1)} \quad (2.20)$$

Denklem 2.20'de tanımlanan kovaryans, $X^{(1)}, X^{(2)}$ değişkenleri arasındaki korelasyonun bir ölçüsüdür. Eğer kovaryans 0' a eşitse, iki değişken arasında korelasyon yoktur, pozitif ise aynı yönde hareket ederler ve negatif ise, ters yönde hareket ederler.

PCA yöntemi, matris 2.26’te verilmekte olan, n-boyutlu uzayda varyansın en yüksek olduğu yönleri tanımlayan yeni ortogonal değişkenleri bulmak için n değişkenli veri setinin kovaryans matrisini köşegenleştirir. Kovaryans matrisi, 2.19 ve 2.20 denklemlerinin ikiden fazla değişkene genelleştirilmiş halidir ve ayrıca kovaryans matrisinin oluşturulmasından önce tüm değişkenler önceden ortalama etrafında ortalanacak ve gerekirse normalize edilecektir.

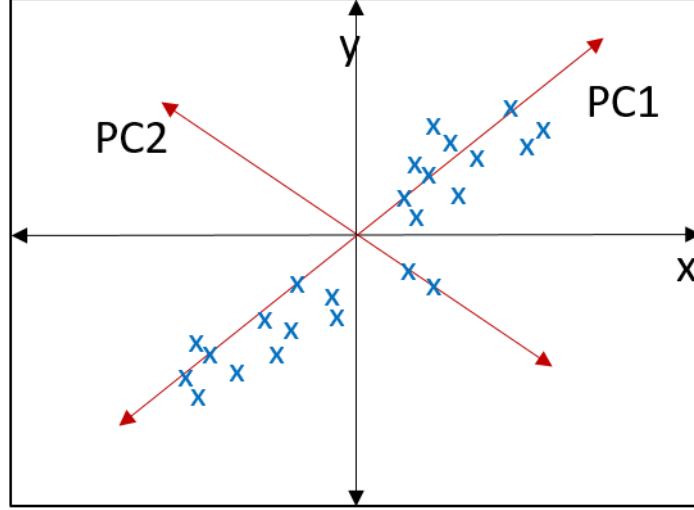
$$\begin{bmatrix} \text{kov}(X^{(1)}, X^{(1)}) & \text{kov}(X^{(1)}, X^{(2)}) & \dots & \text{kov}(X^{(1)}, X^{(p)}) \\ \text{kov}(X^{(2)}, X^{(1)}) & \text{kov}(X^{(2)}, X^{(2)}) & \dots & \text{kov}(X^{(2)}, X^{(p)}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{kov}(X^{(p)}, X^{(1)}) & \text{kov}(X^{(p)}, X^{(2)}) & \dots & \text{kov}(X^{(p)}, X^{(p)}) \end{bmatrix} \xrightarrow{\text{köşegenleştirme}} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \quad (2.26)$$

Yeni değişkenler, kovaryans matrisinin (Denklemler 2.27) özvektörlerine karşılık gelir ve Temel Bileşenler olarak adlandırılır.

Kovaryans matrisinin köşegenleştirilmesinden elde edilen en büyük mutlak değere sahip özdeğer λ_1 , verilerin ilişkili özvektörü boyunca en büyük varyansa sahip olduğunu gösterir ve birinci temel bileşen (PC1) olarak adlandırılır. Temel bileşendeki varyans yüzdesi denklem 2.28 kullanılarak elde edilir. Veri setindeki kalan en büyük varyansın yönünü belirleyen ve PC1'e ortogonal olan temel bileşen PC2 olarak adlandırılır ve bu döngü böyle devam eder. Genel olarak, verilerdeki varyansın çoğunu gözlemlemek için birkaç temel bileşen eksen(PC) yeterlidir (Şekil 2.3.).

$$\overline{\overline{\text{Kovaryans matrisi}}}.PCi = \lambda_i.PCi \quad i = 1, p \quad (2.27)$$

$$\text{var}(PCi) = 100 \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (2.28)$$



Şekil 2.3. 2 boyutlu bir veri kümesinin (değişkenler: X, Y) iki ana ekseninin (PC1 ve PC2) yönlerinin tanımı. PC1 eksenini, örneklerin en büyük varyans yönünü tanımlar, (PC2'ye ortogondur)ve en büyük varyans yönünde hizalanır.

PCA, aynı zamanda örnekleri sınıflara ayırmak için etkili bir şekilde kullanılabilen, denetimsiz etkin bir gruplandırma yöntemidir.

(Uyanık, A, 290-320, İldiz, G.O)

2.3.2. Hiyerarşik Kümeleme Analizi

Hangi kümelerin birleştirilmesi gerektiğine karar vermek için, gözlem setleri arasında bir benzerlik ölçüsü gereklidir. Temel Bileşenler Analizi, birbirine benzer değişkenleri bulmakla beraber bu yöntem her zaman iyi bir ayırım yapamayabilir. Bu, gözlem çiftleri arasındaki mesafenin uygun bir ölçüsü ve setlerdeki örneklerin ikili mesafelerinin bir fonksiyonu olarak setlerin benzerliğini belirten bir bağlantı kriteri kullanılarak elde edilebilir. Hiyerarşik kümeleme, herhangi bir geçerli mesafe ölçüsünün kullanılabilmesi ve gözlemlerin kendisinin gerekli olmaması gibi ayırt edici avantajlara sahiptir: gereken tek şey bir uzaklıklar matrisidir. Genel olarak, hiyerarşik kümelemenin sonuçları, örneklerin ve kümelerin benzerliğini (veya farklılığını) ifade etmenin çok uygun bir yolu olan bir dendrogram şeklinde sunulur. Dendrogramlar, küme füzyonunun sırasını ve her bir füzyonun gerçekleştiği mesafeyi gösterir (Şekil 2.4.).

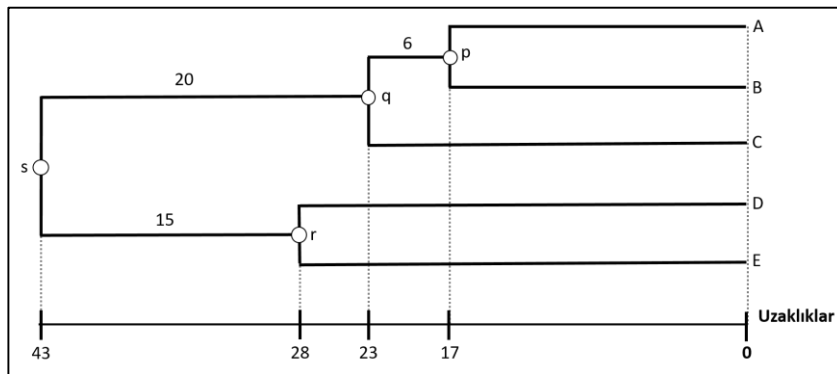
Hem mesafeler hem de bağlantı kriteri farklı şekillerde tanımlanabilir. Öklid mesafesi (Denklem 2.29) açık arayla en sık kullanılan ölçütlerdir, tam bağlantı (veya

en uzak komşu), tek bağlantı (veya en yakın komşu) ve Ward'ın (veya minimum varyans) kriterleri genellikle bağlantı için kullanılır. İlk bağlantı kriterinde, iki küme arasındaki bağlantı, iki kümeye ait tüm eleman çiftlerini içerir ve kümeler arasındaki mesafe, birbirinden en uzak olan bu iki eleman (her bir kümede bir tane) arasındaki mesafeye eşittir. Bu bağlantıların en kısıyası, elemanları dahil olan iki kümenin birleşmesine neden olur.

$$\|X^{(1)} - X^{(2)}\|_2 = \sqrt{\sum_i (X_i^{(1)} - X_i^{(2)})^2} \quad (2.29)$$

İkinci bağlantı kriteri, henüz aynı kümeye ait olmayan en yakın öge çiftini içeren iki kümeyi birleştirir. Ward'un kümeleme işlemi sırasında her adımda birleştirilecek küme çiftini seçmeye yönelik kriteri (J. H. Ward, sy. 236–244), toplam küme içi varyans J'nin en aza indirilmesidir (bkz. Denklem 2.30 (k kümesine ait bir X (i) örneği için $w_{ik} = 1$, aksi takdirde $w_{ik} = 0$ ve C_k , k kümesinin merkez noktası (veya ortalama noktası)). Uygulamada, algoritma her kümenin ortalamasının belirlenmesiyle başlar ve ardından bir kümedeki farklı nesnelere Öklid mesafelerinin karelerinin toplamını karşılık gelen ortalamaya göre hesaplar. Daha sonra kalan kümeler için de aynı işlemi yapar ve J'yi hesaplamak için tüm değerleri ekler. Her adımda birleştirilecek kümeleri seçmek için, olası tüm küme kombinasyonları dikkate alınır ve J'yi en aza indirecek olan seçilir.

$$J = \sum_{i=1}^N \sum_{k=1}^K w_{ik} \|X^i - C_k\|^2 \quad (2.30)$$



Şekil 2.4. A-E örnekleri için hiyerarşik kümeleme dendrogramının açıklayıcı örneği. Kesişim noktaları p, q, r, s olarak belirtilir ve birleştikleri kümelerin ortalama noktalarına eşit mesafede kalan veri kümesi uzayındaki noktalar olarak tanımlanır. Kesişim noktalarının yükseklik değeri, bu durumda birleştirilen kümelerin merkezleri

arasındaki mesafenin yarısıdır. Dendrogram, ultra simetriyi sergiler, çünkü tüm uçlar (A'dan E'ye)) kesişim noktaları s'den eşit uzaklıktadır(Uyanık, A, 290-320, Ildiz, G.O).

2.3.3. Kısmi En Küçük Kareler Ayrıştırma Analizi (PLS-DA)

Sınıflandırma, sınıf üyeliği bilinen gözlemleri (veya örnekleri) içeren bir çalışma veri setine dayanarak yeni bir gözlemin hangi sınıflara (alt popülasyonlara) ait olduğunu belirleme sorunu olarak tanımlanabilir. Dolayısıyla sınıflandırma, takipli çalışmayı ifade eder. Farklı karmaşıklık düzeylerine sahip birçok sınıflandırma yöntemi vardır.

Bu tez çalışması çerçevesinde klasik bir PLS regresyonundan oluşan ancak Y yanıt değişkeninin kategorik olduğu ve örneklerin hangi sınıfa ait olduğunu ifade eden Kısmi En Küçük Kareler Ayrıştırma Analizi (PLS-DA) yöntemi kullanılmıştır. PLS-DA, iki ana prosedürü kapsar: PLS bileşen (yani boyut küçültme) ve tahmin modeli oluşturma (ayrıştırma analizi).

Kısmi en küçük kareler İsveçli istatistikçi Herman Wold tarafından ve oğlu Svante Wold(S. Wold, 109–130, H.Wold, 391–420) tarafından da geliştirildi.

