T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

YAKIN KIZILÖTESİ SPEKTROSKOPİSİNİN ODUN TEŞHİSİNDE KULLANIMI

Fatma Diğdem TUNCER

DANIŞMAN
Prof. Dr.Ayşe Dilek DOĞU

II. DANIŞMAN
Doç. Dr.Esra AKDENİZ

Orman Endüstri Mühendisliği Anabilim Dalı

Orman Endüstri Mühendisliği Programı

İSTANBUL-2020

Bu çalışma 24.04.2020 tarihinde aşağıdaki jüri tarafından Orman Endüstri Mühendisliği Anabilim Dalı, Orman Endüstri Mühendisliği Programında Doktora Tezi olarak kabul edilmiştir.

TEZ JÜRİSİ

Prof. Dr. Ayşe Dilek DOĞU
İstanbul Üniversitesi-Cerrahpaşa
Orman Fakültesi

Prof. Dr. Nural YILGÖR
İstanbul Üniversitesi-Cerrahpaşa
Orman Fakültesi

Prof. Dr. Bilgin İÇEL
Çanakkale Onsekiz Mart Üniversitesi
Çanakkale Teknik Bilimler MYO

Prof. Dr. Coşkun KÖSE
İstanbul Üniversitesi-Cerrahpaşa
Orman Fakültesi

Doç. Dr. Ozan KOCADAĞLI
Mimar Sinan Üniversitesi
Fen Edebiyat Fakültesi

20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi-Cerrahpaşa'nın abonesi olduğu intihal yazılım programı kullanılarak Lisansüstü Eğitim Enstitüsü'nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

Bu tez, 1059B141800316 numaralı TÜBİTAK 2214-A Yurt Dışı Araştırma Burs Programı (Doktora Ögrencileri için) projesi ile desteklenmiştir.

ÖNSÖZ

"Yakın Kızılötesi Spektroskopisinin Odun Teşhisinde Kullanımı" isimli bu çalışma İstanbul Üniversitesi - Cerrahpaşa, Orman Fakültesi, Orman Endüstri Mühendisliği Anabilim Dalı, Orman Endüstri Mühendisliği Programı'nda Doktora Tezi olarak hazırlanmıştır. Bu çalışmanın bir kısmı TÜBİTAK 2214-A Yurt Dışı Araştırma Burs Programı ile desteklenerek Wisconsin Üniversitesi-Madison, ABD 'de gerçekleştirilmiştir.

"Ateşin her bir noktası

bütün öteki noktalarıyla

ilişkilidir" – Oruç Aruoba.

Bu sebepledir ki; danışmanlığımı üstlenmesinin yanı sıra tüm endişelerim ile ilgilenen Prof. Dr. A. Dilek DOĞU 'ya, yoğun çalışma temposuna rağmen ikinci danışmanlığımı kabul eden Doç. Dr. Esra AKDENİZ 'e, tez izleme jürimde bulunarak destek ve yardımlarını esirgemeyen sevgili hocalarım Prof. Dr. Nural YILGÖR ve Prof. Dr. Bilgin İÇEL 'e, beni hep bir sonraki adıma hazırlayan sevgili hocam Prof. Dr. Coşkun KÖSE 'ye, aynı anabilim dalında çalışma imkanı bulduğum bilgi ve tecrübeleriyle her zaman yanımda olan değerli hocalarım Prof. Dr. S. Nami KARTAL, Doç. Dr. Evren TERZİ, Dr. Öğr. Üyesi M. Osman ENGÜR, Arş. Gör. Dr. Davut BAKIR 'a, TÜBİTAK 2214-A Doktora Öğrencileri için Yurt Dışı Araştırma Bursu aracılığıyla çalışmalarımı gerçekleştirmemi sağlayan TÜBİTAK' a, Amerika Birleşik Devletlerde kaldığım süre boyunca benim tatlı ev sahibim olan Mary COLLET 'e, daima yanımda olan sevgili arkadaşım Adriana COSTA 'ya, farklı bilimsel disiplinlerden ve dünyanın dört bir yanından gelen çalışma ekibine beni de kabul eden, farklı yaklaşımları ile ufkumu genişleten Doç Dr. Alex C. Wiedenhoft 'a ve bir süreliğine benimde parçası olduğum USDA-CWAR laboratuvarındaki çalışma arkadaşlarıma, İ.Ü.C., Orman Fakültesi Ahşap Atölyesi çalışanlarına, örneklerin alımında her türlü desteği sağlayan Orman Genel Müdürlüğü çalışanlarına, her koşulda yanımda olan sevgili arkadaşım Arş. Gör. Alper Gün ÖZTURNA 'ya ve ismini sığdıramadığım ama benim için oldukça değerli onlarca kişiye desteklerinden ve yardımlarından dolayı çok teşekkür ederim.

"Yola çıkan kişinin tek 'yardımcı'sı,

Yolu, yanında, onunla birlikte yürüyendir

-yoldaştır…" – Oruç Aruoba.

Ve işte bu sebepledir ki; sevgili eşim Ergün TUNCER 'e, karnımdaki henüz ismi konulmamış kızlarıma, daimi desteklerim annem Sevim KANGALOĞLU ve babam Ahmet Cafer KANGALOĞLU 'na sonsuz şükranlarımı sunarım.

Nisan 2020 Fatma Diğdem TUNCER

İÇİNDEKİLER

Sayfa No

ÖNSÖZ ..... **iv**
İÇİNDEKİLER ..... **v**
ŞEKİL LİSTESİ ..... **viii**
TABLO LİSTESİ ..... **xii**
SİMGE VE KISALTMA LİSTESİ ..... **xvi**
ÖZET ..... **xix**
SUMMARY ..... **xxi**
1. GİRİŞ **1**
2. GENEL KISIMLAR **4**
2.1. AĞAÇ TÜRÜ TEŞHİSİ 4
2.1.1. Ağaç Türü Teşhisinde Yakın Kızılötesi Spektroskopisinin Kullanımı 7
2.2. YAKIN KIZILÖTESİ SPEKTROSKOPİSİ 8
2.3. KEMOMETRİ 10
2.4. VERİ ANALİZİ 11
2.4.1. Ön İşlem Metotları 13
2.4.2. Kalibrasyon Metotları 18
2.4.2.1. Kümeleme Analizi 18
2.4.2.2. Temel Bileşenler Analizi 21
2.4.2.3. Regresyon 22
2.4.2.4. Sınıflandırma 23
2.4.2.5. Diskriminant Analizi 26
2.4.2.6. Kısmi En Küçük Kareler (PLS) 27
2.4.2.7. Sınıf karşılaştırmasının hafif bağımsız modellemesi (SIMCA) 29
2.4.2.8. En Yakın Komşu Metodu (k-NN) 29
2.4.2.9. Ağaç Temelli Yaklaşımlar 30
2.4.2.10. Doğrusal Olmayan Metotlar 30
2.4.3. Model Performansı 34
2.4.4. Yorumlama 43
3. MALZEME VE YÖNTEM **47**
3.1. MALZEME 47

3.2. YÖNTEM 55
3.2.1. Örneklerin Hazırlanması 55
3.2.2. Kullanılan Program ve Paketler 59
3.2.3. Kullanılan Analizler 60
3.2.3.1. Temel Bileşenler Analizi (PCA) 61
3.2.3.2. Çarpımsal Saçılım Düzeltmesi (MSC) 62
3.2.3.3. Standart Normal Değişim (SNV) 63
3.2.3.4. Savitzky-Golay Algoritması ile Düzgünleştirme ve Türev Alma 63
3.2.3.5. Kısmi En Küçük Kareler Diskriminant analizi (PLS-DA) 64
3.2.3.6. Diyagonal Doğrusal Diskriminant Analizi 65
3.2.3.7. Karar Ağaçları (DT) 66
3.2.3.8. Yapay Sinir Ağları (ANN) 66
3.2.3.9. Destek Vektör Makineleri (SVM) 67
3.2.3.10. Model performans değerlendirmesi ve Yorumlama 67
3.2.3.11. Yorumlama 68
4. BULGULAR 69
4.1. İŞLEM GÖRMEMİŞ SPEKTRAL VERİLER 69
4.2. ÖN İŞLEM UYGULANMIŞ SPEKTRAL VERİLER 71
4.3. MODEL PERFORMANSLARI 83
4.3.1. Kısmi en küçük kareler diskriminant analizi 84
4.3.2. Diyagonal doğrusal diskriminant analizi 87
4.3.3. Karar ağaçları 89
4.3.4. Yapay sinir ağları 92
4.3.5. Destek vektör makineleri 95
4.4. YORUMLAMA 97
5. TARTIŞMA VE SONUÇ 104
KAYNAKLAR 117
EKLER 129
EK 1. Çalışmada kullanılan veri formatı. 129
EK 2. Cins ayrımında kullanılan PLS-DA model parametreleri (Ham veri seti). 130
EK 3. Çam türlerinin ayrımında kullanılan PLS-DA model parametreleri (Ham veri seti). 131
EK 4. Meşe türlerinin ayrımında kullanılan PLS-DA model parametreleri (Ham veri seti). 131
EK 5. Cins ayrımında kullanılan PLS-DA korelasyon tabloları (Ham veri seti). 132

EK 6. Çam türlerinin ayrımında kullanılan PLS-DA korelasyon tabloları (Ham veri seti). 133
EK 7. Meşe türlerinin ayrımında kullanılan PLS-DA korelasyon tabloları (Ham veri seti). 134
EK 8. Çalışmada kullanılan DT model parametreleri (Ham veri seti). 135
EK 9. Çalışmada kullanılan DT modellerine ait karmaşıklık parametreleri (CP) (Ham veri seti). 136
EK 10. Çalışmada cins bazında ayrımda kullanılan ANN model parametrelerinin grafiksel gösterimi (Ham veri seti). 137
EK 11. Çalışmada kullanılan ANN model parametreleri (Ham veri seti). 138
EK 12. Çalışmada kullanılan SVM model parametreleri (Ham veri seti). 139
EK 13. Cins ayrımında PLS-DA modelinde kullanılan Dalga boylarının SR aracılığıyla hesaplanan önem düzeyleri sıralaması (ilk 20 dalga boyu) (Ham veri seti). 139
EK 14. Çam türlerinin ayrımında PLS-DA modelinde kullanılan Dalga boylarının SR aracılığıyla hesaplanan önem düzeyleri sıralaması (ilk 20 dalga boyu) (Ham veri seti). 140
EK 15. Meşe türlerinin ayrımında PLS-DA modelinde kullanılan Dalga boylarının SR aracılığıyla hesaplanan önem düzeyleri sıralaması (ilk 20 dalga boyu) (Ham veri seti). 140
EK 16. DT modellerinde kullanılan dalga boylarının değişken önem sıralamaları (Ham veri seti). 141
ÖZGEÇMİŞ 142

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Sarıçam türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200). 5

Şekil 2.2: Karaçam türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200). 5

Şekil 2.3: Macar meşesi türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200).....6

Şekil 2.4: Sapsız meşe türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200). 6

Şekil 2.5: Elektromanyetik spektrum (Davies, 2005; Baier, 2012). 9

Şekil 2.6: Kemometri bilim dalının içeriği (Varmuza ve Filzmoser, 2008). 11

Şekil 2.7: Veri analizi süreci (Schutt ve O'Neil, 2014). 12

Şekil 2.8: Çok değişkenli veri analiz metotları (Brereton, 2007; Elmasry ve ark., 2012; Patel ve Patel, 2014).* 13

Şekil 2.9: Ön işlem amacıyla kullanılan bazı metotların sınıflandırılması (Gholizadeh ve ark., 2015; Rinnan ve ark., 2009). 14

Şekil 2.10: Sınıflandırma analizi şeması (Varmuza ve Filzmoser, 2008).....25

Şekil 2.11: Sınıflandırma amaçlı kullanılan bazı modeller (Brereton, 2007; Nisbet ve ark., 2018; Olivieri, 2018; Varmuza ve Filzmoser, 2008; Wehrens, 2011). 26

Şekil 2.12: Yapay Sinir Ağlarının şematik gösterimi (Varmuza ve Filzmoser, 2008). 31

Şekil 2.13: Çapraz Sorgulamanın (k=3 için) şematik görünümü (Cochrane, 2019). 37

Şekil 2.14: Model karmaşıklığı ve tahmin hatası ilişkisi (Hastie ve ark., 2017, James ve ark., 2013, Varmuza ve Filzmoser, 2008). 38

Şekil 2.15: Karar Vericinin Etkinliği grafiği (James ve ark., 2013). 41

Şekil 2.16: Değişken seçim yöntemlerinin sınıflandırması (Mehmood ve ark., 2012).....44

Şekil 3.1: Sarıçam (a), karaçam (b), sapsız meşe (c) ve Macar meşesi (d) ağaç türlerine ait sürgün fotoğrafı. 48

Şekil 3.2: Disk örnekleri (a) ve artım burgusu aracılığıyla temin edilen artım kalemleri (b). 49

Şekil 3.3: Disk örneklerinin boyutlandırılması. 55

Şekil 3.4: Eclund Brinell Mikroskobu ve artım burgusu örneği. 56

Şekil 3.5: Enine kesit boyunca ölçüm alanlarının belirlenme şablonu. 56

Şekil 3.6: NIRS ölçümleri için disk örneklerinin hazırlanması ve örnek alanların belirlenmesi. 57

Şekil 3.7: Artım kalemlerinin ölçümlere hazırlanması. 57

Şekil 3.8: Antaris FT-NIR Analyser cihazında Yakın Kızılötesi ölçümlerinin alınması.....59

Şekil 3.9: Çalışmada kullanılan ön işlemler ve modeller*.....61

Şekil 4.1: Çam (sarıçam, karaçam) ve meşe (sapsız meşe, Macar meşesi) cinslerine ait yansıtma (1/R) değerleri: a) toplu gösterim b) Çam türleri, c) Meşe türleri (yeşil: sarıçam, siyah: karaçam, pembe: sapsız meşe, gri: Macar meşesi). 69

Şekil 4.2: Sarıçam ve karaçam ağaç türlerine ait yansıtma (1/R) değerleri, a) toplu gösterim (yeşil: sarıçam, siyah: karaçam), b) sarıçam, c) karaçam. 70

Şekil 4.3: Sapsız meşe ve macar meşesi ağaç türlerine ait yansıtma (1/R) değerleri, a) toplu gösterim (pembe: sapsız meşe, gri: Macar meşesi), b) sapsız meşe, c) Macar meşesi. 70

Şekil 4.4: Çam ve meşe ağaç türlerine ait veri setinde SNV dönüşümü. 71

Şekil 4.5: Çam ağaç türlerine ait veri setinde SNV dönüşümü. 72

Şekil 4.6: Meşe ağaç türlerine ait veri setinde SNV dönüşümü. 72

Şekil 4.7: Çam ve meşe ağaç türlerine ait veri setinde MSC dönüşümü. 73

Şekil 4.8: Çam ağaç türlerine ait veri setinde MSC dönüşümü. 73

Şekil 4.9: Meşe ağaç türlerine ait veri setinde MSC dönüşümü. 74

Şekil 4.10: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türev dönüşümü. 74

Şekil 4.11: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türev dönüşümü. 75

Şekil 4.12: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türev dönüşümü. 75

Şekil 4.13: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere SNV uygulanması. 76

Şekil 4.14: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere SNV uygulanması. 76

Şekil 4.15: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere SNV uygulanması. 77

Şekil 4.16: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere MSC uygulanması. 77

Şekil 4.17: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere MSC uygulanması. 78

Şekil 4.18: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere MSC uygulanması. 78

Şekil 4.19: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme. 79

Şekil 4.20: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme. 79

Şekil 4.21: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme. 80

Şekil 4.22: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma. 80

Şekil 4.23: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma. 81

Şekil 4.24: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma. 81

Şekil 4.25: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma. 82

Şekil 4.26: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma. 82

Şekil 4.27: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma. 83

Şekil 4.28: Cins ayrımında PLS-DA skor ve yükleme grafikleri (ham veri seti). 98

Şekil 4.29: Çam türlerinin ayrımında PLS-DA skor ve yükleme grafikleri (ham veri seti). 98

Şekil 4.30: Meşe türlerinin ayrımında PLS-DA skor ve yükleme grafikleri (ham veri seti). 99

Şekil 4.31: Hiç işlem görmemiş çam ve meşe cinsleri için PLS-DA modelinde kullanılan değişkenler. a: 5690 – 6000 cm^{-1}, 1: 4345 cm^{-1}, 2: 4517 cm^{-1}, 3: 5000 cm-1, 4: 8306 cm^{-1}.....100

Şekil 4.32: Hiç işlem görmemiş çam türleri için PLS-DA modelinde kullanılan değişkenler. a: 5680 - 6100cm^{-1}, b:7444 - 8788cm^{-1}, 1: 4336 cm^{-1}, 2: 4504 cm^{-1}, 3: 4966cm^{-1}, 4: 5932cm^{-1}, 5: 7360cm^{-1}. 100

Şekil 4.33: Hiç işlem görmemiş meşe türleri için PLS-DA modelinde kullanılan değişkenler. a: 5930 – 6205 cm^{-1}, b: 7445 – 9170 cm^{-1}, 1: 4517 cm^{-1}, 2: 5310 cm^{-1}, 3: 7617 cm^{-1}, 4: 8823 cm^{-1}.....101

Şekil 4.34: Çam ve meşe cins ayrımında ön işlem görmemiş veri ile DT uygulamaları. 102

Şekil 4.35: Çam türlerinin ayrımında DT uygulamaları. 0: sarıçam, 1:karaçam. 102

Şekil 4.36: Meşe türlerinin ayrımında DT uygulamaları.0: macar meşesi, 1: sapsız meşe. 103

Şekil 5.1: Çam ve meşe cinsleri için ortalama spektrumlar. a: 4180 – 4450 cm^{-1}, b: 4570 – 5255 cm^{-1} , c: 5500 – 7105 cm^{-1}. 112

Şekil 5.2: Çam türleri için ortalama spektrumlar. a: 4167 - 4400 cm^{-1}, b: 4500 – 5267 cm^{-1}, c: 5500 – 6000 cm^{-1}, d: 6267 -7000 cm^{-1}, 1:4733 cm^{-1} , 2:5167 cm^{-1} , 3: 6833 cm^{-1} 113

Şekil 5.3: Meşe türleri için ortalama spektrumlar. a: 4180 – 4420 cm^{-1}, b: 4540 – 5200 cm^{-1}, c: 5380 – 6570 cm^{-1}. 113

TABLO LİSTESİ

Sayfa No

Tablo 2.1: NIRS ile ağaç türü teşhisi konusunda gerçekleştirilen çalışmalarda geçmişten günümüze kullanılan ön işlem metotları. 17

Tablo 2.2: Kümeleme analizi metotları ve en sık kullanılan algoritmalar 20

Tablo 2.3: Regresyon modelleri için x ve y veri tiplerine göre uygun modeller.....23

Tablo 2.4: NIRS ile ağaç türü teşhisi konusunda gerçekleştirilen çalışmalarda geçmişten günümüze kullanılan çok değişkenli analiz metotları 33

Tablo 2.5: Hata matrisi gösterimi. 39

Tablo 2.6: NIRS ile ağaç türü teşhisi konusunda gerçekleştirilen çalışmalarda geçmişten günümüze kullanılan performans değerleri. 42

Tablo 3.1: Çam türlerinin belirlenmesinde kullanılan dış morfolojik özellikler. 47

Tablo 3.2: Meşe türlerinin belirlenmesinde kullanılan dış morfolojik özellikler. 48

Tablo 3.3: Sarıçam ağaç türüne ait kimlik verileri. 51

Tablo 3.4: Karaçam ağaç türüne ait örneklerin kimlik verileri. 52

Tablo 3.5: Sapsız meşe ağaç türüne ait kimlik verileri. 53

Tablo 3.6: Macar meşesi ağaç türüne ait kimlik verileri. 54

Tablo 3.7: NIRS ölçümleri için belirlenen ölçüm alanlarının kimliklendirilmesi. 58

Tablo 3.8: Çalışmada kullanılan ağaç ve örnek sayıları. 59

Tablo 3.9: Savitzky-Golay algoritmasında kullanılan parametreler. 64

Tablo 3.10: Hata matrisi. 67

Tablo 4.1: Eğitim ve test setlerine ayrılan örnek sayıları. 84

Tablo 4.2: Çam ve meşe cins ayrımında eğitim veri seti ve PLS-DA ile kurulan sınıflandırma modelinde elde edilen değerler. 84

Tablo 4.3: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve PLS-DA ile kurulan sınıflandırma modelinde elde edilen değerler. 85

Tablo 4.4: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve PLS-DA ile kurulan sınıflandırma modelinde elde edilen değerler. 85

Tablo 4.5: Çam ve meşe cins ayrımında PLS-DA modelinin test veri seti ile değerlendirilmesi. 86

Tablo 4.6: Sarıçam ve karaçam türlerinin ayrımında PLS-DA modelinin test veri seti ile değerlendirilmesi. 86

Tablo 4.7: Sapsız meşe ve Macar meşesi türlerinin ayrımında PLS-DA modelinin test veri seti ile değerlendirilmesi. 86

Tablo 4.8: Çam ve meşe cins ayrımında eğitim veri seti ve DL-DA ile kurulan sınıflandırma modelinde elde edilen değerler. 87

Tablo 4.9: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve DL-DA ile kurulan sınıflandırma modelinde elde edilen değerler. 87

Tablo 4.10: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve DL-DA ile kurulan sınıflandırma modelinde elde edilen değerler. 88

Tablo 4.11: Çam ve meşe cins ayrımında DL-DA modelinin test veri seti ile değerlendirilmesi. 88

Tablo 4.12: Sarıçam ve karaçam türlerinin ayrımında DL-DA modelinin test veri seti ile değerlendirilmesi. 89

Tablo 4.13: Sapsız meşe ve Macar meşesi türlerinin ayrımında DL-DA modelinin test veri seti ile değerlendirilmesi. 89

Tablo 4.14: Çam ve meşe cins ayrımında eğitim veri seti ve DT ile kurulan sınıflandırma modelinde elde edilen değerler. 90

Tablo 4.15: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve DT ile kurulan sınıflandırma modelinde elde edilen değerler. 90

Tablo 4.16: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve DT ile kurulan sınıflandırma modelinde elde edilen değerler. 90

Tablo 4.17: Çam ve meşe cins ayrımında DT modelinin test veri seti ile değerlendirilmesi. 91

Tablo 4.18: Sarıçam ve karaçam türlerinin ayrımında DT modelinin test veri seti ile değerlendirilmesi. 91

Tablo 4.19: Sapsız meşe ve Macar meşesi türlerinin ayrımında DT modelinin test veri seti ile değerlendirilmesi.....92

Tablo 4.20: Çam ve meşe cins ayrımında eğitim veri seti ve ANN ile kurulan sınıflandırma modelinde elde edilen değerler. 92

Tablo 4.21: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve ANN ile kurulan sınıflandırma modelinde elde edilen değerler. 93

Tablo 4.22: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve ANN ile kurulan sınıflandırma modelinde elde edilen değerler. 93

Tablo 4.23: Çam ve meşe cins ayrımında ANN modelinin test veri seti ile değerlendirilmesi. 94

Tablo 4.24: Sarıçam ve karaçam türlerinin ayrımında ANN modelinin test veri seti ile değerlendirilmesi. 94

Tablo 4.25: Sapsız meşe ve Macar meşesi türlerinin ayrımında ANN modelinin test veri seti ile değerlendirilmesi.....94

Tablo 4.26: Çam ve meşe cins ayrımında eğitim veri seti ve SVM ile kurulan sınıflandırma modelinde elde edilen değerler. 95

Tablo 4.27: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve SVM ile kurulan sınıflandırma modelinde elde edilen değerler. 95

Tablo 4.28: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve SVM ile kurulan sınıflandırma modelinde elde edilen değerler. 96

Tablo 4.29: Çam ve meşe cins ayrımında SVM modelinin test veri seti ile değerlendirilmesi. 96

Tablo 4.30: Sarıçam ve karaçam türlerinin ayrımında SVM modelinin test veri seti ile değerlendirilmesi. 97

Tablo 4.31: Sapsız meşe ve Macar meşesi t türlerinin ayrımında SVM modelinin test veri seti ile değerlendirilmesi. 97

Tablo 5.1: Çam ve meşe türlerinin sınıflandırılmasında kurulan farklı model ve ön işlemlerin doğru tahmin oranları (eğitim seti). 104

Tablo 5.2: Çam türlerinin sınıflandırılmasında kurulan farklı model ve ön işlemlerin doğru tahmin oranları (eğitim seti). 104

Tablo 5.3: Meşe türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (eğitim seti). 105

Tablo 5.4: Çam ve meşe türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (test seti). 105

Tablo 5.5: Çam türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (test seti). 106

Tablo 5.6: Meşe türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (test seti). 106

Tablo 5.7: Çam ve meşe türlerinin sınıflandırılmasında kullanılan ön işlemlerin doğru tahmin oranlarını iyileştirme yüzdeleri. 107

Tablo 5.8: Çam türlerinin sınıflandırılmasında kullanılan ön işlemlerin doğru tahmin oranlarını iyileştirme yüzdeleri.....107

Tablo 5.9: Meşe türlerinin sınıflandırılmasında kullanılan ön işlemlerin doğru tahmin oranlarını iyileştirme yüzdeleri.....108

Tablo 5.10: PLS-DA ve DT aracılığıyla önemli kabul edildiği tespit edilen bantlar / dalga numaraları (cm^{-1}). 111

Tablo 5.11: Literatüre göre band atamaları (Haaland ve Thomas, 1988; Schwanninger ve ark., 2011; Xiaobo ve ark., 2010). 114

Tablo 5.12: Tür bazında ayrımda en yüksek ayırıcı güce sahip dalgaların konumları.....115

SİMGE VE KISALTMA LİSTESİ

Simgeler **Açıklama**

C : Karbon

E() : Beklenen değer

H : Hidrojen

N : Azot

O : Oksijen

R^2 : Determinasyon katsayısı

x/X : Bağımsız değişken/ler (gerçek değer)

y/Y : Bağımlı değişken/ler (gerçek değer)

$\hat{y}_i$: Bağımlı değişkenin tahmini

Kısaltmalar **Açıklama**

AIC : Akaike bilgi kriteri (Akaike Information Criterion)

ANN : Yapay sinir ağları (Artificial Neural Networks)

AUC : Eğrinin altında kalan alan (Area Under Curve)

Bias : Yanlılık

BIC : Bayes bilgi kriteri (Bayesian Information Criterion)

CA : Kümeleme analizi (Cluster Analysis)

CAL : Kalibrasyon (Calibration)

CCA : Kanonik korelasyon analizi (Canonical Correlation Analysis)

CITES :Nesli Tehlike Altında Olan Yabani Hayvan ve Bitki Türlerinin Uluslararası Ticaretine İlişkin Sözleşme (The Convention on International Trade In Endangered Species of Wild Fauna and Flora)

CR : Süreklilik giderme (Continuum removal)

CV : Çapraz sorgulama (Cross Validation)

DA : Diskriminant analizi (Discriminant Analysis)

DL-DA :Diyagonal doğrusal diskriminant analizi (Diagonal Linear Discriminant Analysis)

DN : Doğru negatif (True Negative-TN)

DP : Doğru pozitif (True Positive-TP)

Dr : Türev (Derivative)

DT : Karar ağaçları (Decision Trees)

EÇO : En çok olabilirlik (Maximum Likelihood)

EDC : Sentroidlere Öklid mesafesi (Euclidean distance to centroids)

FT : Fourier dönüşümü (Fourier Transform)

GBM : Gradyan artırma (Gradient Boosting)

HM : Hiyerarşik metotlar (Hierarchical Methods)

IR : Kızılötesi (Infrared)

ILS : Ters en küçük kareler (Inverse Least Squares)

K-means : K-ortalamalar

kNN : k-En yakın komşu (k- Nearest Neighbourhood)

LDA : Doğrusal diskriminant analizi (Linear Discriminant Analysis)

LS : En küçük kareler (Least Squares)

MC : Ortalama merkezleme (Mean Centering)

MLR : Çoklu doğrusal regresyon (Multiple Linear Regresssion)

MSC : Çarpımsal saçılım düzeltmesi (Multiplicative Scatter Correction)

MSE : Hata kareler ortalaması (Mean Squared Error)

NIR : Yakın kızılötesi (Near Infrared)

NIRS : Yakın kızılötesi spektroskopisi (Near Infrared Spectroscopy)

NN : En yakın komşu yöntemleri (Nearest Neighbourhood)

NW : Norris-Williams algoritması

OLS : Basit en küçük kareler regresyon (Ordinary Least Squares)

OSC : Ortogonal sinyal düzeltmesi (Orthogonal Signal Correction)

PCs : Temel bileşenler (Principal Components)

PCA : Temel bileşenler analizi (Principal Component Analysis)

PCR : Temel bileşenler regresyonu (Principal Component Regression)

PLS : Kısmi en küçük kareler (Partial Least Squares)

PLS-DA : Kısmi en küçük kareler diskriminat analizi (Partial Least Squares Discriminant Analysis)

PLS-R : Kısmi en küçük kareler regresyonu (Partial Least Squares Regression)

PRESS : Tahmini artık hata kareleri toplamı (Predicted Residual Error Sum of Squares)

R : Regresyon (Regression)

RF : Rasgele orman (Random Forest)

RMSE : Kök hata kareler ortalaması (Root Mean Squared Error)

ROC : Karar vericinin etkinliği (Receiver Operating Characteristic)

RSS : Artık kareler toplamı (Residual Sum of Squares)

QDA : Karesel diskriminant analizi (Quadratic Discriminant Analysis)

SEP : Standart tahmin hatası (Standard Error of Prediction)

SG : Savitzky-Golay algoritması

SIMCA : Sınıf karşılaştırmasının hafif bağımsız modellemesi (Soft Independent Modelling of Class Analogy)

Sm : Düzgünleştirme (Smoothing)

SNV : Standart normal değişim (Standard Normal Variate)

SR : Seçicilik oranı (Selectivity Ratio)

SVD : Tekil Değer Ayrışma (Singular Value Decomposition)

SVM : Destek vektör makineleri (Support Vector Machines)

TO : Tahmin oranı

TSS : Kareler toplamı (Total Sum of Squares)

YN : Yanlış Negatif (False Negative-FN)

YP : Yanlış Pozitif (False Positive-FP)

YL : Yıllık halka sayısı

ÖZET

DOKTORA TEZİ

YAKIN KIZILÖTESİ SPEKTROSKOPİSİNİN ODUN TEŞHİSİNDE KULLANIMI

Fatma Diğdem TUNCER

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Orman Endüstri Mühendisliği Anabilim Dalı

Danışman : Prof. Dr.Ayşe Dilek DOĞU

II. Danışman :Doç. Dr.Esra AKDENİZ

Bir odun örneğinin hangi ağaç türüne ait olduğunun tespitinde genel kabul gören yöntem mikroskobik teşhistir. Fakat bu yöntemin emek-yoğun olması, nispeten uzun hazırlık ve inceleme sürelerini gerektirmesinin yanı sıra, odun anatomisi uzmanlarına ihtiyaç duyulması sebebiyle son yıllarda daha kısa süreli etkin teşhis yöntemlerinin arayışlarına yönelik çalışmalara yoğunlaşıldığı görülmektedir. Bu çalışmalardan çıkan çözüm önerilerinden biri yakın kızılötesi spektroskopisinin (NIRS) kullanımıdır. Esas olarak ağaç malzemenin kimyasal yapısının analizine dayanan bu yöntemle elde edilen veriler, kemometrik yöntemler aracılığıyla değerlendirilmektedir. Bu çalışmada, NIRS yönteminin çok değişkenli veri analizleri yardımıyla ağaç türü teşhisinde kullanım etkinliğinin belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda dış morfolojik özellikleri aracılığıyla teşhis edilen yerli türlerimizden sarıçam, karaçam, sapsız meşe ve Macar meşesi türlerine ait yakın kızılötesi verileri, 12 000 – 4 000 cm^{-1} spektral aralığında, 4 cm^{-1} çözünürlükle elde edilmiştir. Cins ve tür bazında ayrımın elde edilmesi için doğrusal modellerden kısmi en küçük kareler diskriminat analizi (PLS-DA), diyagonal doğrusal diskriminant analizi (DL-DA), doğrusal olmayan modellerden karar ağaçları (DT), yapay sinir ağları (ANN) ve destek vektör makineleri (SVM) uygulanmıştır. Tüm modeller herhangi bir ön işlem uygulanmamış (ham) veri setlerinde ve standart normal değişim (SNV), çarpımsal saçılım düzeltmesi (MSC), Savizky-Golay algoritması (SG) ile 1.türev (1 Dr), SG 1 Dr + SNV, SG 1 Dr + MSC, düzgünleştirme (Sm), Sm + 1 Dr, Sm + 2 Dr aracılığıyla ön işlem uygulanmış veri setlerinde denenmiştir. Modellerin başarıları kalibrasyona katılmayan

test setleri ile denenmeleri sonucu elde edilen doğru tahmin oranları aracılığıyla belirlenmiştir. Ham veriler kullanılarak DL-DA, PLS-DA, DT, ANN, SVM modellerinin doğru tahmin oranları sırasıyla, çam ve meşe türlerinin cins bazında ayrımı için % 74, % 99, % 94, % 99,5, %100; çam türlerinin birbirinden ayrımında % 47, % 84, % 54, % 50, % 73; meşe türlerinin birbirinden ayrımında % 62, % 96, % 76, % 81, % 91 olarak bulunmuştur. Ön işlem metotlarının doğru tahmin oranlarını iyileştirici etkileri olmasına rağmen, tüm modellerde veya veri setlerinde ortak olarak en iyi denilebilecek bir ön işlem uygulaması tespit edilmemiştir. Denenen modeller arasında ağaç türü ayrımında doğrusal modellerden PLS-DA, doğrusal olmayan modellerden SVM en başarılı modeller olarak belirlenmiştir. Çalışmanın sonucunda NIRS yönteminin çok değişkenli veri analizleri yardımıyla ağaç türlerinin sınıflandırılmasında kullanımının oldukça yüksek başarı oranlarına ulaşılabildiği ortaya konulmuştur.

Nisan 2020, 165. sayfa.

Anahtar kelimeler: NIRS, çam, meşe, odun teşhisi, PLS-DA

SUMMARY

Ph.D. THESIS

UTILIZATION OF NEAR INFRARED SPECTROSCOPY IN WOOD IDENTIFICATION

Fatma Digdem TUNCER

Istanbul University-Cerrahpasa

Institute of Graduate Studies

Department of Forest Industry Engineering

Supervisor : Prof. Dr.Ayse Dilek DOGU

Co-Supervisor :Assoc. Prof. Dr.Esra AKDENİZ

Microscopic identification is generally regarded as the most accepted method to identify wood species. Microscopic evaluations, however, are labor-intensive, and need relatively long preparation and inspection periods, and also require expertise in wood anatomy. Therefore, studies on the search for more effective and faster diagnostic methods have emerged, in recent years. The near infrared spectroscopy (NIRS) is one of the proposed methods. Data obtained by this method based on the chemical analyses of wood were evaluated by chemometric methods. In this study, it was aimed to determine the effectiveness of NIRS method in wood identification with the aid of multivariate data analysis. For this purpose, near infrared data of our native species which identified through their external morphological characteristics, scots pine, black pine, sessile oak and Hungarian oak, were collected in the spectral range of 12 000 - 4 000 cm^{-1} with a resolution of 4 cm^{-1}. In order to achieve wood identification by genus and species, Partial least squares discriminate analysis (PLS-DA), diagonal linear discriminant analysis (DL-DA) from linear models and, decision trees (DT) artificial neural networks (ANN) and support vector machines (SVM) from nonlinear models have been used. All models have been developed on untreated (raw) and pre-treated datasets by standard normal variate (SNV), multiplicative scatter correction (MSC), first derivative(1Dr) using Savizky-Golay (SG), combination of SG 1 Dr + SNV, combination of SG 1 Dr + MSC, Smoothing (Sm) with SG,

combination of Sm + 1 Dr and combination of Sm + 2 Dr. The success of the models was determined by the accuracies of test sets that did not participate in calibration phase. Accuracy of DL-DA, PLS-DA, DT, ANN, SVM models with raw data were 74%, 99%, 94%, 99.5%, 100% for classification of pine and oak species on a genus basis, 47%, 84%, 54%, 50%, 73% for classification of pine species and; 62%, 96%, 76%, 81%, 91% for classification of oak species, respectively. Although pretreatment methods have effects to improve accuracy, no common best pretreatment method has been identified in all models or data sets. PLS-DA from linear models and, SVM from nonlinear models were determined as the most successful models in wood discrimination. As a result of the study, it was revealed that utilization of NIRS in wood classification with the aid of multivariate data analysis could achieve very high success rates.

April 2020, 165 pages.

Keywords: NIRS, pine, oak, wood identification, PLS-DA

1. GİRİŞ

Günümüzde ağaç malzeme sadece endüstriyel bir hammadde ya da ürün değil, aynı zamanda ülkelerin ihracat ve ithalatında önemli yere sahip ticari bir malzeme olarak karşımıza çıkmaktadır. Son yıllarda uluslararası düzeyde tartışılan konular arasında ağaç malzemenin yenilenebilir bir enerji kaynağı olarak kullanımının arttırılması, orman sanayi ürünlerinin üretim, tüketim ve ticareti ile ilgili olarak belirli normlarda istatistiksel bilginin geliştirilmesi, kaçak kesim sonucu üretilmiş ürünlerin ticaretinin önlenmesine yönelik iş birliği ve biyolojik çeşitliliğin korunması gibi konuların önemli bir yer tuttuğu görülmektedir. Ülkemizin de taraf olduğu CITES (Nesli Tehlike Altında Olan Yabani Hayvan ve Bitki Türlerinin Uluslararası Ticaretine İlişkin Sözleşme) Sözleşmesi ile nesilleri tehlike altında bulunan ya da bulunabilecek türlerin uluslararası ticaretinin izlenmesi ve gerekli şartlarda durdurulması, ekolojik dengenin uluslararası ticaret yoluyla istismar edilmesinin önlenmesi ve ülkelerin sahip oldukları biyolojik kaynakların sürdürülebilir kullanımlarının sağlanması amaçlanmaktadır (Doğu, 2013). Ticareti yapılan tomruk ve keresteler ait ağaç türlerinin belirlenmesi ile kaçak kesimlerin tespiti mümkün olmakta ve bu gibi istenmeyen durumların önüne geçilebilmektedir. Kaçak kesimler ve bu kesimlerden elde edilen ürünlerin ticareti hem üretici hem de tüketici ülkeler için ekonomik ve ekolojik açıdan pek çok probleme sebep olmaktadır. Kaçak kesimlerin dünya genelindeki ormansızlaşmanın, sera gazlarındaki artışın temel nedeni olduğu, sürdürülebilir orman ürünleri pazarları için de ciddi dezavantajlara sebep olduğu belirtilmektedir. Bu durumun önüne geçmek için hazırlanan CITES ise nesli tehlike altındaki türlerin ticaretinde resmî belgelerin hazırlanmasını zorunlu kılmaktadır. Suç teşkil eden belgelerde değişikliklerin takibi ve oluşturulan bu belgelerin doğruluğunun kontrolü ise klasik yöntemlerle oldukça uzun sürmektedir (Doğu, 2013; Dormontt ve ark., 2015).

Ağaç türü teşhisi; bitki taksonomisi, ağaç malzeme ticareti, arkeoloji, kültürel varlıkların korunması, sanat tarihi ve hatta kriminoloji gibi alanlarda oldukça önemli bir yer tutmaktadır. Özellikle günümüzde kıt kaynak olarak kabul edilen odun hammaddesinin rasyonel biçimde değerlendirilebilmesi ve kullanım yerinin doğru olarak seçilebilmesi için, ait olduğu ağaç türünün doğru teşhis edilmesi gerekmektedir (Doğu, 2013; Wheeler ve Baas, 1998). Ağaç malzemenin makroskobik ve mikroskobik yapısını incelemeye dayanan klasik odun teşhis yöntemleri günümüzde en güvenilir yöntemler olarak kabul edilmektedir. Fakat bu yöntemlerin

emek-yoğun olması, nispeten uzun hazırlık ve inceleme sürelerini gerektirmesinin yanı sıra, odun anatomisi uzmanlarına ihtiyaç duyulması sebebiyle son yıllarda farklı teşhis yöntemlerine yönelik bir ihtiyaç ortaya çıkmaktadır (Dormontt ve ark., 2015). Dünyada yeni teknolojilerin gelişmesi ve bilgiye hızlı ulaşıma olan beklentimiz hayatın her evresinde olduğu gibi ağaç malzeme ile çalışan endüstriler ve çevre dengesi için de bir ihtiyaç olarak görülmektedir. Bu ihtiyaçların karşılanmasında her türlü endüstriyel uygulama için temel oluşturacak olan tür teşhisinin hızlı ve etkili şekilde yapılmasına yönelik çalışmalar ortaya çıkmaktadır. Yüksek başarı oranları sağlanan ve geliştirilmesi için çalışmaların devam ettiği tür teşhisine yönelik yöntemler: çok değişkenli veri analizlerinin klasik teşhis yöntemlerine uygulanması; odun anatomisi uzmanları tarafından gerekli bilgilerin bilgisayar bazlı kameralı sistemlerine öğretilmesi yolu ile benzer özelliklere sahip ağaç türlerinin ayırt edilmesine olanak sağlayan machine vision; çok küçük örneklerde dahi çalışma imkanı veren fakat işlem süresi birkaç günü bulabilen DNA barkodlama; ekstraktif maddeleri baz alan kütle spektroskopisi olarak sayılabilir. Bu çalışmaların yanı sıra ön plana çıkan çözüm önerilerinden bir diğeri ise yakın kızılötesi spektroskopisinin (NIRS) kullanımıdır. Esas olarak ağaç malzemenin kimyasal yapısının analizine dayanan bu yöntem ile elde edilen verilerin istatistiksel yaklaşımlarla değerlendirilmesi denenmekte ve NIRS yönteminin ağaç türü teşhisinde kullanımı ile ilgili olarak olumlu sonuçlar alınmaktadır (Dormontt ve ark., 2015; Espinoza ve ark., 2014; Hermanson ve Wiedenhoeft, 2011; Jiao ve ark., 2014).

NIRS malzemenin kimyasal bileşenlerini belirlemede kullanılan, genellikle tahribatsız nicel ölçümlerle karakterize edilen analitik bir metottur. Ancak orman ürünleri endüstrisinin daha çok tercih ettiği bir yaklaşım olan, nitel analizlerin yapılması da bu metot ile mümkündür (Cao, 2013; Tsuchikawa ve ark., 2003). NIRS teknolojisi orman ürünleri endüstrisi alanında deneme aşamasında olup, standart metotlar içerisinde henüz yer almamaktadır. Ancak, gerçekleştirilen sınırlı sayıdaki araştırmaların sonuçları bu yöntem ile kısa süreli ve etkili sonuçların alınabildiğini göstermektedir (Brunner ve ark., 1996; Tsuchikawa ve ark., 2003; Flaete ve ark., 2006; Gierlinger ve ark., 2004; Haartveit ve Flæte, 2008; Nisgoski ve ark., 2017; Pastore ve ark., 2011; Russ ve ark., 2009; Schimleck, 2008; Tounis, 2009; Tsuchikawa, 2007; Tsuchikawa ve Kobori, 2015; Tsuchikawa ve Schwanninger, 2013)

Yukarıda bahsedilen bilgiler ışığında, ağaç türü teşhisinde karşılaşılan bazı sorunlar ve bu sorunların çözümlenebilmesi için farklı yönelimlerin mevcut olduğu görülmektedir. Bu çalışma

ile yukarıda sözü edilen tür teşhisine yönelik çözüm arayışlarına destek olunması planlanmıştır. Bu doğrultuda NIRS'ın ülkemizde yetişen ağaç türlerinin birbirinden ayrımında kullanılabilirliğinin değerlendirilmesi, bu çalışmanın temel amacını oluşturmaktadır.

Bu tez çalışması sonucunda NIRS yönteminin ülkemizde kullanımı ve başarı yüzdesi ortaya konulmuştur. Mikroskobik metotlarla ayrımı net olarak yapılamayan sarıçam (*Pinus sylvestris* L.) ve karaçam (*Pinus nigra* J.F. Arnold) türlerinin odunlarında tür bazında ayrım denenmiştir. Sarıçam, karaçam, Macar meşesi (*Quercus frainetto* Ten.) ve sapsız meşe (*Q. petraea* (Mattuschka) Liebl.) ağaç türleri için NIRS yönteminde kullanılmak üzere bilgi bankası oluşturularak, bu ağaç türlerinin kısa süreli ve etkin teşhisleri için NIRS yönteminin kullanılabilirliği araştırılmıştır. Modellemeler aracılığıyla, orman ürünleri endüstrisinde kullanılan pozitif bilimlerin anlamlandırılmasında kullanımı zorunlu olan istatistiksel yöntemlerin gelişimine katkı sağlanmıştır.

"Yakın Kızılötesi Spektroskopisinin Odun Teşhisinde Kullanımı" isimli bu Doktora tez çalışmasında Genel Kısımlar başlığı altında, kullanılan deneme materyali ile ilgili literatür bilgisi ve NIRS'ın tanıtımı ve bu konuda yapılan çalışmaların sonuçları verilmiştir. Ardından Malzeme ve Yöntem başlığı altında, deneme ağaçlarının alınması, ağaçlardan örneklerin elde edilmesi, kimliklendirme çalışmalarının içerikleri, kullanılan analizler ve bu analizlere ait algoritmaların açıklaması yapılmıştır. Bulgular kısmında araştırma sonunda elde edilen veriler ortaya konulmuş, Tartışma ve Sonuç kısmında ise elde edilen bulgular literatür ile karşılaştırılarak yorumlanmıştır.

2. GENEL KISIMLAR

2.1.AĞAÇ TÜRÜ TEŞHİSİ

Odun kendine has özelikleri ile doğal, çevre ile dost, yenilenebilir ve oldukça yüksek talep gören bir hammaddedir. Organik bir materyal olması sebebiyle elde edildiği ağaç türünün özelliklerini genel olarak taşısa da, aynı türden hatta aynı ağaçtan elde edilen hiçbir parça birbiri ile birebir aynı özellikte olmamaktadır (Zobel ve Van Buijtenen, 1989). Bu durum aynı ağaç türü için estetik çeşitliliği arttırırken, yapısal bazı özelliklerde de değişikliklere sebep olmaktadır. Aynı tür içerisinde değişiklik gösteren bu yapısal özellikler, farklı ağaç türlerinde daha geniş çeşitlilikle kendine yer bulmaktadır. Bu sebeple ağaç malzemenin konu olduğu her yerde ağaç türünün teşhis edilmesi öncelikli olarak önem taşımaktadır (Doğu, 2013).

Genel olarak endüstriyel amaçlı çalışmalarda ağaçların sadece hangi cinse ait olduğunun tespiti yeterli olurken, pek çok durumda cins bazında teşhisin yeterli olmadığı görülmektedir (Dormontt ve ark., 2015). Aynı cinse ait bazı türlerin nesli tehlike altında olması sebebiyle ticareti kısıtlanırken, benzer görünüşe sahip bazı türler için ticari bir kısıt bulunmamaktadır (*Dalbergia* spp., *Pterocarpus* spp. gibi) (Gasson, 2011; MacLachlan ve Gasson, 2010; Shou ve ark., 2014; Snel ve ark., 2018). Aynı zamanda kültürel ve tarihi yapılar hakkında bilgiye ihtiyaç duyulduğu pek çok durumda tür tespitinin oldukça önem taşıdığı belirtilmektedir (Horikawa ve ark., 2015; Hwang ve ark., 2016). Ağaç türü teşhisine ihtiyaç duyulan bu gibi durumlarda genel olarak mikroskobik yöntemler kullanılsa da bazı ağaç türlerinin mikroskobik olarak birbirinden ayırt edilebilmesinin güç, hatta bazen imkansız olduğu durumlarla karşılaşılmaktadır: sarıçam ve karaçam türlerinin ayrımı, Macar meşesi ve sapsız meşe türlerinin ayrımı gibi. Bu türlerin karşılaştırmalı odun anatomilerine bakıldığında;

Sert çamlar grubunda yer alan sarıçam ve karaçam türlerinin mikroskobik özellikleri karşılaştırıldığında her iki türünde yıllık halka sınırları belirgin, ilkbahar odunu tabakasından yaz odunu tabakasına geçiş hızlı, çok sayıda bulunan boyuna ve enine reçine kanallarını çevreleyen epitel hücreleri ince çeperli, öz ışınları tek hücre genişliğinde ve heterojen, enine traheid çeperleri dişli yapıda, karşılaşma yeri geçitleri 1-2 adet ve pencere tipindedir (Şekil 2.1 ve Şekil 2.2). Sarıçam ve karaçam türleri için sırasıyla ortalama yaz odunu katılım oranı % 16,80 ve % 18,44, traheid çapı 40,9 ve 43,5 µm, traheid uzunluğu 3100 ve 3400 µm, 1 mm^2'deki traheid sayısı 621 ve 616 adet, öz ışını yüksekliği 175,9 µm (6 hücre) ve 176,1 µm

(6 hücre), 1 mm'deki öz ışını sayısı 5 ve 5 adet, kenarlı geçit çapı 21,7 ve 22,7 µm, porus çapı 6,6 ve 7,6 µm olduğu görülmektedir. Bu iki türün yukarıda belirtildiği gibi mikroskobik özellikleri arasında anlamlı bir farklılık bulunmaması sebebiyle birbirinden ayrılamadıkları belirtilmektedir (Crivellaro ve Schweingruber, 2013; Köksal ve Kılıç Pekgözlü, 2016; Martin-Benito ve ark., 2013; Schoch ve ark. 2004).



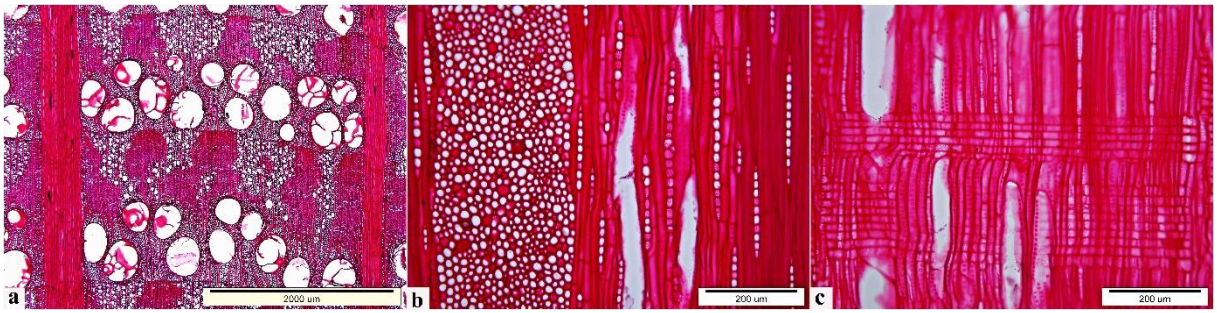
Şekil 2.1: Sarıçam türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200).



Şekil 2.2: Karaçam türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200).

Ak meşeler grubunda yer alan Macar meşesi ve sapsız meşe türlerinin mikroskobik özellikleri karşılaştırıldığında her iki türünde yıllık halka sınırları belirgin, halkalı traheli düzende, ilkbahar odunu traheleri yuvarlak ve oval şekilde, yaz odunu trahelerinin dizilişi radyal/diyagonal (alev) şekilde, perforasyon tablaları basit, trahelerin içi tüllerle doludur. Boyuna paranşimler çok sayıda, apotraheal teğet şeritli ve dağınık, paratraheal kümeli düzendedir. Öz ışınları homojen yapıda, dar ve geniş olmak üzere iki ayrı genişliktedir. Dar öz ışınları 1 hücre genişliğinde 25 hücre yüksekliğinde, geniş öz ışınları 20 hücreden geniş, birkaç cm yüksekliktedirler (Şekil 2.3 ve Şekil 2.4). Macar meşesi ve sapsız meşe türleri için sırasıyla ortalama ilkbahar odunu trahe teğet çapları 200 -400 µm ve 96-355 µm, yaz odunu trahe çapları 20-120 µm ve 15-88 µm, trahe uzunluğu 92-644 µm ve 244-779 µm, mm^2'deki trahe sayısı 43-100 adet ve 10-138 adet, lif uzunluğu 848-1920 ve 703-2009 µm'dir. (Berkel ve Bozkurt, 1961;

Bozkurt ve Erdin, 2000; Merev, 2003; Robert ve ark. 2017; Schweingruber ve Landolt, 2005; Voulgaridis,1990). Bu iki türün belirtilen mikroskobik özellikleri ile birbirinden ayrılmalarının güç olduğu anlaşılmaktadır.



Şekil 2.3: Macar meşesi türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200).



Şekil 2.4: Sapsız meşe türüne ait mikroskopik kesitler a) enine kesit (x40), b) teğet kesit (x200), c) radyal kesit (x200).

Tür teşhisinde günümüzde genel kabul gören ve yaygın olarak kullanılan yöntem ağaç malzemeden hazırlanan ince kesitlerin mikroskop yardımıyla bir uzman tarafından incelenmesini içeren mikroskobik teşhis yöntemidir. Fakat bu yöntem, bir odun anatomisi uzmanını gerektirmesinin yanı sıra nispeten uzun hazırlık ve inceleme aşamalarını da içeren emek yoğun bir yöntemdir. Bu sebeple geçmişten günümüze hızlı ve etkin teşhis yöntemlerine yönelik disiplinler arası çözüm arayışları devam etmektedir (Doğu, 2013; Wheeler ve Baas, 1998). Yapılan pek çok araştırma ve çalışma sonucunda mikroskobik metotlara alternatif olma potansiyeli bulunan yöntemler tespit edilmiş olup (DNA barcoding, kütle spektroskopisi, machine vision gibi), bu konularda çalışmalar sürdürülmektedir. Bu yöntemler arasında öne çıkan çalışmalardan biri de NIRS'ın çok değişkenli analiz metotları yardımı ile orman endüstrisinde kullanılmasıdır (Dormontt ve ark., 2015).

2.1.1. Ağaç Türü Teşhisinde Yakın Kızılötesi Spektroskopisinin Kullanımı

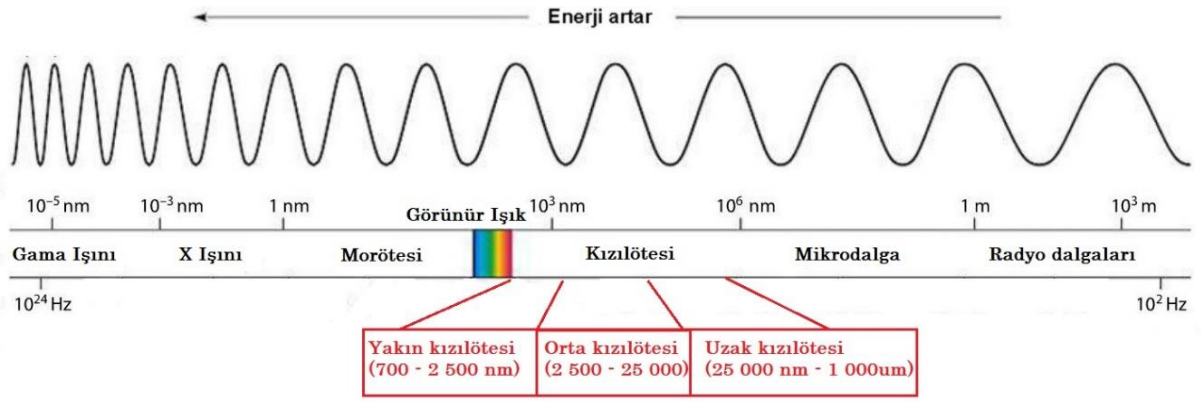
Organik materyallerin analizi için kullanılan NIRS günümüzde aktif olarak tıp, ecza, gıda ve zirai ilaçların kalite kontrolünde aktif olarak kullanılan bir metottur (Cao, 2013). Temel olarak materyali oluşturan moleküllerin gönderilen ışıma karşısındaki absorpsiyonları ve titreşimlerinden alınan sinyallerin işlenmesi mantığıyla çalışmaktadır. Yöntem genel olarak absorblanan ışınlar üzerine kurulmuş olsa da odun, plastik gibi opak malzemelerin yansıtıcı bir aparat ile incelenmesine de imkân sağlamaktadır (Burns ve Ciurczak, 2008). Bu metodun ağaç türlerine dair bilgi toplama amaçlı kullanımı yeni bir uygulama olmamakla beraber (Chow, 1972) gelişen teknoloji ve bilimsel yaklaşımlar ile ağaç türü teşhisinde başarılı ilk uygulamalarından biri Brunner ve ark. (1996) tarafından gerçekleştirilmiştir. Temel bileşenler ve kümeleme analizleri yardımıyla 6 farklı ağaç türü (*Ocotea rubra* Mez, *Trattinickia rhoifolia* Willd., *Liriodendron tulipifera* L., *Andira inermis* (Wright) DC., *Pouteria guianensis* Aubl., *Manilkara bidentata* (A.DC.) A.Chev.) ile gerçekleştirdikleri çalışmada olumlu sonuçlar elde etmelerine karşın, yetersiz örnek sayıları sebebiyle teşhiste % 100 başarıya ulaşamamışlardır. Tsuchikawa ve ark. (2003) ise spektrumların ikinci türevlerini alarak farklı rutubet değerlerinde 6 ağaç türünü karşılaştırmış (*Cryptomeria japonica* D. Don, *Pseudotsuga menziesii* Franco, *Picea sitchensis* Carr., *Tectona grandis* Linn., *Quercus mongolica* Fisch var., *Paulownia tomentosa* Steud., *Quercus gilva* Blume, *Fagus crenata* Blume) ve lif doygunluğu noktasında 1720, 2232 ve 2248 nm dalga boylarında % 100 doğru ayrımın yapılabileceğini ortaya koymuşlardır. Gierlinger ve ark. (2004) 3 farklı melez (*Larix* spp.) türünde yaptıkları çalışmada aynı cins içerisindeki ağaç türlerinin ayrımında NIRS yönteminin başarılı ve güvenilir olduğunu ortaya koymuşlardır. Flaete ve ark. (2006) benzer dış görünüşlere sahip 4 farklı ladin türünü karşılaştırdıkları çalışmalarında NIRS ile tahmin sağlayan modellerin oluşturulmasının ve benzer özelliklere sahip türlerin bu modellerle ayrımının mümkün olduğunu göstermişlerdir. Haartveit ve Flæte (2008) sarıçamda diri odun öz odun ayrımı ve benzer görünüşe sahip ladin türlerinde kısmi en küçük kareler fark analizleri (PLS-DA) kullanarak % 90'ın üzerinde başarılı sınıflandırma yapılabildiğini ortaya koymuşlardır. Russ ve ark. (2009) 7 farklı geniş yapraklı ağaç türüne ait yongaları temel bileşenler analizi (PCA) yardımıyla ayrımın mümkün olup olmadığını araştırdıkları çalışmalarında, yongaların rutubet değerinin ayrım için önemli bir kriter olduğunu ve NIRS 'ın yongalarda dahi başarılı sonuç verdiğini tespit etmişlerdir. Braga ve ark. (2011) CITES türlerinden *Swietenia macrophylla* King (hakiki maun) ve ona benzeyen türlerin arazide teşhis edilebilmesi amacıyla fiber optik bir NIRS sistemini denemişlerdir. PLS-

DA ve PCA kullanarak oluşturdukları modeller ile hakiki maun ve benzeri türlerin birbirinden ayrımının mümkün olduğunu belirtmişlerdir. Pastore ve ark. (2011)'de yaptıkları çalışmada koruma altındaki egzotik türlerin ayrımında anatomik sınıflandırma ile NIRS 'ı karşılaştırmış ve NIRS sisteminin verdiği tahmini değerlerdeki hata payının çok düşük olduğunu ve sınıflandırma için kullanımının umut vaat ettiğini belirtmişlerdir. Horikawa ve ark. (2015) az sayıdaki ksilaryum örnekleri üzerinde gerçekleştirdikleri çalışmada PLS-DA ve PCA kullanarak mikroskobik olarak tür bazında ayrımda bazı sorunlarla karşılaşıldığı *Pinus densiflora* Siebold& Zucc.ve *P. thunbergii* Parl. türlerinin öz odunlarından teşhisin mümkün olduğunu, fakat hava koşulları etkisinde kalmış (yaşlandırılmış) veya arkeolojik örnekler üzerinde bu tarz bir ayrımın, kimyasal yapıda oluşan değişimlerdeki çeşitlilik sebebiyle, oldukça güç olduğunu ortaya koymuşlardır.

