

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI**

**MİKROORGANİZMALARIN ELEKTRONİK BURUN
TEKNOLOJİSİ KULLANILARAK TANIMLANMASI VE
SINIFLANDIRILMASI**

**Hazırlayan
Bekir Hakan AKSEBZECİ**

**Danışmanlar
Prof. Dr. Musa Hakan ASYALI
Prof. Dr. Sadık KARA**

Doktora Tezi

**Temmuz 2011
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI**

**MİKROORGANİZMALARIN ELEKTRONİK BURUN
TEKNOLOJİSİ KULLANILARAK TANIMLANMASI VE
SINIFLANDIRILMASI**

(Doktora Tezi)

**Hazırlayan
Bekir Hakan AKSEBZECİ**

**Danışmanlar
Prof. Dr. Musa Hakan ASYALI
Prof. Dr. Sadık KARA**

**Bu çalışma; Erciyes Üniversitesi Bilimsel Araştırma Projeleri Birimi
tarafından FBT-07-14 kodlu proje ile desteklenmiştir.**

**Temmuz 2011
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Bekir Hakan AKSEBZECİ

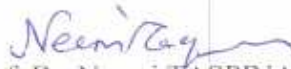
“Mikroorganizmaların Elektronik Burun Teknolojisi Kullanılarak Tanımlanması ve Sınıflandırılması” adlı Doktora tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.



Bekir Hakan AKSEBZECİ



Prof. Dr. Musa Hakan ASYALI



Prof. Dr. Necmi TAŞPINAR
Elektrik - Elektronik Mühendisliği
Anabilim Dalı Başkanı

Prof. Dr. Musa Hakan ASYALI danışmanlığında Bekir Hakan AKSEBZECİ tarafından hazırlanan “Mikroorganizmaların Elektronik Burun Teknolojisi Kullanılarak Tanımlanması ve Sınıflandırılması” adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalında **Doktora** tezi olarak kabul edilmiştir.

14/07/2011

JÜRİ:

Danışman : Prof. Dr. Musa Hakan ASYALI



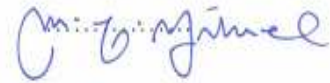
Üye : Prof. Dr. Aydın AKAN



Üye : Prof. Dr. Recai KILIÇ



Üye : Prof. Dr. Mehmet Emin YÜKSEL



Üye : Prof. Dr. Kenan DANIŞMAN



ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun 16.108/2011 tarih ve 2011.128-05... sayılı kararı ile onaylanmıştır.



Prof. Dr. Necmettin MARAŞLI
Enstitü Müdürü

TEŞEKKÜR

Tez konusunun belirlenmesi ve çalışmalarındaki katkılarından dolayı değerli hocam Prof. Dr. Sadık KARA'ya ve tez çalışmasının yürütülmesinde ve sonuçlandırılmasında, fikirleri ve bilgisi ile bana yol göstererek yardımlarını esirgemeyen danışman hocam Prof. Dr. Musa Hakan ASYALI'ya teşekkür ederim.

Tez kapsamında gerçekleştirilen çalışmalarda mikroorganizma kültürlerinin hazırlanmasında ve tezin tıbbi boyutlarındaki yorumları ile katkı sağlayan Prof. Dr. Hatice ÖZBİLGE ve Yrd. Doç. Dr. Esmâ KAYA'ya teşekkür ederim. Ayrıca diş kök kanal enfeksiyonlarına sebep olan mikroorganizmalara yönelik yürüttüğümüz çalışmalarda değerli fikirlerinden ve emeklerinden dolayı Doç. Dr. Özgür ER ve Dr. Yasemin KAHRAMAN'a teşekkür ederim.

Tez çalışmam boyunca maddi ve manevi desteklerini üzerimden eksik etmeyen aileme ve eşime teşekkür ederim.

MİKROORGANİZMALARIN ELEKTRONİK BURUN TEKNOLOJİSİ KULLANILARAK TANIMLANMASI VE SINIFLANDIRILMASI

Bekir Hakan AKSEBZECİ
Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Doktora Tezi, Temmuz 2011
Tez Danışmanları: Prof. Dr. Musa Hakan ASYALI
Prof. Dr. Sadık KARA

ÖZET

Enfeksiyon hastalıklarının tedavisinde, enfeksiyona yol açan hastalık etkeni patojen mikroorganizmanın, en kısa sürede doğru olarak tanımlanması büyük öneme sahiptir. Bazı durumlarda patojen mikroorganizmaların tanımlanması; zahmetli ve zaman alıcı olabilmektedir. Bundan dolayı araştırmacılar, patojen mikroorganizmaları rutin mikrobiyolojik tekniklere göre daha hızlı tespit edebilen yöntemler üzerine yönelmişlerdir. Bu yöntemlerden birisi de, mikroorganizma kültürlerinin ortama yaydıkları kokulardan yararlanarak, elektronik burun cihazı ile mikroorganizma türünün tespiti üzerinedir.

Bu tezde; üç ayrı mikroorganizma seti üzerinde, elektronik burun cihazı kullanılarak, mikroorganizma türüne göre sınıflandırma çalışmaları yapılmıştır. Mikroorganizma setlerinden birincisi, çeşitli enfeksiyon hastalıklarına sebep olan ve rutinde sıklıkla karşılaşılan 8 mikroorganizmadan meydana gelmektedir. İkinci ve üçüncü setler ise, dış kök kanal enfeksiyonlarına sebep olan 7 mikroorganizmadan oluşmaktadır. Mikroorganizma setlerinden elektronik burun cihazı ile elde edilen koku verilerine, sınıflandırma performansına etkilerini incelemek amacıyla; farklı veri ön-işleme, boyut azaltım ve ölçekleme yöntemleri uygulanmıştır. Sınıflandırma aşamasında ise; k -en yakın komşuluk algoritması (k -NN), Ayırma Analizi (DA) ve Destek Vektör Makineleri (SVM) yöntemleri kullanılmıştır.

Bu tez kapsamında gerçekleştirilen çalışmalarda, en başarılı sınıflandırma sonuçlarının genellikle SVM yöntemi ile sağlandığı görülmüştür. Elde edilen sonuçlar, patojen mikroorganizmaların tespitinde, elektronik burun cihazının rutin mikrobiyolojik tekniklere alternatif veya destekleyici bir yöntem olarak kullanılabileceğini göstermiştir.

Anahtar Kelimeler: Mikroorganizma; elektronik burun cihazı; sınıflandırma yöntemleri; k -en yakın komşuluk algoritması; ayırma analizi; destek vektör makineleri.

IDENTIFICATION AND CLASSIFICATION OF MICROORGANISMS USING ELECTRONIC NOSE

Bekir Hakan AKSEBZECİ

Erciyes University, Graduate School of Natural and Applied Sciences

Ph.D. Thesis, July 2011

Thesis Supervisors: Prof. Dr. Musa Hakan ASYALI

Prof. Dr. Sadık KARA

ABSTRACT

In the treatment of infectious diseases, as soon as the correct identification of pathogenic microorganisms that cause infection, has a great importance. In some cases, identification of pathogenic microorganisms can be laborious and time consuming. Therefore, researchers have tried to improve methods that can detect pathogenic microorganisms more rapidly than the routine microbiological techniques. One of these methods is the identification of microorganism species using an electronic nose device by means of the culture odors that they emit to environment.

In this thesis, classification studies were carried out according to the species of microorganism on the 3 different microorganism sets using an electronic nose device. The first of microorganism sets consists of 8 microorganisms which cause various infectious diseases and are frequently encountered in routine. Both the second and third sets were comprised of the 7 microorganisms which caused root canal infections. Odor data which were obtained from microorganism sets using the electronic nose device were applied to different data pre-processing, dimension reduction, and scaling methods in order to investigate the effects on classification performance. In classification stage, *k*-nearest neighbor algorithm (*k*-NN), Discriminant Analysis (DA) and Support Vector Machines (SVM) methods were used.

In studies within the scope of this thesis, the most successful classification results have been generally obtained by SVM method. The results obtained have shown that electronic nose devices can be used as an alternative or supportive method of routine microbiological techniques in identification of pathogenic microorganisms.

Keywords: Microorganism; electronic nose; classification methods; *k*-nearest neighbor algorithm; discriminant analysis; support vector machines.

İÇİNDEKİLER

	<u>Sayfa</u>
BİLİMSEL ETİĞE UYGUNLUK SAYFASI	ii
YÖNERGEYE UYGUNLUK SAYFASI	iii
KABUL VE ONAY SAYFASI	iv
TEŞEKKÜR.....	v
ÖZET	vi
ABSTRACT.....	vii
İÇİNDEKİLER	viii
TABLolar LİSTESİ.....	xi
ŞEKİLLER LİSTESİ	xiv
KISALTMALAR	xvi

GİRİŞ

Genel Bilgiler.....	1
Literatür Taraması.....	2
Tezin Amacı ve Önemi	4

1. BÖLÜM

MATERYAL ve METOT

1.1. Elektronik Burun Cihazı	6
1.1.1. Koku Alma Duyumuz	6
1.1.2. Koku Ölçümü	7
1.1.3. Elektronik Burun Cihazının Yapısı.....	8
1.1.3.1. Örnek Taşıma Birimi.....	9
1.1.3.2. Sensör Birimi	9
1.1.3.3. Örüntü Tanıma Birimi.....	10
1.1.4. Uygulama alanları	11
1.1.5. Cyranose 320	11
1.2. Çalışılan Mikroorganizma Setleri	16

1.2.1. Mikroorganizma Seti-1	16
1.2.2. Mikroorganizma Seti-2	18
1.2.3. Mikroorganizma Seti-3	20
1.3. Sınıflandırma Sürecinin Aşamaları	20
1.4. Mikroorganizma Koku Verilerinin Toplanması.....	21
1.5. Ön-işleme Yöntemleri.....	24
1.5.1. Sensör Cevap Modelleri ve Sensör Referans Direnç Değer Noktaları	24
1.6. Boyut Azaltım Yöntemleri.....	26
1.6.1. Standart Sapmaya Göre Boyut Azaltımı	27
1.6.2. Temel Bileşenler Analizine Göre Boyut Azaltımı.....	28
1.7. Ölçekleme Yöntemleri	30
1.8. Sınıflandırma Yöntemleri.....	31
1.8.1. k -En Yakın Komşuluk Algoritması	32
1.8.2. Ayırma Analizi.....	34
1.8.2.1. Doğrusal Sınıflandırıcı	36
1.8.2.2. Diyagonal Doğrusal Sınıflandırıcı	37
1.8.2.3. İkinci Dereceden Sınıflandırıcı	39
1.8.2.4. Diyagonal İkinci Dereceden Sınıflandırıcı.....	40
1.8.2.5. Mahalanobis Sınıflandırıcı	41
1.8.3. Destek Vektör Makineleri.....	42
1.8.3.1. Doğrusal Destek Vektör Makineleri	43
1.8.3.2. Doğrusal Olmayan Destek Vektör Makineleri.....	48
1.8.3.3. Çok Sınıflı Destek Vektör Makineleri	50
1.8.3.3.1. SVM (1-r) Yöntemi.....	50
1.8.3.3.2. SVM (1-1) Yöntemi	51
1.8.3.4. LIBSVM Kütüphanesi/Yazılımı	52
1.9. Sınıflandırma Performansının Değerlendirilmesi	53
1.9.1. Çapraz Geçerlilik	54

2. BÖLÜM

SINIFLANDIRMA ÇALIŞMALARI

2.1. Sınıflandırma Çalışmalarında Kullanılan Yöntemler.....	57
---	----

2.2. Mikroorganizma Seti-1 için Sınıflandırma Sonuçları	59
2.2.1. k -NN Algoritması ile Sınıflandırma.....	60
2.2.2. DA ile Sınıflandırma	61
2.2.3. SVM ile Sınıflandırma	62
2.2.4. Sınıflandırma Yöntemlerinin Sonuçlarının Karşılaştırılması	66
2.3. Mikroorganizma Seti-2 için Sınıflandırma Sonuçları	67
2.3.1. k -NN Algoritması ile Sınıflandırma.....	67
2.3.2. DA ile Sınıflandırma	70
2.3.3. SVM ile Sınıflandırma	72
2.3.4. Sınıflandırma Yöntemlerinin Sonuçlarının Karşılaştırılması	84
2.4. Mikroorganizma Seti-3 için Sınıflandırma Sonuçları	85
2.4.1. k -NN Algoritması ile Sınıflandırma.....	85
2.4.2. DA ile Sınıflandırma	86
2.4.3. SVM ile Sınıflandırma	87
2.4.4. Sınıflandırma Yöntemlerinin Sonuçlarının Karşılaştırılması	92

3. BÖLÜM

SONUÇLAR ve ÖNERİLER

3.1. Sonuçlar	93
3.2. Öneriler	97

KAYNAKLAR	98
-----------------	----

ÖZGEÇMİŞ	104
----------------	-----

TABLOLAR LİSTESİ

Tablo 1.1. Mikroorganizma Seti-1'i oluşturan mikroorganizma türleri ve özellikleri.....	17
Tablo 1.2. Mikroorganizma Seti-2'yi oluşturan mikroorganizma türleri ve özellikleri.....	19
Tablo 1.3. Çalışılan mikroorganizma setlerinin, içerdikleri tür (sınıf), örnek (numune) ve koku verileri sayıları.	22
Tablo 1.4. Farklı sensör cevap modellerinin eşitlik ifadeleri ve açıklamaları.....	24
Tablo 1.5. Farklı sensör referans direnç değer noktalarının eşitlik ifadeleri ve açıklamaları.	25
Tablo 1.6. Sensör referans direnç değer noktaları ve sensör cevap modellerine göre, 9 farklı ön-işleme yönteminin eşitlik ifadeleri ve kısaltmaları.	26
Tablo 2.1. Veri setlerinin k -NN algoritması kullanılarak 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	60
Tablo 2.2. Veri setlerinin DA yöntemi kullanılarak 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	61
Tablo 2.3. Veri setlerinin MATLAB'deki yerleşik SVM fonksiyonları ile 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	62
Tablo 2.4. Veri setlerinin LIBSVM kütüphanesindeki çekirdek fonksiyonları ile 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Çekirdek fonksiyonları: (a) doğrusal, (b) ikinci dereceden polinom, (c) üçüncü dereceden polinom, (d) RBF.....	63
Tablo 2.5. Veri setlerinin 5-kat CV metoduna göre SVM ile sınıflandırılmasında, çekirdek fonksiyonları için hesaplanan en yüksek başarı oranları (%).	66
Tablo 2.6. Mikroorganizma seti-1 için sınıflandırma yöntemlerine göre 5-kat CV metodu ile elde edilen en yüksek başarı oranları ve bu oranları sağlayan çeşitli yöntemler.	67
Tablo 2.7. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, k -NN algoritması kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	

Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.	68
Tablo 2.8. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, DA yöntemi kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.....	70
Tablo 2.9. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, MATLAB'deki yerleşik SVM fonksiyonları ile 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.	73
Tablo 2.10. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki doğrusal fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.....	76
Tablo 2.11. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki ikinci dereceden fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.....	78
Tablo 2.12. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki üçüncü dereceden fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.....	80
Tablo 2.13. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki RBF fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.	82

Tablo 2.14. Mikroorganizma seti-2'nin 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarına ait örnekleri için, sınıflandırma yöntemlerine göre 5-kat CV metodu ile elde edilen en yüksek başarı oranları ve bu oranları sağlayan çeşitli yöntemler.....	84
Tablo 2.15. Veri setlerinin k-NN algoritması kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	86
Tablo 2.16. Veri setlerinin DA yöntemi kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	87
Tablo 2.17. Veri setlerinin MATLAB'deki yerleşik SVM fonksiyonları ile 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).	88
Tablo 2.18. Veri setlerinin LIBSVM kütüphanesindeki çekirdek fonksiyonları ile 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Çekirdek fonksiyonları: (a) doğrusal, (b) ikinci dereceden polinom, (c) üçüncü dereceden polinom, (d) RBF.....	89
Tablo 2.19. Veri setlerinin 5-kat CV metoduna göre SVM ile sınıflandırılmasında, çekirdek fonksiyonları için hesaplanan en yüksek başarı oranları (%).	91
Tablo 2.20. Mikroorganizma seti-3 için sınıflandırma yöntemlerine göre 5-kat CV metodu ile elde edilen en yüksek başarı oranları ve bu oranları sağlayan çeşitli yöntemler.	92
Tablo 3.1. Literatürdeki elektronik burun cihazı ile mikroorganizmaların sınıflandırılması üzerine yapılmış olan çalışmalardan bazılarının; konuları, kullandıkları sınıflandırma yöntemleri ve elde ettikleri başarı oranları.	96

ŞEKİLLER LİSTESİ

Şekil 1.1.	Elektronik burun cihazını oluşturan 3 temel yapı.	8
Şekil 1.2.	Cyranose 320 elektronik burun cihazı.....	12
Şekil 1.3.	Temizleme çevriminin şematik yapısı [43].	13
Şekil 1.4.	Örnekleme çevriminin şematik yapısı [43].	13
Şekil 1.5.	Bir sensörün, kokuya maruz kalmasıyla meydana gelen polimer yapıdaki ve direnç değerindeki değişim.	14
Şekil 1.6.	Bir koku verisindeki 32 sensörün, PCnose yazılımı tarafından oluşturulan anlık değişim grafiği.	15
Şekil 1.7.	Kanlı agar besiyerinde üremiş Staphylococcus aureus bakteri kültürü.....	18
Şekil 1.8.	Kullanılan steril cam tüpler ve türbidometre cihazı.	19
Şekil 1.9.	Eppendorf tüpte steril dış içeren bir örnek.	20
Şekil 1.10.	Mikroorganizma koku verilerinin sınıflandırma sürecini oluşturan aşamalar ve uygulanan yöntem sayıları.	21
Şekil 1.11.	Mikroorganizma koku verilerinin toplandığı sistem.	22
Şekil 1.12.	Bir koku verisindeki tek bir sensörün direnç değişim grafiği ve koku alma safhaları.	23
Şekil 1.13.	Bir koku veri matrisi ile bir sensöre ait direnç değişim grafiğindeki sensör referans direnç değer noktalarının birlikte gösterimi.	25
Şekil 1.14.	Mikroorganizma seti-1 için, <i>d1</i> ön-işleme metodu uygulanarak elde edilmiş koku veri setindeki, farklı mikroorganizma örneklerine ait sensör direnç değerlerinin değişimleri.	28
Şekil 1.15.	İki boyutlu bir veri setine PCA uygulandıktan sonra elde edilen temel bileşenler (PC1 ve PC2).	29
Şekil 1.16.	Mikroorganizma seti-1 için, <i>d1</i> ön-işleme metodu uygulanarak oluşturulan koku veri setinin, PCA işleminden sonraki temel bileşenlerinin, toplam değişkenliği temsil etme yüzdeleri.	30
Şekil 1.17.	<i>k</i> -NN sınıflandırmasına örnek bir durum ($k = 5$).	33
Şekil 1.18.	Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin doğrusal sınıflandırıcı ile sınıflandırılması.	37

Şekil 1.19. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin diyagonal doğrusal sınıflandırıcı ile sınıflandırılması.	38
Şekil 1.20. Normal dağılıma uygun eşit varyansa sahip iki değişkenli iki sınıfa ait verilerin diyagonal doğrusal sınıflandırıcı ile sınıflandırılması.	39
Şekil 1.21. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin ikinci dereceden sınıflandırıcı ile sınıflandırılması.	40
Şekil 1.22. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin diyagonal ikinci dereceden sınıflandırıcı ile sınıflandırılması.	41
Şekil 1.23. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin Mahalanobis sınıflandırıcı ile sınıflandırılması.	42
Şekil 1.24. İki sınıflı bir veri setinin; (a) farklı doğrusal karar sınırları ile sınıflandırılması, (b) SVM ile en uygun karar sınırı belirlenerek sınıflandırılması.	43
Şekil 1.25. İki sınıflı bir veri setinin doğrusal bir karar sınırı ile ayrılması.	44
Şekil 1.26. İki sınıflı bir veri setinin, hatalı örneklerle izin verecek şekilde doğrusal bir karar sınırı ile ayrılması.	47
Şekil 1.27. Doğrusal olarak sınıflandırılamayan 2 boyutlu verilerin, 3 boyutlu uzaya taşınarak sınıflandırılması.	49
Şekil 1.28. SVM (1-r) sınıflandırmasına örnek bir durum.	51
Şekil 1.29. SVM (1-1) sınıflandırmasına örnek bir durum.	52
Şekil 1.30. İki sınıflı bir veri seti için: (a) aşırı öğrenme durumundaki (over-fitting) bir sınıflandırıcı, (b) genelleştirme performansı yüksek bir sınıflandırıcı.	53
Şekil 1.31. 5-kat çapraz geçerlilik yöntemi [47].	55
Şekil 2.1. Mikroorganizma koku verilerinin sınıflandırılmasında kullanılan yöntemler.	57

KISALTMALAR

ANN	: Yapay Sinir Ağları (<i>Artificial Neural Networks</i>)
ATCC	: Amerikan Kültür Koleksiyonu (<i>American Type Culture Collection</i>)
CFU/mL	: Birim miktarda bulunan mikroorganizma koloni sayısı (<i>Colony-Forming Units per milliliter</i>)
CP	: İletken Polimer (<i>Conducting Polymer</i>)
CV	: Çapraz Geçerlilik (<i>Cross Validation</i>)
DA	: Ayırma Analizi (<i>Discriminant Analysis</i>)
DSMZ	: Alman Kültür Koleksiyonu (<i>Deutsche Sammlung von Mikroorganismen und Zellkulturen</i>)
FCM	: Bulanık C-Ortalamalar (<i>Fuzzy C-Means</i>)
GA	: Genetik Algoritma (<i>Genetic Algorithm</i>)
GC-MS	: Gaz Kromatografisi - Kütle Spektroskopisi (<i>Gas Chromatography - Mass Spectrometry</i>)
k-NN	: k-En Yakın Komşuluk Algoritması (<i>k-Nearest Neighbor algorithm</i>)
LDA	: Doğrusal Ayırma Analizi (<i>Linear Discriminant Analysis</i>)
LOOCV	: Tek bir örneğin/verinin dışarıda bırakıldığı çapraz geçerlilik (<i>Leave One Out Cross Validation</i>)
MLP	: Çok Katmanlı Algılayıcı (<i>Multi Layer Perceptron</i>)
PCA	: Temel Bileşenler Analizi (<i>Principal Component Analysis</i>)
PNN	: Olasılıksal Sinir Ağı (<i>Probabilistic Neural Network</i>)
QCM	: Kuvars Kristal Mikrobaleans (<i>Quartz Crystal Microbalance</i>)
RBFN	: Radyal Tabanlı Fonksiyon Ağı (<i>Radial Basis Function Network</i>)
RSKK	: Refik Saydam Kültür Koleksiyonu
SAW	: Yüzey Akustik Dalga (<i>Surface Acoustic Wave</i>)
SF	: Serum Fizyolojik
SOM	: Özdüzenleyici Haritalar (<i>Self Organizing Maps</i>)
SVM	: Destek Vektör Makineleri (<i>Support Vector Machines</i>)
VOCs	: Uçucu Organik Bileşikler (<i>Volatile Organic Compounds</i>)

GİRİŞ

Genel Bilgiler

Enfeksiyon hastalıkları; bakteri, mantar, parazit ve virüs gibi değişik patojen mikroorganizmalar tarafından oluşturulan bir hastalık grubudur. Bu hastalıkların tedavisinde, öncelikle hastalık etkeni patojen mikroorganizmanın doğru olarak tanımlanması gerekmektedir. Mikrobiyoloji laboratuvarlarında, patojen mikroorganizmaların belirlenmesinde altın standart yöntem; kültür ve kültür sonrası tanımlamadır (identifikasyon). Kültür; patojen mikroorganizmanın uygun besiyerlerine ekilerek saf olarak üretilmesidir. Besiyerlerine ekim, laboratuvara gelen klinik örneğin ve izolasyonu istenen mikroorganizmanın türüne göre yapılır. Besiyerlerine ekilen tüm kültürler, izolasyonu yapılacak mikroorganizmanın türüne göre, 24 saat veya daha uzun bir süre etüvde inkübe edilir. Araştırılan hastalığın özelliğine göre; besin ortamı, üreme sıcaklığı, üreme süresi ve oksijen yoğunluğu ancak uygun koşullarda tutulursa patojen mikroorganizmalar üretilebilir ve tanımlanabilir. Normal flora içermeyen ve steril olması gereken hasta örneklerinde (kan, beyin-omurilik sıvısı gibi), üretilen mikroorganizmalar büyük bir olasılıkla hastalık etkenidirler. Boğaz sürüntüsü, balgam ve dışkı gibi örneklerde ise hastalık etkeni mikroorganizmalar, normal flora ile birlikte üreyeceklerdir. Buradan hastalık etkeni olan mikroorganizmanın saf kültür halinde ayırt edilmesi gerekir. Saf kültür halinde üretilen hastalık etkeni, daha sonra tanımlanma işlemlerine tabii tutulur ve patojen mikroorganizmanın ismi konmaya çalışılır. Mikroorganizmaların tanımlanma işlemlerinde çeşitli besiyerlere ekim yapma, boyama, biyokimyasal testler ve serolojik testler gerektiği şekilde kullanılırlar [1].

Günümüzde rutin mikrobiyolojik tekniklerle hastadan alınan örneklerin işlenmesi ve patojen mikroorganizmanın tespit edilmesi genellikle 24–48 saat sürmektedir [2]. Bazı

mikroorganizmalar için bu süre daha uzun olabilmektedir. Özellikle hızlı antibiyotik tedavisi gerektiren menenjit ve sepsis gibi enfeksiyon hastalıkları için, patojen mikroorganizmanın en kısa sürede tanımlanması hayati öneme sahiptir. Dolayısıyla araştırmacılar, patojen mikroorganizmaları rutin mikrobiyolojik tekniklere göre daha hızlı tespit edebilen yöntemler üzerine yönelmişlerdir. Bu yöntemlerden birisi de, mikroorganizma kültürlerinin ortama yaydıkları kokulardan yararlanarak mikroorganizma türünün tespiti üzerinedir. Araştırmacıların bu konu üzerinde durmalarının sebebi; mikrobiyoloji uzmanlarının, bazı mikroorganizmaları sadece kültürlerinin kokusundan algılayıp ön tanı koyabilmeleridir [3]. Bu amaçla elektronik burun (*electronic nose*) cihazlarının kullanımı araştırılmaya başlanmıştır.

Elektronik burun, insanın koku algılama sistemi temel alınarak geliştirilmiş bir cihazdır [4]. Bu cihazın yapısını; kokuları ve uçucu bileşenleri sensörlerin bulunduğu kısma taşıyan birim, sensör birimi ve örüntü tanıma birimi oluşturmaktadır. Elektronik burun cihazlarının en önemli elemanları kimyasal gaz sensörleridir. Sensörlerden alınan koku verileri (örüntüler - *patterns*), bir veritabanı meydana getirir ve bu veritabanı çeşitli sınıflandırma yöntemlerinde eğitim için kullanılarak örüntü tanıma işlemi gerçekleştirilir. Elektronik burun cihazlarının başlıca uygulama alanları; gıda, çevre sağlığı ve tıptır [5].

Literatür Taraması

Bu tez çalışmasında; elektronik burun cihazı kullanılarak, farklı mikroorganizma türlerine ait kültürlerin kokularına göre sınıflandırılması üzerinde durulmuştur. Literatürde bu konu ile ilgili çeşitli çalışmalar bulunmaktadır. Bunlardan bazıları; üst solunum yolu [6], idrar yolu [7, 8], kulak-burun-boğaz [4, 9] ve göz [9, 10] gibi enfeksiyon hastalıklarına özgü mikroorganizmalar üzerine yapılan çalışmalardır. Diğer çalışmalarda ise belirli mikroorganizmalar üzerine yoğunlaşmıştır [3, 11-13]. Literatürde bu alanda yapılmış olan çalışmalardan bazılarının, konuları ve sınıflandırma yöntemleri aşağıda kısaca özetlenmiştir:

Lai ve arkadaşları tarafından, üst solunum yolu enfeksiyonlarına sebep olan çeşitli patojen bakterilerin algılanması ve ayırt edilmesinde elektronik burun cihazının kullanımı araştırılmıştır [6]. Öznitelik çıkarımı ve sınıflandırıcı olarak sırasıyla Temel

Bileşenler Analizi (*Principal Component Analysis – PCA*) ve Ayırma Analizi (*Discriminant Analysis – DA*) tercih edilmiştir.

İdrar örneklerindeki mikroorganizmaların, elektronik burun cihazı kullanılarak algılanması ile ilgili çalışma, Pavlou ve arkadaşları tarafından yapılmıştır [7]. Yöntem olarak PCA, DA ve genetik algoritma tabanlı sinir ağları (*Genetic Algorithm Based Neural Network – GA-NN*) uygulanmıştır.

Kan ve idrar örneklerindeki bakterilerin elektronik burun cihazı kullanılarak belirlenmesi konusundaki çalışma, Yates ve arkadaşları tarafından yapılmıştır [8]. PCA, Sammon haritalama (*Sammon mapping*) ve doğrusal olmayan zaman serileri modeli (*nonlinear time series model*) yöntemleri uygulanmıştır.

Dutta ve arkadaşları, kulak-burun-boğaz enfeksiyonlarına sebep olan, *Staphylococcus aureus* bakterisinin 3 ayrı alt türünün elektronik burun cihazı kullanılarak tanımlanması üzerine çalışma yapmışlardır [4]. Öznitelik çıkarımı; PCA, Bulanık C-Ortalamalar (*Fuzzy C-Means – FCM*) ve Özdüzenleyici Haritalar (*Self Organizing Maps – SOM*) metodlarının birleştirilmesi ile elde edilmiştir. Sınıflandırmada Çok Katmanlı Algılayıcı (*Multi Layer Perceptron – MLP*), Olasılıksal Sinir Ağı (*Probabilistic Neural Network – PNN*) ve Radyal Tabanlı Fonksiyon Ağı (*Radial Basis Function Network – RBFN*) yöntemleri kullanılmıştır.

Kulak-burun-boğaz ve göz enfeksiyonlarına sebep olan bakterilerin elektronik burun cihazı kullanılarak sınıflandırılması konusundaki çalışma, Boilot ve arkadaşları tarafından gerçekleştirilmiştir [9]. Göz enfeksiyonlarına sebep olan 6 ve kulak-burun-boğaz enfeksiyonlarına sebep olan 4 bakteri olmak üzere, iki ayrı bakteri seti üzerinde çalışılmıştır. PCA, MLP ve RBFN yöntemlerini sınıflandırma amaçlı uygulamışlardır.

Dutta ve arkadaşları, göz enfeksiyonlarına sebep olan 6 farklı bakteri türünün sınıflandırılması ile ilgili çalışmalarında elektronik burun cihazı kullanmışlardır [10]. Sınıflandırmada PCA, FCM, SOM, MLP, PNN ve RBFN yöntemlerini uygulamışlardır.

Elektronik burun cihazı kullanılarak 12 farklı bakteri türü ve 1 patojen mayanın Yapay Sinir Ağları (*Artificial Neural Networks – ANN*) ile sınıflandırılması, Gibson ve arkadaşları tarafından gerçekleştirilmiştir [3].

McEntegart ve arkadaşları, elektronik burun cihazı kullanarak 2 farklı bakteri kültürünün sınıflandırılmasında başarı kaydetmişlerdir [11]. Sınıflandırmada PCA yönteminden yararlanmışlardır.

Klinik olarak 10 önemli mikroorganizmanın elektronik burun cihazı kullanılarak tanımlanması, Moens ve arkadaşları tarafından gerçekleştirilmiştir [12]. Öznelik çıkarmada PCA, sınıflandırmada ANN ve k en yakın komşuluk algoritması (*k-Nearest Neighbor – k-NN*) uygulanmıştır.

Pavlou ve arkadaşları, 2 farklı anaerob bakteri türünün ayırt edilmesinde elektronik burun cihazının kullanımını araştırmışlardır [13]. Yöntem olarak PCA, Genetik Algoritma (*Genetic Algorithm – GA*) ve ANN uygulanmış ve bakteri türlerinin başarılı şekilde sınıflandırıldığı rapor edilmiştir.

Tezin Amacı ve Önemi

Bu tez kapsamında; üç ayrı mikroorganizma seti üzerinde çalışmalar yapılmıştır. Mikroorganizma setlerinden elektronik burun cihazı ile elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması amaçlanmıştır.

Mikroorganizma setlerinden birincisi, çeşitli enfeksiyon hastalıklarına sebep olan ve rutinde sıklıkla karşılaşılan 8 mikroorganizmadan meydana gelmektedir. Bu çalışmada, mikroorganizmaların petri kaplarında hazırlanan kültürlerinden elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması amaçlanmıştır.

Mikroorganizma setlerinden ikincisi, diş-kök kanal enfeksiyonlarına sebep olan 7 mikroorganizmadan meydana gelmektedir. Her mikroorganizmanın, serum fizyolojik içeren 4 farklı konsantrasyonda kültürleri hazırlanmıştır. Bu çalışma ile aynı konsantrasyondaki kültürlerden elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması amaçlanmıştır. Bu sayede farklı konsantrasyonlardaki etkinliğin test edilmesi hedeflenmiştir.

Mikroorganizma setlerinden üçüncüsünde de, ikinci settteki mikroorganizmalar kullanılmıştır. Bu çalışmada ise, diş hekimliğindeki uygulamalara benzer bir durum oluşturabilmek için, her mikroorganizmanın steril diş kökleri ve serum fizyolojik içeren

sabit konsantrasyonda kültürleri hazırlanmıştır. Bu kültürlerden elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması amaçlanmıştır.

Enfeksiyon hastalarına gerekli tedavinin hızlı bir şekilde başlanabilmesi, patojen mikroorganizmanın en kısa sürede doğru olarak tanımlanmasına bağlıdır. Bazı durumlarda patojen mikroorganizmaların tanımlanması; zahmetli ve zaman alıcı olabilmektedir. Bundan dolayı hekimler; acil vakalarda ve patojen mikroorganizmaların zor üretildiği durumlarda, tanımlanma için gerekli süreyi beklemeksizin, tecrübeye ve tahmine dayalı (ampirik) bir tedavi yöntemi uygulamak zorunda kalmaktadırlar. Sonuçta bu seçim; hasta için bazen başarı, bazen de başarısızlıkla neticelenmektedir. Bu tez çalışmasının önemi; enfeksiyonlara yol açan mikroorganizmaların daha kısa sürede tanımlanması için, elektronik burun cihazının kullanımının araştırılmasından kaynaklanmaktadır.