PLS-DA bağımlı ve bağımsız değişkenler arasında maksimum varyansa sahip hiper düzlemler bulmak yerine, tahmin edilen Y ve gözlemlenebilir X değişkenleri yeni bir alana yansıtarak doğrusal bir regresyon modeli oluşturur. Bunu yaparken, varyanslarının çoğunu hesaba katan iki tür değişkeni ilişkilendiren birkaç temel veya gizli faktör bulmaya çalışır.

Pratik açıdan, gizli faktörler, PCA'daki ana bileşenler ile benzer bir rol oynar, ancak farklı sınıflar arasında daha iyi bir ayırım sağlayabilir. (PLS-DA'da (L. C. Lee, v.d, 3526–3539) tahmin (yanıt) matrisi Y, bir sütun vektörüdür ve kategoriktir (iki sınıflı bir deney için, Y değerleri basitçe 0 ve 1 olarak alınır).

Bununla birlikte, PLS regresyonunun Y sürekli bir değişkenmiş gibi çalışmasına izin veren bir gösterge değişkeni aracılığıyla dahili olarak kaydedilir. PLS yöntemiyle geliştirilen modelin başarısı yöntemin iyi sınıflandırma yeteneğinin de bir ölçüsüdür.

Doğrusal regresyon denklem 2.31 ile ifade edilir; burada B, belirlenecek regresyon katsayıları vektörüdür ve F, en aza indirilecek olan bağımlı vektörüdür (geride kalan karelerinin toplamı, $\sum_{i=1}^N f^2 = F^T F$, en aza indirilecek etkili miktardır):

$$Y = \bar{X}B + F \quad (2.31)$$

B katsayıları pratikte 2.34 ifadesinin çözülmesiyle elde edilir.

Bu çözüm, \bar{X} ve Y'nin denklem 2.32 ve 2.33'teki gibi ayrıştırılması ve ardından regresyon sırasında yinelemeli olarak belirlenen ve \bar{W} 'nin ağırlık matrisi olduğu (\bar{W} , $\bar{X}^T Y$ olarak başlatılmaktadır) n denklem 2.34'ün çözülmesiyle elde edilir.

$$\bar{X} = \bar{U}\bar{V}^T + \bar{E} \quad (2.32)$$

$$Y = \bar{P}Q + F \quad (2.33)$$

$$B = \bar{W}(\bar{V}^T \bar{W})^{-1} Q^T \quad (2.34)$$

Denklem 2.32, 2.33 ve 2.34'te bulunan \bar{U} ve \bar{P} , X ve Y skor matrisleridir ve \bar{V} ile Q'ya karşılık gelen değerler sırasıyla X ve Y yüklemeleridir. \bar{E} ve F, sırasıyla X ve Y'deki rezidüellerdir. Regresyon katsayıları elde edildikten sonra, sınıflandırma (tahmin) amacıyla kullanılabilirler.

İlk olarak, bilinmeyen örnekler, tahmin edilen değerleri (y_{tahmin}) üretmek için B kullanılarak yeni düşük boyutlu boşluğa (k PLS gizli faktörleri ile tanımlanan) indirgenir. G sınıflarını içeren bir dizi çalışma verisi verildiğinde, PLS-DA modeli sınıflandırılacak her örnek için G tahmini değerleri ($y_{\text{tahmin}}^{(1)}, y_{\text{tahmin}}^{(2)}, \dots, y_{\text{tahmin}}^{(G)}$) üretir.

İdeal olarak, örneğin o sınıfa ait olup olmadığını belirtmek için mükemmel sınıf üyeliğinin "1" veya "0" olması gerekir. Bununla birlikte, yukarıda belirtildiği gibi, pratikte ortaya çıkan tahmin edilen değerler, bir tam sayı yerine 0 ile 1 arasındaki değerlerdir.

Bu nedenle, tahmin edilen değeri anlamlı bir sınıf üyeliğine çevirmek için bir karar kuralı (DR) uygulanmalıdır. Pek çok farklı karar kuralı kullanılabilir. (L. C. Lee,

v.d, 3526–3539)) Yaygın olarak kullanılan bir DR bir numunenin bu sınıf için öngörülen değer, o sınıfın Y (tamsayı) değerinden $\pm Y / 2$ 'den daha az farklı olması durumunda, A sınıfına ait olarak sınıflandırılmasıdır. Örneğin, bu DR' yi iki sınıf lı model için kullanırsak, için Y = 1 değeri o sınıfa aitliği Y=0 değeri o sınıfa ait olmama durumunu gösterecektir, belirli bir sınıftaki örneğin sınıflandırılması, o sınıfta sınıflandırılması için o sınıf için tahmin edilen y_{tahmin} değerinin 1 ± 0.5 aralığı içinde kalması gerekir

2.3.4. İstatistiksel Model Performansının Değerlendirilmesi

Geliştirilen modellerin performanslarının değerlendirilebilmesi için standart olarak kabul edilmiş belli metriklerin uygulanması gerekmektedir. Duyarlılık, seçicilik, etkinlik, hassasiyet ve doğruluk gibi model başarısını değerlendiren parametreler Karışıklık Tablosu parametrelerinden hesaplanır. Karışıklık tablosu (karışıklık matrisi), yanlış (YP) pozitiflerin, yanlış negatiflerin (YN, gerçek pozitiflerin (DP) ve gerçek negatiflerin (DN) sayısını bildiren iki satır ve iki sütun içeren bir tablodur. Tablonun (matrisin) her satırı tahmin edilen bir sınıftaki örnekleri temsil ederken, her sütun gerçek bir sınıftaki örnekleri temsil eder. Doğru bir şekilde ilgi sınıfta sınıflandırılan örneklere; Doğru Pozitifler (DP), doğru bir şekilde ilgi sınıfta sınıflandırılmamış örneklere; Doğru Negatifler (DN), yanlış bir şekilde ilgi sınıfta sınıflandırılan örneklere; Yanlış Pozitifler (YP) ve yanlış bir şekilde ilgi sınıfta sınıflandırılmayan örneklere; Yanlış Negatifler (YN) karşılık karşılık gelmektedir. (Chao vd., 2014; Campos vd., 2014)

Duyarlılık, modelin herbir sınıfın doğru bir şekilde sınıflandırma kabiliyetini ölçer ve aşağıdaki gibi hesaplanır:

$$\text{Duyarlılık (\%)} = 100 \times \text{DP} / (\text{DP} + \text{YN}) \quad (2.35)$$

burada YN yanlış negatif örnekleri ve DP doğru pozitif örnekleri temsil eder.

Seçicilik, modelin modellenen sınıfa ait olmayan örnekleri, doğru bir şekilde tanımlama kapasitesidir ve aşağıdaki şekilde elde edilir:

$$\text{Seçicilik (\%)} = 100 \times \text{DN} / (\text{DN} + \text{YP}) \quad (2.36)$$

burada YP yanlış pozitif örnekleri, DN ise doğru negatif örnekleri temsil eder.

Hassasiyet, tüm sınıflardan doğru olarak ne kadar tahmin edildiğinin, modelin olumlu tahminlerinin kalitesinin bir ölçüsüdür:

$$\text{Hassasiyet (\%)} = 100 \times \text{DP} / (\text{DP} + \text{YP}) \quad (2.37)$$

Etkinlik ve doğruluk, model performansının bir ölçütüdür.

Etkinlik, duyarlılık ve seçicilik analizleri tarafından elde edilen tüm bilgileri birleştirir, doğruluk ise sınıftan bağımsız olarak doğru sınıflandırmaların oranıdır:

$$\text{Etkinlik (\%)} = 100 \times (\text{duyarlılık} + \text{seçicilik}) / 2 \quad (2.38)$$

$$\text{Doğruluk (\%)} = 100 \times (\text{doğru sınıflandırma}) / \text{toplam örnekler} \quad (2.39)$$

(Chao vd., 2014; Campos vd., 2014)

2.3.5. Sınıflandırma Modellerinin Geliştirilmesi ve Test Edilmesi

Bu tez çalışmasında iki sınıflandırma yöntemi kullanılmıştır. Birincisi açıklanan PCA yöntemi ve ikincisi Kısmi En Küçük Kareler Ayrım Analizi yöntemidir. (PLS-DA) (Barker vd., 2003). PCA, örneklerin sınıf üyeliğine ilişkin herhangi bir ön bilgiye ihtiyaç duymayan denetimsiz bir yöntemdir. Öte yandan PLS-DA yöntemi, çalışma setinde yer alan örneklerin ait olduğu sınıf hakkında bilgi gerektirir.

Modellerin geliştirilmesi için, veri seti (spektrumlar), modelleri oluşturmak için kullanılan kalibrasyon seti ve yeni örneklerin sınıflandırılmasında yöntemin performansını kontrol etmek için kullanılacak olan test seti olmak üzere rastgele iki alt gruba ayrıldı: Kalibrasyon seti çalışma ve validasyon setlerini içermektedir, fakat bu çalışmada geliştirilen tüm modellerde çalışma ve validasyon setlerinin eşit olduğu tam çapraz doğrulama (full cross validation) kullanılmıştır.

Mevcut araştırmada tüm kemometrik analizler Unscrambler™ CAMO yazılımı (Sürüm 5.5) (Unscrambler, 1995) kullanılarak gerçekleştirilmiştir.

PLS-DA, yukarıda belirtildiği gibi, çok değişkenli örnekleri (bizim durumumuzda, spektrumları) uygun sınıfa atayabilen matematiksel modeller bulmayı amaçlayan denetimli bir sınıflandırma yöntemidir. Sınıflandırma modeli kalibre edildikten sonra, bilinmeyen örneklerin sınıfı, geliştirilen matematiksel model kullanılarak tahmin edilebilir.

PLS-DA, incelenen sistemlerin boyutluluğunu (değişkenlerin sayısı; bizim durumumuzda titreşim frekansları), veri setindeki (gizli değişkenler veya gizli faktörler) varyansı en üst düzeye çıkaran çok boyutlu bir uzayda dikey eksenleri tanımlayan değişkenlerin doğrusal kombinasyonlarını bularak, azaltma yönüyle PCA ile aynı şekilde çalışır.