2.2.YAKIN KIZILÖTESİ SPEKTROSKOPİSİ

Spektroskopi, elektromanyetik ışınım ve madde arasındaki ilişkiyi incelemektedir. Herhangi bir maddeye ışın gönderildiğinde madde, gönderilen ışının enerji seviyesi ve yapısında bulunan bağlara bağlı olarak; ışını absorbe edebilir, geçirebilir, yansıtabilir, dağıtabilir veya fotolüminesans (fotonlarla etkileşerek elektronların uyarılması) gerçekleşebilir (Burns ve Ciurczak, 2008; Workman ve Weyer, 2012). Spektroskopik veriler genel olarak maddenin gönderilen ışının dalga boyu veya frekansına bağlı olarak verdiği tepkiyi koordinat düzleminde göstermektedir. Maddeye gönderilen ışının dalga boyu veya frekansı ise elektromanyetik spektrumda bulunduğu yeri ve gönderilen ışının sahip olduğu enerjiyi göstermektedir (Şekil 2.5).

Işınlar sahip oldukları enerji ve dalga boyuna göre elektromanyetik spektrum içerisinde bir alanda yer bulmakta ve bulundukları alana göre de isimlendirilmektedir. Kızılötesi elektromanyetik spektrumda görünür ışık ile radyo dalgaları arasında kalan, 700 nm ile 1 mm arasındaki dalga boylarını içeren bölgedir. Kızılötesi bölgesi kendi içerisinde de üç alana bölünmüştür. Bu alanlar; görünür ışıktan hemen sonra gelen yakın kızıl ötesi bölgesi (700 - 2500 nm), kızılötesi olarak isimlendirilen pek çok cihazın çalışma aralığını içeren mid-IR olarak isimlendirilen orta kızılötesi bölgesi (2500 – 25000 nm) ve orta kızılötesi ile mikrodalga arasında kalan far-IR olarak isimlendirilen uzak kızılötesi (25000 nm – 1000 um) bölgesidir (Workman ve Weyer, 2012) (Şekil 2.5).



Şekil 2.5: Elektromanyetik spektrum (Davies, 2005; Baier, 2012).

NIRS, organik maddelerin moleküler bağlarının maruz kaldığı enerji karşısında gösterdiği titreşim enerjisinin ölçümüne dayanmaktadır. Yakın kızılötesi bölgesinde bulunan elektromanyetik ışınların organik materyallerin temelini oluşturan (Oksijen-O, Karbon-C, Hidrojen-H ve Azot-N) elementler arasında oluşan moleküler bağların (O-H, C-H, C=O vd.) titreşim hareketi için gereken enerji seviyesinde olmaları sebebiyle bu spektroskopi tercih edilmektedir. Analiz amacıyla titreşim enerjisinin kullanımı ile NIRS, moleküler bağlara veya ilgili organik malzemeye zarar vermeden problemlere çözüm arayan tahribatsız bir sistem olarak karşımıza çıkmaktadır (Burns ve Ciurczak, 2008; Workman ve Weyer, 2012).

NIRS odun kimyasında lignin, selüloz, hemiselüloz, ekstraktif maddeler ve selülozun kristallik derecesinin tahmininde (Jiang ve ark., 2007) kullanımının yanı sıra, odunun kimyasal yapısını temel alarak mikrofibril açısı (Schimleck ve ark., 2005; Via ve ark. 2007), traheid morfolojisi (Jones ve ark., 2005; Schimleck ve ark., 2004), mekanik özellikler (Hoffmeyer ve Pedersen, 1995; Kelley ve ark., 2004), yoğunluk, rutubet içeriği (Hans ve ark., 2015), daralma davranışları (Taylor ve ark., 2008), diri odun - öz odun oranı (Sandberg ve Sterley, 2009), basınç odununun tespiti (Chen ve ark., 2007), çürüklük tespiti ve derecesi (Gierlinger ve ark., 2003; So ve ark., 2004), modifikasyon yöntemlerinin etkisi (Bächle ve ark., 2012), kağıt hamurunun içeriği (Dahlman, 2012) gibi konularda da modellerin kurulması ve bu özellikler hakkında yüksek doğruluklarda tahmini bilgi edinilmesini mümkün kılmaktadır (Via ve ark., 2014). Ancak yakın kızılötesi bölgesinde moleküllerden elde edilen sinyaller oldukça sık üst üste binmekte ve birleşmekte, sonuç olarak oldukça geniş ve kompleks spektrumlar elde edilmektedir. Bu durum karmaşık spektrumların ortaya çıkmasına ve belirli bir özelliğin belirli bir bileşenle bağlantısını kurmanın zorlaşmasına sebep olmaktadır. Bu sebeple istenilen kimyasal bilginin spektrumdan

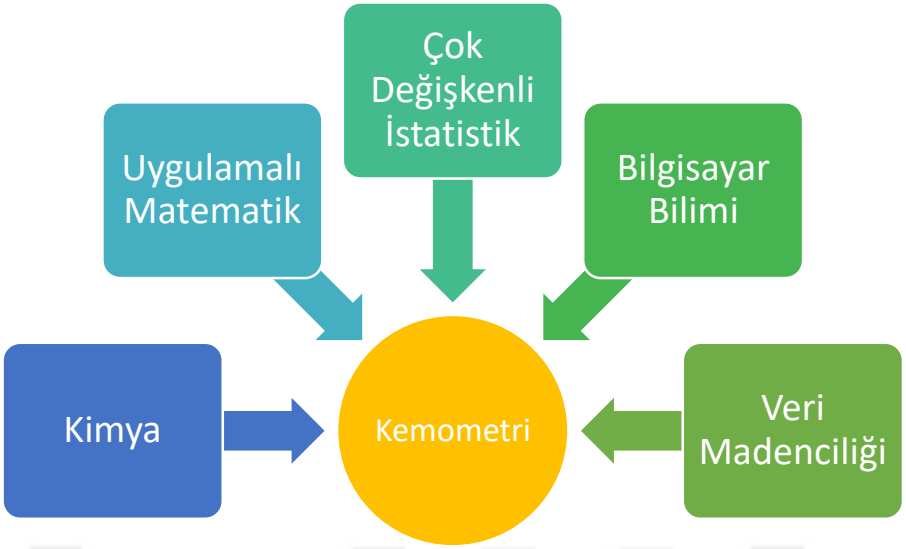
elde edilmesi için kemometrik yöntemlerin kullanılması zorunlu olmaktadır (Burns ve Ciurczak, 2008; Dinç, 2007; Wiesner ve ark., 2014; Workman ve Weyer, 2012).

2.3.KEMOMETRİ

Kemometri; kimyasal verilerden istatiksel ve matematiksel metotları kullanarak istenilen bilgilerin çıkarılması ile ilgilenen çok disiplinli (Şekil 2.6) bir bilim dalıdır (Varmuza ve Filzmoser, 2008). İlk olarak 1970 'lerde kullanılmaya başlayan bu ifade yıllar içerisinde analitik kimyanın önemli bir çalışma alanı olarak kabul görmüş, bu süreçte spektroskopi, kromotografi ve genetik gibi büyük ve karmaşık veri setlerinin işlenme ihtiyacı ile birlikte kemometrik uygulamalar oldukça önemli hale gelmiştir (Brereton, 2007).

Kemometri çözülmek istenen problemlerin tanımlanmasında kullanılabildiği gibi, istenilen özelliklerin tahmin edilmesinde de kullanılmaktadır. Kemometrinin tahmin edici özelliğinin orman endüstrisi için kullanılması ile ilgili ağaç türünün fiziksel özellikleri, mekanik özellikleri, rutubet derecesi, budak tespiti ve benzer özelliklerin yüksek başarı oranları ile tahmini mümkün olmaktadır. Kemometrinin sistemlerin altında yatan ilişkileri ve yapıyı ortaya çıkarmaya yarayan tanımlayıcı özelliği ise orman endüstrisinde teşhis amacıyla kullanımını mümkün kılmaktadır (Tsuchikawa ve ark., 2003).

Kemometri bu özelliklerinin yanı sıra; sinyal işleme, deneysel tasarım, modelleme, kalibrasyon, optimizasyon, örüntü tanıma, sınıflandırma, yapay zeka, görüntü işleme gibi kavram ve uygulamalara dair konuları da içermektedir (Dinç, 2007). Temel olarak kemometri, elde edilen çeşitli kimyasal veriler arasındaki ilişkinin kurulması ve kurulan bu ilişkinin doğruluğunun araştırılmasını içermektedir.

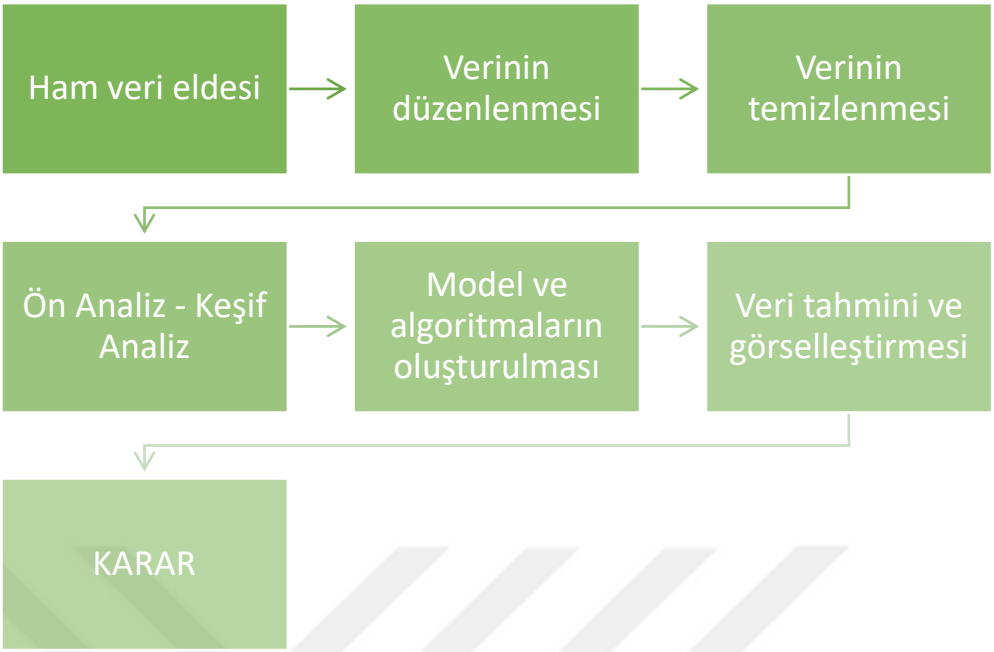


Şekil 2.6: Kemometri bilim dalının içeriği (Varmuza ve Filzmoser, 2008).

NIRS ile elde edilen veriler doğası gereği her örnek için çok sayıda veri noktasının (her dalga boyu için bir veri noktası) toplanmasını içermektedir. NIR modellerinin geliştirilmesi sırasında bağımsız değişkenler (X-matrisi: tüm dalga boylarında ölçülen absorbans/reflektans değerleri) ile bağımlı değişkenler (Y matrisi: örneklerin gelecekte tahmin edilecek değerleri) karşılıklı ilişki içerisindedir. Bunun yanı sıra üst üste binen piklerin varlığı da NIRS ile elde edilen veri setlerini karmaşıklaştırmaktadır. NIR spektrumlarındaki bu karmaşıklık ve farklı örnekler arası görülen spektral değişimler NIR spektroskopisinin analitik bir teknik olarak etkili bir şekilde kullanımı için çok değişkenli veri analizlerinin kullanımını zorunlu hale getirmektedir (Manley, 2014).

2.4.VERİ ANALİZİ

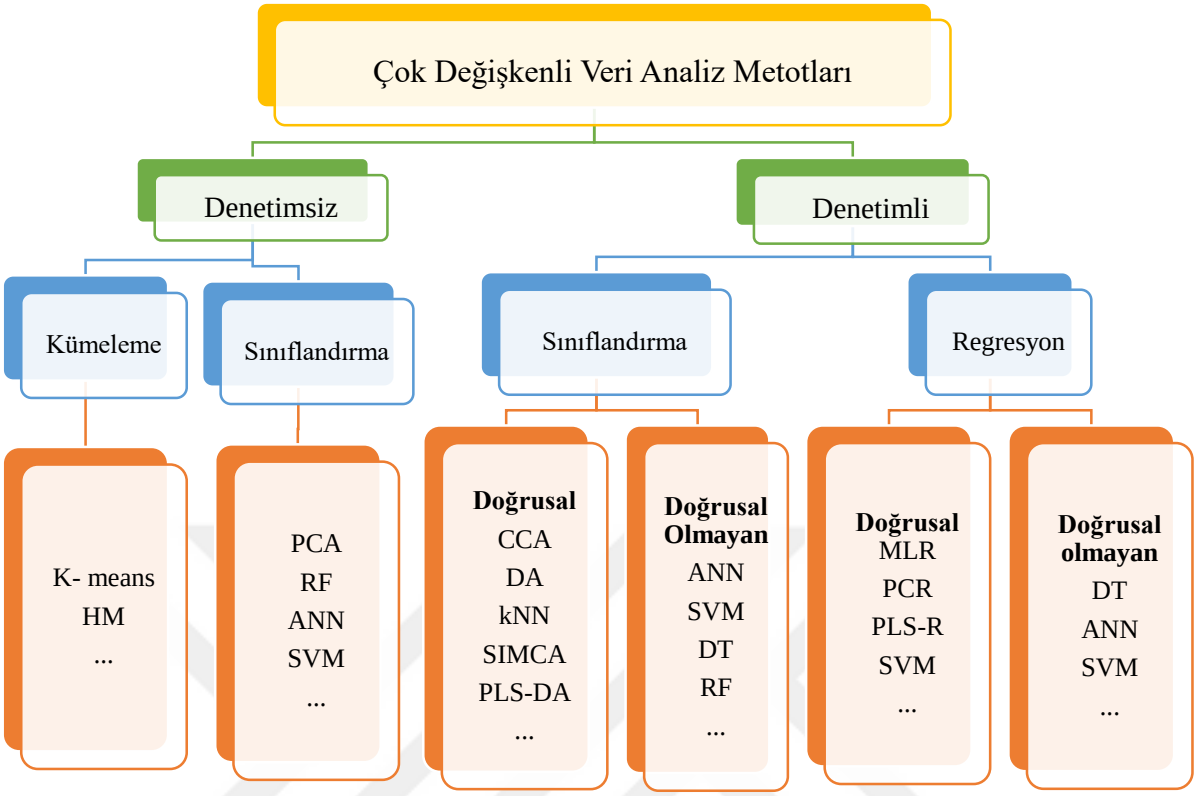
İstatistik, verilerin toplanması, düzenlenmesi, analiz edilmesi ve yorumlanması ile ilgilenen matematik ile ilişkili bir bilim dalıdır (Kendall ve Buckland, 1971). Veri analizlerini içeren uygulamalı istatistik, incelenen toplumun özetlenmesi ile ilgilenen tanımlayıcı (betimleyici) istatistik ve örneklemler aracılığıyla ana kütle hakkında karar verme amacıyla kullanılan çıkarımsal (tahminsel) istatistik olmak üzere iki ana yaklaşıma sahiptir. Çıkarımsal istatistiğin bir parçası olarak veri analizleri; ön işlemler (denetleme, hatalı verilerin ayıklanması, verilerin dönüştürülmesi) veriler arasında ilişkilerin açıklanması (korelasyon) ve modellemeyi içeren bir süreçtir (Şekil 2.7). Bu süreç karar verme mekanizmasını desteklemek için kullanışlı bilgilerin çıkarılmasını amaçlamaktadır (Brereton, 2007; Wehrens, 2011).



Şekil 2.7: Veri analizi süreci (Schutt ve O'Neil, 2014).

İncelenmek istenen birden çok toplum veya değişkenin olduğu durumlarda yani karmaşık veri setlerinin analizinde ise çok değişkenli analiz metotları kullanılmaktadır. Çok değişkenli veri analizlerinde gözlenen verideki tüm değişkenlerin etkileri göz önüne alınarak veri analizleri gerçekleştirilmektedir (Naes ve ark., 2004). Bu analiz metotları genel olarak denetimli ve denetimsiz olmak üzere iki ana gruba ayrılmakta, daha sonra kullanılan algoritmalarına göre sınıflandırılmaktadır. Bu sınıflandırma içerisinde algoritmalar kullanım şekillerine göre, model parametrelerinde ilgili değişikliklerin yapılmasına bağlı olarak denetimli ve/veya denetimsiz analiz gruplarında bulunabilmektedir (Şekil.8) (Brereton, 2007; Olivieri, 2018; Varmuza ve Filzmoser, 2008).

Kullanılacak çok değişkenli analiz metotları verilerden elde edilmek istenen çıktılara (uygulama amacına) göre seçilmekte, bazı durumlardan birden fazla algoritma kullanılarak daha yüksek doğrulukta sonuçlar elde edilmeye çalışılmaktadır. Veri setine uygulanan algoritmalar ve elde edilen sonuçlara göre de modellemeler gerçekleştirilmektedir. Güvenilir bir modelin kurulması için ise öncelikle ham verinin düzenlenmesi, işlenmesi ve veri setinin ihtiyaç duyduğu en sağlam ön işlem metodunun seçilmesi gerekmektedir.



Şekil 2.8: Çok değişkenli veri analiz metotları (Brereton, 2007; Elmasry ve ark., 2012; Patel ve Patel, 2014).*

*ANN: Yapay sinir ağları, CCA: Kanonik korelasyon analizi, DA: Diskriminant analizi, DT: Karar ağaçları, HM: Hiyerarşik metotlar, K-means: K-ortalamalar, kNN: k- En Yakın Komşu, MLR: Çoklu doğrusal regresyon, PCA: Temel bileşenler analizi, PCR: Temel bileşenler regresyonu, PLS-DA: Kısmi en küçük kareler diskriminant analizi, PLS-R: kısmi en küçük kareler regresyonu, RF: Rastgele orman, SIMCA: Sınıf karşılaştırmasının hafif bağımsız modellemesi, SVM: Destek vektör makineleri.

2.4.1. Ön İşlem Metotları

Güvenilir bir modelin kurulması için en sağlam ön işlem metodunun seçimi oldukça önemlidir. Spektral ön işlemler veri seti içerisindeki gürültüyü (noise) azaltma ve ilgilenilen spektral özelliklerin iyileştirilmesi amacıyla kullanılmaktadır. Bir diğer ifadeyle ön işlem metotları, modelleme teknikleri tarafından doğru şekilde ele alınamayan aykırı verilerin etkilerinin azaltılması ve mümkünse kaldırılması amacıyla geliştirilmiştir (Gholizadeh ve ark., 2015; Olivieri, 2018).

Aykırı değerler genel olarak, normal değerlerden önemli ölçüde sapan gözlemler gibi, verilerin çoğunluğu ile tutarlı olmayan gözlemler olarak tanımlanmaktadır. Tüm örnekler için model kurulmadan önce aykırı değerlerin belirlenip uzaklaştırılmasının tahmin doğruluğunu arttırıcı yönde etki yaptığı belirtilmektedir (Gholizadeh ve ark., 2015).

Katı örneklerin yakın kızılötesinde yansıtma verileri alınırken ışığın dağılması ve ışığın izlediği yolda oluşan değişimler sebebiyle sorunlarla karşılaşılabilmektedir. Bu istenmeyen değişimler kaydedilen spektrumlarda temelde kaymalar veya sapmalar şeklinde görülebilmektedir. Gözlenen bu fiziksel fenomenleri (temelde kayma, sapma) sinyallerin filtrelenmesi yoluyla spektrumdan uzaklaştırmak ve böylelikle örneklerin kimyasal kompozisyonları ile ilgili modellerin oluşturulması amacıyla, modelleme ve analizler öncesi ön işlemler uygulanmaktadır (Olivieri, 2018; Rinnan ve ark., 2009). Ön işlem aşaması ile asıl hedeflenen; sonraki aşamada oluşturulan, açıklayıcı/keşfedici analizi veya kalibrasyon modeli veya bir sınıflandırma modelini geliştirmek ve iyileştirmektir (Rinnan ve ark., 2009).

Ön işlem metotlarının özellikle NIR spektroskopisinde çok değişkenli kalibrasyonlar öncesi kullanımı oldukça kabul gören bir yaklaşım olmakla birlikte iki temel probleme de sahiptir. Bu problemlerden ilki standart bir prosedürün olmaması sebebiyle probleme yönelik denemeler sonucu uygulanacak ön işlemlerin seçilmesi iken; diğer problem bazı ön işlemlerin cihazdan kaynaklanan gürültüleri de özellik olarak (gürültünün korelasyonu) görebilmesidir. Ön işlemler bu tarz riskleri taşımasının yanı sıra, olumlu bir özellik olarak, ölçülen sinyalin zaman içerisinde değişmesi veya farklı cihazlardan elde edilen ölçümler arasında oluşan farklılığı tolare edebilmeyi de sağlamaktadır. Bu özelliği sebebiyle kurulan kalibrasyon modellerinin başarısını ve kullanılabilirliğini geliştiren bir etki göstermektedir (Olivieri, 2018).

Ön işlem metotlarının sınıflandırılmasında farklı yaklaşımlar söz konusu olsa da, genel olarak en sık kullanılan metotları örnekleri ile birlikte aşağıdaki gibi gruplandırmak mümkündür (Şekil 2.9);

Ön İşlem Metotları	*Düzgünleştirme/ Spektral türev alma metotları*	Savitzky-Golay (SG) Algoritması	
		Norris-Williams (NW) Algoritması	
	Saçılım Düzeltme metotları	Temel Düzeltme	MSC, ...
		Ölçeklendirme	SNV, ...
		Normalleştirme	CR, ...
		Ortogonal Sinyal Düzeltmesi (OSC)	OSC, ...

Şekil 2.9: Ön işlem amacıyla kullanılan bazı metotların sınıflandırılması (Gholizadeh ve ark., 2015; Rinnan ve ark., 2009).

Düzgünleştirme/ Yumuşatma (Smoothing): Hareketli pencere stratejisini kullanan ön işlem metotlarından biridir. Gürültünün (rasgele ölçüm hatasının) azaltılması amacıyla

kullanılmaktadır. En popüler yaklaşımlardan biri Savitzky-Golay algoritmasının (SG) kullanımıdır. Spektrumun başında belirli bir pencere genişliği (küçük bir spektral alanın) seçilerek bu pencere içerisinde ön işlem gerçekleştirilmekte ve ardından aynı işlem spektrum boyunca her pencerede tekrar edilmektedir.

Sadece düzgünleştirme uygulanacaksa türev derecesi 0 alınarak, belirli bir genişlikteki (tek sayı) hareket eden pencere içerisinde sinyal verileri bir polinomiyal fonksiyona uydurulmakta ve her durum için ayarlanabilir parametreler ortalamalar aracılığıyla tahmin edilmektedir (Olivieri, 2018; Savitzky ve Golay, 1964). Savitzky-Golay algoritması düzgünleştirmenin yanı sıra türev alma özelliği kullanıldığında temel düzeltmesi amacıyla kullanılabilmektedir Rinnan ve ark., 2009).

Temel düzeltmesi (Baseline removal): Bazı spektroskopilerde temelin veya arka plan sinyalinin sıfırdan başlamadığı durumlar söz konusu olabilmektedir. Bu etki piklerin yüksekliğini ve alanını etkilediğinden temel düzeltmesine ihtiyaç duyulmaktadır. Temel düzeltmesi amacıyla uygulanan spektral türev alma dönüşümü oldukça etkili sonuçlar vermektedir. İlk türev, arka plan sinyalinin yaklaşık olarak doğrusal olması şartıyla, örnekler arasında değişen taban hizasının etkisini ortadan kaldırır. İkinci türev ise doğrusal olmayan bir karesel (ikinci derece) fonksiyona benzetilmesiyle, örnekten örneğe değişen doğrusal olmayan arka plan sinyallerini çıkarabilmektedir (Gholizadeh ve ark., 2015; Olivieri, 2018; Rinnan ve ark., 2009).

Spektroskopide saçılım etkisini uzaklaştırmak için kullanılabilecek bir diğer yöntem verideki çarpımsal ve/ veya toplamsal saçılımların etkisini düzeltmek amacıyla uygulanan çarpımsal saçılım düzeltmesi (MSC) ‘dir. MSC, Savitzky-Golay düzgünleştirme veya türev alma filtresinden farklı olarak hareket eden pencere stratejisini kullanmamakta, bu sayede gürültünün korelasyonu etkisine sahip olmamaktadır. Bu metot, ışık saçılması nedeniyle oluşan arka plan sinyallerinin, örnekler arasında değişebilmesine rağmen, birbirlerine göre orantılı veya yaklaşık orantılı olabileceği veya sabit bir değerle daha düşük veya daha yüksek sinyal yoğunluğuna kaydırılabileceği fikrine dayanmaktadır. Bu yaklaşımla, MSC öncelikle spektrumları ortalama kalibrasyon spektrumuna doğrusal olarak en küçük kareler aracılığıyla uydurmakta ve ardından tüm spektrumlardan doğrusal uydurma ile tahmin edilen sinyali çıkarmaktadır (Gholizadeh ve ark., 2015; Olivieri, 2018).

Ölçeklendirme (Scaling): Bir diğer ön işlem metodu ölçeklendirme (range scaling)'dir. Bu metot spektrumdaki toplam yoğunluğun örneğe bağlı olduğu ve yoğunluğun karşılaştırılabilmesi için örneklerin ölçeklendirilmesine ihtiyaç duyulduğu durumlarda kullanılabilmektedir. Bu yöntemler arasında özellikle NIR uygulamalarında tercih edilen standart normal değişim (SNV - standard normal variate); ışık saçılımı ve partikül boyutlarının çarpımsal etkileşiminin düzeltilmesi, her bir spektrumun hizalama ve ölçeklendirmesi aracılığıyla gerçekleştirilmektedir. Her spektrum için ortalama sinyal ve standart sapma tahmin edilmekte; ardından her spektrumdan ortalama sinyal çıkarılmakta ve standart sapmaya bölünmektedir. Sonuçta ölçeklendirmeden sonra her spektrumun ortalaması sıfır ve standart sapması 1 olmaktadır (Gholizadeh ve ark., 2015; Olivieri, 2018).

Normalleştirme (Normalization): Normalleştirme absorpsiyon özelliğinin iyileştirilmesi ve teşhis amacıyla kullanılmaktadır. Bir normalleştirme metodu olarak süreklilik giderme (CR - continuum removal) sürekliliğin zarf eğrisini ham yansıtma spektrumları üzerine bölerek yeni spektral veriler üretmektedir. Bu yöntem, belirli absorpsiyon özelliklerinin izole edilmesinde ve değişen eğimlerin ve tüm yansıtma seviyelerinin etkilerinin giderilmesinde etkili olmaktadır (Gholizadeh ve ark., 2015).

Başarılı bir modelleme aşaması için uygun ön işlemlerin seçimi oldukça önemlidir. Daha önce de bahsedildiği gibi ön işlem metotlarının seçimindeki en önemli sorun, standart bir prosedürün bulunmayışıdır. Bu sebeple genellikle ön işlem metodu seçilirken, klasik bir deneme yanılma araştırması gerçekleştirmektedir. Bu araştırma sırasında farklı prosedürler tek başına veya başka prosedürlerle kombine edilerek iyi bir, çok değişkenli model bulunana kadar (kabul edilebilir ortalama çapraz sorgulama hatası veya latent değişkenleri sayısına sahip) denenmektedir (Olivieri, 2018). Bunun yanı sıra, çok sayıda ön işlemin bir arada uygulanması anlamlı verilerin de kaybolmasına sebep olabileceği için tavsiye edilmemektedir. Seçilen ön işlemin en temel gerekliliği etkili model karmaşıklığını koruması veya azaltmasıdır (Rinnan ve ark., 2009).

Yapılan literatür incelemesinde yukarıda bahsedilen ön işlem metotlarının bazı çalışmalarda aktif bir şekilde kullanılırken bazı çalışmalarda ise işlem görmemiş (ham) veri ile modellerin kurulduğu görülmektedir. NIRS ağaç türü teşhisinde kullanımına dair gerçekleştirilen çalışmalarda kullanılan ön işlem metotları Tablo 2.1' de özetlenmiştir.

Tablo 2.1: NIRS ile ağaç türü teşhisi konusunda gerçekleştirilen çalışmalarda geçmişten günümüze kullanılan ön işlem metotları.

Yayın Yılı	Yazarlar	Ön İşlem Metotları									
		Ham	1. Dr	2. Dr	MSC	SNV	SG	NW	MC	Sm	OSC
1996	Brunner ve ark.										
2003	Tsuchikawa ve ark.			X							
2004	Gierlinger ve ark.			X			X				
2006	Flæte ve ark.										
2008	Haartveit ve Flæte		X						X		
2008	Adedipe ve ark.	X	X			X					
2009	Russ ve ark.		X								
2009	Tounis, E.	X	X	X	X		X				
2011	Braga ve ark.		X	X	X	X					
2011	Pastore ve ark.		X	X	X	X					
2011	Russ ve Fiserova										
2012	Dawson-Andoh ve Adedipe	X	X								
2012	Espinoza ve ark.			X	X						
2014	Shou ve ark.		X				X			X	X
2014	Zhang ve ark.			X				X			
2015	Horikawa ve ark.			X							
2016	Bergo ve ark.		X				X		X	X	
2016	B. de Muñiz ve ark.			X	X		X				
2016	Hwang ve ark.										
2016	Nisgoski ve ark.			X	X	X	X				
2017	Lang ve ark.		X								
2017	Lazarescu ve ark.			X							
2017 a	Nisgoski ve ark.	X			X		X				
2017 b	Nisgoski ve ark.	X		X			X				
2018	Nisgoski ve ark.	X		X	X		X				
2018	Snel ve ark.		X			X	X		X		
2019	Leandro ve ark.	X	X			X	X			X	
2019	Pace ve ark.	X	X	X		X	X				
2019	Li ve ark.	X									
2019	Vieira ve ark.	X		X			X				
2019	Yong ve ark.		X		X	X				X	
2020	Zhou ve ark.		X	X		X				X	

*Dr: Türev, MC: Ortalama merkezleme (mean-centering), MSC: Çarpımsal Saçılım Düzeltmesi, NW: Norris-Williams algoritması, OSC: Ortogonal Sinyal Düzeltmesi, SG: Savitzky-Golay algoritması, Sm: Düzgünleştirme, SNV: Standart Normal Değişim

2.4.2. Kalibrasyon Metotları

Kemometrinin temel uygulamalarından biri olan kalibrasyon, genel olarak bir tip gözlem veya ölçümün (y) tahmini için bir veya birden çok değişkenin (X) değerlerinin veya değerlerin altında yatan parametrelerin kullanılması ile modellenmesini içeren bir süreçtir (Brereton, 2007; Varmuza ve Filzmoser, 2008). Kalibrasyon yöntemleri tek değişkenli ve çok değişkenli olmak üzere iki ana başlıkta incelenmektedir. Her bir örnek için bir özelliğin ölçüldüğü veri setlerinde tek değişkenli kalibrasyon yöntemleri kullanılırken, her bir örnek için pek çok verinin ölçüldüğü (spektrumlar gibi) veri setlerinde ise çok değişkenli kalibrasyon yöntemleri kullanılmaktadır (Olivieri, 2018).

Bağımsız değişkenler aracılığıyla (X) bağımlı değişkenlerin (y/Y) tahminini yapan modellerin oluşturulması veya mevcut modellerin uygulanması ile X ve y arasındaki ilişki ortaya konulmaktadır. Bu ilişki lineer (doğrusal) veya non-lineer (doğrusal olmayan) olabilmekte, kurulan modeller nispeten basit denklemler ile formüle edilebilmekte veya daha az belirgin bir algoritmik yapı olarak uygulanabilmektedir (Varmuza ve Filzmoser, 2008).

Çok değişkenli analiz yöntemleri, ilgilendikleri özelliklere bağlı olarak kantitatif (regresyon) ve kalitatif (sınıflandırma) olmak üzere temelde ikiye ayrılsa da; keşfedici analizleri de içerisinde barındırmaktadır (Rinnan ve ark., 2009). Çok değişkenli analizler aracılığıyla modellerin kalibrasyonu temelde veriler arasındaki ilişkilerin ortaya koyulmasını içermektedir. Bu ilişkilerin açıklanmasında kullanılacak olan metotlar incelenecek verilerin yapısına bağlı olarak tercih edilmektedir (Brereton, 2007; Pasquini, 2018).

2.4.2.1.Kümeleme Analizi

Kümeleme analizi veri içerisindeki benzerlikleri tespit etmek amacıyla kullanılan denetimsiz bir veri analiz metodudur (Brereton, 2007). Yoğunlaşmış gruplar olarak nitelendirilen kümeleme analizinde amaç, grup üyeleri hakkında bilgi sahibi olmadan değişkenlerin yoğunlaştığı alanların gruplar olarak tespit edilmesidir. Kümeleme analizi, analiz gerçekleştirilirken ayrılacak gruplara dair bilgilerin işlenmemesi sebebiyle denetimsiz metotlar içerisinde yer almaktadır (Varmuza ve Filzmoser, 2008). Kümeleme analizi, yoğunlaşmış grupların tespitinde grup yapısının nesneler arasında doğal olarak bulunduğu ve genel olarak bir nesnenin birden fazla grupta bulunabileceği varsayımı üzerine kurulmaktadır. Bu varsayımdan kaynaklanan sorunların çözülmesi için farklı ilkeler ile çalışan çeşitli kümeleme

algoritmaları önerilmektedir. En çok kullanılan kümeleme analizi metotlarını; bölmeli kümeleme, hiyerarşik kümeleme, bulanık kümeleme ve kümelemelerin karşılaştırılması olarak sınıflandırmak mümkündür. Bölmeli kümelemede, her obje bir gruba aittir. Hiyerarşik kümelemede, farklı gruplardaki objeler arasındaki ilişkiler aşamalı olarak dendrogram formunda gösterilmektedir. Bulanık Kümelemede ise, her nesne belirlenen kümelere normalleştirilmiş [0,1] üyelik katsayıları ile atanmakta, böylece bir nesnenin kümelerin herhangi birine atanması olasılığı yorumlanmaktadır (Brereton, 2007; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

Kümeleme yapılırken öncelikle nesneler arasındaki mesafe ve benzerlik ölçülmekte, ardından nesneler arasındaki ilişki ortaya koyulmaktadır. Benzerliklerin ölçülmesinde örnekler arası korelasyon katsayısı, Öklid uzaklığı, Manhattan uzaklığı veya Mahalanobis uzaklığının ölçülmesi gibi yaklaşımlar kullanılabilmektedir. k ve l nesneleri (örnekleri) arasındaki mesafe, j değişkenine göre aşağıdaki şekillerde ifade edilmektedir (Brereton, 2007);

Öklid uzaklığı;

$$d_{kl} = \sqrt{\sum_{j=1}^{J} \left(x_{kj} - x_{lj}\right)^2} \tag{2.1}$$

Manhattan uzaklığı;

$$d_{kl} = \sum_{j=1}^{J} \left|x_{kj-} x_{lj}\right| \tag{2.2}$$

Mahalanobis uzaklığı;

$$d_{kl} = (x_k - \bar{x}_l)^T \sum\nolimits_{i}^{-1} (x_k - \bar{x}_l) \tag{2.3}$$

Nesneler arasındaki uzaklığın ölçülmesi yerine değişkenler arasındaki uzaklığın ölçülmesi tercih edildiğinde kullanılabilecek klasik korelasyon katsayısı, Pearson korelasyon katsayısı, şu şekilde ifade edilmektedir;

$$r_{xy} = \frac{cov_{xy}}{s_x \,.\, s_y} = \frac{\sum_{i=1}^{I}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I}(x_i - \bar{x})^2 \sum_{i=1}^{I}(y_i - \bar{y})^2}} \quad (2.4)$$

Nesneler veya değişkenler arası mesafelerin ölçülmesinin ardından benzerlik veya farklılık matrisi çizilmektedir. Bu aşamadan sonra matristeki objeler arası ilişkiler kurulmaktadır. Bu ilişkinin kurulmasında en yaygın yaklaşım olan yığınsal kümelemede gruplardaki her bir nesne aşamalı olarak birbirlerine bağlanmakta, nesneler arası bağlantıların kurulması ile kümeleme analizi gerçekleştirilmiş olmaktadır (Brereton, 2007). En sık kullanılan kümeleme analizi metotları ve bu metotların kullandıkları algoritmalardan bazıları Tablo 2.2' de verilmiştir (Brereton, 2007; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

Tablo 2.2: Kümeleme analizi metotları ve en sık kullanılan algoritmalar

Kümeleme Metodu	**En sık kullanılan metot**	**Denklemi**	**Denklem No.**
Bölmeli kümeleme	k-ortalamalar	$c_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i^{(j)}$	(2.5)
		$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left\|x_i^{(j)} - c_j\right\|^2 \rightarrow min$	(2.6)
Hiyerarşik kümeleme	Tek bağlı kümeleme	$min_i \left\|x_i^{(j)} - x_i^{(l)}\right\|$	(2.7)
	Ortalama bağlı kümeleme	$ortalama_i \left\|x_i^{(j)} - x_i^{(l)}\right\|$	(2.8)
	Tam bağlamalı kümeleme	$max_i \left\|x_i^{(j)} - x_i^{(l)}\right\|$	(2.9)
	Merkezi yöntem	$\left\|c_i - c_j\right\|$	(2.10)
	Ward Yöntemi	$\left\|c_i - c_j\right\| . \frac{\sqrt{2\, n_j n_l}}{\sqrt{n_j + n_l}}$	(2.11)
Bulanık Kümeleme	Bulanık C-Ortalamalar algoritması	$c_j = \frac{\sum_i u_{ij}^2 \,.x_i}{\sum_i u_{ij}^2}$	(2.12)
		$\sum_{j=1}^{k}\sum_{i=1}^{n_j} u_{ij}^2 \,.\left\|x_i^{(j)} - c_j\right\|^2 \rightarrow min$	(2.13)

2.4.2.2.Temel Bileşenler Analizi

Temel bileşenler analizi (PCA) veri içerisindeki ilişkileri belirlemek amacıyla kemometride en çok kullanılan tekniklerden biridir (Brereton, 2007). Orijinal değişkenlerin doğrusal kombinasyonlarını içeren yeni değişkenlerin oluşturulması ile boyut indirgemeyi amaçlayan bir metottur. Oluşturulan yeni değişkenlerin her biri bir önceki değişkene dikey düzlemdedir, yani ortogonaldir. Gizli değişkenler (latent variables) veya temel bileşenler (PCs) olarak isimlendirilen bu yeni değişkenlerin ilki en çok varyasyonu (maksimum varyans) açıklamaktadır (Varmuza ve Filzmoser, 2008; Wehrens, 2011). PCA yüksek boyutlu verinin orijinal verideki bilgilerin çoğunu içeren düşük boyutlu bir uzaya çevrilmesini sağlamaktadır. Yeni oluşturulan bu uzaydaki örneklerin koordinatları, skorlar (scores) olarak isimlendirilmekte ve T sembolü ile gösterilmektedir. Her bir bileşenin boyutunun ölçülmesi ile elde edilen değerler ise öz değer (eigenvalue) olarak isimlendirilmekte, S sembolü ile gösterilmekte ve genellikle skorların kareler toplamına eşit olmaktadır. Gerçek değişkenler ile gizli değişkenler arasındaki ilişki, yüklemeler (loadings) olarak isimlendirilen, P sembolü ile gösterilen bir grup vektör ile [öz vektörler (eigenvector)] açıklanmaktadır (Brereton, 2007; Olivieri, 2018; Varmuza ve Filzmoser, 2008; Wehrens, 2011). Her yüklemenin bir skor karşılığı olmakta, bu skor ve yükleme çiftine de temel bileşen denilmektedir. Temel bileşenlerin orijinal veri ile ilişkini ortaya koyan formül şu şekildedir:

$$X = P \,.\, T^T \tag{2.14}$$

Bu denklemde; X : Kalibrasyon matrisi, P : Yüklemeleri (öz vektör matrisi),T^T : Skor matrisini göstermektedir.

Hesaplanan yüklemeler sayesinde değişkenlerin indirgenmesi veya indirgenmiş değişkenlerden gerçek değişkenlere dönüştürülmesi mümkün olmaktadır. Bu sayede temel bileşenler analizinin amacı sınıflandırma olmasa da pratikte örneklerin ayrımı içinde kullanımı mümkün olmaktadır. Birbirlerine ortogonal olan temel bileşenlerin koordinat düzleminde gösterimi sonucu bazı durumlarda örneklerin gruplandıkları görülmektedir. Örneklerin grup bilgileri sisteme verilmeden yapılan bu ayrım, temel bileşenler analizinin denetimsiz veri analizi grubu içerisinde kullanılmasını sağlamaktadır (Olivieri, 2018).

Temel bileşenlerin belirlenmesinden sonra oluşturulan yeni gizli değişkenler (temel bileşenler) ile farklı modellerin kurulması da mümkündür. Orman endüstrisinde temel bileşenlerin

diskriminant analizi, yapay sinir ağları gibi çeşitli analizlerde kullanımı ile farklı sınıflandırma temelli modellerin kurulabildiği gibi (Bolzon de Muñiz ve ark., 2016; Nisgoski ve ark., 2018; Russ ve ark., 2009; Russ ve Fišerová, 2011); PCA ve ters en küçük kareler (ILS) yöntemlerinin kombinasyonu olan, temel bileşenler regresyonu (PCR) olarak isimlendirilen, birinci dereceden çok değişkenli regresyon temelli modellerin de kurulduğu görülmektedir (Brunner ve ark., 1996).

Çok değişkenli veri analiz metotları yukarıda bahsedilen denetimsiz metotların yanı sıra, incelenecek gruplara dair bilgilerin (y gözlemi) olduğu durumlarda da kullanılabilmektedir. Ayırt edilmek istenen gruplara dair bilgilerin de işlendiği, bir diğer ifade ile sınıf etiketleri bilinen eğitim veri seti ile model kuran bu analizlere denetimli veri analizleri denilmektedir (Brereton, 2007; Wehrens, 2011). Denetimli analizlerde modellenerek tahmin edilmek istenen sınıf (y) kesikli veya sürekli değerlere sahip olabilmektedir. Kesikli değerlerin tahmin edilmesinde kullanılan modeller sınıflandırma problemlerinin konusuyken, sürekli değerlerin tahmin edilmesinde kullanılan modeller ise regresyon başlığı altında karşımıza çıkmaktadır (Wehrens, 2011).

2.4.2.3.Regresyon

Doğrusal regresyon modeli, basit denetimli bir modeldir. Bağımlı değişken Y ile bağımsız değişkenler X_1, X_2, …, X_p arasındaki ilişkinin doğrusal olduğunu varsayar. Bu ilişki aşağıdaki gibi ifade edilmektedir (James ve ark., 2013);

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{2.15}$$

Burada β_0 (sabit terim) ve β_1 (eğim) katsayıları modelin parametrelerini, ε hata terimini yani modelle açıklanmayan tüm değişkenleri ifade etmektedir. Regresyon modeli kurulurken öncelikle model parametreleri tahmin edilmekte, ardından modelin yeni örnekler için tahmin gücü hesaplanmaktadır (James ve ark., 2013). Regresyon modellerinde veri türüne bağlı olarak farklı yöntemler kullanılabilmektedir. Varmuza ve Filzmoser (2008)'e göre regresyon modelleri için veri türleri ve uygun metotlardan bazıları Tablo 2.3'te verilmiştir.

Tablo 2.3: Regresyon modelleri için x ve y veri tiplerine göre uygun modeller

X değişkenlerinin sayısı	Y değişkenlerinin sayısı	Analiz Tipi	Metot
1	1	Basit	• OLS • Sağlam regresyon
> 1	1	Çoklu	• PLS • PCR • Çoklu OLS • Sağlam regresyon • Lasso regresyon • Ridge Regresyon
> 1	> 1	Çok Değişkenli	• PLS2 • CCA
OLS: Basit En Küçük Kareler Regresyon, Sağlam Regresyon, PLS: Kısmi En Küçük Kareler Regresyon, PCR: Temel Bileşenler Regresyonu, CCA: Kanonik Korelasyon Analizi			

2.4.2.4.Sınıflandırma

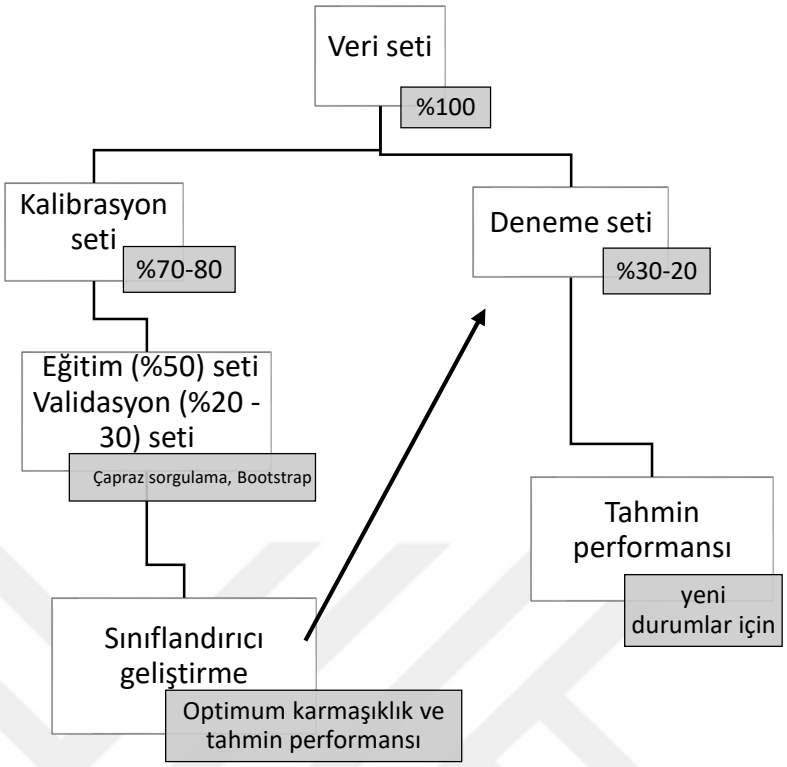
Denetimli veri analiz metotlarından biri olan sınıflandırmadaki amaç, çok değişkenli yapıyı kurallar oluşturmak için uygun şekilde özetlemek ve tahmin performansına göre sınıflar arasında en iyi ayrımı sağlamaktır. Mümkün olduğu kadar yüksek güvenilirlikte sınıflandırma yapılması için grup üyeliğinin bilinmediği yanlış sınıflandırılmış nesnelerin sayısını en aza indiren kurallar kullanılmaktadır. Denetimli sınıflandırmanın kümeleme analizinden farkı, sınıfların sayısının ve sınıfların kendilerinin bilinmesidir. Bir diğer ifade ile sınıflandırmada neyin aranacağı eğitim veri seti aracılığıyla tanımlanmaktadır. İdeal bir sınıflandırma modelinin iyi bir tahmin yeteneğinin yanı sıra farklı sınıfları birbirinden ayıran özellikler hakkında da bilgi vermesi ve eğitim seti dışındaki örnekleri de iyi tahmin etmesi beklenmektedir (Nisbet ve ark., 2018; Olivieri, 2018; Varmuza ve Filzmoser, 2008).

Sınıflandırma yapılırken gerçekleştirilen aşamaları şu şekilde sıralamak mümkündür;

1- Sınıfların belirlenmesi (sınıflandırma probleminin dış hatalarını belirlemek),
2- Sınıfların sınırlarının belirlenmesi (her bir sınıfın sınırlarını girdi değişkenleri açısından tanımlanmak),
3- Karar kurallarının oluşturulması (sınıf sınırlarından her bir sınıfı tanımlamak için karar kuralları oluşturmak),
4- Sınıfların öncelik olasılıklarının belirlenmesi (veri setinde bir sınıfın ortaya çıkma sıklığına bağlı olarak her bir sınıfın öncelik olasılığını belirlemek)

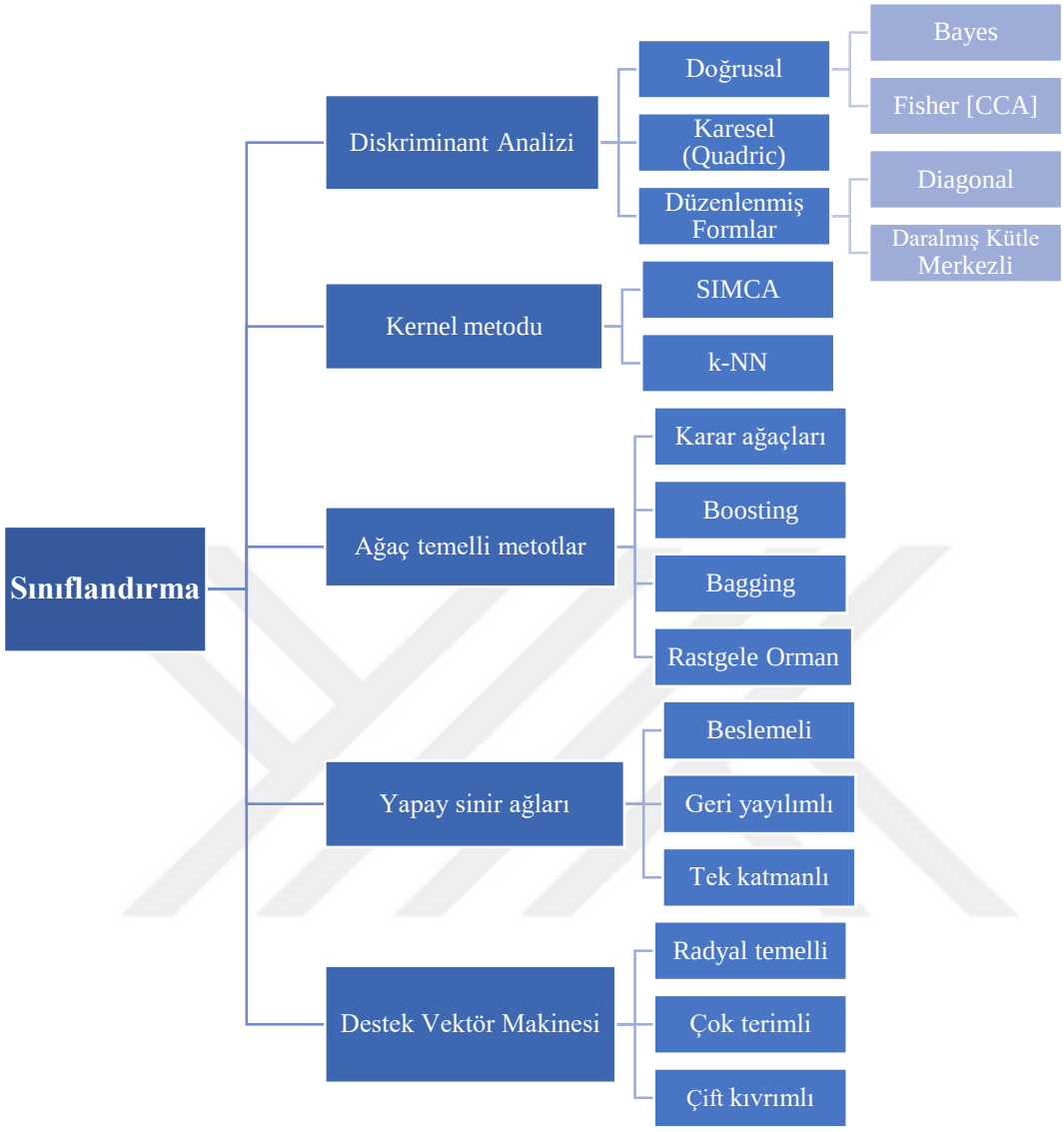
5- Kullanılacak performans kriterlerinin belirlenmesi (her bir sınıfın yanlış sınıflandırma değerlendirmesinde önemli olan kriterleri belirlemek) (Nisbet ve ark., 2018).

İyi bir sınıflandırmada yeni nesnelerin doğru gruplara atanması istendiği için, model kurulurken kullanılan örneklerin modelin değerlendirilmesinde kullanılması yanlış ve yanlı sonuçların elde edilmesine sebep olacaktır. Bu sebeple incelenen veri setlerinin birbiri ile kesişmeyen eğitim (öğrenme) ve test gruplarına ayrılması gerekmektedir. Bu ayrılmanın ardından eğitim setinde bulunan değişkenler ve gözlemler (ölçümler ve örnekler) arasındaki ilişki ortaya konularak model kurulmaktadır (kalibrasyon). Bu ilişki çeşitli nesne sınıflarını ayırmak için özelliklerin nasıl ağırlıklandırılması ve birleştirilmesi gerektiğine karar verilmesi aracılığıyla kurulmaktadır. Modelle açıklanan ilişkinin modelin kurulmasında kullanılan örnekleri ne kadar iyi açıkladığı, bir diğer ifade ile modelin kalibrasyon örneklerini ne kadar iyi tahmin ettiği belirlenmektedir. Bazı durumlarda kalibrasyon için ayrılan veri seti çapraz sorgulama veya bootstrap gibi metotlar aracılığıyla eğitim ve validasyon olmak üzere iki gruba ayrılmaktadır. Bu tarz yaklaşımlarda eğitim setinde belirlenen ağırlıklar, genel sınıflandırma hatasını hesaplamak için eğitime katılmayan bir sete (validasyon seti) uygulanmaktadır. Belirlenen genel sınıflandırma hatası ile model parametrelerinde ayarlama yapılmaktadır. Bu sayede öncelikle modelin optimum karmaşıklığı tahmin edilmekte, ardından optimum karmaşıklık tüm kalibrasyon setine uygulanarak model oluşturulmaktadır. Her iki durumda da, oluşturulan model son olarak modelin eğitiminde hiçbir şekilde kullanılmamış olan test grubu ile tahmin edilmekte, sınıflandırma modeli değerlendirilmektedir. Sınıflandırma analizlerinde izlenen bu genel hat Şekil 2.10 ’da verilmiştir (Nisbet ve ark., 2018; Olivieri, 2018; Varmuza ve Filzmoser, 2008).



Şekil 2.10: Sınıflandırma analizi şeması (Varmuza ve Filzmoser, 2008)

Modelin kurulmasında kullanılmayan yeni verilerin (test veri seti) hangi sınıfa ait olduklarının belirlenmesi için birçok farklı analiz metodu mevcuttur. Bunlardan biri olan diskriminant analizi yöntemlerinde ortalamaların ve kovaryansların parametrik bir tanımı kullanılmakta; gözlemler en yüksek olasılık yoğunluğuna sahip sınıfa atanmaktadır. Bir diğer yöntem olan en yakın komşu yöntemleri (NN-Nearest neighbourhood) ise her bir nesnenin birbiri ile benzerliklerine odaklanarak, bölgesel yoğunlukların tahmini ile nesneleri en olası sınıfa atamaktadır. Bu yöntemlerin yanı sıra nesneler arası benzerliklerin direkt kullanıldığı tipteki kernel metotları (SVM: Destek vektör makinesi),bir dizi ikili karardan oluşan ağaç temelli analizler (DT-Karar ağaçları),nesnelerin aynı anda birden fazla sınıfta bulunmasına imkan veren yumuşatılmış analizler (SIMCA- Sınıf karşılaştırmasının hafif bağımsız modellemesi), ve sinir ağları temelli sınıflandırma analizleri (ANN- Yapay sinir ağları) de kullanılmaktadır (Şekil 2.11) (Brereton, 2007; Wehrens, 2011).



Şekil 2.11: Sınıflandırma amaçlı kullanılan bazı modeller (Brereton, 2007; Nisbet ve ark., 2018; Olivieri, 2018; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

2.4.2.5.Diskriminant Analizi

Diskriminant analizinde p boyutlu veri içerisinde her bir sınıfın normal dağılıma sahip olduğu varsayılırsa şu şekilde gösterilebilir: N_p (μ_k, $\sum_k$). Bu durumda yeni bir nesne p boyutlu uzayda bir nokta olarak görülecektir ve yeni nesnenin bulunduğu nokta en yüksek olasılık yoğunluğuna sahip gruba atanabilmektedir. Bu tarz bir ayrım en çok olabilirlik (EÇO) diskriminant analizi olarak isimlendirilmektedir. Aynı sınıflandırma kuralı bir nesnenin bir sınıf merkezine olan mesafesini o yöndeki standart sapma açısından ifade eden Mahalanobis uzaklığı (denklem 2.4)

ile de açıklanabilmektedir. Sınıf atamaları yapılırken, tüm sınıfların aynı kovaryans matrisine (∑) sahip olduğu varsayıldığında doğrusal diskriminant analizi (LDA), her sınıfın kendi kovaryans matrisi ($\sum_i$) olduğu varsayıldığında ise karesel diskriminant analizi olarak isimlendirilmektedir. Kovaryans matrisinin (S_i) hesaplanması şu şekilde verilebilir (Wehrens, 2011);

$$S = \frac{1}{n-G}\sum_{i=1}^{G} n_i S_i \tag{2.16}$$

Bu denklemde; G: grupları, ni: i grubundaki nesnelerin sayısını göstermektedir.