Özellikle ikinci ve üçüncü mikroorganizma setlerinde; diş kök kanal enfeksiyonlarına sebep olan mikroorganizmaların sınıflandırılmasında, elektronik burun cihazının kullanımı araştırılarak literatüre katkı yapmak hedeflenmiştir. Diş kök kanal enfeksiyonlarında, etken mikroorganizmalar genellikle anaerob (oksijensiz ortamda üreyen/yaşayan) oldukları için üretilmeleri ve tanımlanmaları zordur. Bu nedenle, diş hekimleri hastalarına genellikle ampirik tedavi uygulamaktadırlar. Literatürde bu alanda elektronik burun cihazının kullanıldığı tek çalışma; enfekte olmuş diş kök kanalının kokusunun, elektronik burun cihazıyla analiz edilmesi ve buradan izole edilen bakteri ile ilişkilendirilmesi üzerinedir [14].

Mikroorganizma setlerinden elde edilen koku verilerine, farklı veri ön-işleme, boyut azaltım ve ölçekleme yöntemleri uygulanarak sınıflandırma performansına etkileri incelenmiştir. Bu yöntemlerden sonra sınıflandırma aşamasına geçilmiştir. Literatürdeki mikroorganizma koku verilerinin sınıflandırılması ile ilgili çalışmalar tarandığında, genellikle ANN tabanlı yöntemlerin tercih edildiği görülmüştür. Bu çalışmada ise sınıflandırma yöntemleri olarak; k -NN, DA ve Destek Vektör Makineleri (*Support Vector Machines* – SVM) kullanılmıştır. Bu yöntemlerin her birinin alt metotları da araştırılarak, sınıflandırma performansları karşılaştırılmıştır.

1. BÖLÜM

MATERYAL ve METOT

1.1. Elektronik Burun Cihazı

Elektronik burun, insanın koku algılama sistemi temel alınarak geliştirilmiş bir cihazdır [4]. Elektronik burun cihazları ile ilgili araştırmaların tarihçesi incelendiğinde, 1980’li yılların başlarında Avrupa, Asya ve Kuzey Amerika’da birbirinden bağımsız çalışmalar yürütüldüğü görülmektedir. Avrupalı araştırmacılar, memelilerin koku algılama sistemini taklit eden bir yapıyı meydana getirmek için [18], Kuzey Amerika’daki araştırmacılar zehirli gazların tanımlanması ve niceliğinin belirlenmesinde sahada çalışabilecek taşınabilir bir cihaz geliştirmek için [19] ve Asyalı araştırmacılar ise belirli kokuları algılayan sistemler meydana getirmek için [20] uğraş vermişlerdir [5]. Kimyasal sensörler ve örüntü tanıma birimi içeren, bir elektronik burun cihazının tasarımı ile ilgili yayınlanan ilk çalışma Persaud ve Dodd tarafından 1982 yılında yapılmıştır [18]. Bu tür çalışmalar pek çok ticari teşebbüslere olanak sağlamıştır. Elektronik burun ve kimyasal sensör teknolojileri alanında, NATO destekli bir çalıştay Ağustos 1991’de İzlanda’da düzenlenmiştir. Bu çalıştay, bu alana duyulan ilgiyi daha da artırmış ve şu anda dünya çapında pek çok grup elektronik burun teknolojisi üzerine çalışmalarını sürdürmektedir [15].

1.1.1. Koku Alma Duyumuz

Koku alma duyusu, temel olarak havadaki kimyasalların algılanmasına dayanmaktadır. Bu nedenle yalnızca bazı moleküllerini havaya karıştırabilen/buharlaştırabilen maddelerin kokusu alınabilmektedir. Koku yayan madde teneffüs edildiğinde, koku molekülleri burun boşluğunun üst kısmındaki bölgeye ulaşmaktadırlar. Bu bölge, kokuya duyarlı sinir hücrelerinin bulunduğu, ince yapılı bir zar şeklindedir. Hücrelerin

burun boşluğuna bakan yüzeyleri mukus tabaka ile kaplıdır. Mukus tabakanın burun boşluğuna bakan yüzeyinde, siliya (*cilia*) adı verilen kıvrım şeklinde özel çıkıntılar bulunmaktadır. Bu çıkıntılarla sinir hücrelerinin yüzey alanı artırılarak daha çok koku molekülü ile temas etmesi sağlanmaktadır. Burnumuzun koku bölgesinde yaklaşık 10–20 milyon kadar sinir hücresi bulunmaktadır. Her sinir hücresi, hücre zarına gömülü reseptör adı verilen birçok protein taşımaktadır. Her sinir hücresi sadece bir tek tip reseptör içermektedir. Bu reseptörlerin her biri, kokuları oluşturan belirli kimyasallara özgü tamamlayıcı şekillere sahiptir. Dolayısıyla her koku molekülü her reseptöre bağlanamamaktadır. Koku molekülleri reseptörlerden herhangi birine bağlandığında, hücrenin elektrik yükü değişmektedir. Bu yük değişimi, sinir hücrelerinin uzantılarının bağlı olduğu, koku soğanı (*olfactory bulb*) olarak adlandırılan yapıyı uyarmaktadır. Burada farklı sinir hücrelerinden gelen sinyaller işlenmektedir. Koku soğanında işlenen sinyaller daha sonra beyin koku merkezlerine iletilerek kokunun algılanması ve tanınması sağlanmaktadır [15, 16].

1.1.2. Koku Ölçümü

Günümüzde koku ölçümüne yönelik, genel olarak iki yaklaşım bulunmaktadır. Bunlar;

- insanlardan oluşan bir panelin kokuyu belirlemesine,
- gaz kromatografisi ve kütle spektrometresi (GC-MS) kullanılarak kokuyu oluşturan kimyasal bileşenlerin hassas olarak belirlenmesine,

dayanmaktadır [15].

İnsanlar binlerce farklı kokuyu birbirinden ayırt edebilme kapasitesine sahiptirler. Koku ölçümünde eğitimli kişilerden oluşan bir panelin değerlendirmesi altın standart olarak kabul edilmektedir. Ancak panelistlerle yapılan bu değerlendirmenin bazı zayıf yönleri de mevcuttur. Bunlar arasında;

- kişinin aynı yoğunluktaki kokuyu farklı nitelendirmesi,
- kişiye özgü farklı hassasiyet (öznellik),
- uzun süreli maruziyette daha az duyarlı olma (adaptasyon),
- kişinin fiziksel durumu (yorgunluk, hastalık gibi),
- kişinin zihinsel durumu
- tehlikeli gazlara maruz kalmanın ortaya çıkaracağı sorunlar

sayılabilir [15].

GC-MS kokuyu oluşturan kimyasal bileşenlerin belirlenmesinde yararlanılan hassas tekniklerdir. Ancak bu tekniklerin;

- zaman alıcı olmaları
- pahalı olmaları,
- taşınabilir boyutlarda olmamaları

özellikle gerçek zamanlı (*real time*) uygulamalarda, nadiren kullanılmalarına yol açmaktadır.

Belirtilen bu nedenlerden dolayı, araştırmacılar tarafından elektronik bir sistem kullanılarak kokuların daha hızlı ve ucuz bir şekilde belirlenmesi hedeflenmiştir. Bu amaçla, kimyasal gaz sensörlerinin elektronik burun cihazlarının tasarımında kullanımı için çeşitli araştırmalar yapılmaya başlanmıştır.

1.1.3. Elektronik Burun Cihazının Yapısı

Bir elektronik burun cihazı, Şekil 1.1’de görülen 3 temel yapıdan meydana gelmektedir [5]. Bu yapılar:

- kokuları ve uçucu bileşenleri sensörlerin bulunduğu kısma taşıyan örnek taşıma birimi (*sample delivery unit*),
- sensör birimi (*sensor unit*)
- örüntü tanıma birimi (*pattern recognition unit*)

Sensörlerden alınan koku verileri (örüntüler), bir veritabanı oluşturur ve bu veritabanı eğitim için kullanılarak çeşitli sınıflandırma teknikleri ile örüntü tanıma işlemi gerçekleştirilir.



Şekil 1.1. Elektronik burun cihazını oluşturan 3 temel yapı.

1.1.3.1. Örnek Taşıma Birimi

Bu birim örnek kokuları ve uçucu bileşenleri sensörlerin bulunduğu kısma taşıma işlevini yerine getirmektedir. Örnekler arasındaki nem, sıcaklık ve konsantrasyon gibi parametrelerin değişimini minimum seviyede tutmakla görevli kısımdır.

1.1.3.2. Sensör Birimi

Elektronik burun cihazlarının en önemli elemanları kimyasal gaz sensörleridir. Bu sensörlerin kokuya maruz kalmaları ile meydana gelen değişiklikler, elektriksel sinyallere dönüştürülerek analiz edilmeye uygun hale getirilmektedir. Genelde bu sensörler, tek tip sensörlerden veya değişik tipte sensörlerden oluşan bir dizi (*array*) halinde kullanılırlar [5].

Elektronik burun cihazlarında en çok kullanılan sensörler; iletkenliği değişen sensörler, piezoelektrik sensörler ve metal oksit yarıiletken alan etkili transistör (*Metal Oxide Semiconductor Field-Effect Transistor* - MOSFET) sensörlerdir.

İletkenliği değişen sensörler: Metal oksit yarıiletken (*Metal Oxide Semiconductor* - MOS) ve iletken polimer (*Conducting Polymer* – CP) sensörlerdir. Her ikisi de uçucu organik bileşiklere (*Volatile Organic Compounds* – VOC) maruz kaldığında direnç değişimi göstermektedir. MOS sensörleri temin etmek daha kolay olduğu için, elektronik burun cihazlarında daha çok tercih edilmeye başlanmıştır. CP sensörlerin cevapları zaman ile kayma gösterebilmektedir. CP sensörler, MOS sensörlere göre neme daha duyarlıdır. Bu duyarlılık, uçucu organik bileşiklerin sensör cevaplarını bozucu etki gösterebilmektedir [15].

Piezoelektrik sensörler: Kuvartz kristal mikrodenge (*Quartz Crystal Microbalance* - QCM) yüzey akustik dalga (*Surface Acoustic Wave* - SAW) sensörlerdir. QCM sensörler; her iki tarafında tel ile bağlanmış metal elektrotlar bulunan, birkaç milimetre çapında titreşen bir diskten meydana gelir. Bu sensör karakteristik frekansında (10 MHz ~ 30 MHz) rezonanstır. Gaz molekülleri disk yüzeyinde absorbe olduğunda, diskin kütlesi artmakta, dolayısıyla rezonans frekansı düşmektedir. SAW sensörler daha yüksek frekanslarda çalıştıklarından, frekans değişimleri daha büyük olmaktadır. Tipik bir SAW sensörü 100'lerce MHz frekansında çalışmaktadır. SAW sensörleri, gaza

maruz kaldıklarındaki kütle değişimini QCM ile aynı oranda ölçebilmektedir. Bu nedenle bazı durumlarda SAW sensörler, QCM sensörlere göre daha az hassas olabilirler. QCM ve SAW sensörlerin bir dezavantajı, her ikisinin de iletkenliği değişen sensörlere göre daha karmaşık elektronik devrelere ihtiyaç duymalarıdır [15].

MOSFET sensörler: Bu sensörlerin çalışma prensibi; uçucu organik bileşiklerin MOSFET'in kapı (gate) kısmındaki metal ile reaksiyon oluşturmaya ve bu reaksiyonun sensör çıkışında gerilim değişimi meydana getirmesine dayanmaktadır [15]. MOSFET gaz sensörü ile ilgili yayınlanan ilk çalışma, 1975 yılında Lundstorm tarafından yapılmıştır [42]. Bu çalışmada Hidrojen gazının (H_2) algılanması için MOSFET'in kapı kısmında paladyum malzemesi kullanılmıştır. Daha sonraki çalışmalarda farklı gazların algılanması için, MOSFET'in kapı kısmında farklı metallerin (platin ve iridyum gibi) uygulanması araştırılmıştır [21]. CP sensörlerde olduğu gibi MOSFET sensörlerinde cevapları zaman ile kayma gösterebilmektedir [15].

1.1.3.3. Örüntü Tanıma Birimi

Bu birim, sensörlerden alınan koku verileri üzerinde çeşitli işlemler yaparak muhtemel kokularla ilgili bir veritabanı oluşturmakla ve bu veritabanından yararlanarak örnek (belirlenmesi istenilen) bir kokunun tanımlanmasını sağlamakla görevlidir. Örüntü tanıma birimi 4 ardışık safhadan meydana gelmektedir. Bu safhalar aşağıda kısaca açıklanmıştır.

- Ön-işleme (Pre-processing): Bu safhada; sensör cevaplarının referans noktalarının ayarlanması, tüm sensörler için sensör cevap sahalalarının normalize edilmesi ve sensör cevap modellerinin belirlenmesi için gerekli işlemler gerçekleştirilir.
- Öznitelik çıkarımı (Feature extraction): Öznitelik çıkarımı; veri setinden örüntü tanımaya yardımcı özellikleri belirlemek ve veri setinin boyutunu azaltmak için kullanılır. Öznitelik çıkarımında genellikle, PCA ve doğrusal ayırma analizi (*Linear Discriminant Analysis - LDA*) gibi lineer dönüşüm metotlarından yararlanılır.
- Sınıflandırma (Classification): Veri seti düşük boyutlu uzaya taşındıktan sonra muhtemel koku gruplarına ait örüntüler, çeşitli sınıflandırma yöntemleri

kullanılarak belirlenir. Bu yöntemlere örnek olarak; k -NN, Bayes sınıflandırıcılar ve ANN verilebilir.

- Geçerlilik (Validation): Bir sınıflandırıcının verileri optimum şekilde sınıflandırabilmesi için gerekli parametrelerinin belirlenmesi; veri setinin bir kısmının eğitim, bir kısmının da geçerlilik olarak ayrılması ile sağlanır. Bu sayede genelleştirme performansı yüksek bir sınıflandırıcı belirlenmiş olur.

1.1.4. Uygulama alanları

Elektronik burun cihazlarının başlıca uygulama alanları; gıda, çevre sağlığı ve tıptır. Gıda alanında elektronik burun cihazları; balık ve et gibi ürünlerin tazeliğinin tespitinde [22, 23], bitkisel yağların ve içeceklerin sınıflandırılmasında [24-27], kahve ve tahılların kalite sınıflandırmasında [28-31], peynirlerin ve meyvelerin olgunlaşmasının tespitinde [32-34] kullanılmıştır.

Çevre sağlığı alanında elektronik burun cihazları; içme sularının kalite kontrolünde [35], atık suların biyokimyasal durumunun izlenmesinde [36], çevre kirliliğinin takibinde [37] ve iç mekânlardaki hava kalitesinin izlenmesinde [38] kullanılmıştır.

Tıp alanında elektronik burun cihazları; üst solunum yolu [6], idrar yolu [7, 8], kulak-burun-boğaz [4, 9] ve göz [9, 10] enfeksiyon hastalıklarına sebep olan mikroorganizmaların tespiti, nefes örneklerinden akciğer kanserinin belirlenmesi [39, 40] ve böbrek rahatsızlığı teşhisinde [41] kullanılmıştır.

1.1.5. Cyranose 320

Cyranose 320, Kaliforniya Teknoloji Enstitüsü (Caltech) tarafından 1999 yılında geliştirilmiş bir elektronik burun cihazıdır [17]. Bu cihaz, 32 iletken polimer (CP) sensörden oluşan bir sensör dizisini ve yerleşik örüntü tanıma birimini ihtiva etmektedir. Cyranose 320'nin hafif ve taşınabilir olması diğer elektronik burun cihazlarına karşı üstünlük/avantaj sağlamaktadır. Bu cihazın görünümü Şekil 1.2'de verilmiştir.



Şekil 1.2. Cyranose 320 elektronik burun cihazı.

Cihazın koku tanıma/sınıflandırma işlemi sırasıyla şu şekilde gerçekleştirilmektedir: Cihaz ilk olarak üzerinde çalışılacak kokular/gazlar ile eğitilerek her bir koku grubunun/sınıfının sayısal koku izini (*smellprint*) çıkarmaktadır. Daha sonra cihaza bilinmeyen bir koku tutulduğunda, cihaz bu kokuyu da sayısallaştırmakta ve önceden kayıtlı koku grupları ile karşılaştırarak karar vermektedir.

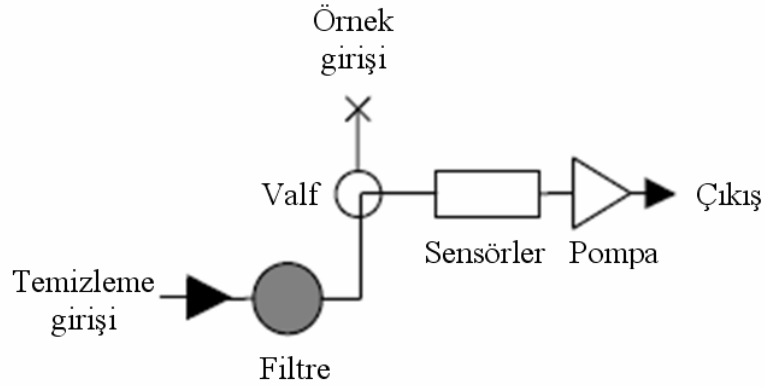
Literatürde, mikroorganizmaların tanımlanmasının/sınıflandırılmasının Cyranose 320 kullanılarak gerçekleştirildiği çeşitli çalışmalar bulunmaktadır [7, 9-12]. Bu nedenden dolayı bu tez çalışmasında elektronik burun cihazı olarak Cyranose 320 tercih edilmiştir. Cyranose 320 temel olarak 3 yapıdan meydana gelmektedir. Bunlar;

- Örneklem birimi
- Sensör birimi
- Örüntü tanıma birimi

Örneklem Birimi: Cyranose 320'nin örneklem sistemi; cihaz içerisindeki pompa ve valf ile ayarlanan iki çevrimden meydana gelmektedir. Bunlar; temizleme çevrimi (purge cycle) ve örneklem çevrimidir (*sampling cycle*).

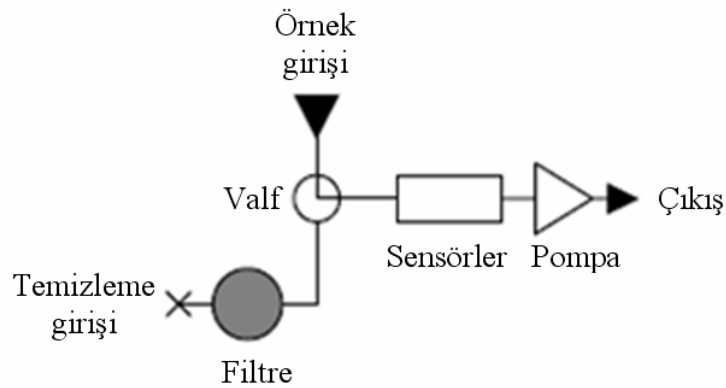
Temizleme çevriminde sensörler örnek kokuya maruz bırakılmadan önce, referans bir gaz veya hava ile temizlenir. Bu çevrim esnasında cihazdaki valf açık pozisyonda olup temizleme girişi kısmındaki gazın sensörlere ulaşmasına olanak sağlar. Sensörlere ulaşan gaz daha sonra cihazın çıkış kısmından dışarı verilir. Bu çevrim sensörlerin referans direnç değerlerinin (*baseline values*) R^0 ölçüldüğü kısımdır. Bazı durumlarda, arka plandaki kimyasalların veya nemin etkisini elimine etmek için opsiyonel bir filtre kullanılabilir. Bu çevrim, örneklem çevriminden sonra sensörleri temizlemek için

(*sensor refresh*) tekrar gerçekleştirilir. Şekil 1.3’de temizleme çevriminin şematik yapısı görülmektedir.



Şekil 1.3. Temizleme çevriminin şematik yapısı [43].

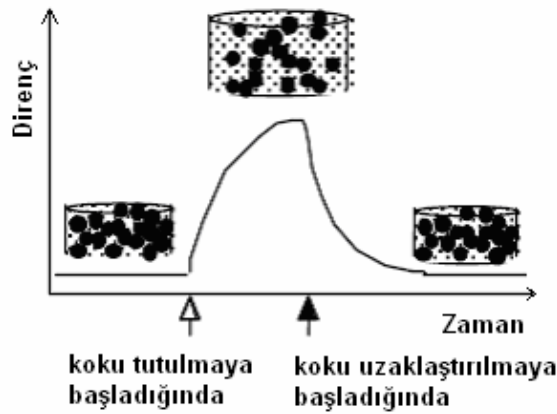
Örnekleme çevriminde sensörlere örnek koku uygulanır. Bu çevrim esnasında cihazdaki valf kapalı pozisyonda olup örnek girişi kısmındaki gazın sensörlere ulaşmasına olanak sağlar. Sensörlere ulaşan gaz daha sonra cihazın çıkış kısmından dışarı verilir. Bu çevrim sensörlerin maksimum direnç değerlerinin $R^{\max}$ ölçüldüğü kısımdır. Şekil 1.4’de örnekleme çevriminin şematik yapısı görülmektedir.



Şekil 1.4. Örnekleme çevriminin şematik yapısı [43].

Sensör Birimi: Cyranose 320’de, 32 iletken polimer sensör içeren bir sensör dizisi bulunmaktadır. Cihazda kullanılan sensörler, alüminyum oksit bir tabaka içerisinde iki elektriksel iletken ihtiva etmektedirler. Sensörler kokuya maruz kaldıklarında, polimer yapıları kokuyu içlerine çekerek sünger gibi şişmekte ve hacimleri artmaktadır. Bu

durumda her bir sensörün iletkenleri arasındaki direnç değeri artmaktadır. Bunun nedeni, iletkenler arasındaki karbon yolun kesintiye uğramasıdır. Örnek koku uzaklaştırıldığında ise, her bir sensördeki polimer yapı büzülerek iletkenler arasındaki karbon yol eski halini almaktadır. Dolayısıyla sensörlerin direnç değerleri de normal duruma dönmektedir. Şekil 1.5’de kokuya maruz kalan bir sensörün direnç değerindeki değişim görülmektedir.



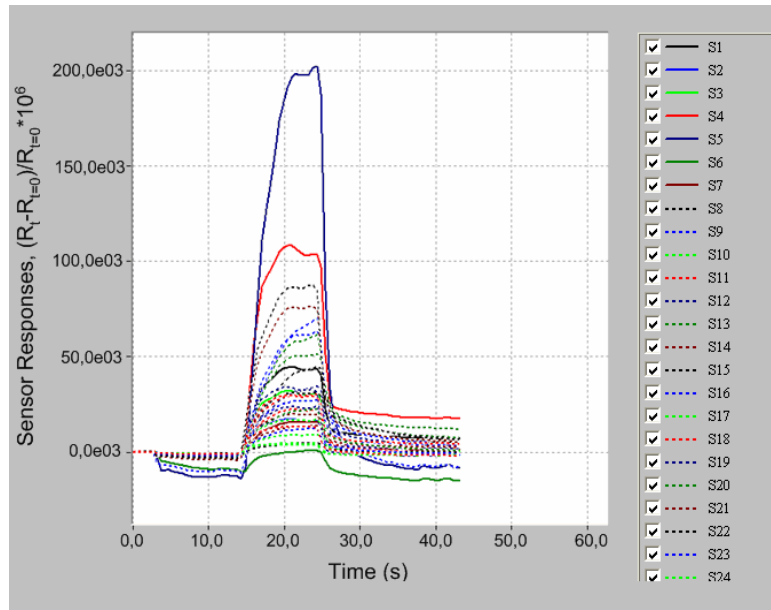
Şekil 1.5. Bir sensörün, kokuya maruz kalmasıyla meydana gelen polimer yapıdaki ve direnç değerindeki değişim.

Cihazın ölçüm sistemi, her bir sensörün bağıl direnç değişiminin $(R^{\max} - R^0)/R^0$ sensör modeline göre belirlenmesine dayanmaktadır. Buradaki;

- R^0 ; sensörlerin referans gaz/hava ile temizlenirkenki ilk direnç değerlerini,
- $R^{\max}$; sensörlerin örnek koku uygulandığındaki maksimum direnç değerlerini

temsil etmektedir. Örnek koku alma işlemi esnasında, cihaz bilgisayardan bağımsız olarak kullanılırken koku verileri bu sensör modeline göre kayıt edilmektedir.

Koku verilerinin cihaz yerine bilgisayarda analizi için, her bir sensördeki direnç değişimi gerçek-zamanlı (*real-time*) olarak bilgisayara aktarılabilir. Bu işlem cihazla birlikte verilen PCnose yazılımı yardımıyla gerçekleştirilmektedir. Bir koku verisindeki 32 sensörün, bu yazılım tarafından $(R^t - R^0)/R^0$ sensör modeline göre oluşturulan, anlık değişim grafiği Şekil 1.6’da görülmektedir. Buradaki R^t ; sensörlerin anlık direnç değerlerini ifade etmektedir.



Şekil 1.6. Bir koku verisindeki 32 sensörün, PCnose yazılımı tarafından oluşturulan anlık değişim grafiği.

Örüntü Tanıma Birimi: Koku verilerinin Cyranose 320 tarafından sınıflandırılabilmesi için bu verilerin, işaret işleme ve veri analiz tekniklerini içeren örüntü tanıma birimine uygulanması gerekmektedir. İşaret işleme kısmında; örnek koku için her bir sensörün direnç değişimi $(R^{\max} - R^0)/R^0$ sensör modeline göre belirlenmektedir. Daha sonra elde edilen sensör direnç değişim değerleri, sınıflandırmaya yardımcı olması beklenen normalizasyon ve ölçekleme yöntemlerine tabi tutulabilmektedir. İşaret işleme kısmından geçen koku verileri, cihaz tarafından k -NN, K-ortalamalar (K-means) veya Kanonik Ayırma Analizi (*Canonical Discriminant Analysis* - CDA) kullanılarak sınıflandırılabilir.

Cyranose 320; önceden seçilen sınıflandırma yöntemine göre, eğitim numunelerinin kokuları üzerinde eğitilir ve daha sonra hangi numune sınıfına/grubuna ait olduğu bilinmeyen bir numunenin kokusuna maruz bırakıldığında, bu numunenin sınıfını/grubunu tanımlamaya çalışır. Cihaz bu numunenin tanımlanmasını,

- 1) Numune sınıfı (*****)
- 2) Numune sınıfı (***)
- 3) Numune sınıfı (*)
- 4) Numune sınıfı
- 5) Numune sınıfı 1, Numune sınıfı 2 (?)

6) *Bilinmeyen (Unknown)*

şeklindeki ifadelerden birisi ile ekranında gösterir. Buradaki yıldızlı gösterim sınıflandırma güvenilirliğini derecelendirmektedir. Beş yıldızlı gösterim en çok güvenilir ve yıldızsız gösterim ise en az güvenilir sınıflandırmayı ifade etmektedir. Yıldızsız gösterim, sadece cihazın yazılımında ‘her zaman bir sınıfa atama yap’ seçeneği seçilirse görüntülenebilir. Eğer numune birden fazla sınıfa ait ise, cihaz numunenin olması mümkün sınıflardan olasılığı en yüksek iki sınıfını, 5. maddede gösterildiği şekilde ekranında belirtir. Eğer numune herhangi bir sınıfa atanamamışsa, cihazın ekranında ‘*Bilinmeyen*’ (*Unknown*) şeklinde gösterilir [43].

Cyranose 320 bilgisayardan bağımsız olarak çalışırken, en fazla 6 sınıfa ait koku verilerini sınıflandırabilmektedir. Ayrıca bu sınıflarda en az 5, en fazla 10 koku verisinin bulunması gerekmektedir. Bu kısıtlamaları aşmak ve farklı veri ön-işleme ve boyut azaltım yöntemleri ile sınıflandırma yöntemlerinin araştırılması için, mikroorganizma koku verileri bilgisayar ortamında analiz edilmiştir.

1.2. Çalışılan Mikroorganizma Setleri

Bu tezdeki mikroorganizma koku verilerinin sınıflandırılmasına ilişkin yapılan çalışmalar, 3 ayrı mikroorganizma seti üzerinde uygulanmıştır. Çalışmadaki mikrobiyolojik işlemler, Erciyes Üniversitesi Eczacılık Fakültesi Farmasötik Mikrobiyoloji Laboratuvarı’nda gerçekleştirilmiştir. Bu mikroorganizma setlerini açıklayıcı bilgiler aşağıda verilmiştir.

1.2.1. Mikroorganizma Seti-1

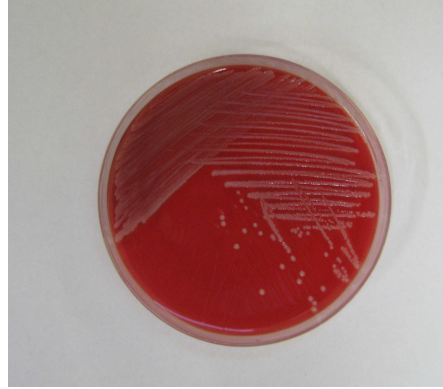
Çeşitli enfeksiyon hastalıklarına sebep olan ve rutinde sıklıkla karşılaşılan mikroorganizmaların sınıflandırılmasında (tanımlanmasında), elektronik burun cihazının kullanımını araştırmak için Mikroorganizma Seti-1 üzerinde çalışma yapılmıştır. Bu mikroorganizma seti; 8 mikroorganizmadan (7 bakteri ve 1 maya) meydana gelmektedir. Bu mikroorganizmalardan 4’ü ATCC (*American Type Culture Collection*) koleksiyonuna ait standart suşlar olup, geri kalanı hastalardan izole edilerek klasik mikrobiyolojik yöntemlerle tanısı konulmuş suşlardır. Buradaki suş ifadesi; bir mikroorganizma türünün, aralarında genetik farklılıklar bulunan alt gruplarını

tanımlamaktadır. Bir mikroorganizma türüne ait değişik suşlar arasında, pek çok yönden (ilaçlara ve dış etkilere karşı dayanıklılık gibi), az ama tanımlanabilir farklılıklar bulunabilir [57]. Tablo 1.1’de bu mikroorganizma setini oluşturan mikroorganizmaların isimleri ve özellikleri görülmektedir.

Tablo 1.1. Mikroorganizma Seti-1’i oluşturan mikroorganizma türleri ve özellikleri.

Mikroorganizma	Türü	Üreme Ortamı	Kültür Besiyeri	Üreme Koşulları	Standart Suş No
<i>Candida albicans</i>	Maya	Aerob	Kanlı agar	37 °C 24 saat	ATCC 90028
<i>Staphylococcus aureus</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	ATCC 25923
<i>Escherichia coli</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	ATCC 25922
<i>Pseudomonas aeruginosa</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	ATCC 27853
<i>Klebsiella pneumoniae</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	-
<i>Acinetobacter baumannii</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	-
<i>Salmonella typhi</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	-
<i>Shigella spp.</i>	Bakteri	Aerob	Kanlı agar	37 °C 24 saat	-

Bu mikroorganizmaların tamamı aerob (oksijenli ortamda üreyen/yaşayan) türdendir. Tüm mikroorganizmalar; petri kaplarında kanlı agar besiyerine ekim yapılarak, inkübatörde 37 °C sıcaklıkta 24 saat bekletilerek üretilmişlerdir. *Staphylococcus aureus* bakterisinin kanlı agar besiyerinde üremiş kolonileri Şekil 1.7’de görülmektedir. Besiyerinin, mikroorganizmaların koku verilerinde farklılık oluşturmaması amacıyla tüm mikroorganizmalar için aynı besiyeri kullanılmıştır. Koku verileri petri kaplarında hazırlanan örneklerden alınmıştır.



Şekil 1.7. Kanlı agar besiyerinde üremiş *Staphylococcus aureus* bakteri kültürü.

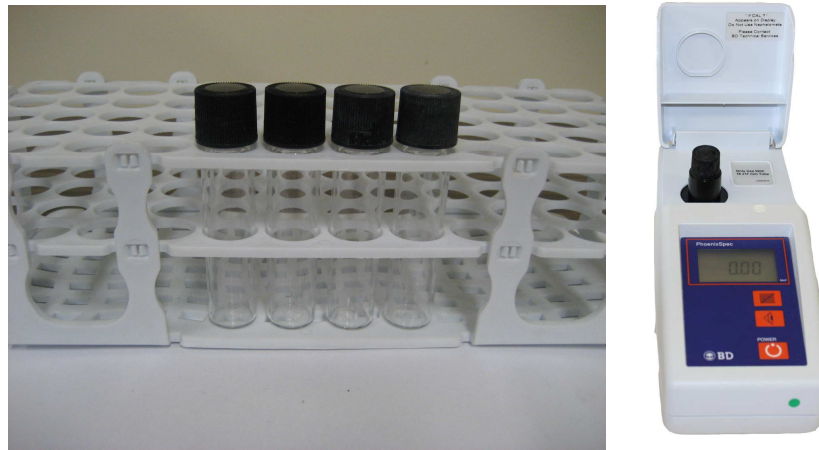
1.2.2. Mikroorganizma Seti-2

Diş kök kanal sisteminden kaynaklanan enfeksiyonların tedavi edilebilmesi için enfeksiyona sebep olan patojen mikroorganizmaların tespit edilmesi gerekmektedir. Ancak bu patojen mikroorganizmalar, genellikle anaerob (oksijensiz ortamda üreyen/yaşayan) oldukları için üretilmeleri ve tanımlanmaları zordur. Bu nedenle diş hekimleri, patojen mikroorganizmayı tespit ederek uygun tedaviyi seçmek yerine tecrübeye dayalı (ampirik) bir tedavi yöntemi uygulamayı tercih etmektedir. Diş hekimliğinin bu alanında, elektronik burun cihazının kullanımını araştırmak için Mikroorganizma Seti – 2 üzerinde çalışma yapılmıştır. Bu mikroorganizma seti; diş kökü kanallarında enfeksiyonlara sebep olan 7 mikroorganizma (5 bakteri ve 2 maya) türünden meydana gelmektedir. Bu mikroorganizmalar; Refik Saydam Kültür Koleksiyonu (RSKK), ATCC ve DSMZ (*the German Collection of Microorganisms and Cell Cultures*) kültür koleksiyonlarından temin edilmiş suşlardır. Çalışılan mikroorganizma türleri, üreme ortamları, ekim yapılan besiyerleri ve koleksiyon suş numaraları Tablo 1.2’de verilmiştir. Bu mikroorganizmaların; ikisi aerob (oksijenli ortamda üreyen/yaşayan), üçü anaerob ve geriye kalan ikisi de fakültatif (hem aerob hem de anaerob koşullarda üreyen/yaşayan) özelliğe sahip türdendir. Tüm mikroorganizmalar, spesifik besiyerlerine ekim yapılarak, inkübatörde 37 °C de üremeye bırakılmıştır. Üreme işleminden sonra, elektronik burun cihazının hangi konsantrasyona daha duyarlı olduğunu tespit etmek için, her bir mikroorganizmanın, 4’er mL serum fizyolojik (SF) içeren ağzı kapaklı steril cam tüplerde, 12×10^8 , 12×10^5 , 12×10^3 ve 12×10^1 CFU/mL (CFU: *Colony Forming Unit*) konsantrasyonlarındaki süspansiyonları hazırlanmıştır. Bu farklı konsantrasyondaki süspansiyonların

hazırlanmasında PhoenixSpec türbidimetre cihazı kullanılmıştır [44]. Şekil 1.8’de kullanılan cam tüpler ve türbidimetre cihazı görülmektedir.