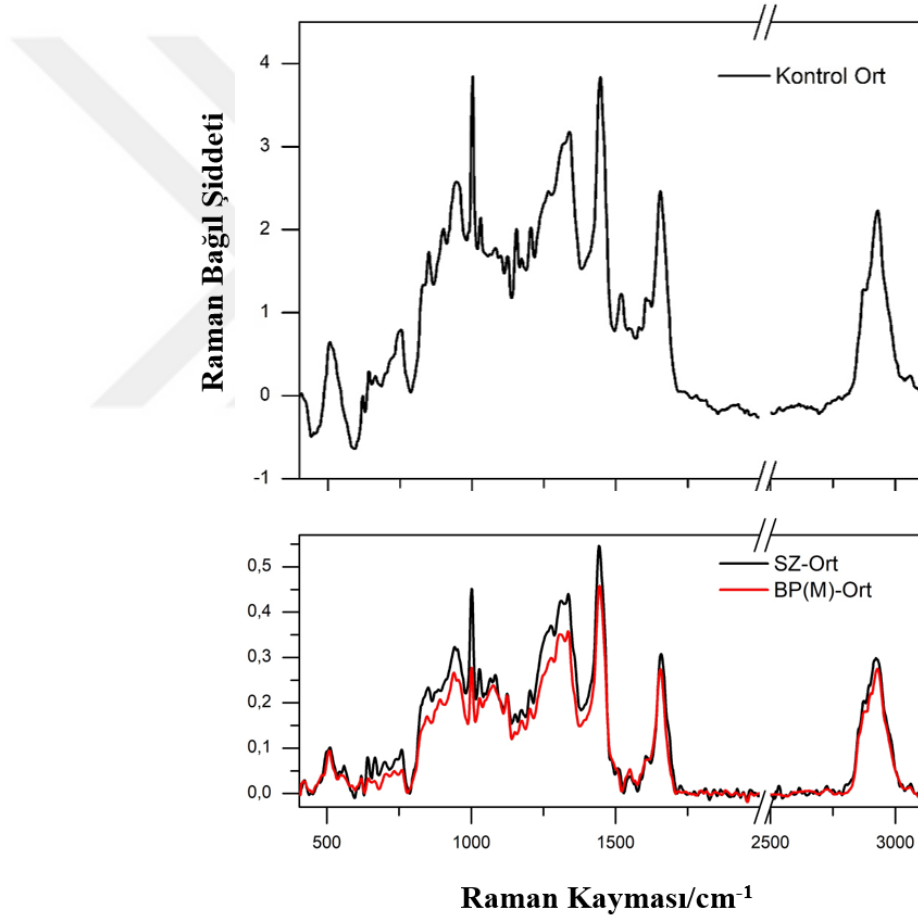
PLS-DA'da, bununla birlikte, PCA'da olanın aksine, sınıfını tanımlayan, çalışma veri setine ait her bir örneğe tahmin edilen bir değişken atanır. Basit bir ifadeyle, bu, maksimum varyans yönü arayışının, örneklere doğru sınıf üyeliklerinin atanmasını zorlayan alanla sınırlı olduğu anlamına gelir. (Ballabio vd., 2013; Luxburg, 2007) Bu nedenle, PLS-DA analizinde, bir veya daha fazla Y değişkeni (sınıflar veya referans değerleri), X değişkenleriyle (veri seti; spektrumlar) aynı anda modellenir. Modelleme en küçük kareler analiz algoritmasının değiştirilmiş bir versiyonu ile yapılır.

Pratik anlamda, sınıf kimliğini tanımlayan Y değişkenleri, belirli bir sınıfa ait bir örnek için ideal "1" değerini ve bu sınıfın üyesi olmayanlar için ideal "0" değerini varsayar (Martens vd., 1989; Esbensen vd., 2005). Tahmin prosedürü sonucunda, her örnek için ideal değerlerden farklı değerler atanır (tahmin edilir), sınıf üyeliklerini belirlemek için bir kriter kullanılması gerekmektedir. Bu çalışmada, o sınıf için öngörülen Y değeri 0.6 ile 1.6 arasında olan bir örneklemin belirli bir sınıfa ait olduğu kabul edilmiştir.

3. BULGULAR

3.1. Kan Raman Spektrumlarının Ön Analizi

Şekil 3.1’de kontrol grubuna (sağlıklı bireyler) ait kan serumunun ortalama Raman spektrumları(üstte) ve bu spektrumun SZ ve BP(M) gruplarının, ortalama spektrumlarından çıkarılarak elde edilen fark spektrumlarını(alta) gösterilmektedir. Kan serumunun Raman spektrumlarının literatüre göre bant işaretlemeleri Tablo 3.1.’de verilmektedir. (Shao vd., 2018; Perez vd., 2018; Atkins vd., 2017; Nargis vd., 2019; Saade vd., 2012)



Şekil 3.1. Kontrol (K) grubunun kan serumunun ortalama Raman Spektrumu (üstte), kontrol (K) grubunun kan serumunun ortalama Raman Spektrumunun, SZ (siyah) ve BP(M) (kırmızı) gruplarının normalleştirilmiş ortalama Raman spektrumlarından çıkarılarak elde edilen fark spektrumları (altta).

Şekil 3.1’de SZ ve BP(M) hastalara ait spektrumların kontrolle kıyaslandığında farklılık gösterdiği görülmektedir, bu da spektrumların kandaki biyokimyasal ortamı,

incelenen grupların her birini karakterize edecek şekilde yansıtması gerektiği tezimizi desteklemektedir.

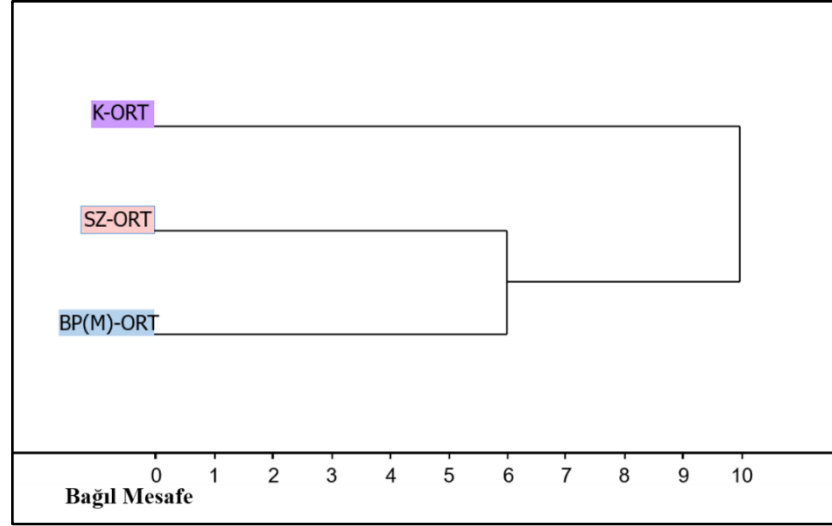
Tablo 3.1. Kan serumu Raman spektrumundaki temel bantların işaretlemeleri

Raman Kayması	İşaretlemeler	Raman Kayması	İşaretlemeler
2929	lipit	1267	fosfolipit $\delta(\text{CH})$
1655	Amit I (protein) $\nu(\text{C}=\text{O})$	1205	$\nu(\text{C}=\text{C})$ amino asitler
1609	fenilalenin $\nu(\text{C}=\text{C})$	1173	sitozin, guanin
1445	fosfolipit, lipoprotein	1002	fenilalenin $\nu(\text{C}-\text{H})$
	$\delta(\text{CH}_2)$, $\delta(\text{CH}_3)$	945	fenilalenin $\nu(\text{C}-\text{C})$
1338	protein (triptofan)	850	tirozin
		754	guanin, timin

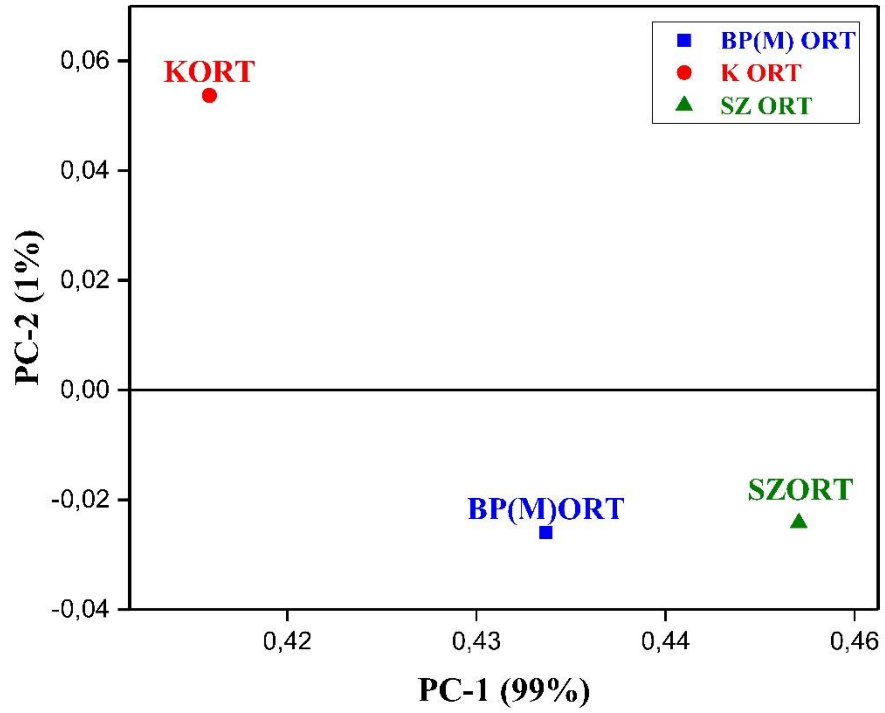
İşaretlemeler literatüre göre yapılmıştır. (Shao vd., 2018; Perez vd., 2018; Atkins vd., 2017; Nargis vd., 2019; Saade vd., 2012) Raman kaymaları cm^{-1} cinsindedir.

Şekil 3.1'den, görüldüğü üzere, SZ ve BP(M) gruplarının ortalama spektrumlarının benzer profiller sergilediği görülmektedir. Bu gözlem, kliniksel/davranışsal semptomlar çerçevesinde BP(M) fazının SZ ile büyük benzerlik göstermesi nedeniyle, beklentilerle uyum içindedir. (Correl vd., 2007)

SZ ve BP(M) grupları arasındaki benzerlik ve farklılıkları daha detaylı incelemek için, Ward Algoritması (Ward, 1963; Murtagh vd., 2014) kullanılarak, kümeleme analizi ve Temel Bileşen Analizi (PCA) yapıldı. BP(M), SZ ve K grubunun ortalama Raman spektrumları üzerinden yapılan Hiyerarşik kümeleme analizinden elde edilen dendrogramın ve Temel Bileşen Analizi sonucu elde edilen skor grafiğinin SZ ve BP(M) gruplarını birlikte ve K grubunu ayrı gruplandığı sırasıyla Şekil 3.2 ve Şekil 3.3'te görülmektedir.