Doğrusal diskriminant analizlerinin bir yaklaşımı olarak Bayes diskriminant analizi veri setinde nesnelerden daha fazla değişken olduğu durumlarda ve değişkenler arasında yüksek oranda korelasyon olduğunda yeterli olmamaktadır. Bu gibi durumlar için farklı metotlar (Fisher yöntemi) ve yaklaşımlar (Düzenlenmiş diskriminant analizleri: diagonal diskriminant analizi, daralmış kütle merkezli diskriminant analizi gibi) geliştirilmiştir (Şekil 2.11) (Wehrens, 2011). Fisher yöntemi bir diğer adıyla kanonik diskriminant analizi çok değişkenli verileri tek değişkenli bir ayırt edici değişkene dönüştürerek, dönüştürülen grupların mümkün olduğunca ayrılmasını sağlar. Bu dönüştürme içinde orijinal x-değişkenlerinin doğrusal kombinasyonu (gizli değişkenler veya kanonik değişkenler) kullanılmaktadır (Varmuza ve Filzmoser, 2008).

2.4.2.6.Kısmi En Küçük Kareler (PLS)

Kısmi en küçük kareler (PLS) yaklaşımı temel bileşenler analizine (PCA) benzer şekilde boyut indirgeme ile ilgilenmektedir. Sadece bağımsız değişkenlerin (X) modellenmesi ile ilgilenen PCA dan farklı olarak, bağımlı değişkenler (y) de boyut indirgemesinde kullanılmaktadır. Bu sayede toplam değişkenliğin yanı sıra gruplar arasındaki değişkenlikte analize katılmış olmaktadır (Liu ve Rayens, 2007; Varmuza ve Filzmoser, 2008).

PLS genel olarak boyut indirgeme ardından regresyon analizi uygulanarak kullanılması sebebiyle temel bileşenler regresyon analizi (PCR) ile karşılaştırılmaktadır. Temelde iki metodun da model yapıları benzerlik göstermektedir. X verisi öncelikle latent (gizli) değişkenlerine (bileşenler) dönüştürülmekte, ardından bu yeni değişkenler regresyon analizi için kullanılmaktadır. İki model arasındaki fark burada ortaya çıkmaktadır. PCR bu dönüşümü sadece X verisi ile dönüştürülmüş temel bileşen skorları ile gerçekleştirirken; PLS X ve y'deki

en yüksek varyansı yakalayarak ve bu matrisler arasındaki korelasyonu (X ve Y arasındaki kovaryansı) en üst seviyeye çıkararak yeni bileşenleri oluşturmaktadır (Varmuza ve Filzmoser, 2008; Wehrens, 2011). Orijinal X matrisindeki maksimum spektral varyansı açıklarken, aynı zamanda X ve y arasındaki maksimum korelasyonu da açıklamak amacıyla, ortogonal ve normalleştirilmiş skorlar ve ortogonal ve normalleştirilmiş olmayan yüklemeler kullanılmaktadır. Genel hatlarıyla regresyon analizinde PLS 'nin kullanımında öncelikle maksimum korelasyona göre ilk PLS bileşenleri hesaplanmaktadır. Ardından X ve Y için hata matrisi oluşturularak yeni PLS bileşenleri bu hata matrisinden belirlenmektedir. Bu işlem y'nin modellenmesinde yapılabilecek iyileştirmeler bitene kadar devam etmektedir. PLS bileşenleri hata matrisinden hesaplandığı için bileşen parametreleri (skorlar, yüklemeler, ağırlıklar) orijinal X verisi ile değil hata matrisi ile ilişkili olmaktadır. Fakat orijinal X verisi ile bileşen yüklemeleri, bileşen skorları arasında ilişkiye ve modelin regresyon katsayılarına ulaşmak mümkündür. PCR gibi doğrusal bir model olan PLS-R, X ve y arasındaki doğrusal kombinasyonu bulmaya çalışmaktadır. PCA 'da kullanılan formüle benzer şekilde, PLS ile ortaya koyulan ilişki şu şekilde ifade edilebilir;

$$X = T.\ P^T + E \tag{2.17}$$

$$Y = U.\ Q^T + F \tag{2.18}$$

Bu denklemde; T : X-skorları, U : Y-skorları, P : X-yüklemeleri, Q : Y-yüklemeleri, E : X-hataları, F : Y-hatalarını göstermektedir.

PLS ile iki grup arasında ayrım hedeflendiği zaman (y-vektörünü kullanan) PLS1 algoritması kullanılırken, birden fazla grubun ayrılmasının hedeflendiği (Y-matrisinin kullanıldığı) durumlarda PLS2 algoritması kullanılmaktadır. PLS1 de PCR da olduğu gibi her bir grup için ayrı modeller kurulurken, PLS2 de tüm gruplar arası korelasyon dikkate alınarak tek bir modelle tahminde bulunulmaktadır (Barker ve Rayens, 2003; Brereton, 2007; Brereton ve Lloyd, 2014; Dinç, 2007; Maitra ve Yan, 2008; Olivieri, 2018; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

PLS asli olarak sınıflandırma ve ayrım için tasarlanmış olmasa da son yıllarda uygulamada oldukça olumlu sonuçlar alınmaktadır. PLS bileşenlerine regresyon analizi yerine diskriminant analizi uygulandığında kısmi en küçük kareler diskriminat analizi (PLS-DA) olarak isimlendirilmektedir. PLS-DA algoritması temelde iki grup arasında düz bir çizgi ile

sınıflandırma yapmayı amaçlamaktadır. Temel bileşenler ile karşılaştırıldığında benzer şekilde boyut indirgeme ve indirgenen boyutlara göre sınıflandırma yapmaya dayanmaktadır. Fakat PCR ile karşılaştırılmasında olduğu gibi PLS-DA bu işlemi gruplar arası ve gruplar içi değişkenliği de hesaba katarak denetimli bir şekilde gerçekleştirmektedir. PLS-DA doğrusal bir ayrım yaptığından sentroidlere Öklid mesafesi (EDC), doğrusal diskriminat analizi (LDA), kanonik korelasyon analizi (CCA), temel bileşenler regresyonu (PCR), doğrusal destek vektör makineleri (SVM) gibi metotlar ile oldukça yakından ilişkilidir. Tüm bu algoritmaların amacı gruplar arasında ayrımı sağlayan ayırıcı/ kanonik değişken/ karar fonksiyonunu tespit etmektir (Barker ve Rayens, 2003; Brereton, 2007; Brereton ve Lloyd, 2014; Geladi ve Kowalski, 1986; Liu ve Rayens, 2007; Olivieri, 2018).

2.4.2.7.Sınıf karşılaştırmasının hafif bağımsız modellemesi (SIMCA)

Keskin ayrımların yapılmasını amaçlayan klasik diskriminant analizlerinin aksine, yumuşatılmış sınıflandırma analizlerinde sınıflar birbirleri ile örtüşebilmekte ve bir nesne aynı anda birden fazla sınıfta bulunabilmektedir. SIMCA metodunun avantajlarından bir diğeri de her bir grup aynı ya da farklı temel bileşenler kullanılarak birbirinden bağımsız olarak modellenebilmektedir. SIMCA analizi ile öncelikle sınıflar birbirinden bağımsız olarak modellenmekte, ardından sınıf uzaklıkları ve modelleme gücü hesaplanmaktadır. Modelleme gücü hesaplanırken ayrım için kullanışlı değişkenler belirlenmekte ve belirlenen değişkenlerin sınıfların ayrımında ne kadar etkili olduğu tespit edilmektedir (Brereton, 2007; Varmuza ve Filzmoser, 2008).

2.4.2.8.En Yakın Komşu Metodu (k-NN)

Herhangi bir dağılım varsayımına dayanmayan en yakın komşu metotları, nesneler arasındaki, özellikle de en yakın nesneler arası mesafelere odaklanmaktadır. Komşuluk ilişkisini tespit etmek için yeni nesne ve tüm eğitim nesneleri arasındaki mesafe genellikle Öklid uzaklığı ile hesaplanmaktadır. *k*-en yakın komşu sınıflandırmasındaki, k yeni grubun sınıf atamasında bakılacak sınıf sayısını belirlemektedir. Yani $k=1$ olduğunda sınıflandırma yeni nesneyi en yakın mesafedeki sınıfa atamaktadır. $k>1$ olduğu durumlarda ise sınıflandırma çoğunluk oyuna göre gerçekleştirilmektedir. *k*-NN ile yeni bir nesne sınıflandırılırken her seferinde eğitim setinin tamamı kullanılmaktadır. Bu sebeple büyük veri setlerinde bu analiz oldukça zaman alıcı olabilmekte ve yanlı bir analiz olmaması için her sınıfta bulunan nesne sayısının birbirine

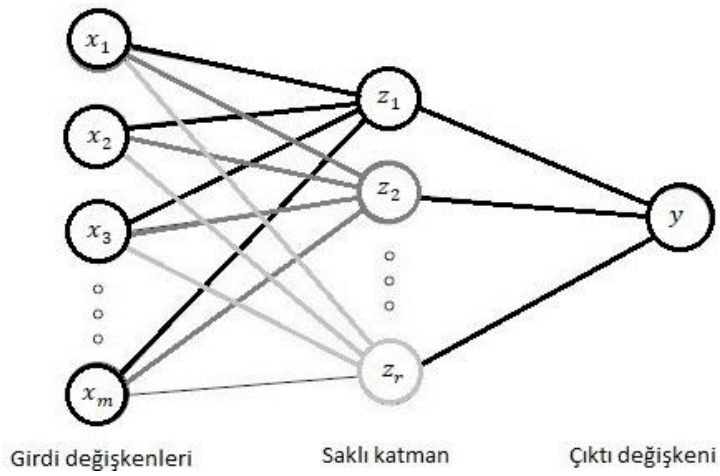
yakın olması gerekmektedir. Eğitim setine bağlı bir sınıflandırma yapıldığından genel bir optimum *k* değeri bulunmamakta, her veri seti için optimum k değerinin tespit edilmesi gerekmektedir. Ayrıca tüm değişkenlerin aynı öneme sahip olduğu varsayıldığı için spektroskobik veriler gibi çok değişkenli veri setlerinde önemli değişkenlerin belirlenmesi veya farklı mesafe ölçüm yöntemlerinin kullanılması gerekmektedir (Brereton, 2007; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

2.4.2.9.Ağaç Temelli Yaklaşımlar

Ağaç temelli yaklaşımlarda veriler (bağımsız değişkenler), sonunda sınıflandırmaya yol açan, bir dizi soru ile sorgulanır. Bu sorgulama bir veri kümesini birkaç dal benzeri parçaya bölerek değişkenler üzerindeki sıralı kararlar almaya dayanmaktadır. Modellemenin temeli hangi soruların en bilgilendirici olduğuna karar vermektir. Bu sebeple sadece sonuca katkıda bulunan değişkenler ağaçta yer almaktadır. Ağaç temelli yaklaşımlarda kullanılan değişkenler; sürekli veya kesikli sayılar olabildiği gibi kategorik veya sıralı değişkenler de olabilmektedir. Ağaç temelli yaklaşımlarda hangi değişkenlerin etkili olduğu açıkça görüldüğü için sonuçların yorumlanması ve görselleştirilmesi oldukça kolay olmaktadır. Ağaç temelli yaklaşımlarda temel metot karar ağaçlarıdır. Karar ağaçları eğitim setine oldukça bağımlı bir yapı göstermesi sebebiyle kurulan modeller genelleme gücünden yoksundur. Bu durum yeni verilerin tahmininin beklenilenden düşük olmasına sebep olmaktadır. Karar ağaçlarının sahip olduğu bu dezavantajların önüne geçmek için, yorumlama yeteneğini biraz daha azaltarak birden fazla karar ağacının kombinasyonlarının kullanıldığı yöntemler geliştirilmektedir (Rasgele Orman, Bagging, Boosting) (James ve ark., 2013; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

2.4.2.10. Doğrusal Olmayan Metotlar

Sınıflandırma amacıyla kullanılan doğrusal yaklaşımlı yöntemlerin yanı sıra oldukça esnek, pek çok karmaşık ilişkiyi modelleme yeteneğine sahip doğrusal olmayan tekniklerde vardır. Bu tekniklerden en aktif kullanılanları destek vektör makinesi (SVM) ve yapay sinir ağları (ANN) metotlarıdır. SVM büyük veri setleri kullanılmadığı durumlarda oldukça kullanışlı olmaktadır. ANN ise ciddi bir eğitim setine ihtiyaç duymasının yanı sıra, yüksek boyutlu verilerin işlenmesi gereken durumlarda öncelikle PCA gibi veri indirgeme yöntemlerinin uygulanmasına ihtiyaç duymaktadır (Varmuza ve Filzmoser, 2008; Wehrens, 2011).



Şekil 2.12: Yapay Sinir Ağlarının şematik gösterimi (Varmuza ve Filzmoser, 2008).

Hem istatistik hem de yapay zeka alanlarında ayrı ayrı geliştirilen ANN pek çok farklı öğrenme yöntemini içeren (beslemeli, geri yayılımlı, tek katmanlı algılayıcı vb.) oldukça esnek modelleme tekniklerine sahiptir (Varmuza ve Filzmoser, 2008; Wehrens, 2011). Öncelikle x değişkenlerinin doğrusal kombinasyonları oluşturulmakta, ardından bu kombinasyonlara doğrusal olmayan fonksiyonlar uygulanarak saklı katmandaki sinirler(z) oluşturulmaktadır (Şekil 2.12). Oluşturulan bu sinirler ile nihai çıktıların(y) tahmininde uygulanmak istenen analizdeki amaca uygun olarak farklı yaklaşımlar (doğrusal ve doğrusal olmayan regresyon, sınıflandırma vb.) kullanılmaktadır. ANN regresyon amaçlı kullanıldığında modelin çıktısı regresyon yanıt değişkeni iken; sınıflandırma amaçlı ANN kullanımında y değişkeni her bir nesne için sınıf üyeliğinin (hedef değişkeni) tahmin edilmesi olarak karşımıza çıkmaktadır (Varmuza ve Filzmoser, 2008; Wehrens, 2011).

Destek vektör makineleri (SVM) diğer modellerden farklı olarak bir sınıfı bir bütün olarak modellemeyi denememekte, sınıflar arasındaki sınırları bulmaya çalışmaktadır. Bu işlem genellikle orijinal x değişkenlerinden daha büyük boyutta olan, dönüştürülmüş x değişkenler uzayında, nesne grupları arasında doğrusal sınırların (sınıf sınırları) oluşturulması ile gerçekleşmektedir. Orijinal değişkenler yerine daha yüksek boyuttaki uzayın (hiper düzlemin) kullanılmasıyla sınıflar arasındaki mesafenin (marj) maksimuma çıkması hedeflenmektedir. Verilerin dönüştürülmesi ve etkin şekilde yüksek boyutlu uzayda temsil edilmesi kernel (çekirdek) fonksiyonları olarak adlandırılan özel bir hesaplama yöntemi aracılığıyla gerçekleştirilmektedir. SVM de kullanılan fonksiyonlar ayarlanabilir parametrelerin sayısına ve sınırların karmaşıklığına bağlı olarak değişiklik göstermekte ve analiz amacına uygun olan

kernel fonksiyonu kullanılmaktadır. En sık kullanılan kernel fonksiyonları; veri hakkında ön bilgiye ihtiyaç duymayan genel amaçlı radyal temelli fonksiyon (Radial Basis Function), genellikle görüntü analizlerinde kullanılan çok terimli fonksiyon (Polynomial Function), yapay sinir ağlarını temsilen kullanılabilen çift kıvrımlı fonksiyon (Sigmoidal Function) olarak karşımıza çıkmaktadır. SVM algoritması iki sınıf arasındaki boşluk mesafesinin maksimum olduğu alan olan, en iyi sınırı aramaktadır. Sınıfları modellemek için tüm eğitim seti yerine sadece sınır örnekleri(destek vektörler) yeterlidir. Bu sebeple destek vektörlerinin belirlenmesi SVM metotlarının en önemli kısımlarından biridir. Yeni bir örnek sisteme geldiğinde, mesafe ölçen diğer modellerde olduğu gibi, sınırın hangi tarafında olduğuna göre bir sınıfa atanmaktadır (Brereton, 2007; Varmuza ve Filzmoser, 2008; Wehrens, 2011).

Genel olarak bakıldığında sınıflandırma amaçlı pek çok farklı yöntem bulunmaktadır. Pratikte sınıflandırma için uygulanabilecek en iyi yaklaşım olarak farklı metotların denenerek, sonuçların karşılaştırılması önerilmektedir (Nisbet ve ark., 2018). Farklı yaklaşımlar ile kurulan modellerin performanslarında farklılıklar gözlense de, genellikle bu farklılıkların kullanılan algoritmanın kendisinden öte; deney tasarımı, ölçeklendirme, uzaklık ölçümleri, değişken seçimi, ön işlem kullanımı, doğrulama yöntemi gibi ölçümlerden kaynaklandığı görülmektedir (Brereton, 2007).

Yapılan literatür incelemesinde NIRS ile ağaç türü teşhisinde yukarıda bahsedilen çok değişkenli analizlerden bazıları üzerinde yoğunlaşıldığı görülmektedir. Regresyon yaklaşımıyla ağaç türlerinin NIRS aracılığıyla ayrımında ilk olarak denetimsiz metotlardan PCR ‘nin kullanıldığı (Brunner ve ark., 1996) ardından gerçekleşen pek çok çalışmada bu analizin yerini denetimli metotlardan kısmi en küçük kareler regresyon analizine (PLS-R) bıraktığı (Bergo ve ark., 2016; Braga ve ark., 2011; Brunner ve ark., 1996; Flaete ve ark., 2006; Haartveit ve Flæte, 2008; Pastore ve ark., 2011; Tounis, 2009; Yang ve ark., 2016) görülmektedir. Ağaç türlerinin yakın kızılötesi verileri kullanılarak diskriminant yaklaşımı ile ayrılması üzerine gerçekleştirilen çalışmaların hem ham veri ile (Espinoza ve ark., 2012; Nisgoski ve ark., 2016; Tsuchikawa ve ark., 2003) hem de PCA ve PLS aracılığıyla indirgenmiş veriler ile (Bolzon de Muñiz ve ark., 2016, Nisgoski ve ark., 2018; Russ ve ark., 2009; Russ ve Fišerová, 2011) gerçekleştirildiği görülmektedir. NIRS’ın ağaç türü teşhisinde kullanımına dair gerçekleştirilen çalışmalarda kullanılan çok değişkenli analiz metotları Tablo 2.4’ de özetlenmiştir.

Tablo 2.4: NIRS ile ağaç türü teşhisi konusunda gerçekleştirilen çalışmalarda geçmişten günümüze kullanılan çok değişkenli analiz metotları

Yayın Yılı	Yazar(lar)	Çok Değişkenli Analizler						
		CA	PCA	- R	- DA	SIMCA	ANN	SVM
1996	Brunner ve ark.	X		PCR,PLS-R				
2003	Tsuchikawa ve ark.				DA			
2004	Gierlinger ve ark.	X	X			X		
2006	Flæte ve ark.		X	PLS-R	PLS-DA			
2008	Haartveit ve Flæte			PLS-R	PLS-DA			
2008	Adedipe ve ark.		X			X		
2009	Russ ve ark.		X		PCA-DA			
2009	Tounis, E.			PLS-R	PLS-DA			
2011	Braga ve ark.			PLS-R	PLS-DA			
2011	Pastore ve ark.			PLS-R	PLS-DA			
2011	Russ ve Fiserova				PCA-DA			
2012	Dawson-Andoh ve Adedipe		X			X		
2012	Espinoza ve ark.		X		DA			
2014	Shou ve ark.		X			X		
2014	Zhang ve ark.		X					
2015	Horikawa ve ark.		X		PLS-DA			
2015	Lang ve ark.		X		LDA			
2015	Yang ve ark.		X	PLS-R	PLS-DA			
2016	Bergo ve ark.			PLS-R	PLS-DA			
2016	Bolzon de Muñiz ve ark.		X		PCA-LDA	X		
2016	Hwang ve ark.		X		PLS-DA			
2016	Nisgoski ve ark.		X		LDA DA			
2017	Lang ve ark.		X					
2017	Lazarescu ve ark.		X		PLS-DA		X	
2017a	Nisgoski ve ark.		X					
2017b	Nisgoski ve ark.		X			X	X	
2018	Nisgoski ve ark.		X		PCA-LDA	X		
2018	Snel ve ark.				PLS-DA	X		
2019	Leandro ve ark.		X		PLS-DA			
2019	Pace ve ark.		X	PLS-R	PLS-DA			
2019	Li ve ark.							SVM
2019	Vieira ve ark.		X		LDA			
2019	Yong ve ark.		X		PLS-DA	X		
2020	Zhou ve ark.				PLS-LDA			RBF-SVM

*ANN: Yapay Sinir Ağları, CA: Kümeleme analizleri, DA: Diskriminant Analizleri, LDA: Doğrusal Diskriminant Analizleri, PCA: Temel Bileşenler Analizi, PCR : Temel Bileşenler Regresyonu, PLS: Kısmi En Küçük Kareler, PLS-DA: Kısmi En Küçük Kareler Diskriminat Analizi, PLS-R: Kısmi En Küçük Kareler Regresyonu, R: Regresyon, SIMCA: Sınıf Karşılaştırmasının Hafif Bağımsız Modellemesi

2.4.3. Model Performansı

Belirli bir veri setiyle (eğitim seti) kurulan kalibrasyon modelinin performansını değerlendirmek için, modelin tahminleri ile gerçek değerlerin ne kadar uyumlu olduğunun belirlenmesine ihtiyaç vardır. Bir diğer ifade ile belirli bir gözlem için tahmin edilen değerin o gözlemin gerçek değerine ne kadar yakın olduğunun ölçülmesi gereklidir. Bu ölçüm hem model kurmada kullanılan eğitim seti hem de modele katılmayan test seti için ayrı ayrı hesaplanmaktadır. Model kurmada kullanılan eğitim verilerinin tahmini ile bu verilerin gerçek değerleri arasındaki fark eğitim hatası olarak isimlendirilirken, model kalibrasyonuna katılmayan yeni örneklerin gerçek değerleri ile modelin tahmin değerleri arasındaki fark test hatası olarak isimlendirilmektedir (Hastie ve ark., 2017; James ve ark., 2013).

Model performans değerlendirmesinde kabaca izlenen yol, sayısal bağımlı değişken tahmininde regresyon modeli kurulumu için, öncelikle model parametrelerinin tespit edilmesi, ardından yeni örnekler için nihai modelin tahmin gücünün hesaplanması aşamalarını içermektedir. Model parametrelerinin tahmininde genellikle artıkların kareleri toplamı (RSS) veya en küçük kareler (LS) yöntemi ile RSS 'nı minimum yapan katsayıların hesaplanması yaklaşımları kullanılmaktadır. Katsayıların tahmininden sonra tahmin edilen katsayı değerinin gerçek değerden uzaklığı hesaplanarak standart hata (SEP) tespit edilmektedir. Standart hatanın belirlenmesi ile güven aralıklarının tespiti ve hipotez testlerinin kurulması mümkün olmakta ve bu sayede açıklayıcı değişkenler ve önemli değişkenler belirlenmektedir. İlgili model kurulduktan sonra eğitim setinin ne kadar iyi modellendiği belirlenerek ilgili kriterler ile en uygun model seçilmektedir (James ve ark., 2013).

Tüm performans kriterlerinin temeli, bağımsız bir test setine veya validasyon setine ayrılan örneklerin gerçek değerlerinden, kurulan modelin tahmin değerlerinin çıkarılması ile bulunan $(y_i - \hat{y}_i)$ tahmin hatalarıdır (artıklar). Tahmin hataları belirlendikten sonra performans değerlerinin belirlenmesinde çeşitli hata hesaplamaları kullanılmaktadır. Bunlardan en çok kullanılanları; standart tahmin hatası(SEP), hata kareler ortalaması (MSE), kök hata kareler ortalaması (RMSE), artık kareler toplamı (RSS), tahmini artık hata kareleri toplamı (PRESS) olarak verilebilmektedir (James ve ark., 2013; Nisbet ve ark., 2018; Olivieri, 2018; Varmuza ve Filzmoser, 2008).

Hata dağılımındaki artıkların standart sapması, uygulamada standart tahmin hatası (SEP) olarak isimlendirilmekte ve şu şekilde hesaplanmaktadır (Varmuza ve Filzmoser, 2008);

$$SEP = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i - bias)^2} \tag{2.19}$$

Yanlılık denklemi (bias) ile birlikte,

$$bias = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{2.20}$$

Burada y_i gerçek değerleri, $\hat{y}_i$ modelin tahmin değerlerini, n gözlem sayısını, bias (yanlılık) tahmin hatalarının aritmetik ortalamasını ifade etmektedir. Tahmin hatalarının aritmetik ortalamasının karesi ise hata kareler ortalaması (MSE) olarak isimlendirilmektedir.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.21}$$

Hata karelerinin ortalamasının kare kökü "kök hata kareler ortalaması" (RMSE) olarak isimlendirilmektedir.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.22}$$

Modellerin geliştirilmesi ve optimizasyonu sırasında MSE kullanımı tercih edilse de, tahmin edilen özelliklere dair birimleri içermediği için uygulamada daha az fayda sağlamaktadır. Çapraz sorgulamada sıklıkla uygulanılan bir diğer benzer ölçü "tahmini artık hata kareleri toplamı" (PRESS) dır.

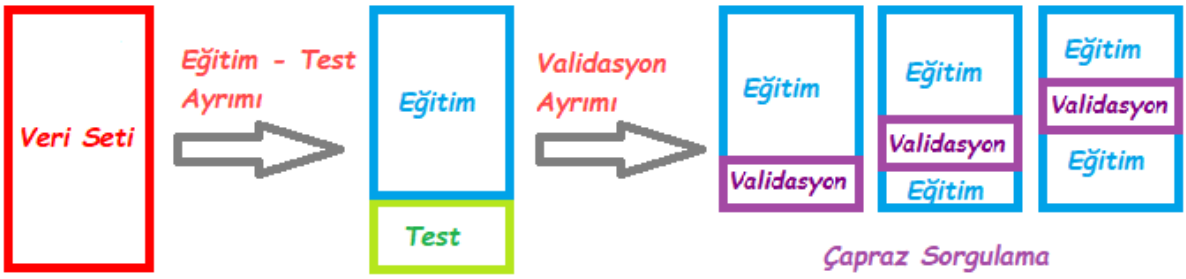
$$PRESS = \sum_{i=1}^{n}(y_i - \hat{y}_{i,-i})^2 \tag{2.23}$$

$\hat{y}_{i,-i}$: i'nci gözlem çıkarıldığında yi 'nin tahmini

Model performansına bakılırken iki farklı amaç söz konusudur: modelin seçimi ve modelin değerlendirmesi. Model seçiminde, en iyi modelin seçilmesi için farklı model performansları tahmin edilmekte, en iyi model parametreleri belirlenmektedir. Model değerlendirmesinde, seçilen nihai modelin yeni veriler için tahmin hatası belirlenmektedir. Her iki amaç için de örnek sayısının yeterli olduğu durumlarda, oluşturulan modelin iyileştirilmesi ve model hakkında daha fazla bilgi alınması amacıyla; modelin kurulmasında (kalibrasyon) eğitim seti (% 75), model seçimi için hata tahmininin belirlenmesinde validasyon seti (% 25), seçilen nihai modelin değerlendirilmesinde test seti (% 25) olmak üzere veri setinin üç parçaya ayrılması önerilmektedir. Yukarıda verilen tüm tahmin hataları her veri seti için verilebilmektedir. Hesaplanan hata tahmininin hangi veri setine ait olduğu alt indis ile gösterilmektedir (MSE_{CAL}, $RMSE_{CV}$, SEP_{TEST} gibi) (Hastie ve ark., 2017; Varmuza ve Filzmoser, 2008).

Validasyon aşamasının oluşturulmasında en sık kullanılan yeniden örnekleme metotları Çapraz Sorgulama (Cv: Cross Validation) ve Bootstrap yöntemleri olarak karşımıza çıkmaktadır. Çapraz sorgulamada kabaca eğitim seti içerisindeki örnekleri farklı alt gruplandırmalar ile tekrarlı bir şekilde modele uygulamaktadır. Model rasgele olarak ikiye ayrılan eğitim seti ile kurulmakta ve oluşturulan model sınama setinin (validasyon) tahmininde kullanılmaktadır (Şekil 2.13). Sınama seti için hesaplanan tahmin hatası ile modelin kurulmasına hiç katılmayan test setinin tahmin hatası, parametre tahminlerinin standart hatası ve yanlılıkları hakkında fikir yürütülebilmekte ve bu tahminler en iyi modelin seçilmesinde kullanılmaktadır. K parçalı çapraz sorgulama yönteminde; eğitim seti rastgele olarak eşit büyüklükte k adet gruba veya parçaya bölünmektedir. İlk parça validasyon için ayrılır (k), kalan kısım (k-1) modelin kurulmasında kullanılır. Bu işlem k kez tekrarlanır, her seferinde farklı bir grup validasyon için kullanılır. İşlem sonunda k kez hata kareleri ortalaması tahmin edilir ve en sonunda çapraz sorgulama tahmini ($CV_{(k)}$) hesaplanan tüm hata kareler ortalaması alınarak hesaplanır (James ve ark., 2013; Nisbet ve ark., 2018; Olivieri, 2018).

$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k} MSE_i \qquad (2.24)$$



Şekil 2.13: Çapraz Sorgulamanın (k=3 için) şematik görünümü (Cochrane, 2019).

Validasyon aşaması (uygun modeller içerisinde en uygun olanın belirlenmesi) için genellikle yeniden örnekleme yöntemleri (CV, bootstrap) tercih edilse de analitik olarak farklı kriterler kullanılarak (R^2, Bayes Bilgi Kriteri (BIC), Mallow CP, Akaike Bilgi Kriteri (AIC) gibi) bu hesaplamanın yapılması mümkündür. Bu yöntemler aracılığıyla parametre tahminlerinin standart hatası ve yanlılıklarına dair tahminler hesaplanmaktadır (Hastie ve ark., 2017). Modelin veriyi ne kadar iyi modellediğinin tahmin edilmesinde kullanılan kriterlerden biri olarak, açıklanan varyans olarak da isimlendirilen, R^2'nin hesaplanması;

$$R^2 = 1 - \frac{RSS}{TSS} \tag{2.25}$$

Burada RSS artık kareler toplamı,

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.26}$$

TSS Kareler toplamı,

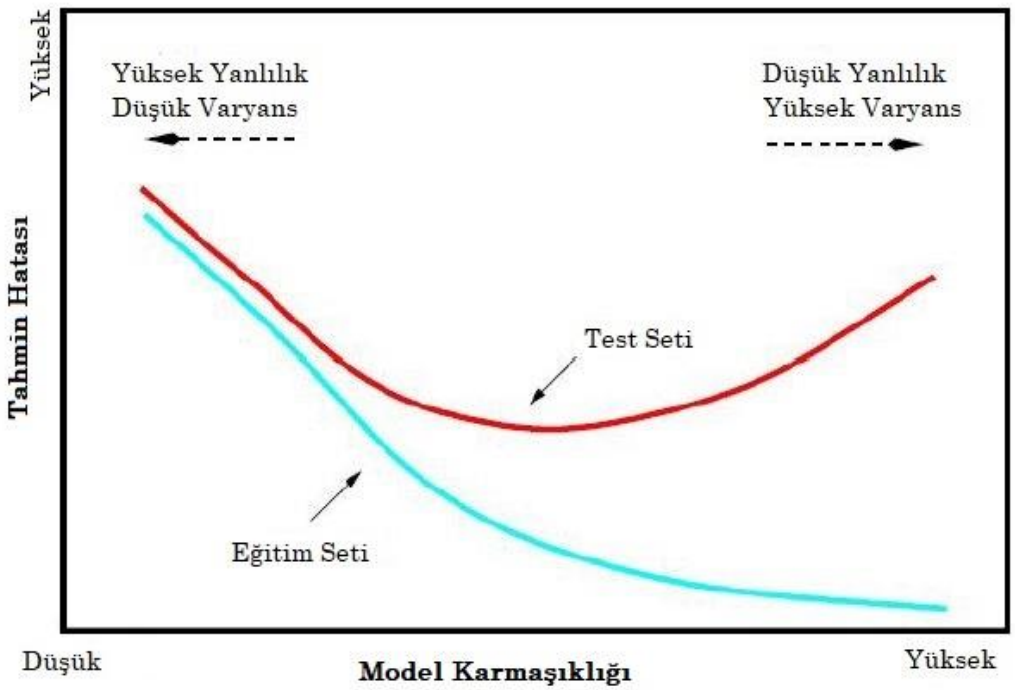
$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{2.27}$$

Model performansı belirlenirken asıl istenilen kurulan modelin yeni örnekler karşısında başarılı olmasıdır. Bu sebeple test hatası en önemli kriterlerden biri olarak karşımıza çıkmaktadır. Test hatasının ölçümü bağımlı değişkenin (Y) sayısal veya kategorik olmasına göre değişiklik göstermektedir. Genellikle sayısal bağımlı değişken için hata kareleri ortalaması (MSE) kullanılırken, kategorik bağımlı değişken için yanlış sınıflandırma oranı ölçüt olarak kullanılmaktadır (James ve ark., 2013). Genel olarak MSE 'nin en düşük değer alması istense de aynı zamanda modelin eğitim veri seti ile birebir uyumlu olması (overfit) istenmemektedir. Kurulan model eğitim setine birebir uyduğunda modelin esnekliği artmakta ve eğitim seti için

MSE ‘ı azalmaktadır. Fakat bir modelin başarısı için asıl kıstas test setinin yani yeni değerlerin tahmini ile hesaplanan MSE ’nin en küçük değeri almasıdır ve genellikle eğitim MSE ’nin en düşük olduğu nokta da test setine ait MSE en düşük olmamaktadır. Test setine ait beklenen MSE, varyans, yanlılık ve varyans hatasının bileşkesinden oluşmaktadır. Beklenen test hatasını ($E\ (y_i - \hat{y}_i)^2$) minimum yapmak için varyansın ve yanlılığın düşük olması gerekmektedir (Hastie ve ark., 2017; James ve ark., 2013; Varmuza ve Filzmoser, 2008).

$$E\ (y_i - \hat{y}_i)^2 = Var(\hat{y}_i) + [Bias(\hat{y}_i)]^2 + Var(\varepsilon) \tag{2.28}$$

Burada varyans ($Var(\varepsilon)$), farklı bir eğitim veri seti kullanarak tahminde bulunulduğunda $\hat{y}_i$'nın değişeceği miktarı ifade ederken; yanlılık ($Bias(\hat{y}_i)$), yeni örneğe yaklaşmakla ortaya çıkan hatayı ifade etmektedir. Genellikle kurulan modelin esnekliği arttıkça varyans artmakta, yanlılık ise azalmaktadır. Bu sebeple bu iki miktarın karşılıklı değişim ve etki oranı test MSE nin optimum değerini belirlemede kullanılmaktadır (Şekil 2.14). Dolayısıyla regresyon temelli bir model kurulurken parametrelerin tahmini için hedeflenen varyansın ve yanlılığın karesinin birlikte en düşük olduğu modeli bulmaktır (Hastie ve ark., 2017; James ve ark., 2013).



Şekil 2.14: Model karmaşıklığı ve tahmin hatası ilişkisi (Hastie ve ark., 2017, James ve ark., 2013, Varmuza ve Filzmoser, 2008).

Bağımlı değişkenin kategorik olduğu sınıflandırma temelli bir model kurulduğunda modelin performansı için en basit yaklaşım doğru sınıflandırılmış gözlemlerin oranına (Doğru Tahmin Oranı) bakmaktır. Sınıflandırma problemlerinin çoğu hedef değişkende (bağımlı değişken) sadece iki sınıfa sahiptir ve bu durum ikili sınıflandırma problemi olarak isimlendirilmektedir. İkili sınıflandırmanın doğruluğu, öngörülen (tahmin edilen) sınıf ile gerçek sınıf arasındaki ilişki analiz edilerek değerlendirilmektedir. Bu ilişkinin değerlendirilmesinde dört farklı sonuç durumu ile karşılaşılmaktadır. Bu sonuç durumlarında Doğru (1) terimi bir sınıfı temsil ederken Yanlış (0) terimi diğer sınıfı ifade etmektedir ve her iki sınıf için de tahmin edilen sınıf ve gözlenen sınıf aynı olduğu durumlarda tahmin Pozitif, farklı olduğu durumlarda tahmin Negatif olarak isimlendirilmektedir. Daha ayrıntılı bir ifadeyle; gerçek sınıfı Doğru olan örneğin tahmini Doğru sınıfında olduğunda "Doğru Pozitif"; gerçek sınıfı Doğru olan örneğin tahmini Yanlış sınıfında olduğunda "Doğru Negatif"; gerçek sınıfı Yanlış olan örneğin tahmini Yanlış sınıfında olduğunda "Yanlış Pozitif" ve gerçek sınıfı Yanlış olan örneğin tahmini Doğru sınıfında olduğunda "Yanlış Negatif" olarak isimlendirilmektedir (Nisbet ve ark., 2018). İkili sınıflandırma problemlerinde olduğu gibi, genellikle mevcut probleme göre sadece doğru tahmin oranı yeterli olmamaktadır. Daha ayrıntılı bir yaklaşım ihtiyacı duyulduğunda her bir yanlış sınıflandırma türüne (Doğru Negatif, Yanlış Negatif) bir değer atanmakta veya duyarlılık, özgüllük, pozitif tahmin oranı ve negatif tahmin oranı gibi değerler kullanılmaktadır. Kurulan sınıflandırma modelinin başarısını ortaya koymak için bahsi geçen değerlerin hesaplanmasında genellikle hata matrisleri oluşturulmaktadır. Hata matrisi modelden elde edilen tahminler ile gerçek değerlerin 2 x 2 lik bir matriste gösterilmesi ile elde edilmektedir (Tablo 2.5) (Hastie ve ark., 2017; James ve ark., 2013; Varmuza ve Filzmoser, 2008).

Tablo 2.5: Hata matrisi gösterimi.

Hata Matrisi		Gerçek Değer	
		Pozitif (1)	Negatif (0)
Tahmin Edilen değer	Pozitif (1)	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif (0)	Yanlış Negatif(YN)	Doğru Negatif (DN)

Doğruluk/Doğru Tahmin oranı (Accuracy) $=\frac{DP+DN}{DP+DN+YP+YN}$ (2.29)

Sensitivite/Duyarlılık (Sensitivity) $=\frac{DP}{DP+YN}$ (2.30)

Spesifite/Özgünlük (Spesifity) $=\frac{DN}{DN+YP}$ (2.31)

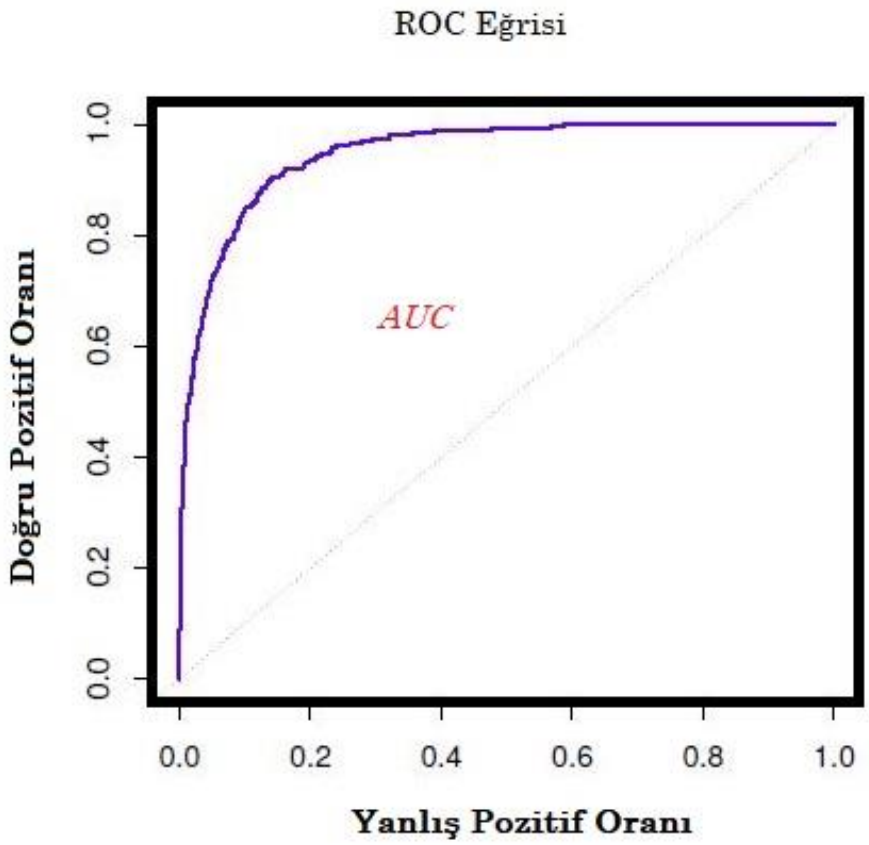
Yanlış Sınıflandırma oranı (Misclassification) $=\frac{YN+YP}{DP+DN+YP+YN}$ (2.32)

Pozitif Tahmin Oranı $=\frac{DP}{DP+YP}$ (2.33)

Negatif Tahmin Oranı $=\frac{DN}{DN+YN}$ (2.34)

Prevalans $=\frac{DP+YN}{Toplam}$ (2.35)

Bir gözlem için niteliksel bir cevap tahmin etmek, gözlemi bir kategoriye veya sınıfa atamayı içerdiğinden, bu gözlemin sınıflandırılması olarak adlandırılmaktadır. Öte yandan, çoğu zaman sınıflandırma için kullanılan yöntemler regresyon yöntemlerinde olduğu gibi, öncelikle nitel değişken sınıflarının her birinin olasılığını tahmin ve benzer şekilde veri setinin eğitim (eğitim + validasyon) ve test setine ayrılması aşamalarını içermektedir. Sınıflandırma problemlerinde model kurulurken genellikle En Çok Olabilirlik (EÇO) veya En Küçük Kareler (LS) metotları tercih edilmektedir. Eğitim seti ile β_0 ve β_1 katsayılar benzerlik fonksiyonunu en yüksek yapacak şekilde tahmin edilmekte ardından olasılık hesaplanmaktadır. Olasılık hesaplanırken bir bağımlı değişkenin hangi sınırı geçtiği zaman hangi sınıfa atanacağı yani hata oranlarının eşik değerinin belirlenmesi gerekmektedir. İkili sınıflandırma problemlerinde yanlış sınıflandırma oranının görselleştirilmesi ve olası tüm eşikler için iki tür hatayı aynı anda görüntülemek için doğru pozitif oranı ile yanlış pozitif oranının grafiği kullanılmaktadır. Bu grafik "Karar Vericinin Etkinliği" grafiği olarak isimlendirilmektedir (Receiver Operating Characteristic- ROC). Bu grafikten elde edilen eğrinin (ROC) altında kalan alan (AUC) tüm olası eşiklerle özetlenen sınıflandırıcının performansını vermektedir (Şekil 2.15) (James ve ark., 2013; Wehrens, 2011).



Şekil 2.15: Karar Vericinin Etkinliği grafiği (James ve ark., 2013).

Bu grafiğe bakılarak en düşük yanlış pozitif oranı ile en yüksek doğru pozitif oranının kesiştiği nokta modelin performansının en iyi olduğu nokta olarak kabul edilmektedir. Model seçiminde genellikle, hata tahmini gibi kriterlere dayanarak, bir dizi olasılık arasından en iyi modelin seçilmesi amaçlanmaktadır. Fakat model seçimi sırasında hata tahminlerinin kendilerinden öte hataların boyutları ile ilgilenildiğinden modelin en küçük hatayla seçilmesi istenmektedir (Wehrens, 2011).

Yapılan literatür incelemesinde NIRS ile ağaç türü teşhisinde yukarıda bahsedilen performans ölçütlerinden bazıları üzerinde yoğunlaşıldığı görülmektedir. NIRS yönteminin ağaç türü teşhisinde kullanımına dair gerçekleştirilen çalışmalarda kullanılan performans ölçüm metotları Tablo 2.6’ de özetlenmiştir.

Tablo 2.6: NIRS ile ağaç türü teşhisi konusunda gerçekleştirilen çalışmalarda geçmişten günümüze kullanılan performans değerleri.

Yayın Yılı	Yazar(lar)	Model Performans Değerleri				
		SEP	**RMSE**	**PRESS**	**R^2**	**Doğru TO**
1996	Brunner ve ark.					
2003	Tsuchikawa ve ark.					X
2004	Gierlinger ve ark.				X	
2006	Flæte ve ark.					X
2008	Haartveit ve Flæte					X
2008	Adedipe ve ark.					X
2009	Russ ve ark.				X	X
2009	Tounis, E.	X	X		X	
2011	Braga ve ark.		X			
2011	Pastore ve ark.		X			
2011	Russ ve Fiserova					X
2012	Dawson-Andoh ve Adedipe					X
2012	Espinoza ve ark.					X
2014	Shou ve ark.			X		X
2014	Zhang ve ark.					
2015	Horikawa ve ark.		X			X
2015	Lang ve ark.					X
2015	Yang ve ark.	X			X	X
2016	Bergo ve ark.					X
2016	Bolzon de Muñiz ve ark.					X
2016	Hwang ve ark.		X		X	
2016	Nisgoski ve ark.					
2017	Lang ve ark.					X
2017	Lazarescu ve ark.				X	X
2017a	Nisgoski ve ark.					X
2017b	Nisgoski ve ark.					
2018	Nisgoski ve ark.					X
2018	Snel ve ark.		X			X
2019	Leandro ve ark.	X				X
2019	Pace ve ark.		X		X	X
2019	Li ve ark.					X
2019	Vieira ve ark.					
2019	Yong ve ark.					X
2020	Zhou ve ark.					X

*TO: Tahmin oranı, PRESS: Tahmini artık hata kareleri toplamı, RMSE: Kök hata kareler ortalaması, SEP: Standart tahmin hatası.

2.4.4. Yorumlama

NIRS'da olduğu gibi yüksek boyutlu veri setleri genellikle çok sayıda değişkene ve değişkenden daha az sayıda örneğe sahip olma eğilimindedirler. Bu durum bağımlı değişkenlerin tahminini zorlaştırmaktadır. Değişken seçimleri, PCA veya PLS gibi uygun projeksiyon veya çeşitli seçim yöntemleri aracılığıyla gerçekleştirilebilmektedir (Balabin ve Smirnov 2011; Farres ve ark., 2015; Mehmood ve ark., 2012; Mehmood ve ark., 2020; Xiaobo ve ark., 2010).

Değişken veya özellik seçimi spektroskopik verilere uygulandığında frekans veya dalga boyu seçimi olarak da isimlendirilmektedir. Değişken seçim yöntemleri, kalibrasyon prosedürü sırasında verilerin kalitesinin etkileşimli olarak iyileştirilmesini sağladığı için veri analizinde kritik bir adım olarak kabul edilmektedir (Balabin ve Smirnov 2011, Xiaobo ve ark. 2010). Değişken seçim yöntemleri, orijinal değişkenlerin arasından daha az sayıda değişkeni seçme ilkesine dayanarak, belirli bir yanıt değişkeni ile ilişkili daha küçük bir ilgili tahmin değişkenleri kümesinin seçilmesine (boyut indirgenmesine), yardımcı olmaktadır. Bu sayede çok değişkenli analizlerde değişken seçimi, modelin tahmin yeteneğinin arttırılması ve model karmaşıklığının azaltılmasına olanak sağlayacağından daha kolay yorumlama yapma olanağı sağlamaktadırlar (Farres ve ark., 2015; Mehmood ve ark., 2012; Mehmood ve ark., 2020; Xiaobo ve ark., 2010).

PLS, yüksek derecede çoklu bağlantıya sahip çok sayıda tahmin değişkeni ile başa çıkabilme yeteneğine sahiptir. Ancak, orijinal değişkenler yerine latent değişkenlerin kullanılması da model yorumlamasında zorluk yaratır. Bu karmaşıklık, PLS'nin oluşturduğu latent değişkenlerinin sadece X'in y cevabı ile korelasyonunu maksimize etmekle kalmayıp aynı zamanda X'te açıklanan varyansı maksimize etmeye çalışmasından kaynaklanmaktadır. Bu nedenle, ağırlık ve yük gibi model parametrelerini kullanarak modeli doğrudan yorumlayamayız. Bu özellikle birden çok varyasyon kaynağından etkilenen analitik verilerde geçerlidir. Bu durum PLS model parametrelerinin ve bu parametrelere dayalı değişken önemin yorumlanmasını zorlaştırmaktadır (Tran ve ark., 2014) Orijinal formundaki PLS'nin değişken seçimin doğrudan bir uygulaması yoktur, ancak PLS'de değişken seçimini ele alan çok çeşitli yöntemler önerilmektedir. Genel olarak PLS'de değişken seçim yöntemleri filtre yöntemleri, sarma yöntemleri ve gömülü yöntemler olmak üzere üç ana kategoriye ayrılmaktadır (Şekil 2.16) (Mehmood ve ark. 2012; Mehmood ve ark., 2020).



Şekil 2.16: Değişken seçim yöntemlerinin sınıflandırması (Mehmood ve ark., 2012).

Filtre yöntemleri değişkenleri iki adımda seçmektedir. Öncelikle PLS modeli verilere uygulanmakta ardından PLS modelinden elde edilen bazı ilgili ölçülere bir eşik verilerek değişken seçimi gerçekleştirilmektedir. Bu yöntemlere dair başlıca örnekler; yükleme ağırlıkları (loading weights), regresyon katsayıları (Regression coefficients), projeksiyonda değişken önemi (variable importance on projection-VIP), Seçicilik oranı (Selectivity Ratio-SR), Çok değişkenli korelasyon anlamlılığı (Significance multivariate correlation – sMC).

Sarma yöntemleri çoğunlukla filtre yöntemlerini yinelemeli bir şekilde kullanır. Bu yöntemler modelin denetimli öğrenme yaklaşımı ile değişken seçim algoritması içerisinde sarılarak tekrardan kalibre edilmesine dayanmaktadır. Bu yöntemlere dair başlıca örnekler; Genetik algoritması (Genetic Algorithm), bilgi vermeyen değişken eleme (uninformative variable elimination), alt pencere permutasyon analizi (sub-window permutation analysis), yinelemeli tahmin ağırlığı (Iterative predictor weighting), geriye yönelik değişken elemesi (backward variable elimination), düzenlenmiş değişken eleme prosedürü (regularized variable elimination procedure), Hotelling T^2 değerine dayalı değişken seçimi (Hotelling T^2 based variable selection).

Gömülü yöntemlerde değişken seçimi ve modelleme tek bir adımda ve bileşenler seviyesinde gerçekleşmektedir. Optimal değişken alt kümesinin araştırılması, PLS modelinin her bileşeni üzerinde tek bir yinelemeli prosedür ile gerçekleştirilmektedir. Bu yöntemlere dair başlıca örnekler; etkileşimli değişken seçimi (Interactive variable selection), yumuşak eşik (Soft-Threshold), değişken seçim için dağıtım tabanlı kesme (Distribution-based truncation for

variable selection), ağırlıklandırılmış değişken seçim katkısı (Weighted variable contribution) (Mehmood ve ark., 2012; Mehmood ve ark., 2020; Tran ve ark., 2014).

Yukarıda da gösterildiği gibi, değişken seçiminde pek çok farklı metot bulunsa da, en çok kullanılan değişken seçim yöntemleri VIP ve SR değerleridir. VIP yöntemi, X değişkenlerindeki temel varyasyona en çok katkıda bulunan X değişkenlerini seçmektedir. Bu seçim, y ile ilişkili olmayan, ortogonal varyasyon olarak adlandırılan etkileşimleri tanımlayan varyasyonu içermektedir. VIP skorları y varyansını en iyi açıklayan X tahmin değişkenlerini anlamada yararlı olmaktadır (Farres ve ark., 2015).

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 .\ SSY_f\ .\ J}{SSY_{toplam}\ F}} \tag{2.36}$$

Burada w_{jf} j değişkeni ve f bileşeninin ağırlık değerini, SSY_f f. bileşenin açıklanan varyansının karelerinin toplamını, J X değişkenlerinin sayısını, SSY_{toplam}, açıklanan bağımlı değişkenin kareler toplamını, F toplam bileşen sayısını ifade etmektedir. Tek boyutlu Y düzleminde,

$$SSY_f = b_f^2\ t_f'\ t_f \tag{2.37}$$

$$SSY_{toplam} = b^2\ T'\ T \tag{2.38}$$

Burada T X skor matrisini, b katsayıların PLS iç ilişki vektörünü ifade etmektedir. Ortalama VIP skorlarının karesi 1'e eşit olduğundan, genellikle değişken seçim ölçütü olarak 1'den büyük kuralı uygulanmaktadır (Farres ve ark., 2015; Mehmood ve ark., 2020; Tran ve ark., 2014).

Çok değişkenli veri analizinde değişken seçim için bir araç olarak önerilen bir diğer popüler yöntem ise Seçicilik oranıdır (SR). Bu yöntem ile y değişkeninin açıklanmasına ilişkin X değişken sıralaması için kullanılmakta ve özellikle tahmin için yararlı olmaktadır. SR değişkenleri, açıklanan y değişkeninin X bileşeninin kalıntı varyansına oranının hesaplanmasıyla elde edilmektedir (Farres ve ark., 2015).

$$SR_i = \frac{SS_{i,açıklanan}}{SS_{i,artık}} \tag{2.39}$$

Burada SS i, açıklanan açıklanan varyans oranını, SSi,artık her değişkenin artık varsyansını kareler toplamı olarak ifade etmektedir.

$$SS_{i,açıklanan} = \|t_{TP}\, p'_{TPi}\|^2 \tag{2.40}$$

$$SS_{i,artık} = \|e_{TPi}\|^2 \tag{2.41}$$

Burada t_{TP} X üzerinde yansıtılan skor vektörünü, p_{TP} X üzerinde yansıtılan yüklemeyi ifade etmektedir.

$$t_{TP} = \frac{X\, b_{PLS}}{\|b_{PLS}\|} \tag{2.42}$$

$$p_{TP} = \frac{X'\, t_{TP}}{(t'_{TP}\, t_{TP})} \tag{2.43}$$

Değişken önemi belirlemek için farklı yaklaşımlar mevcuttur. Bunlardan biri kritik eşik değerinin SR_i'nin n - 2 ve n - 3 serbestlik dereceli F dağılımına göre değerlendirilmesi önerilmektedir (Farres ve ark. 2015; Rajalahti ve ark. 2009;Tran ve ark., 2014). Bir diğeri ise, SR değerlerine göre değişkenleri yüksekten düşüğe sıralayarak kullanıcı tarafından amaca uygun değişkenlerin ileriye veya geriye doğru seçimini gerçekleştirmektir (Kvalheim, 2020).

3. MALZEME VE YÖNTEM

Yakın Kızılötesi Spektroskopisinin Odun Teşhisinde Kullanımı isimli bu Doktora çalışmasında dört farklı ağaç türü, üç farklı örnek alanı ve iki farklı örnek toplama yöntemi kullanılmıştır.

3.1.MALZEME

Çalışmada, iğne yapraklı ağaç türlerini temsilen sarıçam (*Pinus sylvestris* L.) ve karaçam (*Pinus nigra*J.F. Arnold); geniş yapraklı ağaç türlerini temsilen Macar meşesi (*Quercus frainetto* Ten.) ve sapsız meşe (*Quercus petraea* (Mattuschka) Liebl.) ağaç türleri kullanılmıştır. Örnek ağaçlar İstanbul Orman Bölge Müdürlüğüne bağlı Bahçeköy Orman İşletme Müdürlüğü ve Kastamonu Orman Bölge Müdürlüğüne bağlı Ayancık Orman İşletme Müdürlüğünün idaresi altındaki ormanlık alanlardan alınmıştır. Örneklerin toplanması sırasında ağaç türlerinin tespiti dış morfolojik özellikleri temel alınarak gerçekleştirilmiştir (Tablo 3.1, Tablo 3.2) (Yaltırık ve Efe, 2000).

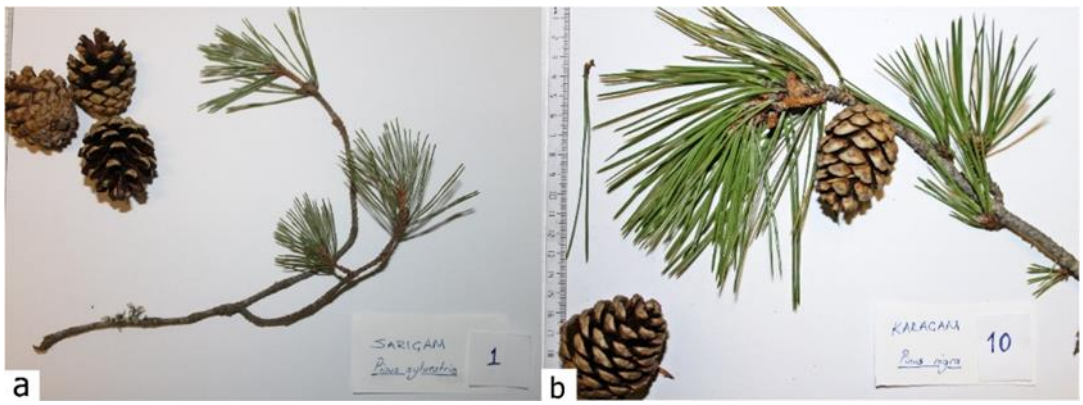
Tablo 3.1: Çam türlerinin belirlenmesinde kullanılan dış morfolojik özellikler.