Tablo 1.2. Mikroorganizma Seti-2’yi oluşturan mikroorganizma türleri ve özellikleri.

Mikroorganizma	Türü	Üreme Ortamı	Kültür Besiyeri	Üreme Koşulları	Standart Suş No
<i>Candida albicans</i>	Maya	Aerob	Sabouraud dextrose agar	37 °C 24-48 saat	ATCC 90028
<i>Candida glabrata</i> (<i>Torulopsis glabrata</i>)	Maya	Aerob	Sabouraud dextrose agar	37 °C 24-48 saat	RSKK 04019
<i>Fusobacterium nucleatum</i>	Bakteri	Anaerob	Anaerobik kanlı agar	37 °C 4-6 gün	DSMZ 20482
<i>Porphyromonas gingivalis</i>	Bakteri	Anaerob	Anaerobik kanlı agar	37 °C 4-6 gün	ATCC 33277
<i>Pseudoramibacter alactolyticus</i>	Bakteri	Anaerob	Anaerobik kanlı agar	37 °C 4-6 gün	DSMZ 3980
<i>Streptococcus sanguinis</i>	Bakteri	Fakültatif Aerob	Tryptic soy agar	37 °C 4 gün %5 CO ₂	DSMZ 20567
<i>Enterococcus faecalis</i>	Bakteri	Fakültatif Aerob	Kanlı agar	37 °C 24 saat	ATCC 29212

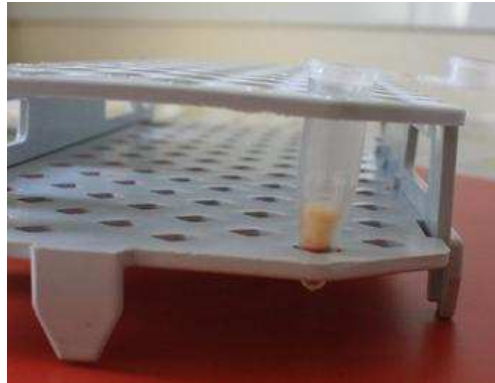


Şekil 1.8. Kullanılan steril cam tüpler ve türbidimetre cihazı.

1.2.3. Mikroorganizma Seti-3

Bu mikroorganizma seti; ikinci setteki mikroorganizmalarla aynı olup, sabit konsantrasyondaki daha fazla örnek üzerinde çalışma yapmak için oluşturulmuştur. Bunun sebebi; elektronik burun cihazı ile elde edilen mikroorganizma koku verilerinin, sınıflandırma başarısının güvenilirliğini artırmaktır.

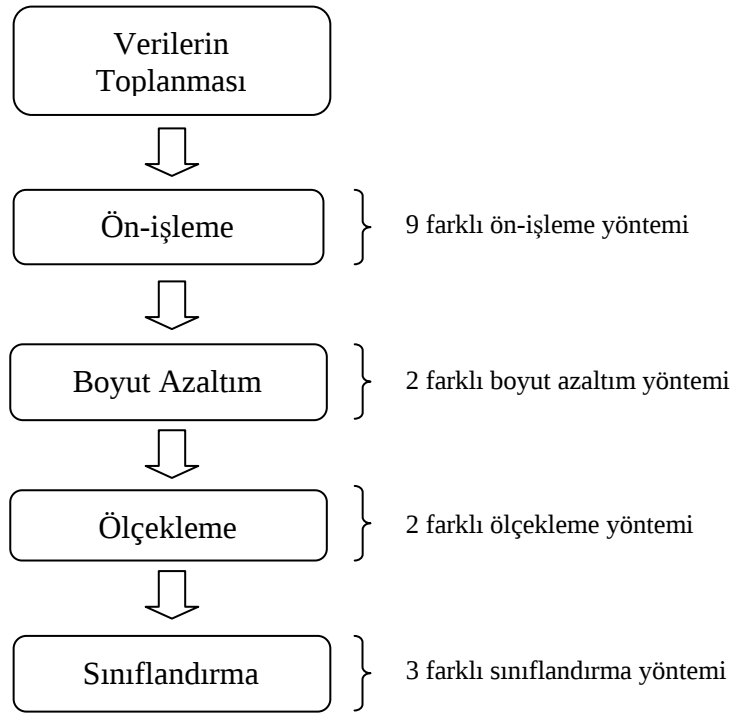
Buradaki mikroorganizmalar spesifik besiyerlerine ekim yapılarak üretildikten sonra, her bir mikroorganizmanın, SF içeren ağız kapaklı steril cam tüplerde, 1.5×10^8 CFU/mL konsantrasyonlarındaki süspansiyonları hazırlanmıştır. Gerçeğe yakın bir durum ortaya koymak için, ependorf tüplerde steril diş içeren örnekler oluşturulmuştur. Bu örneklerden biri Şekil 1.9’da görülmektedir. Daha sonra hazırlanan mikroorganizma süspansiyonları, her örnekteki dişlerin dentin tübüllerine yayılması için, her seferinde 2’şer μ L olacak şekilde toplam 20 μ L dağıtılmıştır.



Şekil 1.9. Ependorf tüpte steril diş içeren bir örnek.

1.3. Sınıflandırma Sürecinin Aşamaları

Bu tez çalışmasında, 3 ayrı mikroorganizma setinin her birinde, elektronik burun cihazı ile elde edilen mikroorganizma koku verilerinin sınıflandırılmasına ilişkin uygulamalar geliştirilmiştir. Mikroorganizma koku verilerine çeşitli ön-işleme, boyut azaltım, ölçekleme ve sınıflandırma yöntemleri uygulanarak bu yöntemlerin sınıflandırma performansına etkileri araştırılmıştır. Verilere uygulanan bu yöntemlerin sıralamasını ifade eden yapı, Şekil 1.10’da görülmektedir.



Şekil 1.10. Mikroorganizma koku verilerinin sınıflandırma sürecini oluşturan aşamalar ve uygulanan yöntem sayıları.

1.4. Mikroorganizma Koku Verilerinin Toplanması

Mikroorganizma koku verileri, Farmasötik Mikrobiyoloji Laboratuvarı'nda, mikroorganizma setlerindeki örneklerden Cyranose 320 elektronik burun cihazı ile toplanmıştır. Mikroorganizma setlerinden elde edilen koku verileri ile ilgili bilgiler aşağıda verilmiş olup, ayrıntıları Tablo 1.3'de görülmektedir.

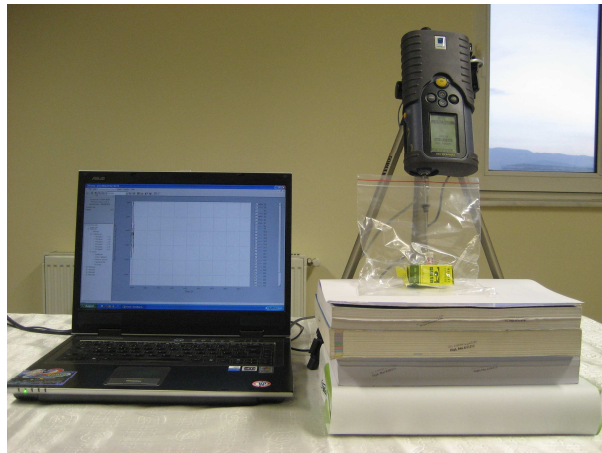
- Mikroorganizma seti-1'deki 8 mikroorganizmanın her biri, 8 ayrı petriye ekim yapılmıştır. Dolayısıyla toplam 64 petri üzerinde çalışılmış ve her birinden birer koku verisi alınmıştır.
- Mikroorganizma seti-2'deki 7 mikroorganizmanın her birinin 4 farklı konsantrasyonda steril cam tüplerde süspansiyonları hazırlanmıştır. Bu cam tüplerin her birinden 5 koku verisi alınmıştır. Dolayısıyla her bir konsantrasyon için toplamda 35 koku verisi üzerinde çalışılmıştır.
- Mikroorganizma seti-3'deki 7 mikroorganizmanın her birinin sabit konsantrasyonda steril cam tüplerde süspansiyonları hazırlanmıştır. Daha sonra her mikroorganizma süspansiyonu, endorf tüplerde steril dış içeren 12 örneğe

eklenmiştir. Dolayısıyla toplamda 84 örnek üzerinde çalışılmış ve her örnekten birer koku verisi alınmıştır.

Tablo 1.3. Çalışılan mikroorganizma setlerinin, içerdikleri tür (sınıf), örnek (numune) ve koku verileri sayıları.

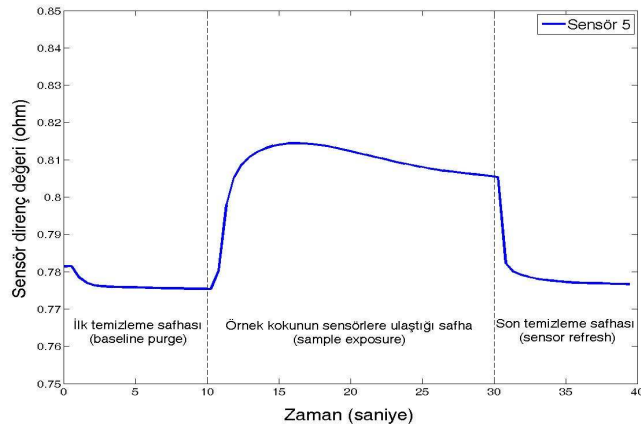
Mikroorganizma seti	Tür sayısı	Her türden alınan örnek sayısı	Her örnekten alınan koku verisi sayısı	Elde edilen toplam koku verisi
Mikroorganizma seti-1	8	8	1	64
Mikroorganizma seti-2 (Her konsantrasyon için)	7	1	5	35
Mikroorganizma seti-3	7	12	1	84

Elektronik burun cihazı ile koku verileri alınmaya başlanmadan önce, her örneğin (petri kabı/cam tüp/ependorf tüp) kapağı açılarak 3 dakika ağzı kilitli bir naylon torbada bekletilmiştir. Bu sayede numunenin kokusu torba içine dağılmış ve dış etkilerden izole edilmiştir. Daha sonra elektronik burun cihazının probu torba içerisine geçirilerek veri alınmaya başlanmıştır. Veriler alınırken elektronik burun cihazının tüm sensörleri (32 sensör) aktif halde tutulmuştur. Mikroorganizma koku verilerinin Cyranose 320 cihazı ile toplandığı sistem Şekil 1.11’de görülmektedir.



Şekil 1.11. Mikroorganizma koku verilerinin toplandığı sistem.

Bir koku verisinin alımı, toplam 40 saniye sürecektir ve 3 safhada gerçekleşecek şekilde cihazın yazılımı aracılığıyla ayarlanmıştır. Bu ayarlamaya göre elektronik burun cihazı; ilk 10 saniye sensörlerini ortam havası ile temizlemekte (*baseline purge*), sonraki 20 saniye boyunca numunenin kokusunu almakta (*sample exposure*) ve son 10 saniyede de tekrar sensörlerini ortam havası ile temizlemektedir (*sensor refresh*). Sensörlerin temizlenmesinde ortam havasının tercih edilmesinin sebebi; literatürdeki çeşitli çalışmalarda kullanılmış olmasındandır [4, 9, 10]. Cihaz, bir koku verisinin alımı süresince, toplam 74 örnekleme (sensör okuması) yaparak tüm sensörlerin anlık direnç değerlerini içeren dosyayı bilgisayara transfer etmektedir. Şekil 1.12’de bir koku verisindeki tek bir sensörün direnç değişim grafiği ve koku alma safhaları görülmektedir.



Şekil 1.12. Bir koku verisindeki tek bir sensörün direnç değişim grafiği ve koku alma safhaları.

Her bir koku verisi, elektronik burun cihazındaki tüm sensörlerin direnç değerlerini içeren 74 okumadan oluşmaktadır. Dolayısıyla her bir koku verisi 74×32 'lik bir matris formundadır (R_i) ve bu matristeki her bir sütun bir sensöre karşılık gelmekte olup, koku alma süresince o sensörün anlık direnç değerlerini ($r_{i,k,j}$) göstermektedir. Buradaki i koku veri indeksini, $k = 1, 2, \dots, 74$ örnekleme (sensör okuma) indeksini ve $j = 1, 2, \dots, 32$ sensör indeksini temsil etmektedir. Eşitlik 1.1’de bu matrisin yapısı görülmektedir.

$$R_i = \begin{bmatrix} r_{i,1,1} & r_{i,1,2} & \dots & r_{i,1,j} & \dots & r_{i,1,32} \\ r_{i,2,1} & r_{i,2,2} & \dots & \dots & \dots & r_{i,2,32} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & r_{i,k,j} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{i,74,1} & r_{i,74,2} & \dots & r_{i,74,j} & \dots & r_{i,74,32} \end{bmatrix} \quad (1.1)$$

1.5. Ön-işleme Yöntemleri

Gerçekleştirilen bu çalışmada, mikroorganizma koku verilerine uygulanan farklı ön-işleme yöntemlerinin sınıflandırma performansına etkileri değerlendirilmiştir. Ön-işleme yöntemleri olarak; her bir koku verisine (veri matrisine) 3 ayrı sensör cevap modeli uygulanmış ve her sensör cevap modeli için de 3 ayrı sensör referans direnç değer noktası göz önüne alınmıştır. Dolayısıyla toplamda 9 farklı ön-işleme yöntemi uygulanmıştır.

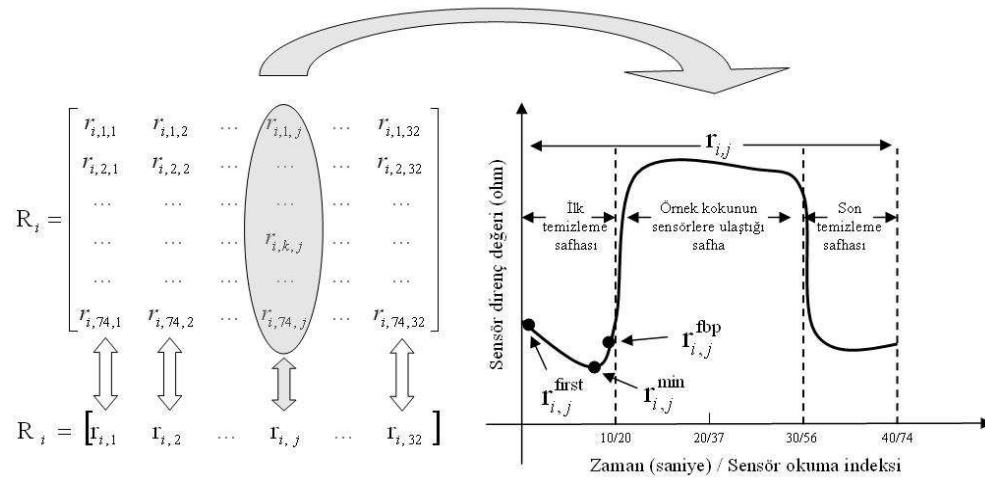
1.5.1. Sensör Cevap Modelleri ve Sensör Referans Direnç Değer Noktaları

Sensör cevap modelleri, koku verilerindeki zaman (örnekleme) boyutunu yok etmek ve koku örnekleri arasındaki sıcaklık farklarının verilere etkisinin azaltılması için kullanılmıştır. Uygulanan sensör cevap modellerinin eşitlikleri ve açıklamaları Tablo 1.4’de verilmiştir. Buradaki i koku veri indeksini, R_i ise 32 sensörün anlık direnç değerlerini içeren 74 okumayı kapsayan bir koku veri matrisini ifade etmektedir.

Tablo 1.4. Farklı sensör cevap modellerinin eşitlik ifadeleri ve açıklamaları.

Sensör cevap modeli	Eşitlik ifadesi	Açıklama
Fark modeli	$x_i = R_i^{\max} - R_i^o = \Delta R_i$	Sensörlerin maksimum ve referans direnç değerleri arasındaki fark
Oransal fark modeli	$x_i = \Delta R_i / R_i^o$	Sensörlerin direnç değişimlerinin referans değere oranı
Normalize oransal fark modeli	$x_i = (\Delta R_i / R_i^o) / (\Delta R_i / R_i^o)^{\max}$	Oransal fark modelinin maksimum değişim gösteren sensöre göre normalizasyonu

Cyranose 320 elektronik burun cihazının kullanım kılavuzunda sensör referans değer noktası ile ilgili açık bir bilgi olmamasından dolayı, farklı referans noktalarının da sınıflandırma performansına etkileri incelenmek istenmiştir. Bir koku veri matrisi ile bir sensöre ait direnç değişim grafiğindeki sensör referans değer noktaları, Şekil 1.13’de birlikte görülmektedir. Uygulanan sensör referans direnç değer noktalarının eşitlikleri ve açıklamaları Tablo 1.5’de verilmiştir.



Şekil 1.13. Bir koku veri matrisi ile bir sensöre ait direnç değişim grafiğindeki sensör referans direnç değer noktalarının birlikte gösterimi.

Tablo 1.5. Farklı sensör referans direnç değer noktalarının eşitlik ifadeleri ve açıklamaları.

Sensör referans direnç değeri noktası	Eşitlik ifadesi	Açıklama
Sensör referans değeri-1	$R_i^o = R_i^{\min}$	Bir koku verisindeki sensörlerin, minimum değerleri referans kabul edilerek
Sensör referans değeri-2	$R_i^o = R_i^{\text{first}}$	Bir koku verisindeki sensörlerin, ilk sensör okuma değerleri referans kabul edilerek
Sensör referans değeri-3	$R_i^o = R_i^{\text{fbp}}$	Bir koku verisindeki sensörlerin, cihazın kokuyu içine almaya başlamadan önceki son sensör okuma değerleri referans kabul edilerek

Böylece Cyranose 320 elektronik burun cihazı ile elde edilen mikroorganizma koku verilerine, toplamda 9 farklı ön-işleme yöntemi uygulanmıştır. Bu yöntemlerin tanımları, eşitlik ifadeleri ve kısaltmaları Tablo 1.6’da görülmektedir.

Tablo 1.6. Sensör referans direnç değer noktaları ve sensör cevap modellerine göre, 9 farklı ön-işleme yönteminin eşitlik ifadeleri ve kısaltmaları.

Ön-işleme yöntemleri		Eşitlik ifadesi	Kısaltma
Sensör referans direnç değer noktası	Sensör cevap modeli		
Sensör referans değeri-1 $R_i^o = R_i^{\min}$	Fark modeli	$x_i = R_i^{\max} - R_i^{\min} = \Delta R_i$	$d1$
	Oransal fark modeli	$x_i = \Delta R_i / R_i^{\min}$	$fd1$
	Normalize oransal fark modeli	$x_i = (\Delta R_i / R_i^{\min}) / (\Delta R_i / R_i^{\min})^{\max}$	$nfd1$
Sensör referans değeri-2 $R_i^o = R_i^{\text{first}}$	Fark modeli	$x_i = R_i^{\max} - R_i^{\text{first}} = \Delta R_i$	$d2$
	Oransal fark modeli	$x_i = \Delta R_i / R_i^{\text{first}}$	$fd2$
	Normalize oransal fark modeli	$x_i = (\Delta R_i / R_i^{\text{first}}) / (\Delta R_i / R_i^{\text{first}})^{\max}$	$nfd2$
Sensör referans değeri-3 $R_i^o = R_i^{\text{fbp}}$	Fark modeli	$x_i = R_i^{\max} - R_i^{\text{fbp}} = \Delta R_i$	$d3$
	Oransal fark modeli	$x_i = \Delta R_i / R_i^{\text{fbp}}$	$fd3$
	Normalize oransal fark modeli	$x_i = (\Delta R_i / R_i^{\text{fbp}}) / (\Delta R_i / R_i^{\text{fbp}})^{\max}$	$nfd3$

1.6. Boyut Azaltım Yöntemleri

Bir veri setindeki öznitelik sayısı, aynı zamanda bu veri setinin boyut sayısını ifade etmektedir. Çok fazla öznitelige sahip veri setlerinde, sınıflandırma hatası artabilmekte ve dolayısıyla sınıflandırma performansı düşebilmektedir. Meydana gelen bu olaya boyutsallık belası (*curse of dimensionality*) adı verilmektedir [45].

Genellikle iyi bir sınıflandırma performansı elde etmek için, veri setindeki örnek sayısının (n) öznitelik sayısından (p) daha fazla olması istenir. Jain ve arkadaşları bu konuda şu uyarılarda bulunmuşlardır [46]: “Yanlış sınıflandırma olasılığı, eğitim veri setindeki örnek sayısı ve öznitelik sayısı ile sınıf koşullu yoğunlukların doğru parametreleri arasındaki gerçek ilişkinin belirlenmesi çok güçtür. Bununla birlikte, veri

setinin örnek ve öznitelik sayıları arasındaki oranı tanımlayan bazı ön kabuller yapılmıştır. Sınıf başına düşen eğitim verisi/örneği sayısının, öznitelik sayısının 10 katından daha fazla olması genel kabul görmektedir. Bu durum Eşitlik 1.2’de ifade edilmiştir.

$$\frac{n}{K \times p} > 10 \quad (1.2)$$

Burada n ; veri setindeki örnek sayısını, K ; sınıf sayısını ve p ; öznitelik sayısını temsil etmektedir. Eşitlik 1.2, sınıflandırıcı tasarımında izlenmesi gereken iyi bir yöntemi göstermektedir. Bir sınıflandırıcı ne kadar kompleks/karmaşık olursa, boyutsallık belasına yakalanmamak için veri setinin örnek ve öznitelik sayıları arasındaki oran o kadar fazla olmalıdır.” Bu oran ile ilgili Eşitlik 2.2’de belirtilen koşul bazı durumlarda çok kısıtlayıcı olacağından, Jain ve arkadaşları tarafından, ilgili makalelerindeki [46] öznitelik seçimi kısmında, aşağıdaki eşitlik yeterli/tatmin edici bulunmaktadır [45].

$$\frac{n}{K \times p} > 5 \quad (1.3)$$

Koku verileri elektronik burun cihazı ile toplanırken, cihazın 32 sensörünün de aktif halde tutulduğu daha önce belirtilmişti. Bu sensörlerin direnç değişimleri kaydedildikten sonra, koku verilerine çeşitli ön-işleme yöntemleri uygulanarak 32 boyutlu veri setleri elde edilmiştir. Boyut sayısının veri setlerindeki örnek sayısına nazaran fazla olduğu görülmüştür. Bu nedenle; boyutsallık belasına yakalanmamak ve etkili bir sınıflandırma gerçekleştirebilmek için 2 farklı boyut azaltım yöntemi üzerinde durulmuştur. Bunlar;

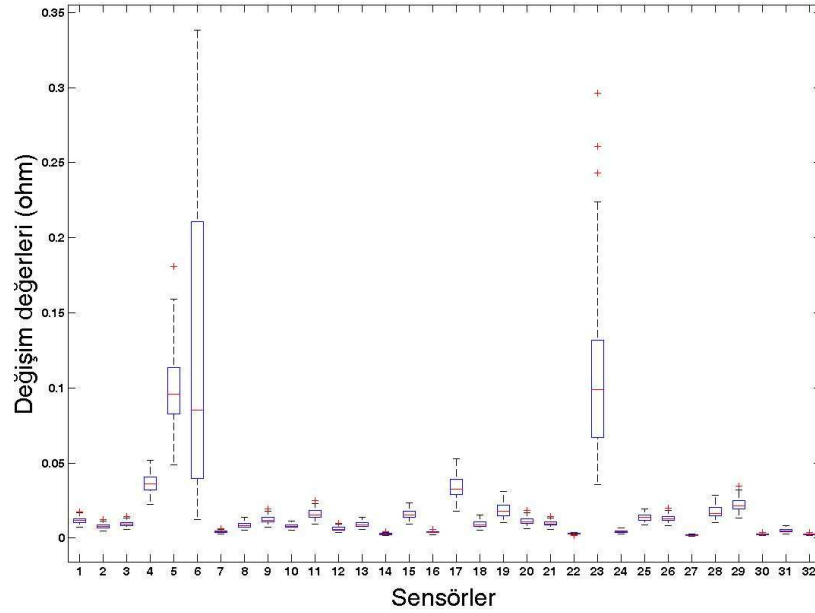
- Standart sapmaya göre
- Temel bileşenler analizine (Principal Component Analysis - PCA) göre

boyut azaltımıdır.

1.6.1. Standart Sapmaya Göre Boyut Azaltımı

Bu yöntemde her koku veri setindeki 32 sensörün her birinin standart sapmaya göre değişimleri incelenmiştir. Bu sensörlerden en fazla değişime sahip olanların, veri seti hakkında en fazla bilgiyi taşıdığı ve koku verilerinin sınıflandırılmasında en ayırt edici

özellikler olduğu düşünülmüştür. Şekil 1.14’de; mikroorganizma seti-1 için, $d1$ (sensör referans değeri-1 ile fark modeli) ön-işleme metodu uygulanarak elde edilmiş bir koku veri setindeki, farklı mikroorganizma numunelerine ait sensör değerlerinin değişimleri görülmektedir.



Şekil 1.14. Mikroorganizma seti-1 için, $d1$ ön-işleme metodu uygulanarak elde edilmiş koku veri setindeki, farklı mikroorganizma örneklerine ait sensör direnç değerlerinin değişimleri.

Üzerinde çalışılan mikroorganizma setlerinde, standart sapmaya göre boyut azaltımı yapılırken en fazla değişim gösteren 3 sensör göz önüne alınmıştır. Bunun nedeni PCA’da 3 öz niteliğin seçilmesi dolayısıyla boyut azaltım yöntemlerinin kendi aralarında kıyaslanmasında öz nitelik sayısı açısından fark olmaması içindir. PCA’da neden 3 öz nitelik seçildiği bir sonraki kısımda açıklanmıştır.

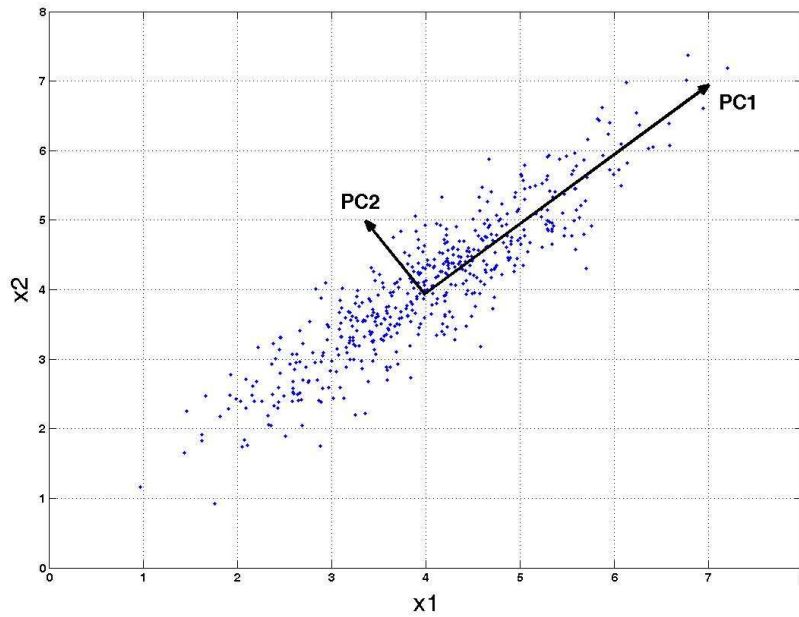
1.6.2. Temel Bileşenler Analizine Göre Boyut Azaltımı

PCA, çok değişkenli veri analiz ve boyut azaltım tekniklerinin, örüntü tanımadaki en eski ve en yaygın kullanılan yöntemlerinden birisidir. PCA’daki temel fikir; veri seti hakkındaki bilginin, verilerin öz nitelik değerlerindeki değişimlerde saklandığı düşüncesidir. Dolayısıyla bir öz nitelik ne kadar fazla değişime sahip ise veri seti hakkında o kadar fazla bilgi taşımaktadır [47]. Böylece PCA, p öz niteliğe (değişkene)

sahip bir veri setinde, bu özneliliklerin belirlediği toplam değişkenliği ifade etmek üzere d sayıda ($d < p$) ana bileşen bulmak için gerçekleştirilir. Böylece daha az sayıda değişken ile p boyutlu uzay yerine d boyutlu farklı bir uzayda çalışma yapılabilir. Bu sayede boyut indirgeme işlemi gerçekleştirilmiş olur.

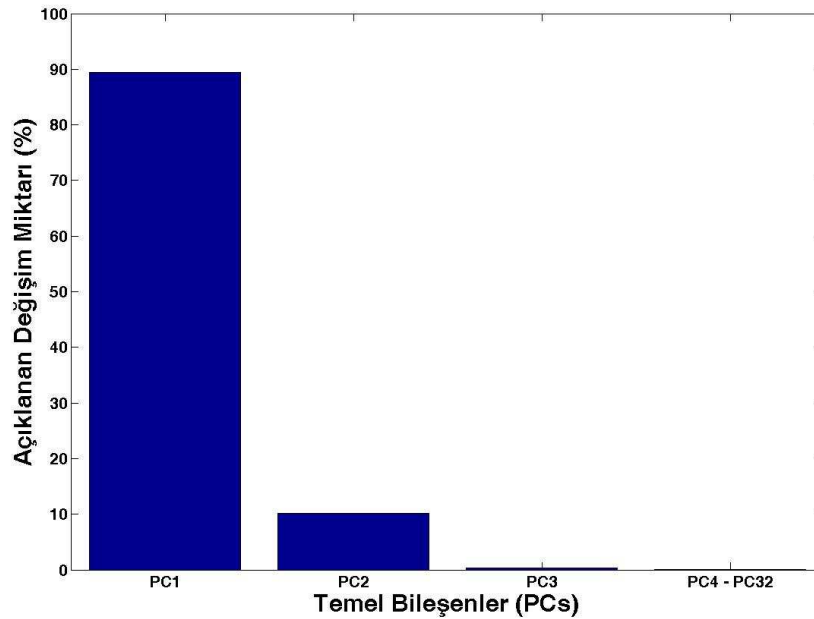
PCA'yı aralarında yüksek korelasyon bulunan p sayıda değişkeni, bu değişkenlerin doğrusal bileşenleri olan ve aralarında korelasyon bulunmayan d sayıda yeni değişkenlerle ifade etmek olarak da tanımlamak mümkündür.

Bir veri setinin en fazla değişim gösterdiği doğrultu temel bileşen-1 (*principal component-1* – PC1) eksenini olarak adlandırılır. Veri setinin boyutuna bağlı olarak en fazla değişim gösteren doğrultudan en aza doğru, temel bileşen eksenleri belirlenir. Şekil 1.15'de iki boyutlu bir veri setine PCA uygulandıktan sonra elde edilen temel bileşenlerin, veri setinin en fazla değişim gösterdiği doğrultular olduğu görülmektedir. Burada PC2 eksenini doğrultusundaki değişimin çok küçük olduğu düşünülerek, veri sadece PC1 eksenini üzerinde gösterilebilir ve bu sayede iki boyuttan tek boyuta boyut azaltımı gerçekleştirilmiş olur.



Şekil 1.15. İki boyutlu bir veri setine PCA uygulandıktan sonra elde edilen temel bileşenler (PC1 ve PC2).

Ön-işlemden geçirilmiş koku verilerinden oluşan mikroorganizma veri setlerine, PCA uygulanırken ilk 3 temel bileşen (PC1 – PC2 – PC3) göz önüne alınmıştır. Bunun nedeni; ilk 3 temel bileşenin her bir veri setindeki toplam değişkenliğin %99’undan fazlasını temsil etmesinden dolayıdır. Şekil 1.16’da; mikroorganizma seti-1 için, $d1$ (sensör referans değeri-1 ile fark modeli) ön-işleme metodu uygulanarak oluşturulan koku veri setinin, PCA uygulandıktan sonra elde edilen temel bileşenlerinin, toplam değişkenliği temsil etme yüzdeleri görülmektedir.



Şekil 1.16. Mikroorganizma seti-1 için, $d1$ ön-işleme metodu uygulanarak oluşturulan koku veri setinin, PCA işleminden sonraki temel bileşenlerinin, toplam değişkenliği temsil etme yüzdeleri.

1.7. Ölçekleme Yöntemleri

Her mikroorganizma seti için 9 farklı ön-işleme ve 2 farklı boyut azaltım metodu uygulanarak 18 ayrı veri seti türetilmiştir. Bir veri seti; n adet veri içeren d boyutlu bir veri matrisi olan X ile ifade edilebilir. Bu matrisin sütunları, veri setinin öz nitelik vektörlerine karşılık gelmekte olup, x_i ile gösterilebilir. Eşitlik 1.4’de veri matrisi ifadesi görülmektedir.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & \dots & x_{2,d} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & x_{n,d} & \dots & \dots \\ x_{n,1} & x_{n,2} & \dots & \dots & x_{n,d} \end{bmatrix} = [x_1 \quad x_2 \quad \dots \quad x_i \quad \dots \quad x_d] \quad (1.4)$$

Veri setleri sınıflandırıcıya verilmeden önce 2 farklı ölçeklemeye tabii tutularak, ölçeklemenin sınıflandırma performansına etkisi incelenmiştir. Kullanılan ölçekleme metotları;

- [0-1] arasına ölçekleme: Bu metoda göre, sınıflandırılacak verinin öznitelikleri [0-1] arasına ölçeklenir. Eşitlik 1.5’de bu ölçekleme ifadesi görülmektedir.

$$(x_i - x_i^{\min}) / (x_i^{\max} - x_i^{\min}) \quad (1.5)$$

- z-skoru/normalizasyonu ile ölçekleme: Bu metoda göre, sınıflandırılacak verinin öznitelik vektör elemanlarının ortalama (μ_i) etrafındaki dağılımları, standart sapmaya (σ_i) göre normalize edilir. Eşitlik 1.6’da bu ölçekleme ifadesi görülmektedir.

$$z_i = (x_i - \mu_i) / \sigma_i \quad (1.6)$$

1.8. Sınıflandırma Yöntemleri

Bu tez çalışmasında, elektronik burun cihazı ile elde edilen mikroorganizma koku verilerinin sınıflandırılması için, 3 ayrı sınıflandırma yöntemi üzerinde durulmuştur. Bunlar:

- k-En Yakın Komşuluk Algoritması
- Ayırma Analizi
- Destek Vektör Makineleri

Kullanılan bu sınıflandırma yöntemlerinin çeşitli alt metotları da incelenerek, kapsamlı bir araştırma yapılmaya çalışılmıştır.