Şekil 3.2. BP(M), SZ ve K grubunun ortalama Raman Hiyerarşik kümeleme analizi dendrogramı



Şekil 3.3 SZ, BP(M) ve K gruplarının alan normalize ortalama kan serumu Raman spektrumlarının PCA skorları (PC-2'ye karşı PC-1)

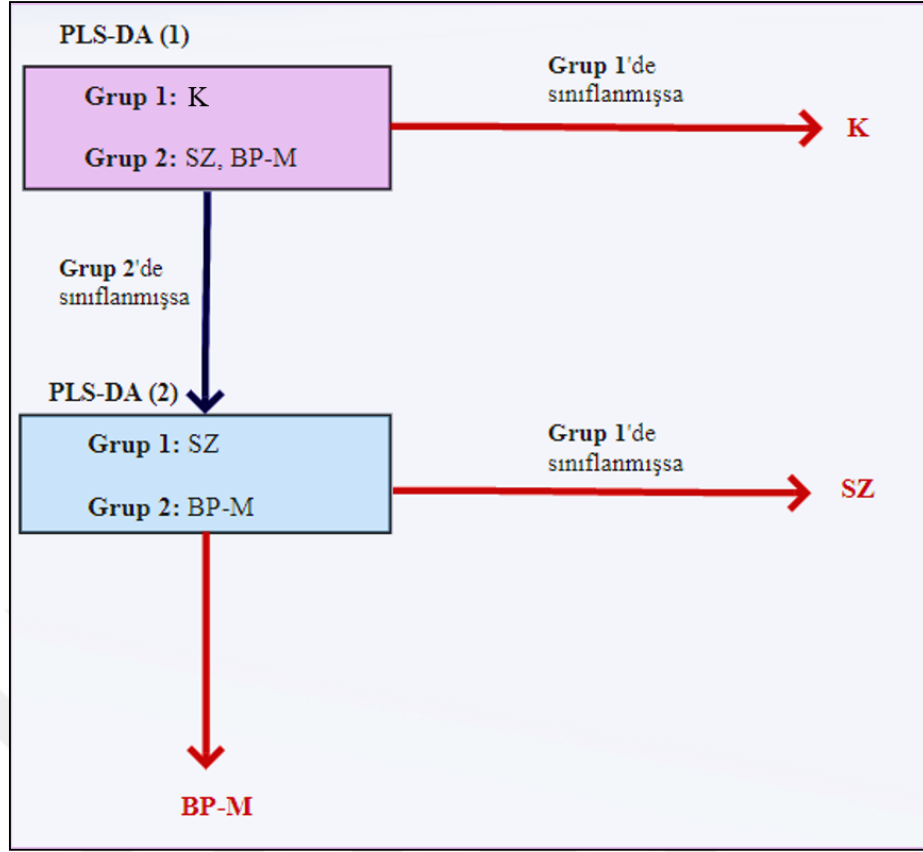
Şekil 3.3.'de PCA skorları grafiğinde gösterilen eğilimler göz önüne alındığında, PC-2 ve PC-1 yüklemelerini (bunlar Raman dalgasayıları ile ilişkilidir), araştırılan grupların kan serumunun spesifik biyomoleküllere bağlı faktörlerle ilişkilendirmeye çalışmak cazip görünmektedir. Ancak, örneklerin ortalama spektrumları gruplar

arasındaki deęişimleri yansıtmaz ki bu alıřma erevesinde, arařtırılan her iki bozukluęa sahip, hastalar aralarında kayda deęer bir eřitlilik ile karakterize edildięi iin nemli olması beklenebilir.

Bu alıřma erevesinde hastalıkların biyobelirteleri olarak iřlev gorebilecek spesifik moleklleri tanımlamaya alıřmak yerine, tm spektroskopik verilere dayanan, bireyleri hastalıklarına gore sınıflandıran ve dolayısıyla hastalıkların teřhisi iin yardımcı bir ara olarak kullanılabilir btncl bir yaklařım saęlamaktır. PCA analizinden elde edilen sonular ortalamalar zerinden genel bir yaklařım olsada, incelenen birey gruplarının kan serumunda bulunan biyokimyasal ortamlar arasında benzerlik veya farklılıkları aıka yansıttıęı grlmektedir. (Palazzo vd., 2017; Torres vd., 2007) BP manik fazının řizofreni ile benzer semtoplara sahip olması nedeniyle kan rneklelerinin biyokimyasal zelliklerinde benzerlikler beklenebilir. (Correl vd., 2007)

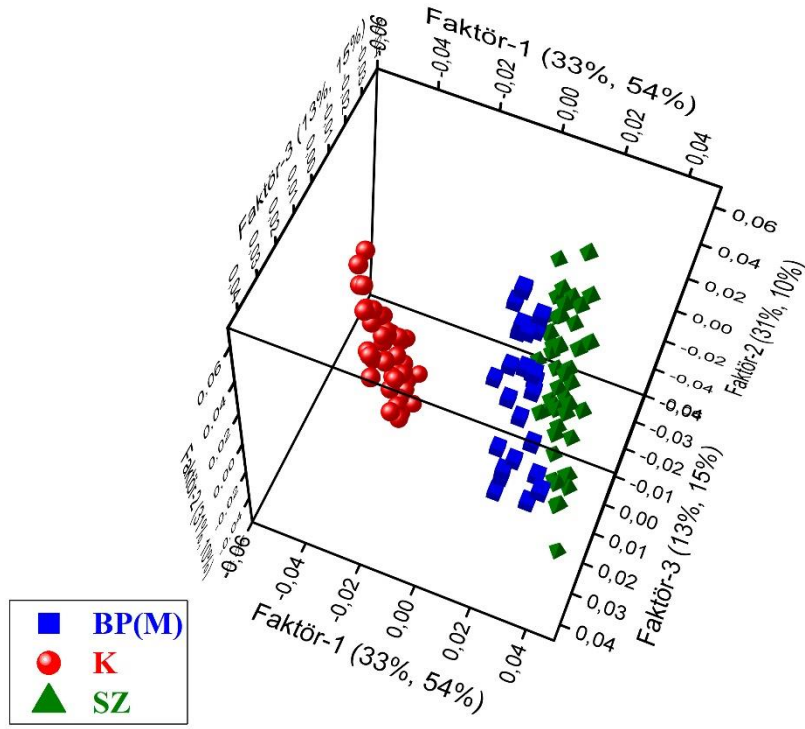
3.2. Raman Spektrum Tabanlı Sınıflandırma Modellerinin Geliřtirilmesi (3 sınıflı model ve hiyerarřik 2x 2 sınıflı model)

Sınıflandırma modellerini geliřtirmek iin Raman spektroskopik teknięi kullanıldı. İlk olarak Kısmı En Kk Kareler Ayırma Analizi (PLS-DA) yntemiyle 3 sınıflı bir model (BP(M), SZ ve K) geliřtirilerek, Raman spektroskopik teknięinin hastalıkları birbirlerinden ve saęlıklı kontrollerden ayırt etme gc incelendi (Barker, 2003). Bu geliřtirilen ilk model, kalibrasyon setindeki tm rnekle aynı anda kullanılarak oluřturulan 3 sınıflı bir modeldir (3S Modeli). Ayrıca BP(M)'yi SZ'den ve kontrollerden ayırmak iin, grupların ortalama spektrumlarının (řekil 3.2) hiyerarřik kmelenmesinden kaynaklanan benzerlik řeması gz nnde bulundurulur ikili adımlardan oluřan (řekil 3.4) 2 sınıflı iki ařamalı ikinci bir model daha geliřtirilerek (2 / 2S Model) model performansları karřılařtırıldı. Her iki modelde de kalibrasyon sırasında apraz validasyon (full cross-validation) kullanılmıřtır. Modellerin kalibrasyonu iin kullanılan veri seti 45 SZ, 45 K ve 25 BP(M)'den elde edilen 115 rneęi iermektedir.



Şekil 3.4. 2/2S Modelinin ikili adımları. İkili adımlar oluşturulurken, grupların ortalama spektrumlarının (bakınız Şekil 3.2.) hiyerarşik kümelenmesindeki dallanma temel almıştır.

Şekil 3.5'te en büyük varyans değerlerini (Faktör-1, Faktör-2, Faktör-3) açıklayan 3 faktör için 3S Modeli 3 boyutlu skor grafiğini göstermektedir. Bu faktörler çalışma seti (training seti) için sırasıyla, X değişkenindeki (toplam 77 % varyans) varyansın 33, 31 ve 13 % 'ünü ve Y'deki varyansın 54, 10 ve 15 %'ini (toplam 79%) açıklamaktadır. Validasyon seti (geçerlilik seti) için de benzer sayılar (X değişkeninde % 31, 32, 13 ve Y'de % 54, 8, 13) gözlenmektedir. Model, çalışma seti için sırasıyla % 77 ve % 79 toplam X ve Y varyansını hesaba katan beş gizli değişken kullanılarak geliştirilmiştir (validasyon: % 76 ve % 75). Çalışma ve validasyon için hata karalar ortalamasının karekökü (RMSE) sırasıyla, K için 0.07 ve 0.08, SZ için 0.13 ve 0.14, BP(M) için ise 0.16 ve 0.17'dir.

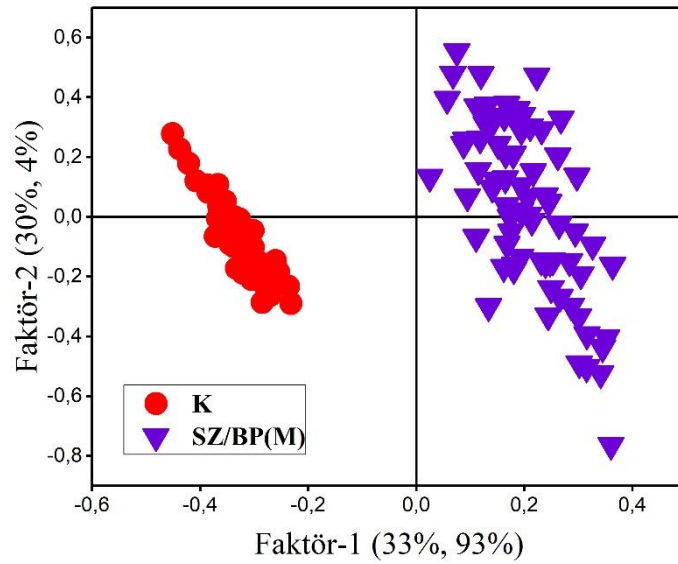


Şekil 3.5. 3S Modeli için 3D skor grafiği (Faktör-1'e Faktör-2'ye Faktör-3).