	Sarıçam	***Karaçam***
Kozalak	En küçük boyutlu 3-7 cm	5-10 cm
	Asimetrik	Simetrik
	Saplı ve aşağı sarkık	Çok kısa saplı, sapsız dik ve yükselici
	Işık gören kısımları gelişkin	Konik, yumurta biçimli
	Boz mat renkte Karpeller çıkıntısız	Saman sarısı renkte Karpellerin ortası iğne şeklinde çıkıntılı
Tomurcuk	Kırmızımsı	Uçları sivri
	Yumurta biçimli	Silindirimsi biçimli
	Çoğunlukla reçinesiz	Büyük ve reçineli
Yaprak (2li)	3-8cm kısa	8-15cm uzun
	mavimsi (donuk) yeşil	koyu yeşil
	uçları sivri, batıcı	tomurcuğa doğru yönelmiş
	kenarları dişli	çanak gibi boşluk oluşturmuş
	180º kıvrık	
Sürgün	kırmızımsı	siyahımsı
Kabuk	Tilki sarısı renkte	Grimsi yeşilimsi siyah
	ince levhalar halinde ayrılır	Çatlaklı
	yaşlılarda çatlaklı	
Tepe Tacı	Sivri, piramidal tepeli	Gençlerde konik,
	yaşlılarda yuvarlaklaşır	yaşlılarda geniş ve dağınık

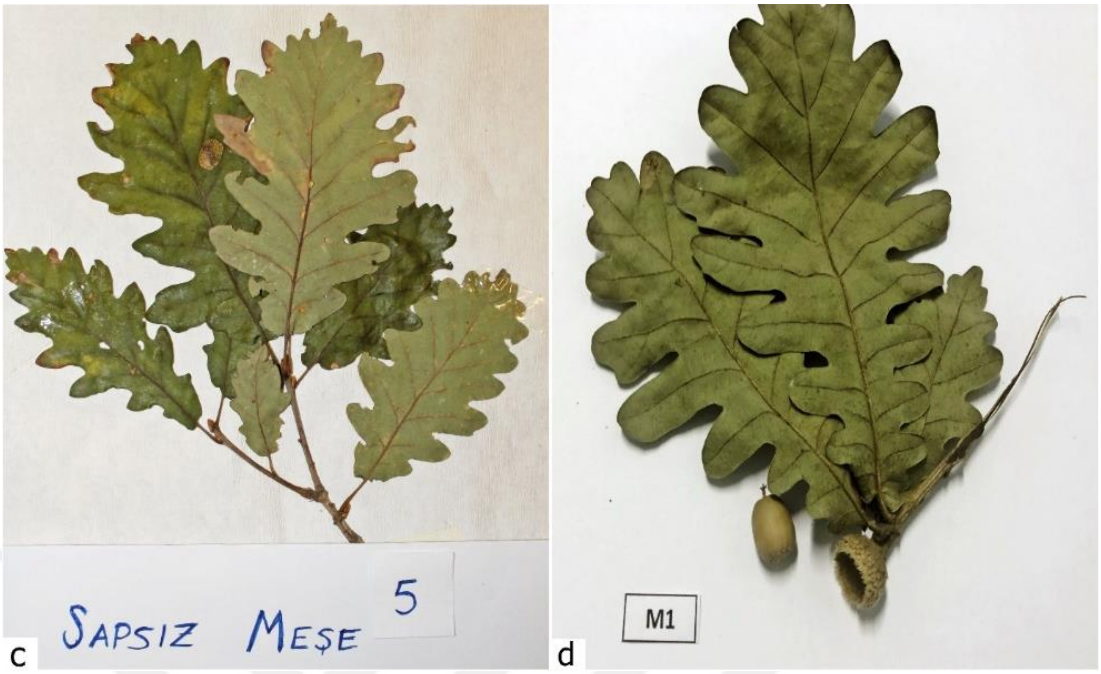
Tablo 3.2: Meşe türlerinin belirlenmesinde kullanılan dış morfolojik özellikler.

	Macar Meşesi	***Sapsız Meşe***
Yaprak	18-20 cm uzunluk 13 cm genişlik	5-17 cm uzunluk 3-9 cm genişlik
	sapsız	saplı 1-3,5 cm
	kenarlar düzenli loplu	kenarlar düzenli sığ loplu
	loplar derin ve dar oyuntulu	ters yumurta, eliptik, dikdörtgensi
	üst yüzü yıldız tüylerle kaplı, bazen çıplak	üst yüzü çıplak, parlak
	alt yüzü yıldız tüylerle kaplı	alt yüzü çıplak , soluk
	dipte iki kulakçık sapı örter	dipte kama gibi daralır
	tam kenarlı lop 6-9 adet	tam kenarlı lop 5-9 adet
	yan damar sayısı 1-3 adet	yan damar sayısı 5-11
	Ters yumurta şeklinde (20x13 cm)	
Meyve	2,5 cm, sapsız	sapsız
	Kadeh 1,5 cm çapında, pulları kısa, boz tüylü, gevşek	kadeh palamutun en az yarısını içine alır
Palamut	uzun yumurta şeklinde	yarım küre veya armut şeklinde
	2-3 cm	1-2 cm

Dış morfolojik özelliklerine göre belirlenen ağaç türlerinden seçilen örneklerin her biri fotoğraflanarak kayıt altına alınmış, yaprak, sürgün, meyve veya kozalak örnekleri toplanmıştır (Şekil 3.1).



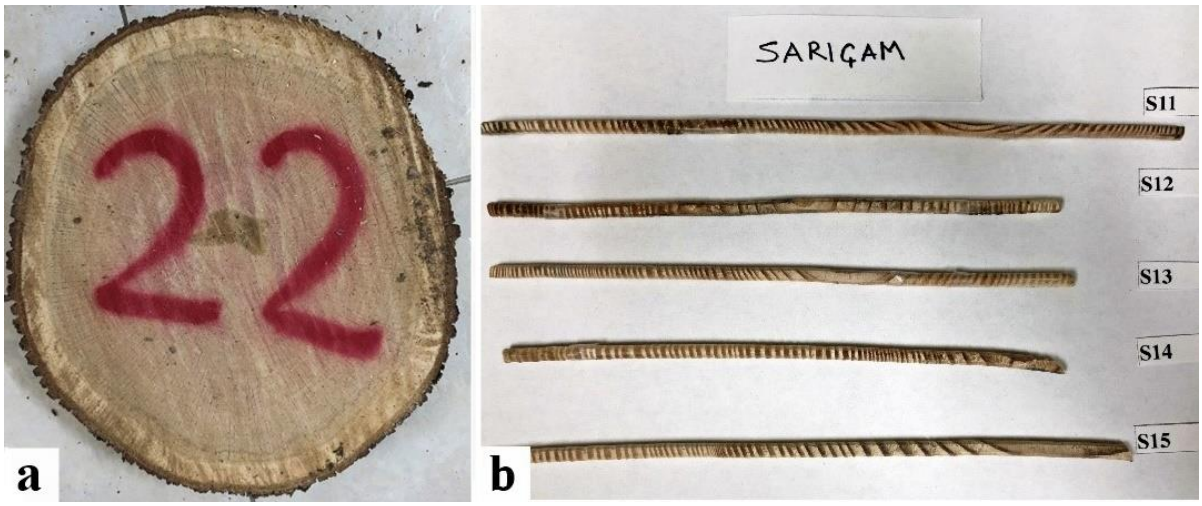
Şekil 3.1: Sarıçam (a), karaçam (b), sapsız meşe (c) ve Macar meşesi (d) ağaç türlerine ait sürgün fotoğrafı.



Şekil 3.1 (devam): Sarıçam (a), karaçam (b), sapsız meşe (c) ve Macar meşesi (d) ağaç türlerine ait sürgün fotoğrafı.

Çalışmada kullanılan örnekler iki farklı tip örnekleme metodu ile ağaçların göğüs yüksekliğinden (1,30 m) temin edilmiştir (Şekil 3.2). Bu metotlar:

- deneme ağaçlarının kesilmesi sonrası 1,30 m 'den disk formunda örneklerin alınması,
- artım burgusu yardımıyla 1,30 m 'den öze kadar giden artım kalemlerinin alınmasıdır.



Şekil 3.2: Disk örnekleri (a) ve artım burgusu aracılığıyla temin edilen artım kalemleri (b).

Gerçekleştirilen bu çalışmada iki farklı örnekleme tipinin seçilmesine yönelten farklı kısıtlar ve yaklaşımlar bulunmaktadır. Bu yaklaşımlar genel olarak şu şekildedir:

- Disk halinde örnek alımı boyutsal büyüklük sebebiyle rahat çalışma imkanı sağladığı için tercih edilmektedir.
- Planlanan bu çalışmanın içeriği gereği farklı birey sayısının arttırılması başarılı sonuçların alınması için gereklidir.
- Göğüs yüksekliğinden alınan disklerin elde edilebilecek yüksek değerli tomruklarda değer düşüşüne sebep olması tüm örneklerin disk ile temininde sıkıntılar yaratmaktadır.
- Çalışma için alınan kesim izinleri sırasında da ağaç sayısında kısıtlamaya gidilmiştir.
- Aynı ağaçtan ve nispeten aynı yükseklikten alınan disk örnekleri ve artım kalemi örneklerinin NIRS verileri arasında anlamlı bir farklılık bulunmamıştır.

Yukarıda sıralanan sebepler ve kısıtlar ışığında örneklemenin mümkün olan en çok sayıda disk örneği ile gerçekleştirilmesi ve bunun mümkün olmadığı hallerde ise ağaçlar dikili haldeyken çalışma imkanı sağlayan artım burgusu ile örnek sayısının arttırılması yoluna gidilmiştir.

NIRS ağaç malzemenin kimyasal yapısındaki değişimleri ve ilişkileri incelemektedir. Farklı örnekleme tiplerinin, aynı yükseklikten elde edilen örneklerin kimyasal yapıları üzerinde bir etkisi olmaması sebebiyle bahsi geçen iki farklı örnekleme tipinin kullanılması çalışmanın amacına uygun bir yaklaşım olarak kabul edilmiştir (Schweingruber, 2007; Zobel ve Jett, 1995).

Çalışmada kullanılan ağaç türlerine dair örnek alımı gerçekleştirilmeden önce ve sonra kimlik verileri (ağaç çapları ve koordinatlar) kayıt altına alınmıştır. Çap ölçümü kumpas yardımıyla, göğüs yüksekliğinden (1.30 m) iki farklı yönde alınan ölçümlerin ortalanması ile belirlenmiştir. Çalışmada kullanılan ağaçlara ait kimlik bilgileri ağaç türleri göz önüne alınarak Tablo 3.3, Tablo 3.4, Tablo 3.5 ve Tablo 3.6’ da verilmiştir.

Tablo 3.3: Sarıçam ağaç türüne ait kimlik verileri.

Ağaç türü	Örnek temini	İşletme Şefliği	Ağaç No	1.30 m Çap (cm)	Ağaç Yaşı	Koordinatlar	
Sarıçam	Disk Örnekleri	(Sinop-Ayancık: Kepez Orman İşletme Şefliği)	S1	62	114	34.532131	41.766068
			S2	56	138	34.532128	41.755062
			S3	54	142	34.532144	41.766152
			S4	48	129	34.532155	41.766140
			S5	46	121	34.532188	41.766112
			S6	50	137	34.532188	41.766066
			S7	62	139	34.532165	41.766013
			S8	54	126	34.532170	41.765995
			S9	54	132	34.532165	41.765918
			S10	36	138	34.532148	41.765977
	Artım Burgusu Örnekleri	(İstanbul-Bahçeköy: Bentler Orman İşletme Şefliği)	S11	40	63	28 9882 156	41.1755 548
			S12	38	58	28.9883 700	41.1753 410
			S13	32	61	28.9884 696	41.1753 757
			S14	44	70	28.9883 973	41.1756 475
			S15	44	62	28.9884 778	41.1757 5
			S16	34	65	28.9887 654	41.1754 801
			S17	36	66	28.9889 469	41.1759 666
			S18	36	57	28.9874 645	41.1785 613
			S19	31	63	28.9873 406	41.1784 888
			S20	38	64	28.9872 852	41.1784 201
			S21	40	60	28.9871 709	41.1784 79
			S22	40	59	28.9871 635	41.1767 316
			S23	36	66	28.9876 135	41.1767 215
			S24	42	66	28.9878 576	41.1768 542
			S25	38	60	28.9879 091	41.1769 735

Tablo 3.4: Karaçam ağaç türüne ait örneklerin kimlik verileri.

Ağaç türü	Örnek temini	İşletme Şefliği	Ağaç No	1.30 m Çap (cm)	Ağaç Yaşı	Koordinatlar	
Karaçam	Disk Örnekleri	Sinop-Ayancık: Kepez Orman İşletme Şefliği	K1	60	109	34 532100	41.773382
			K2	66	117	34532124	41.773365
			K3	64	127	34 532131	41.773301
			K4	46	122	34 532189	41.773195
			K5	50	138	34532204	41.773114
			K6	58	142	34 532146	41.773167
			K7	66	147	34 532130	41.773148
			K8	74	134	34 532148	41.773137
			K9	64	155	34 532165	41.773199
			K10	64	162	34 532141	41.773194
	Artım Burgusu Örnekleri	İstanbul-Bahçeköy: Bentler Orman İşletme Şefliği	K11	42	68	28 9884 797	41.1755 729
			K12	38	72	28 9869 191	41.1756 155
			K13	58	67	28 9869 014	41.1783 159
			K14	58	58	28 9869 014	41.1782 833
			K15	54	53	28 9872 683	41.1783 291
			K16	42	57	28 9871 077	41.1783 550
			K17	32	49	28 9871 559	41.1784 00
			K18	44	55	28 9876 897	41.1773 978
			K19	40	62	28 9876 005	41.1773 504
			K20	38	62	28 9872 864	41.1786 318
			K21	36	60	28 9873 745	41.1786 303
			K22	40	64	28 9873 149	41.1789 665
			K23	38	57	28 9871 626	41.1789 387
			K24	40	61	28 9873 103	41.1789 337
			K25	42	64	28 9875 773	41.1788 513

Tablo 3.5: Sapsız meşe ağaç türüne ait kimlik verileri.

Ağaç türü	Örnek temini	İşletme Şefliği	Ağaç No	1.30 m Çap (cm)	Ağaç Yaşı	Koordinatlar	
Sapsız Meşe	Artım Burgusu Örnekleri	İstanbul-Bahçeköy: Bentler Orman İşletme Şefliği	P1	64	88	28.9399610	41.1992716
			P2	67	110	28.9398435	41.1992415
			P3	65	125	28.9397452	41.1992201
			P4	57	106	28.9391691	41.1989986
			P5	56	114	28.9391089	41.1992127
	Disk Örnekleri		P6	57	175	28.9398245	41.1974521
			P7	64	180	28.9401498	41.1976865
			P8	53	143	28.9402107	41.1985102
			P9	63	183	28.9408146	41.1983511
			P10	37	83	28.9411857	41.1984981
			P11	52	168	28.9412731	41.1984922
			P12	52	182	28.9420937	41.1984263
			P13	45	180	28.9434211	41.1980898
			P14	60	178	28.9436411	41.1981812
			P15	52	179	28.9438206	41.1983834
			P16	50	181	28.9433083	41.1984842
			P17	54	182	28.9433024	41.1984918
			P18	60	189	28.9432923	41.1990159

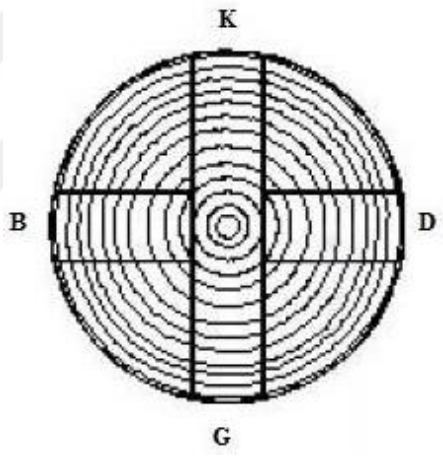
Tablo 3.6: Macar meşesi ağaç türüne ait kimlik verileri.

Ağaç türü	Örnek temini	İşletme Şefliği	Ağaç No	1.30 m Çap (cm)	Ağaç Yaşı	Koordinatlar	
Macar Meşesi	Disk Örnekleri	İstanbul-Sarıyer: Sarıyer Orman İşletme Şefliği	M1	38	90	28 889.729	41 232.466
			M2	56	217	28 889.960	41 232.218
			M3	45	38	28 890.778	41 233.805
			M4	53	227	28 889.776	41 231.998
			M5	47	209	28 889.781	41 232.073
			M6	70	216	28 889.799	41 232.604
			M7	63	221	28 891.251	41 233.809
			M8	58	217	28 891.479	41 233.837
			M9	43	216	28 891.479	41 233.860
			M10	38	215	28 892.038	41 233.969
			M11	48	219	28 891.698	41 234.465
			M12	32	69	28 892.655	41 234.315
			M13	32	68	28 892.889	41 234.442
			M14	52	226	28 893.116	41 234.410
			M15	40	209	29 005 745	41 178 507
			M16	46	37	29 005 809	41 176819
			M17	42	35	29 005 691	41 177328
			M18	44	35	29 006 775	41 179 524
			M19	34	35	29 007 022	41 180 857

3.2.YÖNTEM

3.2.1. Örneklerin Hazırlanması

Ormandan yaş halde temin edilen deneme örnekleri İstanbul Üniversitesi-Cerrahpaşa, Orman Fakültesi, Odun Mekaniği ve Teknolojisi Ana Bilim Dalı bünyesinde bulunan iklimlendirme kabinine konulmuştur. İklimlendirme kabinine yerleştirilen örnekler 20 ± 2 ºC sıcaklık ve % 65 ± 5 bağıl nemde değişmez ağırlığa ulaşıncaya kadar klimatize edilmiştir. Kabin içerisinde % 12 hava kurusu rutubete getirilen diskler İstanbul Üniversitesi-Cerrahpaşa, Orman Fakültesi Ahşap Atölyesi'nde boyutlandırılmıştır (Şekil 3.3). Boyutlandırılan örneklerin yüzeyleri, yıllık halka genişliklerinin ölçümü için sırası ile 80 ve 180 kumluk zımpara kağıtlarıyla zımparalanarak düzeltilmiştir.



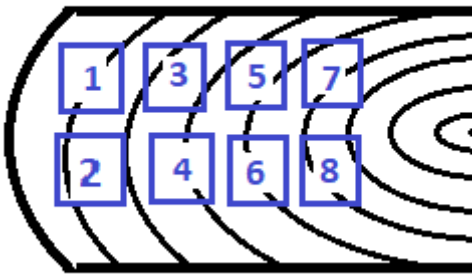
Şekil 3.3: Disk örneklerinin boyutlandırılması.

Kimliklendirme işlemlerinin bir parçası olarak örneklerin yaşları, yaz odunu ve ilkbahar odunu tabakalarının genişlikleri ile yıllık halka genişlikleri ölçülmüştür. Bu amaçla yüzeyi düzeltilen örneklerden özden kabuğa kadar doğrusal bir hat içeren şeritler çıkarılmıştır. Ölçümler bu şeritler üzerinde 0,01 mm hassasiyette Eclund Brinell Mikroskobu yardımıyla gerçekleştirilmiştir (Şekil 3.4).



Şekil 3.4: Eclund Brinell Mikroskobu ve artım burgusu örneği.

Ağaç malzemenin sahip olduğu kimyasal özellikler özden kabuğa doğru (radyal yönde) gidildikçe değişiklik gösterdiğinden (Schweingruber, 2007; Zobel ve Buijtenen, 1989), disk örnekleri üzerinde radyal yönde birden fazla ölçüm alanı belirlenmiş, artım kalemi örneklerinde ise radyal yönde tüm alan taranmıştır (Şekil 3. 5).



Şekil 3.5: Enine kesit boyunca ölçüm alanlarının belirlenme şablonu.

Ağaç çaplarına bağlı olarak örnek boyutlarındaki değişimler sebebiyle, belirlenen ölçüm alanlarının sayısı 10 ila 20 adet arasında farklılık göstermiştir. Spektral ölçümlerin alınacağı bu alanlar, cihazın ölçüm alanı boyutu olan 1 cm çapında daireye uygun olacak şekilde işaretlenmiştir (Şekil 3.6, Şekil 3.7).



Şekil 3.6: NIRS ölçümleri için disk örneklerinin hazırlanması ve örnek alanların belirlenmesi.



Şekil 3.7: Artım kalemlerinin ölçümlere hazırlanması.

Yakın kızılötesi (NIRS) ölçümleri için belirlenen (1 cm çapında) alanların içerdiği yıllık halka sayıları ve bu alanların sahip olduğu diğer özellikler kimlik bilgilerine eklenmiştir (Tablo 3.7). Özden 2-3 yıllık halka sayıldıktan sonraki 1 cm alanlar, yıllık halka genişlikleri de göz önünde bulundurularak, genç odun olarak kimliklendirilmiş; geçiş zonundan kaçınmak için özden 25 yıllık halka sonrasında alınan alanlar ergin odun olarak isimlendirilmiştir (Rowell, 2005; Zobel ve Van Buijtenen, 1989).

Tablo 3.7: NIRS ölçümleri için belirlenen ölçüm alanlarının kimliklendirilmesi.

Örnek kodu	Ağaç Türü	Ağaç kodu	Ergin odun (M) / Genç odun (J)	Öz odun (H)/ Diri Odun (S)/ Geçiş (T)	Yıllık Halka Sayıları
S8-1	*P.sylvestris*	S8	M	S	14
S8-2	*P.sylvestris*	S8	M	S	14
S8-3	*P.sylvestris*	S8	M	S	7
S8-4	*P.sylvestris*	S8	M	S	7
S8-5	*P.sylvestris*	S8	M	S	6
S8-6	*P.sylvestris*	S8	M	S	6
S8-7	*P.sylvestris*	S8	M	T	6
S8-8	*P.sylvestris*	S8	M	T	6
S8-9	*P.sylvestris*	S8	M	H	5
S8-10	*P.sylvestris*	S8	M	H	5
S8-11	*P.sylvestris*	S8	M	H	5
S8-12	*P.sylvestris*	S8	M	H	5
S8-13	*P.sylvestris*	S8	M	H	3
S8-14	*P.sylvestris*	S8	M	H	3
S8-15	*P.sylvestris*	S8	J	H	4
S8-16	*P.sylvestris*	S8	J	H	4

Örnekler ölçümler için hazırlandıktan sonra spektroskopik veriler Amerika Birleşik Devletleri, Wisconsin Üniversitesi-Madison, Botanik Bölümü Laboratuvarı ve Birleşik Devletler Orman Teşkilatı, Orman Ürünleri Laboratuvarı, Odun Anatomisi Araştırmaları Merkezi'nde bulunan Thermo Nicolet Scientific marka Antaris FT-NIR analyzer cihazında (Şekil 3.8) aşağıdaki ayarlar ile EK-1'de gösterilen formatta toplanmıştır:

- Örnekleme Tekniği: Integrating sphere
- Arka Plan Tarama Sayısı: Örnek ile aynı
- Çözünürlük: 4.0 cm^{-1}
- Spektral Aralık: 4 000 – 12 000 cm^{-1} (Yakın Kızılötesi)
- Örnek Tarama Sayısı: 64
- Veri biçimi: log 1/R
- Arka Plan Sıklığı: Her saat



Şekil 3.8: Antaris FT-NIR Analyser cihazında Yakın Kızılötesi ölçümlerinin alınması.

Tüm bu ön hazırlık aşamalarının ardından gerçekleştirilen çalışmada kullanılan ağaç ve örnek sayıları, temin metoduna göre Tablo 3.8 ’de verilmiştir.

Tablo 3.8: Çalışmada kullanılan ağaç ve örnek sayıları.

	Ağaç			**Örnek**
	Disk	**Artım burgusu**	**Toplam**	**Spektrum**
P. sylvestris (S)	10	15	25	382
P. nigra (K)	10	15	25	356
Q. petraea (P)	18	-	18	287
Q. frainetto (M)	18	-	18	322
		Toplam	*86*	*1347*

3.2.2. Kullanılan Program ve Paketler

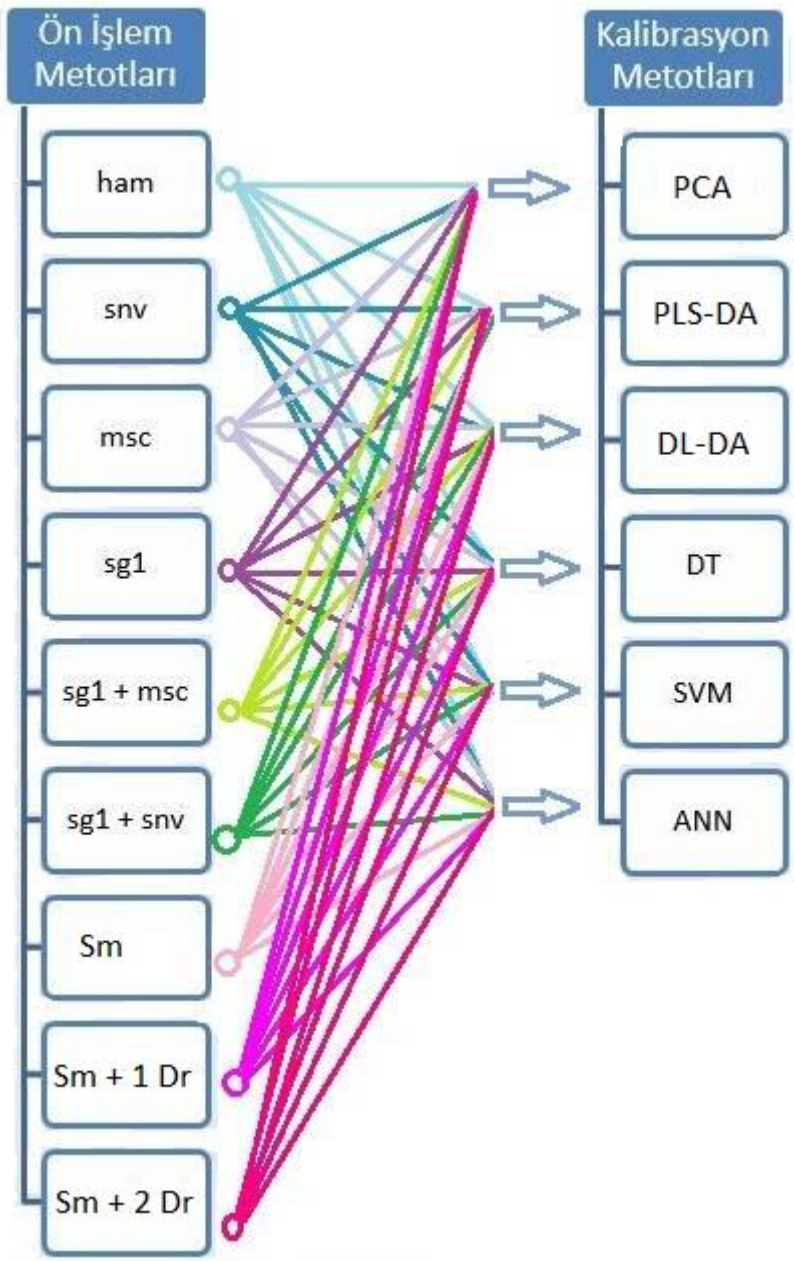
Gerçekleştirilen çalışmada elde edilen verilerin analizleri kodlama ağırlıklı bir program olan R aracılığıyla gerçekleştirilmiştir (R Core Team, 2018). Açık kaynaklı bir program olan R ile gerçekleştirilecek veri analizleri için ilgili analizlerin fonksiyonlarının oluşturulduğu paketlerin kullanımı gerekmektedir. Verilerin düzenlenmesi, programa çağrılması, ön işlemlerin uygulanması, modellerin kurulması, model parametrelerin optimizasyonu ve model performanslarının belirlenmesinde kullanılan paketler şunlardır:

- ChemometricsWithR (Wehrens, 2017)
- prospectr (Stevens ve Ramirez-Lopez, 2013)
- pls (Mevik, Wehrens ve Hovde Liland, 2019)
- signal (SignalDevelopers, 2014)
- mdatools (Kucheryavskiy, 2018)
- e1071 (Meyer ve ark., 2019)
- sfsmisc (Maechler, 2019)
- rpart (Therneau and Atkinson, 2019)
- kohonen (Wehrens ve Kruisselbrink, 2018)
- neuralnet (Fritsch, Guetner ve Wright, 2019)
- caret (Kuhn, 2019)
- kernlab (Karatzoglou ve ark., 2004)
- ggplot2 (Wickham, 2016)

3.2.3. Kullanılan Analizler

NIRS ile gerçekleştirilen analizler ön işlemler, model kurmada kullanılan kalibrasyonlar ve model performansının belirlenmesinde kullanılan parametreler olmak üzere üç ana başlık altında verilmiştir.

Çalışmada kullanılan ön işlem metotları boyut indirgeme amacıyla temel bileşenler analizi (PCA), saçılım düzeltmesi ve gürültünün uzaklaştırılması amacıyla çarpımsal saçılım düzeltmesi (MSC), standart normal değişim (SNV), düzgünleştirme (Sm) ve türev (Dr) alma amacıyla Savitzky-Golay (SG) algoritmaları ve bu ön işlem metotlarının kombinasyonlarıdır (Şekil 3.9).



Şekil 3.9: Çalışmada kullanılan ön işlemler ve modeller*.

*Dr: Türev, Msc: Çarpımsal saçılım düzeltmesi, Sm: Düzgünleştirme, Snv: Standart normal değişim, SG: Savizky-Golay algoritması; ANN: Yapay sinir ağları, DA: Diskriminant analizi, DL: Diyagonal linear, DT: Karar ağaçları, PCA: Temel bileşenler analizi, PLS: Kısmi en küçük kareler, SVM: Destek vektör makineleri.

3.2.3.1. Temel Bileşenler Analizi (PCA)

Gerçekleştirilen çalışmada temel bileşenler analizi (PCA) hem sınıflandırma amaçlı modelleme aşmasında, hem de yapay sinir ağları (ANN) için ön işlem aşaması olarak kullanılmıştır. Çalışmada kullanılan veri setinde gözlemlerden daha çok sayıda bağımsız değişken olması

sebebiyle temel bileşenler (PCs), Tekil Değer Ayrışma (SVD) algoritması aracılığıyla belirlenmiştir. SVD aracılığıyla X matrisi (n*m boyutlarındaki) 3 matrise ayrıştırılabilir (Varmuza ve Filzmoser, 2008);

$$X = T_0 \,.\, S.\, P^T \tag{3.1}$$

Burada, Ortalama merkezli X için T_0 matrisi (n*m) 1 uzunluğunda normalize edilmiş PCA skorlarını içermektedir. S diyagonal matrisi (m*m) skorların standart sapmalarına ($\sqrt{\lambda_j}$) eşit tekil değerleri içermektedir. P^T matrisi (m*m) devrik (transpoze) PCA yüklemelerini içermektedir. PCA skorlarının (T) hesaplanması,

$$T = T_0 \,.\, S \tag{3.2}$$

3.2.3.2. Çarpımsal Saçılım Düzeltmesi (MSC)

Özellikle NIR verilerine uygulanması için geliştirilen bu metotta, her bir spektrum için uygulanan MSC modeli şu şekildedir (Geladi ve Kowalski, 1986; Naes ve ark., 2004);

$$x_{ik} = a_i + b_i \bar{x}_k + e_{ik} \,(i = 1, \dots, N\,; k = 1, \dots, K) \tag{3.3}$$

Burada i örnek numarası, k dalga boyu numarası, a_i arttırıcı etkiyi gösteren katsayı, $b_i i$ örneği için çarpımsal etkiyi gösteren katsayı, e_{ik} modellenemeyen hatayı gösteren ifadelerdir ve ortalama ($\bar{x}_k$);

$$\bar{x}_k = \frac{1}{N}\sum_{i=1}^{N} x_{ik} \tag{3.4}$$

Bu hesaplamada bilinmeyen a_i ve b_i katsayıları en küçük kareler yöntemi ile tahmin edilmekte ve MSC dönüşümünde kullanılmaktadır. Sonuçta spektrumun MSC si her bir değişken değerinin x_{ik} (absorbansın) düzeltilmesiyle gerçekleşmektedir;

$$x_{ik}^{*}(MSC) = \frac{x_i - \widehat{a_i}}{\widehat{b_i}} \tag{3.5}$$

3.2.3.3. Standart Normal Değişim (SNV)

MSC ile oldukça benzer etkilere sahip SNV dönüşümde her bir spektrum ortalanır ve ölçeklendirilir (auto-scaling) (Barnes ve ark., 1989).

$$x_{ik}^{*}(SNV) = \frac{x_{ik} - \bar{x}_i}{s_i} \tag{3.6}$$

Burada *i.* örneğin *k.* dalga boyundaki spektral ölçümü, x_{ik}; *i* örneğinin *k* spektral ölçümündeki ortalaması, $\bar{x}_i$; standart sapması s_i olarak gösterilmiştir. Bu dönüşüm işlemi sonunda her bir spektrum için dönüştürülmüş absorbans değerlerinin ortalaması 0 standart sapması 1 olmaktadır.

3.2.3.4. Savitzky-Golay Algoritması ile Düzgünleştirme ve Türev Alma

Matematiksel olarak, x_j değeri olan her *j* noktası için, komşu değerlerin ağırlıklı bir toplamı hesaplanmakta (doğrusal bir kombinasyon) ve ağırlıklar yumuşatmanın yapılıp yapılmayacağını veya türevin hesaplanıp hesaplanmadığını belirlemektedir. Komşuların sayısı ve polinomun derecesi, düzleştirme gücünü kontrol etmektedir. Bir vektör bileşeni olan x_j 'nin dönüştürülmesi;

$$x_j^{*} = \frac{1}{N} \sum_{h=-k}^{k} c_h \, x_j + h \tag{3.7}$$

Burada x_j^{*} yeni değer (düzleştirilmiş eğri veya türev), *N* normalleştirme sabiti, *k j*'nin her iki yanındaki komşu değerlerin sayısı, c_h kullanılan polinom derecesine ve amaca (düzleştirme, birinci veya ikinci türev) bağlı katsayılardır. Savitzky-Golay algoritması ile birinci dereceden türev alındığında temel uzaklaştırılmakta, ikinci dereceden türev alındığında ise temel ve doğrusal eğilim uzaklaştırılmaktadır (Naes ve ark., 2004; Olivieri, 2018; Savitzky ve Golay, 1964)

İlk pencerede sinyale ikinci dereceden bir polinom uygulandığında fonksiyonel açıklanması;

$$y = ax^2 + bx + c \tag{3.8}$$

Burada x dalga boyu, y spektral absorbans, a, b ve c en küçük kareler uygulamasıyla tahmin edilen, ayarlanabilir parametrelerdir. Bu durumda birinci türev hesaplaması (3.7) ve ikinci türev hesaplaması (3.8);

$$\frac{dy}{dx} = 2ax + b \tag{3.9}$$

$$\frac{d^2y}{d\,x^2} = 2a \tag{3.10}$$

Çalışmada Savitzky-Golay (SG) algoritması aracılığıyla farklı polinom dereceleri ve pencere aralıkları ile düzgünleştirme ve birinci ve ikinci dereceden türev alma yaklaşımları denenmiştir. Çalışmada kullanılan SG algoritmalarının parametreleri Tablo 3.9 'de verilmiştir.

Tablo 3.9: Savitzky-Golay algoritmasında kullanılan parametreler.

	Polinom derecesi	Pencere/ Filtre uzunluğu	Türev derecesi
SG1	1	1	-
Sm	1	15	-
Sm + 1 Dr	1	3	1
Sm + 2 Dr	2	5	2

3.2.3.5.Kısmi En Küçük Kareler Diskriminant analizi (PLS-DA)

Çalışmada gerçekleştirilen kısmi en küçük kareler diskriminant analizi (PLS-DA) için SIMPLS algoritması kullanılmıştır. SIMPLS algoritmasının temel aşamaları şu şekildedir;

1- PLS ağırlık vektörünün hesaplanması,

$$w = X'\,y \tag{3.11}$$

2- Skorların hesaplanması,

$$t = \frac{X.w}{\sqrt{\sum w^2}} \tag{3.12}$$

3- x yüklemelerinin hesaplanması,

$$p = \frac{t'.X}{\sum t^2} \tag{3.13}$$

4- c yüklemelerinin hesaplanması,

$$q = \frac{y'.t}{\sum t^2} \tag{3.14}$$

5- Artık veri matrisi elde edilmesi (veri matrisinden yeni PLS bileşeninin çıkarılması ile),

$$X_{res} = X - t.p \tag{3.15}$$

6- c 'nin artık değerinin hesaplanması,

$$y_{res} = y - t.q \tag{3.16}$$

7- Başka bileşenlere ihtiyaç duyulduğunda X ve c artıkları ile yer değiştirerek 6 basamağın tekrar edilmesi.

Burada X veri setini, T skor matrisini (PLS bileşenleri), p X yüklemelerini, y bağımlı değişkenleri (grup bilgilerini), q y yüklemelerini göstermektedir. Kullanılan temel PLS-DA denklemleri (Brereton ve Lloyd, 2014),

$$X = T.P + E \tag{3.17}$$

$$Y = T.q + f \tag{3.18}$$

Çalışmada kullanılan ön işlem uygulanmamış PLS-DA model parametreleri ve korelasyon tabloları cins, çam türleri ve meşe türleri olmak üzere sırasıyla EK-2, EK-3, EK-4, EK-5, EK-6 ve EK-7'de verilmiştir.

3.2.3.6.Diyagonal Doğrusal Diskriminant Analizi

Bu çalışmada diskriminant analizinin düzenlenmiş formalarından diyagonal doğrusal diskriminant analizi (DL-DA) doğrusal modellere alternatif olarak kullanılmıştır (Friedman, 1988; Guo ve ark., 2005; Hastie ve ark., 2017; Pang ve ark., 2009; Wehrens, 2011).

Karesel diskriminant analizi olarak da isimlendirilen (QDA) DL-DA diyagonal sınıf kovaryansı için Gauss en çok olabilirlik ayrım kuralını kullanmaktadır. Diskriminant fonksiyonu,

$$d_k(x) = -\sum_{j=1}^{p} \frac{\left(x_j^* - \bar{x}_{kj}\right)^2}{s_j^2} + 2log\pi_k \quad (3.19)$$

Burada denklemin ilk parçası x* den k. kütle merkezine (sınıfa) olan negatif standartlaştırılmış mesafenin karesini, ikinci parçası öncelikli olasılık sınıfına (π_k) dayanan düzeltmeyi temsil etmektedir. Çalışmada tüm veri setlerinde kovaryans matrisleri diyagonal ve sabit kabul edilerek, doğrusal diskriminant kuralı ve Gauss en çok olabilirlik fonksiyonu (Naive Bayes) kullanılmıştır.

3.2.3.7.Karar Ağaçları (DT)

Çalışmada kullanılan kalibrasyon metotlarından karar ağaçlarında amaç x değişkenlerinin alanının koordinat sistemi boyunca, yanlış sınıflandırma ölçüsünü en aza indirgeyecek şekilde, bölgelere ($R_1, \ldots, R_r$) bölmektir (Hastie ve ark., 2017; Varmuza ve Filzmoser, 2008). $x_i = x_1, \ldots, x_n$ e kadar bağımsız değişkenleri, $y_i = y_1, \ldots, y_n$ bağımlı değişkenleri (sınıf) göstermektedir. y_i $1,2,\ldots,k$ ya kadar k adet sınıfın değerlerin almakta, n_l nesnesi R_l alanında $n_1+\ldots-n_r=n$ kadar değer almaktadır. Indeks fonksiyonu $I(y_i=j)$ şeklinde gösterildiğinde ($y_i=j$ ise 1, $y_i \neq j$ ise 0), l bölgesindeki j grubunun göreceli frekansı p_{lj},

$$p_{lj} = \frac{1}{n_l} \sum_{x_i \in R_l} I(y_i = j) \quad (3.20)$$

Çalışmada kullanılan DT model parametreleri ve karmaşıklık parametreleri EK-8 ve EK-9'da verilmiştir.

3.2.3.8.Yapay Sinir Ağları (ANN)

Doğrusal olmayan yaklaşımlardan yapay sinir ağlarında (ANN) temel yaklaşım, girdilerin (bağımsız değişkenlerin) doğrusal kombinasyonları ile türetilmiş yeni özelliklerin oluşturulması ve daha sonra bu yeni özelliklerin doğrusal olmayan bir fonksiyon ile modellenmesidir (Hastie ve ark., 2017). ANN de hata fonksiyonu denklemin ilk kısmı ile gösterilmekte, ikinci kısımda modelin eğitim veri seti ile fazla uyumlu olmasını engellemek amacıyla "ağırlık azalımı (λ)" olarak isimlendirilen düzenlenmiş kriter eklenmektedir (Varmuza ve Filzmoser, 2008).

$$-\sum_{i=1}^{n}\sum_{j=1}^{k}\widehat{y_{ij}}\log\hat{y}_{ij}+\lambda\sum(parametreler)^2\rightarrow min \quad (3.21)$$

Çalışmada kullanılan ANN model parametreleri EK-10 ve EK-11'de verilmiştir.

3.2.3.9.Destek Vektör Makineleri (SVM)

Doğrusal olmayan yaklaşımlardan destek vektör makineleri (SVM) analizinde doğrusal kernel (çekirdek) kullanımı (u*v) ile model kurumuştur. Sınıflandırmada kullanılan temel fonksiyon,

$$K(x_i, x_j) = h(x_i)^T\, h\,(x_j) \quad (3.22)$$

Burada K dönüştürülen uzaydaki verileri hesaplayan çekirdek fonksiyonu, h (x_i) dönüştürülmüş özellik vektörünü belirtmektedir (Hastie ve ark., 2017; Varmuza ve Filzmoser, 2008).

Çalışmada kullanılan SVM model parametreleri EK-12'de verilmiştir.

3.2.3.10. Model performans değerlendirmesi ve Yorumlama

Çalışmada kullanılan modellerin performans değerlendirilmesinde hata matrisi formu (Tablo 3.10) kullanılmıştır (Banks ve ark., 2004).

Tablo 3.10: Hata matrisi.

Hata Matrisi		Gerçek Değer	
		Pozitif (1)	Negatif (0)
Tahmin Edilen Değer	Pozitif (1)	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif (0)	Yanlış Negatif(YN)	Doğru Negatif (DN)

Doğru Tahmin Oranı $=\frac{DP+DN}{DP+DN+YP+YN}$ (3.23)

Sensitivite/Duyarlılık $=\frac{DP}{DP+YN}$ (3.24)

Spesifite/Özgünlük$=\frac{DN}{DN+YP}$ (3.25)

Yanlış Sınıflandırma oranı $=\frac{YN+YP}{DP+DN+YP+YN}$ (3.26)

3.2.3.11. Yorumlama

Çalışmada kullanılan modellerden PLS-DA'nın yorumlanmasında Seçicilik Oranı (SR) kullanılmıştır (Farres ve ark. 2015; Rajalahti ve ark. 2009;Tran ve ark., 2014).

$$SR_i = \frac{SS_{i,açıklanan}}{SS_{i,artık}} \tag{3.27}$$

Burada SS i, açıklanan açıklanan varyans oranını, SSi,artık her değişkenin artık varsyansını kareler toplamı olarak ifade etmektedir.

$$SS_{i,açıklanan} = \|t_{TP}\, p'_{TPi}\|^2 \tag{3.28}$$

$$SS_{i,artık} = \|e_{TPi}\|^2 \tag{3.29}$$

Burada t_{TP} X üzerinde yansıtılan skor vektörünü, p_{TP} X üzerinde yansıtılan yüklemeyi ifade etmektedir.

$$t_{TP} = \frac{X\, b_{PLS}}{\|b_{PLS}\|} \tag{3.30}$$

$$p_{TP} = \frac{X'\, t_{TP}}{(t'_{TP}\, t_{TP})} \tag{3.31}$$

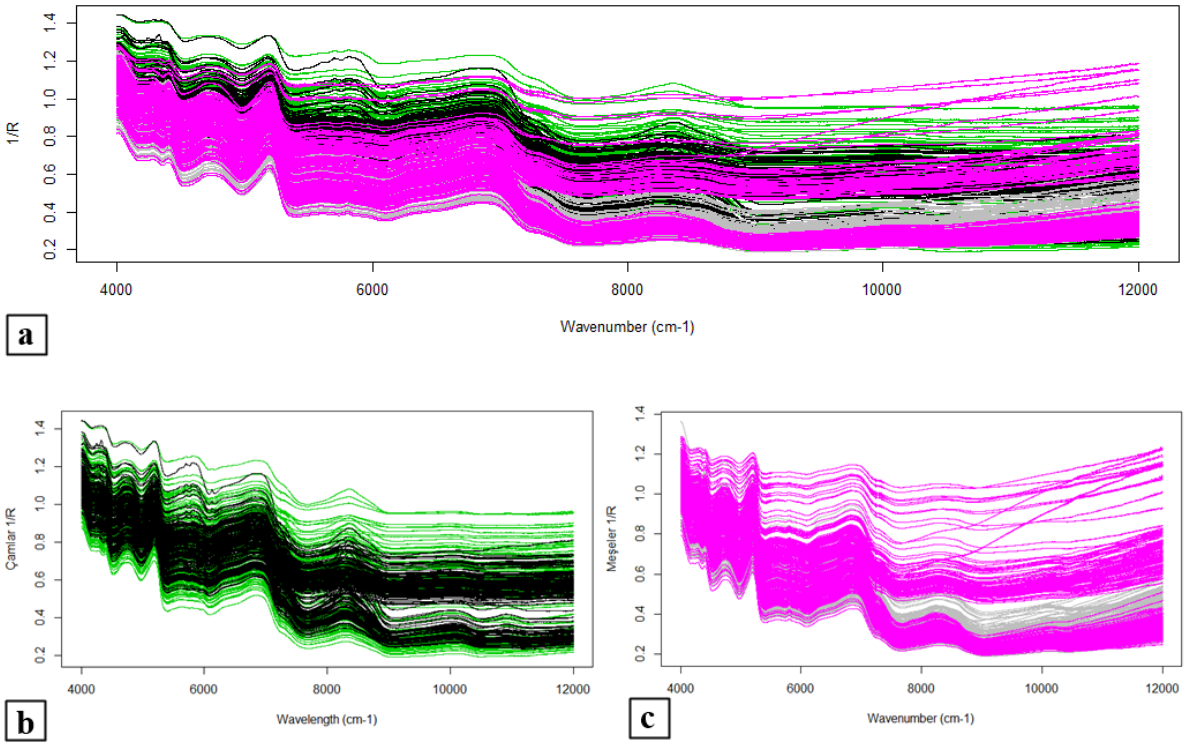
Değişken önemi belirlerken, SR değerlerine göre değişkenler yüksekten düşüğe doğru sıralanmıştır (Kvalheim, 2020).

4. BULGULAR

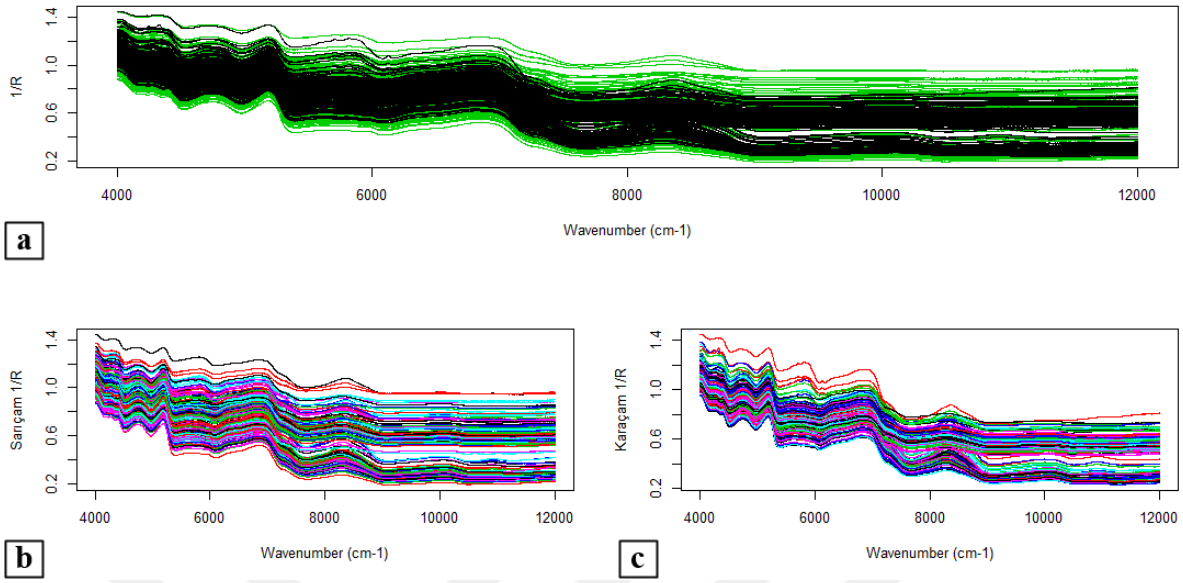
Çalışmada sarıçam, karaçam, sapsız meşe ve Macar meşesi olmak üzere dört farklı ağaç türü kullanılmıştır. Bu bölümde ağaç türlerinin teşhisinde NIRS yönteminin kullanım etkisi sırasıyla, önce cins bazında, ardından ikili gruplar halinde tür bazında ortaya konulmuştur.

4.1.İŞLEM GÖRMEMİŞ SPEKTRAL VERİLER

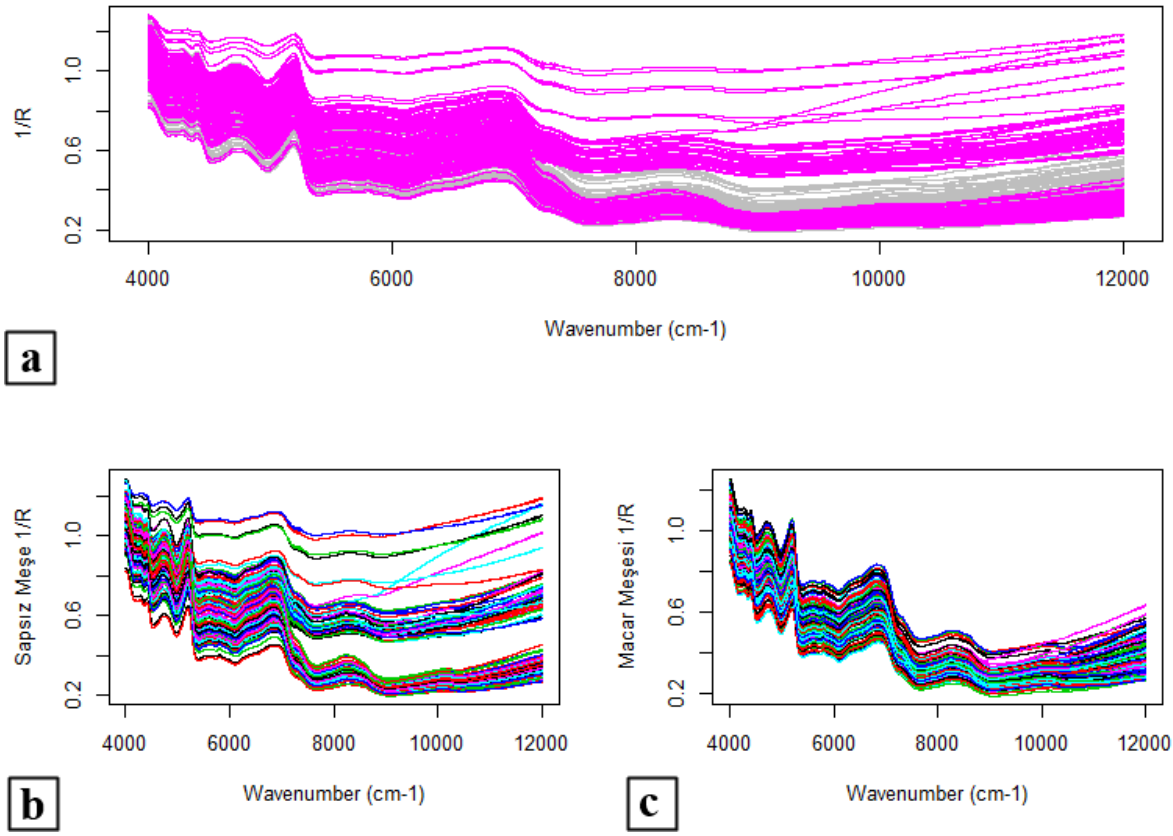
Cins bazında toplanan işlem görmemiş (ham) spektral verilere ait gösterim Şekil 4.1 de, çam türlerinden toplanan ham spektrumlar Şekil 4.2 de, meşe türlerinden toplanan ham spektrumlar Şekil 4.3 de verilmiştir.



Şekil 4.1: Çam (sarıçam, karaçam) ve meşe (sapsız meşe, Macar meşesi) cinslerine ait yansıtma (1/R) değerleri: a) toplu gösterim b) Çam türleri, c) Meşe türleri (yeşil: sarıçam, siyah: karaçam, pembe: sapsız meşe, gri: Macar meşesi).



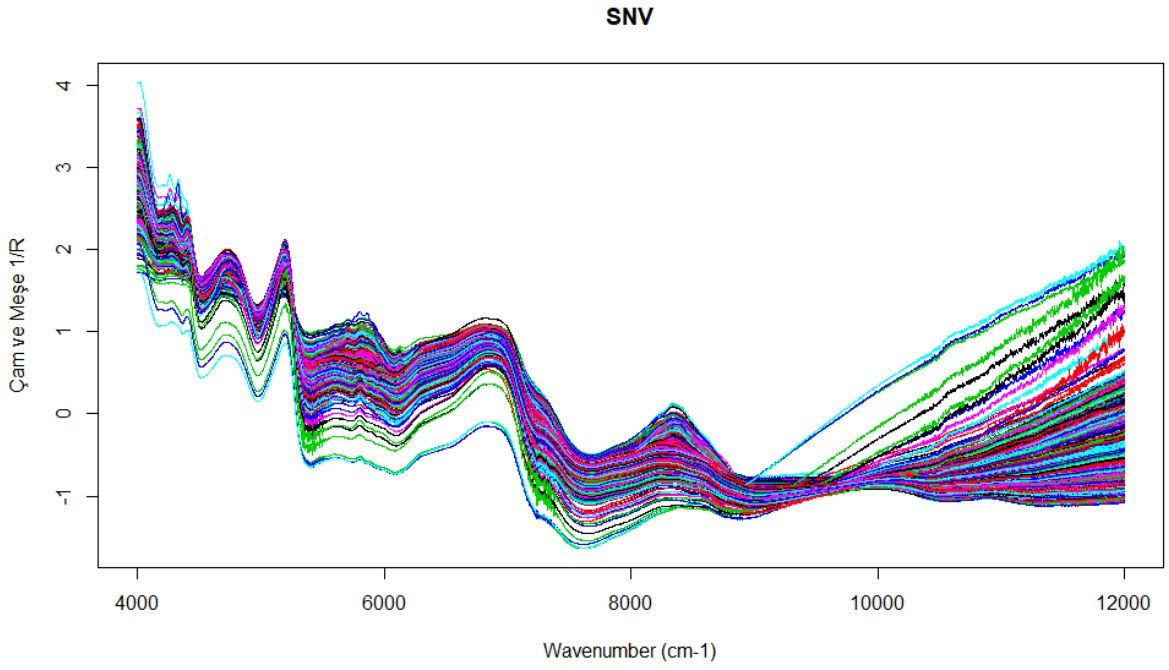
Şekil 4.2: Sarıçam ve karaçam ağaç türlerine ait yansıtma (1/R) değerleri, a) toplu gösterim (yeşil: sarıçam, siyah: karaçam), b) sarıçam, c) karaçam.



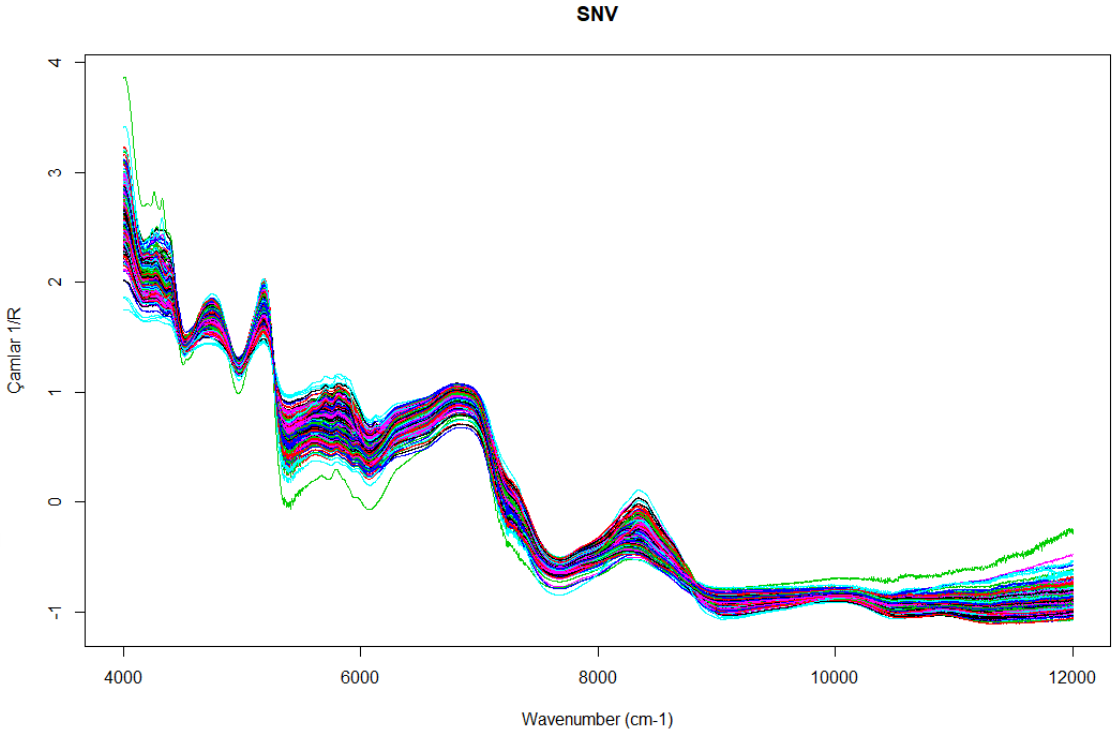
Şekil 4.3: Sapsız meşe ve macar meşesi ağaç türlerine ait yansıtma (1/R) değerleri, a) toplu gösterim (pembe: sapsız meşe, gri: Macar meşesi), b) sapsız meşe, c) Macar meşesi.