1.8.1. k -En Yakın Komşuluk Algoritması

Bu sınıflandırıcı (k -Nearest Neighbor - k -NN), en yaygın kullanılan sınıflandırma algoritmalarından birisidir. Bunun nedenleri arasında; bu algoritmanın sezgisel olması, kolay uygulanabilmesi ve zor/zahmetli problemlerde (*demanding applications*) fark edilir bir şekilde iyi performans göstermesi sayılabilir. k -NN sınıflandırıcı bir $\mathbf{x}$ test verisini/örneğini alır ve bu verinin eğitim veri setindeki en yakın k adet komşusunu bulur. Daha sonra bu $\mathbf{x}$ verisini, bu k adet komşu en fazla hangi sınıfa ait ise o sınıfa atar. k -NN algoritmasını aşağıdaki gibi özetlemek mümkündür [47]:

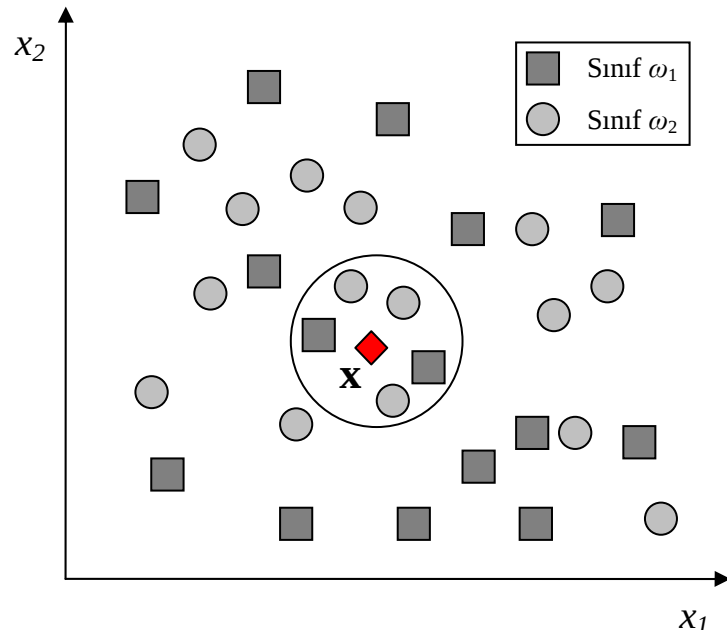
- N adet eğitim verisinden, $\mathbf{x}$ test verisine en yakın k adet komşu, belirli bir uzaklık ölçüm şekli seçilerek (örneğin Euclid uzaklığı/mesafesi gibi) belirlenir. k için tek (veya asal) olan bir sayı seçilir.
- Her ω_j sınıfına ait k_j örnek/veri sayısı belirlenir. Burada $\sum_{j=1}^c k_j = k$ ifadesi geçerlidir. Buradaki c , veri setindeki sınıf sayısını temsil etmektedir.
- Daha sonra $\mathbf{x}$ test verisi, en büyük k_j sayısına sahip sınıfa atanır. Bu işlem, test verisinin en fazla oya sahip sınıfa atanması olarak tanımlanabilir (*majority rule*).

İki sınıflı bir veri setinde, $k = 5$ için k -NN sınıflandırmasına örnek bir uygulama Şekil 1.17’de görülmektedir. k -NN algoritmasına göre $\mathbf{x}$ test verisinin sınıflandırılması; $\mathbf{x}$ ’e en yakın 5 eğitim verisinin belirlenmesi ve bu verilerin çoğu hangi sınıfa ait ise, $\mathbf{x}$ test verisinin o sınıfa atanması ile gerçekleştirilmektedir. Buradaki durumda $k_1 = 2$, $k_2 = 3$ ve $k_2 > k_1$ olduğu için, $\mathbf{x}$ test verisi k -NN sınıflandırıcı tarafından ω_2 sınıfına atanmaktadır.

Burada k sayısının tek olarak seçilmesinin tesadüfi/rasgele olmadığı belirtilmelidir. Muhtemel en büyük k_j sayılarındaki eşitliği önlemek için genellikle;

- $c = 2$ sınıflı problemler için tek,
- $c > 2$ sınıflı problemler için ise c sayısına bölünemeyen

bir k sayısı seçilir [47]. İki den fazla sınıfa sahip bir veri seti için, uygun bir k sayısı seçilmesine rağmen en fazla oya sahip eşit, iki ya da daha fazla sınıf ortaya çıkabilir. Bu durum test verisinin eşit oya sahip sınıflardan, en yakın veya koşullu olasılık yoğunluk fonksiyonu (*posterior probability*) $P(\omega_i | \mathbf{x})$ en büyük olanına atanarak aşılabılır [48].



Şekil 1.17. k -NN sınıflandırmasına örnek bir durum ($k = 5$).

k -NN için sınırlandırıcı bir durum; $k = 1$ olarak seçilmesidir. Bu durumda sınıflandırma algoritması x test verisini/örneğini, eğitim veri setindeki en yakın komşusunun sınıfına atar. k sayısının büyük seçilmesi ise, düzgün bir karar sınırının oluşmasına sebep olur. Bununla birlikte bu durum sınıflandırma işleminde; hesaplama karmaşıklığını (*computational complexity*) artırmakta, daha da önemlisi, yerel bilgi kaybı oluşturmaktadır [47].

k -NN sınıflandırıcıda, aslında herhangi bir öğrenme gerçekleşmemektedir. Bilinmeyen bir verinin sınıflandırılması istenene kadar sınıflandırıcıda bir işlem yapılmamaktadır. Bilinmeyen bir verinin sınıflandırılması, bu verinin eğitim veri setindeki verilerle karşılaştırılarak sağlanmaktadır. Bu nedenle bu yaklaşım, bilinmeyen bir verinin sınıflandırılması istenmeden önce karar sınırının belirlendiği istekli öğrenme algoritmalarının (*eager learning algorithms*) aksine göre işlev gördüğünden, tembel öğrenme (*lazy learning*) veya hafıza tabanlı öğrenme (*memory based learning*) olarak adlandırılmaktadır [47].

1.8.2. Ayırma Analizi

Ayırma Analizi (*Discriminant Analysis - DA*), danışmanlı sınıflandırma (*supervised classification*) yöntemlerinden olup, Bayes olasılık kuralı temelli bir sınıflandırıcıdır (*Bayesian classification*). Çok sınıflı bir veri setinde;

- $\mathbf{x}$, d -boyutlu öznitelik uzayındaki (*feature space*) bir gözlemi
- $i = 1, 2, \dots, c$ sınıf indeksini
- R_i , i . sınıfa ait bölgeyi
- ω_i , i . sınıfa ait verileri
- $p(\mathbf{x} | \omega_i)$, ω_i sınıfı için $\mathbf{x}$ 'in koşullu olasılık yoğunluk fonksiyonunu (*likelihood*)
- $P(\omega_i)$, ω_i sınıfı için önsel olasılık yoğunluk fonksiyonunu (*prior probability*)
- $p(\mathbf{x})$, $\mathbf{x}$ için önsel olasılık yoğunluk fonksiyonunu (*evidence factor*)
- $P(\omega_i | \mathbf{x})$, $\mathbf{x}$ için ω_i sınıfının koşullu olasılık yoğunluk fonksiyonunu (*posterior probability*)

temsil etmektedir.

Önsel olasılık yoğunluk fonksiyonu $P(\omega_i)$ ve koşullu olasılık yoğunluk fonksiyonu $p(\mathbf{x} | \omega_i)$ bilindiği takdirde, ω_i sınıfının koşullu olasılık yoğunluk fonksiyonu $P(\omega_i | \mathbf{x})$, Eşitlik 1.7'deki Bayes formülü ile hesaplanır.

$$P(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)P(\omega_i)}{p(\mathbf{x})} \quad (1.7)$$

$$p(\mathbf{x}) = \sum_{i=1}^c p(\mathbf{x} | \omega_i)P(\omega_i) \quad (1.8)$$

$$P(\omega_1 | \mathbf{x}) + P(\omega_2 | \mathbf{x}) + \dots + P(\omega_c | \mathbf{x}) = 1 \quad (1.9)$$

Bayes kuralına göre, $\mathbf{x}$ en yüksek $P(\omega_i | \mathbf{x})$ olasılık değerini sağlayan veri sınıfı ω_i 'ye aittir. Bu kural Eşitlik 1.10'da verilmiştir.

$$\text{Her } j \neq i \text{ için, } P(\omega_i | \mathbf{x}) > P(\omega_j | \mathbf{x}) \rightarrow \mathbf{x} \text{ verisi } \omega_i \text{ 'ye aittir.} \quad (1.10)$$

Eşitlik 1.10'daki koşullu olasılık yoğunluk fonksiyonlarının sınıflara göre karşılaştırılmasında, eşitlik 1.7'den yararlanılabilir. Bu durumda $p(\mathbf{x})$ ifadesi sınıflardan bağımsız olduğu için ihmal edilir. Dolayısıyla eşitlik 1.10'daki ifade, eşitlik 1.11'deki gibi yazılabilir.

$$\text{Her } j \neq i \text{ için, } p(\mathbf{x} | \omega_i)P(\omega_i) > p(\mathbf{x} | \omega_j)P(\omega_j) \rightarrow \mathbf{x} \text{ verisi } \omega_i \text{ 'ye aittir.} \quad (1.11)$$

Minimum sınıflandırma hatası elde etmek için, $P(\omega_i | \mathbf{x})$ koşullu olasılık yoğunluk fonksiyonunun doğal logaritması alınır. Eşitlik 1.12, $P(\omega_i | \mathbf{x})$ ifadesini maksimize ederek maksimum ayırma fonksiyonunu tanımlar.

$$g_i(\mathbf{x}) = \ln p(\mathbf{x} | \omega_i) + \ln P(\omega_i) \quad (1.12)$$

Eğer koşullu olasılık yoğunluk fonksiyonu $p(\mathbf{x} | \omega_i)$, çok değişkenli normal dağılıma sahip ise,

$$p(\mathbf{x} | \omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (1.13)$$

$$p(\mathbf{x} | \omega_i) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)}{2}\right) \quad (1.14)$$

olur. Burada $\boldsymbol{\mu}_i$, $d \times 1$ 'lik ortalama vektörünü, $\boldsymbol{\Sigma}_i$, $d \times d$ 'lik kovaryans matrisini ve $|\boldsymbol{\Sigma}_i|$ kovaryans matrisinin determinantını temsil etmektedir. Eşitlik 1.14, eşitlik 1.12'de yerine konulduğunda ayırma fonksiyonu

$$g_i(\mathbf{x}) = -\frac{1}{2} [d \ln(2\pi) + \ln |\boldsymbol{\Sigma}_i| + (\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)] + \ln P(\omega_i) \quad (1.15)$$

olarak elde edilir. Bu ifadedeki ilk terim veri sınıflarından bağımsız olduğu için ihmal edilebilir. Ayrıca önsel olasılık yoğunluk fonksiyonu $P(\omega_i)$ de, her sınıf için eşit olduğu düşünülerek ihmal edilebilir. Dolayısıyla eşitlik 1.15,

$$g_i(\mathbf{x}) = -\frac{1}{2} [\ln |\boldsymbol{\Sigma}_i| + (\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)] \quad (1.16)$$

halini alır.

Farklı ayırma fonksiyonları ile elde edilen farklı sınıflandırıcılar mevcuttur. Bunlar;

- Doğrusal (*Linear*)
- Diyagonal doğrusal (*Diagonal linear*)
- İkinci dereceden (*Quadratic*)
- Diyagonal ikinci dereceden (*Diagonal quadratic*)
- Mahalanobis

sınıflandırıcılardır. Diyagonal sınıflandırıcılara, Naive Bayes sınıflandırıcılar da denmektedir [45].

1.8.2.1. Doğrusal Sınıflandırıcı

Her veri sınıfının kovaryans matrisinin birbirine eşit olduğu varsayımı ($\Sigma_i = \Sigma$), sınıflandırıcıyı doğrusal hale getirir. Dolayısıyla eşitlik 1.16'daki ilk terim, her veri sınıfı için aynı olacağından ihmal edilir.

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) \quad (1.17)$$

Bu ifade genişletilerek eşitlik 1.20'deki forma dönüştürülebilir. Eşitlik 1.20'deki ilk terim, veri sınıflarından bağımsız olduğu için dikkate alınmaz ve eşitlik 1.21 elde edilir.

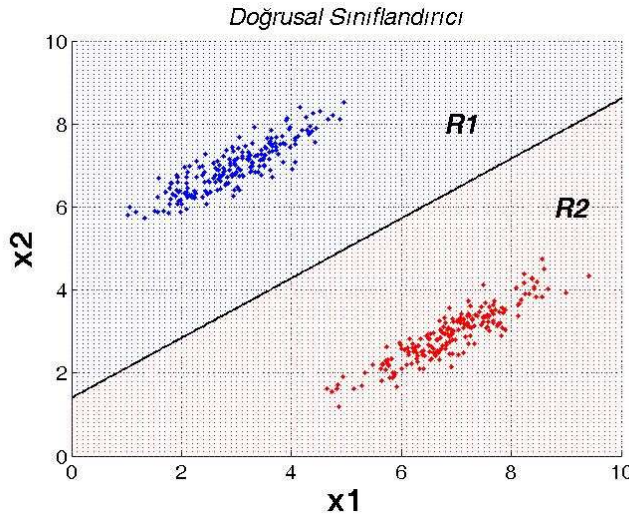
$$(\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) = \mathbf{x}^t \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \quad (1.18)$$

$$(\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) = \mathbf{x}^t \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2\boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \quad (1.19)$$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}^t \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2\boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i) \quad (1.20)$$

$$g_i(\mathbf{x}) = \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \quad (1.21)$$

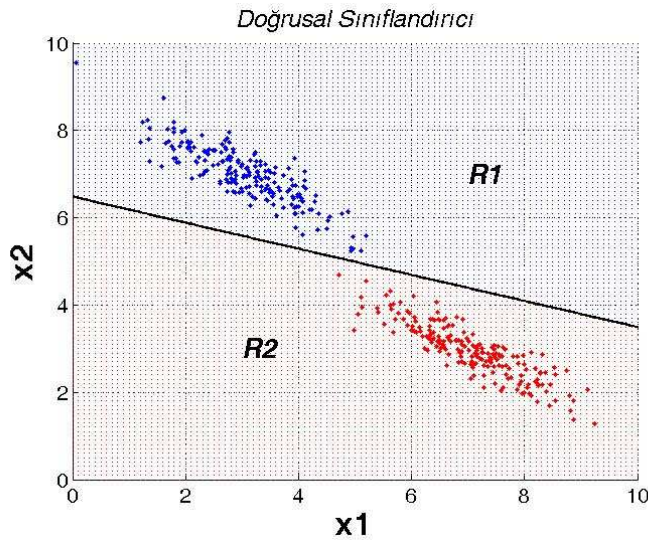
Eşitlik 1.21'deki $g_i(\mathbf{x})$ ifadesi, $\mathbf{x}$ 'in doğrusal bir fonksiyonu olduğu için, bu fonksiyonu kullanan sınıflandırıcılara doğrusal sınıflandırıcı (*linear classifier*) adı verilmektedir. Normal dağılıma uygun eşit sayıda veri içeren, iki değişkenli iki sınıfın doğrusal sınıflandırıcı ile sınıflandırılmasına örnek grafikler Şekil 1.18'de verilmiştir.



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.8 & 0.5 \\ 0.5 & 0.4 \end{bmatrix}$$



$$P(\omega_1) = P(\omega_2) = 0.5$$

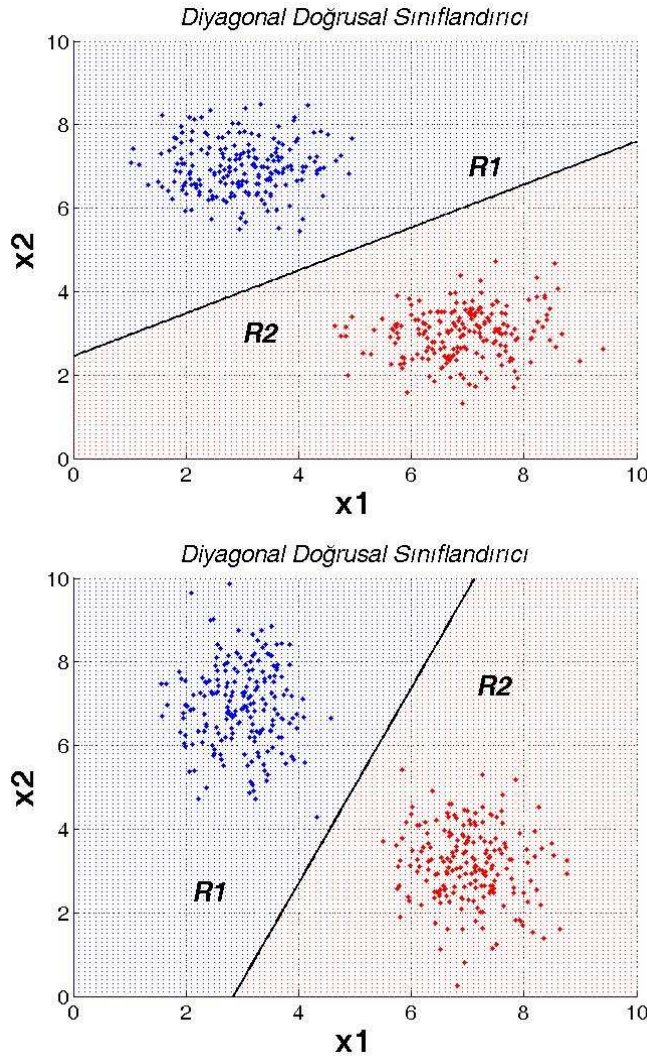
$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.8 & -0.5 \\ -0.5 & 0.4 \end{bmatrix}$$

Şekil 1.18. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin doğrusal sınıflandırıcı ile sınıflandırılması.

1.8.2.2. Diyagonal Doğrusal Sınıflandırıcı

Bu sınıflandırıcı doğrusal sınıflandırıcının farklı bir çeşidi olup, her veri sınıfının kovaryans matrisinin birbirine eşit ve diyagonal olduğu varsayılır (*diagonal linear classifier*). Dolayısıyla öznitelikler istatistiksel olarak birbirinden bağımsızdır. Normal dağılıma uygun eşit sayıda veri içeren, iki değişkenli iki sınıfın diyagonal doğrusal sınıflandırıcı ile sınıflandırılmasına örnek grafikler Şekil 1.19’da görülmektedir.



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.4 \end{bmatrix}$$

$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.8 \end{bmatrix}$$

Şekil 1.19. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin diyagonal doğrusal sınıflandırıcı ile sınıflandırılması.

Eğer her öznitelik eşit varyansa sahip olursa; kovaryans matrisi $\Sigma_i = \sigma^2 \mathbf{I}$ formunu alır.

Kovaryans matrisinin determinanı $|\Sigma_i| = \sigma^{2d}$ ve tersi $\Sigma_i^{-1} = (\frac{1}{\sigma^2}) \mathbf{I}$ olur. Eşitlik

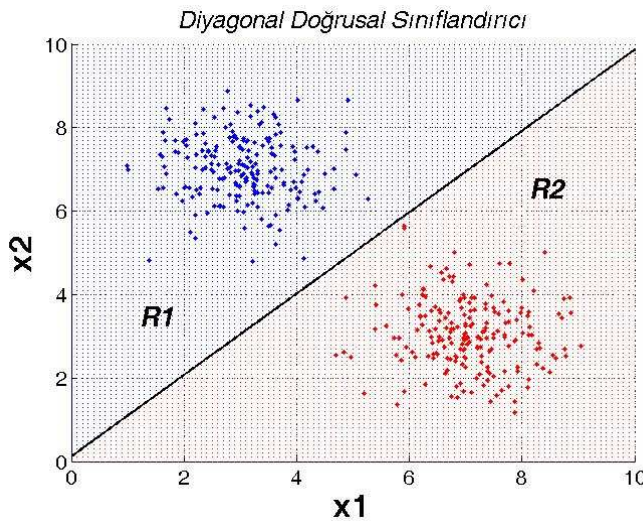
1.16'deki ilk terim, her veri sınıfı için aynı olacağından ihmal edilir ve fonksiyon eşitlik 1.24'deki forma dönüşür.

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \frac{\mathbf{I}}{\sigma^2} (\mathbf{x} - \mu_i) \quad (1.22)$$

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(\mathbf{x} - \mu_i)^t (\mathbf{x} - \mu_i) \quad (1.23)$$

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \boldsymbol{\mu}_i\|^2}{2\sigma^2} \quad (1.24)$$

Bu durumda bir $\mathbf{x}$ öznitelik vektörünün sınıflandırılması; $\mathbf{x}$ 'in her bir veri sınıfı ortalamasına olan Euclid mesafesinin $\|\mathbf{x} - \boldsymbol{\mu}_i\|$ hesaplanarak, ortalaması en yakın veri sınıfına atanması ile sağlanır [48]. Normal dağılıma uygun eşit sayıda veri içeren, eşit varyansa sahip iki değişkenli, iki sınıfın diyagonal doğrusal sınıflandırıcı ile sınıflandırılmasına örnek grafik Şekil 1.20'de görülmektedir.



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$$

Şekil 1.20. Normal dağılıma uygun eşit varyansa sahip iki değişkenli iki sınıfa ait verilerin diyagonal doğrusal sınıflandırıcı ile sınıflandırılması.

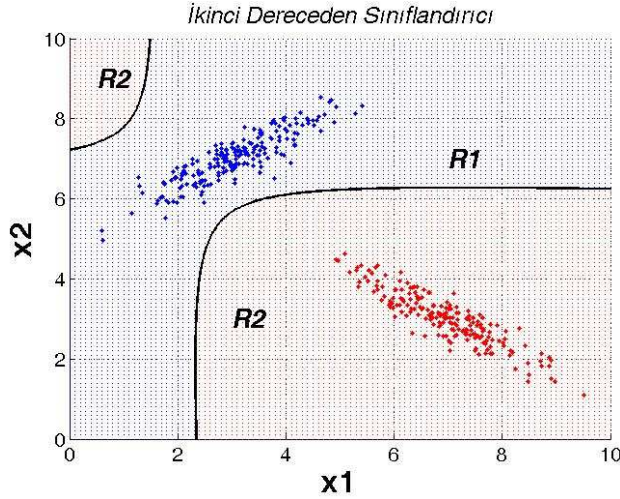
1.8.2.3. İkinci Dereceden Sınıflandırıcı

Her veri sınıfının kovaryans matrisinin farklı olduğu varsayımı, sınıflandırıcıyı ikinci dereceden bir sınıflandırıcı haline getirir. Dolayısıyla eşitlik 1.16'daki ifade;

$$g_i(\mathbf{x}) = -\frac{1}{2}[\ln |\Sigma_i| + (\mathbf{x}^t \Sigma_i^{-1} \mathbf{x} - 2\boldsymbol{\mu}_i^t \Sigma_i^{-1} \mathbf{x} + \boldsymbol{\mu}_i^t \Sigma_i^{-1} \boldsymbol{\mu}_i)] \quad (1.25)$$

halini alır. Buradaki $g_i(\mathbf{x})$ ifadesi, $\mathbf{x}$ 'in ikinci dereceden bir fonksiyonu olduğu için, bu fonksiyonu kullanan sınıflandırıcılara ikinci dereceden sınıflandırıcı (*quadratic classifier*) adı verilmektedir. Normal dağılıma uygun eşit sayıda veri içeren, iki

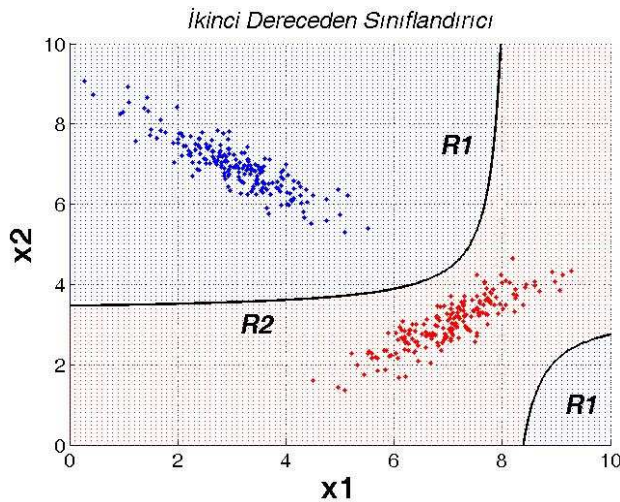
değişkenli iki sınıfın ikinci dereceden sınıflandırıcı ile sınıflandırılmasına örnek grafikler Şekil 1.21’de verilmiştir.



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 0.8 & 0.5 \\ 0.5 & 0.4 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.8 & -0.5 \\ -0.5 & 0.4 \end{bmatrix}$$



$$P(\omega_1) = P(\omega_2) = 0.5$$

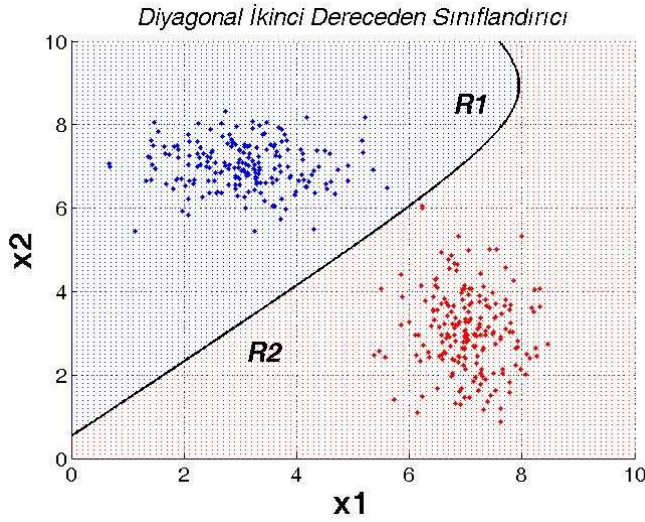
$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 0.8 & -0.5 \\ -0.5 & 0.4 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.8 & 0.5 \\ 0.5 & 0.4 \end{bmatrix}$$

Şekil 1.21. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin ikinci dereceden sınıflandırıcı ile sınıflandırılması.

1.8.2.4. Diyagonal İkinci Dereceden Sınıflandırıcı

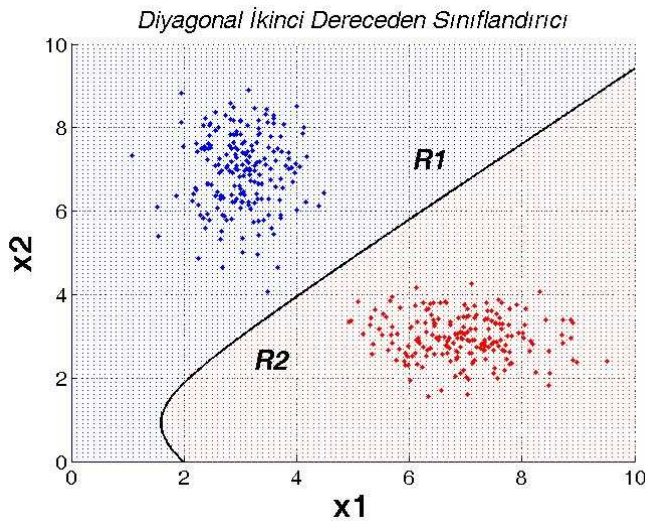
Bu sınıflandırıcı, ikinci dereceden sınıflandırıcının farklı bir çeşidi olup, her veri sınıfının kovaryans matrisinin farklı ve diyagonal olduğu varsayılır (*diagonal quadratic classifier*). Dolayısıyla öznelilikler istatistiksel olarak birbirinden bağımsızdır. Normal dağılıma uygun eşit sayıda veri içeren, iki değişkenli iki sınıfın diyagonal ikinci dereceden sınıflandırıcı ile sınıflandırılmasına örnek grafikler Şekil 1.22’de görülmektedir.



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.3 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.8 \end{bmatrix}$$



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.8 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.3 \end{bmatrix}$$

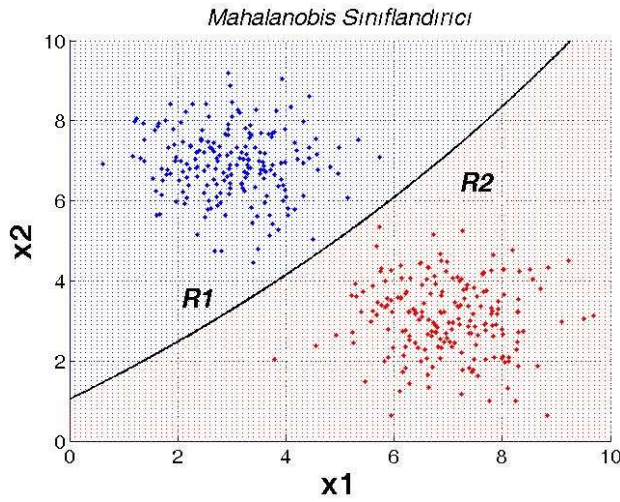
Şekil 1.22. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin diyagonal ikinci dereceden sınıflandırıcı ile sınıflandırılması.

1.8.2.5. Mahalanobis Sınıflandırıcı

Bu sınıflandırıcıda bir $\mathbf{x}$ öznitelik vektörünün sınıflandırılması; $\mathbf{x}$ 'in her bir veri sınıfı ortalamasına olan karesel Mahalanobis mesafesi $(\mathbf{x} - \mu_i)^t \Sigma_i^{-1} (\mathbf{x} - \mu_i)$ hesaplanılarak, bu mesafeye göre ortalaması en yakın veri sınıfına atanması ile sağlanır [48]. Mahalanobis sınıflandırıcı için ayırma fonksiyonu;

$$g_i(\mathbf{x}) = (\mathbf{x} - \mu_i)^t \Sigma_i^{-1} (\mathbf{x} - \mu_i) \quad (1.26)$$

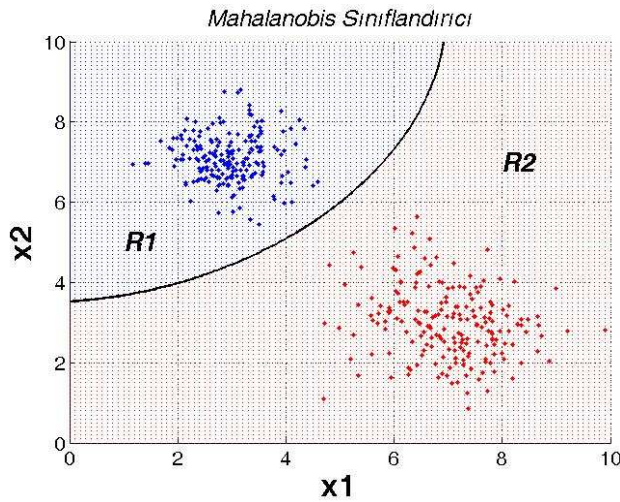
halini alır. Normal dağılıma uygun eşit sayıda veri içeren, iki değişkenli iki sınıfın Mahalanobis sınıflandırıcı ile sınıflandırılmasına örnek grafikler Şekil 1.23'de verilmiştir.



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$$



$$P(\omega_1) = P(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}$$

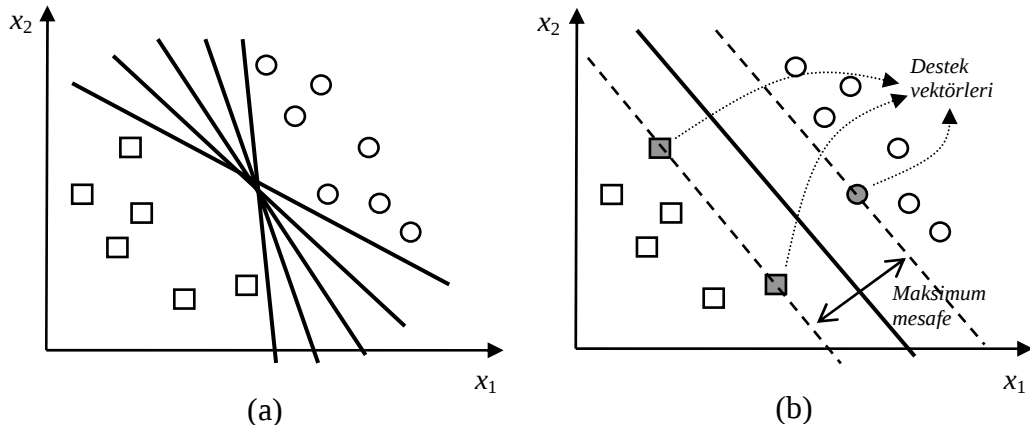
Şekil 1.23. Normal dağılıma uygun iki değişkenli iki sınıfa ait verilerin Mahalanobis sınıflandırıcı ile sınıflandırılması.

1.8.3. Destek Vektör Makineleri

Destek vektör makineleri (*Support Vector Machines - SVM*), sınıflandırma ve regresyon problemlerinin çözümü için, Vapnik tarafından geliştirilmiş danışmanlı (*supervised*) bir öğrenme metodudur. SVM nesne tanıma, el yazısı, ses ve yüz tanıma gibi çeşitli örüntü tanıma uygulamalarında yaygın olarak kullanılmaktadır. Bu tür uygulamalarda SVM'in genelleştirme performansının (*generalization performance*), başka bir ifadeyle test verilerindeki hata oranının, diğer öne çıkan metotlarla aynı yada belirgin bir şekilde daha iyi olduğu görülmektedir [49-51, 55].

İki sınıflı bir veri setinin farklı doğrusal karar sınırları (*decision boundary / hyper plane*) ile ayrıldığı durum Şekil 1.24 (a)'da görülmektedir. SVM'in temel amacı; veri

sınıflarının sınırları arasındaki mesafeyi maksimum yapan en uygun karar sınırını bulmaktır. SVM, her bir sınıfın sınırlarını belirleyen örnekleri/verileri (destek vektörler – *support vectors*) tanımlayarak bu işlemi gerçekleştirmektedir [47]. Şekil 1.24 (b)'de SVM ile en uygun karar sınırının belirlendiği durum görülmektedir.