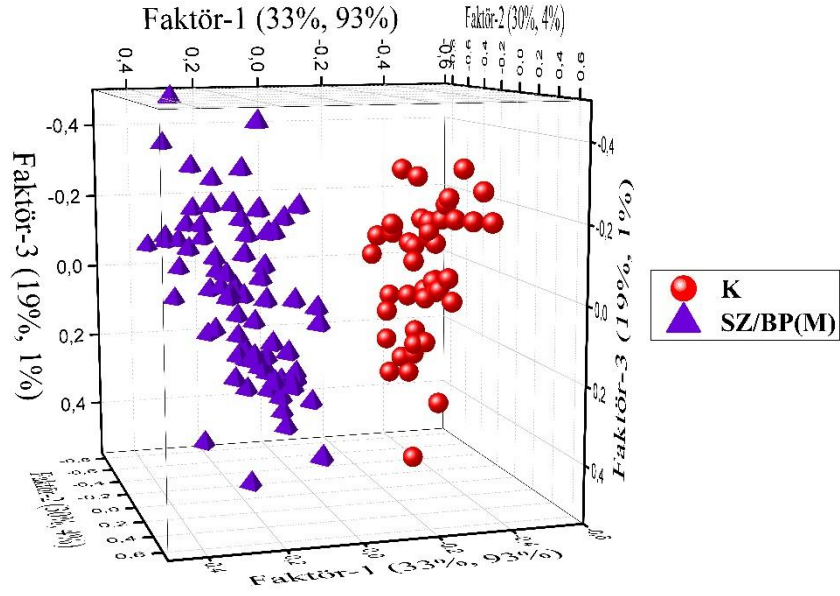
Skor grafiği, örneklerin tabiatlarına göre SZ, BP(M) ve K olmak üzere üç grup biçiminde kümelendiğini göstermektedir. SZ ve BP(M) örnekleri pozitif Faktör-1 eksenini boyunca kümelendirken, K örnekleri negatif Faktör-1 eksenini boyunca gruplanmaktadır. Bununla birlikte, BP(M) örneklerinin SZ örneklerine kıyasla Faktör-1 eksenini boyunca daha negatif skor değerleri aldığı görülmektedir. Her bir gruba ait örnekler hem Faktör-2 hem Faktör-3 eksenleri boyunca dağınık bir şekilde yayılmıştır. Skor grafiklerinden de görüldüğü üzere iki adımlı üç sınıflı 3S modelinin örnekleri gruplarına göre kümelemektedir, fakat SZ ve BP(M) örneklerini içeren gruplara ait kümeler arası mesafe oldukça küçüktür ve örneklerin üst üste binme durumu gözlenmektedir. Bu nedenle SZ ve BP(M) gruplarının daha iyi ayrıştırılmasına olanak sağlayabilecek, iki sınıf iki adımdan oluşan ikinci bir model (2/2S) geliştirilmiştir (Şekil 3.4.).

2/2S Modeli için elde edilen skor grafikleri Şekil 3.6. ve Şekil 3.7.'de gösterilmektedir. Bu model her biri iki sınıflı iki adımdan oluşan bir modeldir. Geliştirilen modelin birinci basamağı (2 sınıflı PLS-DA basamağı (PLS-DA (1))), iki

sınıftan oluşmaktadır. Birinci sınıf K örneklerini içerirken, ikinci sınıf BP(M) ve SZ örneklerinden oluşmaktadır. Şekil 3.6'da verilen skor grafikleri incelendiğinde toplamda Faktör-1 ve Faktör-2 çalışma (training) seti için, X ve Y verilerindeki varyansın % 62'ünü ve % 97'ünü açıklarken doğrulama (validasyon) seti için X ve Y verilerindeki varyansın sırasıyla 61% ve 97%'ünü açıklamaktadır. Bununla birlikte bu modelde kullanılan beş faktör çalışma (training) seti için X ve Y'deki varyansın sırasıyla %87 ve %99'unu (doğrulama (validasyon) : 86% ve 99%) açıklamaktadır. Hem çalışma hem de validasyon seti için RMSE 0.09'tur. Skor grafiğinden PLS-DA'in, Faktör-1 ekseninde, birinci sınıfa (K) ait skorların negatif ve ikinci sınıfa (SZ/BP(M)) ait skorların pozitif değer alacak şekilde net olarak ayırdığı görülmektedir (Şekil 3.6 ve Şekil 3.7).



Şekil 3.6. 2/2S Modelin ilk PLS-DA (1) adımı için skor grafiği (Faktör-2'ye karşı Faktör-1)

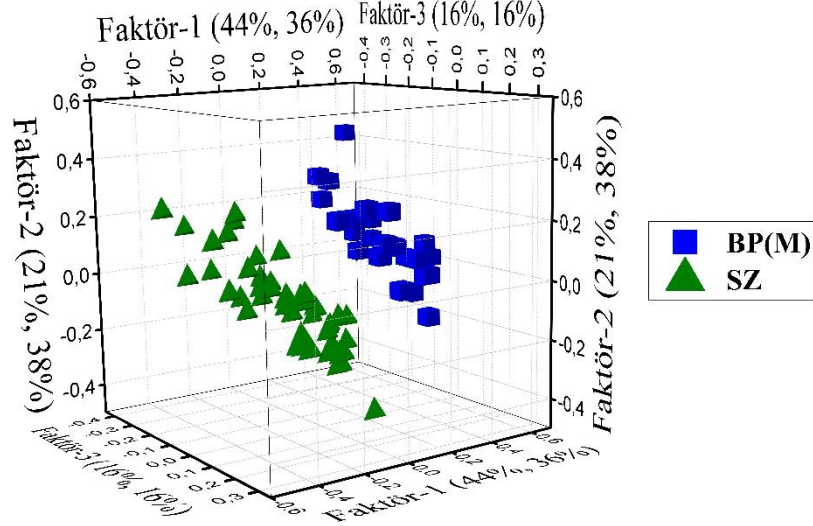


Şekil 3.7. 2/2S Modelin ilk PLS-DA (1) adımı için 3D skor grafiği (Faktör-1'e Faktör-2'ye Faktör-3)

Modelin ikinci adımı için 2 sınıflı PLS-DA basamağı (PLS-DA(2)) Şekil 3.8.'de gösterilmektedir. Bu aşamada birinci sınıf SZ, ikinci sınıf BP(M) numulerinden oluşmaktadır.

PLS-DA(2) adımı için verilen skor grafikleri incelendiğinde toplamda Faktör-1, Faktör-2 ve Faktör-3 çalışma (training) seti için, X verilerindeki varyansın sırasıyla %44, %21 ve %16 'sını, Y'deki varyansın sırasıyla 36%, 38% ve 16%) (toplam X: 81%, Y: 90%), doğrulama (validasyon) seti için X verilerindeki varyansın sırasıyla 42%, 20% ve 17% 'sini, Y'deki varyansın sırasıyla 32%, 37% ve 19% (toplam X: 79%, Y: 88%)'ini açıklamaktadır. Bununla birlikte bu modelde kullanılan beş faktör çalışma (training) seti için X ve Y'deki varyansın sırasıyla 84% ve 82%'ini (doğrulama (validasyon) : 98% ve 98%) açıklamaktadır. Çalışma seti ve doğrulama seti için sırasıyla hata karaler ortalamasının karekökü RMSE 0.6 ve 0.7'dir. Skor grafiğinden PLS-DA'in, Faktör-1 eksininde, birinci sınıfa ait skorların negatif ve ikinci sınıfa ait skorların pozitif değer alacak şekilde net olarak ayırdığı görülmektedir. (Şekil 3.8.). BP(M) örnekleri pozitif Faktör-1, Faktör-2 ve Faktör-3 skor bölgesinde

kümelenirken, SZ örnekleri, Faktör-1 eksenini boyunca yayılarak negatif Faktör-3 ve Faktör-2 skor bölgelerinde kümelenmiştir.



Şekil 3.8. 2/ 2S Modelinin ikinci adımı PLS-DA(2) için 3D skor grafiği (Faktör-1 / Faktör-2 / Faktör-3).

3.3. Modellerin Tahmin Yeteneğinin Test Edilmesi

Modellerin tahmin yeteneğinin ölçülebilmesi için, modellerin kalibrasyonunda kullanılmayan her bir gruba ait 15 örnek doğrulama (test) seti olarak kullanılmıştır. Tahminler için kullanılacak doğrulama setine ait Raman spektrumları, modellerin kalibrasyonunda kullanılan örneklerle aynı adımları izleyerek önceden işlenmiştir. 3S Modeli kullanılarak yapılan tahminler için elde edilen sonuçlar Tablo 3.3 ve tahminler Şekil 9-11'de verilmektedir. Modelin değerlendirilmesi için bölüm 2.3.3.'te bahsedilen duyarlılık, seçicilik, hassasiyet, etkinlik ve doğruluk sınıflandırma performans parametreleri (Chao vd., 2014; Campos vd., 2014) Karışıklık Tablosu değerlerinden (bakınız Tablo 3) hesaplanmıştır.

Örneklerin sınıf atamasının yapıldığı PLS-DA tahmin aşamasında, her bir örnek için tahmin edilen Y değerinin eşik değeri 0.6'dir. Tahmin edilen Y değerleri

tüm sınıflar için 0,6' dan küçük veya birden fazla sınıf için bu değerden daha büyük olan örneklerin, model tarafından “aykırı değer” olarak tahmin edilmektedir.

Tablo 3.2. Karışıklık (Confusion) Tablosu ve 3S Modeli için istatistiksel model performans göstergeleri^a

Gerçek Sınıf	Tahmin edilen Sınıf				Duyarlılık (%)	Seçicilik (%)	Hassasiyet (%)	Etkinlik (%)
	K	SZ	BP(M)	AYKIRI				
K	14	0	0	1	93,3	100	100	96,6
SZ	0	8	0	7	53,3	100	100	76,6
BP(M)	0	0	15	0	100	100	100	100
Toplam					82,2	100	100	91

^a AYKIRI (aykırı değer), tahmin edilen Y değerleri tüm sınıflar için 0,6'ten küçük veya birden fazla sınıf için bu değerden daha büyük olan örneklere karşılık gelir.

Tablo 3.3. K , SZ ve BP(M) için Tablo 3.2'de verilen karışıklık tablosu verileri kullanılarak elde edilen DP, DN, YP ve YN değerleri (3S Modeli için)

	Doğru Pozitif	Doğru Negatif	Yanlış Pozitif	Yanlış Negatif
K	14	30	0	1
SZ	8	30	0	7
BP(M)	15	30	0	0

Tablo 3.3' te sunulan sonuçlardan, 3S Modelinin, seçicilik ve hassasiyetinin iyi olduğu görülmektedir. Fakat, yanlış negatif değerlerdeki fazlalık nedeniyle model (BP(M) dışında) düşük duyarlılık ve etkinliğe sahiptir.

“Aykırı değer” olarak sınıflandırılan K ve SZ örneklerinin çoğunun, yine de, bu grupların üyesi olarak sınıflandırılmak için belirlenmiş kriterlere oldukça yakın olduğunun ve Y sınıfı için ilgili değerlerin 0.6' nın üzerinde olduğunun belirtilmesi yerinde olacaktır.

Sonuç olarak, 3S Modelinin sadece BP(M) örnekleri için kabul edilebilir tahmin performansına sahip olduğu görülmüştür. Model doğruluğu % 82,2' dir, bu da modelin tahmin yeteneğinin iyi fakat yeterli olmadığını göstermektedir.