4.2.ÖN İŞLEM UYGULANMIŞ SPEKTRAL VERİLER

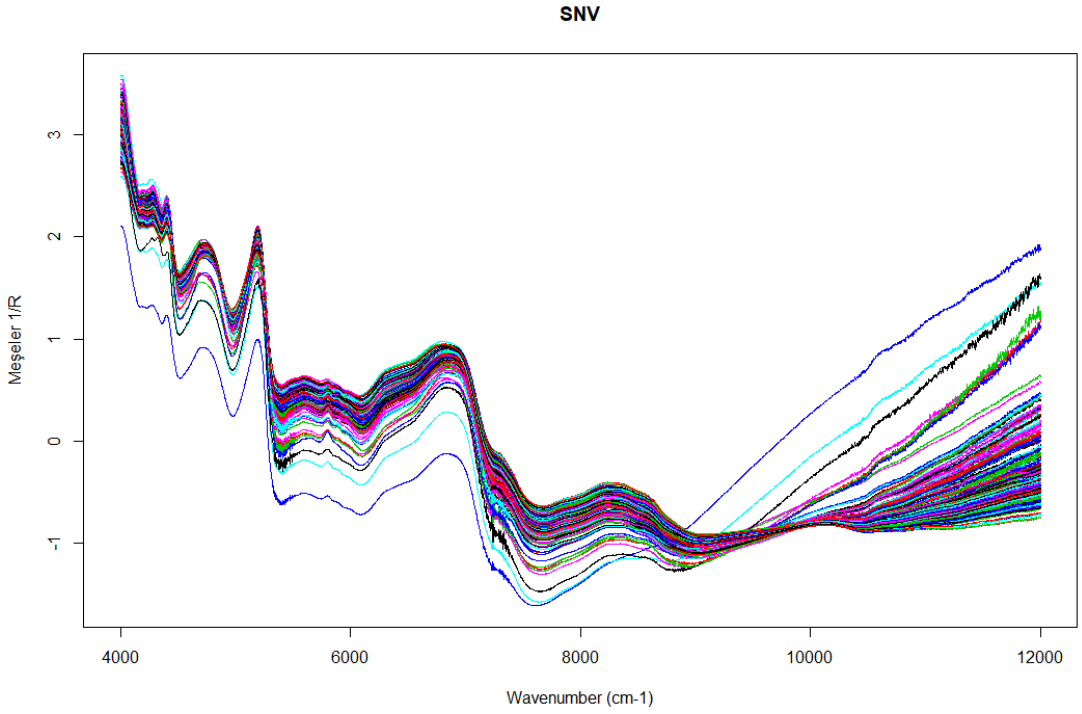
Gerçekleştirilen çalışmada, ağaç türü teşhisinde yakın kızıl ötesi spektroskopisinden elde edilen veriler arasındaki ilişki ham verilerin yanı sıra ön işlem uygulamaları ardından oluşturulan yeni veri setleri ile de ortaya koyulmuştur. Uygulanan ön işlemler sonrası veri setlerindeki SNV dönüşümleri Şekil 4.4, Şekil 4.5 ve Şekil 4.6'da, MSC dönüşümleri Şekil 4.7, Şekil 4.8 ve Şekil 4.9'da, Savitzky-Golay algoritması ile 1. türev dönüşümleri Şekil 4.10, Şekil 4.11 ve Şekil 4.12'de, Savitzky-Golay algoritması ile 1. türevi alınan verilerde SNV dönüşümleri Şekil 4.13, Şekil 4.14 ve Şekil 4.15'de, Savitzky-Golay algoritması ile 1. türevi alınan verilerde MSC dönüşümleri Şekil 4.16, Şekil 4.17 ve Şekil 4.18'de, Savitzky-Golay algoritması ile düzgünleştirme Şekil 4.19, Şekil 4.20 ve Şekil 4.21'de, Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma Şekil 4.22, Şekil 4.23 ve Şekil 4.24'de, Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma dönüşümleri Şekil 4.25, Şekil 4.26 ve Şekil 4.27'de gösterilmiştir.



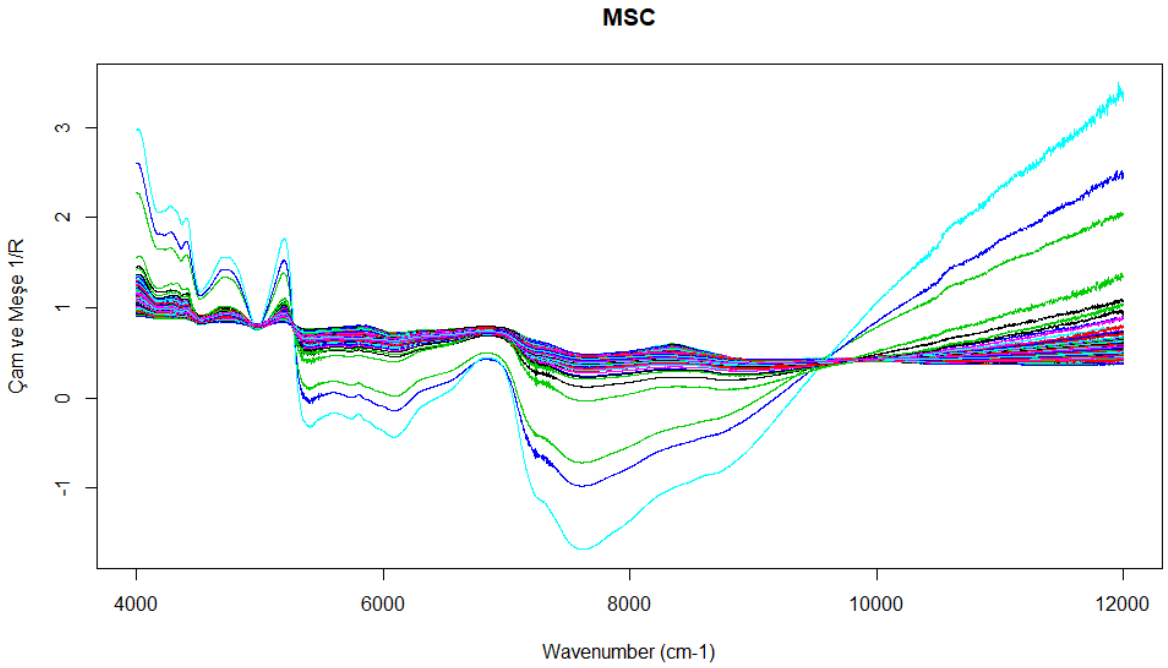
Şekil 4.4: Çam ve meşe ağaç türlerine ait veri setinde SNV dönüşümü.



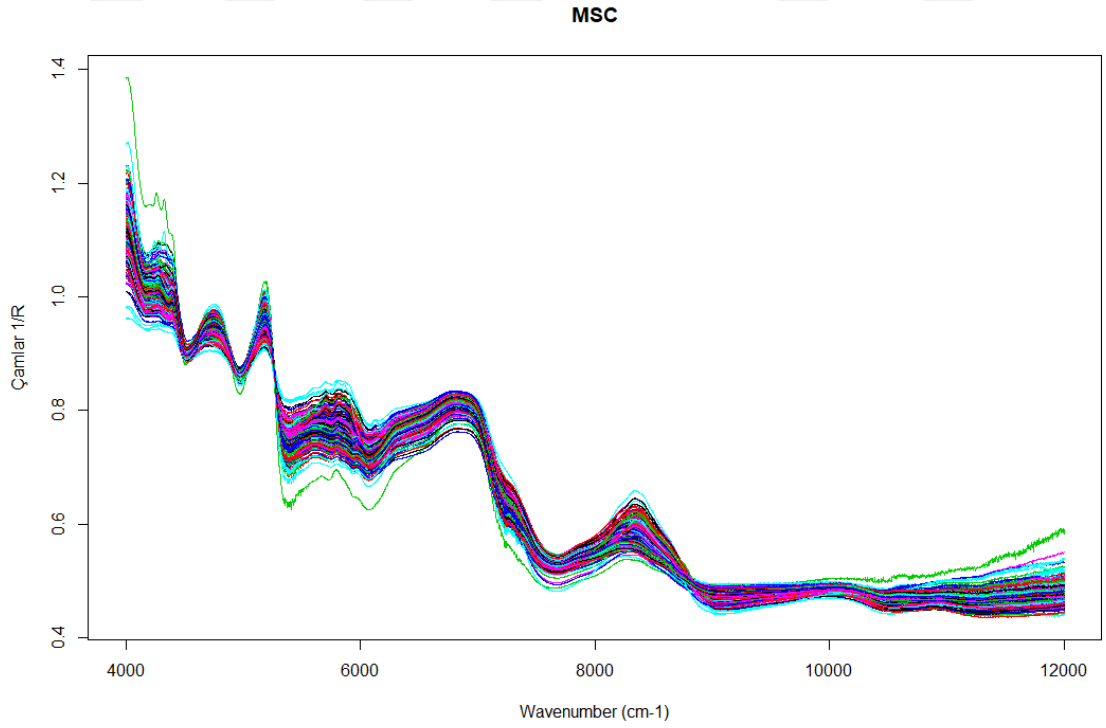
Şekil 4.5: Çam ağaç türlerine ait veri setinde SNV dönüşümü.



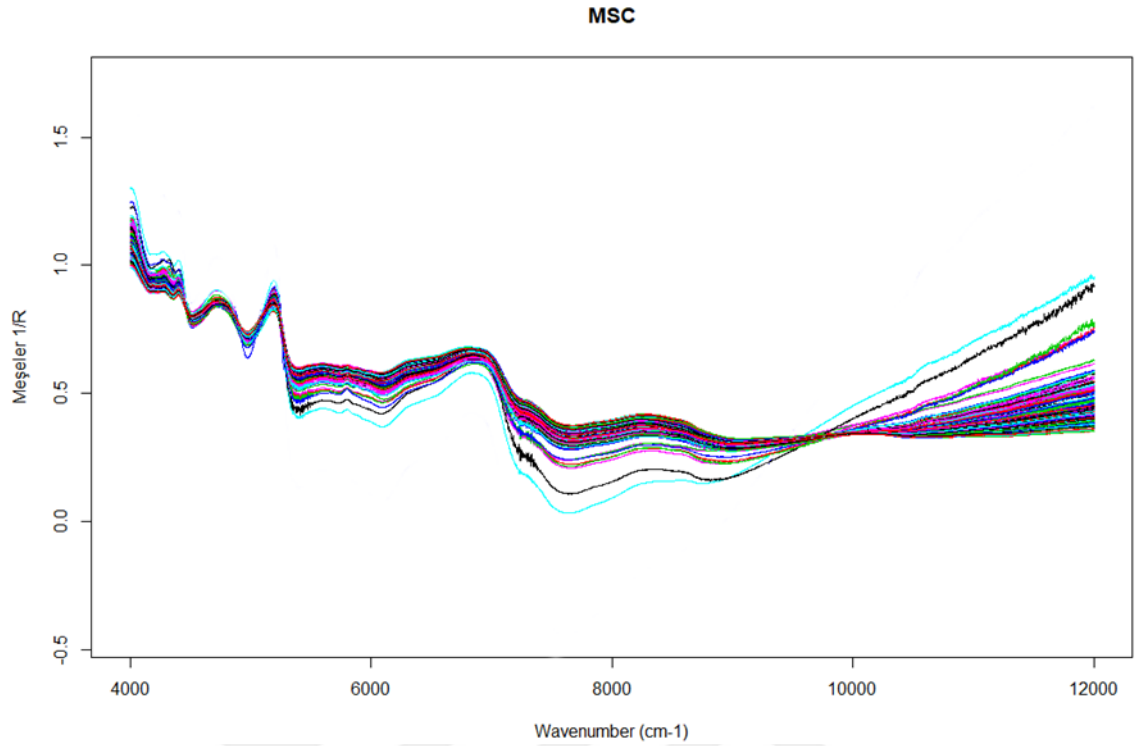
Şekil 4.6: Meşe ağaç türlerine ait veri setinde SNV dönüşümü.



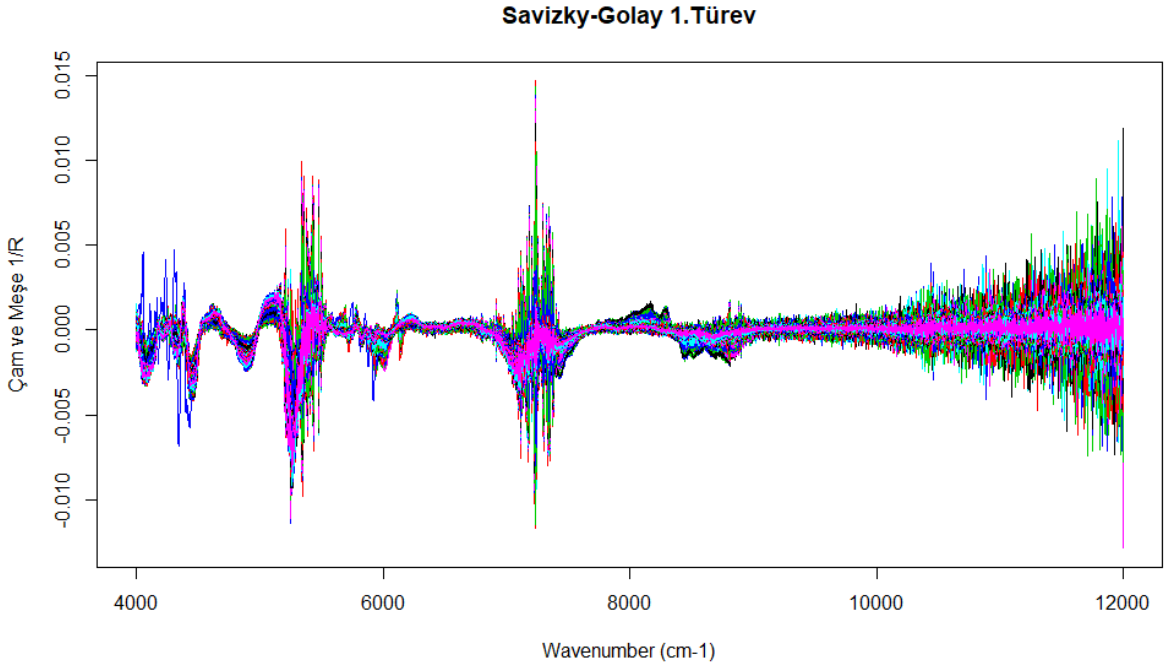
Şekil 4.7: Çam ve meşe ağaç türlerine ait veri setinde MSC dönüşümü.



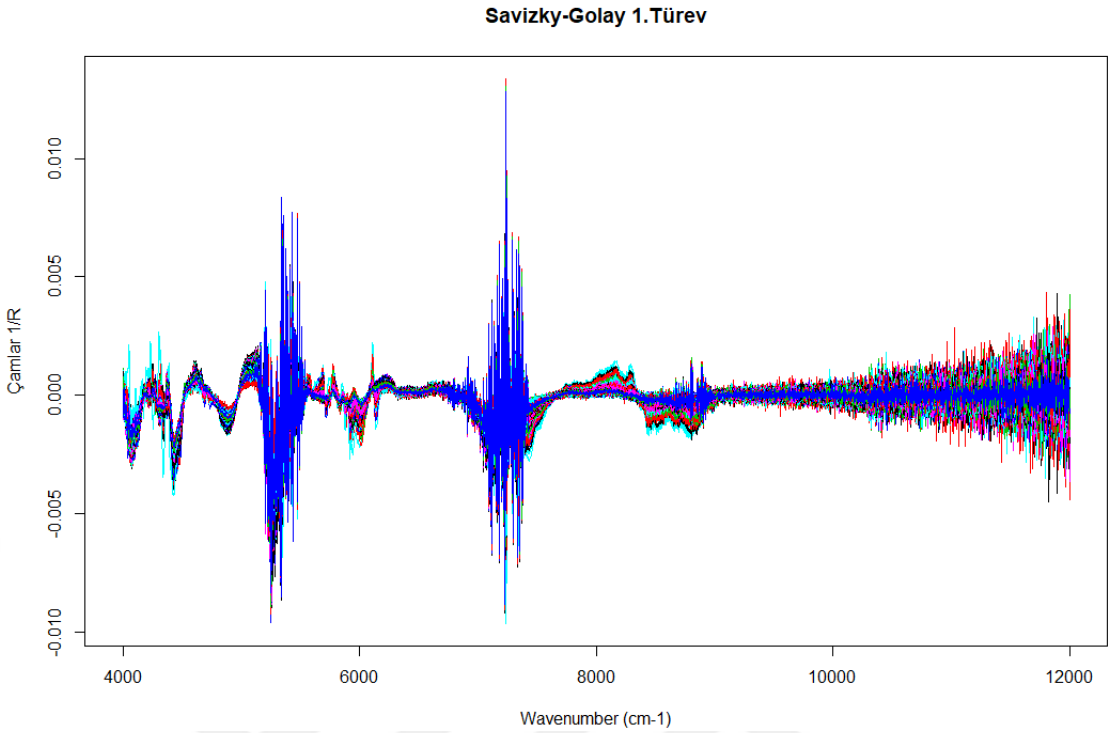
Şekil 4.8: Çam ağaç türlerine ait veri setinde MSC dönüşümü.



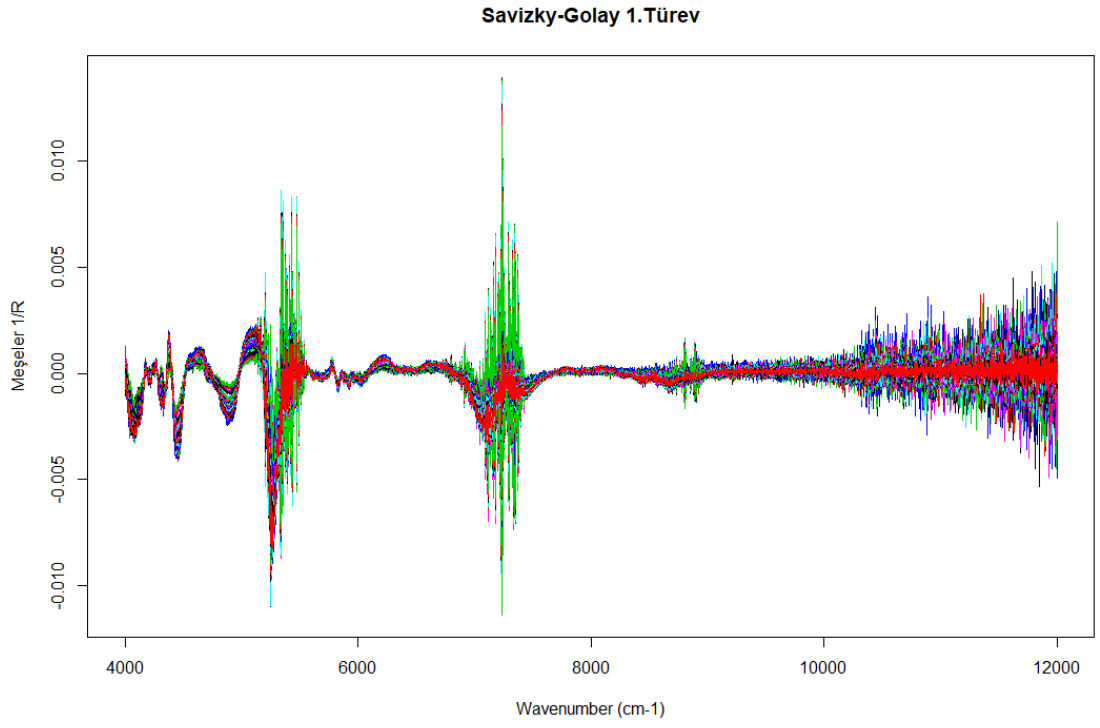
Şekil 4.9: Meşe ağaç türlerine ait veri setinde MSC dönüşümü.



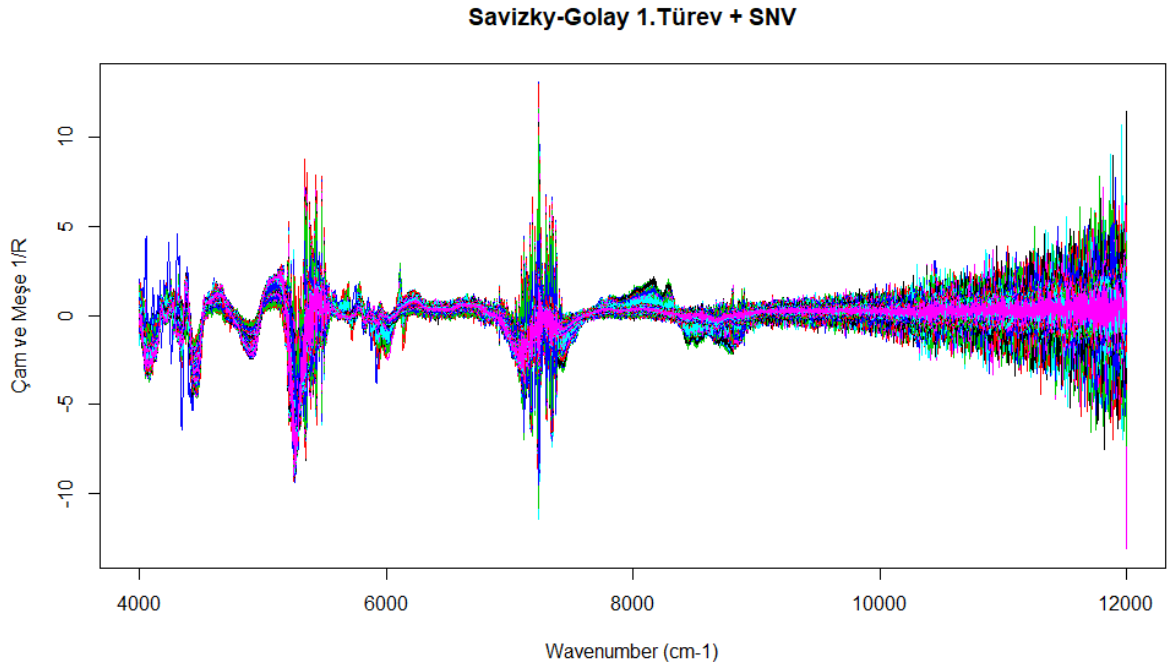
Şekil 4.10: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türev dönüşümü.



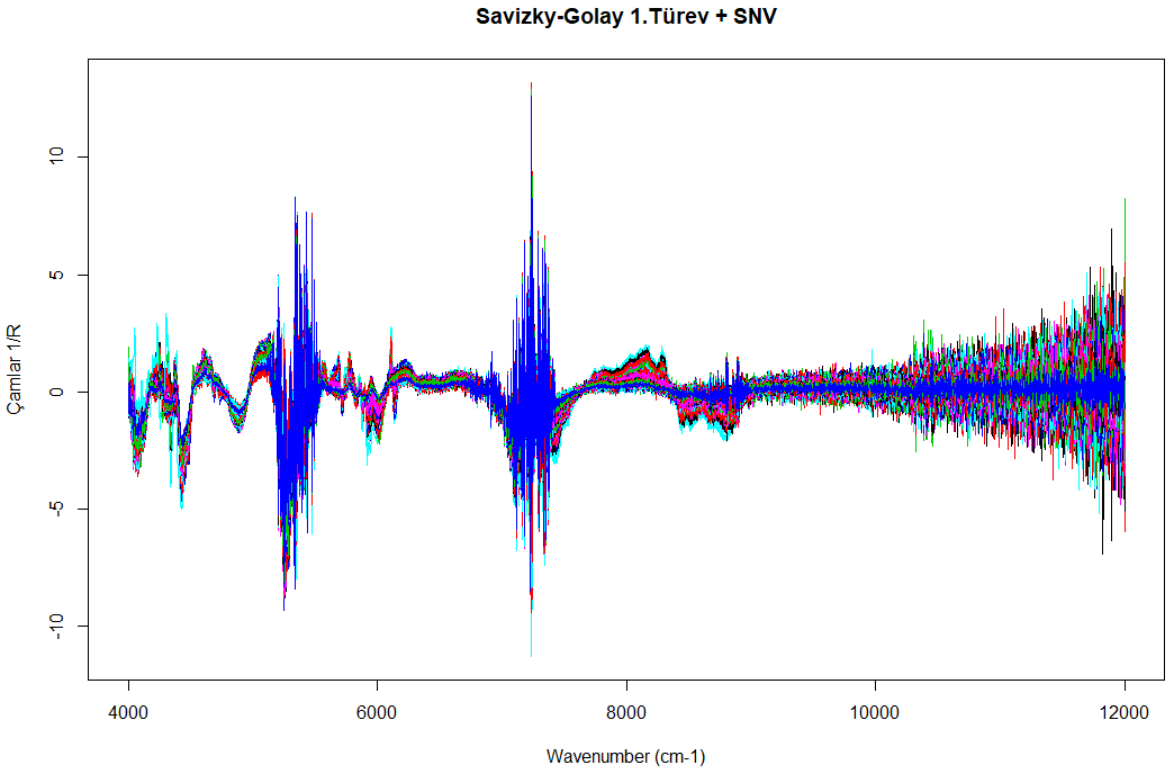
Şekil 4.11: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türev dönüşümü.



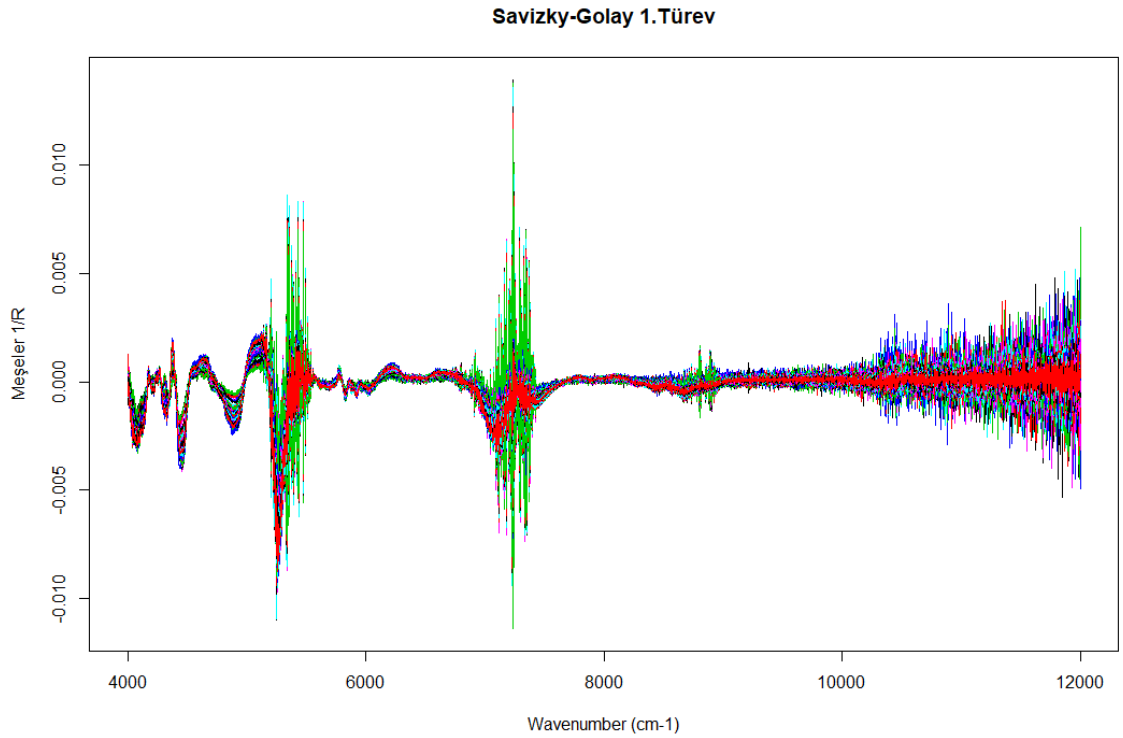
Şekil 4.12: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türev dönüşümü.



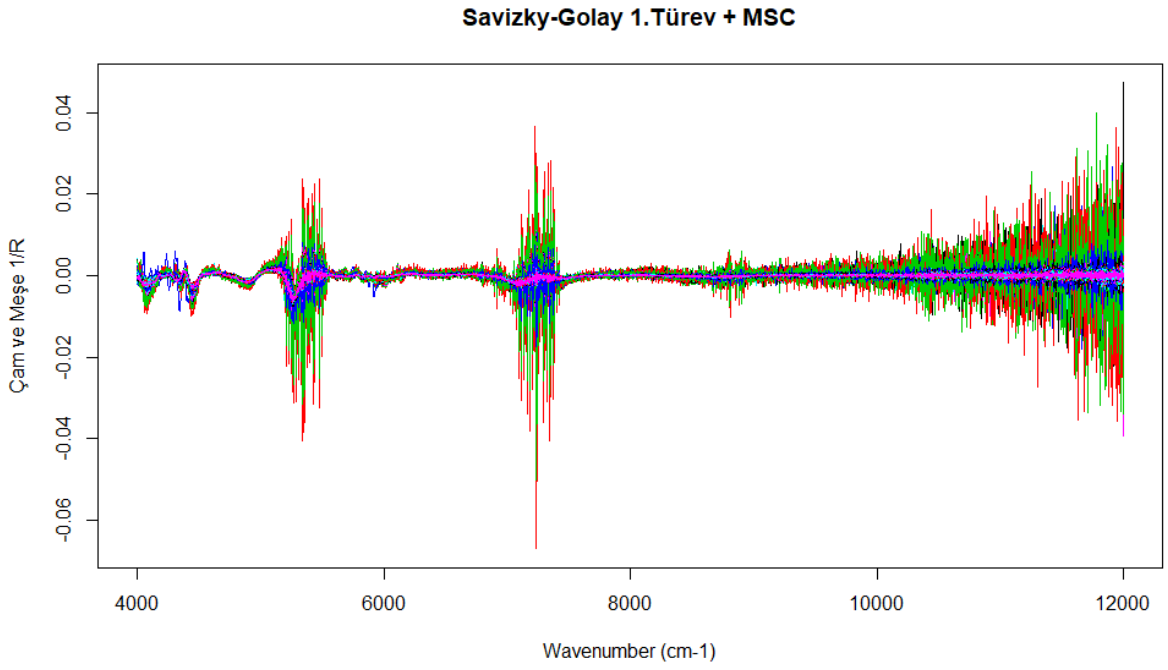
Şekil 4.13: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere SNV uygulanması.



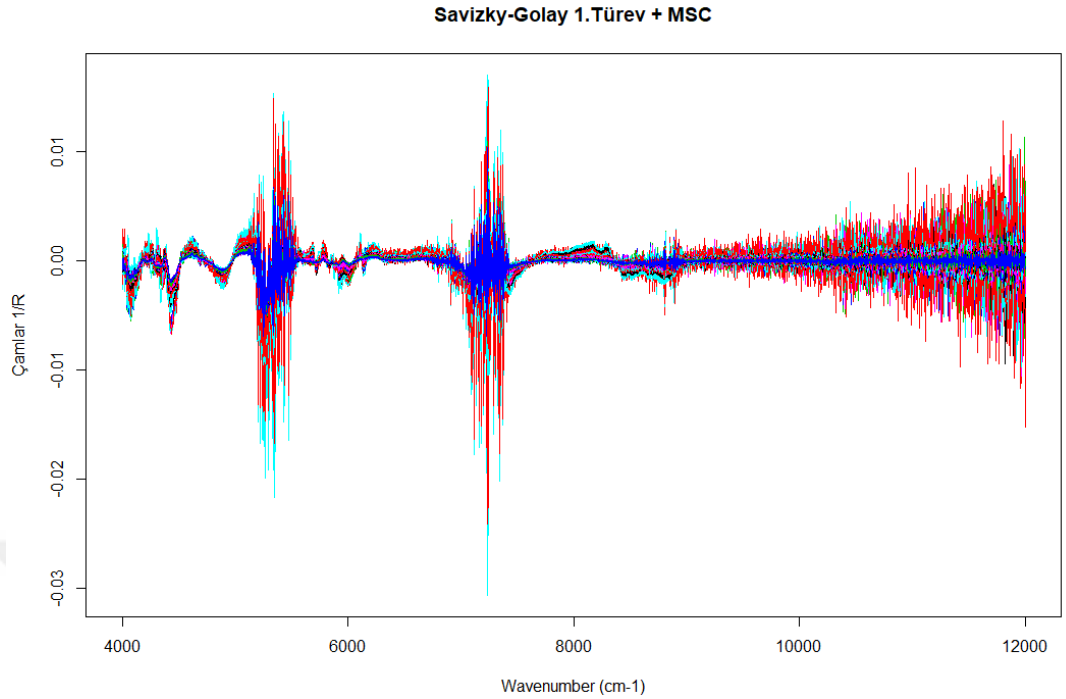
Şekil 4.14: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere SNV uygulanması.



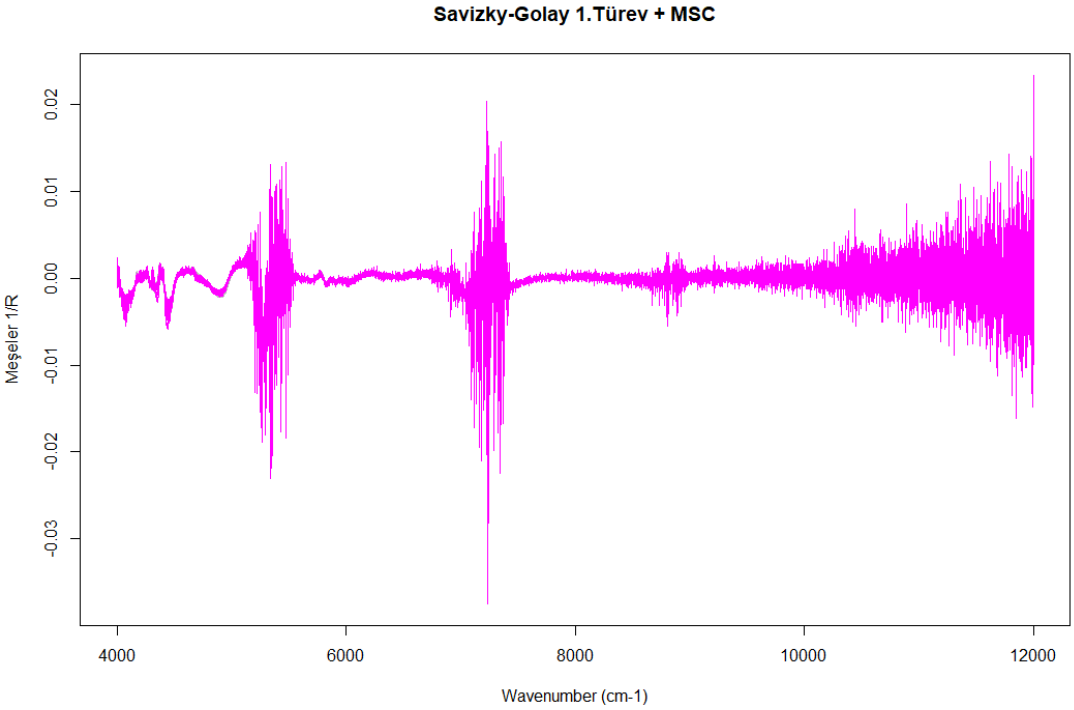
Şekil 4.15: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere SNV uygulanması.



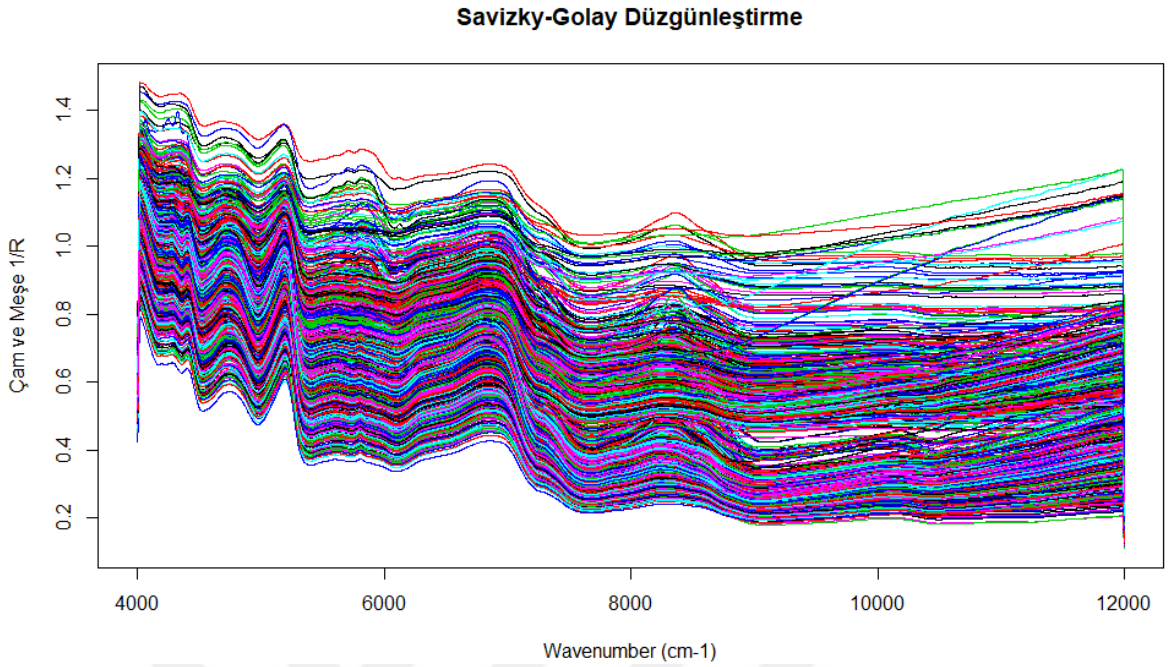
Şekil 4.16: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere MSC uygulanması.



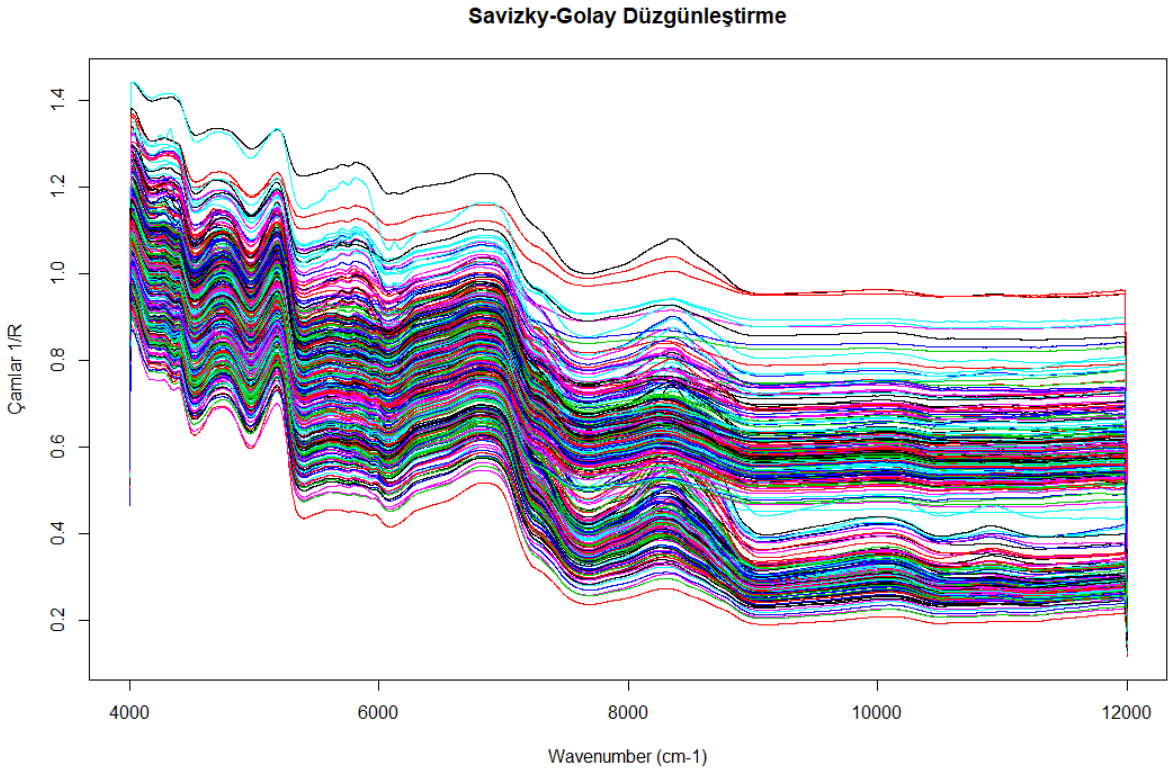
Şekil 4.17: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere MSC uygulanması.



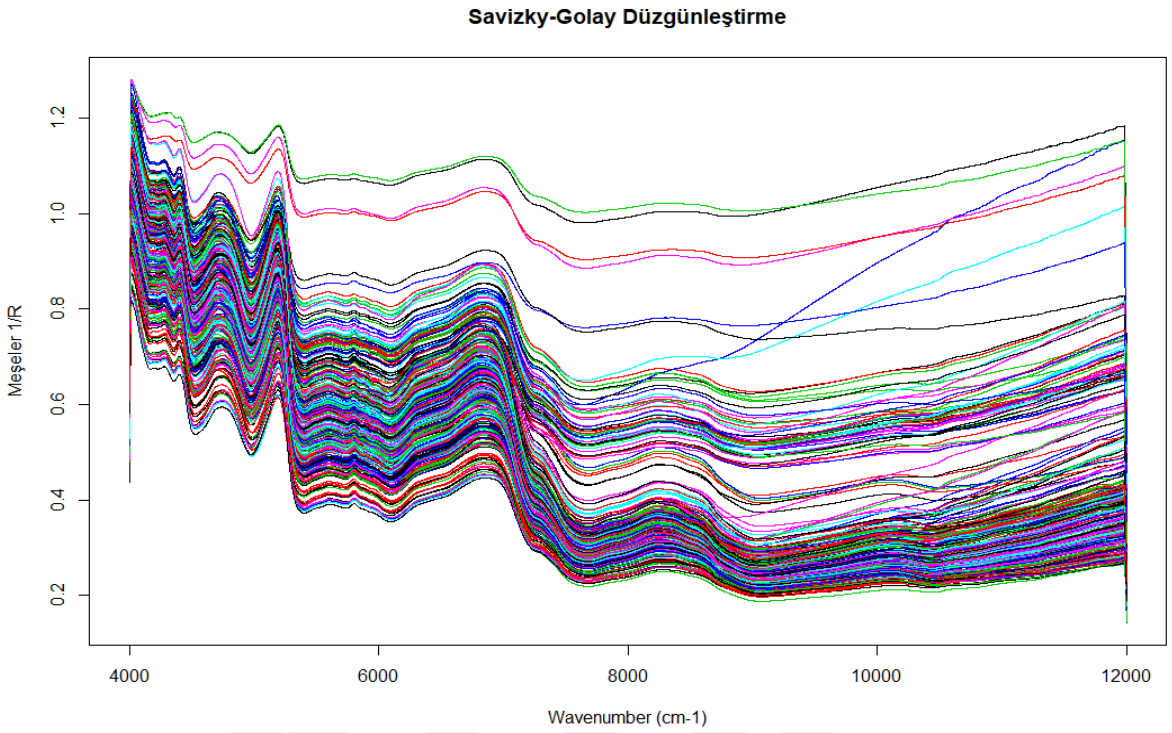
Şekil 4.18: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile 1. türevi alınan verilere MSC uygulanması.



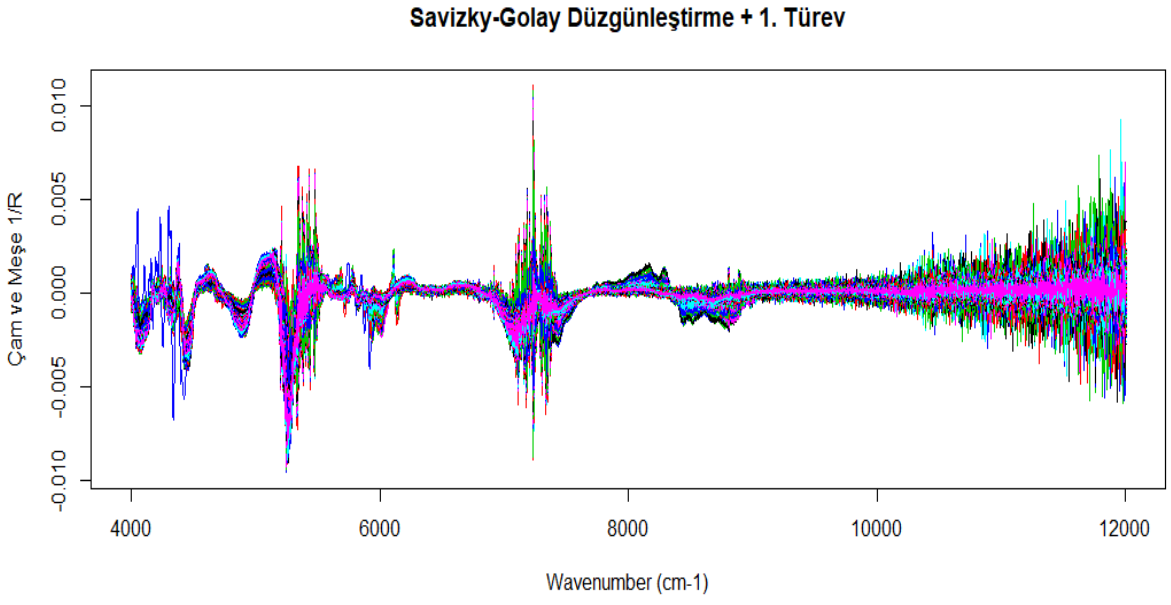
Şekil 4.19: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme.



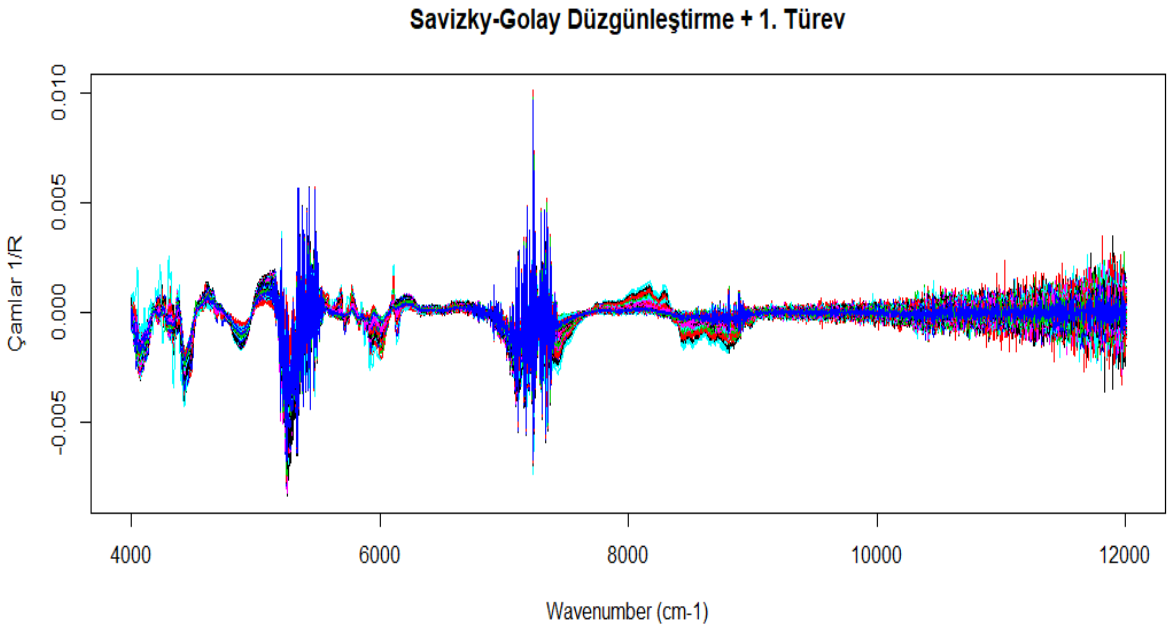
Şekil 4.20: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme.



Şekil 4.21: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme.



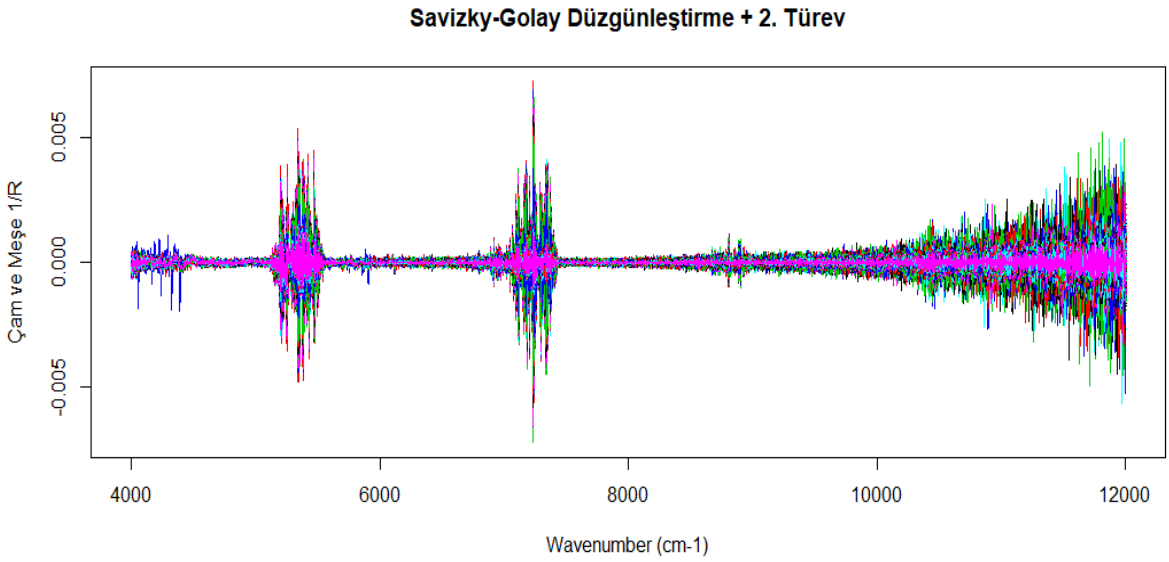
Şekil 4.22: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma.



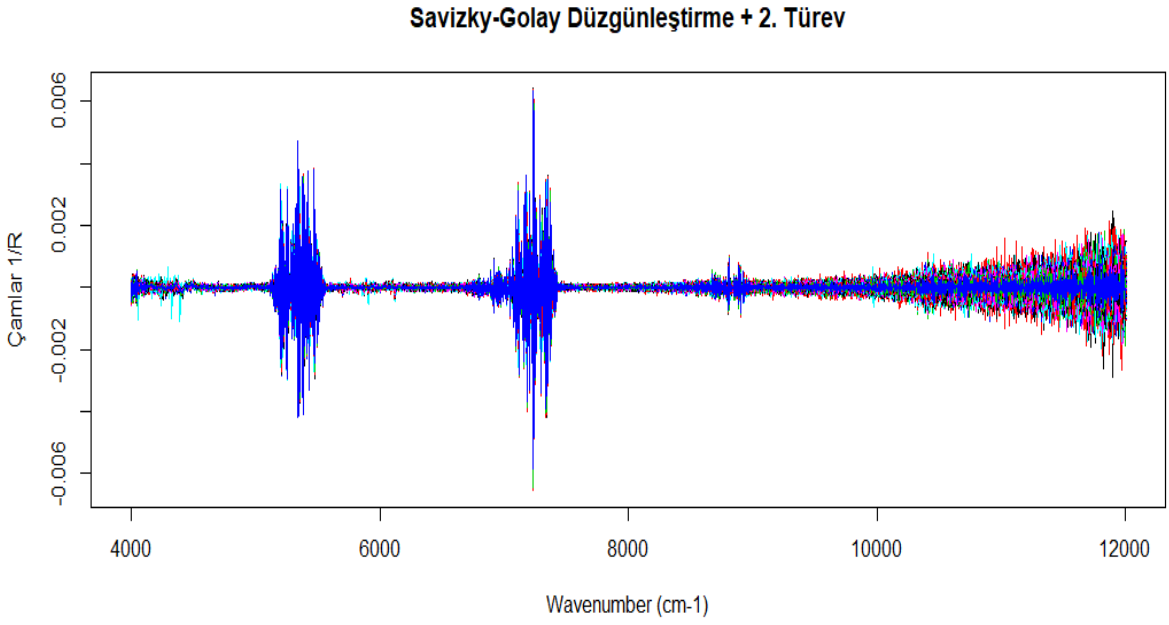
Şekil 4.23: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma.



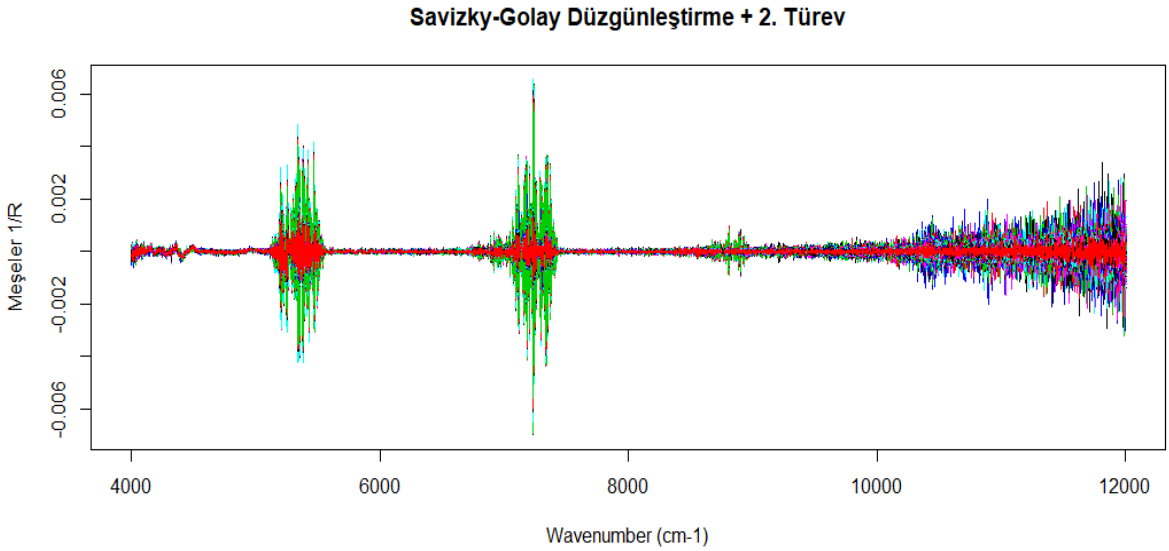
Şekil 4.24: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 1. türev alma.



Şekil 4.25: Çam ve meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma.



Şekil 4.26: Çam ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma.



Şekil 4.27: Meşe ağaç türlerine ait veri setinde Savitzky-Golay algoritması ile düzgünleştirme ve 2. türev alma.

4.3.MODEL PERFORMANSLARI

NIRS yönteminin ağaç türü teşhisinde kullanım etkinliğini ortaya koymak amacıyla doğrusal ve doğrusal olmayan çeşitli modeller denenmiştir. Bu denemeler; ham veri setleri ve ön işlem uygulandıktan sonra elde edilen yeni veri setleri ile gerçekleştirilmiştir. Kurulan modellerin ağaç türü teşhisinde kullanımı açısından başarısı ve uygulanan ön işlemlerin modellerin doğru tahmin oranları üzerindeki etkisi eğitim ve test veri setleri ile sınanmıştır (Tablo 4.1).

Toplam veri setinin ortalama % 70 ± 3' ünü içeren eğitim veri setleri aracılığıyla, modeldeki hatanın en düşük olduğu yer bulunmuş ve bu şekilde modeller kurulmuştur. Daha sonra bu modeller, model kurma aşamasına katılmayan test setleri ile denenmiş ve modellerin başarısı doğru tahmin oranları aracılığıyla belirlenmiştir. Modellerin oluşturulması aşamasında eğitim veri seti ile kurulan modellerin doğrulukları genellikle pratikte karşılaşılacak değerlerden yüksek çıkmaktadır. Bu sebeple model kurulması aşamasında kullanılmayan örnekler (test örnekleri) ile kurulan modellerin denenmesi gerekmektedir. Bu amaçla ayrılan ve model kurulma aşamasında kullanılmayan test veri setleri ile modellerin gerçek hayatta beklenen başarılarına bakılmıştır.

Tablo 4.1: Eğitim ve test setlerine ayrılan örnek sayıları.

4150 dalga boyu /değişken		**Eğitim**	**Test**	**Toplam**
ÇAM	**sarıçam**	143	52	195
	karaçam	128	53	181
	toplam	*271*	*105*	*376*
MEŞE	**Macar meşesi**	113	48	161
	sapsız meşe	99	48	147
	toplam	*212*	*96*	*308*
	Genel toplam	*483*	*201*	*684*

4.3.1. Kısmi en küçük kareler diskriminant analizi

Kısmi en küçük kareler diskriminat analizi (PLS-DA) aracılığıyla oluşturulan sınıflandırma modellerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.2 , Tablo 4.3 ve Tablo 4.4'de verilmiştir. PLS-DA aracılığıyla eğitim setleri ile oluşturulan modellerin performanslarının çok iyi olduğu görülmektedir. PLS-DA aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile kurulduğunda; kurulan modellerin çam-meşe cins ayrımında % 100, çam türlerinin ayrımında % 97,4, meşe türlerinin ayrımında % 100 lük doğru tahmin oranlarına sahip oldukları görülmektedir.

Tablo 4.2: Çam ve meşe cins ayrımında eğitim veri seti ve PLS-DA ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	**DP**	**YP**	**DN**	**YN**	**Özgüllük**	**Duyarlılık**	**Doğru TO**	**Yanlış TO**
Ham	212	0	271	0	1	1	1	0
SNV	212	0	271	0	1	1	1	0
MSC	212	0	271	0	1	1	1	0
SG 1. Dr	212	0	271	0	1	1	1	0
SG 1. Dr + SNV	212	0	271	0	1	1	1	0
SG 1. Dr + MSC	212	0	271	0	1	1	1	0
Sm	212	0	271	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	212	0	271	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	212	0	271	0	1	1	1	0

Tablo 4.3: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve PLS-DA ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	121	0	143	7	1	0,945	0,974	0,026
SNV	122	0	143	6	1	0,953	0,978	0,022
MSC	125	0	143	3	1	0,977	0,989	0,011
SG 1. Dr	128	0	143	0	1	1	1	0
SG 1. Dr + SNV	128	0	143	0	1	1	1	0
SG 1. Dr + MSC	128	0	143	0	1	1	1	0
Sm	105	11	132	23	0,923	0,820	0,875	0,126
Sm + 1. Dr (w=3,p=1)	128	0	143	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	128	0	143	0	1	1	1	0

Tablo 4.4: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve PLS-DA ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	99	0	113	0	1	1	1	0
SNV	99	0	113	0	1	1	1	0
MSC	99	0	113	0	1	1	1	0
SG 1. Dr	99	0	113	0	1	1	1	0
SG 1. Dr + SNV	99	0	113	0	1	1	1	0
SG 1. Dr + MSC	99	0	113	0	1	1	1	0
Sm	98	0	113	1	1	0,990	0,995	0,005
Sm + 1. Dr (w=3,p=1)	99	0	113	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	99	0	113	0	1	1	1	0

PLS-DA aracılığıyla oluşturulan sınıflandırma modellerinin test setleri ile analizlerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.5, Tablo 4.6 ve Tablo 4.7'de verilmiştir. PLS-DA aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin cins ayrımında % 99, çam türleri için % 83,8, meşe türleri için % 95,8 'lik doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahminin cins ayrımında üç farklı ön işlemle sağlandığı (%100- MSC, birinci türevi alınmış SNV ve birinci türevi alınmış MSC); tür ayrımında her iki deneme grubu için de düzgünleştirilmiş ve birinci türevi alınmış veri setleri kullanıldığında elde edildiği tespit edilmiştir.