Şekil 1.24. İki sınıflı bir veri setinin; (a) farklı doğrusal karar sınırları ile sınıflandırılması, (b) SVM ile en uygun karar sınırı belirlenerek sınıflandırılması.

1.8.3.1. Doğrusal Destek Vektör Makineleri

Hatasız sınıflandırılabilen durum:

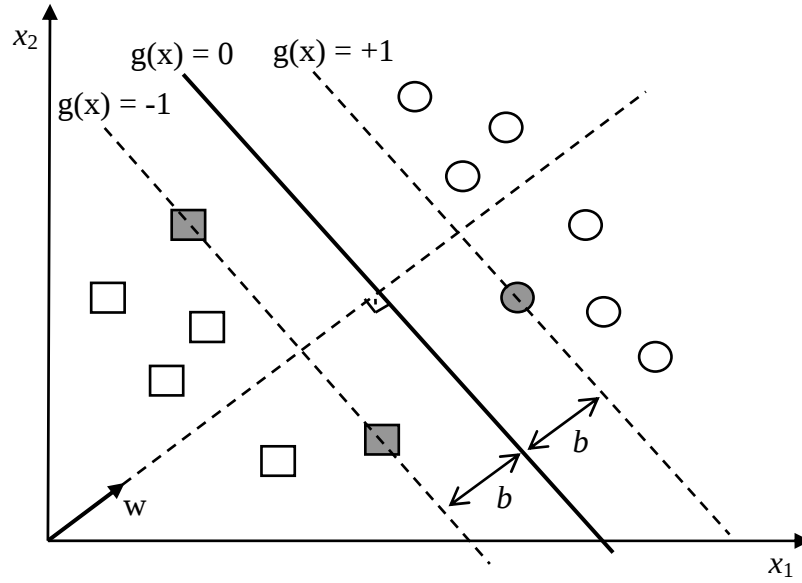
Bir veri setinde;

- $k = 1, 2, \dots, n$ veri indeksini
- $\mathbf{x}_k \in \mathbf{R}^d$; d -boyutlu öznitelik uzayındaki (*feature space*) her bir veriyi
- $i = 1, 2$; sınıf indeksini
- ω_i ; i . sınıfa ait verileri
- $z_k \in \{-1, +1\}$; ait olduğu sınıf etiketini

temsil etmek üzere, iki sınıfa ait verileri sınıflandıracak karar sınırı (hiper düzlem), eşitlik 1.27'deki formül ile tanımlanır.

$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0 \quad (1.27)$$

Burada $\mathbf{w}$ ve w_0 sırasıyla; ayırma düzleminin normal vektörünü ve eşik değerini temsil etmektedir. İki sınıflı bir veri setinin, doğrusal bir karar sınırı ile hatasız olarak ayrıldığı durum Şekil 1.25'de görülmektedir.



Şekil 1.25. İki sınıflı bir veri setinin doğrusal bir karar sınırı ile ayrılması.

Karar sınırı $g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0 = 0$ doğrusudur. Bu karar sınırının ayırdığı uzaylardan her biri bir sınıf belirtir. Bu durumda aşağıdaki koşulların gerçekleşmesi gerekir:

$$\omega_1 \text{ için } g(\mathbf{x}) > 0 \text{ ve } \omega_2 \text{ için } g(\mathbf{x}) < 0 \quad (1.28)$$

$$z_k = \begin{cases} +1; x_k \in \omega_1 \\ -1; x_k \in \omega_2 \end{cases} \quad (1.29)$$

$$z_k g(\mathbf{x}_k) > 0, \quad k = 1, 2, \dots, n \quad (1.30)$$

SVM ile amaç, veri sınıflarına eşit ve maksimum mesafedeki karar sınırının belirlenmesidir (*maximal margin classifier*). Sınıflar arasındaki mesafe ne kadar çok olursa, sınıflandırıcının o kadar iyi genelleştirme (*generalization*) sağlayacağı beklenir [48]. Karar sınırına en yakın verilerin bu düzleme olan en kısa mesafeleri;

$$\frac{z_k g(\mathbf{x}_k)}{\|\mathbf{w}\|} \geq b, \quad b > 0 \quad (1.31)$$

eşitliği ile ifade edilir. Buradaki $\|\mathbf{w}\|$ normal vektörünün Euclid normunu temsil etmektedir. Buradaki ifadeden b yi maksimize etmek için uygun bir normal vektörünün bulunması gerektiği anlaşılmaktadır. Ancak b için sonsuz sayıda çözüm olduğundan;

$$b\|\mathbf{w}\| = 1 \quad (1.32)$$

sınırlaması/kısıtlaması getirilerek tek bir çözüm sağlanmış olur. Eşitlik 1.31 ve eşitlik 1.32 birlikte düzenlenerek,

$$z_k g(\mathbf{x}_k) \geq 1, \quad k = 1, 2, \dots, n \quad (1.33)$$

halini alır. b yi maksimize etmek için $\|\mathbf{w}\|$ ifadesi eşitlik 1.33'deki koşulu sağlayacak şekilde minimize edilebilir. Buradan çözülecek problemi;

$$\frac{1}{2} \|\mathbf{w}\|^2 \text{ ifadesini,} \quad (1.34)$$

$$z_k g(\mathbf{x}_k) \geq 1 \text{ koşulu ile minimize etmek}$$

şeklinde tanımlayabiliriz. Bu ifade ikinci dereceden bir optimizasyon problemidir. Bu problemi çözebilmek için Lagrange optimizasyonu kullanılır. Lagrange çarpanları α_k kullanılarak, bu problem eşitlik 1.35'deki sınırlamasız/kısıtlamasız hale dönüştürülür.

$$L_p(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{k=1}^n \alpha_k [z_k (\mathbf{w}^t \mathbf{x}_k + w_0) - 1], \quad \alpha_k \geq 0, \quad k = 1, 2, \dots, n \quad (1.35)$$

Bu ifade $\mathbf{w}$, w_0 'a göre minimize ve $\alpha_k \geq 0$ için maksimize edilmelidir. L_p fonksiyonunun $\mathbf{w}$, w_0 'a göre kısmi türevleri alınarak 0'a eşitlenir.

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{k=1}^n \alpha_k z_k \mathbf{x}_k \quad (1.36)$$

$$\frac{\partial L_p}{\partial w_0} = 0 \Rightarrow \sum_{k=1}^n \alpha_k z_k = 0 \quad (1.37)$$

Eşitlik 1.36 ve eşitlik 1.37, eşitlik 1.35'de yerine konularak Lagrange fonksiyonunun ikincil formu (dual form) L_d elde edilir.

$$L_d(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \mathbf{w}^t \sum_{k=1}^n \alpha_k z_k \mathbf{x}_k + w_0 \sum_{k=1}^n \alpha_k z_k + \sum_{k=1}^n \alpha_k$$

$$L_d(\mathbf{w}, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^t \mathbf{w} - \mathbf{w}^t \mathbf{w} + \sum_{k=1}^n \alpha_k$$

$$L_d(\mathbf{w}, \boldsymbol{\alpha}) = -\frac{1}{2} \mathbf{w}^t \mathbf{w} + \sum_{k=1}^n \alpha_k \quad (1.38)$$

Buradan çözülecek problemi;

$$L_d(\alpha) = \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,j}^n \alpha_k \alpha_j z_k z_j \mathbf{x}_j^t \mathbf{x}_k \text{ fonksiyonunu,}$$

$$\sum_{k=1}^n \alpha_k z_k = 0 \text{ ve } \alpha_k \geq 0 \text{ koşulları ile maksimize etmek} \quad (1.39)$$

şeklinde ifade edebiliriz. Bu ifade α_k için çözüldüğünde, örneklerin çoğunun Lagrange çarpan değerinin $\alpha_k = 0$ olduğu, çok az bir kısmının ise $\alpha_k > 0$ ifadesini sağladığı görülür. $\alpha_k > 0$ ifadesini sağlayan $\mathbf{x}_k$ örnekleri, ayırma düzlemine eşit mesafedeki en yakın veriler olup destek vektörleri (*support vectors*) olarak tanımlanmaktadır. Şekil 9'daki koyu renkli şekiller bu vektörleri göstermektedir. Eşitlik 1.33'deki $\mathbf{w}$ ifadesinin hesaplanmasında, sadece destek vektörlerinin işleme etkileri vardır. Dolayısıyla karar sınırının belirlenmesinde yalnız destek vektörleri etkilidirler.

Destek vektörleri, en uygun karar sınırının belirlenmesini sağlamakla birlikte sınıflandırılması en güç verilerdir. Bu veriler sınıflandırma uygulamasında en fazla bilgi içeren noktalardır [48]. Destek vektörleri;

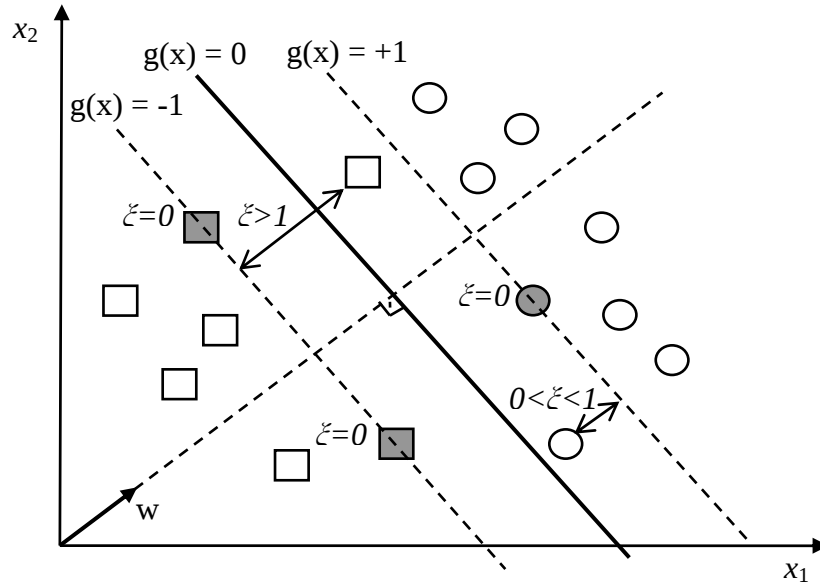
$$z_k (\mathbf{w}^t \mathbf{x}_k + w_0) = 1 \quad (1.40)$$

eşitliğini sağlayan ve karar sınırından b mesafesi kadar uzaklıkta bulunan örneklerdir. w_0 ifadesi ise, herhangi bir destek vektörü yardımıyla ve eşitlik 1.40'dan türetilen eşitlik 1.41 ile hesaplanabilir.

$$w_0 = \frac{1 - z_k \mathbf{w}^t \mathbf{x}_k}{z_k} = z_k - \mathbf{w}^t \mathbf{x}_k \quad (1.41)$$

Hatasız sınıflandırılmayan durum:

Doğrusal sınıflandırmanın hatasız olarak gerçekleştirilemediği durumlarda, doğrusal sınıflandırma en az sayıda hataya izin verecek şekilde yapılandırılabilir (*soft margin classifier*). İki sınıflı bir veri setinin, hatalı örneklerle izin veren, doğrusal bir karar sınırı ile ayrıldığı durum Şekil 1.26'da görülmektedir.



Şekil 1.26. İki sınıflı bir veri setinin, hatalı örneklerle izin verecek şekilde doğrusal bir karar sınırı ile ayrılması.

Hata terimleri olarak $\xi_k \geq 0$ tanımlanarak eşitlik 1.30'daki ifade yeniden düzenlenirse,

$$z_k g(\mathbf{x}_k) \geq 1 - \xi_k, \quad \xi_k \geq 0, \quad k = 1, 2, \dots, n \quad (1.42)$$

elde edilir. Burada hata terimi için 3 durum söz konusudur:

- $\xi_k = 0 \Rightarrow \mathbf{x}_k$ verisi doğru sınıflandırılmıştır.
- $0 < \xi_k < 1 \Rightarrow \mathbf{x}_k$ verisi doğru sınıflandırılmış ancak sınır bölgesindedir.
- $\xi_k \geq 1 \Rightarrow \mathbf{x}_k$ verisi yanlış sınıflandırılmıştır.

Çözülecek problemi;

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{k=1}^n \xi_k \text{ ifadesini,}$$

$$z_k g(\mathbf{x}_k) \geq 1 - \xi_k \text{ ve } \xi_k \geq 0 \text{ koşulları ile minimize etmek} \quad (1.43)$$

şeklinde tanımlayabiliriz. Buradaki C ifadesi, ceza parametresi (*penalty factor*) olarak işlev görmekte olup, sınıflar arasındaki mesafe artarken, hatalı sınıflandırmaların az olması için düzenleme yapmaktadır (*trade-off*) [52]. Lagrange çarpanları α_k ve μ_k kullanılarak, bu problem;

$$L_p(\mathbf{w}, w_0, \xi, \alpha, \mu) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{k=1}^n \xi_k - \sum_{k=1}^n \alpha_k [z_k (\mathbf{w}^t \mathbf{x}_k + w_0) - 1 + \xi_k] - \sum_{k=1}^n \mu_k \xi_k \quad (1.44)$$

halini alır. Bu fonksiyonun çözümü için gerekli şartlar aşağıda belirtilmiştir.

$$\begin{aligned} z_k (\mathbf{w}^t \mathbf{x}_k + w_0) - 1 + \xi_k &\geq 0 \\ \alpha_k [z_k (\mathbf{w}^t \mathbf{x}_k + w_0) - 1 + \xi_k] &= 0 \\ \alpha_k &\geq 0, \mu_k \geq 0, \xi_k \geq 0, \mu_k \xi_k = 0, k = 1, 2, \dots, n \end{aligned} \quad (1.45)$$

Bu fonksiyonun w , w_0 ve ξ 'a göre kısmi türevleri alınarak 0'a eşitlenir.

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{k=1}^n \alpha_k z_k \mathbf{x}_k \quad (1.46)$$

$$\frac{\partial L_p}{\partial w_0} = 0 \Rightarrow \sum_{k=1}^n \alpha_k z_k = 0 \quad (1.47)$$

$$\frac{\partial L_p}{\partial \xi} = 0 \Rightarrow \alpha_k + \mu_k = C \quad (1.48)$$

Bu sonuçlar kullanılarak, eşitlik 1.38'deki fonksiyon ile aynı olan, Lagrange fonksiyonunun ikincil formu (*dual form*) L_d elde edilir. Bu fonksiyonun eşitlik 1.38'den tek farkı, koşullarının farklı olmasıdır. Buradan çözülecek problemi;

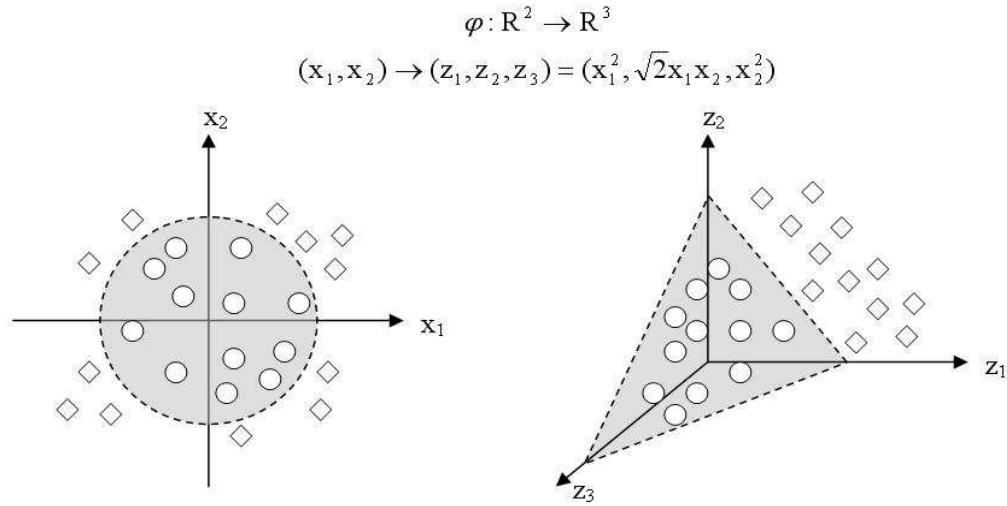
$$\begin{aligned} L_d(\alpha) &= \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,j} \alpha_k \alpha_j z_k z_j \mathbf{x}_j^t \mathbf{x}_k \text{ fonksiyonunu,} \\ \sum_{k=1}^n \alpha_k z_k &= 0 \text{ ve } 0 \leq \alpha_k \leq C \text{ koşulları ile maksimize etmek} \end{aligned} \quad (1.49)$$

şeklinde ifade edilebilir. Bu ifade, bir önceki kısımdaki işlem adımları ile çözülerek karar sınırının parametreleri belirlenir.

1.8.3.2. Doğrusal Olmayan Destek Vektör Makineleri

Giriş uzayında (*input space*) doğrusal olarak sınıflandırılmayan veriler, doğrusal olmayan bir ϕ fonksiyonu ile daha yüksek boyutlu bir uzaya (*feature space*) taşınarak,

bu uzayda doğrusal olarak sınıflandırılabilir. Şekil 1.27’de bu uygulamaya örnek bir durum görülmektedir.



Şekil 1.27. Doğrusal olarak sınıflandırılmayan 2 boyutlu verilerin, 3 boyutlu uzaya taşınarak sınıflandırılması.

Çekirdek fonksiyonları:

Eşitlik 1.38 ve eşitlik 1.48’deki SVM optimizasyon problemlerinde, $\mathbf{x}_j^t \mathbf{x}_k$ ifadesi nokta çarpımı (*inner product*) tanımlamaktadır. Veriler doğrusal olmayan bir ϕ fonksiyonu ile daha yüksek boyutlu bir uzaya taşındıklarında, bu ifade

$$K(\mathbf{x}_k, \mathbf{x}_j) = \phi(\mathbf{x}_k) \cdot \phi(\mathbf{x}_j) = \phi(\mathbf{x}_k)^t \phi(\mathbf{x}_j) \quad (1.50)$$

halini alır. Buradaki $K(\mathbf{x}_k, \mathbf{x}_j)$ ifadesi, çekirdek fonksiyonu (*kernel function*) olarak adlandırılır. Belirli bir çekirdek fonksiyonu kullanılarak, eşitlik 1.50’deki nokta çarpımların yapılmasına gerek kalmaz. Diğer bir deyişle, çekirdek fonksiyonu kullanılarak, ϕ fonksiyonu hesaplanmadan veriler yüksek boyutlu bir uzaya taşınabilmektedir. Çekirdek fonksiyonu ile Eşitlik 1.39 ve eşitlik 1.49’deki ifadeler,

$$L_d(\boldsymbol{\alpha}) = \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,j} \alpha_k \alpha_j z_k z_j K(\mathbf{x}_k, \mathbf{x}_j) \quad (1.51)$$

formuna dönüşür. Ayırma düzleminin eşitlik 1.27’deki ifadesi ise, eşitlik 1.36 ve eşitlik 1.46’dan yararlanılarak eşitlik 1.52 şeklinde yazılabilir.

$$g(\mathbf{x}) = \sum_{k=1}^n \alpha_k z_k K(\mathbf{x}_k, \mathbf{x}) + w_0 \quad (1.52)$$

En yaygın kullanılan çekirdek fonksiyonları aşağıda verilmiştir:

- Doğrusal (*Linear*): $K(\mathbf{x}_k, \mathbf{x}_j) = \mathbf{x}_k^t \mathbf{x}_j$ (1.53)

- Polinom (*Polynomial*): $K(\mathbf{x}_k, \mathbf{x}_j) = (\mathbf{x}_k^t \mathbf{x}_j + 1)^d$, $d \geq 2$ (1.54)

- Gauss RBF (*Gaussian Radial Basis Function*):

$$K(\mathbf{x}_k, \mathbf{x}_j) = \exp(-\|\mathbf{x}_k - \mathbf{x}_j\|^2 / 2\sigma^2) \quad (1.55)$$

SVM'in önemli bir özelliği de; sınıflandırıcının kompleksliğinin/karmaşıklığının; veri setinin taşındığı uzayın boyutu yerine destek vektörlerinin sayısına bağlı olmasıdır. Bunun sonucu olarak SVM, diğer metotlara nazaran aşırı öğrenmeye (*over-fitting*) daha az meyillidir [48].

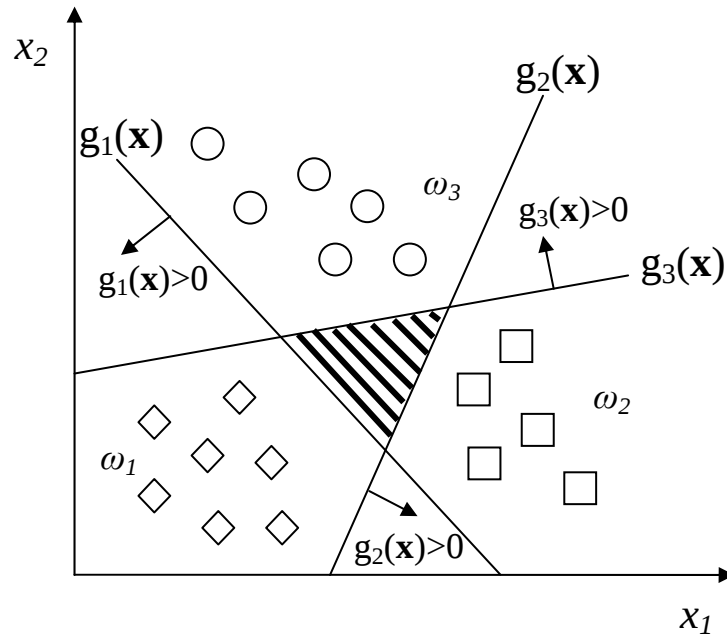
1.8.3.3. Çok Sınıflı Destek Vektör Makineleri

SVM temelde ikili sınıflandırıcı olduğundan çok sınıflı problemler, veri setinde çeşitli yöntemlerle (değişik şekillerde) ikili sınıflar oluşturularak SVM algoritmasının uygulanması ve elde edilen sonuçların incelenmesi ile çözümlenir. Çok sınıflı verilerin SVM ile sınıflandırılması için kullanılan 2 genel yöntem aşağıdaki kısımda açıklanmıştır.

1.8.3.3.1. SVM (1-r) Yöntemi

SVM (1-r) (*one versus rest*); veri setindeki bir sınıfın verilerine karşılık, diğer sınıflara ait tüm verilerin sınıflandırıcıya verildiği yöntemdir. Bu yöntemde göre sınıflandırmaya örnek bir durum, Şekil 1.28'de görülmektedir. Bu yöntemde; c sınıfa sahip bir veri seti için, c adet ikili sınıflandırma işlemi SVM ile yapılmaktadır. Bir test verisi, bu sınıflandırma işlemlerinin sonucu incelenerek ilgili sınıfa atanır. Eğer bu test verisi ω_i sınıfına ait ise bu sınıflandırıcıya ait fonksiyon ifadesi $g_i(\mathbf{x}) > 0$ ve diğer tüm sınıflandırıcıların fonksiyon ifadeleri $g_i(\mathbf{x}) < 0$ olmalıdır. Birden fazla sınıflandırıcıya ait fonksiyon ifadesinin $g_i(\mathbf{x}) > 0$ veya tüm sınıflandırıcıların fonksiyon ifadelerinin

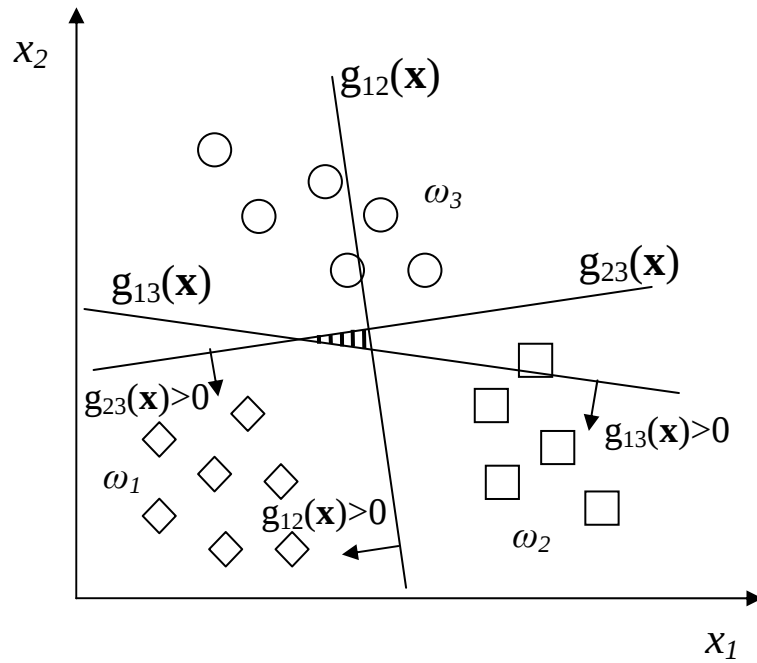
$g_i(\mathbf{x}) < 0$ olması durumunda test verisi bir sınıfa atanamaz. Şekil 1.28’de görülen taralı alan sınıflandırılmayan bölgeyi temsil etmektedir. Bu problem, $g_i(\mathbf{x})$ ifadesi en büyük olan sınıflandırıcının sınıfına atama yapılarak aşılabilir [56].



Şekil 1.28. SVM (1-r) sınıflandırmasına örnek bir durum.

1.8.3.3.2. SVM (1-1) Yöntemi

SVM (1-1) (*one versus one*); veri setindeki sınıfların ikili olarak birbirleri ile karşılaştırıldığı yöntemdir. Bu yöntemde göre sınıflandırmaya örnek bir durum, Şekil 1.29’da görülmektedir. Bu yöntemde; c sınıfa sahip bir veri seti için, $c(c-1)/2$ adet ikili sınıflandırma işlemi SVM ile yapılmaktadır. Bir test verisi, bu ikili sınıflandırıcıların sonuçları incelenerek en fazla oyu alan sınıfa atanır. Eğer en fazla oya sahip birden fazla sınıf var ise; test verisi bu sınıflar arasında en düşük indeksli sınıfa veya önsel olasılık yoğunluk fonksiyonu (*prior probability*) $P(\omega_i)$ en yüksek olan sınıfa atanır [53, 56]. Şekil 1.29’da görülen taralı alan, eşit oya sahip sınıfların bulunduğu bölgeyi temsil etmektedir.



Şekil 1.29. SVM (1-1) sınıflandırmasına örnek bir durum.

1.8.3.4. LIBSVM Kütüphanesi/Yazılımı

SVM yöntemi ile yapılan sınıflandırma çalışmalarında,

- MATLAB'deki yerleşik SVM fonksiyonları (komutları)
- LIBSVM kütüphanesi

kullanılarak, MATLAB ortamında ayrı ayrı programlar geliştirilmiştir [58]. LIBSVM, SVM yönteminin sınıflandırma ve regresyon uygulamalarında kullanılmasını sağlayan, çeşitli araçlara sahip bir kütüphanedir [54]. LIBSVM'de, sınıflandırma performansına etkisi bulunan, ceza ve çekirdek fonksiyonu parametrelerinin ayarlanmasına kolaylıkla erişilebilmektedir. MATLAB'deki yerleşik SVM fonksiyonlarında (komutlarında) ise bu parametrelere erişim mümkün olmamaktadır. Ayrıca MATLAB'deki bu fonksiyonlar (komutlar) sadece ikili sınıflandırma yapmaya olanak sağladığı için, çok sınıflı problemler ikili sınıflar halinde düzenlenerek programlar yazılmak zorundadır. LIBSVM ise, çok sınıflı problemlerin sınıflandırılmasına uygun bir yapıya sahiptir ve sınıflandırmada SVM (1-1) yöntemi kullanılmaktadır [54]. Her iki yazılım arasındaki bu farklılıklardan dolayı; MATLAB'deki yerleşik fonksiyonlar (komutlar) kullanılarak hazırlanan SVM programları ile LIBSVM kütüphanesi kullanılarak hazırlanan SVM programlarının, sınıflandırma performanslarının incelenmesine karar verilmiştir.

LIBSVM’de, eşitlik 1.43’deki optimizasyon probleminin çözümü aranmaktadır. LIBSVM’de kullanılan çekirdek fonksiyonları; eşitlik 1.53, 1.54 ve 1.55’deki fonksiyonlarda çeşitli parametreler tanımlanarak türetilmişlerdir.

- Doğrusal: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ (1.56)

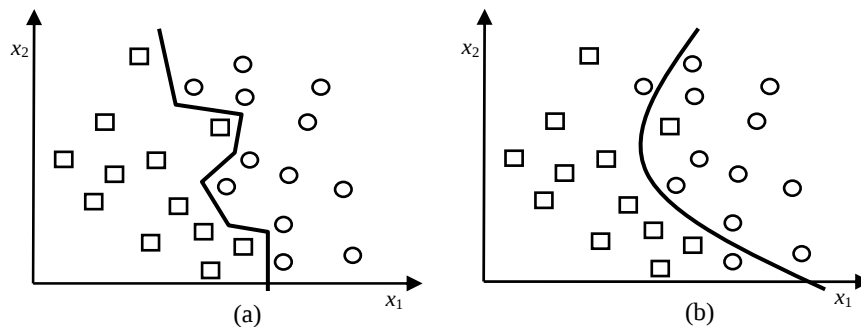
- Polinom: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0, d \geq 2$ (1.57)

- Radyal tabanlı fonksiyon (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0$ (1.58)

Buradaki γ, r ve d çekirdek parametreleridir [54].

1.9. Sınıflandırma Performansının Değerlendirilmesi

Sınıflandırma modeli seçilip eğitim tamamlandıktan sonra, bu modelin gerçek performansının tahmini için daha önce görülmemiş/verilmemiş verilerle değerlendirilmesi gerekmektedir. Bu değerlendirme işlemine, genelleştirme performansı (*generalization performance*) adı verilmektedir. Bir sınıflandırıcının bir eğitim veri setindeki tüm örnekleri doğru sınıflandıracak şekilde eğitilmesi mümkündür. Ancak eğer sınıflandırıcı örneklerin genel karakteristiklerinden daha fazlası ile öğrenme gerçekleştirirse, daha önce görülmemiş/verilmemiş verilerde başarısız bir performans sergileyebilmektedir. Bu tür bir aşırı öğrenme (*over-training/over-fitting*), sınıflandırıcının eğitim veri setinin genelindeki bir gürültüyü modellemesine/öğrenmesine ve dolayısıyla yeni örneklerdeki tahmin performansının düşmesine neden olabilmektedir. Bu sebeple bir sınıflandırıcıdan, yeni örnekleri yeterince iyi sınıflandırabilmesi için, eğitim veri setinden sadece gerektiği kadar öğrenme yapması beklenmektedir [45]. Burada açıklanan duruma örnek teşkil edecek grafikler Şekil 1.30’da görülmektedir.



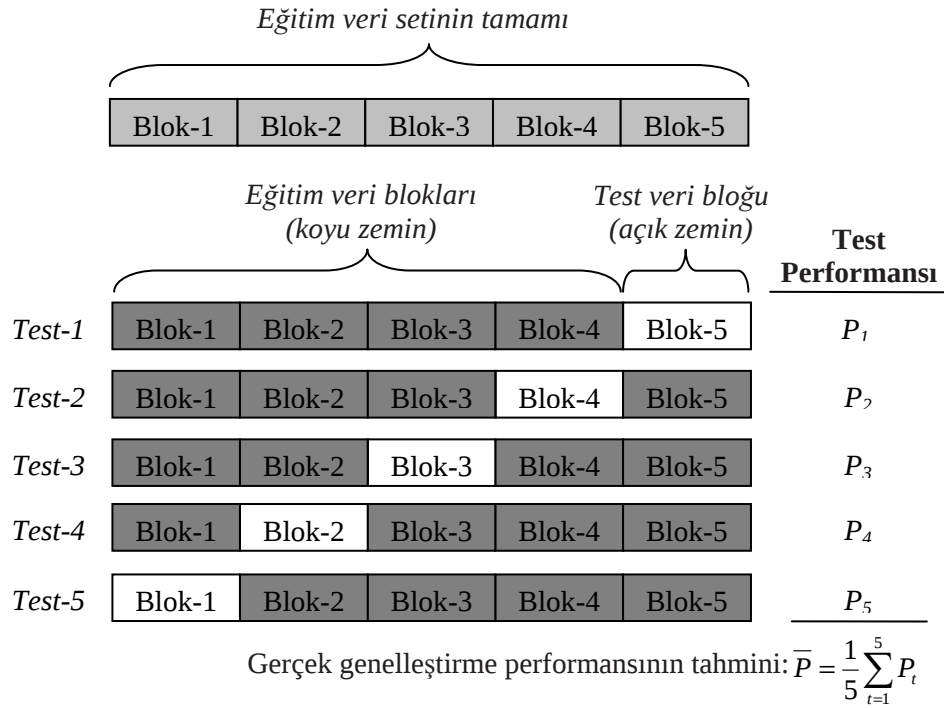
Şekil 1.30. İki sınıflı bir veri seti için: (a) aşırı öğrenme durumundaki (*over-fitting*) bir sınıflandırıcı, (b) genelleştirme performansı yüksek bir sınıflandırıcı.

Genelleştirme performansını değerlendirmek için tercih edilen en basit ve dolayısıyla en yaygın metotlardan birisi, eğitim veri setinin tamamının iki kısma ayrılarak, ilk kısmın eğitim ve ikinci kısmın da algoritmanın performansını test için kullanılmasıdır. Veri setinin ikinci kısmındaki performans, algoritmanın gerçek (ve bilinmeyen verilerden elde etmesi beklenen) performansının tahmini olarak değerlendirilir. Bununla birlikte, algoritmanın gelecekteki performansının bu yolla tahmininde, hala akla gelen bazı endişeler bulunmaktadır [47]:

1. Veri setinin bir kısmı test için ayrıldığından, orijinal veri örneklerinin tamamı eğitimde kullanılamamaktadır. Başarılı bir eğitim için yeterli ve tanımlayıcı (*representative*) eğitim verisinin bulunması başlıca öneme sahip olduğundan, bu yaklaşım en uygun (*optimal*) performansın altında sonuçlanabilir.
2. Orijinal veri setinin ne kadarı test için ayrılmalıdır? Eğer veri setinin küçük bir kısmı test için ayrılırsa, genelleştirme performansının tahmini güvenilir olmayabilir. Eğer veri setinin büyük bir kısmı test için ayrılırsa, eğitimde daha az veri kullanılacağından, bu durum zayıf öğrenmeye yol açabilir.
3. Veri sayısından bağımsız olarak, test için ayrılan verilerin, gelecekte karşılaşılabilecek verileri ifade edip edemeyeceği asla bilinemez. Test veri setindeki örnekler karar sınırına yakın veya uzakta büyük bir dağılım gösterebilirler. Örneklerin karar sınırına yakın olması test veri setinin sınıflandırılmasını zorlaştırır. Örneklerin karar sınırına uzak olması ise test veri setinin sınıflandırılmasını kolaylaştırır.