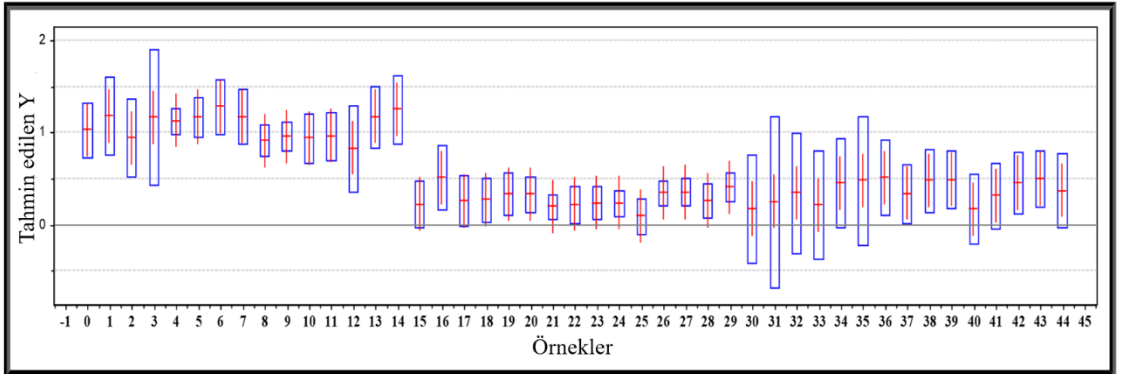
2/2S modeli kullanılarak yapılan tahminlerin sonuçları Tablo 3.4. ve Şekil 12-15 'de verilmektedir. Tablo 3.4, her bir 2 sınıflı PLS-DA adımı için karışıklık tablosundan elde edilen verilere dayanarak hesaplanan sınıflandırma performansı parametrelerini, şekiller ise modelin her bir adımında örnekler için tahmin edilen Y değerlerini ve sapmaları göstermektedir.(De Vries vd., 1995)

Tablo 3.4. Karışıklı (Confusion) Tabloları ve 2 / 2S Modeli için istatistiksel model performans göstergeleri. ^a

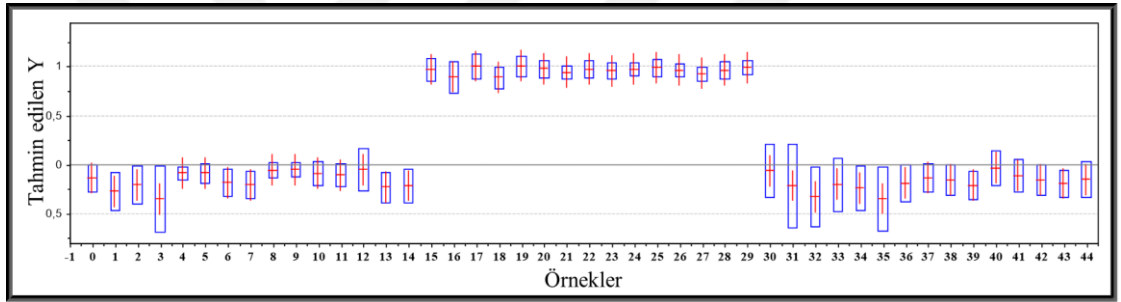
Gerçek Sınıf	Tahmin Edilen Sınıf			Duyarlılık (%)	Seçicilik (%)	Hassasiyet (%)	Etkinlik (%)
1.adım	BP(M)/SZ	K	AYKIRI				
BP(M)/SZ	15	0	0	100	100	100	100
K	0	15	0	100	100	100	100
2.adım	SZ	BP(M)	AYKIRI				
SZ	15	0	0	100	100	100	100
BP(M)	0	15	0	100	100	100	100
Toplam				100	100	100	100

^a AYKIRI, tahmin edilen Y değerleri tüm sınıflar için 0,60' tan küçük veya birden fazla sınıf için bu değerden daha büyük olan örneklere karşılık gelir.

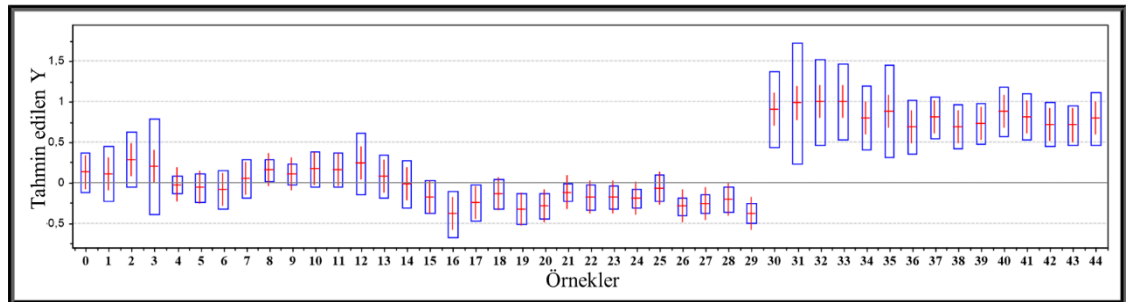
Sonuçlar tüm adımların grupları iyi ayırt edebildiğini göstermektedir; modelin doğruluğu % 100.0 olarak hesaplanmıştır



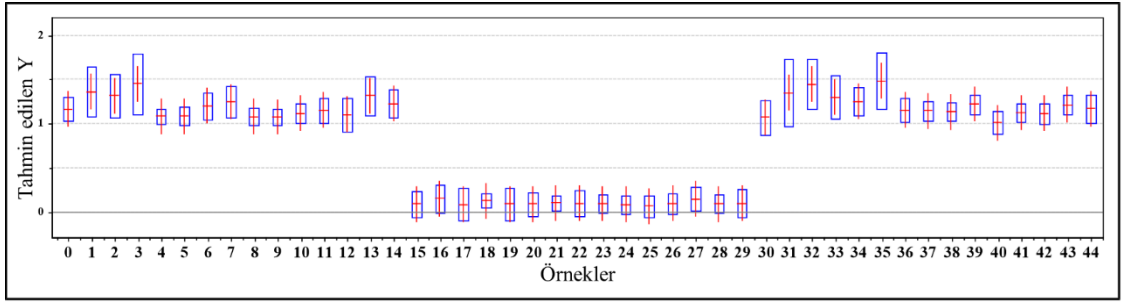
Şekil 3.9.: 3S modelindeki örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, BP(M) grubuna ait örnekler sınıf (Y için değer 1) ve SZ ve K'ya ait örnekler sınıf 2 (Y için değer 0) olarak tanımlanmıştır. 45 test örneğinde:1-14 BP(M)'ye; 15-29 K'ya; 30-45 SZ'ye karşılık gelmektedir.



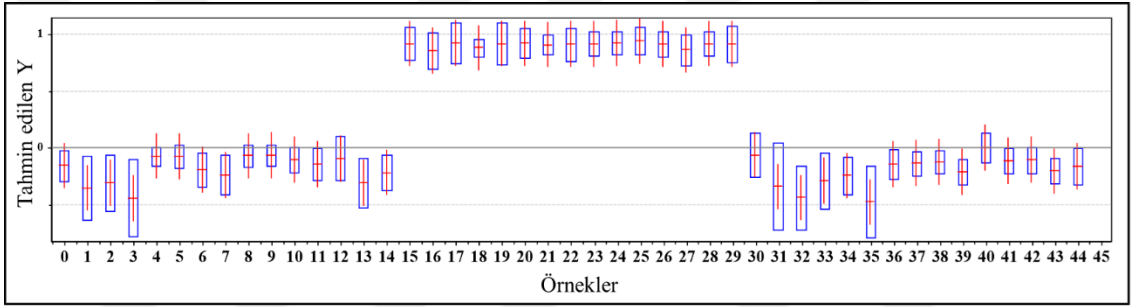
Şekil 3.10.: 3S modelindeki örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, K grubuna ait örnekler sınıf (Y için değer 1) ve SZ, BP(M)'ye ait örnekler sınıf 2 (Y için değer 0) olarak tanımlanmıştır. 45 test örneğinde:1-14 BP(M)'ye; 15-29 K'ya; 30-45 SZ'ye karşılık gelmektedir.



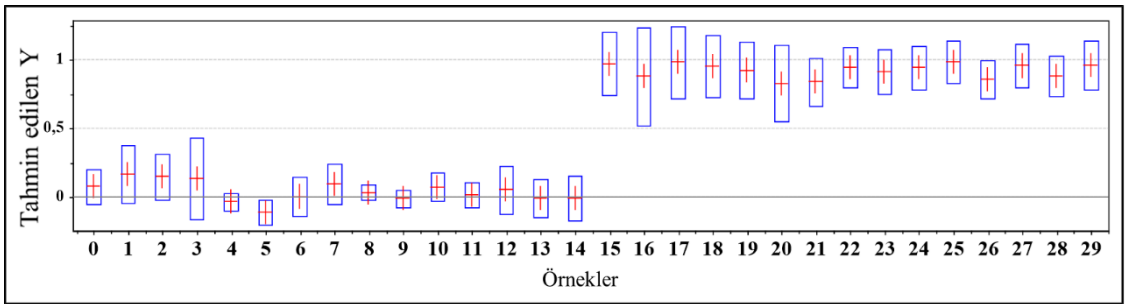
Şekil 3.11.: 3S modelindeki göre, örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, SZ grubuna ait örnekler sınıf (Y için değer 1) ve BP(M), K'ya ait örnekler sınıf 2 (Y için değer 0) olarak tanımlanmıştır. 45 test örneğinde:1-14 BP(M)'ye; 15-29 K'ya; 30-45 SZ'ye karşılık gelmektedir.



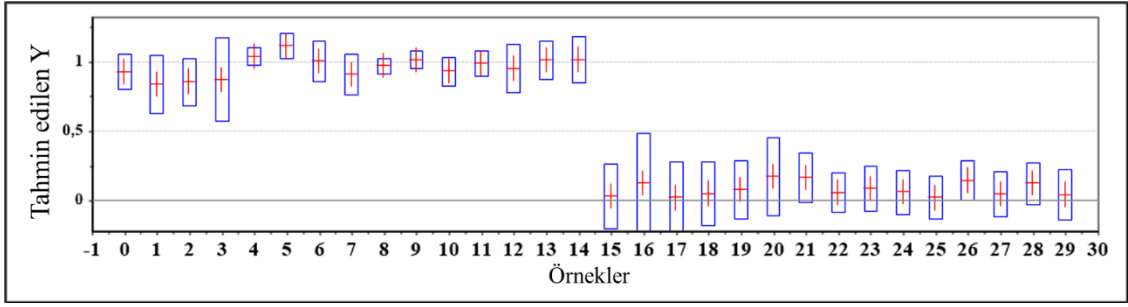
Şekil 3.12.: 2/2S modelinin 1. adımındaki örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, SZ ve BP(M) gruplarına ait örnekler sınıf 1(Y değeri 1) ve K sınıf 2 (Y değeri 0) olarak tanımlanmıştır. 45 test örneğinde:1-14 BP(M)'ye; 15-29 K'ya; 30-45 SZ'ye karşılık gelmektedir.