Tablo 4.5: Çam ve meşe cins ayrımında PLS-DA modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	94	2	105	0	0,981	1	0,990	0,010
SNV	94	2	105	0	0,981	1	0,990	0,010
MSC	96	0	105	0	1	1	1	0
SG 1. Dr	95	1	105	0	0,991	1	0,995	0,005
SG 1. Dr + SNV	96	0	105	0	1	1	1	0
SG 1. Dr + MSC	96	0	105	0	1	1	1	0
Sm	95	1	105	0	0,991	1	0,995	0,005
Sm + 1. Dr (w=3,p=1)	95	1	105	0	0,991	1	0,995	0,005
Sm + 2. Dr (w=5,p=2)	90	6	104	1	0,945	0,989	0,965	0,035

Tablo 4.6: Sarıçam ve karaçam türlerinin ayrımında PLS-DA modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	45	8	43	9	0,843	0,833	0,838	0,162
SNV	40	13	37	15	0,740	0,727	0,733	0,267
MSC	43	10	45	7	0,818	0,860	0,838	0,162
SG 1. Dr	46	7	52	0	0,881	1	0,933	0,067
SG 1. Dr + SNV	47	6	51	1	0,895	0,979	0,933	0,067
SG 1. Dr + MSC	47	6	52	0	0,897	1	0,943	0,057
Sm	43	10	38	14	0,792	0,754	0,771	0,229
Sm + 1. Dr (w=3,p=1)	48	5	52	0	0,912	1	0,952	0,048
Sm + 2. Dr (w=5,p=2)	44	9	52	0	0,852	1	0,914	0,086

Tablo 4.7: Sapsız meşe ve Macar meşesi türlerinin ayrımında PLS-DA modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	48	4	44	0	0,917	1	0,958	0,042
SNV	48	3	45	0	0,938	1	0,969	0,031
MSC	48	5	43	0	0,896	1	0,948	0,052
SG 1. Dr	42	0	48	5	1	0,875	0,938	0,063
SG 1. Dr + SNV	43	0	48	5	1	0,896	0,948	0,052
SG 1. Dr + MSC	39	0	48	9	1	0,813	0,906	0,094
Sm	48	4	44	0	0,917	1	0,958	0,042
Sm + 1. Dr (w=3,p=1)	46	0	48	2	1	0,958	0,979	0,021
Sm + 2. Dr (w=5,p=2)	39	0	48	9	1	0,813	0,906	0,094

4.3.2. Diyagonal doğrusal diskriminant analizi

Diyagonal doğrusal diskriminant analizi (DL-DA) aracılığıyla oluşturulan sınıflandırma modellerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.8, Tablo 4.9 ve Tablo 4.10'de verilmiştir. DL-DA aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 49,8, meşe türleri için % 54,7 lik doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahmin oranı çam türleri için düzgünleştirilmiş ve birinci türevi alınmış veri setinde (% 78,2), meşe türleri için düzgünleştirilmiş ve ikinci türevi alınmış veri setinde (% 96,2) tespit edilmiştir.

Tablo 4.8: Çam ve meşe cins ayrımında eğitim veri seti ve DL-DA ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	181	90	171	41	0,655	0,815	0,729	0,271
SNV	267	4	203	9	0,981	0,967	0,973	0,027
MSC	266	5	202	10	0,976	0,964	0,969	0,031
SG 1. Dr	264	7	208	4	0,967	0,985	0,977	0,023
SG 1. Dr + SNV	270	1	211	1	0,995	0,996	0,996	0,004
SG 1. Dr + MSC	268	3	212	0	0,986	1	0,994	0,006
Sm	181	90	171	41	0,655	0,815	0,729	0,271
Sm + 1. Dr (w=3,p=1)	264	7	209	3	0,968	0,989	0,979	0,021
Sm + 2. Dr (w=5,p=2)	265	6	191	21	0,970	0,927	0,944	0,056

Tablo 4.9: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve DL-DA ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	57	71	78	65	0,524	0,467	0,498	0,502
SNV	79	49	76	67	0,608	0,541	0,572	0,428
MSC	79	49	67	76	0,608	0,541	0,572	0,428
SG 1. Dr	103	25	108	35	0,812	0,745	0,779	0,221
SG 1. Dr + SNV	76	52	108	35	0,675	0,685	0,679	0,321
SG 1. Dr + MSC	80	48	101	42	0,678	0,656	0,668	0,332
Sm	57	71	78	65	0,524	0,467	0,498	0,502
Sm + 1. Dr (w=3,p=1)	103	25	109	34	0,813	0,752	0,782	0,218
Sm + 2. Dr (w=5,p=2)	89	39	119	24	0,753	0,788	0,768	0,233

Tablo 4.10: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve DL-DA ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	41	58	75	38	0,564	0,519	0,547	0,453
SNV	71	28	91	22	0,765	0,763	0,764	0,236
MSC	67	32	92	21	0,742	0,761	0,750	0,250
SG 1. Dr	84	15	108	5	0,878	0,944	0,906	0,094
SG 1. Dr + SNV	79	20	112	1	0,849	0,988	0,901	0,099
SG 1. Dr + MSC	87	12	106	7	0,898	0,926	0,910	0,090
Sm	41	58	75	38	0,564	0,519	0,547	0,453
Sm + 1. Dr (w=3,p=1)	85	14	106	7	0,883	0,924	0,901	0,099
Sm + 2. Dr (w=5,p=2)	91	8	113	0	0,934	1	0,962	0,038

DL-DA aracılığıyla oluşturulan sınıflandırma modellerinin test setleri ile analizlerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.11, Tablo 4.12 ve Tablo 4.13'de verilmiştir. DL-DA aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 46,7, meşe türleri için % 61,5 lik doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahmin oranı çam türleri için düzgünleştirilmiş ve 1.türevi alınmış veri setinde (% 73,3), meşe türleri için 1. Türevi alınmış ve SNV uygulanmış veri setinde (% 77,1) tespit edilmiştir.

Tablo 4.11: Çam ve meşe cins ayrımında DL-DA modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	71	34	171	41	0,694	0,789	0,736	0,264
SNV	104	1	203	9	0,990	0,990	0,990	0,010
MSC	104	1	202	10	0,989	0,912	0,945	0,055
SG 1. Dr	105	0	208	4	1,000	0,913	0,950	0,050
SG 1. Dr + SNV	102	3	211	1	0,969	0,971	0,970	0,030
SG 1. Dr + MSC	101	4	212	0	0,960	1,000	0,980	0,020
Sm	71	34	171	41	0,694	0,789	0,736	0,264
Sm + 1. Dr (w=3,p=1)	105	0	209	3	1,000	0,929	0,960	0,040
Sm + 2. Dr (w=5,p=2)	93	12	191	21	0,882	0,939	0,910	0,090

Tablo 4.12: Sarıçam ve karaçam türlerinin ayrımında DL-DA modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	21	32	28	24	0,467	0,467	0,467	0,533
SNV	30	23	29	23	0,558	0,566	0,562	0,438
MSC	36	17	17	35	0,500	0,507	0,505	0,495
SG 1. Dr	37	16	38	14	0,704	0,725	0,714	0,286
SG 1. Dr + SNV	31	22	32	20	0,593	0,608	0,600	0,400
SG 1. Dr + MSC	34	19	32	20	0,627	0,630	0,629	0,371
Sm	21	32	28	24	0,467	0,467	0,467	0,533
Sm + 1. Dr (w=3,p=1)	39	14	38	14	0,731	0,736	0,733	0,267
Sm + 2. Dr (w=5,p=2)	32	21	37	15	0,638	0,681	0,657	0,343

Tablo 4.13: Sapsız meşe ve Macar meşesi türlerinin ayrımında DL-DA modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	20	28	39	9	0,582	0,690	0,615	0,385
SNV	34	14	29	19	0,674	0,642	0,656	0,344
MSC	14	34	48	0	0,585	1	0,646	0,354
SG 1. Dr	31	17	34	14	0,667	0,689	0,677	0,323
SG 1. Dr + SNV	27	21	47	1	0,691	0,964	0,771	0,229
SG 1. Dr + MSC	35	13	32	16	0,711	0,686	0,698	0,302
Sm	20	28	39	9	0,582	0,690	0,615	0,385
Sm + 1. Dr (w=3,p=1)	35	13	33	15	0,717	0,700	0,708	0,292
Sm + 2. Dr (w=5,p=2)	19	29	48	0	0,623	1	0,698	0,302

4.3.3. Karar ağaçları

Doğrusal olmayan modellerin arasında yorumlaması en kolay modellerden biri olan karar ağaçları (DT) aracılığıyla oluşturulan sınıflandırma modellerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.14, Tablo 4.15 ve Tablo 4.16'de verilmiştir. DT aracılığıyla oluşturulan sınıflandırma modelleri tüm veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 100, meşe türleri için % 100 lük doğru tahmin oranlarına sahip oldukları görülmektedir.

Tablo 4.14: Çam ve meşe cins ayrımında eğitim veri seti ve DT ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	210	6	265	2	0,978	0,991	0,983	0,017
SNV	210	3	268	2	0,989	0,991	0,990	0,010
MSC	212	0	271	0	1	1	1	0
SG 1. Dr	212	0	271	0	1	1	1	0
SG 1. Dr + SNV	212	0	271	0	1	1	1	0
SG 1. Dr + MSC	212	0	271	0	1	1	1	0
Sm	210	6	265	2	0,978	0,991	0,983	0,017
Sm + 1. Dr (w=3,p=1)	212	0	271	0	1	1	1	0,000
Sm + 2. Dr (w=5,p=2)	212	3	268	0	0,989	1	0,994	0,006

Tablo 4.15: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve DT ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	128	0	143	0	1	1	1	0
SNV	128	0	143	0	1	1	1	0
MSC	128	0	143	0	1	1	1	0
SG 1. Dr	128	0	143	0	1	1	1	0
SG 1. Dr + SNV	128	0	143	0	1	1	1	0
SG 1. Dr + MSC	128	0	143	0	1	1	1	0
Sm	128	0	143	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	128	0	143	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	128	0	143	0	1	1	1	0

Tablo 4.16: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve DT ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	99	0	113	0	1	1	1	0
SNV	99	0	113	0	1	1	1	0
MSC	99	0	113	0	1	1	1	0
SG 1. Dr	99	0	113	0	1	1	1	0
SG 1. Dr + SNV	99	0	113	0	1	1	1	0
SG 1. Dr + MSC	99	0	113	0	1	1	1	0
Sm	99	0	113	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	99	0	113	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	99	0	113	0	1	1	1	0

DT aracılığıyla oluşturulan sınıflandırma modellerinin test setleri ile analizlerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.17, Tablo 4. 18 ve Tablo 4.19'da verilmiştir. DT aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 54,3, meşe türleri için % 76,0 lık doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahmin oranı çam türleri için birinci türevi alınmış veri setinde (% 83,8), meşe türleri için hiçbir ön işlem uygulanmamış (ham) veri setinde (% 76,0) tespit edilmiştir.

Tablo 4.17: Çam ve meşe cins ayrımında DT modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	87	3	102	9	0,971	0,906	0,940	0,060
SNV	85	0	105	11	1	0,885	0,945	0,055
MSC	95	0	105	1	1	0,990	0,995	0,005
SG 1. Dr	90	0	105	6	1	0,938	0,970	0,030
SG 1. Dr + SNV	87	0	105	9	1	0,906	0,955	0,045
SG 1. Dr + MSC	91	1	104	5	0,990	0,948	0,970	0,030
Sm	87	3	102	9	0,971	0,906	0,940	0,060
Sm + 1. Dr (w=3,p=1)	90	0	105	6	1	0,938	0,970	0,030
Sm + 2. Dr (w=5,p=2)	92	5	100	4	0,952	0,958	0,955	0,045

Tablo 4.18: Sarıçam ve karaçam türlerinin ayrımında DT modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	25	20	32	28	0,615	0,472	0,543	0,457
SNV	40	23	29	13	0,558	0,755	0,657	0,343
MSC	35	33	19	18	0,365	0,660	0,514	0,486
SG 1. Dr	39	3	49	14	0,942	0,736	0,838	0,162
SG 1. Dr + SNV	42	7	45	11	0,865	0,792	0,829	0,171
SG 1. Dr + MSC	37	2	50	16	0,962	0,698	0,829	0,171
Sm	28	23	29	25	0,558	0,528	0,543	0,457
Sm + 1. Dr (w=3,p=1)	38	16	36	15	0,692	0,717	0,705	0,295
Sm + 2. Dr (w=5,p=2)	40	5	47	13	0,904	0,755	0,829	0,171

Tablo 4.19: Sapsız meşe ve Macar meşesi türlerinin ayrımında DT modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	34	9	39	14	0,813	0,708	0,760	0,240
SNV	32	16	33	16	0,673	0,667	0,670	0,330
MSC	23	8	40	25	0,833	0,479	0,656	0,344
SG 1. Dr	30	9	39	18	0,813	0,625	0,719	0,281
SG 1. Dr + SNV	35	12	36	13	0,750	0,729	0,740	0,260
SG 1. Dr + MSC	40	22	26	8	0,542	0,833	0,688	0,313
Sm	32	9	39	16	0,813	0,667	0,740	0,260
Sm + 1. Dr (w=3,p=1)	21	5	43	27	0,896	0,438	0,667	0,333
Sm + 2. Dr (w=5,p=2)	22	9	39	26	0,813	0,458	0,635	0,365

4.3.4. Yapay sinir ağları

Yapay sinir ağları (ANN) aracılığıyla oluşturulan sınıflandırma modellerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.20, Tablo 4.21 ve Tablo 4.22'de verilmiştir. ANN aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 70,5, meşe türleri için % 85,4 lük doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahmin oranı çam türleri(% 87,5) ve meşe türleri (% 99,5) için düzgünleştirilmiş ve ikinci türevi alınmış veri setlerinde tespit edilmiştir.

Tablo 4.20: Çam ve meşe cins ayrımında eğitim veri seti ve ANN ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	212	0	268	3	1	0,986	0,994	0,006
SNV	212	0	268	3	1	0,986	0,994	0,006
MSC	212	0	269	2	1	0,991	0,996	0,004
SG 1. Dr	212	0	271	0	1	1	1	0
SG 1. Dr + SNV	212	0	271	0	1	1	1	0
SG 1. Dr + MSC	212	0	271	0	1	1	1	0
Sm	212	0	271	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	212	0	271	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	212	0	271	0	1	1	1	0

Tablo 4.21: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve ANN ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	73	55	118	25	0,682	0,745	0,705	0,295
SNV	75	53	109	34	0,673	0,688	0,679	0,321
MSC	77	51	116	27	0,695	0,740	0,712	0,288
SG 1. Dr	73	55	82	61	0,599	0,545	0,572	0,428
SG 1. Dr + SNV	65	63	11	35	0,638	0,650	0,642	0,358
SG 1. Dr + MSC	79	49	118	25	0,707	0,760	0,727	0,273
Sm	91	37	114	29	0,755	0,758	0,756	0,244
Sm + 1. Dr (w=3,p=1)	73	55	113	30	0,673	0,709	0,686	0,314
Sm + 2. Dr (w=5,p=2)	117	11	120	23	0,916	0,836	0,875	0,125

Tablo 4.22: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve ANN ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	79	20	102	11	0,836	0,878	0,854	0,146
SNV	76	23	102	11	0,816	0,874	0,840	0,160
MSC	83	16	105	8	0,868	0,912	0,887	0,113
SG 1. Dr	82	17	98	15	0,852	0,845	0,849	0,151
SG 1. Dr + SNV	75	24	107	6	0,817	0,926	0,858	0,142
SG 1. Dr + MSC	95	4	106	7	0,964	0,931	0,948	0,052
Sm	95	4	89	24	0,957	0,798	0,868	0,132
Sm + 1. Dr (w=3,p=1)	96	3	104	9	0,972	0,914	0,943	0,057
Sm + 2. Dr (w=5,p=2)	98	1	113	0	0,991	1,000	0,995	0,005

ANN aracılığıyla oluşturulan sınıflandırma modellerinin test setleri ile analizlerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.23, Tablo 4.24 ve Tablo 4.25'de verilmiştir. ANN aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 49,5, meşe türleri için % 81,3 lük doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahmin oranı çam türleri için düzgünleştirilmiş ve ikinci türevi alınmış veri setinde (% 75,2), meşe türleri için sadece düzgünleştirilmiş veri setinde (% 83,3) tespit edilmiştir.

Tablo 4.23: Çam ve meşe cins ayrımında ANN modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	96	0	104	1	1	0,990	0,995	0,005
SNV	96	0	104	1	1	0,990	0,995	0,005
MSC	96	0	103	2	1	0,980	0,990	0,010
SG 1. Dr	52	44	69	36	0,611	0,591	0,602	0,398
SG 1. Dr + SNV	93	3	98	7	0,970	0,930	0,950	0,050
SG 1. Dr + MSC	96	0	97	8	1	0,923	0,960	0,040
Sm	96	0	105	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	93	3	104	1	0,972	0,989	0,980	0,020
Sm + 2. Dr (w=5,p=2)	96	0	97	8	1	0,923	0,960	0,040

Tablo 4.24: Sarıçam ve karaçam türlerinin ayrımında ANN modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	14	39	38	14	0,494	0,500	0,495	0,505
SNV	36	17	20	32	0,541	0,529	0,533	0,467
MSC	35	18	26	26	0,591	0,574	0,581	0,419
SG 1. Dr	35	18	26	26	0,591	0,574	0,581	0,419
SG 1. Dr + SNV	14	39	33	19	0,458	0,424	0,448	0,552
SG 1. Dr + MSC	36	17	18	34	0,514	0,514	0,514	0,486
Sm	14	39	40	12	0,506	0,538	0,514	0,486
Sm + 1. Dr (w=3,p=1)	25	28	37	15	0,569	0,625	0,590	0,410
Sm + 2. Dr (w=5,p=2)	33	20	46	6	0,697	0,846	0,752	0,248

Tablo 4.25: Sapsız meşe ve Macar meşesi türlerinin ayrımında ANN modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	37	11	41	7	0,788	0,841	0,813	0,188
SNV	40	8	35	13	0,814	0,755	0,781	0,219
MSC	6	42	48	0	0,533	1	0,563	0,438
SG 1. Dr	15	33	37	11	0,529	0,577	0,542	0,458
SG 1. Dr + SNV	30	18	36	12	0,667	0,714	0,688	0,313
SG 1. Dr + MSC	29	19	37	11	0,661	0,725	0,688	0,313
Sm	45	3	35	13	0,921	0,776	0,833	0,167
Sm + 1. Dr (w=3,p=1)	22	26	39	9	0,600	0,710	0,635	0,365
Sm + 2. Dr (w=5,p=2)	20	28	48	0	0,632	1	0,708	0,292

4.3.5. Destek vektör makineleri

Destek vektör makineleri (SVM) aracılığıyla oluşturulan sınıflandırma modellerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.26, Tablo 4.27 ve Tablo 4.28'de verilmiştir. DT aracılığıyla oluşturulan sınıflandırma modelleri tüm veri setleri ile denendiğinde; kurulan modellerin her iki veri grubu içinde ham veri ve düzgünleştirme ön işlemi dışında % 100'lük doğru tahmin oranlarına sahip oldukları görülmektedir.

Tablo 4.26: Çam ve meşe cins ayrımında eğitim veri seti ve SVM ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	212	0	271	0	1	1	1	0
SNV	212	0	271	0	1	1	1	0
MSC	212	0	271	0	1	1	1	0
SG 1. Dr	212	0	271	0	1	1	1	0
SG 1. Dr + SNV	212	0	271	0	1	1	1	0
SG 1. Dr + MSC	212	0	271	0	1	1	1	0
Sm	212	0	271	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	212	0	271	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	212	0	271	0	1	1	1	0

Tablo 4.27: Sarıçam ve karaçam türlerinin ayrımında eğitim veri seti ve SVM ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	89	5	138	39	0,965	0,695	0,838	0,162
SNV	128	0	143	0	1	1	1	0
MSC	128	0	143	0	1	1	1	0
SG 1. Dr	128	0	143	0	1	1	1	0
SG 1. Dr + SNV	128	0	143	0	1	1	1	0
SG 1. Dr + MSC	128	0	143	0	1	1	1	0
Sm	87	7	136	41	0,951	0,680	0,822	0,177
Sm + 1. Dr (w=3,p=1)	128	0	143	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	128	0	143	0	1	1	1	0

Tablo 4.28: Sapsız meşe ve Macar meşesi türlerinin ayrımında eğitim veri seti ve SVM ile kurulan sınıflandırma modelinde elde edilen değerler.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	99	1	112	0	0,991	1	0,995	0,005
SNV	99	0	113	0	1	1	1	1
MSC	99	0	113	0	1	1	1	1
SG 1. Dr	99	0	113	0	1	1	1	1
SG 1. Dr + SNV	99	0	113	0	1	1	1	1
SG 1. Dr + MSC	99	0	113	0	1	1	1	1
Sm	98	1	112	1	0,991	0,990	0,991	0,009
Sm + 1. Dr (w=3,p=1)	99	0	113	0	1	1	1	1
Sm + 2. Dr (w=5,p=2)	99	0	113	0	1	1	1	1

SVM aracılığıyla oluşturulan sınıflandırma modellerinin test setleri ile analizlerine ait hata matrisleri ve bazı kriter değerleri Tablo 4.29, Tablo 4.30 ve Tablo 4.31'de verilmiştir. SVM aracılığıyla oluşturulan sınıflandırma modelleri hiçbir ön işlem görmemiş ham veri setleri ile denendiğinde; kurulan modellerin çam türleri için % 73,3, meşe türleri için % 90,6'lık doğru tahmin oranlarına sahip oldukları görülmektedir. Uygulanan çeşitli ön işlemler sonrasında genellikle bu doğruluk oranının arttığı gözlenmiştir. En yüksek doğru tahmin oranı çam türleri için SNV uygulanmış veri setinde (% 94,3), meşe türleri için birinci türevi alınmış ve SNV uygulanmış veri setinde (% 96,9) tespit edilmiştir.

Tablo 4.29: Çam ve meşe cins ayrımında SVM modelinin test veri seti ile değerlendirilmesi.

Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	96	0	105	0	1	1	1	0
SNV	93	0	105	3	1	0,969	0,985	0,015
MSC	92	0	105	4	1	0,958	0,980	0,020
SG 1. Dr	96	0	105	0	1	1	1	0
SG 1. Dr + SNV	96	0	105	0	1	1	1	0
SG 1. Dr + MSC	96	0	105	0	1	1	1	0
Sm	96	0	105	0	1	1	1	0
Sm + 1. Dr (w=3,p=1)	96	0	105	0	1	1	1	0
Sm + 2. Dr (w=5,p=2)	95	0	105	1	1	0,990	0,995	0,005

Tablo 4.30: Sarıçam ve karaçam türlerinin ayrımında SVM modelinin test veri seti ile değerlendirilmesi.

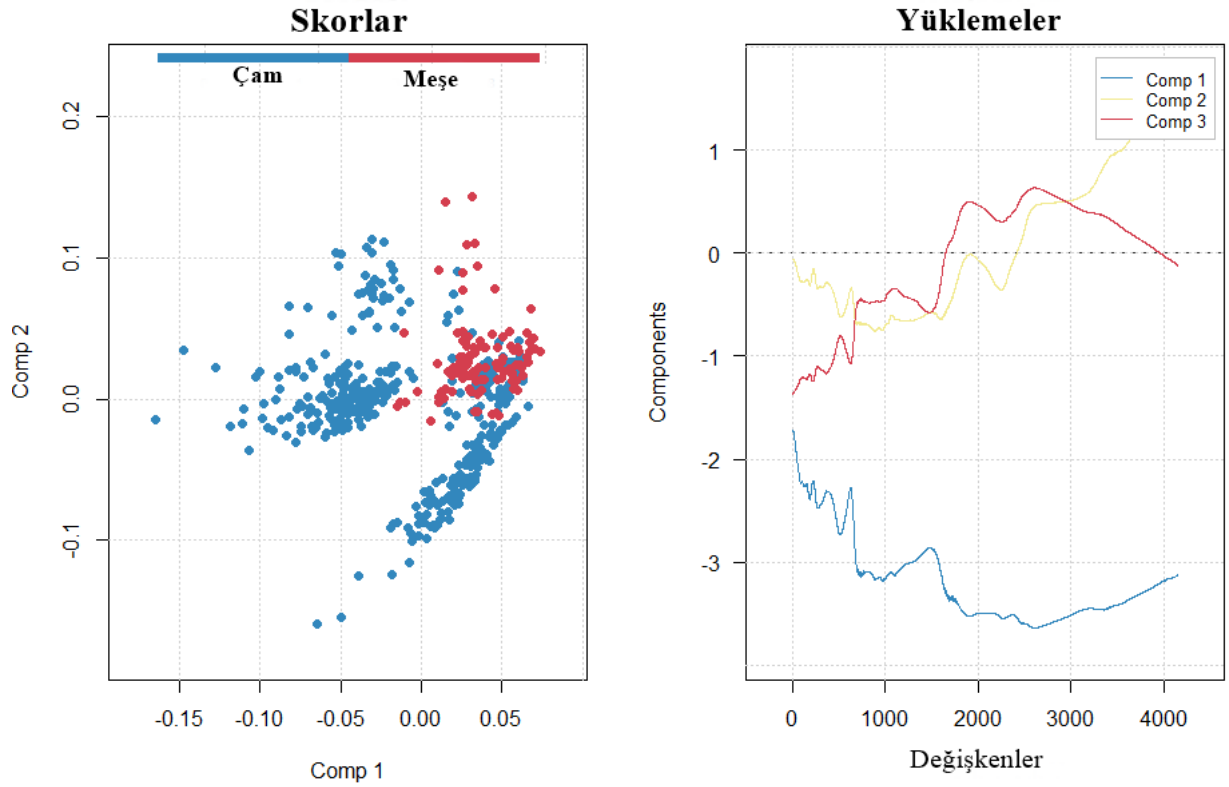
Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	38	13	39	15	0,750	0,717	0,733	0,267
SNV	48	1	51	5	0,981	0,906	0,943	0,057
MSC	53	50	2	0	0,038	1	0,524	0,476
SG 1. Dr	43	0	52	10	1,000	0,811	0,905	0,095
SG 1. Dr + SNV	45	0	52	8	1,000	0,849	0,924	0,076
SG 1. Dr + MSC	43	1	51	10	0,981	0,811	0,895	0,105
Sm	41	15	37	12	0,712	0,774	0,743	0,257
Sm + 1. Dr (w=3,p=1)	43	0	52	10	1	0,811	0,905	0,095
Sm + 2. Dr (w=5,p=2)	40	0	52	13	1	0,755	0,876	0,124

Tablo 4.31: Sapsız meşe ve Macar meşesi t türlerinin ayrımında SVM modelinin test veri seti ile değerlendirilmesi.

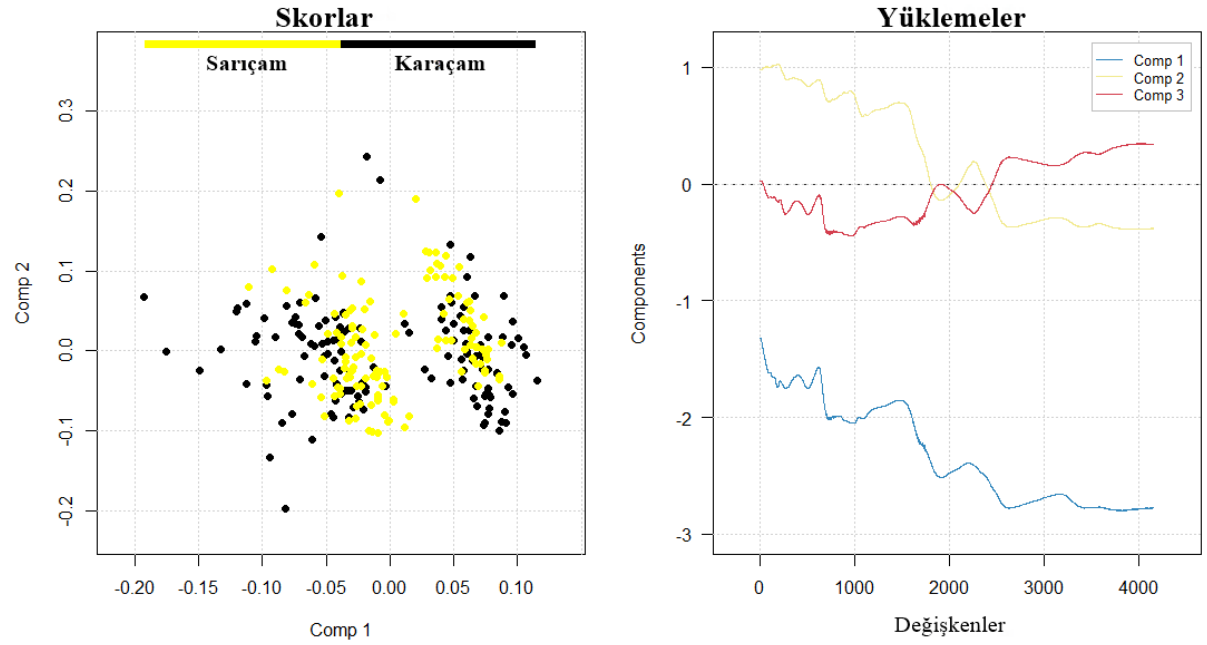
Ön İşlemler	DP	YP	DN	YN	Özgüllük	Duyarlılık	Doğru TO	Yanlış TO
Ham	45	6	42	3	0,875	0,938	0,906	0,094
SNV	48	4	44	0	0,917	1	0,958	0,042
MSC	13	0	48	35	1,000	0,271	0,635	0,365
SG 1. Dr	42	2	46	6	0,958	0,875	0,917	0,083
SG 1. Dr + SNV	45	0	48	3	1	0,938	0,969	0,031
SG 1. Dr + MSC	41	3	45	7	0,938	0,854	0,896	0,104
Sm	45	7	41	3	0,854	0,938	0,896	0,104
Sm + 1. Dr (w=3,p=1)	42	3	45	6	0,938	0,875	0,906	0,094
Sm + 2. Dr (w=5,p=2)	37	0	48	11	1	0,771	0,885	0,115

4.4.YORUMLAMA

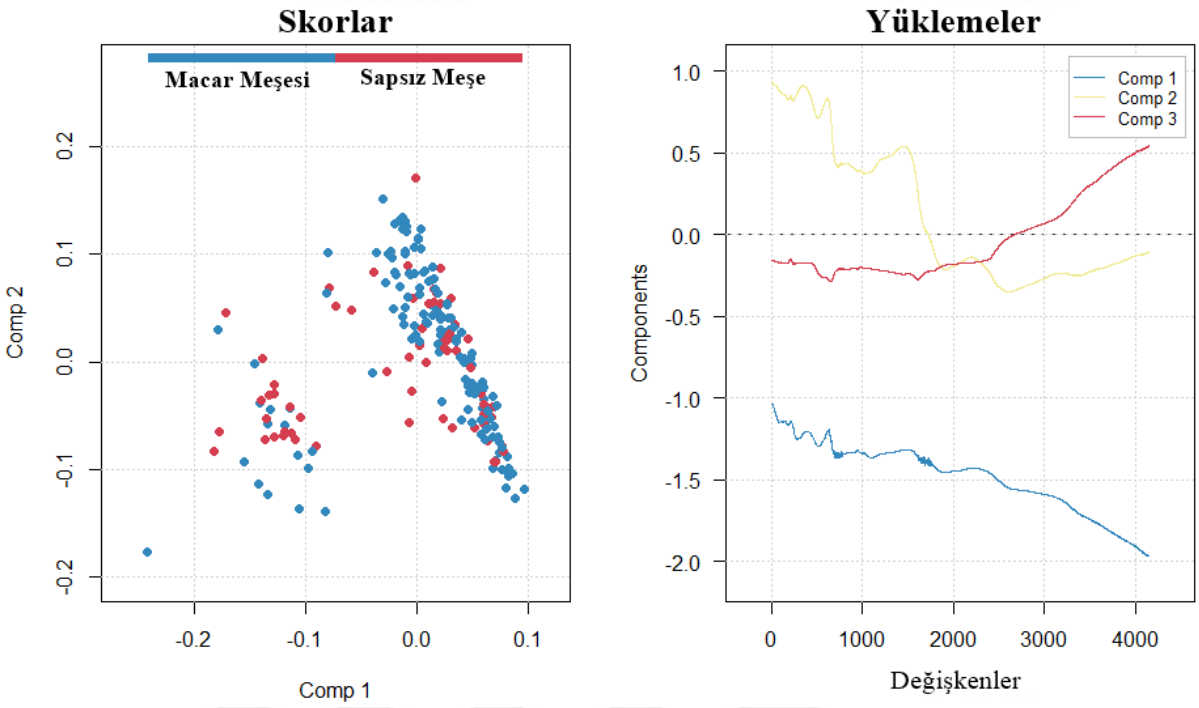
Çalışmada kullanılan modellerin doğrululuk oranları karşılaştırıldığında PLS-DA'nın oldukça başarılı olarak diğer modeller arasında öne çıktığı görülmüştür. Bu sebeple bu modelin daha ayrıntılı bir şekilde incelenmesine ve modellenen ilişkinin anlaşılması için kullanılan değişkenlerin belirlenmesine karar verilmiştir. İşlem görmemiş ham veri setlerinde cins bazında, çam türleri ve meşe türlerinin ayrımında kullanılan PLS-DA modelinde skor ve yükleme grafikleri sırasıyla Şekil 4.28, Şekil 4.29 ve Şekil 4.30 'da verilmiştir.



Şekil 4.28: Cins ayrımında PLS-DA skor ve yükleme grafikleri (ham veri seti).



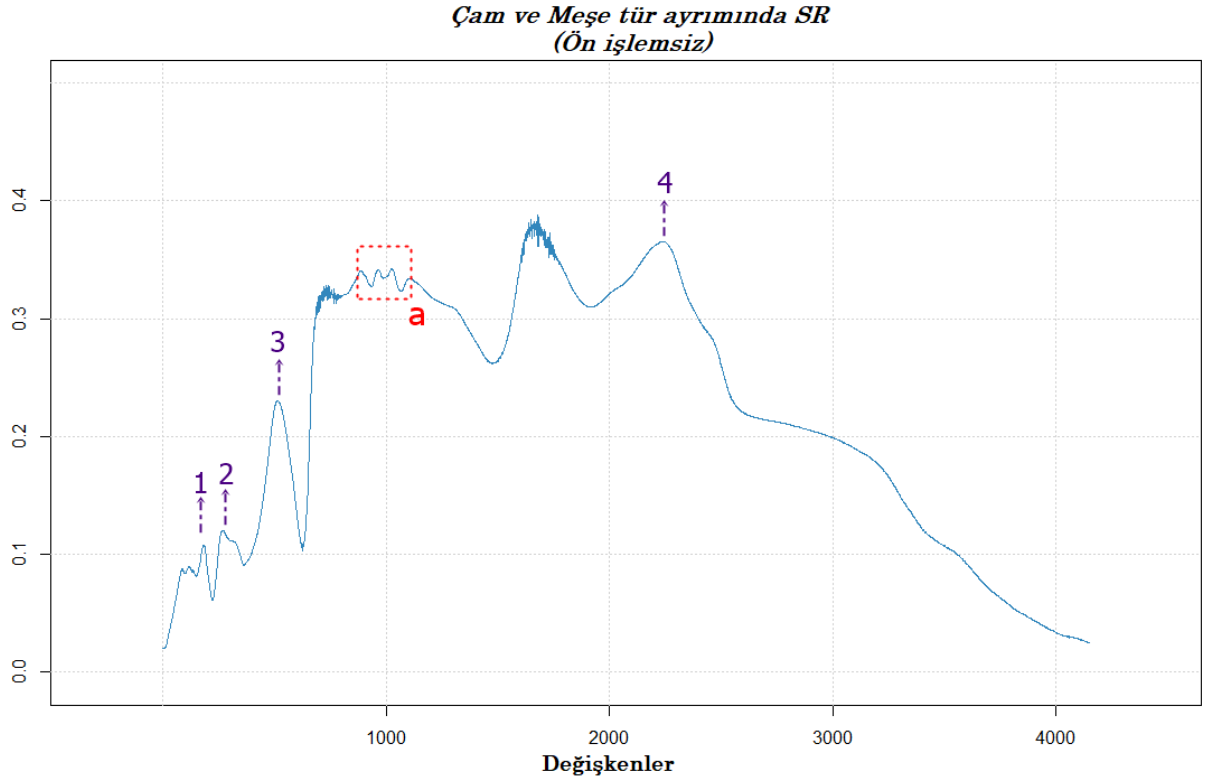
Şekil 4.29: Çam türlerinin ayrımında PLS-DA skor ve yükleme grafikleri (ham veri seti).



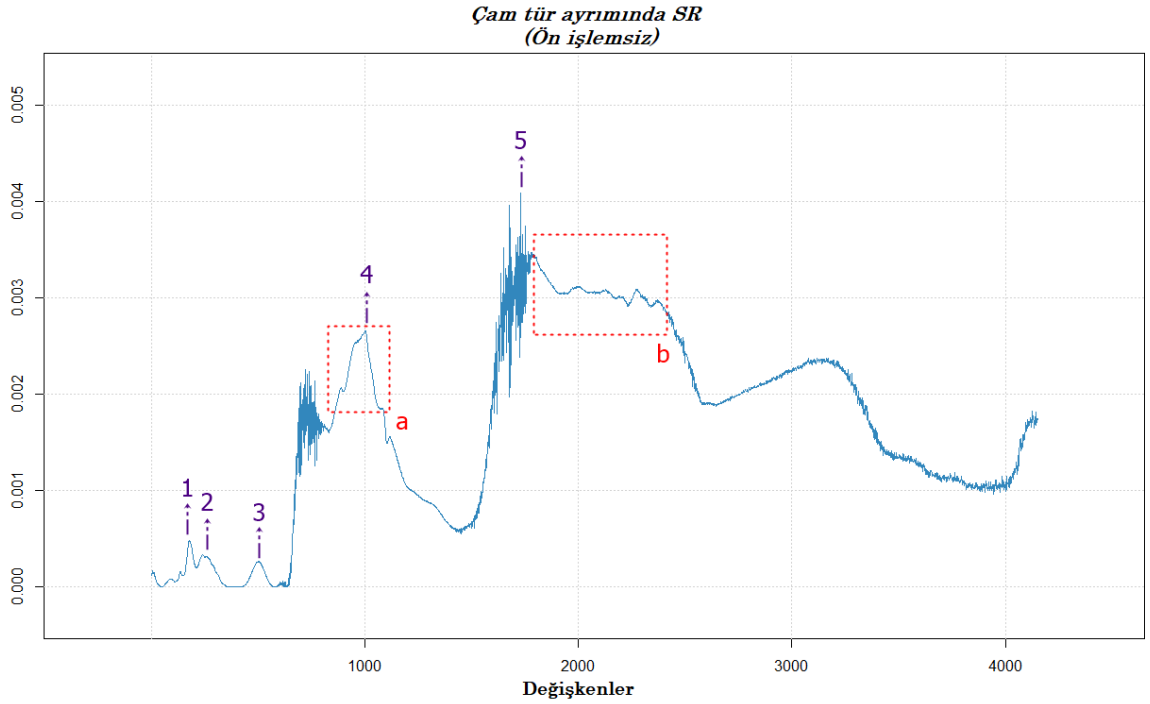
Şekil 4.30: Meşe türlerinin ayrımında PLS-DA skor ve yükleme grafikleri (ham veri seti).

PLS-DA'da kullanılan latent değişkenlerinin seçtiği dalga boylarının belirlenmesinde seçicilik oranlarına (SR) bakılmıştır. İşlem görmemiş ham veri setlerinde cins bazında, çam türleri ve meşe türlerinin ayrımında kullanılan değişkenleri gösteren SR grafikleri sırasıyla Şekil 4.31, Şekil 4.32 ve Şekil 4.33'de verilmiştir.

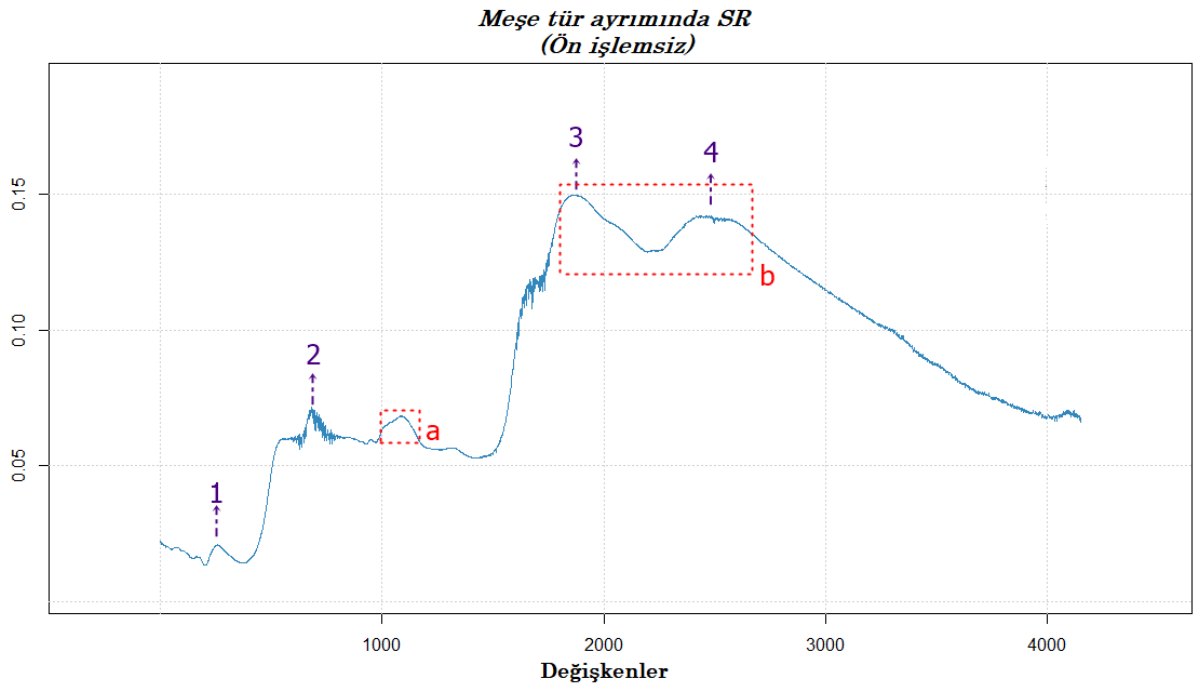
Cins bazındaki ayrımda PLS-DA'da kullanılan değişkenlerin seçicilik oranları (SR) büyükten küçüğe 8306 cm^{-1}, 5690 - 6000cm^{-1}, 5000cm^{-1}, 4517cm^{-1} ve 4345cm^{-1} olarak tespit edilmiştir. Çam türlerinin ayrımında kullanılan değişkenlerin SR sırasıyla 7444 – 8788 cm^{-1}, 7360cm^{-1}, 5932cm^{-1}, 5680 – 6100 cm^{-1}, 4336cm^{-1}, 4504cm^{-1} ve 4966cm^{-1} ; meşe türlerinin ayrımında PLS-DA'da kullanılan değişkenlerin SR 7614cm^{-1}, 8823cm^{-1}, 7445 – 9170 cm^{-1}, 5310cm^{-1}, 5930 – 6205 cm^{-1}, 4517cm^{-1} de en yüksek olduğu görülmüştür. Ayrıca, PLS-DA modellerinde latent değişkenlerin kullandığı ilk 20 dalga boyunun SR aracılığıyla hesaplanan önem düzeyi sıralamaları EK 13, EK 14 ve EK 15'de verilmiştir.



Şekil 4.31: Hiç işlem görmemiş çam ve meşe cinsleri için PLS-DA modelinde kullanılan değişkenler. a: 5690 – 6000 cm^{-1}, 1: 4345 cm^{-1}, 2: 4517 cm^{-1}, 3: 5000 cm-1, 4: 8306 cm^{-1}.

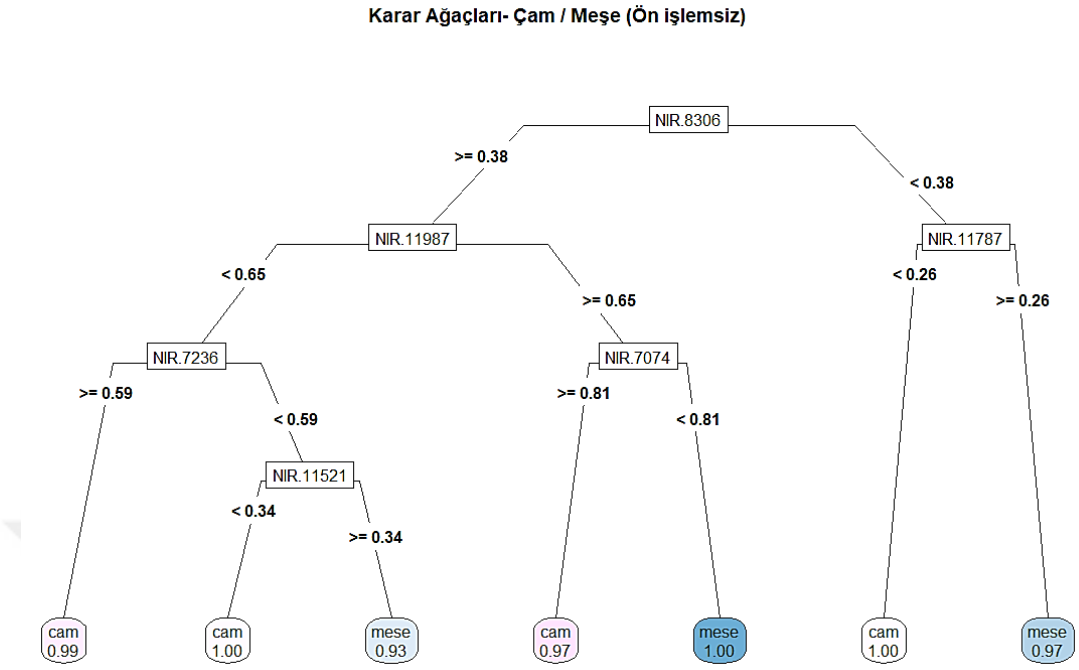


Şekil 4.32: Hiç işlem görmemiş çam türleri için PLS-DA modelinde kullanılan değişkenler. a: 5680 - 6100cm^{-1}, b:7444 - 8788cm^{-1}, 1: 4336 cm^{-1}, 2: 4504 cm^{-1}, 3: 4966cm^{-1}, 4: 5932cm^{-1}, 5: 7360cm^{-1}.

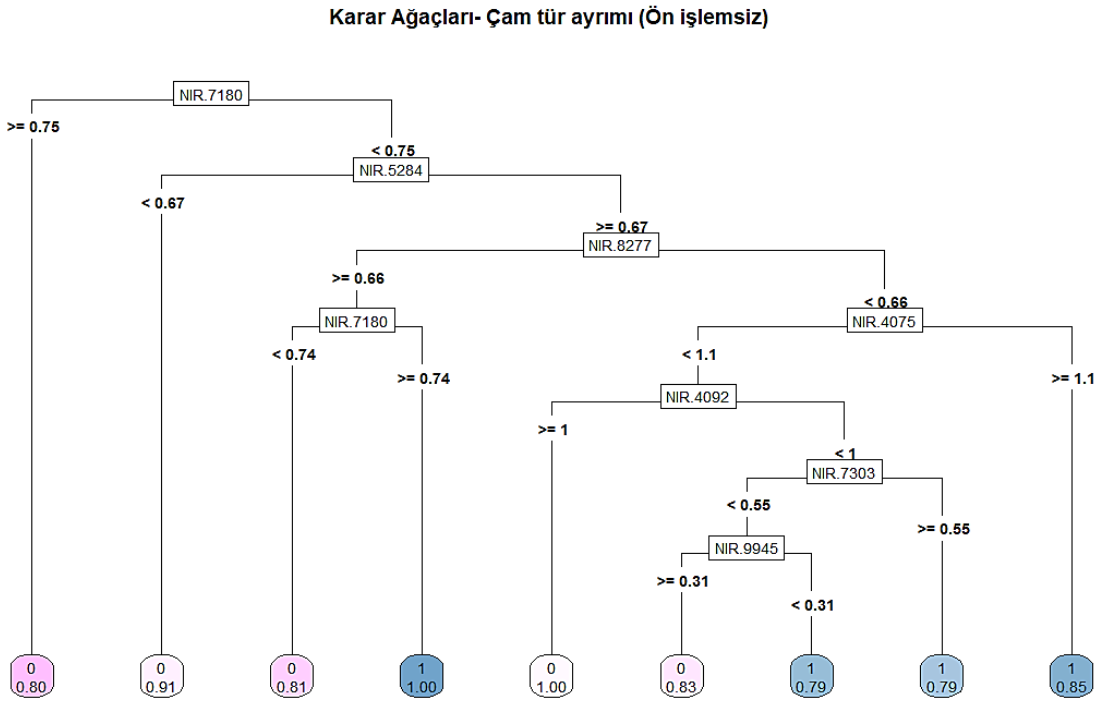


Şekil 4.33: Hiç işlem görmemiş meşe türleri için PLS-DA modelinde kullanılan değişkenler. a: 5930 – 6205 cm^{-1}, b: 7445 – 9170 cm^{-1}, 1: 4517 cm^{-1}, 2: 5310 cm^{-1}, 3: 7617 cm^{-1}, 4: 8823 cm^{-1}.

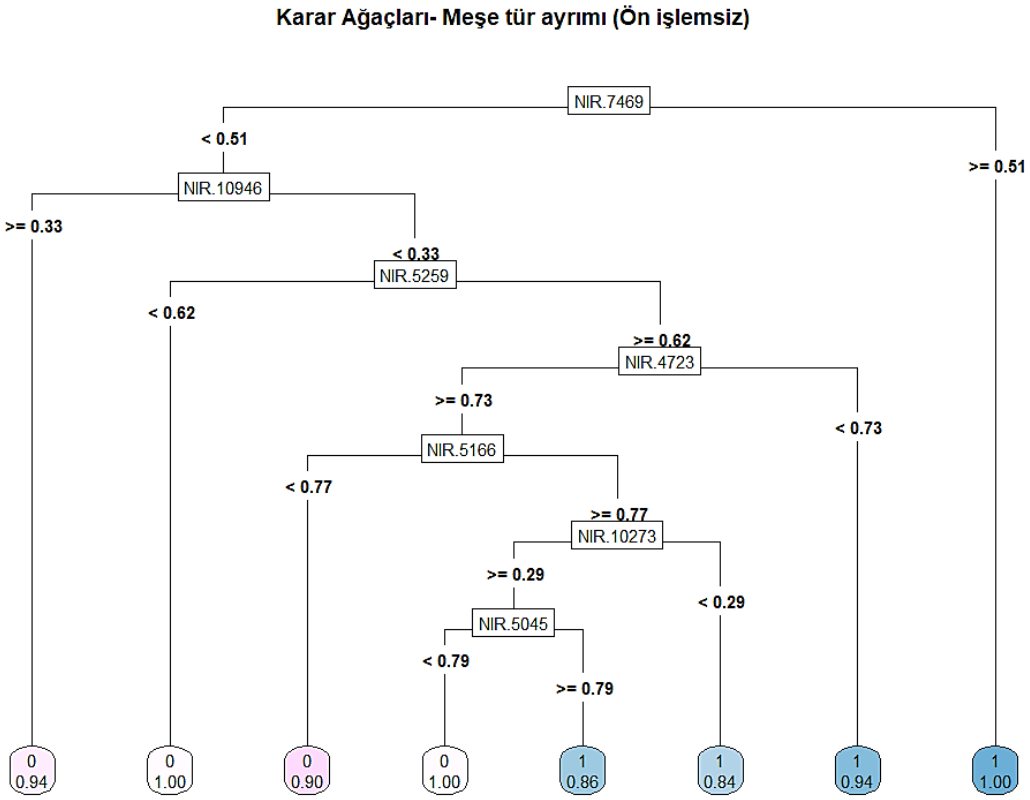
Gerçekleştirilen çalışmada ağaç türü teşhisi amacıyla NIRS 'dan elde edilen verilerin ilişkilendirilmesinde doğrusal olmayan modellerden DT, ANN ve SVM kullanılmıştır. Bu metotlardan ANN ve SVM, kurulan modellerde kullanılan ilişkilerin yorumlanması oldukça güç olan, bağımsız değişkenlerin her birinin bağımlı değişkenleri ne kadar etkilediğini bilemediğimiz, bu özelliklerinden dolayı da kapalı kutu / kara kutu olarak isimlendirilen modellerdir (Ciaburro ve Venkateswaran, 2017; Xiaobo ve ark., 2010). Bu sebeple çalışmada kullanılan doğrusal olmayan modellerden değişken seçiminin yorumlanması oldukça kolay olan DT aracılığıyla önemli kabul edilen değişkenler Şekil 4.34, Şekil 4.35 ve Şekil 4. 36'de doğru tahmin olasılıkları ile birlikte gösterilmiştir.



Şekil 4.34: Çam ve meşe cins ayrımında ön işlem görmemiş veri ile DT uygulamaları.



Şekil 4.35: Çam türlerinin ayrımında DT uygulamaları. 0: sarıçam, 1:karaçam.



Şekil 4.36: Meşe türlerinin ayrımında DT uygulamaları.0: macar meşesi, 1: sapsız meşe.

Şekillerden de anlaşılacağı gibi, belirli bir önem derecesine bağlı kalmaksızın DT aracılığıyla seçilen değişkenler gruplara göre şu şekildedir;

çam ve meşe cins ayrımında: 7074cm^{-1}, 7236cm^{-1}, 8306cm^{-1}, 11521cm^{-1}, 11787cm^{-1}, 11987cm^{-1} (Şekil 4.31);

çam türlerinin ayrımında: 4075cm^{-1}, 4092cm^{-1}, 5284cm^{-1}, 7180cm^{-1}, 7303cm^{-1}, 8277cm^{-1}, 9945cm^{-1}(Şekil 4.32);

meşe türlerinin ayrımında: 4723cm^{-1}, 5045cm^{-1}, 5166cm^{-1}, 5259cm^{-1}, 7469cm^{-1}, 10273cm^{-1}, 10946cm^{-1} (Şekil 4.33).

Ayrıca, DT modellerinde kullanılan dalga boylarının değişken önem sıralamaları EK 16'da verilmiştir.

5. TARTIŞMA VE SONUÇ

Genel olarak kurulan modellerin verileri iyi bir şekilde temsil etmesi istenilse de, modelin eğitim veri seti ile aşırı uyumlu olması (overfit), modelin genelleştirilmesine engel olması sebebiyle bir problem olarak karşımıza çıkmaktadır (James ve ark., 2013). Gerçekleştirilen denemelerde eğitim veri setleri ile kurulan modellerin aşırı uyumlu olma şüphesi doğrusal modellerden PLS-DA'da, doğrusal olmayan modellerden DT ve SVM'de ve cins bazında ayrım için ANN 'de karşımıza çıkmaktadır (Tablo 5.1 , Tablo 5.2 ve Tablo 5.3).

Tablo 5.1: Çam ve meşe türlerinin sınıflandırılmasında kurulan farklı model ve ön işlemlerin doğru tahmin oranları (eğitim seti).

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
Ham	1	0,73	0,98	0,99	1
SNV	1	0,97	0,99	0,99	1
MSC	1	0,97	1,00	0,996	1
SG 1. Dr	1	0,98	1,00	1	1
SG 1. Dr + SNV	1	1,00	1,00	1	1
SG 1. Dr + MSC	1	0,99	1,00	1	1
Sm	1	0,73	0,98	1	1
Sm 1. Dr	1	0,98	1,00	1	1
Sm 2. Dr	1	0,94	0,99	1	1

Tablo 5.2: Çam türlerinin sınıflandırılmasında kurulan farklı model ve ön işlemlerin doğru tahmin oranları (eğitim seti).

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
Ham	0,97	0,50	1	0,71	0,84
SNV	0,98	0,57	1	0,68	1
MSC	0,99	0,57	1	0,71	1
SG 1. Dr	1	0,78	1	0,57	1
SG 1. Dr + SNV	1	0,68	1	0,64	1
SG 1. Dr + MSC	1	0,67	1	0,73	1
Sm	0,88	0,50	1	0,76	0,82
Sm 1. Dr	1	0,78	1	0,69	1
Sm 2. Dr	1	0,77	1	0,88	1

Tablo 5.3: Meşe türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (eğitim seti).

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
Ham	1	0,55	1	0,85	0,995
SNV	1	0,76	1	0,84	1
MSC	1	0,75	1	0,89	1
SG 1. Dr	1	0,91	1	0,85	1
SG 1. Dr + SNV	1	0,90	1	0,86	1
SG 1. Dr + MSC	1	0,91	1	0,95	1
Sm	0,995	0,55	1	0,87	0,99
Sm 1. Dr	1	0,90	1	0,94	1
Sm 2. Dr	1	0,96	1	0,995	1

Çalışmada kullanılan modellerin eğitim veri setlerine fazla uyumlu olup olmadıkları test setleri ile gerçekleştirilen denemelerin başarısı ile ortaya koyulmaktadır. Eğitim veri seti ile çok uyumlu kurulan modeller, eğitime katılmayan yeni veri setleri (test seti) ile denendiğinde nispeten düşük tahmin oranlarına sahip oldukları görülmektedir. Çalışmada uygulanan modellerin yeni veri setleri ile denenmesi sonucu elde edilen doğruluk oranları Tablo 5.4, Tablo 5.5 ve Tablo 5.6 da verilmiştir. Bu tablolar aracılığıyla eğitim ve test setleri karşılaştırıldığında sadece DT ile kurulan bazı modellerde (ham, MSC, Sm gibi) eğitim veri setine çok uyumlu modellerin kurulduğu düşünülmektedir.

Tablo 5.4: Çam ve meşe türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (test seti).

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
Ham	0,99	0,74	0,94	0,995	1
SNV	0,99	0,99	0,95	0,995	0,985
MSC	1	0,95	0,995	0,99	0,980
SG 1. Dr	0,995	0,95	0,97	0,60	1
SG 1. Dr + SNV	1	0,97	0,96	0,95	1
SG 1. Dr + MSC	1	0,98	0,97	0,96	1
Sm	0,995	0,74	0,94	1	1
Sm 1. Dr	0,995	0,96	0,97	0,98	1
Sm 2. Dr	0,97	0,91	0,96	0,96	0,995

Tablo 5.5: Çam türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (test seti).

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
Ham	0,84	0,47	0,54	0,50	0,73
SNV	0,73	0,56	0,66	0,53	0,94
MSC	0,84	0,51	0,51	0,58	0,52
SG 1. Dr	0,93	0,71	0,84	0,58	0,91
SG 1. Dr + SNV	0,93	0,60	0,83	0,45	0,92
SG 1. Dr + MSC	0,94	0,63	0,83	0,51	0,90
Sm	0,77	0,47	0,54	0,51	0,74
Sm 1. Dr	0,96	0,73	0,71	0,60	0,91
Sm 2. Dr	0,91	0,66	0,83	0,75	0,88

Tablo 5.6: Meşe türlerinin sınıflandırılmasında kullanılan farklı model ve ön işlemlerin doğru tahmin oranları (test seti).