1.9.1. Çapraz Geçerlilik

Gerçek genelleştirme performansının kabul edilebilir derecede iyi bir tahminini sağlayan bir yaklaşım da çapraz geçerlilik (*cross validation*) yöntemidir. k -kat (k -fold) çapraz geçerlilikte eğitim veri setinin tamamı $k > 2$ kısma bölünür ve k adet veri bloğu oluşur. Bu k adet veri bloğundan, $k - 1$ adedi eğitim, geri kalan k . blok test için kullanılır. Bu yöntem, her seferinde farklı bir veri bloğu test için ayrılarak, k kere tekrarlanır. Elde edilen k adet test performans değerinin ortalaması alınır ve bu değer algoritmanın gerçek genelleştirme performansının tahmini olarak kabul edilir [47]. k -kat çapraz geçerlilik yönteminin $k = 5$ için grafiksel anlatımı, Şekil 1.31’de görülmektedir.



Şekil 1.31. 5-kat çapraz geçerlilik yöntemi [47].

Bu algoritmanın birçok avantajı bulunmaktadır. Bunlar [47]:

1. Tüm veri örnekleri aynı anda olmamak şartıyla, hem eğitim hem de test için kullanılmaktadır.
2. Bu yöntem k -kere tekrarlandığından, sınıflandırmaya yatkın olan veya olmayan blokların meydana gelme olasılığı ortalama alınarak düşürülmektedir.
3. Gerçek genelleştirme performansının güven aralığı; $k > 30$ ise çift kuyruk z -testi, aksi takdirde çift kuyruk t -testi ile belirlenebilmektedir.

Veri bloklarının oluşturulması sırasında; örneklerin bloklara rasgele atanmasına, her blokta her sınıftan veri bulunmasına ve bloklar arası sınıf dağılımının dengeli olmasına dikkat edilmelidir.

Çapraz geçerlilikte k 'nın seçimi veri setine bağlıdır. k çok büyük seçilirse; veri setini çok fazla kısma böler ve büyük miktarda verinin eğitim için kullanılmasına olanak sağlar. Bununla birlikte bu durumda çok az test örneği üzerinde çalışılması, test performanslarında büyük değişimler meydana getirir ve güven aralığının artmasına neden olur. Diğer yandan k 'nın çok küçük seçilmesi; veri setini daha az kısma böler ve eğitim için ayrılan veri sayısının azalmasına sebep olur. Bu durum ise test

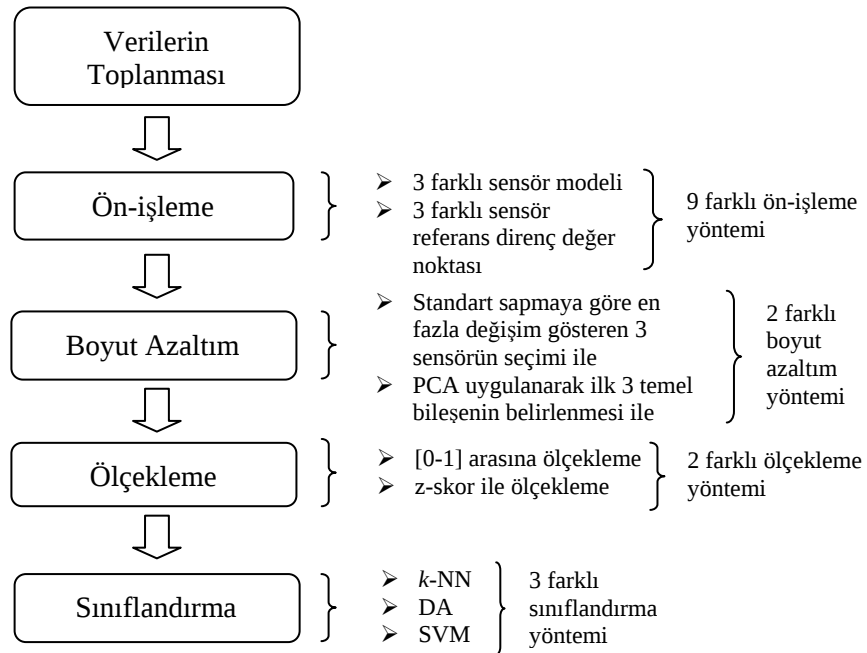
performanslarında küçük deęişimler meydana getirir. Yeterince büyük veri setleri için k genellikle 5 veya 10 olarak seçilir [45]. Daha küçük veri setlerinde ise, eğitim verilerini artırmak için k daha büyük seçilebilir. Veri setindeki örnek (veri) sayısı N ile ifade edilecek olursa; $k = N$ durumu, tek bir örneğin dışarıda bırakıldığı çapraz geçerlilięi (*leave one out cross validation* - LOOCV) tanımlamaktadır. Bu durumda N kez tekrarlanan bir süreç içerisinde, test için ayrılan bir örnek hariç tüm veri seti eğitim için kullanılmaktadır. Bu yöntem genelleştirme performansının en iyi tahminini sağlamasına rağmen, genellikle, istenmeyen bir durum olan güven aralığının geniş bir alanda oluşmasına yol açmaktadır. Buna ek olarak hesaplama karmaşıklığını en üst seviyeye getirmesi de dezavantajlarından birisidir [47].

2. BÖLÜM

SINIFLANDIRMA ÇALIŞMALARI

2.1. Sınıflandırma Çalışmalarında Kullanılan Yöntemler

Bu tez çalışmasında, üç ayrı mikroorganizma seti üzerinde çalışılmıştır. Her bir mikroorganizma setinde, elektronik burun cihazı ile elde edilen mikroorganizma koku verilerinin sınıflandırılmasına ilişkin uygulamalar geliştirilmiştir. Mikroorganizma koku verilerine çeşitli ön-işleme, boyut azaltım, ölçekleme ve sınıflandırma yöntemleri uygulanmıştır. Bu yöntemlerin sınıflandırma performansına etkileri araştırılmıştır. Kullanılan yöntemler, Şekil 2.1’de görülmekte olup, bu yöntemler hakkında ayrıntılı bilgiler Bölüm 1’de verilmiştir.



Şekil 2.1. Mikroorganizma koku verilerinin sınıflandırılmasında kullanılan yöntemler.

Sınıflandırma aşamasında, 3 farklı yöntem üzerinde durulmuştur. Bu yöntemlerin çeşitli alt metotları da incelenerek sınıflandırma performansları karşılaştırılmıştır. Kullanılan yöntemler ve bu yöntemlerin alt metotları aşağıda maddeler halinde belirtilmiştir:

1. *k*-En Yakın Komşuluk Algoritması (*k*-NN): *k*-NN ile sınıflandırma işleminde, Euclid mesafesine (*Euclidean distance*) göre $k = 1, 3, 5$ değerleri için sınıflandırma başarıları hesaplanmıştır.
2. Ayırma Analizi (DA): DA ile sınıflandırma işleminde;
 - Doğrusal (*Linear*)
 - Diyagonal doğrusal (*Diagonal linear*)
 - İkinci dereceden (*Quadratic*)
 - Diyagonal ikinci dereceden (*Diagonal quadratic*)
 - Mahalanobis

fonksiyonları olmak üzere 5 ayrı alt metot kullanılmıştır.

3. Destek Vektör Makineleri (SVM): SVM ile sınıflandırma işleminde çekirdek fonksiyonu olarak;
 - Doğrusal (*Linear*)
 - İkinci dereceden polinom (*Quadratic*)
 - Üçüncü dereceden polinom
 - RBF

uygulanmıştır. Yapılan sınıflandırma çalışmalarında,

- MATLAB'deki yerleşik SVM fonksiyonları (komutları)
- LIBSVM kütüphanesi

kullanılarak, MATLAB ortamında ayrı ayrı programlar geliştirilmiştir.

MATLAB'deki yerleşik SVM fonksiyonları (komutları), sadece ikili sınıflandırma yapmaya olanak sağladığı için, çok sınıflı problemler ikili sınıflar halinde düzenlenerek programlar hazırlanmıştır. Bu programlarda SVM (1-r) ve SVM (1-1) yöntemleri kullanılmıştır. LIBSVM ise, çok sınıflı problemlerin sınıflandırılmasına uygun bir yapıya sahiptir ve sınıflandırmayı SVM (1-1) yöntemine göre gerçekleştirmektedir. Ayrıca LIBSVM, MATLAB yazılımından farklı olarak, sınıflandırma performansına etkisi bulunan ceza (C) ve çekirdek fonksiyonu parametrelerinin (d , r , γ) ayarlanmasını mümkün kılmaktadır.

LIBSVM ile yapılan sınıflandırma çalışmalarında, aşağıdaki işlem sırası takip edilmiştir [59]:

- a) Veri setleri LIBSVM formatına uygun hale getirilmiştir.
- b) Veri setleri ölçeklenmeden ve ölçeklenerek sınıflandırıcıya verilmiştir. Bu sayede ölçeklemenin sınıflandırma performansına etkisi incelenmiştir.
- c) Çekirdek fonksiyonu parametreleri; ikinci dereceden polinom için $d = 2$, $r = 1$ ve üçüncü dereceden polinom için $d = 3$, $r = 1$ olarak seçilmiştir. Bu değerlerin böyle seçilmesinin sebebi; çekirdek fonksiyonlarının MATLAB ortamında kullanılan SVM fonksiyonları ile tutarlı olması içindir.
- d) LIBSVM içerisindeki yerleşik çapraz geçerlilik uygulaması aktif hale getirilerek, veri setlerinin çekirdek fonksiyonlarına göre sınıflandırılmasındaki en başarılı sonucu sağlayan C ve γ parametreleri belirlenmiştir. Çapraz geçerlilik için 5-kat CV metodu seçilmiştir. Her veri seti için çekirdek fonksiyonlarına göre en başarılı sınıflandırma sonucunu sağlayan (C, γ) çifti belirlenirken, öncelikle geniş aralıklarda taramalar (*a coarse grid search*) yapılmıştır. Bu taramalar $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ ve $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ değerleri için gerçekleştirilmiştir. Başarılı sonucu sağlayan (C, γ) çifti bulunduğundan sonra, bu parametrelerin yakın komşuluklarında daha dar kapsamda taramalar (*a finer grid search*) yapılmıştır. Bu sayede en başarılı sınıflandırma sonucunu sağlayan (C, γ) çifti belirlenmiştir.

2.2. Mikroorganizma Seti-1 için Sınıflandırma Sonuçları

Çeşitli enfeksiyon hastalıklarına sebep olan ve rutinde sıklıkla karşılaşılan 8 mikroorganizmadan oluşan Mikroorganizma Seti-1'de, elektronik burun cihazının kullanımı araştırılmıştır [60]. Her mikroorganizma, 8 ayrı petriye ekim yapılarak toplam 64 petri üzerinde çalışılmış ve her birinden birer koku verisi alınmıştır. Dolayısıyla her mikroorganizmadan 8 koku verisi olmak üzere toplam 64 koku verisi üzerinde çalışılmıştır. Farklı yöntemlere göre elde edilen sınıflandırma sonuçları, aşağıdaki kısımlarda açıklanmıştır.

2.2.1. *k*-NN Algoritması ile Sınıflandırma

k-NN algoritması kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.1’de görülmektedir. Bu tablodaki koyu değerler, *k* değerlerine göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.1. Veri setlerinin *k*-NN algoritması kullanılarak 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			Ortalama
		<i>k</i> = 1	<i>k</i> = 3	<i>k</i> = 5	<i>k</i> = 1	<i>k</i> = 3	<i>k</i> = 5	<i>k</i> = 1	<i>k</i> = 3	<i>k</i> = 5	
Sensör Referans değeri - 1	<i>d1 & std</i>	64.06	67.19	64.06	60.94	68.75	65.63	64.06	64.06	65.63	64.93
	<i>d1 & pca</i>	65.63	67.19	65.63	73.44	71.88	70.31	67.19	62.50	68.75	68.06
	<i>fd1 & std</i>	59.38	60.94	59.38	64.06	68.75	64.06	65.63	67.19	67.19	64.06
	<i>fd1 & pca</i>	59.38	60.94	59.38	70.31	68.75	65.63	67.19	68.75	60.94	64.58
	<i>nfd1 & std</i>	56.25	48.44	45.31	59.38	51.56	50.00	60.94	50.00	50.00	52.43
	<i>nfd1 & pca</i>	59.38	51.56	50.00	65.63	60.94	62.50	64.06	60.94	62.50	59.72
	Ortalama	60.68	59.38	57.29	65.63	65.10	63.02	64.84	62.24	62.50	62.30
Sensör Referans değeri - 2	<i>d2 & std</i>	57.81	60.94	64.06	65.63	62.50	62.50	65.63	59.38	62.50	62.33
	<i>d2 & pca</i>	59.38	64.06	64.06	60.94	62.50	59.38	62.50	59.38	60.94	61.46
	<i>fd2 & std</i>	59.38	59.38	56.25	62.50	62.50	60.94	65.63	68.75	64.06	62.15
	<i>fd2 & pca</i>	60.94	60.94	57.81	62.50	64.06	64.06	57.81	62.50	60.94	61.28
	<i>nfd2 & std</i>	50.00	43.75	45.31	51.56	45.31	48.44	50.00	43.75	46.88	47.22
	<i>nfd2 & pca</i>	53.13	48.44	43.75	59.38	60.94	59.38	59.38	59.38	60.94	56.08
	Ortalama	56.77	56.25	55.21	60.42	59.64	59.11	60.16	58.85	59.38	58.42
Sensör Referans değeri - 3	<i>d3 & std</i>	64.06	67.19	62.50	59.38	67.19	65.63	64.06	64.06	65.63	64.41
	<i>d3 & pca</i>	62.50	64.06	65.63	71.88	70.31	70.31	65.63	65.63	67.19	67.01
	<i>fd3 & std</i>	59.38	60.94	59.38	64.06	68.75	64.06	64.06	68.75	65.63	63.89
	<i>fd3 & pca</i>	60.94	60.94	59.38	70.31	65.63	65.63	68.75	65.63	60.94	64.24
	<i>nfd3 & std</i>	56.25	48.44	45.31	57.81	50.00	50.00	60.94	48.44	50.00	51.91
	<i>nfd3 & pca</i>	59.38	51.56	51.56	65.63	60.94	62.50	60.94	60.94	62.50	59.55
	Ortalama	60.42	58.85	57.29	64.84	63.80	63.02	64.06	62.24	61.98	61.83

Bu tablo incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Veri setlerinin sınıflandırma sonuçlarının sensör referans değer noktalarına göre ortalamaları alındığında, sensör referans değeri-1’in daha başarılı olduğu görülmüştür.
- Sensör referans değeri-1 için en başarılı ön-işleme ve boyut azaltım yönteminin, ortalamalar dikkate alındığında, *d1&pca* olduğu görülmüştür.
- Ölçeklemenin sınıflandırma başarısını etkilediği görülmüştür.
- En yüksek sınıflandırma başarı oranı; [0-1] ölçekleme yöntemine göre, *d1&pca* yöntemi ve *k* = 1 değeri için %73.44 olarak elde edilmiştir.

2.2.2. DA ile Sınıflandırma

DA yöntemi kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.2’de görülmektedir. Boyut azaltım yöntemi olarak sadece PCA uygulanan veri setlerinin sınıflandırılması gerçekleştirilmiştir. Standart sapmaya göre boyut azaltımı yapılan veri setlerinin sınıflandırılması ise yapılmamıştır. Bunun nedeni; standart sapmaya göre boyut azaltımı yapılan bazı veri setlerin, grup (sınıflara ait) kovaryans matrislerinin pozitif tanımlı olmamasından dolayı, ikinci dereceden ve Mahalanobis fonksiyonlarla sınıflandırılmamasından kaynaklanmaktadır. Ölçekleme yapılmış ve yapılmamış veri setlerinin sınıflandırma sonuçları aynı olduğu için, burada ayrıca verilmemiştir. Bu tablodaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.2. Veri setlerinin DA yöntemi kullanılarak 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Doğrusal	Diyagonal doğrusal	İkinci dereceden	Diyagonal ikinci dereceden	Mahalanobis	Ortalama
Sensör Referans değeri	<i>d1 & pca</i>	67.19	68.75	76.56	68.75	68.75	70.00
	<i>fd1 & pca</i>	65.63	60.94	73.44	68.75	68.75	67.50
	<i>nfd1 & pca</i>	60.94	59.38	60.94	62.50	57.81	60.31
	Ortalama	64.58	63.02	70.31	66.67	65.10	65.94
Sensör Referans değeri	<i>d2 & pca</i>	59.38	59.38	71.88	67.19	57.81	63.13
	<i>fd2 & pca</i>	60.94	57.81	67.19	67.19	56.25	61.88
	<i>nfd2 & pca</i>	54.69	54.69	64.06	60.94	64.06	59.69
	Ortalama	58.33	57.29	67.71	65.10	59.38	61.56
Sensör Referans değeri	<i>d3 & pca</i>	67.19	68.75	75.00	68.75	70.31	70.00
	<i>fd3 & pca</i>	65.63	62.50	73.44	68.75	68.75	67.81
	<i>nfd3 & pca</i>	57.81	56.25	59.38	60.94	57.81	58.44
	Ortalama	63.54	62.50	69.27	66.15	65.63	65.42

Bu tablo incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Tablo genelinde ortalamalar dikkate alındığında, ikinci dereceden fonksiyon kullanılarak elde edilen sonuçlar daha başarılıdır.

- Veri setlerinin sınıflandırma sonuçlarının sensör referans değer noktalarına göre ortalamaları alındığında, sensör referans değeri-1'in daha başarılı olduğu görülmüştür.
- Sensör referans değeri-1 için en başarılı ön-işleme ve boyut azaltım yönteminin, ortalamalar dikkate alındığında, *d1&pca* olduğu görülmüştür.
- En yüksek sınıflandırma başarı oranı; ikinci dereceden fonksiyon kullanılarak, *d1&pca* yöntemi ile %76.56 olarak elde edilmiştir.

2.2.3. SVM ile Sınıflandırma

MATLAB'deki yerleşik SVM fonksiyonları kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.3'de görülmektedir. Ölçekleme yapılmış ve yapılmamış veri setlerinin sınıflandırma sonuçları aynı olduğu için, burada ayrıca verilmemiştir. Bu tablodaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.3. Veri setlerinin MATLAB'deki yerleşik SVM fonksiyonları ile 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	SVM (1-r)				SVM (1-1)				Ortalama
		Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	
Sensör Referans değeri - 1	<i>d1 & std</i>	29.84	55.70	62.30	35.97	77.74	72.47	66.53	69.69	58.78
	<i>d1 & pca</i>	44.05	64.71	63.31	52.37	79.77	69.74	71.76	68.53	64.28
	<i>fd1 & std</i>	27.12	50.59	57.31	30.36	74.41	74.01	64.15	68.03	55.75
	<i>fd1 & pca</i>	38.53	63.80	63.46	46.70	75.96	69.89	64.43	66.01	61.10
	<i>nfd1 & std</i>	34.75	42.88	52.01	36.63	59.69	66.16	58.68	57.86	51.08
	<i>nfd1 & pca</i>	38.37	55.06	59.85	45.87	60.37	64.26	64.09	66.67	56.82
	Ortalama	35.44	55.46	59.71	41.32	71.32	69.42	64.94	66.13	57.97
Sensör Referans değeri - 2	<i>d2 & std</i>	24.51	53.18	58.04	30.28	76.50	72.55	68.80	73.95	57.23
	<i>d2 & pca</i>	32.26	58.93	59.64	42.08	70.80	65.40	63.50	64.30	57.11
	<i>fd2 & std</i>	24.51	52.69	56.04	27.59	74.96	74.75	69.44	73.95	56.74
	<i>fd2 & pca</i>	33.37	57.49	59.51	41.38	74.75	65.28	61.37	64.43	57.20
	<i>nfd2 & std</i>	34.75	42.88	49.69	33.94	59.93	63.71	57.38	55.21	49.69
	<i>nfd2 & pca</i>	38.37	58.53	59.80	44.63	62.62	64.16	62.67	57.99	56.10
	Ortalama	31.30	53.95	57.12	36.65	69.93	67.64	63.86	64.97	55.68
Sensör Referans değeri - 3	<i>d3 & std</i>	29.84	55.70	62.30	34.63	77.74	71.04	65.63	69.69	58.32
	<i>d3 & pca</i>	43.85	66.94	64.67	51.33	78.10	70.02	71.76	67.00	64.21
	<i>fd3 & std</i>	28.37	50.59	57.31	30.36	74.41	72.58	64.15	68.03	55.72
	<i>fd3 & pca</i>	38.53	63.80	62.62	48.16	73.82	70.15	65.15	67.68	61.24
	<i>nfd3 & std</i>	34.75	42.69	51.23	35.29	59.69	64.50	57.01	59.29	50.55
	<i>nfd3 & pca</i>	38.18	58.48	59.80	45.87	61.56	62.28	59.22	65.95	56.42
	Ortalama	35.59	56.37	59.65	40.94	70.89	68.43	63.82	66.27	57.74

Bu tablo incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- SVM (1-1), SVM (1-r)'ye göre daha başarılı sonuçlar vermiştir.

- En başarılı sınıflandırma fonksiyonu, ortalamalar dikkate alındığında,
 - SVM (1-r) için üçüncü dereceden polinom,
 - SVM (1-1) için doğrusal fonksiyon
 olarak elde edilmiştir.
- Veri setlerinin sınıflandırma sonuçlarının sensör referans değer noktalarına göre ortalamaları alındığında, sensör referans değeri-1'in daha başarılı olduğu görülmüştür.
- En yüksek sınıflandırma başarı oranı; SVM (1-1)'de doğrusal fonksiyon kullanılarak *d1&pca* yöntemi ile %79.77 olarak elde edilmiştir.

LIBSVM kütüphanesindeki yerleşik SVM fonksiyonları kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırma başarı oranları hesaplanmıştır. Kullanılan 4 farklı çekirdek fonksiyonu için elde edilen sonuçlar, Tablo 2.4'de görülmektedir. Bu tablolardaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.4. Veri setlerinin LIBSVM kütüphanesindeki çekirdek fonksiyonları ile 8 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Çekirdek fonksiyonları: (a) doğrusal, (b) ikinci dereceden polinom, (c) üçüncü dereceden polinom, (d) RBF.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz		Ölçekleme [0-1]		Ölçekleme [z-skor]		Doğrusal
		$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	13.25	76.56	11.25	73.44	7.00	75.00	75.00
	<i>d1 & pca</i>	15.75	79.69	17.00	78.13	17.00	78.13	78.65
	<i>fd1 & std</i>	11.00	73.44	11.25	73.44	17.00	73.44	73.44
	<i>fd1 & pca</i>	11.25	73.44	5.00	75.00	11.00	71.88	73.44
	<i>nfd1 & std</i>	11.50	67.19	10.00	67.19	6.50	67.19	67.19
	<i>nfd1 & pca</i>	10.25	70.31	6.50	71.88	3.00	71.88	71.35
	Ortalama		73.44		73.18		72.92	73.18
Sensör Referans değeri - 2	<i>d2 & std</i>	15.25	75.00	17.00	75.00	17.00	76.56	75.52
	<i>d2 & pca</i>	16.75	76.56	8.50	75.00	4.25	76.56	76.04
	<i>fd2 & std</i>	11.25	71.88	7.25	78.13	3.00	76.56	75.52
	<i>fd2 & pca</i>	17.00	73.44	12.00	75.00	7.75	75.00	74.48
	<i>nfd2 & std</i>	17.00	62.50	17.00	62.50	3.25	64.06	63.02
	<i>nfd2 & pca</i>	7.00	65.63	5.75	67.19	1.25	68.75	67.19
	Ortalama		70.83		72.14		72.92	71.96
Sensör Referans değeri - 3	<i>d3 & std</i>	14.00	73.44	6.75	75.00	7.00	73.44	73.96
	<i>d3 & pca</i>	15.75	78.13	11.50	78.13	3.00	79.69	78.65
	<i>fd3 & std</i>	11.00	73.44	11.25	73.44	17.00	73.44	73.44
	<i>fd3 & pca</i>	15.50	73.44	4.25	78.13	10.25	71.88	74.48
	<i>nfd3 & std</i>	11.50	67.19	10.25	67.19	3.75	68.75	67.71
	<i>nfd3 & pca</i>	11.00	67.19	6.50	68.75	3.00	68.75	68.23
	Ortalama		72.14		73.44		72.66	72.74

(a)

(Tablo 2.4'ün devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			İkinci dereceden polinom
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	13.50	-1.00	79.69	11.00	-0.25	78.13	11.50	-5.50	79.69	79.17
	<i>d1 & pca</i>	12.75	2.75	81.25	17.00	-5.75	78.13	4.25	1.75	79.69	79.69
	<i>fd1 & std</i>	15.00	0.25	75.00	15.00	-2.75	75.00	10.75	-4.75	79.69	76.56
	<i>fd1 & pca</i>	14.25	0.75	78.13	17.00	-13.00	75.00	9.75	-1.00	75.00	76.04
	<i>nfd1 & std</i>	17.00	-6.75	70.31	17.00	-8.25	70.31	7.00	-3.75	71.88	70.83
	<i>nfd1 & pca</i>	7.75	0.50	71.88	17.00	-11.50	71.88	9.75	-7.50	75.00	72.92
	Ortalama			76.04			74.74			76.82	75.87
Sensör Referans değeri - 2	<i>d2 & std</i>	17.00	-2.00	78.13	15.00	-5.75	78.13	9.50	-5.00	79.69	78.65
	<i>d2 & pca</i>	16.75	-3.00	78.13	15.25	-6.00	76.56	17.00	-12.50	76.56	77.08
	<i>fd2 & std</i>	17.00	-1.50	78.13	13.75	-4.25	76.56	16.75	-14.75	76.56	77.08
	<i>fd2 & pca</i>	13.00	0.75	79.69	8.00	-1.75	75.00	17.00	-3.50	75.00	76.56
	<i>nfd2 & std</i>	13.00	-2.50	67.19	16.75	-1.75	67.19	6.25	-3.00	70.31	68.23
	<i>nfd2 & pca</i>	17.00	1.00	71.88	-1.00	2.25	68.75	16.50	-16.75	70.31	70.31
	Ortalama			75.52			73.70			74.74	74.65
Sensör Referans değeri - 3	<i>d3 & std</i>	13.00	2.75	76.56	11.00	-0.50	78.13	11.50	-5.50	79.69	78.13
	<i>d3 & pca</i>	17.00	-2.75	78.13	17.00	-6.25	78.13	3.00	1.75	81.25	79.17
	<i>fd3 & std</i>	15.00	0.25	76.56	8.75	-1.00	76.56	10.50	-4.50	79.69	77.60
	<i>fd3 & pca</i>	15.00	0.00	78.13	3.00	0.00	76.56	9.75	-1.00	75.00	76.56
	<i>nfd3 & std</i>	16.00	-5.50	70.31	13.25	-5.00	70.31	8.75	-5.25	70.31	70.31
	<i>nfd3 & pca</i>	6.75	1.00	71.88	17.00	-11.75	70.31	8.75	-6.50	73.44	71.88
	Ortalama			75.26			75.00			76.56	75.61

(b)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			Üçüncü dereceden polinom
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	12.75	-1.00	79.69	12.75	-2.75	78.13	12.50	-5.25	79.69	79.17
	<i>d1 & pca</i>	14.50	-2.50	81.25	17.00	-6.75	78.13	17.00	-15.00	79.69	79.69
	<i>fd1 & std</i>	14.00	0.00	78.13	12.25	-2.25	78.13	10.75	-5.50	79.69	78.65
	<i>fd1 & pca</i>	14.75	-0.50	78.13	17.00	-7.00	73.44	7.50	-9.00	75.00	75.52
	<i>nfd1 & std</i>	-2.25	2.75	73.44	17.00	-8.75	70.31	15.00	-10.25	70.31	71.35
	<i>nfd1 & pca</i>	17.00	-8.25	70.31	16.75	-11.75	71.88	9.75	-8.25	73.44	71.88
	Ortalama			76.82			75.00			76.30	76.04
Sensör Referans değeri - 2	<i>d2 & std</i>	13.00	-0.75	79.69	17.00	-8.25	78.13	17.00	-12.50	78.13	78.65
	<i>d2 & pca</i>	11.00	1.75	78.13	17.00	-8.75	75.00	16.75	-12.00	76.56	76.56
	<i>fd2 & std</i>	15.00	-1.50	76.56	12.75	-7.00	78.13	14.00	-11.00	76.56	77.08
	<i>fd2 & pca</i>	11.00	1.75	79.69	17.00	-10.50	75.00	17.00	-14.00	73.44	76.04
	<i>nfd2 & std</i>	17.00	-5.50	67.19	2.75	0.25	68.75	3.00	-1.75	70.31	68.75
	<i>nfd2 & pca</i>	17.00	0.00	70.31	3.00	2.75	70.31	-1.00	-1.00	71.88	70.83
	Ortalama			75.26			74.22			74.48	74.65
Sensör Referans değeri - 3	<i>d3 & std</i>	15.00	0.00	76.56	12.75	-3.00	78.13	12.25	-5.00	79.69	78.13
	<i>d3 & pca</i>	9.25	4.25	81.25	7.00	-2.25	79.69	15.00	-6.00	79.69	80.21
	<i>fd3 & std</i>	14.50	-0.25	78.13	12.50	-2.50	76.56	10.75	-5.50	79.69	78.13
	<i>fd3 & pca</i>	15.25	-1.00	78.13	7.25	1.00	76.56	17.00	-1.50	73.44	76.04
	<i>nfd3 & std</i>	-1.75	2.50	71.88	-1.00	2.25	70.31	1.75	-1.25	70.31	70.83
	<i>nfd3 & pca</i>	7.50	-0.25	71.88	16.50	-11.75	70.31	9.75	-8.25	73.44	71.88
	Ortalama			76.30			75.26			76.04	75.87

(c)

(Tablo 2.4'ün devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			RBF
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	13.25	2.50	82.81	10.75	0.00	82.81	11.75	-4.25	82.81	82.81
	<i>d1 & pca</i>	15.75	0.50	81.25	10.25	-3.50	79.69	10.50	-8.00	79.69	80.21
	<i>fd1 & std</i>	13.00	1.00	76.56	11.00	-0.75	79.69	12.25	-6.00	79.69	78.65
	<i>fd1 & pca</i>	13.00	1.25	76.56	6.75	-2.75	78.13	15.50	-5.00	75.00	76.56
	<i>nfd1 & std</i>	9.25	-0.25	71.88	17.00	-8.25	70.31	5.00	-0.25	73.44	71.88
	<i>nfd1 & pca</i>	7.75	2.50	71.88	17.00	-11.50	71.88	17.00	-14.75	71.88	71.88
	Ortalama			76.82			77.08			77.08	77.00
Sensör Referans değeri - 2	<i>d2 & std</i>	11.25	3.00	81.25	14.00	-6.00	79.69	12.00	-7.00	79.69	80.21
	<i>d2 & pca</i>	16.00	2.25	78.13	14.00	-5.75	76.56	17.00	-11.50	76.56	77.08
	<i>fd2 & std</i>	14.00	0.25	78.13	13.25	-7.00	78.13	17.00	-15.00	76.56	77.60
	<i>fd2 & pca</i>	14.75	-1.00	79.69	11.00	-4.00	75.00	9.00	-7.25	75.00	76.56
	<i>nfd2 & std</i>	17.00	-5.50	67.19	4.75	3.25	70.31	15.00	-4.00	67.19	68.23
	<i>nfd2 & pca</i>	17.00	0.25	68.75	5.00	2.50	71.88	4.75	-1.25	71.88	70.83
	Ortalama			75.52			75.26			74.48	75.09
Sensör Referans değeri - 3	<i>d3 & std</i>	13.25	2.50	81.25	10.75	0.00	82.81	12.25	-6.00	81.25	81.77
	<i>d3 & pca</i>	11.75	3.25	81.25	7.75	2.75	81.25	17.00	-15.00	79.69	80.73
	<i>fd3 & std</i>	12.75	1.00	78.13	11.75	-1.25	79.69	12.00	-5.75	79.69	79.17
	<i>fd3 & pca</i>	16.00	-1.00	76.56	6.75	-3.00	78.13	9.00	-2.50	75.00	76.56
	<i>nfd3 & std</i>	17.00	-6.75	70.31	7.75	-0.25	71.88	5.00	-0.25	73.44	71.88
	<i>nfd3 & pca</i>	7.75	1.00	71.88	9.50	-3.50	71.88	16.50	-14.50	70.31	71.35
	Ortalama			76.56			77.60			76.56	76.91

(d)

Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Tüm fonksiyonlarda, sensör referans değeri-1 ve sensör referans değeri-3'ün başarılı olduğu görülmüştür.
- Tüm fonksiyonlarda, en başarılı sonuçların çoğunun PCA yöntemi ile sağlandığı görülmüştür.
- Ölçeklemenin sınıflandırma başarısını etkilediği görülmüştür.
- Fonksiyonlara göre en yüksek sınıflandırma başarı oranları;
 - Doğrusal fonksiyon için; *d1&pca* ve *d3&pca* yöntemlerinde %79.69
 - İkinci dereceden polinom için; *d1&pca* ve *d3&pca* yöntemlerinde %81.25
 - Üçüncü dereceden polinom için; *d1&pca* ve *d3&pca* yöntemlerinde %81.25
 - RBF fonksiyonu için; *d1&std* ve *d3&std* yöntemlerinde %82.81
 olarak elde edilmiştir.

MATLAB ve LIBSVM kütüphanesi kullanılarak gerçekleştirilen 5-kat CV metoduna göre SVM ile sınıflandırmada, çekirdek fonksiyonları için hesaplanan en yüksek başarı oranları Tablo 2.5'de verilmiştir. Bu tablodaki koyu değerler, kullanılan yöntemlere göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.5. Veri setlerinin 5-kat CV metoduna göre SVM ile sınıflandırılmasında, çekirdek fonksiyonları için hesaplanan en yüksek başarı oranları (%).