Şekil 3.13.: 2/2S modelinin 1. adımındaki örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, SZ/ BP(M) gruplarına ait örnekler sınıf 1(Y değeri 0) ve K sınıf 2 (Y değeri 1) olarak tanımlanmıştır. 45 test örneğinde:1-14 BP(M)'ye; 15-29 K'ya; 30-45 SZ'ye karşılık gelmektedir.



Şekil 3.14.: 2/2S modelinin 2. adımındaki örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, BP(M) gruplarına ait örnekler sınıf 1(Y değeri 0) ve SZ sınıf 2 (Y değeri 1) olarak tanımlanmıştır. 30 test örneğinde: 1-14 BP(M)'ye; 15-30 SZ'ye karşılık gelmektedir.



Şekil 3.15.: 2/2S modelinin 2. adımındaki örnekler için tahmin edilen Y değerleri. Tahmin edilen değerler kırmızı çizgilerle, sapmalar mavi kutularla gösterilmektedir. Modelde, BP(M) gruplarına ait örnekler sınıf 1(Y değeri 1) ve SZ sınıf 2 (Y değeri 0) olarak tanımlanmıştır. 30 test örneğinde: 1-14 BP(M)'ye; 15-30 SZ'ye karşılık gelmektedir.

Tüm bu sonuçlar, 2/2S Modelinin, yüksek bir tahmin gücü olan güvenilir bir model olduğunu göstermektedir. Buna ek olarak, hiçbir örnek model tarafından “aykırı” olarak sınıflandırılmamakla beraber tasarlanan modelin (ikili algoritmaya dayalı) analizi gerçekleştirme ve farklı örnekleri sınıflandırmadaki gücünü ortaya koymaktadır.

4. TARTIŞMA VE SONUÇ

Bu tezin asıl sonucu, hastaların kan serumu Raman spektrumlarına dayalı olarak SZ ve BP(M)'nin yardımcı ayırıcı tanısı için analitik bir modelin geliştirilmiş olmasıdır. Bu model aynı zamanda SZ ve BP(M)'yi, sağlıklıdan ayırt edebilmektedir. Model, 2 adet 2 sınıflı PLS-DA adımının (2/2SModel) hiyerarşik bir dizisine dayanmaktadır. Bu model, örneklerin sınıflandırmasını gerçekleştirmek için, örneklerin tümü aynı anda kullanılarak geliştirilen klasik 3 sınıflı bir PLS-DA modeline (3S Model) göre önemli ölçüde daha iyi olduğu gösterilmiştir.

Elde edilen bulgular ışığında $400-3100\text{ cm}^{-1}$ Raman spektral aralığının BP ve SZ bozuklukları için hassas bir araç olduğu ve bu hastalıklar için bütünsel bir spektroskopik biyobelirteç görevi gördüğü sonucuna varılmıştır. Bu yaklaşım kavramsal olarak, belirli moleküler biyobelirteçleri bulmak için karmaşık biyolojik materyallerin spektroskopik verilerini kullanmayı amaçlayan alternatiflerden daha tutarlıdır, çünkü sınıflandırma tüm spektroskopik bilgileri dikkate alır ve kanda varsayılabilecek farklı biyobelirteçlerin düşük konsantrasyonuna ilişkili ortaya çıkacak sorunları önler.

Araştırılan farklı grupların (Raman spektrumları ile temsil edilen) kan serumundaki biyokimyasal ortamın benzerliğine ilişkin elde edilen sonuçlar, psikiyatrik çerçeveye uygun olup, şizofreni grubu verileri sistematik olarak BP manik hastalarınıninkine benzerken, kontrol grubuna karşılık gelenlere daha uzaktır. PCA ve PLS-DA sınıflandırma yöntemleri kullanılarak geliştirilen bu iki modelin, mükemmel sınıflandırma / tahmin performansları sergiledikleri görülmüştür.

Tez kapsamında oluşturulan metodoloji güvenilir, hızlı, ucuz, esasen non-invaziv olup, psikiyatristlerin, kişisel ve toplumsal önemi büyük olan nöropsikiyatrik bozuklukların erken tanısına destek olarak daha kesin bir şekilde tespit etmelerine yardımcı olmak için klinik ortamda kolaylıkla uygulanabilir.

KAYNAKLAR

Ahmed, S.S.J., Santosh, W., Kumar, S., Christlet, T.H.T. 2010. "Neural network algorithm for the early detection of Parkinson's disease from blood plasma by FTIR micro-spectroscopy", *Vibrational Spectroscopy* 53, 181-188.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders. (Fifth Edition)* Arlington, VA, USA American Psychiatric Association.

American Psychiatric Association. 2000. *American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders. (Fourth Edition) (Revised)*. Washington DC (U.S.A.)

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders. (Fifth Edition)*. Arlington. VA (U.S.A.).

Atkins, C.G., Buckley, K., Blades, M.W., Turner, R.F.B. 2017. "Raman Spectroscopy of Blood and Blood Components", *Appl. Spectrosc.*, 71 767.

Bahn, S., Schwarz, E. 2012. "Serum-based biomarkers for psychiatric disorders", *Nervenarzt*, 82, 1395.

Ballabio, D., Consonni, V. 2013. "Classification tools in chemistry. Part 1: linear models. PLS-DA", *Analytical Methods*, 5, 3785-3789.

Barker, M., Rayens, W. 2003. "Partial least squares for discrimination", *J. Chemom.*, 17, 166.

Bayari Haman S., Sevinc, A., Tiftikcioglu Irem B., Zorlu, Y., Severcan, F. 2015. Rio de Janeiro, Brazil. 9th IBRO World Congress on Neuroscience.

Bayari Haman S., Krafft C., İde, S., Poop, J., Guven, G., Cehreli, Z.C., Soylu, E.H. 2012. "Investigation of adhesive-dentin interface using Raman microspectroscopy and SAXS", *Journal of Raman Spectroscopy* , 43 6-15.

Bayari S., Severcan, F. 2005. "FTIR study of biodegradable biopolymers: P (3HB), P(3HB-co-4HB)and P(3HB-co-4HV)", *J.Mol.Struct.*, 744-747, 529-534.

Bayari Haman S, Utku H., Ikemoto, Y., Celasun, B., Kömürçü, M., Atik, A. 2007. "Synchrotron FT-IR microspectroscopic analysis of necrotic bone", *Spectroscopy: An International Journal*, 21(4), 227-234.

Bayari Haman S. 2012. "Applications of Vibrational Spectroscopy in Diagnosis and Screening of Dental Tissues, *Vibrational Spectroscopy in Diagnosis and Screening*", Editorler: Severcan F, Haris P.I. IOS Press, 321 – 349.

Bellivier, F., Geoffroy, P.A., Scott, J., Schurhoff, F., Leboyer, M., Etain, B. 2013. "Biomarkers of bipolar disorder: specific or shared with schizophrenia?", *Front. Biosci.*, E5, 845.

Berrettini, W. 2003. "Evidence for shared susceptibility in bipolar disorder and schizophrenia", *Am. J. Med. C Semin. Med. Genet.*, 123C, 59.

Bright, A., Devi, R.S.T., Gunasekaran, S. 2011. "Spectroscopical vibrational band assignment and qualitative analysis of biomedical compounds with cardiovascular activity", *Int. J. Pharma and Biosci.* 2 3 , 346-354.

Bozkurt, O., Bayari Haman S., Severcan, M., Krafft, C., Popp J., Severcan F. 2012. "Structural alterations in rat liver proteins due to STZ-induced diabetes and the

recovery effect of selenium: An FTIR microspectroscopy and neural network study”,
Journal of Biomedical Optics 17(7), 076023

C.V. Raman and K.S. Krishnan, Nature, 121, 501 (1928).

Carey, P.R., Biochemical Applications of Raman and Resonance Raman Spectroscopies, Academic Press (1982)]

Campos, N.D.S., Oliveira, K.S., Almeida, M.R., Stephani, R., Oliveira, L.F.C. 2014. “Classification of Frankfurters by FT-Raman Spectroscopy and Chemometric Methods”, Molecules, 19, 18980.

Carpenter, W. T., Tandon, R. 2013. “Psychotic disorders in DSM-5: summary of changes”, Asian journal of psychiatry, 6(3), 266-268.

Chao, L., Mei, H., Pan, X.C., Tan, W., Liu, T.F., Yang, L. 2014. “Combinations of fragment descriptors for improved prediction of CYP2C19 inhibitors”, Chemom. Intell. Lab. Syst. , 130, 109.

Conti, C., Giorgini, E., Pieramici, T., Rubini, C., Tosi, G. 2005. “FT-IR microscopy imaging on oral cavity tumours, II”, J. Mol. Struc. 744-747, 187-193.

Correll, C.U., Penzner, J.B., Frederickson, A.M., Richter, J.J., Auther, A.M., Smith, C.W., Kane, J.M., Cornblatt, B.A. 2007. “Differentiation in the preonset phases of schizophrenia and mood disorders: evidence in support of a bipolar mania prodrome”, Schizophr. Bull. , 33, 703.

Çağlayan, Çağlar "Fluorourasil molekülünün su komplekslerinin titreşim frekanslarının AB-INITO DFT yöntemi ile hesaplanması", Y.Lisans Tezi. TC. İstanbul Kültür Üniversitesi Fen bilimleri Enstitüsü, İstanbul, 2015

De Vries, S., Ter Braak, J.F. 1995. "Prediction error in partial least squares regression: a critique on the deviation used in The Unscrambler", *Chemom. Intell. Lab. Syst.*, 30, 239.

Erukhimovitch, V., Talyshinsky, M., Souprun, Y., Huleihel, M. 2006. "*FTIR spectroscopy examination of leukemia patients plasm*", *Vibrational Spectroscopy*, 40, 40-46.

E. Bair, T. Hastie, D. Paul and R. Tibshirani, Prediction by Supervised Principal Components. *J. Am. Stat. Assoc.*, vol. 101, pp. 119–137, 2006.

Esbensen, K.H. 2005. In: *Multivariate Data Analysis - In Practice*. (5th Edition). Esbjerg, Denmark: CAMO Process AS.

WHO Region. 2000-2012. *Global Health Estimates 2014 Summary Tables: YLD by cause, age and sex*.

González-Pinto, A., Gutierrez, M., Mosquera, F., Ballesteros, J., López, P., Ezcurra, J., Figuerido, J.L., de Leon, J. 1998. "First episode in bipolar disorder: misdiagnosis and psychotic symptoms", *J. Affect. Disord.*, 50, 41.

H. Hotelling, Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.*, vol. 24, pp. 417–441, 1933.

H. Hotelling, (1933). Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.*, vol. 24, pp. 498–520, 1933.

H. Hotelling, Relations Between Two Sets of Variates. *Biometrika*, vol. 28, pp. 321–377, 1936.]

H. Wold, Estimation of Principal Components and Related Models by Iterative Least Squares. In: P. R. Krishnaiah Ed., *Multivariate Analysis*. Academic Press, New York, USA, pp. 391–420, 1966.