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
Ham	0,96	0,62	0,76	0,81	0,91
SNV	0,97	0,66	0,67	0,78	0,96
MSC	0,95	0,65	0,66	0,56	0,64
SG 1. Dr	0,94	0,68	0,72	0,54	0,92
SG 1. Dr + SNV	0,95	0,77	0,74	0,69	0,97
SG 1. Dr + MSC	0,91	0,70	0,69	0,69	0,90
Sm	0,96	0,62	0,74	0,83	0,90
Sm 1. Dr	0,98	0,71	0,67	0,64	0,91
Sm 2. Dr	0,91	0,70	0,64	0,71	0,89

Ön işlem metotlarının modellerin doğru tahmin oranlarına etkileri, ön işlem görmemiş veri setlerindeki doğru tahmin oranlarını iyileştirme yüzdelerine göre incelenmiştir (Tablo 5.7, Tablo 5.8, Tablo 5.9). Tablolar incelendiğinde bir ön işlem metodu aynı veri setinde kullanılırken dahi bir modelin doğru tahmin oranını arttırıcı etki gösterirken, başka bir modelin doğru tahmin oranını azaltabilmektedir. Örneğin Tablo 5.7 de SG 1. türeve bakıldığında DL-DA'nin doğru tahmin oranını % 28,38 oranında iyileştirirken, ANN'nın doğru tahmin oranını % 39,7 oranında azalttığı, SVM'nin doğru tahmin oranında ise hiçbir etkisinin olmadığı görülmektedir. Benzer şekilde bir ön işlem metodunun aynı model üzerindeki etkisine farklı bir veri seti ile bakıldığında bir veri setinin doğru tahmin oranını arttırırken diğer veri setinde

azaltıcı etki gösterebildiği görülmektedir. Örneğin Tablo 5.7, Tablo 5.8 ve Tablo 5.9 da ön işlemlerden SNV'nin PLS-DA üzerindeki etkisine bakıldığında çam ve meşe cins ayrımında hiçbir etkisinin olmadığı, çam türlerinin ayrımında % 13,10 oranında doğru tahmin oranını negatif yönde etkilediği, meşe türlerinin ayrımında ise doğru tahmin oranını % 1,04 oranında iyileştirdiği görülmektedir.

Tablo 5.7: Çam ve meşe türlerinin sınıflandırılmasında kullanılan ön işlemlerin doğru tahmin oranlarını iyileştirme yüzdeleri.

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
SNV	0,00	33,78	1,06	0,00	-1,50
MSC	1,01	28,38	5,85	-0,50	-2,00
SG 1. Dr	0,51	28,38	3,19	-39,70	0,00
SG 1. Dr + SNV	1,01	31,08	2,13	-4,52	0,00
SG 1. Dr + MSC	1,01	32,43	3,19	-3,52	0,00
Sm	0,51	0,00	0,00	0,50	0,00
Sm 1. Dr	0,51	29,73	3,19	-1,51	0,00
Sm 2. Dr	-2,02	22,97	2,13	-3,52	-0,50

Tablo 5.8: Çam türlerinin sınıflandırılmasında kullanılan ön işlemlerin doğru tahmin oranlarını iyileştirme yüzdeleri.

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
SNV	-13,10	19,15	22,22	6,00	28,77
MSC	0,00	8,51	-5,56	16,00	-28,77
SG 1. Dr	10,71	51,06	55,56	16,00	24,66
SG 1. Dr + SNV	10,71	27,66	53,70	-10,00	26,03
SG 1. Dr + MSC	11,90	34,04	53,70	2,00	23,29
Sm	-8,33	0,00	0,00	2,00	1,37
Sm 1. Dr	14,29	55,32	31,48	20,00	24,66
Sm 2. Dr	8,33	40,43	53,70	50,00	20,55

Tablo 5.9: Meşe türlerinin sınıflandırılmasında kullanılan ön işlemlerin doğru tahmin oranlarını iyileştirme yüzdeleri.

Ön İşlemler	PLS-DA	DL-DA	DT	ANN	SVM
SNV	-1,04	4,84	-13,16	-30,86	-29,67
MSC	-2,08	9,68	-5,26	-33,33	1,10
SG 1. Dr	-1,04	24,19	-2,63	-14,81	6,59
SG 1. Dr + SNV	-5,21	12,90	-9,21	-14,81	-1,10
SG 1. Dr + MSC	0,00	0,00	-2,63	2,47	-1,10
Sm	2,08	14,52	-11,84	-20,99	0,00
Sm 1. Dr	-5,21	12,90	-15,79	-12,35	-2,20
Sm 2. Dr	1,04	6,45	-11,84	-3,70	5,49

Yapılan çalışma sonucunda doğrusal modeller arasında en etkili sonuçların PLS-DA modelinde, doğrusal olmayan modeller arasındaki en etkili sonuçların SVM aracılığıyla elde edildiği görülmektedir. Literatürdeki sonuçlara bakıldığında PLS-DA kullanılarak ağaç türü teşhisinde geçekleştirilen bu çalışmaya paralel, oldukça başarılı sonuçlar elde edildiği görülmüştür. Flaete ve ark. (2006) üç farklı ladin türünü birinci türevlerini kullanarak PLS-DA ile sınıflandırmada % 94 doğruluğa ulaşmıştır. Tounis (2009) çam, ladin ve göknar türlerini sınıflandırdığı tez çalışmasında PLS-DA ile % 93,33 doğru tahmin değerine ulaşmıştır. Horikawa ve ark. (2015) ksilaryumdan alınan iki farklı çam türüne ait örneği düzgünleştirip ikinci türevleri alarak % 89,6 – 93,8 doğrulukla sınıflandırmıştır. Yang ve ark. (2015) üç farklı egzotik tür için PLS-DA aracılığıyla sınıflandırma modelleri oluşturmuş ve % 60, % 83 ve % 100 oranlarında ağaç türüne bağlı olarak değişken başarılar elde etmişlerdir. Bergo ve ark. (2016) hakiki maun (*Swietenia macrophylla* King) ve benzer görünüş özelliklerine sahip 3 farklı türde yaptıkları çalışmada düzeltilmiş, merkezileştirilmiş ve birinci türevi alınmış veri setleri ile ilgili 4 türü % 100 doğrulukta sınıflandırmışlardır. Hwang ve ark. (2016) çam türlerinde yaptıkları çalışmada ön işlem uygulanmamış örneklerde % 99,7 - 100, ikinci türevi alınmış örneklerde %100 oranında doğru tahmin değerlerine ulaşmışlardır. Lazarescu ve ark. (2017) PLS-DA aracılığıyla tsuga ve göknar türlerinin sınıflandırılmasına baktıkları çalışmada; ortalama merkezleme uyguladıkları göknarda % 90,3, tsugada % 94,3, birinci türevi alınan

göknarda % 88,2, tsugada % 99,4 oranlarında doğru tahminler elde etmişlerdir. Snel ve ark. (2018) CITES ile koruma altına alınmış bazı *Dalbergia* spp. türlerini ortalama merkezleme, düzgünleştirme ve birinci türevlerini alarak PLS-DA ile % 96,77 oranında doğru tahminle sınıflandırmışlardır. Leandro ve ark. (2019) altı farklı egzotik ağaç türünü ayırdıkları çalışmalarında Düzgünleştirme, SNV ve birinci türev alarak bu altı türün birbirinden % 88,89 oranında doğru tahmin ile ayrılabildiklerini belirtmişlerdir. Pace ve ark. (2019) on iki yerli ağaç türünü PLS-DA aracılığıyla sınıflandırdıkları çalışmada ön işlem uygulamadan kurdukları modelin % 91 oranında başarılı tahminde bulunduğunu; bu açıdan bakıldığında ağaç türü teşhisinde NIRS'in kullanımının sınıflandırma yeteneğinin oldukça yeterli olduğu sonucuna varmışlardır. Yong ve ark. (2019) farklı ön işlem metotlarının PLS-DA aracılığıyla sınıflandırmada doğru tanıma oranları; ön işlem uygulanmamış örneklerde % 88,2, düzgünleştirme ardından % 88,2, dalgacık dönüşümü ardından % 87,2, SNV ardından % 84,5, MSC ardından % 84,2 ve birinci türevi alındıktan sonra % 82,7 olarak tespit etmişlerdir. Zhou ve ark. (2020) göknar ve tsuga türlerinin PLS-LDA aracılığıyla sınıflandırılmasına dair yaptıkları çalışmada düzgünleştirme uygulanmış örneklerde % 100, düzgünleştirme ve SNV uygulanmış örneklerde ise % 98,75 oranlarında doğru tahmine sahip olduklarını göstermişlerdir. Son 15 yılda yapılan tüm bu çalışmalara bakıldığında PLS-DA aracılığıyla sınıflandırmada doğru tahmin oranları % 82,7 ile % 100 arasında değişiklik gösterdiği tespit edilmiştir. Bu doğruluk oranlarının da bazı durumlarda uygulanan ön işlemler veya bu ön işlemlerin kombinasyonları ile iyileştirilebildiği görülmektedir. Gerçekleştirilen bu çalışmada PLS-DA aracılığıyla en yüksek doğru tahminin düzgünleştirilmiş ve birinci türevi alınmış veri setlerinde (çam türlerinde % 96, meşe türlerinde % 98) elde edildiği belirlenmiştir. Doğru tahmin oranlarında görülen bu değişiklik ağaç türlerine ve uygulanan ön işlem metotlarına bağlı olmakla birlikte son 15 yılda yapılan çalışmalar ve tarafımızca gerçekleştirilen bu çalışma sonucunda görülmüştür ki; farklı ağaç türlerinin NIRS aracılığıyla tespit edilmesi ve özellikle PLS-DA aracılığıyla sınıflandırılması oldukça başarılıdır.

Literatüre bakıldığında genel olarak yorumlanması nispeten daha kolay olan doğrusal modellerin (PLS-DA, PLS-R gibi) tercih edildiği görülmektedir. Son yıllarda bu tercihlerin yanı sıra, kapalı kutu olarak yorumlanan doğrusal olmayan modellerin ağaç türü teşhisinde kullanımı da denenmektedir. Ağaç türü teşhisinde NIRS ile kullanılan doğrusal olmayan yöntemlerin başında ise ANN ve SVM metotları gelmektedir. Lazarescu ve ark. (2017) doğrusal olmayan modellerden ANN kullanılarak tsuga ve göknar türlerinin sınıflandırılmasına baktıkları

çalışmada; ortalama merkezleme uyguladıkları göknarda % 89,8, tsugada % 99,1, birinci türevi alınan göknarda % 86,7, tsugada % 98,9 oranlarında doğru tahminler elde etmişlerdir. Nisgoski ve ark. (2017b) dört farklı egzotik türü ANN aracılığıyla ayırdıkları çalışmada ön işlem görmemiş örneklerle % 100, ikinci türevi alınmış örneklerle % 76.07, MSC uygulanmış örneklerde % 99,21 oranlarında başarılı bir sınıflandırma yapmışlardır. Jin ve ark. (2017), üç farklı *Miscanthus* türünü ANN aracılığıyla sınıflandırdıkları çalışmada % 99,42 lik bir doğru sınıflandırma oranına ulaşmışlardır. Jin ve ark. (2017) tarafından gerçekleştirilen aynı çalışmada iki farklı SVM yaklaşımı da aynı ağaç türleri için denemiş ve ANN ile aynı doğru tahmin oranlarına ulaşılmışlardır (% 99,42). Çalışma sonunda aynı doğrululuk oranlarına sahip olmalarına rağmen daha hızlı sonuçlar alınabildiği için SVM metodunu önermişlerdir. Li ve ark. (2019) 3 farklı ağaç türünün (melez, karaağaç, kavak) sınıflandırılmasında SVM metodunu uyguladıkları çalışmalarında, ön işlem uygulanmamış veri seti ile % 84,7, dalgacık katsayısı (wavelet coefficients) dönüşümü ile ön işlem uygulanmış veri setlerinde % 98,6 oranlarında doğru tahmin değerlerine ulaşmışlardır. Zhou ve ark. (2020) göknar ve tsuga türlerinin SVM aracılığıyla sınıflandırılmasına dair yaptıkları çalışmada düzgünleştirme ve düzgünleştirme ile SNV uygulayarak bu iki ağaç türünü % 100 oranında doğru tahminlerini elde etmişlerdir. Doğrusal olmayan bu modellerdeki doğrululuk oranlarının % 76,07 ile % 100 arasında değişen değerlerde seyrettiği görülmektedir. Tarafımızca gerçekleştirilen bu çalışmada farklı ön işlem uygulamaları ile ANN uygulandığında çam türlerinde % 45 – 75, meşe türlerinde % 54 – 83 aralığında doğru sınıflandırma oranlarına ulaşılmıştır. Yine farklı ön işlem uygulamaları ile SVM uygulandığında ise çam türlerinde % 52 – 94, meşe türlerinde % 64 – 97 aralığında doğru sınıflandırma oranlarına ulaşılmıştır. Literatür ile karşılaştırıldığında bu değerlerin özellikle ANN'nin doğru tahmin oranlarının nispeten düşük olduğu görülmektedir. Bu durum oldukça karmaşık sayılabilecek bu makine öğrenmesi tekniklerinin kullanımında model optimizasyonlarının ve parametre tahminlerinin öneminden kaynaklandığı düşünülmektedir. Bunun yanı sıra doğrusal olmayan bir yaklaşım olan ANN'nin doğrusal bir yaklaşıma sahip PCA ile kombinasyonunun başarılı sonuç vermediği söylenebilir.

Modellenen ilişkinin anlaşılması ve ihtiyaç duyulduğu takdirde verinin boyutunun yeni bileşenlere indirgenmesi için orijinal değişkenlerin seçimi şarttır (Kvalheim, 2020; Mehmood ve ark., 2020; Xiaobo ve ark., 2010). Modellenen ilişkinin anlamlandırılmasında doğrusal modeller ve doğrusal olmayan modellerin seçimlerine dair birer örnekle açıklanması yaklaşımı izlenmiştir. Doğrusal modellerden PLS-DA da kullanılan latent değişkenler seçicilik oranları

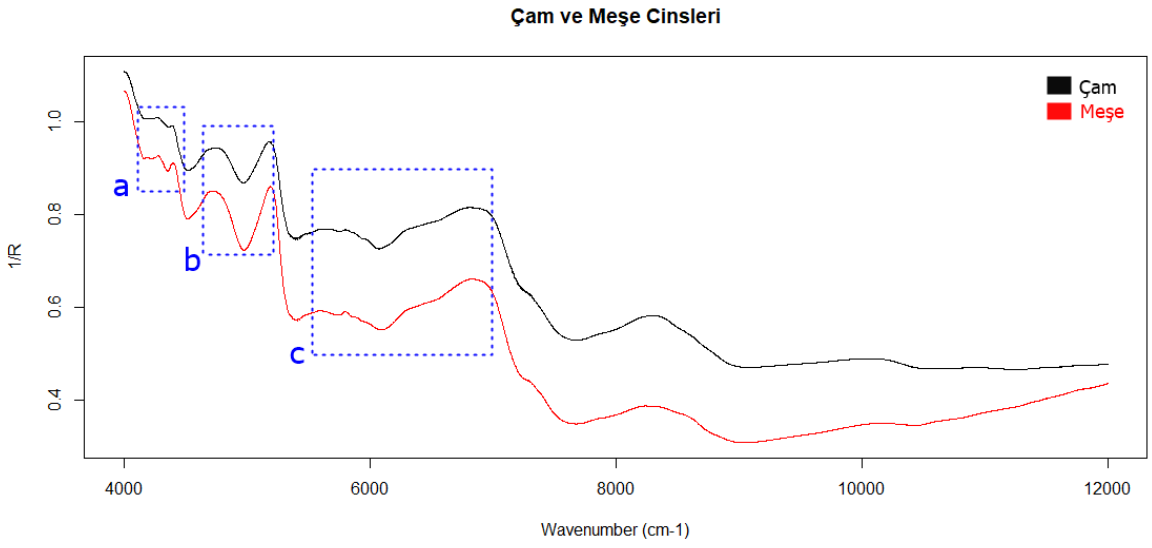
(SR) aracılığıyla belirlenmiş, doğrusal olmayan modeller ise karar ağaçları (DT) aracılığıyla tespit edilmiştir. Çalışmada ilgili yöntemlerle önemli kabul edildiği tespit edilen dalga numaraları ve ortalama spektrumlarda farklılıkların görüldüğü alanlardaki dalga numaraları Tablo 5.10 'da verilmiştir.

Tablo 5.10: PLS-DA ve DT aracılığıyla önemli kabul edildiği tespit edilen bantlar / dalga numaraları (cm^{-1}).

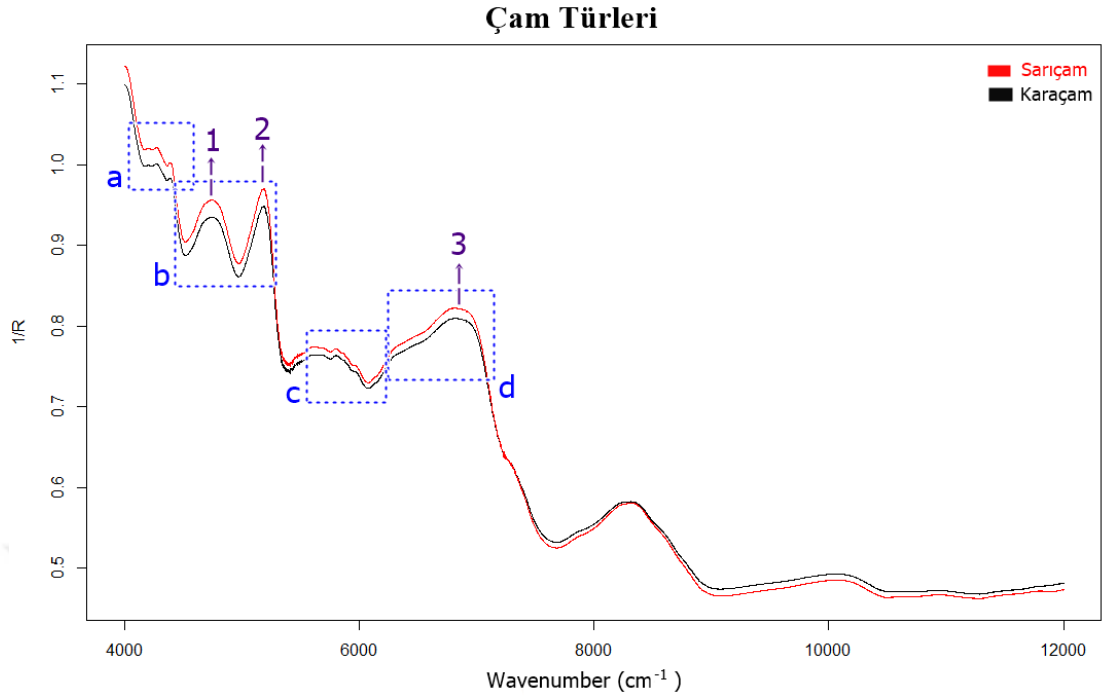
CİNS			ÇAM			MEŞE		
Ortalama	**PLS-DA**	**DT**	**Ortalama**	**PLS-DA**	**DT**	**Ortalama**	**PLS-DA**	**DT**
4180 – 4450	4345	-	4167 - 4400	4336	-	4180 - 4420	-	-
4570 – 5255	4517 5000	-	4500 – 5267	4504	5284	4540 – 5200	4517	4723 5045 5166
5500 – 7105	5690 – 6000	7074	5500 – 6000	5680 - 6100 5932	-	5380 – 6570	5930 – 6205	-
-	8306	8306	-	7444 - 8788	8277	-	7445 – 9170 7617 8823	7469
Ortak olmayanlar			***Ortak olmayanlar***			***Ortak olmayanlar***		
		7236	4733	7360	4075		5310	5259
		11521	5167		4092			10273
		11787	6833		7180			10946
		11987	6267 -7000		7303			
					9945			

Ortalama spektrumlara bakıldığında (Şekil 5.1, Şekil 5.2, Şekil 5.3) ağaç türleri ve cinsler arasında gözle görülür bir şekilsel farklılık tespit edilememektedir. Görsel olarak birbirine benzeyen bu türlere ait spektrumlar karşılaştırıldığında, farklılıkların absorpsiyon miktarlarından kaynaklandığı görülmektedir. Bu absorpsiyon farklılıklarının görüldüğü alanlar ile PLS-DA ve DT modelleri için önemli kabul edilen değişkenlerin bulunduğu alanlar karşılaştırıldığında birbirleri ile oldukça uyumlu oldukları görülmektedir (4 000 – 9 000 cm^{-1}) (Tablo 5.10). Absorpsiyon farklılıkları cins bazında karşılaştırıldığında (Şekil 5.1) oldukça belirgin bir şekilde bu farklılıkların tespiti mümkün olmaktadır. Diğer yandan, odun dokularının yanı sıra mikroskobik özellikleri dahi birbirine çok benzeyen sarıçam ve karaçam türlerinde bu farklılığın oldukça minimal düzeyde kaldığı görülmektedir (Şekil 5.2). Meşe türlerinin ayrımına bakıldığında ise absorpsiyon farklılıklarının cins ayrımından daha az fakat çam türlerinden daha fazla olduğu görülmektedir (Şekil 5.3).

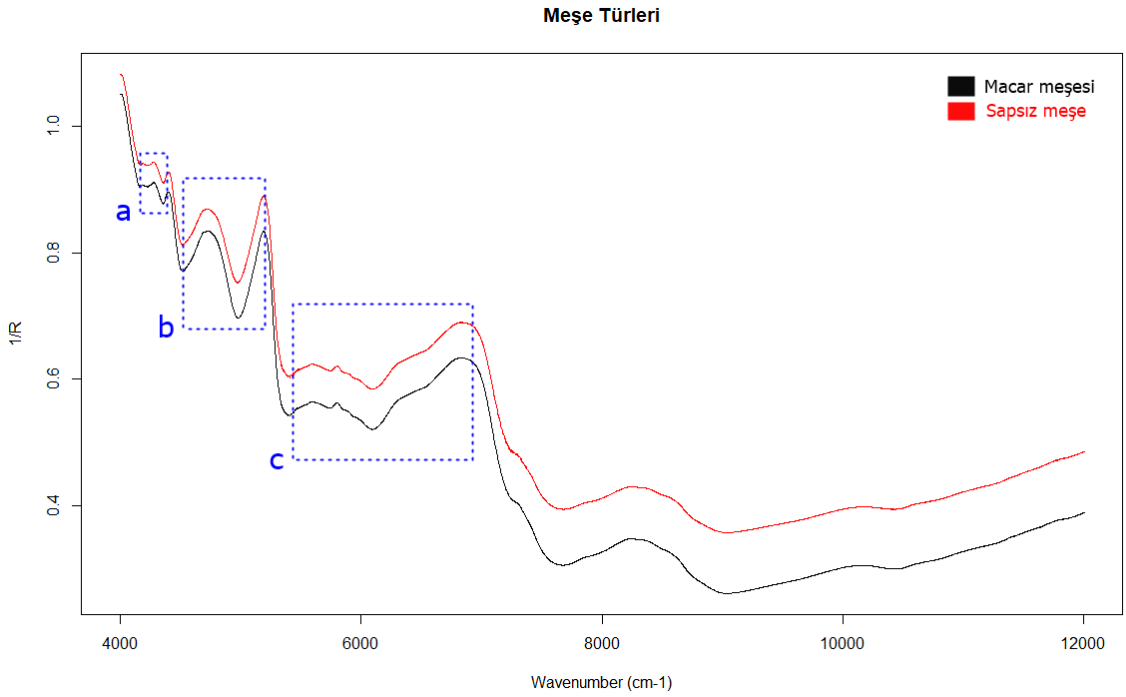
Absorpsiyon düzeylerinde tespit edilen bu farklılıkların iğne yapraklı (çam türleri gibi) ve geniş yapraklı (meşe türleri gibi) ağaçların kimyasal yapılarındaki nitel ve nicel farklılıklardan (özellikle lignin ve polyoz/hemiselüloz bileşenlerinden) kaynaklandığı bilinmektedir. Geniş yapraklı ağaçlar, daha fazla miktarda polyoz içerirken, iğne yapraklı ağaçlarda daha fazla miktarda lignin bulunmaktadır. Öte yandan iğne yapraklı ve geniş yapraklı ağaçların yapılarında bulunan polyoz ve lignin bileşenlerinin kimyasal yapıları birbirlerinden oldukça faklıdır. Odun karbonhidratlarından polyozlar, (6C-glukoz, mannoz, galaktoz, 5C- ksiloz, arabinoz), selülozdan daha kısa molekül zincirlerine sahiptir ve yan gruplar içererek dallanma göstermektedirler. İğne yapraklı ağaçlar yüksek oranda mannoz birimleri içerirken; geniş yapraklı ağaçlarda ise yüksek oranda ksiloz birimleri ve iğne yapraklı ağaçlara göre daha fazla oranda asetil grupları bulunmaktadır. Lignin ise fenilpropan birimlerinden oluşmuş aromatik bir polimerdir. İğne yapraklı ağaçlarda ve geniş yapraklı ağaçlarda bulunan ligninler (guayisil, siringil ve p- hidroksifenil) içerikleri açısından önemli farklılıklar göstermektedirler. Geniş yapraklı ağaçlarda bulunan ligninin daha farklı fonksiyonel grupları olsa da kızılötesi absorpsiyonlarının iğne yapraklı ağaçlardan daha düşük olduğu gözlenmektedir (Fengel ve Wegener, 1983). Bahsi geçen bu yapısal farklılıkları, farklı fonksiyonel grupların farklı dalga boylarında görülmesi ile açıklanması mümkündür. Bu bilgiler ışığında cins ayrımında belirgin farklılıkların farklı dalga boyları boyunca görülmesi, cins içi türlerin ayrımında ise farklılıkların absorpsiyon değerleri arasında görülmesi beklenmektedir.



Şekil 5.1: Çam ve meşe cinsleri için ortalama spektrumlar. a: 4180 – 4450 cm^{-1}, b: 4570 – 5255 cm^{-1}, c: 5500 – 7105 cm^{-1}.



Şekil 5.2: Çam türleri için ortalama spektrumlar. a: 4167 - 4400 cm^{-1}, b: 4500 – 5267 cm^{-1}, c: 5500 – 6000 cm^{-1}, d: 6267 -7000 cm^{-1}, 1:4733 cm^{-1} , 2:5167 cm^{-1} , 3: 6833 cm^{-1} .



Şekil 5.3: Meşe türleri için ortalama spektrumlar. a: 4180 – 4420 cm^{-1}, b: 4540 – 5200 cm^{-1}, c: 5380 – 6570 cm^{-1}.

Bantların hangi kimyasallar veya bağlar ile ilgili olduklarına dair atamaları kızılötesi spektroskopisinin başlangıcından beri hesaplansa da, basit moleküller için dahi bantların sayısı ve yoğunlukları oldukça değişkenlik gösterdiğinden oldukça zor bir aşamadır. Odun gibi karmaşık materyallerde hatta odun bileşenlerinde (lignin, selüloz vb.) bu tarz hesaplamalar çok daha karmaşık ve uygulaması oldukça güç kabul edilmektedir. Bu durum lignin gibi moleküllerin daha karmaşık moleküler yapılarının yanı sıra basit bir fonksiyonel grubun dahi (metil veya hidroksil) moleküler ortamı oldukça değiştirmesinden de kaynaklanmaktadır. Örneğin metil CH geriliminin 1.overtonunda 4 bant ile görülmekte, hidroksil grupları hidrojen bağları ile ilişkili olduğundan selüloz bantlarını etkilemekte (OH geriliminin 1.overtonunda), tüm bunların yanı sıra suyun hidrojen bağları NIR spektrumunda birden fazla alanda yer bulmaktadır (Burns ve Ciurczak, 2008; Schwanninger ve ark., 2011; Xiaobo ve ark., 2010). Bu sebeple yakın kızılötesi spektrumunda bantların ilgili oldukları kimyasal bileşenler yerine bu alanlardaki bağlar ve titreşimler verilmiştir (Tablo 5.11).

Tablo 5.11: Literatüre göre band atamaları (Haaland ve Thomas, 1988; Schwanninger ve ark., 2011; Xiaobo ve ark., 2010).

Dalga numarası (cm^{-1})	**İlgili Bağlar**
12 000 – 7 000	1.overtone + 2. overtone OH, NH gerilmesi 1. overtone CH kombinasyonu 2.overtone + 3. overtone CH, $_{Ar}$CH, NH gerilmesi
9 000 – 6 200	2. overtone bölgesi: H_2O, RCH, $_{Ar}$OH, $CCNH_2$, RNH, CONHR, CH, CH_2, CH_3
7 000 – 6 000	1. overtone OH (karbonhidratlar), OH, NH gerilmesi
6 500 – 5 000	1.overtone bölgesi : H_2O, RCO_2H, RCOR, $_{Ar}$CH, SH, $CONH_2$, CH, CH_2, CH_3 1. overtone CH gerilmesi, OH kombinasyonu (selüloz, hemiselüloz, lignin, ekstraktif)
~5 000 (5200 – 4900)	Su
5 000 – 4 000	Kombinasyonlar bölgesi NH, H_2O, ROH, RNH_2, $CONH_2$, CHO, CH, CH_2, CH_3, $CONH_2$ OH deformasyon ve gerilmesinin kombinasyonu (selüloz, hemiselüloz, lignin)

Ağaç türlerinin ayrımına dair gerçekleştirilen yakın kızılötesi çalışmalarına bakıldığında genel olarak tür bazında ayrımlar için 4 000 – 7 000 cm^{-1} aralığının kullanıldığı ve önemli kabul edilen dalga boylarının ilgili ağaç türüne bağlı olarak değişiklik gösterdiği görülmektedir (Tablo 5.12). Bu durum 12 000 – 7000 cm^{-1} arasının genel olarak nicel amaçla kalibrasyon modellerinin kurulmasında kullanılan bir alan olması ile ilişkilendirilmiştir. Bu alanın partikül büyüklüğü ve renk gibi fiziksel değişimlerden etkilenen kısım olması sebebiyle, termal işlem görmüş odunlar gibi çalışma konularında tercih edilmektedir (Schwanninger ve ark., 2004; Bächle ve ark., 2010; Schwanninger ve ark., 2011; Bächle ve ark., 2012; Popescu ve ark., 2018).

Tablo 5.12: Tür bazında ayrımda en yüksek ayırıcı güce sahip dalgaların konumları.

Literatür	***Dalga numarası (cm^{-1})***
Gierlinger ve ark. (2004)	6700 - 6300
Braga ve ark. (2011)	8995 – 7498
	8246 – 7498
	7502 – 4247
Pastore ve ark. (2011)	4249 – 7501
	5448 – 6100
	5448 – 5774
	4249 – 4600
Espinoza ve ark. (2012)	5405 – 4000
Shou ve ark. (2014)	9090 – 6250
	9090 – 7142
	10000 – 6060
Lang ve ark. (2015)	4826 – 4103
Snel ve ark. (2018)	6097 – 5682
	5348 – 5236
	4808 - 4630

NIRS yönteminin ağaç türü teşhisinde kullanım etkinliğinin belirlenmesi için gerçekleştirilen bu çalışmanın sonucunda elde edilen bulguları şu şekilde özetlemek mümkündür:

- NIRS'dan elde edilen veriler ile ağaç türleri arasında anlamlı bir ilişki vardır.
- NIR Spektroskopisinin, çok değişkenli analiz yöntemleri ile birlikte bazı yerli ağaç türlerinin ayrımında kullanımı başarılıdır.
- NIRS, Macar meşesi ve sapsız meşe ile sarıçam ve karaçam türlerinin ayrımında kullanılabilir bir metottur.

- Uygulanan modeller arasında en yüksek doğru tahmin oranına sahip model PLS-DA ’dır.
    - Denenen doğrusal modeller arasında en etkili sonuçlar PLS-DA ‘dan elde edilmiştir.
    - Denenen doğrusal olmayan modeller arasında en etkili sonuçlar SVM ’den elde edilmiştir.
- Denenen ön işlem metotları modellerin doğru tahmin oranları üzerinde etkilidir.
    - Bu etkinin her zaman olumlu olmadığı belirlenmiştir.
    - İlgili ağaç türleri ve en etkili modeller için hiçbir ön işlem metodu uygulanmadan da yüksek başarı oranları elde edilmiştir.
    - Her model ve veri seti için geçerli bir en iyi ön işlem metodu tespit edilememiştir.

Yukarıda bahsedilen sonuçlar ışığında çalışmanın geliştirilmesi için gelecekte yapılması önerilen araştırmalar şu şekildedir;

- Diğer yerli ağaç türleri içinde cins bazında teşhislerin gerçekleştirilmesi (PLS-DA),
- PLS-DA için elde edilen NIRS sonuçlarının kullanıcı dostu bir uygulamaya taşınarak, uygulamaya kullanılan algoritmaların gömülmesi ile çalışma sonuçlarının pratikte kullanılabilirliğinin sağlanması,
- Daha kısıtlı dalga aralığında/daha az değişkenle çalışan portatif sistemlerin ayrım başarılarının denenmesi.

KAYNAKLAR

Cochrane, C., 2019, Time Series Nested Cross-Validation, https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9, [Ziyaret tarihi: 17 Eylül 2019].

Adedipe, O. E., Dawson-Andoh A, B., Slahor, J. ve Osborn A, L., 2008, Classification of red oak (*Quercus rubra*) and white oak (*Quercus alba*) wood using a near infrared spectrometer and soft independent modelling of class analogies, *Journal of Near Infrared Spectroscopy*, 16(1), 49–57. https://doi.org/10.1255/jnirs.760

Bächle, H., Zimmer, B. ve Wegener, G., 2012, Classification of thermally modified wood by FT-NIR spectroscopy and SIMCA, *Wood Science and Technology*, 1181–1192. https://doi.org/10.1007/s00226-012-0481-z

Bächle, H., Zimmer, B., Windeisen, E. ve Wegener, G., 2010, Evaluation of thermally modified beech and spruce wood and their properties by FT-NIR spectroscopy, *Wood Science and Technology*, 421–433. https://doi.org/10.1007/s00226-010-0361-3

Baier, S., 2012, What is Full Spectrum Lighting?, http://lumenistics.com/what-is-full-spectrum-lighting/ , [Ziyaret tarihi: 21 Haziran 2018].

Balabin, R. M. ve Smirnov, S.V., 2011, Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data, *Analytica Chimica Acta*, 692, 63–72. doi:10.1016/j.aca.2011.03.006

Banks, D., House, L., McMorris, F. R., Arabie, P. ve Gaul, W., 2004, *Classification, Clustering, and Data Mining Applications*, Springer, Berlin Heidelberg, ISBN: 3-540-22014-3.

Barker, M. ve Rayens, W., 2003, Partial least squares for discrimination, *Journal of Chemometrics*, 17(3), 166–173. https://doi.org/10.1002/cem.785

Barnes, R. J., Dhanoa, M. S. ve Lister, S. J., 1989, Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772–777.

Bergo, M. C. J., Pastore, T. C. M., Coradin, V. T. R., Wiedenhoeft, A. C. ve Braga, J. W. B., 2016, Nirs Identification of Swietenia Macrophylla Is Robust Across Specimens From 27 Countries, *IAWA Journal*, 37(3), 420–430. https://doi.org/10.1163/22941932-20160144

Berkel, A. ve Bozkurt Y., 1961, *Türkiyenin önemli bazı meşe türleri odunlarının makroskopik ve mikroskobik çzellikleri hakkında araştırmalar,* İstanbul Üniveristesi Orman Fakültesi Yayınları, İ.Ü. Yayın no:922, O.F. Yayın No:73.

Bolzon de Muñiz, G. I., Carneiro, M. E., Ribeiro Batista, F. R., Zatt Schardosin, F. ve Nisgoski, S., 2016, Wood and charcoal identification of five species from the miscellaneous group known in Brazil as “Angelim” by Near-IR and wood anatomy, *Maderas. Ciencia y tecnología*, 18(3), 505–522. https://doi.org/10.4067/s0718-221x2016005000045

Bozkurt, Y. ve Erdin, N., 2000, Odun Anatomisi, Bozkurt, A. Y. Erdin, N., 2000, Odun Anatomisi, İstanbul Üniversitesi Orman Fakültesi Yayınları,İstanbul, ISBN 975-404-592-5.

Braga, J. W. B., Pastore, T. C. M., Coradin, V. T. R., Camargos, J. A. A. ve Da Silva, A. R. 2011, the Use of Near Infrared Spectroscopy To Identify Solid Wood Specimens of *Swietenia macrophylla*(CITES Appendix II), *IAWA Journal*, 32(2), 285–296.

Brereton, R. G., 2007, *Applied Chemometrics for Scientists*, John Wiley& Sons, West Sussex, UK, ISBN-10: 0-470-01686-8.

Brereton, R. G. ve Lloyd, G. R., 2014, Partial least squares discriminant analysis: Taking the magic away, *Journal of Chemometrics*, 28(4), 213–225. https://doi.org/10.1002/cem.2609

Brunner, M., Eugster, R., Trenka, E. ve Bergamin-Strotz, L., 1996, FT-NIR spectroscopy and wood identification, *Holzforschung*, 50(2), 130–134. https://doi.org/10.1515/hfsg.1996.50.2.130

Burns, D. A. ve Ciurczak, E. W., 2008, *Handbook of Near Infrared Analysis*, Taylor & Francis, Boca Raton, ISBN: 978-0-8493-7393-0.

Cao, N., 2013, *Calibration optimization and efficiency in near infrared spectroscopy*, Thesis (Graduate), Iowa State University.

Chen, Q., Hu, Z., Chang, H. ve Li, B., 2007, Micro Analytical Methods for Determination of Compression Wood Content in Loblolly Pine, *Journal of Wood Chemistry and Technology*, 27(3–4), 169–178. https://doi.org/10.1080/02773810701700810

Chow, S. Z., 1972, Infrared spectral study of woody tissues from four conifers, *Wood Science*, 5(1), 27–33.

Ciaburro, G. ve Venkateswaran, B., 2017, *Neural network with R*, Packt Publishing, Birgmingham, UK. ISBN 978-1-78839-787-2.

Crivellaro, A. ve Schweingruber, F. H., 2013, *Atlas of wood bark and stem anatomy of eastern mediterranean trees and shrubs with a special focus on Cyprus*, Springer-Verlag, Berlin, ISBN: 978-3-642-37235-3.

Dahlman, N. 2012, *Near Infrared Spectroscopy – An Introductory Study On Measurement Techniques And Tools For Analysing Moisture Content In Pulp And Paper Media*, Thesis (M.Sc.), Karlstads University.

Davies, A.M.C., 2005, An introduction to near infrared spectroscopy, *NIR news*, 16(7), 9- 11.

Dawson-Andoh, B. ve Adedipe, O. E., 2012, Rapid spectroscopic separation of three Canadian softwoods, *Wood Science and Technology*, 46(6), 1193–1202. https://doi.org/10.1007/s00226-012-0468-9

Dinç, E., 2007, Kemometri Çok Değişkenli Kalibrasyon Yöntemleri. *Hacettepe Üniversitesi Eczacılık Fakültesi Dergisi*, 27(1), 61–92.

Doğu, A. D., 2013, Ahşabı Tanımak, *Restorasyon ve Konservasyon Çalışmaları Dergisi*, 16, 59–71.

Dormontt, E. E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J. C., Koch, G., Lee, S. L., Kanashiro, M., Rimbawanto, A., Thomas, D., Wiedenhoeft, A. C., Yin, Y., Zahnen, J. ve Lowe, A. J., 2015, Forensic timber identification: It's time to integrate disciplines to combat illegal logging, *Biological Conservation*, 191, 790–798. https://doi.org/10.1016/j.biocon.2015.06.038

Elmasry, G., Kamruzzaman, M., Sun, D. W. ve Allen, P., 2012. Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review. *Critical Reviews in Food Science and Nutrition*, 52(11), 999–1023. https://doi.org/10.1080/10408398.2010.543495

Espinoza, J. A., Hodge, G. R. ve Dvorak, W. S., 2012, The Potential Use of near Infrared Spectroscopy to Discriminate between Different Pine Species and Their Hybrids, *Journal of Near Infrared Spectroscopy*, 20(4), 437–447. https://doi.org/10.1255/jnirs.1006

Espinoza, E. O., Lancaster, C. A., Kreitals, N. M., Hata, M., Cody, R. B. ve Blanchette, R. A., 2014, Distinguishing wild from cultivated agarwood (*Aquilaria* spp.) using direct analysis in real time and time of-flight mass spectrometry, *Rapid Communications in Mass Spectrometry*, 28(3), 281-289. https://doi.org/10.1002/rcm.6779

Farres, M., Platikanov, S., Tsakovski, S. ve Tauler, R., 2015, Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation, *Journal of Chemometrics*, 29, 528–536. DOI: 10.1002/cem.2736

Fengel D. ve Wegener, G., 1983, *Wood: chemistry, ultrastructures, reactions*, Walter de Gruyter and Co., Berlin, ISBN: 3-11-008481-3

Flaete, P. O., Haartveit, E. Y. ve Vadla, K., 2006, Near infrared spectroscopy with multivariate statistical modelling as a tool for differentiation of wood from tree species with similar appearance, *New Zealand Journal of Forestry Science*, 36(2), 382–392.

Friedman, J., 1988, Regularized Discriminant Analysis, *Journal of the American Statistical Association*, 4389(July), 1–32.

Fritsch, S., Guetner, F. ve Wright, M. N., 2019, neuralnet: Training of Neural Networks. https://cran.r-project.org/package=neuralnet.

Gasson, P., 2011, How precise can wood identification be? wood anatomy's role in support of the legal timber trade, especially cites. *IAWA Journal*, 32(2), 137–154. https://doi.org/10.1163/22941932-90000049

Geladi, P. ve Kowalski, B. R., 1986, Partıal Least-Squares Regression: A Tutorial, *Analytica Chimica Acta*, 185, 1–17.

Gholizadeh, A., Boruvka, L., Saberioon, M. M., Kozák, J., Vašát, R. ve Nemecek, K., 2015, Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features, *Soil and Water Research*, 10(4), 218–227. https://doi.org/10.17221/113/2015-SWR

Gierlinger, N., Jacques, D., Schwanninger, M., Wimmer, R., Hinterstoisser, B. ve Paques, L. E., 2003, Rapid prediction of natural durability of larch heartwood using Fourier transform near-infrared spectroscopy, *Canadian Journal of Forest Research*, 33, 1727–1736.

Gierlinger, N., Schwanninger, M. ve Wimmer, R., 2004, Characteristics and classifi cation of Fourier- transform near infrared spectra of the heartwood of different larch species (Larix sp.), *Journal of Near Infrared Spectroscopy*, 12, 113–119.

Guo, Y., Hastie, T. ve Tibshirani, R., 2005, Regularized Discriminant Analysis and Its Application in Microarrays, *Biostatistics*, 1(1), 1–18. https://doi.org/10.1155/2015/797280

Haaland, D. M. ve Thomas, E. V., 1988, Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information, *Analytical Chemistry*, 60(11), 1193–1202. https://doi.org/10.1021/ac00162a020

Haartveit, Y. E. ve Flæte, P. O., 2008, Near Infrared Spectroscopy (NIRS) as a Tool for Effective Classification of Wood. *Proceedings of the 51st International Convention of Society of Wood Science and Technology*, Concepcion, Chile, 10–12.

Hans, G., Leblon, B., Cooper, P., La Rocque, A. ve Nader, J., 2015, Determination of moisture content and basic specific gravity of *Populus tremuloides* (Michx.) and *Populus balsamifera* (L.) logs using a portable near-infrared spectrometer, *Wood Material Science & Engineering*, 10(1), 3–16. https://doi.org/10.1080/17480272.2014.916349

Hastie, T., Tibshirani, R. ve Friedman, J., 2017, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, Stanford, California, ISBN: 978-0-387-84858-7.

Hermanson, J. ve Wiedenhoeft, A., 2011, A brief review of machine vision in the context of automated wood identification systems, IAWA Journal, 32(2), 233-250. https://doi.org/10.1163/22941932-90000054

Hoffmeyer, P. ve Pedersen, J. G., 1995, Evaluation of density and strength of Norway spruce wood by near infrared reflectance spectroscopy, *Holz als Roh- und Werkstoff*, (53), 165–170.

Horikawa, Y., Mizuno-Tazuru, S. ve Sugiyama, J., 2015, Near-infrared spectroscopy as a potential method for identification of anatomically similar Japanese diploxylons, *Journal of Wood Science*, 61(3), 251–261. https://doi.org/10.1007/s10086-015-1462-2

Hwang, S. W., Horikawa, Y., Lee, W. H. ve Sugiyama, J., 2016, Identification of Pinus species related to historic architecture in Korea using NIR chemometric approaches, *Journal of Wood Science*, 62(2), 156–167. https://doi.org/10.1007/s10086-016-1540-0

James, G., Witten, D., Hastie, T. ve Tibshirani, R., 2013, A*n Introduction to Statistical Learning*, Springer, New York, NY. https://doi.org/10.1007/978-1-4614-7138-7

Jiang, Z. H., Yang, Z., So, C. L. ve Hse, C. Y., 2007, Rapid prediction of wood crystallinity in Pinus elliotii plantation wood by near-infrared spectroscopy. *Journal of Wood Science*, , 449–453. https://doi.org/10.1007/s10086-007-0883-y

Jiao, L., Yin, Y., Cheng, Y. ve Jiang, X., 2014, DNA barcoding for identification of the endangered species Aquilaria sinensis: comparison of data from heated or aged wood samples, *Holzforschung*, 68(4), 487-494. https://doi.org/10.1515/hf-2013-0129

Jin, X., Chen, X., Xiao, L., Shi, C., Chen, L., Yu, B., Yi, Z., Yoo, J. H., Heo, K., Yu, C. Y., Yamada, T., Sacks, E. J. ve Peng, J., 2017, Application of visible and near-infrared spectroscopy to classification of Miscanthus species, *PLoS ONE*, 12(4), 1–12. https://doi.org/10.1371/journal.pone.0171360

Jones, P. D., Schimleck, L. R., Peter, G. F., Daniels, R. F. ve Clark, A., 2005, Non-destructive estimation of *Pinus taeda* L. tracheid morphological characteristics for samples from a wide range of sites in Georgia, *Wood Science and Technology*, 39(7), 529–545. https://doi.org/10.1007/s00226-005-0021-1

Karatzoglou, A., Smola, A., Hornik, K. ve Zeileis, A., 2004, An S4 Package for Kernel Methods in R, *Journal of Statistical Software*, 11(9), 1–20.

Kelley, S. S., Rials, T. G., Groom, L. R. ve So, C. L., 2004, Use of near infrared spectroscopy to predict the mechanical properties of six softwoods, *Holzforschung*, 58, 252–260.

Kendall, M. G. ve Buckland, W. R., 1971, *A Dictionary of Statistical Terms*, Hafner Publishing, New York. https://doi.org/10.2307/2986804

Köksal, S.E. ve Kılıç Pekgözlü, A., 2016, Sarıçam (*Pinus sylvetris* L), Karaçam (*Pinus nigra* Arnold.) ve Kızılçam (*Pinus brutia* Ten.) Basınç Odununun Mikroskobik Yapısı, Düzce Üniversitesi Ormancılık Dergisi, 12(1), 72- 82.

Kucheryavskiy, S., 2018, mdatools: Multivariate Data Analysis for Chemometrics. https://cran.r-project.org/package=mdatools

Kuhn, M., 2019, caret: Classification and Regression Training. https://cran.r-project.org/package=caret, [Ziyaret tarihi: 04.04.2019].

Kvalheim, O. M., 2020, Variable importance: Comparison of selectivity ratio and significance multivariate correlation for interpretation of latent-variable regression models, *Journal of Chemometrics*, (December 2019), 1–10. https://doi.org/10.1002/cem.3211

Lang, C., Almeida, D. R. A. ve Costa, F. R. C., 2017, Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR), *Forest Ecology and Management*, 406(August), 219–227. https://doi.org/10.1016/j.foreco.2017.09.003

Lang, C., Costa, F. R. C., Camargo, J. L. C., Durgante, F. M. ve Vicentini, A., 2015, Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree species, *PLoS ONE*, 10(8), 1–15. https://doi.org/10.1371/journal.pone.0134521

Lazarescu, C., Hart, F., Pirouz, Z., Panagiotidis, K., Mansfield, S. D., Barrett, J. D. ve Avramidis, S., 2017, Wood species identification by near-infrared spectroscopy, *International Wood Products Journal*, 8(1), 32–35. https://doi.org/10.1080/20426445.2016.1242270

Leandro, J. G. R., Gonzaga, F. B., ve Latorraca, J. V. de F., 2019, Discrimination of wood species using laser-induced breakdown spectroscopy and near-infrared reflectance spectroscopy, *Wood Science and Technology*, 53(5), 1079–1091. https://doi.org/10.1007/s00226-019-01119-7

Li, Y., Via, B. K., Young, T. ve Li, Y., 2019, Visible-Near Infrared Spectroscopy and Chemometric Methods for Wood Density Prediction and Origin/Species Identification, *Forests*, 10(12), 1078. https://doi.org/10.3390/f10121078

Liu, Y. ve Rayens, W., 2007, PLS and dimension reduction for classification, *Computational Statistics*, 22(2), 189–208. https://doi.org/10.1007/s00180-007-0039-y

MacLachlan, I. R. ve Gasson, P., 2010, PCA of CITES listed Pterocarpus santalinus (Leguminosae) wood, *IAWA Journal*, 32(2), 121-138. https://doi.org/10.1163/22941932-90000010

Maechler, M., 2019, sfsmisc: Utilities from "Seminar fuer Statistik" ETH Zurich. https://cran.r-project.org/package=sfsmisc, [Ziyaret tarihi: 08.05.2019].

Maitra, S. ve Yan, J., 2008, Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. *2008 Casualty Actuarial Society Discussion Paper Program - Applying Multivariate Statistical Models*, 79–90. https://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf

Manley, M., 2014, Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials, *Chemical Society Reviews*, 43(24), 8200–8214. https://doi.org/10.1039/c4cs00062e

Martin-Benito, D., Beeckman, H. ve Cañellas, I., 2013, Influence of drought on tree rings and tracheid features of *Pinus nigra* and *Pinus sylvestris* in a mesic Mediterranean forest. *Eur J Forest Res*, 132, 33–45. https://doi.org/10.1007/s10342-012-0652-3

Mehmood, T., Liland, K.H., Snipen, L. ve Saebo, S., 2012, A review of variable selection methods in Partial Least Squares Regression, *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69. http://dx.doi.org/10.1016/j.chemolab.2012.07.010

Mehmood, T., Saebo, S. ve Liland, K. H., 2020, Comparison of variable selection methods in partial least squares regression, *Journal of Chemometrics*, (December 2019), 1–14. https://doi.org/10.1002/cem.3226

Merev, N., 2003, *Odun Anatomisi ve Odun Tanıtımı,* Karadeniz Teknik Üniversitesi Matbaası, Trabzon, ISBN:975-6983-30-2.

Mevik, B., Wehrens, R. ve Hovde Liland, K., 2019, pls: Partial Least Squares and Principal Component Regression. https://cran.r-project.org/package=pls, [Ziyaret tarihi: 12.01.2019].

Meyer, D., Dimitriadu, E., Hornik, K., Weingessel, A. ve Leisch, F., 2019, e1071: Misc Functions of the Department of Statistics, Probability Theory Group. https://cran.r-project.org/package=e1071, [Ziyaret tarihi: 04.04.2019].

Naes, T., Isaksson, T., Fearn, T. ve Davies, T., 2004, *A user-friendly guide to Multivariate Calibration and Classification,* Biddles, Chichester, West Sussex, UK, ISBN:0 9528666 2 5.

Nisbet, R., Miner, G. ve Yale, K., 2018, *Handbook of Statistical Analysis and Data Mining Applications,* Elsevier. https://doi.org/10.1016/C2012-0-06451-4

Nisgoski, S, De Muñiz, G., Gonçalves, T. ve Ballarin, A., 2017a, Use of Visible and Near-Infrared Spectroscopy for Discrimination of Eucalypt Species by Examination of Solid Samples, *Journal Of Tropical Forest Science*, 29(3), 371–379. https://doi.org/10.26525/jtfs2017.29.3.371379

Nisgoski, S., Batista, F. R. R., Naide, T. L., Laube, N. C. C., Leão, A. C. R. ve Muñiz, G. I. B. de, 2018, Discrimination of wood and charcoal from six Caatinga species by near-infrared spectroscopy, *Maderas. Ciencia y tecnología*, 20(2), 199–210. https://doi.org/10.4067/S0718-221X2018005002401

Nisgoski, S., de Oliveira, A. A. ve de Muñiz, G. I. B., 2017b, Artificial neural network and SIMCA classification in some wood discrimination based on near-infrared spectra, *Wood Science and Technology*, 51(4), 929–942. https://doi.org/10.1007/s00226-017-0915-8

Nisgoski, Silvana, Schardosin, F. Z., Batista, F. R. R., de Muñiz, G. I. B. ve Carneiro, M. E., 2016, Potential use of NIR spectroscopy to identify Cryptomeria japonica varieties from southern Brazil. *Wood Science and Technology*, 50(1), 71–80. https://doi.org/10.1007/s00226-015-0783-z

Olivieri, A. C., 2018, *Introduction to Multivariate Calibration,* Cham, Springer International Publishing. https://doi.org/10.1007/978-3-319-97097-4

Pace, J. H. C., Latorraca, J. V. F., Hein, P. R. G., Castro, J. P., Carvalho, A. M. ve Silva, C. E. S., 2019, Wood species identification from atlantic forest by near infrared spectroscopy, *Forest Systems*, 28(3), 1-10. https://doi.org/10.5424/fs/2019283-14558

Pang, H., Tong, T. ve Zhao, H., 2009, Shrinkage-based diagonal discriminant analysis and its applications in high-dimensional data, *Biometrics*, 65(4):, 1021–1029. https://doi.org/10.1111/j.1541-0420.2009.01200.x

Pasquini, C., 2018, Near infrared spectroscopy: A mature analytical technique with new perspectives – A review, *Analytica Chimica Acta*, 1026, 8–36.

https://doi.org/10.1016/j.aca.2018.04.004

Pastore, T. C. M., Braga, J. W. B., Coradin, V. T. R., Magalhães, W. L. E., Okino, E. Y. A., Camargos, J. A. A., De Muñiz, G. I. B., Bressan, O. A. ve Davrieux, F., 2011, Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixá, *Holzforschung*, 65(1), 73–80. https://doi.org/10.1515/HF.2011.010

Patel, H. ve Patel, D., 2014, A Brief survey of Data Mining Techniques Applied to Agricultural Data, *International Journal of Computer Applications*, 95(9), 6–8. https://doi.org/10.5120/16620-6472

Popescu, C.-M., Navi, P., Placencia Peña, M. I. ve Popescu, M.-C., 2018, Structural changes of wood during hydro-thermal and thermal treatments evaluated through NIR spectroscopy and principal component analysis, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 191, 405–412. https://doi.org/10.1016/j.saa.2017.10.045

R Core Team, 2018, R: A language and environment for statistical computing. Vienna, Austria. https://www.r-project.org/ , [Ziyaret tarihi: 21.06.2018].

Rajalahti, T., Arneberg, R., Kroksveen, A.C., Berle, M., Myhr, K.-M.O. ve Kvalheim, M., 2009, Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles, *Analytical Chemistry*, 81(7), 2581–2590.

Rinnan, Å., van den Berg, F. ve Engelsen, S. B., 2009, Review of the most common pre-processing techniques for near-infrared spectra, *TrAC - Trends in Analytical Chemistry*, 28(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Robert, E.M.R., Mencuccini, M. Ve Martinez-Vilalta, J., 2017, *The Anatomy and Functioning of the Xylem in Oaks, Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.*, In: Gil-Pelegrín, E., Peguero-Pina, J.J. ve Sancho-Knapik, D.(ed.), Chapter 8, Springer International Publishing, Cham, Switzerland, 261-302.

Rowell, R.M., 2005, *Handbook of Wood Chemistry and Wood Composites*, CRC Press, Boca Raton, Florida, ISBN: 0-8493-1588-3.

Russ, A., Fierova, M. ve Gigac, J., 2009, Preliminary study of wood species identification by NIR spectroscopy, *Wood Research*, 54(4), 23–32.

Russ, A. ve Fišerová, M. 2011, Estimation of Hardwood Species in Mixture By Near Infrared Spectroscopy, *Wood Research*, 56(1), 93–104.

Sandberg, K. ve Sterley, M., 2009, Separating Norway spruce heartwood and sapwood in dried condition with near-infrared spectroscopy and multivariate data analysis, *European Journal of Forest Research*, 475–481. https://doi.org/10.1007/s10342-009-0296-0

Savitzky, A. ve Golay, M. J. E., 1964, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Analytical Chemistry*, 36(8), 1627–1639.

Schimleck, L. R., 2008, Near-infrared spectroscopy: A rapid non-destructive method for measuring wood properties and its application to tree breeding, *New Zealand Journal of Forestry Science*, 38(1), 14–35.

Schimleck, L. R., Evans, R., Jones, P.D., Daniels, R.F., Peter, G.F. ve Clark III, A., 2005, Estimation of microfibril angle and stiffness by near infrared spectroscopy using sample sets having limited wood density variation, *IAWA Journal*, 26(2),175-187.

Schimleck, L. R., Jones, P. D., Peter, G. F., Daniels, R. F. ve ClarkIII, A., 2004, Nondestructive estimation of tracheid length from sections of radial wood strips by near infrared spectroscopy, *Holzforschung*, 58, 375–381.

Schoch, W., Heller, I., Schweingruber, F.H. ve Kienast, F., 2004, *Wood anatomy of central European Species.* Online version http:www.woodanatomy.ch.

Schutt, R. ve O'Neil, C., 2014, *Doing data science*, O'Reilly Media, USA. https://doi.org/10.5860/choice.51-6803

Schwanninger, M., Hinterstoisser, B., Gierlinger, N., Wimmer, R. ve Hanger, J., 2004, Application of Fourier Transform Near Infrared Spectroscopy (FT-NIR) to thermally modified wood, *Holz als Roh- und Werkstoff*, 62(6), 483–485. https://doi.org/10.1007/s00107-004-0520-z

Schwanninger, M., Rodrigues, J. C. ve Fackler, K., 2011, A review of band assignments in near infrared spectra of wood and wood components, *Journal of Near Infrared Spectroscopy*, 19(5), 287–308. https://doi.org/10.1255/jnirs.955

Schweingruber, F.H., 2007, *Wood Structure and Enviroment*, Springer-Verlag, Berlin Heidelberg, ISBN-13 978-3-540-48299-4.