	MATLAB		LIBSVM		
	SVM (1-r)	SVM (1-1)	Ölçeklemesiz	Ölçekleme [0 1]	Ölçekleme [z-skor]
Doğrusal	44.05	79.77	79.69	78.13	79.69
İkinci dereceden polinom	66.94	74.75	81.25	78.13	81.25
Üçüncü dereceden polinom	64.67	71.76	81.25	79.69	79.69
RBF	52.37	73.95	82.81	82.81	82.81

Bu tablodan, LIBSVM ile gerçekleştirilen sınıflandırma uygulamalarının MATLAB ortamında hazırlananlardan daha başarılı olduğu görülmektedir. Bunun nedeni; LIBSVM ortamında hazırlanan programların, her veri seti için çekirdek fonksiyonlarına göre en başarılı sınıflandırma sonucunu sağlayan (C, γ) çiftini belirlemesinden dolayıdır.

2.2.4. Sınıflandırma Yöntemlerinin Sonuçlarının Karşılaştırılması

Kullanılan sınıflandırma yöntemlerine göre 5-kat CV metodu ile hesaplanan en yüksek başarı oranları ve bu oranları sağlayan alt metot, ölçekleme, ön-işleme ve boyut azaltım yöntemleri Tablo 2.6'da verilmiştir. Burada sınıflandırma yöntemleri arasında en yüksek başarıyı; RBF çekirdek fonksiyonu kullanılarak %82.81 ile LIBSVM yazılımının sağladığı görülmüştür. Tablo genelinde, $d1$ ön-işleme yöntemi başarılı sonuçları sağlamıştır. Bu yöntem; sensör cevap modeli olarak fark modelini $(x_i = R_i^{\max} - R_i^0 = \Delta R_i)$ ve sensör referans değeri-1'i $(R_i^0 = R_i^{\min})$ temsil etmektedir. Ayrıca LIBSVM hariç diğer tüm sınıflandırıcılarda, boyut azaltımında PCA yönteminin en başarılı sonuçları verdiği görülmüştür.

Tablo 2.6. Mikroorganizma seti-1 için sınıflandırma yöntemlerine göre 5-kat CV metodu ile elde edilen en yüksek başarı oranları ve bu oranları sağlayan çeşitli yöntemler.

Yöntem	k-NN	DA	MATLAB SVM	LIBSVM
Ön-işleme *	<i>d1</i>			<i>d1, d3</i>
Boyut azaltım **	PCA			<i>Std</i>
Ölçekleme	[0-1] ölçekleme	-		Ölçeklemesiz [0-1] ölçekleme z-skor ölçekleme
Metot	$k = 1$	İkinci dereceden fonksiyon	SVM (1-1) ve doğrusal fonksiyon	RBF
Sınıflandırma başarıları (%)	73.44	76.56	79.77	82.81

(*) Ön-işleme kısaltmaları; sensör cevap modeli ve sensör referans değer noktasını tanımlamaktadır. Eşitlik ifadeleri için Tablo 1.6'ya bakınız.

(**) Boyut azaltım kısmında; *Std*, standart sapmaya göre en fazla değişim gösteren 3 sensörün belirlenmesini ifade etmektedir.

2.3. Mikroorganizma Seti-2 için Sınıflandırma Sonuçları

Mikroorganizma Seti-2; dış kök kanallarında enfeksiyonlara sebep olan 7 mikroorganizmadan meydana gelmektedir. Bu mikroorganizma seti üzerinde elektronik burun cihazının kullanımı araştırılmıştır [61]. Her mikroorganizmanın 4 farklı konsantrasyonda (12×10^8 , 12×10^5 , 12×10^3 ve 12×10^1 CFU/mL) steril cam tüplerde süspansiyonları hazırlanmış ve her birinden 5 koku verisi alınmıştır. Dolayısıyla her bir konsantrasyon için toplamda 35 koku verisi üzerinde çalışılmıştır. Her konsantrasyon için, farklı yöntemlere göre elde edilen sınıflandırma sonuçları aşağıdaki kısımlarda açıklanmıştır.

2.3.1. k-NN Algoritması ile Sınıflandırma

Sabit konsantrasyondaki mikroorganizma örneklerinden elde edilen veri setlerinin, k-NN algoritması kullanılarak, 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.7'de görülmektedir. Bu tablolardaki koyu değerler, k değerlerine göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.7. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, k -NN algoritması kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ⁸
		k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	Ortalama
Sensör Referans değeri - 1	d1 & std	54.29	48.57	54.29	51.43	51.43	51.43	51.43	51.43	51.43	51.75
	d1 & pca	60.00	54.29	51.43	88.57	82.86	74.29	88.57	80.00	74.29	72.70
	fd1 & std	51.43	54.29	54.29	65.71	62.86	54.29	62.86	57.14	60.00	58.10
	fd1 & pca	57.14	60.00	51.43	91.43	82.86	77.14	91.43	82.86	77.14	74.60
	nfd1 & std	85.71	77.14	71.43	82.86	77.14	68.57	82.86	77.14	68.57	76.83
	nfd1 & pca	82.86	77.14	68.57	85.71	85.71	77.14	82.86	82.86	74.29	79.68
	Ortalama	65.24	61.90	58.57	77.62	73.81	67.14	76.67	71.90	67.62	68.94
Sensör Referans değeri - 2	d2 & std	48.57	45.71	54.29	45.71	45.71	45.71	45.71	45.71	42.86	46.67
	d2 & pca	48.57	40.00	51.43	71.43	65.71	62.86	71.43	62.86	65.71	60.00
	fd2 & std	54.29	51.43	54.29	60.00	51.43	54.29	60.00	48.57	48.57	53.65
	fd2 & pca	51.43	51.43	54.29	68.57	71.43	71.43	68.57	71.43	71.43	64.44
	nfd2 & std	71.43	65.71	60.00	62.86	65.71	60.00	62.86	68.57	60.00	64.13
	nfd2 & pca	65.71	65.71	54.29	74.29	77.14	80.00	74.29	77.14	77.14	71.75
	Ortalama	56.67	53.33	54.76	63.81	62.86	62.38	63.81	62.38	60.95	60.11
Sensör Referans değeri - 3	d3 & std	54.29	48.57	54.29	60.00	57.14	51.43	57.14	54.29	51.43	54.29
	d3 & pca	57.14	48.57	54.29	82.86	82.86	74.29	82.86	80.00	71.43	70.48
	fd3 & std	54.29	57.14	54.29	68.57	62.86	57.14	68.57	65.71	57.14	60.63
	fd3 & pca	54.29	57.14	54.29	88.57	80.00	80.00	88.57	80.00	77.14	73.33
	nfd3 & std	85.71	74.29	74.29	82.86	77.14	68.57	82.86	77.14	68.57	76.83
	nfd3 & pca	80.00	77.14	68.57	88.57	85.71	80.00	85.71	82.86	74.29	80.32
	Ortalama	64.29	60.48	60.00	78.57	74.29	68.57	77.62	73.33	66.67	69.31

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ⁵
		k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	Ortalama
Sensör Referans değeri - 1	d1 & std	54.29	54.29	51.43	51.43	51.43	51.43	54.29	51.43	54.29	52.70
	d1 & pca	54.29	51.43	51.43	80.00	74.29	68.57	80.00	82.86	80.00	69.21
	fd1 & std	51.43	54.29	51.43	54.29	57.14	57.14	54.29	57.14	57.14	54.92
	fd1 & pca	51.43	54.29	54.29	82.86	68.57	74.29	82.86	71.43	77.14	68.57
	nfd1 & std	80.00	74.29	68.57	80.00	80.00	74.29	80.00	80.00	74.29	76.83
	nfd1 & pca	77.14	74.29	71.43	91.43	91.43	91.43	88.57	88.57	85.71	84.44
	Ortalama	61.43	60.48	58.10	73.33	70.48	69.52	73.33	71.90	71.43	67.78
Sensör Referans değeri - 2	d2 & std	54.29	54.29	54.29	54.29	54.29	62.86	54.29	54.29	62.86	56.19
	d2 & pca	57.14	57.14	51.43	74.29	74.29	68.57	71.43	74.29	71.43	66.67
	fd2 & std	54.29	54.29	48.57	54.29	57.14	57.14	54.29	57.14	57.14	54.92
	fd2 & pca	54.29	54.29	48.57	80.00	74.29	77.14	80.00	74.29	77.14	68.89
	nfd2 & std	62.86	68.57	65.71	65.71	74.29	68.57	71.43	74.29	68.57	68.89
	nfd2 & pca	71.43	68.57	68.57	85.71	80.00	82.86	85.71	82.86	82.86	78.73
	Ortalama	59.05	59.52	56.19	69.05	69.05	69.52	69.52	69.52	70.00	65.71
Sensör Referans değeri - 3	d3 & std	54.29	54.29	57.14	60.00	60.00	62.86	60.00	60.00	62.86	59.05
	d3 & pca	51.43	51.43	54.29	82.86	80.00	80.00	82.86	82.86	91.43	73.02
	fd3 & std	54.29	54.29	54.29	54.29	54.29	57.14	54.29	54.29	57.14	54.92
	fd3 & pca	51.43	54.29	54.29	82.86	77.14	77.14	82.86	82.86	82.86	71.75
	nfd3 & std	80.00	68.57	71.43	82.86	77.14	77.14	80.00	74.29	74.29	76.19
	nfd3 & pca	74.29	71.43	71.43	88.57	94.29	91.43	88.57	94.29	85.71	84.44
	Ortalama	60.95	59.05	60.48	75.24	73.81	74.29	74.76	74.76	75.71	69.89

(b)

(Tablo 2.7'nin devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ³
		k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	
Sensör Referans değeri - 1	d1 & std	25.71	25.71	25.71	28.57	28.57	28.57	28.57	28.57	28.57	27.62
	d1 & pca	22.86	22.86	28.57	62.86	57.14	54.29	65.71	60.00	51.43	47.30
	fd1 & std	25.71	28.57	25.71	31.43	34.29	37.14	31.43	34.29	37.14	31.75
	fd1 & pca	22.86	22.86	25.71	65.71	62.86	62.86	65.71	65.71	68.57	51.43
	nfd1 & std	34.29	31.43	37.14	37.14	37.14	40.00	40.00	37.14	37.14	36.83
	nfd1 & pca	37.14	45.71	37.14	60.00	60.00	65.71	65.71	62.86	68.57	55.87
	Ortalama	28.10	29.52	30.00	47.62	46.67	48.10	49.52	48.10	48.57	41.80
Sensör Referans değeri - 2	d2 & std	28.57	28.57	34.29	25.71	28.57	34.29	25.71	28.57	31.43	29.52
	d2 & pca	31.43	34.29	40.00	48.57	54.29	57.14	51.43	51.43	60.00	47.62
	fd2 & std	25.71	25.71	34.29	31.43	34.29	37.14	31.43	34.29	37.14	32.38
	fd2 & pca	34.29	34.29	40.00	60.00	57.14	51.43	57.14	57.14	57.14	49.84
	nfd2 & std	40.00	48.57	42.86	40.00	54.29	42.86	40.00	51.43	42.86	44.76
	nfd2 & pca	45.71	48.57	45.71	65.71	60.00	57.14	65.71	62.86	54.29	56.19
	Ortalama	34.29	36.67	39.52	45.24	48.10	46.67	45.24	47.62	47.14	43.39
Sensör Referans değeri - 3	d3 & std	22.86	22.86	20.00	22.86	25.71	20.00	22.86	25.71	20.00	22.54
	d3 & pca	25.71	25.71	22.86	68.57	68.57	62.86	71.43	71.43	65.71	53.65
	fd3 & std	22.86	22.86	22.86	34.29	34.29	28.57	34.29	34.29	28.57	29.21
	fd3 & pca	22.86	20.00	22.86	68.57	71.43	71.43	71.43	74.29	71.43	54.92
	nfd3 & std	37.14	48.57	45.71	42.86	48.57	45.71	42.86	48.57	45.71	45.08
	nfd3 & pca	54.29	60.00	48.57	71.43	82.86	80.00	68.57	80.00	77.14	69.21
	Ortalama	30.95	33.33	30.48	51.43	55.24	51.43	51.90	55.71	51.43	45.77

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ¹
		k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	k = 1	k = 3	k = 5	
Sensör Referans değeri - 1	d1 & std	28.57	37.14	37.14	34.29	42.86	42.86	34.29	42.86	42.86	38.10
	d1 & pca	34.29	42.86	40.00	80.00	74.29	68.57	82.86	77.14	74.29	63.81
	fd1 & std	28.57	34.29	34.29	34.29	40.00	42.86	34.29	40.00	42.86	36.83
	fd1 & pca	34.29	37.14	40.00	80.00	80.00	60.00	82.86	80.00	65.71	62.22
	nfd1 & std	62.86	62.86	60.00	68.57	68.57	60.00	68.57	68.57	65.71	65.08
	nfd1 & pca	60.00	57.14	48.57	77.14	80.00	71.43	74.29	80.00	68.57	68.57
	Ortalama	41.43	45.24	43.33	62.38	64.29	57.62	62.86	64.76	60.00	55.77
Sensör Referans değeri - 2	d2 & std	25.71	28.57	34.29	31.43	37.14	40.00	31.43	37.14	42.86	34.29
	d2 & pca	28.57	34.29	31.43	54.29	60.00	48.57	54.29	57.14	54.29	46.98
	fd2 & std	20.00	31.43	31.43	34.29	34.29	37.14	34.29	34.29	37.14	32.70
	fd2 & pca	22.86	34.29	31.43	68.57	68.57	74.29	68.57	74.29	71.43	57.14
	nfd2 & std	60.00	57.14	57.14	65.71	60.00	62.86	65.71	60.00	62.86	61.27
	nfd2 & pca	60.00	60.00	54.29	77.14	80.00	80.00	74.29	82.86	82.86	72.38
	Ortalama	36.19	40.95	40.00	55.24	56.67	57.14	54.76	57.62	58.57	50.79
Sensör Referans değeri - 3	d3 & std	25.71	25.71	37.14	28.57	31.43	34.29	28.57	31.43	37.14	31.11
	d3 & pca	34.29	40.00	40.00	62.86	68.57	51.43	62.86	62.86	62.86	53.97
	fd3 & std	28.57	22.86	34.29	28.57	34.29	37.14	28.57	34.29	37.14	31.75
	fd3 & pca	34.29	31.43	34.29	74.29	82.86	80.00	77.14	82.86	77.14	63.81
	nfd3 & std	62.86	57.14	57.14	62.86	60.00	57.14	65.71	60.00	57.14	60.00
	nfd3 & pca	62.86	57.14	54.29	85.71	80.00	85.71	85.71	82.86	85.71	75.56
	Ortalama	41.43	39.05	42.86	57.14	59.52	57.62	58.10	59.05	59.52	52.70

(d)

Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Ölçeklemenin sınıflandırma başarısını etkilediği görülmüştür.
- Tüm konsantrasyonlar için, sensör referans değeri-1 ve sensör referans değeri-3'ün başarılı olduğu görülmüştür.
- En yüksek sınıflandırma başarı oranları;

- 12×10^8 CFU/mL konsantrasyonda, [0-1] ve z-skor ölçekleme ile $k = 1$ değeri kullanılarak %91.43,
- 12×10^5 CFU/mL konsantrasyonda, [0-1] ve z-skor ölçekleme ile $k = 3$ değeri kullanılarak %94.29,
- 12×10^3 CFU/mL konsantrasyonda, [0-1] ölçekleme ile $k = 3$ değeri kullanılarak %82.86,
- 12×10^1 CFU/mL konsantrasyonda, [0-1] ve z-skor ölçekleme ile $k = 1$ ve $k = 5$ değerleri kullanılarak %85.71 olarak elde edilmiştir.

2.3.2. DA ile Sınıflandırma

Sabit konsantrasyondaki mikroorganizma örneklerinden elde edilen veri setlerinin, DA yöntemi kullanılarak, 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.8’de görülmektedir. Ölçekleme yapılmış ve yapılmamış veri setlerinin sınıflandırma sonuçları aynı olduğu için, burada ayrıca verilmemiştir. Bu tablolardaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.8. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, DA yöntemi kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Doğrusal	Diyagonal doğrusal	İkinci dereceden	Diyagonal ikinci dereceden	Mahalanobis	12×10^8 Ortalama
Sensör Referans değeri - 1	d1 & std	68.57	51.43	82.86	51.43	80.00	66.86
	d1 & pca	82.86	80.00	71.43	68.57	71.43	74.86
	fd1 & std	71.43	57.14	74.29	57.14	77.14	67.43
	fd1 & pca	77.14	71.43	77.14	74.29	77.14	75.43
	nfd1 & std	82.86	74.29	68.57	85.71	68.57	76.00
	nfd1 & pca	85.71	77.14	74.29	80.00	74.29	78.29
	Ortalama	78.10	68.57	74.76	69.52	74.76	73.14
Sensör Referans değeri - 2	d2 & std	51.43	37.14	68.57	37.14	68.57	52.57
	d2 & pca	54.29	40.00	74.29	45.71	74.29	57.71
	fd2 & std	65.71	42.86	65.71	40.00	62.86	55.43
	fd2 & pca	71.43	51.43	68.57	57.14	68.57	63.43
	nfd2 & std	80.00	60.00	60.00	65.71	65.71	66.29
	nfd2 & pca	82.86	74.29	68.57	68.57	65.71	72.00
	Ortalama	67.62	50.95	67.62	52.38	67.62	61.24
Sensör Referans değeri - 3	d3 & std	62.86	51.43	85.71	48.57	85.71	66.86
	d3 & pca	80.00	77.14	80.00	65.71	80.00	76.57
	fd3 & std	77.14	57.14	74.29	57.14	68.57	66.86
	fd3 & pca	71.43	71.43	74.29	74.29	74.29	73.14
	nfd3 & std	82.86	74.29	74.29	80.00	71.43	76.57
	nfd3 & pca	85.71	77.14	80.00	80.00	77.14	80.00
	Ortalama	76.67	68.10	78.10	67.62	76.19	73.33

(a)

(Tablo 2.8'in devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Doğrusal	Diyagonal doğrusal	İkinci dereceden	Diyagonal ikinci dereceden	Mahalanobis	12x10 ⁵ Ortalama
Sensör Referans değeri - 1	d1 & std	62.86	48.57	68.57	42.86	65.71	57.71
	d1 & pca	62.86	62.86	60.00	54.29	57.14	59.43
	fd1 & std	68.57	51.43	57.14	42.86	60.00	56.00
	fd1 & pca	71.43	68.57	65.71	48.57	65.71	64.00
	nfd1 & std	100.00	62.86	68.57	60.00	68.57	72.00
	nfd1 & pca	97.14	97.14	71.43	80.00	71.43	83.43
	Ortalama	77.14	65.24	65.24	54.76	64.76	65.43
Sensör Referans değeri - 2	d2 & std	51.43	51.43	45.71	51.43	45.71	49.14
	d2 & pca	77.14	65.71	57.14	62.86	54.29	63.43
	fd2 & std	68.57	62.86	48.57	57.14	45.71	56.57
	fd2 & pca	80.00	65.71	60.00	60.00	60.00	65.14
	nfd2 & std	80.00	57.14	54.29	51.43	57.14	60.00
	nfd2 & pca	85.71	88.57	77.14	74.29	77.14	80.57
	Ortalama	73.81	65.24	57.14	59.52	56.67	62.48
Sensör Referans değeri - 3	d3 & std	62.86	48.57	57.14	45.71	57.14	54.29
	d3 & pca	62.86	68.57	60.00	48.57	57.14	59.43
	fd3 & std	71.43	51.43	71.43	51.43	68.57	62.86
	fd3 & pca	68.57	65.71	77.14	54.29	77.14	68.57
	nfd3 & std	97.14	62.86	62.86	60.00	62.86	69.14
	nfd3 & pca	85.71	88.57	62.86	77.14	62.86	75.43
	Ortalama	74.76	64.29	65.24	56.19	64.29	64.95

(b)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Doğrusal	Diyagonal doğrusal	İkinci dereceden	Diyagonal ikinci dereceden	Mahalanobis	12x10 ³ Ortalama
Sensör Referans değeri - 1	d1 & std	54.29	25.71	42.86	17.14	34.29	34.86
	d1 & pca	57.14	42.86	48.57	57.14	42.86	49.71
	fd1 & std	57.14	25.71	57.14	20.00	62.86	44.57
	fd1 & pca	57.14	60.00	54.29	62.86	45.71	56.00
	nfd1 & std	68.57	42.86	45.71	31.43	37.14	45.14
	nfd1 & pca	65.71	65.71	40.00	57.14	42.86	54.29
	Ortalama	60.00	43.81	48.10	40.95	44.29	47.43
Sensör Referans değeri - 2	d2 & std	51.43	28.57	48.57	22.86	51.43	40.57
	d2 & pca	54.29	54.29	34.29	51.43	34.29	45.71
	fd2 & std	40.00	28.57	31.43	28.57	31.43	32.00
	fd2 & pca	54.29	54.29	37.14	51.43	37.14	46.86
	nfd2 & std	57.14	40.00	48.57	34.29	48.57	45.71
	nfd2 & pca	62.86	60.00	40.00	57.14	34.29	50.86
	Ortalama	53.33	44.29	40.00	40.95	39.52	43.62
Sensör Referans değeri - 3	d3 & std	54.29	25.71	42.86	20.00	37.14	36.00
	d3 & pca	62.86	57.14	65.71	57.14	60.00	60.57
	fd3 & std	60.00	25.71	51.43	20.00	51.43	41.71
	fd3 & pca	68.57	68.57	51.43	62.86	48.57	60.00
	nfd3 & std	71.43	42.86	45.71	40.00	48.57	49.71
	nfd3 & pca	77.14	71.43	51.43	62.86	48.57	62.29
	Ortalama	65.71	48.57	51.43	43.81	49.05	51.71

(c)

(Tablo 2.8'in devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Doğrusal	Diyagonal doğrusal	İkinci dereceden	Diyagonal ikinci dereceden	Mahalanobis	12x10 ¹ Ortalama
Sensör Referans değeri - 1	d1 & std	57.14	34.29	57.14	31.43	51.43	46.29
	d1 & pca	65.71	68.57	51.43	57.14	51.43	58.86
	fd1 & std	80.00	45.71	65.71	34.29	65.71	58.29
	fd1 & pca	74.29	71.43	80.00	65.71	80.00	74.29
	nfd1 & std	88.57	62.86	65.71	48.57	68.57	66.86
	nfd1 & pca	85.71	88.57	65.71	71.43	62.86	74.86
	Ortalama	75.24	61.90	64.29	51.43	63.33	63.24
Sensör Referans değeri - 2	d2 & std	57.14	40.00	37.14	37.14	40.00	42.29
	d2 & pca	51.43	51.43	48.57	37.14	45.71	46.86
	fd2 & std	74.29	31.43	48.57	34.29	51.43	48.00
	fd2 & pca	71.43	57.14	60.00	48.57	57.14	58.86
	nfd2 & std	77.14	40.00	54.29	42.86	51.43	53.14
	nfd2 & pca	80.00	74.29	65.71	68.57	62.86	70.29
	Ortalama	68.57	49.05	52.38	44.76	51.43	53.24
Sensör Referans değeri - 3	d3 & std	65.71	37.14	45.71	31.43	48.57	45.71
	d3 & pca	57.14	54.29	45.71	45.71	45.71	49.71
	fd3 & std	80.00	45.71	57.14	34.29	54.29	54.29
	fd3 & pca	80.00	85.71	54.29	65.71	57.14	68.57
	nfd3 & std	85.71	57.14	48.57	51.43	48.57	58.29
	nfd3 & pca	85.71	82.86	62.86	65.71	60.00	71.43
	Ortalama	75.71	60.48	52.38	49.05	52.38	58.00

(d)

Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Tüm konsantrasyonlar için, sensör referans değeri-1 ve sensör referans değeri-3'ün başarılı olduğu görülmüştür.
- En yüksek sınıflandırma başarı oranları;
 - 12x10⁸ CFU/mL konsantrasyonda, diyagonal doğrusal hariç tüm fonksiyonlarda %85.71,
 - 12x10⁵ CFU/mL konsantrasyonda, doğrusal fonksiyon kullanılarak %100,
 - 12x10³ CFU/mL konsantrasyonda, doğrusal fonksiyon kullanılarak %77.14,
 - 12x10¹ CFU/mL konsantrasyonda, doğrusal ve diyagonal doğrusal fonksiyonlar kullanılarak %88.57 olarak elde edilmiştir.

2.3.3. SVM ile Sınıflandırma

MATLAB'deki yerleşik SVM fonksiyonları kullanılarak, sabit konsantrasyondaki mikroorganizma örneklerinden elde edilen her bir veri setinin, 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.9'da görülmektedir. Ölçekleme yapılmış ve yapılmamış veri setlerinin sınıflandırma sonuçları aynı olduğu için, burada ayrıca

verilmemiştir. Bu tablolardaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.9. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, MATLAB'deki yerleşik SVM fonksiyonları ile 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	SVM (1-r)				SVM (1-1)				12x10 ⁸
		Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	
Sensör Referans değeri - 1	d1 & std	40.41	60.20	62.86	40.00	69.05	68.57	64.76	55.71	57.70
	d1 & pca	67.35	80.82	88.57	72.24	94.29	85.71	82.86	91.43	82.91
	fd1 & std	52.65	76.33	71.22	42.45	77.62	76.19	71.43	71.43	67.41
	fd1 & pca	64.90	81.63	90.00	72.65	79.52	82.86	80.00	78.57	78.77
	nfd1 & std	57.96	59.39	61.43	45.71	83.81	84.29	76.67	82.38	68.95
	nfd1 & pca	60.61	69.73	79.39	62.45	79.52	87.14	77.14	85.71	75.21
	Ortalama	57.31	71.35	75.58	55.92	80.63	80.79	75.48	77.54	71.83
Sensör Referans değeri - 2	d2 & std	29.80	36.94	44.29	26.94	43.81	58.57	58.57	48.57	43.44
	d2 & pca	41.63	65.31	65.03	55.10	78.57	74.29	77.14	71.43	66.06
	fd2 & std	36.12	46.53	60.41	33.06	65.71	68.57	70.00	62.86	55.41
	fd2 & pca	41.63	62.45	73.27	57.14	74.29	79.52	83.81	78.57	68.84
	nfd2 & std	37.96	48.78	47.35	43.27	74.29	71.43	70.00	71.43	58.06
	nfd2 & pca	49.39	72.65	72.18	54.69	73.81	80.00	65.71	78.57	68.38
	Ortalama	39.42	55.44	60.42	45.03	68.41	72.06	70.87	68.57	60.03
Sensör Referans değeri - 3	d3 & std	37.96	58.16	63.27	40.00	69.05	68.57	62.86	55.71	56.95
	d3 & pca	69.80	82.65	87.55	72.65	88.57	90.00	78.57	85.24	81.88
	fd3 & std	52.65	77.76	71.22	42.45	75.71	76.67	72.86	71.43	67.59
	fd3 & pca	64.90	85.10	86.12	74.69	79.52	85.71	81.43	81.43	79.86
	nfd3 & std	60.41	61.84	60.41	48.16	84.29	76.67	78.10	82.38	69.03
	nfd3 & pca	66.94	74.01	81.22	64.90	78.57	85.71	85.71	85.71	77.85
	Ortalama	58.78	73.25	74.97	57.14	79.29	80.56	76.59	76.98	72.19

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	SVM (1-r)				SVM (1-1)				12x10 ⁵
		Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	
Sensör Referans değeri - 1	d1 & std	26.12	39.18	55.85	34.29	62.86	58.10	51.43	57.14	48.12
	d1 & pca	56.05	59.52	79.12	55.92	78.19	65.71	82.86	71.43	68.60
	fd1 & std	26.12	34.69	53.78	37.96	60.95	59.52	53.81	57.14	48.00
	fd1 & pca	56.05	59.52	75.41	55.51	76.19	68.57	76.90	68.10	67.03
	nfd1 & std	50.61	66.33	67.89	58.37	82.86	77.14	75.71	80.95	69.98
	nfd1 & pca	67.76	80.61	76.94	75.51	94.29	77.14	88.57	84.29	80.64
	Ortalama	47.12	56.64	68.16	52.93	75.89	67.70	71.55	69.84	63.73
Sensör Referans değeri - 2	d2 & std	26.12	37.96	55.31	35.92	60.00	64.29	60.00	52.38	49.00
	d2 & pca	48.78	55.71	53.47	50.61	63.81	68.10	73.81	66.67	60.12
	fd2 & std	26.12	37.55	57.14	38.37	68.57	58.10	60.48	58.10	50.55
	fd2 & pca	50.20	52.79	55.78	51.02	75.71	71.43	67.14	60.71	60.60
	nfd2 & std	45.71	56.94	59.52	53.06	74.29	63.81	68.57	71.43	61.67
	nfd2 & pca	60.41	74.49	75.92	70.20	84.29	75.24	69.52	82.86	74.12
	Ortalama	42.89	52.57	59.52	49.86	71.11	66.83	66.59	65.36	59.34
Sensör Referans değeri - 3	d3 & std	26.12	37.76	55.44	36.73	62.86	60.00	52.86	57.14	48.61
	d3 & pca	58.50	62.45	74.22	53.47	81.43	69.52	74.76	66.67	67.63
	fd3 & std	26.12	34.69	56.22	37.96	60.95	61.43	58.10	57.14	49.08
	fd3 & pca	56.60	64.01	67.65	53.47	78.10	75.24	77.14	77.14	68.67
	nfd3 & std	50.61	63.47	63.95	55.51	85.71	74.29	76.67	75.24	68.18
	nfd3 & pca	60.82	79.18	78.37	75.51	87.14	79.52	84.29	80.00	78.10
	Ortalama	46.46	56.93	65.98	52.11	76.03	70.00	70.63	68.89	63.38

(b)

(Tablo 2.9'un devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	SVM (1-r)				SVM (1-1)				12x10 ³
		Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Ortalama
Sensör Referans değeri - 1	d1 & std	26.53	23.47	29.55	24.08	40.00	27.62	35.24	28.57	29.38
	d1 & pca	42.65	57.14	60.00	43.27	70.00	60.00	72.38	54.29	57.47
	fd1 & std	24.08	28.37	29.69	24.08	42.86	28.57	39.52	34.29	31.43
	fd1 & pca	45.10	57.69	59.12	42.45	72.38	65.71	66.19	60.00	58.58
	nfd1 & std	28.57	28.78	31.56	23.27	60.00	63.81	58.57	48.57	42.89
	nfd1 & pca	48.37	62.93	54.18	35.51	74.29	58.10	59.05	62.86	56.91
	Ortalama	35.88	43.06	44.02	32.11	59.92	50.63	55.16	48.10	46.11
Sensör Referans değeri - 2	d2 & std	24.69	27.89	29.18	23.27	38.57	28.57	37.14	40.00	31.16
	d2 & pca	36.33	40.54	47.55	35.10	46.67	51.43	50.00	44.29	43.99
	fd2 & std	23.27	30.20	26.73	23.27	43.81	38.57	41.43	28.57	31.98
	fd2 & pca	35.31	53.20	51.84	30.61	65.71	52.38	54.76	55.71	49.94
	nfd2 & std	28.57	27.76	33.33	28.57	57.14	57.14	65.71	54.29	44.06
	nfd2 & pca	35.31	54.52	53.13	37.55	72.86	55.24	60.95	63.81	54.17
	Ortalama	30.58	39.02	40.29	29.73	54.13	47.22	51.67	47.78	42.55
Sensör Referans değeri - 3	d3 & std	26.12	22.99	34.08	24.08	45.71	27.14	45.71	28.57	31.80
	d3 & pca	43.06	52.24	58.57	45.31	72.38	62.38	63.33	58.10	56.92
	fd3 & std	24.08	24.49	29.39	21.63	45.71	28.10	43.81	34.29	31.44
	fd3 & pca	43.47	58.98	60.14	45.71	74.29	72.38	62.86	65.71	60.44
	nfd3 & std	28.57	25.85	31.36	23.67	65.71	60.95	64.29	51.43	43.98
	nfd3 & pca	51.43	54.01	52.76	40.00	77.14	63.81	62.38	65.71	58.41
	Ortalama	36.12	39.76	44.38	33.40	63.49	52.46	57.06	50.63	47.16

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	SVM (1-r)				SVM (1-1)				12x10 ¹
		Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Ortalama
Sensör Referans değeri - 1	d1 & std	18.78	37.35	43.61	23.67	48.57	52.86	55.24	45.71	40.72
	d1 & pca	45.92	56.05	69.35	45.31	78.10	75.24	69.05	70.00	63.63
	fd1 & std	26.12	41.84	48.78	25.71	51.43	55.24	57.14	42.86	43.64
	fd1 & pca	52.24	66.67	76.77	55.10	80.95	80.00	78.10	71.43	70.16
	nfd1 & std	39.80	46.94	49.52	38.78	80.00	65.71	57.14	64.29	55.27
	nfd1 & pca	50.61	67.14	65.37	66.94	80.95	71.43	74.29	72.38	68.64
	Ortalama	38.91	52.66	58.90	42.59	70.00	66.75	65.16	61.11	57.01
Sensör Referans değeri - 2	d2 & std	23.88	37.96	40.41	28.16	46.67	40.95	40.48	31.43	36.24
	d2 & pca	32.45	34.73	42.34	29.80	51.43	52.38	59.52	54.29	44.62
	fd2 & std	22.24	40.82	48.78	26.53	48.10	43.81	49.52	38.57	39.80
	fd2 & pca	40.82	50.75	60.14	52.24	70.00	65.71	68.57	68.57	59.60
	nfd2 & std	35.92	30.20	36.33	31.02	80.00	63.81	70.95	68.57	52.10
	nfd2 & pca	47.35	72.59	63.40	67.76	78.57	72.86	78.10	74.29	69.36
	Ortalama	33.78	44.51	48.56	39.25	62.46	56.59	61.19	55.95	50.29
Sensör Referans değeri - 3	d3 & std	26.12	32.86	37.28	23.27	51.43	43.57	46.67	42.86	38.01
	d3 & pca	36.46	45.10	54.42	36.33	68.10	54.29	64.29	57.14	52.02
	fd3 & std	26.12	41.84	45.92	26.12	51.43	47.62	55.71	40.00	41.85
	fd3 & pca	48.78	69.25	75.44	59.59	77.14	68.57	68.57	60.95	66.04
	nfd3 & std	36.94	44.90	51.84	36.33	77.14	70.00	74.29	68.57	57.50
	nfd3 & pca	54.29	76.33	76.12	66.94	88.10	74.29	77.14	78.57	73.97
	Ortalama	38.12	51.71	56.84	41.43	68.89	59.72	64.44	58.02	54.90

(d)

Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Tüm konsantrasyonlar için, sensör referans değeri-1 ve sensör referans değeri-3'ün başarılı olduğu görülmüştür.