Ildiz, G.O., Bayari, Aksoy U.M., Yoruguner, N., Bulut H., Yilmaz, S.S., Halimoglu, G., Kabuk N.H., Yavuz G., Fausto, R. 2020. “Auxiliary Differential Diagnosis of Schizophrenia and Phases of Bipolar Disorder Based on the Blood Serum Raman Spectra”, *Journal of Raman Spectroscopy*, 20-0137.

Ildiz, G.O., Bayari, S., Karadag, A., Kaygisiz, E., Fausto, R. 2020. “Fourier Transform Infrared Spectroscopy Based Complementary Diagnosis Tool for Autism Spectrum Disorder in Children and Adolescents”, *Biomedical Applications of Infrared and Raman Spectroscopy, Molecules*, 25 (9), 2079.

Ildiz, G.O., Arslan, M., Unsalan, O., Kurt, E., Karatepe, H.T., Yilmaz, A., Yalcinkaya, O.B., Herken, H. 2016. “FT-IR spectroscopy and multivariate analysis as an auxiliary tool for diagnosis of mental disorders: bipolar and schizophrenia cases”, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 152, 551-556.

Ildiz, G.O., S. Bayari, S., Yorguner N., Fausto R. 2020. *Blood Serum Infrared Spectra Based Chemometric Models for Auxiliary Diagnosis of Autism Spectrum Disorder*. *Autism Spectrum Disorder: Diagnosis and Treatment*. Amsterdam, Hollanda: Elsevier. (Baskıda)

J. H. Ward, Hierarchical Grouping to Optimize an Objective Function, *J. Am. Stat. Assoc.*, Vol. 58, pp. 236–244, 1963.

Joyce, P. R. 1984. “Age of onset in bipolar affective disorder and misdiagnosis as schizophrenia”, *Psychol. Med.*, 1984,14, 145.

Kanagathara, N., Thirunavukkarasu, M., Jeyanthi, C.E., Shenbagarajan, P. 2011. "FTIR and UV-visible spectral study on normal blood samples", Int. J. Pharm. Bio. Sci. 1, 74-81

K. Pearson, K, On Lines and Planes of Closest Fit to Systems of Points in Space. Phil. Magazine., vol. 2, pp. 559–572, 1901.

Köksal, F. and Köseoğlu, R., 2010. *Spektroskopi Ve Lazerlere Giriş*. 1nci basım. Ankara: Nobel yayın dağıtım, sy.209-216.

L. C. Lee, C.-Y. Liong and A. A. Jemain, Partial Least Squares-Discriminant Analysis (PLS-DA) for Classification of High-Dimensional (HD) Data: A Review of Contemporary Practice Strategies and Knowledge Gaps. Analyst, vol. 143, pp. 3526–3539, 2018.

Lobo, D.M., Agius, M. 2012. "The mental illness spectrum ", Psychiatria Danubina 24 (1) 157-160.

Luxburg, U. 2007. "A tutorial on spectral clustering, Statistics and Computing", 17(4), 395-416.

Martens, C., Murray, L. 1989. "The abnormal T lymphocytes in lpr mice transcribe interferon- γ and tumor necrosis factor- α genes spontaneously in vivo", European Journal of Immunology, 41.

Meyer, F., Meyer, T.D.2009. "The misdiagnosis of bipolar disorder as a psychotic disorder: some of its causes and their influence on therapy", J. Affect. Disord.,112, 174.

Miller, C.E., Naes, T. 1990. "A Pathlength Correction Method for Near-Infrared Spectroscopy", Appl. Spectrosc., 44, 895.

Mueser, K. T., Drake, R. E., Wallach, M. A.1998. "Dual diagnosis: a review of etiological theories", Addict. Behav, 23, 717.

Murray, M.R., Sham, P., Os, V.J., Zanelli, J., Cannon, M., McDonald, C. 2004. "A developmental model for similarities and dissimilarities between schizophrenia and bipolar disorder", *Schizophrenia Research* 71, 405 -416.

Murtagh F., Legendre, P. 2014. "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?", *J. Classification*, 31, 274.

Nargis, H.F., Nawaz, H., Ditta, A., Mahmood, T., Majeed, M.I., Rashid, N., Muddassar, M., Bhatti, H.N., Saleem, M., Jilani, K., Bonnier, F., Byrne, H.J. 2019. "Raman spectroscopy of blood plasma samples from breast cancer patients at different stages", *Spectrochim Acta A Mol Biomol Spectrosc*, 222, 117210.

Nusslock, R., Young, C.B., Damme, K.S.F. 2014. "Elevated reward-related neural activation as a unique biological marker of bipolar disorder: assessment and treatment implications", *Behav. Res. Ther.*, 62, 74.

OMNIC (Version 8.3.103) "Thermo Fisher Scientific Inc.". MA (U.S.A.), Waltham.

Palazzo, M.C., Arici, C., Cremaschi, L., Cristoffanini, M., Dobrea, C., Dell'Osso, B., Altamura, A.C. 2017. "Cognitive Performance in Euthymic Patients with Bipolar Disorder vs Healthy Controls: A Neuropsychological Investigation", *Clin. Pract. Epidemiol. Ment. Health.*, 13, 71.

Öztürk, Tuğçe P., "Embriyo Atık Kültür Sıvısının Raman Spektroskopisi İle Analizi", Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi.Fen Bilimleri Enstitüsü, İstanbul, Mayıs, 2015.

Pearlson, G.D. 2015. "Etiologic, phenomenologic, and endophenotypic overlap of schizophrenia and bipolar disorder", *Annu. Rev. Clin. Psychol.*, 2015 11, 251.

Pérez, A., Prada, Y.A., Cabanzo, R., González, C.I., Mejía-Ospín, E. 2018. "Diagnosis of chagas disease from human blood serum using surface-enhanced Raman scattering (SERS) spectroscopy and chemometric methods", *Sens. Bio-Sens. Res.*, 21, 40.

Rouaud, Mathieu (2013). *Probability, Statistics and Estimation*, sy.7.

Saade, J., Silva, J.N., Farias, P.M.A., Lopes, D.F., Santos, C.T., Farias, B.A., Rodrigues K.C., Martin, A.A. 2012. "Glicemical Analysis of Human Blood Serum Using FT-Raman: A New Approach", *Photomed. Laser Surg.*, 30, 388.

Scola, G., Andreatza, A. C. 2014. "Current State of Biomarkers in Bipolar Disorder", *Curr. Psychiatry Rep.*, 16, 514.

Shao, L., Zhang, A., Rong, Z., Wang, C., Jia, X., Zhang, K., Xiao, R., Wang, S. 2018. "Fast and non-invasive serum detection technology based on surface-enhanced Raman spectrond multivariate statistical analysis for liver disease", *Nanomed.: Nanotech. Biol. Med.*, 14, 451.

Sheng, D., Liu, X., Li, W., Yang, Y., Chen, X., Wang, X. 2013. "Distinction of leukemia patients' and healthy persons' serum using FTIR spectroscopy", *Spectrochim. Acta Part A: Mol. and Biomol. Spectrosc.*, 101, 228–232

Strakowski, S. M. 2003. "Diagnostic boundaries between bipolar disorder and schizophrénia: Implications for pharmacologic intervention", *Adv. Stud. Med.* 3", 792–798.

Strakowski, Stephen M. 2003. "Diagnostic boundaries between bipolar disorder and schizophrénia: implications for pharmacologic intervention", Johns Hopkins University School of Medicine Advanced Studies in Medicine 3.8C. S792-S798.

Strakowski S. 1999. *The Co-occurrence of Depression in Schizophrénia*. London: Science Press, Ltd.

Strakowski, S. M., Williams, J. R., Sax, K. W., Fleck, D. E., DelBello, M. P., Bourne, M. L. 2000. "Is impaired outcome following a first manic episode due to mood-incongruent psychosis?", *Journal of affective disorders*, 61(1), 87-94.

Strakowski, Stephen M. 1994. "Diagnostic validity of schizophreniform disorder", *American Journal of Psychiatry*, 151.6, 815-824.

Smekal A G. The quantum theory of dispersion. *Nature*, 1924,114(2861):310-311.

Sudworth, D.C., Archer, J.K. John, Mann, D. 2005. "Near infrared Raman spectroscopy for Alzheimer's disease detection", *Proc. SPIE 5862*, 586203.

S. Wold, M. Sjöström and L. Eriksson, *PLS-Regression: A Basic Tool of Chemometrics*. *Chemom. Intell. Lab. Sys.*, vol. 58, pp. 109–130, 2001.

Tas, C., Brown, E.C., Aydemir, O., Brüne, M., Lysaker, P.H. 2014. "Metacognition in psychosis: comparison of schizophrenia with bipolar disorder", *Psychiatry Res.*, 219, 464.

The Unscrambler (Version 5.5). 1995. CAMO A/S, Trondheim, Norway.

Tomasik, J., Schwarz, E., Guest, P.C., Bahn S. 2012. "Blood test for schizophrenia", *Eur. Arch. Psychiatry Clin. Neurosci*, 262, 79.

Torrey, E.F. 1999. "Epidemiological comparison of schizophrenia and bipolar disorder", *Schizophr. Res*, 39, 101.

Torrey, E. F., Knable, M. B. 1999. "Are schizophrenia and bipolar disorder one disease or two?", *Schizophr. Res.*, 39, 93.

Torres, I.J., Boudreau, V.G., Yatham, L.N. 2007. "Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis", *Acta Psychiatr. Scand. Suppl.*, 434, 17.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Sydney, Australia: Addison-Wesley.

Uyanık, A. Miler, J. and Miler, J., 2012. "Analitik kimya için istatistik ve kemometri".(6. Baskı). Türkiye: Ankara: Pegem akademi, pp.290-320.

W. Berrettini, *Am. J. Med. C Semin. Med. Genet.* 2003, 123C, 59.

Ward, J.H. 1963. "Hierarchical Grouping to Optimize an Objective Function", *J. Am. Stat. Assoc.*, 58, 236.

Wold, H. 1975. *Path models with latent variables: The NIPALS approach*, in H.M. Blalock Ed. *Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building*. NY (U.S.A.): Academic Press. 307-357.

V. Ramesh, Ed. 2019. *Biomolecular and Bioanalytical Techniques: Theory*. Wiley, Hoboken, NJ: *Methodology and Applications*.

Yang, E., Tadin, D., Glasser, D., Hong, S. W., Blake, R., Park, S. 2013. "Visual context processing in bipolar disorder: a comparison with schizophrenia", *Front. Psychol.*, 4, 569.

Yurdakul, Ş., 2010. *Spektroskopi Ve Grup Teorisinin Temelleri*. Ankara: Gazi Kitabevi, sy.77-81.

Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A. 1978. "A rating scale for mania: reliability, validity and sensitivity", *Brit. J. Psych.*, 133, 429.