Schweingruber, F. and Landolt, W., 2005, *The Xylem Database*, https://www.wsl.ch/dendropro/xylemdb/index.php, [Ziyaret tarihi: 15.02.2020].

Shou, G., Zhang, W., Gu, Y. ve Zhao, D., 2014, Application of near infrared spectroscopy for discrimination of similar rare woods in the Chinese market, *Journal of Near Infrared Spectroscopy*, 22(6), 423–432. https://doi.org/10.1255/jnirs.1136

SignalDevelopers, 2014, signal: Signal processing. http://r-forge.r-project.org/projects/signal/, [Ziyaret tarihi: 18.12.2018].

Snel, F. A., Braga, J. W. B., da Silva, D., Wiedenhoeft, A. C., Costa, A., Soares, R., Coradin, V. T. R. ve Pastore, T. C. M., 2018, Potential field-deployable NIRS identification of seven Dalbergia species listed by CITES, *Wood Science and Technology*, 52(5), 1411–1427. https://doi.org/10.1007/s00226-018-1027-9

So, C. L., Lebow, S. T., Groom, L. H. ve Rials, T. G., 2004, The Application of Near Infrared (NIR) Spectroscopy to Inorganic Preservative-Treated Wood, *Wood And Fiber*, 36(3), 329–336.

Stevens, A. ve Ramirez-Lopez, L., 2013, An introduction to the prospectr package. R package Vignette.

Taylor, A.M., Baek, S.H., Jeong, M.K. ve Nix, G., 2008, Wood shrinkage prediction using NIR spectroscopy, *Wood and Fiber Science*, 40(2), 301 – 307.

Therneau, T. ve Atkinson, B., 2019, rpart: Recursive Partitioning and Regression Trees. https://cran.r-project.org/package=rpart, [Ziyaret tarihi: 17.05.2019].

Tounis, E., 2009, I*nvestigation of Nir Spectroscopy for Identifying and Sorting Wood With Respect To Species, Moisture Content, and Weathering*, Thesis (M.Sc.), Ryerson University.

Tran, T.N., Afanador, N.L., Buydens, L.M.C. ve Blanchet, L., 2014, Interpretation of variable importance in Partial Least Squares with Significance Multivariate Correlation (sMC), *Chemometrics and Intelligent Laboratory Systems*, 138, 153–160. http://dx.doi.org/10.1016/j.chemolab.2014.08.005

Tsuchikawa, S., Inoue, K., Noma, J. ve Hayashi, K., 2003, Application of near-infrared spectroscopy to wood discrimination, *Journal of Wood Science*, 49(1), 29–35. https://doi.org/10.1007/s100860300005

Tsuchikawa, S., Yamato, K. ve Inoue, K., 2003, Discriminant Analysis of Wood-Based Materials Using Near-Infrared Spectroscopy, *Journal of Wood Science*, 49, 275–280.

Tsuchikawa, S., 2007, A review of recent near infrared research for wood and paper, *Applied Spectroscopy Reviews*, 42(1), 43–71. https://doi.org/10.1080/05704920601036707

Tsuchikawa, S. ve Schwanninger, M., 2013, A review of recent near-infrared research for wood and paper (Part 2), *Applied Spectroscopy Reviews*, 48(7), 560–587. https://doi.org/10.1080/05704928.2011.621079

Tsuchikawa, S. ve Kobori, H., 2015, A review of recent application of near infrared spectroscopy to wood science and technology, *Journal of Wood Science*, 61(3), 213–220. https://doi.org/10.1007/s10086-015-1467-x

Varmuza, K. ve Filzmoser, P., 2008, *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Boca Raton. https://doi.org/10.1201/9781420059496

Via, B., So, C.L., Groom, L.H., Shupe, T.F., Stine, M. ve Wikaira, J., 2007, Within Tree Variation of Lignin, Extractives, and Microfibril Angle Coupled With the Theoretical and Near Infrared Modeling of Microfibril Angle, *IAWA Journal*, 28(2), 189-209.

Via, B. K., Zhou, C., Acquah, G., Jiang, W. ve Eckhardt, L., 2014, Near infrared spectroscopy calibration for wood chemistry: Which chemometric technique is best for prediction and interpretation?, *Sensors (Switzerland)*, 14(8), 13532–13547. https://doi.org/10.3390/s140813532

Vieira, H. C., Santos, J. X. dos, Lopes da Silva, E., D' Angelo Rios, P., de Muñiz, G. I. B., Ribeiro Morrone, S. ve Nisgoski, S., 2019, Potential of the near-infrared spectroscopy for the discrimination of wood and charcoal of four native Myrtaceae species in southern Brazil. *Wood Material Science and Engineering*. https://doi.org/10.1080/17480272.2019.1689296

Voulgaridis, E., 1990, Wood cell morphology characteristics of some oak species and mediterranean shrubs, *Holz als Rob- und Werkstoff,* 48, 261-267.l

Wehrens, R., 2011, *Chemometrics with R* (R. Gentleman, K. Hornik, & G. Parmigiani, Ed.), Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17841-2

Wehrens, R., 2017, *Package 'ChemometricsWithR*'. https://github.com/rwehrens/CWR BugReports, [Ziyaret tarihi: 24.11.2018].

Wehrens, R. ve Kruisselbrink, J., 2018, Flexible Self-Organizing Maps in kohonen 3.0. *Journal of Statistical Software*, 87(7). https://doi.org/10.18637/jss.v087.i07

Wheeler, E. A. ve Baas, P., 1998, Wood Identification -A Review, *IAWA Journal*, 19(3), 241–264. https://doi.org/10.1163/22941932-90001528

Wickham, H., 2016, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York. https://doi.org/http://ggplot2.org

Wiesner, K., Fuchs, K., Gigler, A. M. ve Pastusiak, R., 2014, Trends in near infrared spectroscopy and multivariate data analysis from an industrial perspective, *Procedia Engineering*, 87, 867–870. https://doi.org/10.1016/j.proeng.2014.11.292

Workman, J. J. ve Weyer, L., 2012, *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy,* Taylor & Francis, Boca Raton.

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M. ve Hanpin, M., 2010, Variables selection methods in near-infrared spectroscopy, *Analytica Chimica Acta*, 667(1–2), 14–32. https://doi.org/10.1016/j.aca.2010.03.048

Yaltırık, F. ve Efe, A., 2000, *Dendroloji Ders Kitabı,* Gymnospermae -Angiospermae (Orman Endüstri Mühendisliği Bölümü öğrencileri için), İ.Ü. Orman Fakültesi, İstanbul.

Yang, Z., Liu, Y., Pang, X. ve Li, K., 2015, Preliminary Investigation into the Identification of Wood Species from Different Locations by Near Infrared Spectroscopy, *BioResources*, 10(4), 8505–8517. https://doi.org/10.15376/biores.10.4.8505-8517

Yang, Z., Li, K., Zhang, M., Xin, D. ve Zhang, J., 2016, Rapid determination of chemical composition and classification of bamboo fractions using visible-near infrared spectroscopy coupled with multivariate data analysis, *Biotechnology for Biofuels*, 9(1), 1–18. https://doi.org/10.1186/s13068-016-0443-z

Yong, H., Qing-yuan, S., Min, R. ve Yuan, H., 2019, Identification of Wood Species Based on Near Infrared Spectroscopy and Pattern Recognation Method, *Spectroscopy and Spectral Analysis*, 39(03), 705–710.

Zhang, X., Yu, H., Li, B., Li, W. J., Li, X. ve Bao, C., 2014, Discrimination of *Pinus yunnanensis, P. kesiyaand P. densata* by FT-NIR, *Journal of Chemical and Pharmaceutical Research*, 6(4), 142–149.

Zhou, Z., Rahimi, S. ve Avramidis, S., 2020, On-line species identification of green hem-fir timber mix based on near infrared spectroscopy and chemometrics, *European Journal of Wood and Wood Products*, 78(1), 151–160. https://doi.org/10.1007/s00107-019-01479-8

Zobel, B.J. ve Van Buijtenen, J.P., 1989, *Wood Variation: Its Causes and Control*, Springer-Verlag, Berlin Heidelberg, ISBN-13: 978-3-642-74071-8

Zobel, B.J. ve Jett, J.B., 1995, *Genetics of Wood Production,* Springer-Verlag, Berlin Heidelberg, ISBN: 978-3-642-79516-9

EKLER

EK 1. Çalışmada kullanılan veri formatı.

Örnek	Sarıçam	Karaçam	Macar Meşesi	Sapsız Meşe
Örnek kodu	S1-1	K1-1	M1-1	P1-1
Wavenumber (cm-1)	Log (1/R)	Log (1/R)	Log (1/R)	Log (1/R)
4000	1,000155	1,053231	1,085770	1,060389
4002	1,000267	1,053095	1,085720	1,060744
4003	1,000170	1,053304	1,085091	1,061001
4005	1,000010	1,053294	1,084726	1,061117
4007	0,999171	1,052099	1,083830	1,060133
4009	0,999018	1,052089	1,083447	1,060907
4011	0,998612	1,052338	1,083236	1,061042
4013	0,998126	1,052171	1,082936	1,061087
...	...	...	...	...
5001	0,703707	0,782411	0,740203	0,763533
5002	0,704393	0,783056	0,740878	0,764102
5004	0,705143	0,783795	0,741721	0,764788
5006	0,705931	0,784610	0,742559	0,765475
5008	0,706796	0,785411	0,743508	0,766202
5010	0,707611	0,786191	0,744411	0,766848
5012	0,708579	0,787112	0,745353	0,767682
5014	0,709399	0,787991	0,746229	0,768474
5016	0,710307	0,788923	0,747211	0,769376
5018	0,711237	0,789865	0,748149	0,770200
5020	0,712284	0,790872	0,749238	0,771132
...	...	...	...	...
11980	0,276351	0,284390	0,512147	0,621132
11982	0,277222	0,285551	0,512391	0,619058
11983	0,276102	0,286171	0,512651	0,617743
11985	0,275953	0,286654	0,511889	0,617187
11987	0,276073	0,285977	0,512614	0,618163
11989	0,276210	0,285574	0,512414	0,619712
11991	0,277396	0,282926	0,513106	0,620887
11993	0,277783	0,283217	0,513615	0,621115
11995	0,276008	0,284671	0,513425	0,622082
11997	0,276142	0,284756	0,512682	0,622097
11999	0,276878	0,284677	0,514501	0,621372
12001	0,276717	0,284483	0,514766	0,620497

EK 2. Cins ayrımında kullanılan PLS-DA model parametreleri (Ham veri seti).

Cins Ayrımı	Yüklemeler (Loadings)				Eğim (Slope)			
	Sarıçam	Karaçam	Sapsız M.	Macar M.	Sarıçam	Karaçam	Sapsız M.	Macar M.
Comp.1	-6,4851	-4,9828	7,5452	3,9228	0,1045	0,0660	0,1808	0,0444
Comp.2	-7,4514	-6,4758	5,9067	8,0205	0,2423	0,1774	0,2916	0,2302
Comp.3	4,9275	3,0881	-3,7864	-4,2292	0,3026	0,2028	0,3372	0,2819
Comp.4	1,0887	-0,8206	7,6462	-7,9142	0,3056	0,2045	0,5229	0,4628
Comp.5	0,0933	-2,2839	-4,7593	6,9499	0,3056	0,2184	0,5948	0,6023
Comp.6	-0,8933	-4,2907	3,2962	1,8878	0,3076	0,2673	0,6293	0,6126
Comp.7	0,4280	4,7464	-3,5556	-1,6188	0,3080	0,3272	0,6695	0,6201
Comp.8	6,5343	-4,5721	-2,4110	0,4489	0,4141	0,3827	0,6880	0,6207
Comp.9	2,3571	-1,2518	1,6501	-2,7554	0,4279	0,3869	0,6966	0,6426
Comp.10	2,1106	-1,5609	-3,3517	2,8021	0,4389	0,3934	0,7323	0,6653
Comp.11	-0,4712	3,3159	0,5722	-3,4169	0,4395	0,4226	0,7333	0,6990
Comp.12	3,1255	-1,1599	1,8319	-3,7975	0,4638	0,4262	0,7440	0,7407
Comp.13	-0,3321	-1,2392	-2,2834	3,8546	0,4640	0,4303	0,7605	0,7836
Comp.14	3,6284	-3,4380	3,0991	-3,2896	0,4967	0,4617	0,7911	0,8148
Comp.15	3,7126	-3,6795	0,5959	-0,6290	0,5310	0,4976	0,7922	0,8160
Cins Ayrımı	Bias				R2			
	Sarıçam	Karaçam	Sapsız M.	Macar M.	Sarıçam	Karaçam	Sapsız M.	Macar M.
Comp.1	-1,54E-17	5,20E-18	-1,62E-17	-1,00E-17	11,3478	6,7057	19,0894	4,5813
Comp.2	-1,23E-16	1,10E-16	-2,37E-17	-1,42E-16	24,9982	17,8365	30,0348	23,1331
Comp.3	-1,03E-16	-1,10E-17	-3,92E-17	-4,25E-17	30,9675	20,3677	34,5325	28,2912
Comp.4	4,52E-17	2,93E-17	6,36E-17	-9,58E-17	31,2589	20,5464	52,8744	46,3548
Comp.5	1,19E-16	-2,64E-17	5,53E-17	-3,22E-17	31,2610	21,9309	59,9805	60,2844
Comp.6	2,98E-17	3,63E-17	2,64E-16	-4,24E-16	31,4572	26,8176	63,3892	61,3121
Comp.7	1,36E-16	1,25E-17	-1,03E-16	-6,21E-17	31,5022	32,7972	67,3553	62,0678
Comp.8	-7,24E-16	-1,41E-17	-8,37E-19	-1,71E-16	41,9992	38,3457	69,1790	62,1260
Comp.9	-3,51E-16	1,76E-16	-5,40E-17	9,34E-17	43,3651	38,7616	70,0332	64,3155
Comp.10	3,85E-16	4,23E-16	3,41E-16	-2,88E-16	44,4602	39,4083	73,5577	66,5798
Comp.11	7,10E-17	-1,36E-16	-3,03E-16	-2,81E-16	44,5148	42,3268	73,6604	69,9468
Comp.12	-6,43E-16	-4,72E-16	-5,29E-16	1,56E-15	46,9165	42,6838	74,7132	74,1057
Comp.13	4,23E-17	-5,24E-16	-6,22E-17	1,51E-15	46,9436	43,0914	76,3489	78,3906
Comp.14	-3,18E-16	-6,38E-16	-1,15E-15	1,50E-15	50,1804	46,2288	79,3621	81,5114
Comp.15	-1,34E-16	4,75E-16	-1,55E-15	1,54E-15	53,5691	49,8224	79,4735	81,6255

EK 3. Çam türlerinin ayrımında kullanılan PLS-DA model parametreleri (Ham veri seti).

Çam Türleri	Yüklemeler	Eğim	Bias	R2
Comp.1	0,6378	0,0015	1,54E-12	0,4565
Comp.2	1,6114	0,0111	3,76E-12	1,4146
Comp.3	3,0361	0,0452	-9,84E-12	4,8159
Comp.4	3,1121	0,0811	-8,79E-11	8,3899
Comp.5	3,3399	0,1224	5,38E-10	12,5060
Comp.6	5,6638	0,2411	-2,74E-10	24,3433
Comp.7	4,9866	0,3331	7,05E-10	33,5190
Comp.8	2,5731	0,3577	1,15E-09	35,9621
Comp.9	3,5034	0,4031	-1,74E-09	40,4911
Comp.10	2,3814	0,4241	3,44E-09	42,5837
Comp.11	2,8907	0,4550	3,73E-09	45,6672
Comp.12	4,8287	0,5413	-1,03E-09	54,2711
Comp.13	3,4468	0,5853	-6,94E-10	58,6549
Comp.14	4,3817	0,6563	-2,04E-09	65,7395
Comp.15	4,2933	0,7246	-5,94E-10	72,5412

EK 4. Meşe türlerinin ayrımında kullanılan PLS-DA model parametreleri (Ham veri seti).

Meşe Türleri	Yüklemeler	Eğim	Bias	R2
Comp.1	4,2554	0,0858	1,09E-17	8,5790
Comp.2	3,2487	0,1358	-2,61E-18	13,5791
Comp.3	3,7657	0,2030	5,46E-17	20,2975
Comp.4	7,7407	0,4868	-3,28E-17	48,6847
Comp.5	4,9916	0,6049	-5,27E-17	60,4890
Comp.6	3,3378	0,6577	2,46E-16	65,7671
Comp.7	3,0765	0,7025	-1,76E-15	70,2512
Comp.8	3,5835	0,7633	1,39E-16	76,3350
Comp.9	2,5086	0,7932	-1,35E-16	79,3164
Comp.10	3,5133	0,8516	7,55E-16	85,1640
Comp.11	1,8595	0,8680	7,27E-17	86,8023
Comp.12	2,3973	0,8952	3,71E-17	89,5250
Comp.13	1,8697	0,9118	7,12E-16	91,1812
Comp.14	1,7576	0,9264	1,22E-16	92,6447
Comp.15	2,1185	0,9477	-2,16E-15	94,7711

EK 5. Cins ayrımında kullanılan PLS-DA korelasyon tabloları (Ham veri seti).

Cins		Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
Yüklemeler	Comp.1	1														
	Comp.2	-0,3849	1													
	Comp.3	-0,9696	0,49578	1												
	Comp.4	-0,1551	0,00431	-0,00176	1											
	Comp.5	-0,1416	-0,00013	0,01153	0,71809	1										
	Comp.6	0,1079	-0,00285	0,03498	-0,13507	-0,15684	1									
	Comp.7	0,0670	-0,00103	0,00353	0,01070	-0,02224	0,76477	1								
	Comp.8	0,0027	-0,00140	-0,01524	-0,04702	-0,10248	0,06312	0,11661	1							
	Comp.9	0,0336	-0,00081	-0,00046	-0,00795	0,05056	0,01012	0,02355	-0,27532	1						
	Comp.10	0,0549	-0,00104	0,00078	0,00150	-0,02332	-0,00173	-0,09110	-0,34700	0,28655	1					
	Comp.11	0,0540	-0,00115	-0,00017	-0,00433	0,00584	0,00574	-0,11828	0,00104	0,04916	0,45307	1				
	Comp.12	-0,0074	0,0001	-0,0001	0,0002	-0,00127	-0,00032	0,00152	-0,01470	-0,53072	-0,15088	-0,10828	1			
	Comp.13	0,0236	-0,0005	0,0003	-0,0003	0,00032	0,00038	-0,00943	-0,00228	-0,08149	-0,15402	0,32773	0,37546	1		
	Comp.14	0,0193	-0,0004	0,0004	0,0002	-0,00016	-0,00022	0,00361	-0,00049	-0,01813	0,04349	0,18283	-0,18926	0,31663	1	
	Comp.15	-0,0313	0,0006	-0,0006	-0,0002	0,00004	0,00020	-0,00290	-0,00022	-0,00761	-0,04226	-0,00866	-0,05908	-0,17368	-0,53950	1
Skorlar	Comp.1	1														
	Comp.2	2E-15	1													
	Comp.3	9E-16	-3E-16	1												
	Comp.4	-4E-15	7E-16	9E-16	1											
	Comp.5	7E-15	-5E-15	3E-17	7E-16	1										
	Comp.6	-2E-14	1E-15	-4E-17	1E-15	-1E-17	1									
	Comp.7	8E-15	6E-18	-1E-15	9E-16	-1E-15	2E-16	1								
	Comp.8	1E-14	6E-15	5E-15	4E-15	1E-15	5E-16	2E-17	1							
	Comp.9	-4E-14	1E-14	5E-15	-3E-15	2E-15	-2E-15	-2E-15	7E-15	1						
	Comp.10	1E-14	7E-15	2E-15	-7E-16	-6E-17	-4E-15	-1E-15	8E-15	4,8E-15	1					
	Comp.11	7E-15	-4E-16	-1E-15	-4E-15	-4E-15	-1E-15	3E-15	-1E-14	-2E-15	-4E-15	1				
	Comp.12	-5E-15	-4E-16	1E-15	-7E-15	8E-16	8E-16	-1E-14	1E-14	1E-14	-3E-15	1,1E-14	1			
	Comp.13	6E-14	2E-15	1E-14	3E-15	-3E-15	-4E-16	6E-15	-4E-15	-9E-15	1,4E-15	1,3E-14	-2E-14	1		
	Comp.14	-3E-14	-5E-15	-1E-14	6E-17	9E-16	3E-17	9E-16	2E-14	2,1E-14	2,3E-14	5,5E-15	9,4E-14	-5E-14	1	
	Comp.15	-5E-14	3E-15	2E-14	-9E-15	8E-15	-2E-15	5E-15	8E-15	5,1E-15	2,6E-14	-4E-15	8,4E-15	-4E-14	1,1E-13	1
Ağırlıklar	Comp.1	1														
	Comp.2	0,8764	1													
	Comp.3	-0,8354	-0,4799	1												
	Comp.4	-0,0272	-0,0097	0,0111	1											
	Comp.5	-0,0052	0,0189	-0,0295	-0,7073	1										
	Comp.6	0,0716	0,0299	-0,0583	0,0976	0,0902	1									
	Comp.7	-0,0320	-0,0218	0,0426	-0,1002	-0,0479	-0,7813	1								
	Comp.8	-0,0082	-0,0043	0,0130	-0,0279	0,1066	0,0654	-0,1297	1							
	Comp.9	0,0063	-0,0032	0,0072	0,0975	-0,1018	0,0351	-0,0761	0,23409	1						
	Comp.10	0,0088	-0,0004	0,0030	-0,0871	0,1282	-0,0073	0,0081	0,34214	-0,2786	1					
	Comp.11	0,0053	-0,0043	0,0076	0,0498	-0,1000	-0,1270	0,1575	-0,2145	0,2014	-0,6098	1				
	Comp.12	-0,0002	-0,0027	0,0057	0,0639	-0,0705	0,0054	-0,0280	0,1485	0,6586	-0,2160	0,2801	1			
	Comp.13	0,0066	0,0019	-0,0030	-0,0567	0,0805	0,0251	-0,0233	0,0738	-0,3652	0,4613	-0,5386	-0,5790	1		
	Comp.14	-0,0003	-0,0004	0,0012	0,0340	-0,0294	0,0365	-0,0576	0,1051	0,3633	-0,0773	-0,0014	0,5277	-0,3854	1	
	Comp.15	-0,0042	-0,0002	0,0008	0,0099	0,0015	0,0311	-0,0476	0,1353	0,2154	0,1150	-0,1674	0,3040	-0,0106	0,6596	1

EK 6. Çam türlerinin ayrımında kullanılan PLS-DA korelasyon tabloları (Ham veri seti).

	Çam	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
Yüklemeler	Comp.1	1														
	Comp.2	0,98636	1													
	Comp.3	-0,80414	-0,85025	1												
	Comp.4	0,05911	-0,00367	-0,16784	1											
	Comp.5	-0,12237	0,00759	-0,00332	-0,52110	1										
	Comp.6	0,07809	-0,00484	0,00212	-0,00016	-0,59540	1									
	Comp.7	-0,04604	0,00286	-0,00125	0,00009	-0,00019	-0,33297	1								
	Comp.8	0,05025	-0,00312	0,001364	-0,000100	0,000208	-0,000133	-0,459597	1							
	Comp.9	0,00536	-0,00033	0,000145	-0,000011	0,000022	-0,000014	0,000008	-0,323504	1						
	Comp.10	-0,03200	0,00199	-0,000869	0,000064	-0,000132	0,000084	-0,000050	0,000054	-0,597798	1					
	Comp.11	0,02771	-0,00172	0,000752	-0,000055	0,000114	-0,000073	0,000043	-0,000047	-0,000005	-0,451366	1				
	Comp.12	-0,01889	0,00117	-0,000513	0,000038	-0,000078	0,000050	-0,000029	0,000032	0,000003	-0,000020	-0,394017	1			
	Comp.13	0,01149	-0,00071	0,000312	-0,000023	0,000047	-0,000030	0,000018	-0,000019	-0,000002	0,000012	-0,000011	-0,652815	1		
	Comp.14	-0,00045	0,00003	-0,000012	0,000001	-0,000002	0,000001	-0,000001	0,000001	0,000000	0,000000	0,000000	0,000000	-0,432309	1	
	Comp.15	-0,00326	0,00020	-0,000089	0,000007	-0,000013	0,000009	-0,000005	0,000006	0,000001	-0,000004	0,000003	-0,000002	0,000001	-0,51689	1
Skorlar	Comp.1	1														
	Comp.2	1,61E-15	1													
	Comp.3	1,42E-14	4,23E-16	1												
	Comp.4	-2,19E-15	-1,35E-15	-8,51E-16	1											
	Comp.5	-8,47E-15	-4,59E-15	-4,51E-15	-1,57E-15	1										
	Comp.6	5,89E-15	-1,40E-15	2,16E-15	3,64E-16	-4,60E-16	1									
	Comp.7	-3,32E-14	2,08E-14	1,74E-15	-2,51E-15	-1,42E-15	-1,36E-14	1								
	Comp.8	2,01E-14	-4,78E-15	-3,85E-15	-4,34E-15	-1,88E-15	-6,12E-15	-5,18E-15	1							
	Comp.9	-2,13E-14	-1,73E-14	-6,79E-15	-7,35E-16	-4,46E-15	-6,31E-15	-2,59E-15	7,58E-15	1						
	Comp.10	-1,05E-14	1,50E-14	1,39E-15	-4,65E-15	1,51E-16	-6,05E-15	2,78E-16	1,19E-14	1,19E-14	1					
	Comp.11	7,09E-15	-5,33E-15	3,30E-15	2,94E-15	3,17E-15	-1,06E-14	-6,71E-15	5,06E-15	1,30E-14	1,66E-14	1				
	Comp.12	1,66E-15	1,54E-14	1,73E-14	-6,39E-15	-5,84E-15	-3,09E-14	-5,91E-14	3,74E-15	3,77E-14	1,06E-15	1,11E-13	1			
	Comp.13	3,11E-14	8,59E-15	-7,76E-15	-7,65E-15	-2,11E-15	-3,24E-14	-4,73E-14	-2,57E-15	2,86E-14	4,25E-15	6,75E-14	1,97E-13	1		
	Comp.14	-6,43E-14	2,90E-14	-5,42E-15	-5,60E-15	-1,02E-14	-5,36E-14	-2,68E-14	1,99E-14	5,50E-14	2,39E-14	8,52E-14	2,94E-13	2,51E-13	1	
	Comp.15	-4,06E-14	1,58E-14	3,75E-15	-1,37E-14	-8,19E-15	-4,45E-14	-6,09E-14	1,57E-14	-3,23E-15	-1,13E-14	9,33E-14	2,70E-13	2,03E-13	2,04E-13	1
Ağırlıklar	Comp.1	1														
	Comp.2	0,933365	1													
	Comp.3	0,652167	0,879224	1												
	Comp.4	0,283106	0,386817	0,43979	1											
	Comp.5	0,176827	0,280455	0,317648	0,721982	1										
	Comp.6	0,156371	0,216552	0,246117	0,559603	0,775221	1									
	Comp.7	0,079179	0,121389	0,1376	0,312778	0,433239	0,55891	1								
	Comp.8	0,063271	0,071957	0,082264	0,187162	0,259349	0,334482	0,59849	1							
	Comp.9	0,030689	0,041418	0,047106	0,107115	0,148392	0,191414	0,342486	0,572276	1						
	Comp.10	0,019611	0,033743	0,038147	0,086686	0,120056	0,154896	0,277134	0,463101	0,809206	1					
	Comp.11	0,021611	0,026351	0,030059	0,068372	0,094733	0,122186	0,218624	0,365299	0,638336	0,788849	1				
	Comp.12	0,014401	0,021771	0,024687	0,056118	0,077732	0,100278	0,179418	0,299805	0,523876	0,647398	0,820692	1			
	Comp.13	0,015332	0,019958	0,022722	0,051673	0,071589	0,092341	0,165221	0,276072	0,482413	0,596161	0,755737	0,920855	1		
	Comp.14	0,013149	0,017549	0,019966	0,045401	0,062898	0,081132	0,145166	0,242564	0,423858	0,523799	0,664006	0,809082	0,87862	1	
	Comp.15	0,011407	0,015529	0,017657	0,04015	0,055621	0,071748	0,128374	0,214506	0,37483	0,46321	0,5872	0,715494	0,776989	0,884329	1

EK 7. Meşe türlerinin ayrımında kullanılan PLS-DA korelasyon tabloları (Ham veri seti).

	Meşe	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
Yüklemeler	Comp.1	1														
	Comp.2	0,772844	1													
	Comp.3	-0,928673	-0,553972	1												
	Comp.4	0,235184	-0,008666	-0,350080	1											
	Comp.5	0,168994	-0,006227	0,003235	-0,236397	1										
	Comp.6	-0,155310	0,005723	-0,002973	0,000753	-0,557943	1									
	Comp.7	-0,017549	0,000647	-0,000336	0,000085	0,000061	-0,206357	1								
	Comp.8	0,034508	-0,001272	0,000661	-0,000167	-0,000120	0,000110	-0,623564	1							
	Comp.9	-0,020117	0,000741	-0,000385	0,000098	0,000070	-0,000064	-0,000007	-0,581158	1						
	Comp.10	-0,033187	0,001223	-0,000635	0,000161	0,000116	-0,000106	-0,000012	0,000024	-0,333288	1					
	Comp.11	0,042483	-0,001565	0,000813	-0,000206	-0,000148	0,000136	0,000015	-0,000030	0,000018	-0,517322	1				
	Comp.12	0,000781	-0,000029	0,000015	-0,000004	-0,000003	0,000002	0,000000	-0,000001	0,000000	0,000001	-0,469020	1			
	Comp.13	-0,011875	0,000438	-0,000227	0,000058	0,000041	-0,000038	-0,000004	0,000008	-0,000005	-0,000008	0,000010	-0,367604	1		
	Comp.14	0,011059	-0,000408	0,000212	-0,000054	-0,000039	0,000035	0,000004	-0,000008	0,000005	0,000008	-0,000010	0,000000	-0,392311	1	
	Comp.15	-0,007588	0,000280	-0,000145	0,000037	0,000026	-0,000024	-0,000003	0,000005	-0,000003	-0,000005	0,000007	0,000000	-0,000002	-0,680227	1
Skorlar	Comp.1	1														
	Comp.2	2,74E-15	1													
	Comp.3	4,45E-15	2,11E-17	1												
	Comp.4	-4,50E-15	7,15E-16	1,32E-15	1											
	Comp.5	2,38E-15	8,82E-16	1,33E-15	1,19E-16	1										
	Comp.6	-1,85E-14	-3,94E-15	2,25E-15	-4,79E-17	4,55E-16	1									
	Comp.7	-5,22E-14	3,02E-15	-5,11E-16	2,65E-15	6,72E-15	-1,79E-17	1								
	Comp.8	-8,33E-15	-1,84E-15	1,93E-15	1,65E-15	2,45E-15	7,79E-16	6,65E-15	1							
	Comp.9	-1,75E-14	-9,02E-15	-1,38E-15	-1,08E-15	6,03E-16	2,33E-16	-4,22E-15	-2,51E-15	1						
	Comp.10	1,52E-14	-1,33E-14	-1,06E-15	1,79E-15	8,29E-15	9,30E-16	2,26E-16	8,61E-15	-5,00E-16	1					
	Comp.11	-4,63E-14	4,49E-16	1,08E-14	-3,01E-15	1,62E-15	6,47E-16	-2,52E-15	-8,05E-16	-4,29E-15	-8,62E-15	1				
	Comp.12	1,19E-14	-1,04E-14	1,10E-14	-2,50E-16	-9,16E-16	7,17E-15	5,27E-15	5,56E-16	-1,09E-14	-3,70E-14	-3,06E-14	1			
	Comp.13	6,24E-14	2,23E-14	-2,12E-15	1,57E-15	3,47E-15	-4,18E-15	7,97E-15	2,66E-15	-4,09E-18	-2,00E-15	-4,06E-15	5,60E-15	1		
	Comp.14	5,03E-14	2,16E-14	-5,53E-15	-7,07E-16	1,24E-14	-5,20E-15	8,96E-15	-1,01E-15	-4,66E-15	-5,24E-14	-2,12E-14	-1,70E-14	-3,16E-14	1	
	Comp.15	1,23E-13	-1,61E-14	-5,80E-15	6,26E-15	1,22E-15	1,00E-14	2,75E-14	7,03E-15	-1,56E-14	-6,56E-14	-1,30E-14	-1,63E-14	-4,60E-14	-1,34E-13	1
Ağırlıklar	Comp.1	1														
	Comp.2	0,8521	1													
	Comp.3	0,1433	0,6186	1												
	Comp.4	0,1968	0,2745	0,4487	1											
	Comp.5	0,1093	0,0886	0,1465	0,32451	1										
	Comp.6	-0,0028	0,0555	0,0892	0,20061	0,61832	1									
	Comp.7	0,0009	0,0241	0,0387	0,08698	0,26809	0,43360	1								
	Comp.8	0,0080	0,0201	0,0326	0,07304	0,22512	0,36415	0,83979	1							
	Comp.9	-0,0005	0,0170	0,0272	0,06125	0,18878	0,30531	0,70414	0,83848	1						
	Comp.10	-0,0022	0,0110	0,0177	0,03979	0,12266	0,19836	0,45748	0,54477	0,64971	1					
	Comp.11	0,0168	0,0077	0,0130	0,02845	0,08766	0,14191	0,32719	0,38959	0,46465	0,71516	1				
	Comp.12	0,0100	0,0052	0,0087	0,01920	0,05915	0,09574	0,22076	0,26287	0,31351	0,48253	0,67472	1			
	Comp.13	0,0027	0,0032	0,0052	0,01166	0,03592	0,05812	0,13402	0,15959	0,19034	0,29295	0,40964	0,60712	1		
	Comp.14	0,0054	0,0023	0,0040	0,00868	0,02673	0,04328	0,09979	0,11883	0,14172	0,21813	0,30501	0,45205	0,74457	1	
	Comp.15	0,0032	0,0021	0,0035	0,00779	0,02402	0,03887	0,08962	0,10672	0,12728	0,19590	0,27393	0,40599	0,66871	0,89811	1

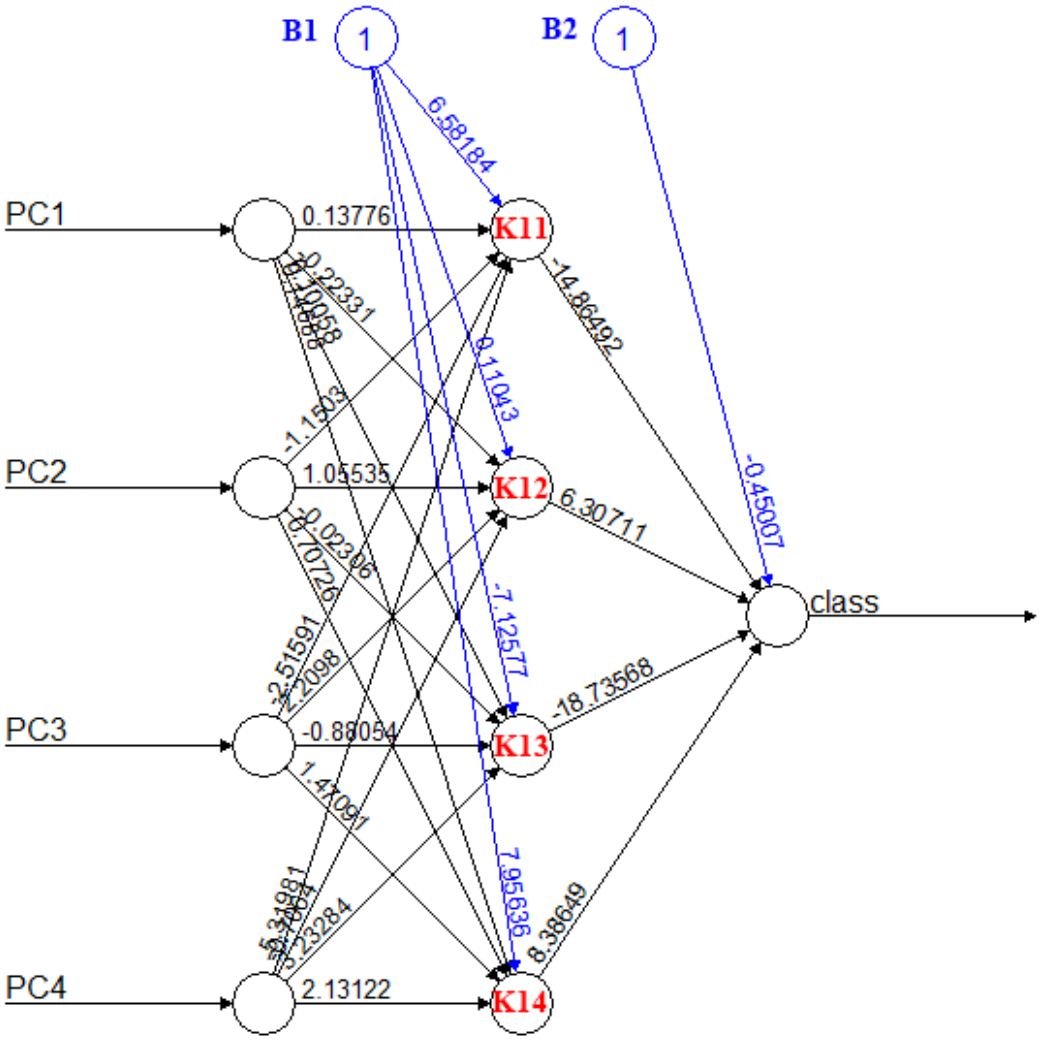
EK 8. Çalışmada kullanılan DT model parametreleri (Ham veri seti).

	Cins	Çam	Meşe
Ön İşlem	Ham	Ham	Ham
Metot	class	class	class
numresp	4	4	4
parms$split	1	1	1
parms$prior 1	0,5610766	0,5276753	0,5330189
parms$prior 2	0,4389234	0,4723247	0,4669811
minsplit	20	2	2
minbucket	7	1	1
cp	0,01	0	0
maxcompete	4	4	4
maxsurrogate	5	5	5
usesurrogate	2	2	2
surrugatestyle	0	0	0
maxdepth	30	30	30
xval	10	10	10

EK 9. Çalışmada kullanılan DT modellerine ait karmaşıklık parametreleri (CP) (Ham veri seti).

		CP	nsplit	rel error	xerror	xstd
Cins	1	0,64151	0	1	1	0,05145
	2	0,07547	1	0,35849	0,37736	0,03854
	3	0,06604	2	0,28302	0,31132	0,03561
	4	0,05660	4	0,15094	0,20755	0,02983
	5	0,01000	6	0,03774	0,15566	0,02615
		CP	**nsplit**	**rel error**	**xerror**	**xstd**
Çam	1	0,17578	0	1	1	0,06421
	2	0,05078	2	0,64840	0,73438	0,06122
	3	0,03906	4	0,54687	0,78125	0,06206
	4	0,02344	8	0,35156	0,76562	0,06179
	5	0,01563	9	0,32810	0,76562	0,06179
	6	0,01172	13	0,26562	0,80469	0,06243
	7	0,00977	19	0,19531	0,81250	0,06254
	8	0,00781	23	0,15625	0,81250	0,06254
	9	0,00391	39	0,03125	0,85156	0,06306
	10	0,00000	47	0	0,85156	0,06306
		CP	**nsplit**	**rel error**	**xerror**	**xstd**
Meşe	1	0,35354	0	1	1	0,07338
	2	0,11616	1	0,64646	0,66667	0,06810
	3	0,08586	3	0,41414	0,56566	0,06484
	4	0,04545	5	0,24242	0,47475	0,06109
	5	0,01212	7	0,15152	0,42424	0,05862
	6	0,01010	12	0,09091	0,46465	0,06062
	7	0,00505	18	0,03030	0,49495	0,06200
	8	0,00000	24	0	0,48485	0,06155

EK 10. Çalışmada cins bazında ayrımda kullanılan ANN model parametrelerinin grafiksel gösterimi (Ham veri seti).



EK 11. Çalışmada kullanılan ANN model parametreleri (Ham veri seti).

Veri Seti		Cins Ayrımı	Çam Türleri	Meşe Türleri
Ön İşlem		Ham	Ham	Ham
Saklı katman	*(hidden layer)*	4	4	4
Hata	*(error)*	0,0012	25,2270	5,2102
Ulaşılan eşik	*(threshold)*	0,0065	0,0095	0,0094
Adımlar	*(steps)*	282	4643	1200
Bias	B1 - K11	6,5818	37,0448	-8,6152
Ağırlıklar (Weights)	PC1 - K11	0,1378	-1,7229	0,2718
	PC2 -K11	-1,1503	6,9964	0,1839
	PC3 - K11	-2,5159	4,4925	-3,0715
	PC4 - K11	-5,3198	4,2945	16,5958
Bias	B1 - K12	0,1104	13,4674	6,1526
Ağırlıklar (Weights)	PC1 - K12	-0,2233	0,4763	-0,0663
	PC2 -K12	1,0553	-2,8184	0,9954
	PC3 - K12	2,2098	3,5775	1,2463
	PC4 - K12	-0,7064	-4,4511	2,9194
Bias	B1 - K13	-7,1258	-89,9119	-4,2367
Ağırlıklar (Weights)	PC1 - K13	0,1006	18,7444	1,5777
	PC2 -K13	-0,0231	8,3949	-0,9552
	PC3 - K13	-0,8805	24,4005	1,7943
	PC4 - K13	3,2328	-13,6209	-1,1621
Bias	B1 - K14	7,9564	41,8467	2,7411
Ağırlıklar (Weights)	PC1 - K14	0,1469	5,4834	-0,3894
	PC2 -K14	0,7073	-5,3397	-0,8675
	PC3 - K14	1,4709	-18,4265	1,6943
	PC4 - K14	2,1312	-48,8974	-12,0058
Bias	B2 - class	-0,4501	6,2003	-13,1227
Ağırlıklar (Weights)	K11 - class	-14,8649	-6,0905	-32,5365
	K12 - class	6,3071	-7,1904	8,6828
	K13 - class	-18,7357	-3,5777	6,9420
	K14 - class	8,3865	4,0708	8,1266

EK 12. Çalışmada kullanılan SVM model parametreleri (Ham veri seti).

	Cins ayrımı	Çam türleri	Meşe türleri
Tür	C-Classification	C-Classification	C-Classification
Kernel	Linear	Linear	Linear
C sabiti (C-costant)	1	1	1
Destek Vektör Sayısı	10 (4 6)	175 (89 86)	32 (15 17)
Polinomiyal derecesi	3	3	3
gamma	0,000240964	0,000240964	0,000240964
nu	0,5	0,5	0,5
epsilon	0,1	0,1	0,1
rho (negatif intercept)	-0,2035154	-0,1967259	-1,021235

EK 13. Cins ayrımında PLS-DA modelinde kullanılan Dalga boylarının SR aracılığıyla hesaplanan önem düzeyleri sıralaması (ilk 20 dalga boyu) (Ham veri seti).

Cins	Comp.1		Comp.2		Comp.3		Comp.4		...	Comp.15	
SR önem düzeyleri	**NIR.7405**	68,4428	**NIR.5702**	1,8206	**NIR.7230**	0,9409	**NIR.7230**	0,9299	...	**NIR.7230**	0,3882
	NIR.7407	68,2169	**NIR.5701**	1,8206	**NIR.7239**	0,9384	**NIR.7239**	0,9264	...	**NIR.7239**	0,3856
	NIR.7400	68,0907	**NIR.5704**	1,8201	**NIR.7201**	0,9383	**NIR.7224**	0,9245	...	**NIR.7180**	0,3842
	NIR.7419	68,0053	**NIR.5699**	1,8199	**NIR.7224**	0,9380	**NIR.7201**	0,9228	...	**NIR.7201**	0,3824
	NIR.7415	67,9221	**NIR.5706**	1,8197	**NIR.7211**	0,9373	**NIR.7211**	0,9222	...	**NIR.7224**	0,3823
	NIR.7398	67,8902	**NIR.5697**	1,8189	**NIR.7199**	0,9364	**NIR.7218**	0,9220	...	**NIR.7158**	0,3818
	NIR.7421	67,8659	**NIR.5708**	1,8189	**NIR.7209**	0,9364	**NIR.7216**	0,9212	...	**NIR.7178**	0,3817
	NIR.7413	67,7769	**NIR.5695**	1,8180	**NIR.7180**	0,9362	**NIR.7209**	0,9207	...	**NIR.7251**	0,3805
	NIR.7417	67,7488	**NIR.5710**	1,8177	**NIR.7218**	0,9356	**NIR.7212**	0,9205	...	**NIR.7211**	0,3805
	NIR.7394	67,7259	**NIR.5693**	1,8164	**NIR.7216**	0,9355	**NIR.7228**	0,9201	...	**NIR.7218**	0,3802
	NIR.7411	67,6936	**NIR.5712**	1,8160	**NIR.7193**	0,9355	**NIR.7199**	0,9200	...	**NIR.7199**	0,3798
	NIR.7403	67,6751	**NIR.5691**	1,8148	**NIR.7178**	0,9354	**NIR.7238**	0,9199	...	**NIR.7193**	0,3798
	NIR.7409	67,6551	**NIR.5714**	1,8142	**NIR.7212**	0,9354	**NIR.7203**	0,9196	...	**NIR.7203**	0,3795
	NIR.7382	67,6456	**NIR.5689**	1,8133	**NIR.7191**	0,9352	**NIR.7180**	0,9193	...	**NIR.7249**	0,3794
	NIR.7396	67,5473	**NIR.5716**	1,8123	**NIR.7197**	0,9352	**NIR.7214**	0,9189	...	**NIR.7216**	0,3792
	NIR.7401	67,4583	**NIR.5687**	1,8118	**NIR.7187**	0,9348	**NIR.7251**	0,9188	...	**NIR.7187**	0,3792
	NIR.7390	67,3379	**NIR.5718**	1,8105	**NIR.7203**	0,9342	**NIR.7193**	0,9186	...	**NIR.7191**	0,3792
	NIR.7392	67,2629	**NIR.5685**	1,8101	**NIR.7189**	0,9338	**NIR.7249**	0,9186	...	**NIR.7209**	0,3792
	NIR.7423	67,2317	**NIR.5683**	1,8089	**NIR.7228**	0,9337	**NIR.7222**	0,9186	...	**NIR.7238**	0,3790

EK 14. Çam türlerinin ayrımında PLS-DA modelinde kullanılan Dalga boylarının SR aracılığıyla hesaplanan önem düzeyleri sıralaması (ilk 20 dalga boyu) (Ham veri seti).

Çam	Comp.1		Comp.2		Comp.3		Comp.4		...	Comp.15	
SR önem düzeyleri	NIR.7992	397,5298	NIR.11779	0,3142	NIR.7338	0,0607	NIR.8404	0,0415	...	NIR.7338	0,0041
	NIR.8003	397,2161	NIR.11725	0,3141	NIR.7230	0,0604	NIR.8402	0,0415	...	NIR.7230	0,0040
	NIR.8001	397,1912	NIR.11590	0,3141	NIR.7326	0,0595	NIR.8406	0,0415	...	NIR.7373	0,0037
	NIR.7993	396,8996	NIR.11789	0,3140	NIR.7305	0,0594	NIR.8408	0,0415	...	NIR.7239	0,0037
	NIR.7999	396,8321	NIR.11748	0,3140	NIR.7239	0,0594	NIR.8400	0,0415	...	NIR.7353	0,0037
	NIR.8005	396,7990	NIR.11733	0,3139	NIR.7292	0,0592	NIR.8397	0,0415	...	NIR.7326	0,0036
	NIR.8007	396,7769	NIR.11735	0,3139	NIR.7353	0,0587	NIR.8398	0,0415	...	NIR.7305	0,0036
	NIR.7990	396,7436	NIR.11833	0,3139	NIR.7297	0,0586	NIR.8395	0,0415	...	NIR.7292	0,0036
	NIR.7997	396,4334	NIR.11633	0,3139	NIR.7241	0,0586	NIR.8410	0,0415	...	NIR.7378	0,0035
	NIR.7995	396,3642	NIR.11775	0,3139	NIR.7251	0,0585	NIR.8414	0,0415	...	NIR.7180	0,0035
	NIR.7988	396,0233	NIR.11609	0,3138	NIR.7320	0,0584	NIR.8393	0,0414	...	NIR.7403	0,0035
	NIR.8009	395,8316	NIR.11787	0,3138	NIR.7322	0,0584	NIR.8412	0,0414	...	NIR.7425	0,0035
	NIR.7986	395,4096	NIR.11777	0,3138	NIR.7286	0,0583	NIR.8416	0,0414	...	NIR.7434	0,0035
	NIR.8011	394,8269	NIR.11741	0,3137	NIR.7303	0,0581	NIR.8391	0,0414	...	NIR.7401	0,0035
	NIR.8647	394,2859	NIR.11625	0,3137	NIR.7261	0,0580	NIR.8387	0,0414	...	NIR.7374	0,0035
	NIR.7984	394,0522	NIR.11781	0,3137	NIR.7336	0,0579	NIR.8418	0,0414	...	NIR.7444	0,0035
	NIR.8645	393,8121	NIR.11636	0,3137	NIR.8331	0,0579	NIR.8389	0,0414	...	NIR.7359	0,0035
	NIR.8653	393,8057	NIR.11750	0,3136	NIR.8339	0,0579	NIR.8385	0,0414	...	NIR.7436	0,0035
	NIR.7982	393,5351	NIR.11731	0,3136	NIR.8329	0,0579	NIR.8383	0,0414	...	NIR.7452	0,0035

EK 15. Meşe türlerinin ayrımında PLS-DA modelinde kullanılan Dalga boylarının SR aracılığıyla hesaplanan önem düzeyleri sıralaması (ilk 20 dalga boyu) (Ham veri seti).

Meşe	Comp.1		Comp.2		Comp.3		Comp.4		...	Comp.15	
SR önem düzeyleri	NIR.8236	39,6565	NIR.8996	4,7523	NIR.7633	1,5816	NIR.7573	0,3379	...	NIR.7585	0,1498
	NIR.8242	39,6447	NIR.9006	4,7522	NIR.7629	1,5815	NIR.7569	0,3379	...	NIR.7583	0,1497
	NIR.8229	39,6436	NIR.8998	4,7516	NIR.7631	1,5813	NIR.7575	0,3379	...	NIR.7594	0,1497
	NIR.8227	39,6324	NIR.8994	4,7513	NIR.7639	1,5812	NIR.7571	0,3379	...	NIR.7596	0,1497
	NIR.8235	39,6311	NIR.9004	4,7506	NIR.7637	1,5812	NIR.7560	0,3378	...	NIR.7587	0,1497
	NIR.8231	39,6179	NIR.9000	4,7496	NIR.7617	1,5812	NIR.7558	0,3378	...	NIR.7608	0,1497
	NIR.8225	39,6175	NIR.9002	4,7493	NIR.7621	1,5812	NIR.7562	0,3378	...	NIR.7571	0,1497
	NIR.8244	39,6165	NIR.8987	4,7491	NIR.7635	1,5811	NIR.7556	0,3378	...	NIR.7602	0,1497
	NIR.8240	39,6154	NIR.8979	4,7489	NIR.7619	1,5809	NIR.7577	0,3377	...	NIR.7600	0,1497
	NIR.8233	39,6132	NIR.8991	4,7485	NIR.7627	1,5809	NIR.7567	0,3377	...	NIR.7590	0,1497
	NIR.8238	39,6101	NIR.9008	4,7480	NIR.7648	1,5808	NIR.7563	0,3377	...	NIR.7577	0,1497
	NIR.8221	39,6096	NIR.8989	4,7479	NIR.7623	1,5808	NIR.7554	0,3377	...	NIR.7573	0,1496
	NIR.8254	39,6043	NIR.8985	4,7478	NIR.7641	1,5808	NIR.7583	0,3376	...	NIR.7589	0,1496
	NIR.8223	39,5906	NIR.8992	4,7477	NIR.7616	1,5807	NIR.7585	0,3376	...	NIR.7579	0,1496
	NIR.8250	39,5906	NIR.8983	4,7472	NIR.7625	1,5806	NIR.7565	0,3376	...	NIR.7592	0,1496
	NIR.8256	39,5876	NIR.9010	4,7467	NIR.7650	1,5806	NIR.7579	0,3376	...	NIR.7575	0,1496
	NIR.8252	39,5803	NIR.8981	4,7461	NIR.7646	1,5806	NIR.7548	0,3376	...	NIR.7610	0,1496
	NIR.8248	39,5739	NIR.9012	4,7447	NIR.7614	1,5805	NIR.7552	0,3376	...	NIR.7598	0,1496
	NIR.8246	39,5710	NIR.8977	4,7443	NIR.7644	1,5805	NIR.7550	0,3376	...	NIR.7604	0,1496

EK 16. DT modellerinde kullanılan dalga boylarının değişken önem sıralamaları (Ham veri seti).

Cins		Çam		Meşe	
Değişken	Değişken Önemi	Değişken	Değişken Önemi	Değişken	Değişken Önemi
8306 - 8316	110,872	4000	22,419	7469-7479	23,820
11521 - 11546	35,650	5301	19,158	10946-10967	17,934
7074 - 7076	28,930	5297 -5299	18,098	5166-5176	12,014
11787 - 11970	27,655	7180	17,627	5259-5276	10,510
11987 - 11999	12,650	7193	16,644	4000	7,143
7236 - 7232	6,636	5284 - 5292	15,256	5045-5055	5,786
		4002 - 4009	11,668	10273 - 10292	5,094
		7197 - 7203	10,746	4002 - 4009	4,810
		4075 - 4081	9,681	4723 - 4732	3,898
		9945 - 9970	9,453	9876- 9897	3,103
		4102	7,954	4306 - 4316	2,857
		4092 - 4100	7,268	7583 - 7592	2,000
		5305 - 5311	6,153	…	…
		7158 - 7178	5,898		
		4083 - 4084	5,163		
		4071 - 4073	4,518		
		5922 - 5934	4,200		
		5303	3,902		
		7303 - 7319	3,507		
		8277 - 8287	3,500		
		…	…		

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Fatma Diğdem TUNCER
Doğum Yeri	Fatih
Doğum Tarihi	23.10.1987
Uyruğu	☑T.C.❒ Diğer:
Telefon	0212 3382400
E-Posta Adresi	digdem.kangaloglu@istanbul.edu.tr
Web Adresi	http://avesis.istanbulc.edu.tr/digdem.kangaloglu/



Eğitim Bilgileri	
Lisans	
Üniversite	İstanbul Üniversitesi
Fakülte	Orman Fakültesi
Bölümü	Orman Endüstri Mühendisliği Bölümü
Mezuniyet Yılı	2010

Yüksek Lisans	
Üniversite	İstanbul Üniversitesi-Cerrahpaşa
Enstitü Adı	Lisansüstü Eğitim Enstitüsü
Anabilim Dalı	Orman Endüstri Mühendisliği Anabilim Dalı
Programı	Orman Biyolojisi ve Odun Koruma Teknolojisi

Makale ve Bildiriler

Doğu, A.D., Bakir, D., Tirak Hizal, K., Tuncer, F.D., Candan, Z., Ünsal, Ö., 2015, Anatomical Structures of Thermally Compressed Poplar (Populus Spp.) Wood, *Pro Ligno,* 11, 366-372.

Doğu, A.D., Bakir, D., Tirak Hizal, K., Tuncer, F.D., Candan, Z., Ünsal, Ö., 2015, Anatomical Structures of Thermally Compressed Poplar (*Populous* spp.) Wood, *The 10th International Conference Wood Science and Engineering in The Third Millennium,* Braşov, ROMANYA, 5-7 Kasım 2015, 382-389.

Doğu, A.D., Bakir, D., Tuncer, F.D., Korkut, S., Candan, Z., 2017, Microstructure of Heat treated Wood Material, *International Conference On Advance and Innovations in Engineering ICAIE* , Elazığ, Türkiye, 10-12 Mayıs 2017, 1079-1080.

Doğu, A.D., Bakir, D., Tuncer, F.D., Sözen, E., Yilgör, N., 2017, Microscopic Evaluation of a Historical Wood from Mudurnu in Turkey, *Ecology 2017,* Kayseri, Türkiye, 11-13 Mayıs 2017, 239.

Doğu, A.D., Bakir, D., Tuncer, F.D., Tırak Hızal, K., Ünsal, Ö., Candan, Z., 2016, Microscopic Investigation of Defects in Thermally Compressed Poplar Wood Panels, *Maderas-Ciencia Y Tecnologia,* 18(2), 337-348.

Doğu, A.D., Tirak Hizal, K., Bakir, D., Tuncer, F.D., Candan, Z., Ünsal, Ö., 2015, Anatomical Structures of Thermally Compressed Paulownia (*Paulownia* spp.) Wood, *Proceedings of the 58th International Convention of Society of Wood Science and Technology,* Wyoming, ABD, 7-12 Haziran 2015, 304.

Doğu, A.D., Tirak Hizal, K. , Tuncer, F.D., Bakangil, A., 2017, Structural Evaluation of Natural Weathered Wooden Window Frames, *4th International Furniture And Decoration Congress (UMDK2017),* Düzce, Türkiye, 19-21 Ekim 2017.

Doğu, A.D., Tirak Hizal, K., Tuncer, F.D., Bakir, D., 2017, Comparative Investigations on Microstructure of Thermally Compressed Wood Species, *Uluslararası Mühendislik Araştırmaları Sempozyumu (UMAS),* Düzce, Türkiye, 11-13 Eylül 2017, 131-133.

Doğu, A.D., Tuncer, F.D., 2012, Historical Journey to Istanbul, *2nd Symposium of Archaeological and Historical Wood Utilization,* Egmond aan Zee, Hollanda, 28-30 Eylül 2012.

Doğu, A.D., Tuncer, F.D., Bakir, D., Candan, Z., 2017, Characterizing Microscopic Changes of Paulownia Wood under Thermal Compression, *Bioresources*, 12 (3), 5279-5295.

Doğu, A.D., Yilgör, N., Mantanis, G., Tuncer, F.D., 2017, Structural Evaluation of a Timber Construction Element Originating from the Great Metéoron Monastery in Greece, *Bioresources,* 12(2), 2433-2451.

Tuncer, F. D., Akdeniz, E., Doğu, A.D., 2019, The Feasibility of Near Infrared Spectroscopy for Classification of Pine Species, *International Conference on Data Science, Machine Learning and Statistics 2019 (DMS-2019*), Van, Türkiye, 26 - 29 June 2019, 70.

Tuncer, F.D., Doğu, A.D., 2017, The Interaction of the Physical and Macroscopic Properties against Heat Treatment, *13th IUFRO International Wood Drying Conference "Wood Drying in Developing Countries",* İstanbul, Türkiye, 13-16 Eylül 2017.

Tuncer, F.D., Doğu, A.D., 2018, Effects of heat treatment on some macroscopic and physical properties of Scots pine sapwood and heartwood, *Forestist,* 68(2), 93-100.