- Tüm konsantrasyonlarda SVM (1-1), SVM (1-r)'ye göre daha başarılı sonuçlar vermiştir.
- Her konsantrasyonda tüm fonksiyonlar için, en başarılı sonuçların PCA yöntemi ile sağlandığı görülmüştür.
- En yüksek sınıflandırma başarı oranları;
 - 12×10^8 CFU/mL konsantrasyonda, SVM (1-1) ve doğrusal fonksiyon kullanılarak %94.29,
 - 12×10^5 CFU/mL konsantrasyonda, SVM (1-1) ve doğrusal fonksiyon kullanılarak %94.29,
 - 12×10^3 CFU/mL konsantrasyonda, SVM (1-1) ve doğrusal fonksiyon kullanılarak %77.14,
 - 12×10^1 CFU/mL konsantrasyonda, SVM (1-1) ve doğrusal fonksiyon kullanılarak %88.10 olarak elde edilmiştir.

LIBSVM kütüphanesindeki yerleşik SVM fonksiyonları kullanılarak, sabit konsantrasyondaki mikroorganizma örneklerinden elde edilen her bir veri setinin 5-kat CV metoduna göre sınıflandırma başarı oranları hesaplanmıştır. Kullanılan 4 farklı çekirdek fonksiyonu için elde edilen sonuçlar, Tablo 2.10 ile Tablo 2.13 arasındaki tablolarda görülmektedir. Bu tablolardaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir. Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Her konsantrasyonda tüm fonksiyonlar için, sensör referans değeri-1 ve sensör referans değeri-3'ün başarılı olduğu görülmüştür.
- Her konsantrasyonda tüm fonksiyonlar için, en başarılı sonuçların çoğunun PCA yöntemi ile sağlandığı görülmüştür.
- Ölçeklemenin sınıflandırma başarısını etkilediği görülmüştür.
- Fonksiyonlara göre en yüksek sınıflandırma başarı oranları;
 - Doğrusal fonksiyon için; 12×10^5 CFU/mL konsantrasyonda %94.29
 - İkinci dereceden polinom için; 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarında %97.14
 - Üçüncü dereceden polinom için; 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarında %97.14

- RBF fonksiyonu için; 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarında %97.14 olarak elde edilmiştir.

Tablo 2.10. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki doğrusal fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz		Ölçekleme [0-1]		Ölçekleme [z-skor]		12×10^8
		$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	16.75	77.14	17.00	88.57	17.00	85.71	83.81
	<i>d1 & pca</i>	17.00	82.86	17.00	91.43	17.00	91.43	88.57
	<i>fd1 & std</i>	17.00	74.29	6.25	85.71	17.00	82.86	80.95
	<i>fd1 & pca</i>	17.00	80.00	17.00	91.43	17.00	91.43	87.62
	<i>nfd1 & std</i>	13.50	91.43	10.25	91.43	6.25	91.43	91.43
	<i>nfd1 & pca</i>	13.25	91.43	10.00	88.57	6.00	85.71	88.57
	Ortalama		82.86		89.52		88.10	86.83
Sensör Referans değeri - 2	<i>d2 & std</i>	17.00	51.43	17.00	60.00	17.00	60.00	57.14
	<i>d2 & pca</i>	17.00	54.29	17.00	74.29	17.00	74.29	67.62
	<i>fd2 & std</i>	17.00	74.29	11.75	71.43	17.00	71.43	72.38
	<i>fd2 & pca</i>	17.00	57.14	17.00	80.00	17.00	77.14	71.43
	<i>nfd2 & std</i>	15.00	82.86	9.50	82.86	5.50	82.86	82.86
	<i>nfd2 & pca</i>	12.00	80.00	9.75	80.00	1.50	80.00	80.00
	Ortalama		66.67		74.76		74.29	71.90
Sensör Referans değeri - 3	<i>d3 & std</i>	17.00	77.14	17.00	85.71	17.00	85.71	82.86
	<i>d3 & pca</i>	17.00	82.86	17.00	85.71	3.00	88.57	85.71
	<i>fd3 & std</i>	17.00	74.29	6.50	82.86	17.00	85.71	80.95
	<i>fd3 & pca</i>	17.00	77.14	17.00	91.43	17.00	91.43	86.67
	<i>nfd3 & std</i>	13.25	91.43	9.75	91.43	5.75	91.43	91.43
	<i>nfd3 & pca</i>	15.25	88.57	9.00	88.57	4.50	88.57	88.57
	Ortalama		81.90		87.62		88.57	86.03

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz		Ölçekleme [0-1]		Ölçekleme [z-skor]		12×10^5
		$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	17.00	68.57	17.00	85.71	17.00	85.71	80.00
	<i>d1 & pca</i>	17.00	71.43	17.00	88.57	17.00	85.71	81.90
	<i>fd1 & std</i>	17.00	62.86	14.50	77.14	14.00	82.86	74.29
	<i>fd1 & pca</i>	17.00	65.71	9.25	94.29	17.00	91.43	83.81
	<i>nfd1 & std</i>	13.75	91.43	8.25	94.29	5.75	91.43	92.38
	<i>nfd1 & pca</i>	17.00	88.57	17.00	94.29	17.00	94.29	92.38
	Ortalama		74.76		89.05		88.57	84.13
Sensör Referans değeri - 2	<i>d2 & std</i>	17.00	60.00	13.75	74.29	17.00	77.14	70.48
	<i>d2 & pca</i>	16.75	68.57	5.25	80.00	1.75	77.14	75.24
	<i>fd2 & std</i>	17.00	60.00	12.75	77.14	9.75	77.14	71.43
	<i>fd2 & pca</i>	17.00	60.00	7.50	80.00	17.00	80.00	73.33
	<i>nfd2 & std</i>	11.25	82.86	7.25	82.86	3.00	82.86	82.86
	<i>nfd2 & pca</i>	13.00	82.86	17.00	85.71	4.75	91.43	86.67
	Ortalama		69.05		80.00		80.95	76.67
Sensör Referans değeri - 3	<i>d3 & std</i>	16.75	65.71	14.25	82.86	11.00	82.86	77.14
	<i>d3 & pca</i>	17.00	68.57	17.00	94.29	17.00	94.29	85.71
	<i>fd3 & std</i>	17.00	60.00	13.75	82.86	13.00	80.00	74.29
	<i>fd3 & pca</i>	17.00	65.71	17.00	94.29	17.00	94.29	84.76
	<i>nfd3 & std</i>	13.25	85.71	7.00	91.43	4.75	85.71	87.62
	<i>nfd3 & pca</i>	12.25	88.57	6.25	91.43	5.50	91.43	90.48
	Ortalama		72.38		89.52		88.10	83.33

(b)

(Tablo 2.10'un devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz		Ölçekleme [0-1]		Ölçekleme [z-skor]		12x10 ³
		log ₂ C	% Başarı	log ₂ C	% Başarı	log ₂ C	% Başarı	Ortalama
Sensör Referans değeri - 1	d1 & std	17.00	42.86	10.75	68.57	17.00	65.71	59.05
	d1 & pca	16.50	40.00	17.00	68.57	3.00	71.43	60.00
	fd1 & std	17.00	40.00	11.50	74.29	7.50	74.29	62.86
	fd1 & pca	17.00	40.00	6.00	71.43	1.75	71.43	60.95
	nfd1 & std	14.50	62.86	7.75	65.71	4.00	65.71	64.76
	nfd1 & pca	17.00	62.86	4.50	68.57	17.00	65.71	65.71
	Ortalama		48.10		69.52		69.05	62.22
Sensör Referans değeri - 2	d2 & std	14.50	25.71	13.00	57.14	9.00	60.00	47.62
	d2 & pca	17.00	34.29	17.00	48.57	17.00	54.29	45.71
	fd2 & std	17.00	34.29	17.00	48.57	13.50	48.57	43.81
	fd2 & pca	16.75	37.14	7.00	62.86	17.00	60.00	53.33
	nfd2 & std	14.00	65.71	14.25	65.71	10.25	65.71	65.71
	nfd2 & pca	17.00	74.29	7.25	68.57	3.00	68.57	70.48
	Ortalama		45.24		58.57		59.52	54.44
Sensör Referans değeri - 3	d3 & std	17.00	31.43	10.00	74.29	9.75	68.57	58.10
	d3 & pca	17.00	37.14	7.75	74.29	3.25	74.29	61.90
	fd3 & std	16.75	40.00	11.75	77.14	7.75	77.14	64.76
	fd3 & pca	16.75	40.00	7.25	77.14	2.75	77.14	64.76
	nfd3 & std	15.75	68.57	9.75	65.71	5.75	65.71	66.67
	nfd3 & pca	13.00	74.29	5.00	80.00	1.50	82.86	79.05
	Ortalama		48.57		74.76		74.29	65.87

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz		Ölçekleme [0-1]		Ölçekleme [z-skor]		12x10 ¹
		log ₂ C	% Başarı	log ₂ C	% Başarı	log ₂ C	% Başarı	Ortalama
Sensör Referans değeri - 1	d1 & std	17.00	45.71	15.50	68.57	9.75	71.43	61.90
	d1 & pca	17.00	51.43	11.00	74.29	3.75	74.29	66.67
	fd1 & std	17.00	45.71	16.25	80.00	13.00	80.00	68.57
	fd1 & pca	17.00	51.43	10.00	82.86	5.25	82.86	72.38
	nfd1 & std	14.75	85.71	11.50	82.86	6.75	85.71	84.76
	nfd1 & pca	12.00	82.86	17.00	77.14	1.25	80.00	80.00
	Ortalama		60.48		77.62		79.05	72.38
Sensör Referans değeri - 2	d2 & std	16.25	37.14	10.25	48.57	7.00	51.43	45.71
	d2 & pca	16.00	42.86	5.25	48.57	1.50	51.43	47.62
	fd2 & std	17.00	31.43	17.00	60.00	7.50	68.57	53.33
	fd2 & pca	14.50	34.29	13.00	65.71	9.25	65.71	55.24
	nfd2 & std	14.50	80.00	14.75	80.00	10.50	80.00	80.00
	nfd2 & pca	13.75	80.00	6.50	82.86	3.25	85.71	82.86
	Ortalama		50.95		64.29		67.14	60.79
Sensör Referans değeri - 3	d3 & std	17.00	40.00	17.00	60.00	14.75	60.00	53.33
	d3 & pca	17.00	48.57	17.00	65.71	17.00	71.43	61.90
	fd3 & std	17.00	48.57	17.00	77.14	14.00	77.14	67.62
	fd3 & pca	17.00	48.57	17.00	85.71	17.00	85.71	73.33
	nfd3 & std	13.75	85.71	17.00	80.00	4.50	82.86	82.86
	nfd3 & pca	14.75	85.71	6.25	91.43	3.50	91.43	89.52
	Ortalama		59.52		76.67		78.10	71.43

(d)

Tablo 2.11. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki ikinci dereceden fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12×10^8
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	5.00	85.71	17.00	-4.00	88.57	17.00	-2.75	82.86	85.71
	d1 & pca	17.00	4.75	91.43	17.00	5.00	94.29	17.00	-3.25	94.29	93.33
	fd1 & std	17.00	5.00	85.71	15.50	-5.00	91.43	16.75	-8.50	88.57	88.57
	fd1 & pca	17.00	2.50	91.43	17.00	5.00	94.29	17.00	1.00	97.14	94.29
	nfd1 & std	15.50	-3.00	91.43	17.00	-7.75	91.43	17.00	-11.75	91.43	91.43
	nfd1 & pca	13.50	-1.00	91.43	17.00	-1.25	91.43	9.00	-2.25	94.29	92.38
	Ortalama			89.52			91.90			91.43	90.95
Sensör Referans değeri - 2	d2 & std	17.00	5.00	71.43	17.00	-4.00	74.29	16.75	-9.00	74.29	73.33
	d2 & pca	17.00	3.00	74.29	17.00	5.00	85.71	17.00	-2.50	82.86	80.95
	fd2 & std	17.00	5.00	80.00	16.75	-5.00	82.86	13.00	-7.75	80.00	80.95
	fd2 & pca	17.00	5.00	80.00	17.00	5.00	85.71	17.00	-4.00	82.86	82.86
	nfd2 & std	17.00	-3.00	82.86	17.00	-8.50	82.86	9.00	-5.50	82.86	82.86
	nfd2 & pca	16.00	-5.00	80.00	7.25	-1.25	85.71	17.00	5.00	85.71	83.81
	Ortalama			78.10			82.86			81.43	80.79
Sensör Referans değeri - 3	d3 & std	17.00	5.00	88.57	17.00	-4.50	88.57	17.00	-8.25	88.57	88.57
	d3 & pca	17.00	5.00	91.43	17.00	5.00	94.29	17.00	-3.00	97.14	94.29
	fd3 & std	17.00	5.00	82.86	17.00	-0.25	88.57	17.00	-3.75	88.57	86.67
	fd3 & pca	17.00	5.00	88.57	17.00	5.00	91.43	17.00	0.00	94.29	91.43
	nfd3 & std	14.50	-3.00	91.43	17.00	-8.25	91.43	0.00	2.75	94.29	92.38
	nfd3 & pca	17.00	-2.75	88.57	9.00	0.25	91.43	9.00	-2.25	94.29	91.43
	Ortalama			88.57			90.95			92.86	90.79

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12×10^5
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	3.50	82.86	17.00	-3.25	82.86	9.00	0.00	80.00	81.90
	d1 & pca	17.00	2.50	82.86	17.00	5.00	88.57	17.00	-0.25	88.57	86.67
	fd1 & std	15.00	5.00	74.29	17.00	3.50	82.86	17.00	0.00	82.86	80.00
	fd1 & pca	17.00	3.75	82.86	17.00	1.00	94.29	17.00	-0.50	91.43	89.52
	nfd1 & std	17.00	-4.50	91.43	17.00	-10.00	94.29	17.00	-12.25	91.43	92.38
	nfd1 & pca	17.00	4.75	88.57	17.00	-0.50	97.14	17.00	-1.25	97.14	94.29
	Ortalama			83.81			90.00			88.57	87.46
Sensör Referans değeri - 2	d2 & std	17.00	3.75	77.14	15.50	-2.75	74.29	13.00	-3.75	74.29	75.24
	d2 & pca	16.50	1.00	71.43	17.00	-12.75	80.00	17.00	-16.25	77.14	76.19
	fd2 & std	16.75	1.00	77.14	14.00	-3.00	77.14	16.75	-12.75	77.14	77.14
	fd2 & pca	17.00	1.00	74.29	17.00	-10.50	80.00	17.00	-7.75	80.00	78.10
	nfd2 & std	17.00	-6.75	82.86	17.00	-10.75	82.86	17.00	-15.00	82.86	82.86
	nfd2 & pca	17.00	-5.00	82.86	7.00	2.50	91.43	13.00	-4.00	91.43	88.57
	Ortalama			77.62			80.95			80.48	79.68
Sensör Referans değeri - 3	d3 & std	17.00	5.00	82.86	17.00	-0.25	82.86	17.00	-1.25	82.86	82.86
	d3 & pca	17.00	5.00	88.57	17.00	-1.75	94.29	17.00	-0.75	97.14	93.33
	fd3 & std	17.00	3.00	68.57	17.00	5.00	82.86	17.00	-3.00	82.86	78.10
	fd3 & pca	17.00	5.00	82.86	17.00	5.00	94.29	17.00	0.75	94.29	90.48
	nfd3 & std	14.75	-5.00	88.57	17.00	-11.00	91.43	17.00	-13.25	85.71	88.57
	nfd3 & pca	17.00	-5.75	88.57	8.25	-1.25	94.29	17.00	-12.50	91.43	91.43
	Ortalama			83.33			90.00			89.05	87.46

(b)

(Tablo 2.11'in devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ³
		log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	3.25	68.57	15.50	-4.25	68.57	15.00	-8.00	68.57	68.57
	d1 & pca	17.00	5.00	71.43	17.00	-0.50	80.00	17.00	-0.25	77.14	76.19
	fd1 & std	17.00	5.00	68.57	17.00	-0.75	74.29	17.00	-2.25	71.43	71.43
	fd1 & pca	17.00	3.00	71.43	17.00	5.00	71.43	17.00	-16.25	71.43	71.43
	nfd1 & std	17.00	3.00	62.86	17.00	-10.50	65.71	17.00	-8.00	65.71	64.76
	nfd1 & pca	14.50	-5.00	68.57	4.25	-1.25	74.29	4.25	-5.00	71.43	71.43
	Ortalama			68.57			72.38			70.95	70.63
Sensör Referans değeri - 2	d2 & std	17.00	4.50	62.86	17.00	-6.50	65.71	16.75	-10.25	65.71	64.76
	d2 & pca	16.25	4.75	57.14	1.75	2.75	60.00	17.00	0.00	65.71	60.95
	fd2 & std	17.00	5.00	48.57	15.00	-2.00	60.00	11.75	-6.00	60.00	56.19
	fd2 & pca	17.00	3.00	54.29	3.00	2.00	71.43	15.00	-8.00	71.43	65.71
	nfd2 & std	17.00	-2.25	68.57	2.00	2.50	74.29	12.50	-7.75	71.43	71.43
	nfd2 & pca	17.00	3.00	77.14	14.25	-11.00	71.43	9.00	-1.50	71.43	73.33
	Ortalama			61.43			67.14			67.62	65.40
Sensör Referans değeri - 3	d3 & std	17.00	3.00	74.29	16.50	-7.25	77.14	16.00	-11.00	77.14	76.19
	d3 & pca	17.00	1.75	74.29	9.00	-2.50	74.29	17.00	-14.75	74.29	74.29
	fd3 & std	17.00	4.25	77.14	17.00	-2.25	77.14	17.00	-6.25	77.14	77.14
	fd3 & pca	17.00	3.50	77.14	13.00	-6.75	77.14	13.00	-11.25	77.14	77.14
	nfd3 & std	16.50	-2.00	71.43	13.00	-4.25	74.29	13.00	-8.00	74.29	73.33
	nfd3 & pca	12.25	-0.25	77.14	16.75	-12.75	80.00	1.75	-3.25	82.86	80.00
	Ortalama			75.24			76.67			77.14	76.35

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ¹
		log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	5.00	68.57	15.25	-2.75	74.29	17.00	-6.75	74.29	72.38
	d1 & pca	17.00	5.00	77.14	17.00	-7.00	74.29	17.00	4.00	85.71	79.05
	fd1 & std	17.00	5.00	77.14	11.00	0.00	80.00	13.00	-4.75	80.00	79.05
	fd1 & pca	17.00	5.00	77.14	9.25	-1.25	85.71	17.00	1.25	85.71	82.86
	nfd1 & std	17.00	-3.50	85.71	16.00	-5.50	85.71	16.75	-11.00	85.71	85.71
	nfd1 & pca	17.00	-6.00	82.86	3.00	-2.25	82.86	7.00	-6.75	80.00	81.90
	Ortalama			78.10			80.48			81.90	80.16
Sensör Referans değeri - 2	d2 & std	17.00	3.75	48.57	17.00	-0.75	60.00	17.00	-3.75	60.00	56.19
	d2 & pca	17.00	5.00	51.43	17.00	2.00	51.43	17.00	1.75	57.14	53.33
	fd2 & std	17.00	4.50	65.71	15.50	-1.00	74.29	17.00	-3.25	71.43	70.48
	fd2 & pca	17.00	5.00	62.86	16.50	-5.00	68.57	-1.25	-1.00	68.57	66.67
	nfd2 & std	16.50	-3.00	80.00	17.00	-8.00	82.86	17.00	-2.50	80.00	80.95
	nfd2 & pca	17.00	-4.25	80.00	-1.00	1.75	85.71	17.00	-14.75	85.71	83.81
	Ortalama			64.76			70.48			70.48	68.57
Sensör Referans değeri - 3	d3 & std	17.00	5.00	60.00	14.25	1.50	62.86	7.00	4.50	62.86	61.90
	d3 & pca	17.00	5.00	62.86	17.00	5.00	68.57	10.75	-3.50	74.29	68.57
	fd3 & std	17.00	5.00	74.29	17.00	2.50	85.71	14.75	-5.00	85.71	81.90
	fd3 & pca	17.00	5.00	71.43	17.00	5.00	85.71	17.00	-5.00	91.43	82.86
	nfd3 & std	17.00	-4.25	85.71	16.50	-9.00	85.71	17.00	-13.50	82.86	84.76
	nfd3 & pca	17.00	-3.25	85.71	17.00	-11.75	91.43	17.00	-14.50	91.43	89.52
	Ortalama			73.33			80.00			81.43	78.25

(d)

Tablo 2.12. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki üçüncü dereceden fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12×10^8
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	4.75	88.57	17.00	-1.00	82.86	17.00	-4.25	82.86	84.76
	d1 & pca	17.00	3.75	91.43	7.00	0.25	97.14	17.00	-4.50	94.29	94.29
	fd1 & std	16.25	1.75	85.71	16.00	-6.00	91.43	15.00	-8.00	88.57	88.57
	fd1 & pca	17.00	5.00	88.57	7.00	1.50	97.14	17.00	-0.25	97.14	94.29
	nfd1 & std	7.00	2.25	91.43	17.00	-8.50	91.43	17.00	-12.25	91.43	91.43
	nfd1 & pca	14.75	-3.00	91.43	15.00	-1.25	91.43	10.00	-3.75	94.29	92.38
	Ortalama			89.52			91.90			91.43	90.95
Sensör Referans değeri - 2	d2 & std	17.00	4.50	71.43	17.00	-5.00	74.29	17.00	-7.00	71.43	72.38
	d2 & pca	17.00	5.00	71.43	17.00	0.00	85.71	17.00	-5.25	82.86	80.00
	fd2 & std	17.00	5.00	80.00	17.00	-6.00	82.86	15.00	-9.25	80.00	80.95
	fd2 & pca	17.00	5.00	80.00	17.00	5.00	85.71	17.00	-5.00	82.86	82.86
	nfd2 & std	17.00	-3.75	82.86	16.75	-9.00	82.86	17.00	-13.00	82.86	82.86
	nfd2 & pca	17.00	-6.75	82.86	17.00	3.00	82.86	17.00	-1.75	85.71	83.81
	Ortalama			78.10			82.38			80.95	80.48
Sensör Referans değeri - 3	d3 & std	17.00	4.25	88.57	17.00	-4.25	88.57	17.00	-9.00	88.57	88.57
	d3 & pca	17.00	4.25	91.43	6.50	-1.25	97.14	5.00	-4.00	97.14	95.24
	fd3 & std	17.00	4.25	82.86	15.50	-5.75	91.43	17.00	-4.75	88.57	87.62
	fd3 & pca	17.00	5.00	88.57	17.00	4.25	94.29	17.00	-1.50	94.29	92.38
	nfd3 & std	17.00	-6.00	91.43	17.00	-8.75	91.43	17.00	-12.75	91.43	91.43
	nfd3 & pca	17.00	-3.25	88.57	10.50	-1.75	91.43	10.25	-4.00	94.29	91.43
	Ortalama			88.57			92.38			92.38	91.11

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12×10^5
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	1.75	82.86	17.00	-4.00	82.86	17.00	-7.25	82.86	82.86
	d1 & pca	15.00	4.25	85.71	17.00	1.00	88.57	17.00	0.25	88.57	87.62
	fd1 & std	17.00	5.00	77.14	17.00	1.50	82.86	17.00	-2.25	82.86	80.95
	fd1 & pca	14.75	4.50	85.71	8.75	-2.50	97.14	17.00	-0.25	91.43	91.43
	nfd1 & std	17.00	-5.00	91.43	14.00	-5.50	94.29	17.00	-12.75	91.43	92.38
	nfd1 & pca	17.00	2.75	88.57	17.00	-2.00	97.14	17.00	-2.50	97.14	94.29
	Ortalama			85.24			90.48			89.05	88.25
Sensör Referans değeri - 2	d2 & std	17.00	3.00	77.14	16.25	-4.00	74.29	-0.25	2.75	77.14	76.19
	d2 & pca	15.00	2.00	71.43	8.50	-5.00	80.00	17.00	-16.50	74.29	75.24
	fd2 & std	17.00	3.50	74.29	14.75	-4.25	77.14	-1.00	2.50	80.00	77.14
	fd2 & pca	16.50	1.00	74.29	17.00	-11.00	80.00	17.00	-8.50	80.00	78.10
	nfd2 & std	3.00	2.25	82.86	7.00	-2.00	85.71	16.50	-15.00	82.86	83.81
	nfd2 & pca	17.00	5.00	82.86	9.00	0.00	91.43	17.00	-3.00	91.43	88.57
	Ortalama			77.14			81.43			80.95	79.84
Sensör Referans değeri - 3	d3 & std	17.00	5.00	82.86	17.00	-1.75	82.86	17.00	-4.00	82.86	82.86
	d3 & pca	17.00	5.00	88.57	17.00	-2.75	94.29	17.00	-1.25	97.14	93.33
	fd3 & std	17.00	5.00	77.14	17.00	0.75	82.86	17.00	-4.00	82.86	80.95
	fd3 & pca	17.00	5.00	85.71	17.00	0.00	94.29	17.00	0.00	97.14	92.38
	nfd3 & std	16.25	-7.00	88.57	16.25	-11.00	91.43	9.50	-7.00	88.57	89.52
	nfd3 & pca	17.00	-6.50	88.57	7.00	-1.25	94.29	17.00	-13.00	91.43	91.43
	Ortalama			85.24			90.00			90.00	88.41

(b)

(Tablo 2.12'nin devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ³
		log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	4.75	71.43	17.00	-6.25	68.57	15.75	-8.00	68.57	69.52
	d1 & pca	17.00	5.00	74.29	17.00	0.25	77.14	17.00	-2.50	77.14	76.19
	fd1 & std	17.00	4.75	71.43	17.00	-3.00	74.29	17.00	-7.00	74.29	73.33
	fd1 & pca	16.75	4.75	71.43	-3.00	1.00	74.29	17.00	-16.50	71.43	72.38
	nfd1 & std	7.00	2.25	65.71	17.00	-11.00	65.71	11.75	-7.50	68.57	66.67
	nfd1 & pca	16.00	-7.00	68.57	-5.75	3.25	80.00	2.50	-3.75	74.29	74.29
	Ortalama			70.48			73.33			72.38	72.06
Sensör Referans değeri - 2	d2 & std	17.00	5.00	65.71	14.75	-4.25	65.71	14.75	-8.00	65.71	65.71
	d2 & pca	17.00	3.50	57.14	17.00	5.00	60.00	17.00	-1.75	65.71	60.95
	fd2 & std	17.00	3.00	48.57	17.00	-4.00	60.00	17.00	-7.75	60.00	56.19
	fd2 & pca	17.00	4.75	62.86	6.75	-2.25	74.29	17.00	-9.75	71.43	69.52
	nfd2 & std	17.00	1.75	71.43	13.00	-5.00	71.43	12.75	-8.75	71.43	71.43
	nfd2 & pca	14.75	-1.00	77.14	7.50	-3.00	71.43	17.00	-2.50	71.43	73.33
	Ortalama			63.81			67.14			67.62	66.19
Sensör Referans değeri - 3	d3 & std	17.00	2.50	74.29	17.00	-8.25	77.14	16.50	-12.00	77.14	76.19
	d3 & pca	17.00	1.25	74.29	17.00	-10.75	74.29	8.50	-6.00	77.14	75.24
	fd3 & std	17.00	5.00	77.14	17.00	-3.00	77.14	17.00	-7.00	77.14	77.14
	fd3 & pca	17.00	2.75	77.14	17.00	-11.25	77.14	17.00	-15.75	77.14	77.14
	nfd3 & std	15.50	-2.75	71.43	14.25	-6.00	74.29	14.00	-9.75	74.29	73.33
	nfd3 & pca	12.75	-1.25	77.14	16.50	-12.50	80.00	16.75	-17.00	82.86	80.00
	Ortalama			75.24			76.67			77.62	76.51

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ¹
		log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	3.00	71.43	17.00	-4.75	74.29	14.25	-6.50	74.29	73.33
	d1 & pca	17.00	4.50	80.00	13.00	-4.00	74.29	17.00	1.25	82.86	79.05
	fd1 & std	17.00	5.00	77.14	12.00	-1.50	80.00	15.00	-6.25	80.00	79.05
	fd1 & pca	17.00	5.00	77.14	8.75	-1.75	85.71	17.00	-1.75	85.71	82.86
	nfd1 & std	7.00	2.25	88.57	16.25	-6.25	85.71	16.25	-10.50	85.71	86.67
	nfd1 & pca	17.00	-6.50	82.86	5.50	-5.00	82.86	-0.25	-3.00	85.71	83.81
	Ortalama			79.52			80.48			82.38	80.79
Sensör Referans değeri - 2	d2 & std	17.00	3.25	51.43	17.00	-1.75	60.00	17.00	-4.75	60.00	57.14
	d2 & pca	17.00	5.00	57.14	13.00	-0.75	54.29	17.00	-0.25	62.86	58.10
	fd2 & std	17.00	5.00	71.43	17.00	3.00	71.43	-2.75	2.75	71.43	71.43
	fd2 & pca	17.00	5.00	62.86	17.00	-6.00	68.57	-2.00	-1.00	68.57	66.67
	nfd2 & std	5.00	3.00	80.00	17.00	-8.50	82.86	17.00	-5.00	80.00	80.95
	nfd2 & pca	17.00	-4.75	80.00	17.00	-12.25	82.86	17.00	-15.25	85.71	82.86
	Ortalama			67.14			70.00			71.43	69.52
Sensör Referans değeri - 3	d3 & std	15.25	4.75	60.00	10.75	0.00	62.86	17.00	5.00	62.86	61.90
	d3 & pca	17.00	4.75	62.86	17.00	2.50	71.43	17.00	3.25	74.29	69.52
	fd3 & std	17.00	5.00	77.14	17.00	-4.25	85.71	17.00	-3.50	82.86	81.90
	fd3 & pca	17.00	5.00	71.43	17.00	-15.00	85.71	15.00	-6.00	91.43	82.86
	nfd3 & std	17.00	-5.00	85.71	16.00	-9.00	85.71	9.00	-6.25	82.86	84.76
	nfd3 & pca	17.00	-4.00	85.71	17.00	-12.25	91.43	17.00	-15.00	91.43	89.52
	Ortalama			73.81			80.48			80.95	78.41

(d)

Tablo 2.13. Konsantrasyona göre mikroorganizma örneklerinden elde edilen veri setlerinin, LIBSVM kütüphanesindeki RBF fonksiyonlar kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Konsantrasyonlar: (a) 12×10^8 CFU/mL, (b) 12×10^5 CFU/mL, (c) 12×10^3 CFU/mL, (d) 12×10^1 CFU/mL.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12×10^8
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	3.00	88.57	17.00	-4.25	85.71	17.00	-9.00	85.71	86.67
	d1 & pca	17.00	3.25	91.43	17.00	-1.50	94.29	17.00	-5.00	94.29	93.33
	fd1 & std	17.00	1.50	85.71	17.00	-6.50	91.43	15.00	-7.75	88.57	88.57
	fd1 & pca	17.00	5.00	88.57	17.00	2.75	94.29	17.00	-1.75	94.29	92.38
	nfd1 & std	15.50	-3.00	91.43	17.00	-7.75	91.43	17.00	-11.75	91.43	91.43
	nfd1 & pca	15.50	-3.00	91.43	11.00	0.00	94.29	7.75	-1.75	94.29	93.33
	Ortalama			89.52			91.90			91.43	90.95
Sensör Referans değeri - 2	d2 & std	17.00	3.75	71.43	17.00	-2.75	71.43	17.00	-10.25	71.43	71.43
	d2 & pca	17.00	5.00	71.43	13.75	-5.50	85.71	17.00	-6.00	82.86	80.00
	fd2 & std	17.00	5.00	80.00	14.00	-5.00	82.86	15.00	-9.50	80.00	80.95
	fd2 & pca	17.00	5.00	80.00	17.00	-2.50	85.71	9.00	-6.50	85.71	83.81
	nfd2 & std	17.00	-3.00	82.86	17.00	-8.50	82.86	17.00	-12.50	82.86	82.86
	nfd2 & pca	16.00	-5.00	80.00	17.00	4.00	82.86	17.00	0.00	82.86	81.90
	Ortalama			77.62			81.90			80.95	80.16
Sensör Referans değeri - 3	d3 & std	14.75	2.75	88.57	17.00	-5.00	88.57	17.00	-8.50	85.71	87.62
	d3 & pca	17.00	3.50	91.43	9.00	-1.25	97.14	17.00	-4.50	97.14	95.24
	fd3 & std	17.00	5.00	82.86	17.00	-1.50	88.57	17.00	-5.25	88.57	86.67
	fd3 & pca	17.00	5.00	88.57	17.00	2.50	94.29	17.00	1.00	94.29	92.38
	nfd3 & std	15.25	-3.00	91.43	17.00	-8.25	91.43	3.25	1.50	94.29	92.38
	nfd3 & pca	17.00	-2.75	88.57	11.00	-0.25	94.29	11.00	-4.25	94.29	92.38
	Ortalama			88.57			92.38			92.38	91.11

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12×10^5
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	3.50	82.86	17.00	-4.00	82.86	16.75	-7.00	80.00	81.90
	d1 & pca	17.00	2.75	82.86	17.00	1.25	91.43	17.00	-2.25	88.57	87.62
	fd1 & std	15.25	4.75	80.00	17.00	0.25	82.86	17.00	-3.25	82.86	81.90
	fd1 & pca	16.00	4.00	85.71	17.00	1.50	97.14	17.00	-3.00	94.29	92.38
	nfd1 & std	17.00	-4.25	91.43	17.00	-9.75	94.29	17.00	-12.25	91.43	92.38
	nfd1 & pca	17.00	2.00	88.57	17.00	1.00	97.14	17.00	-3.75	94.29	93.33
	Ortalama			85.24			90.95			88.57	88.25
Sensör Referans değeri - 2	d2 & std	17.00	4.75	77.14	15.75	-3.00	74.29	16.25	-7.00	74.29	75.24
	d2 & pca	16.50	1.00	71.43	17.00	-12.75	80.00	17.00	-16.25	77.14	76.19
	fd2 & std	17.00	4.00	74.29	15.75	-4.75	77.14	15.75	-12.00	77.14	76.19
	fd2 & pca	17.00	1.00	74.29	17.00	-10.50	80.00	17.00	-8.25	80.00	78.10
	nfd2 & std	17.00	-6.75	82.86	17.00	-10.75	82.86	17.00	-15.00	82.86	82.86
	nfd2 & pca	17.00	4.00	82.86	15.00	0.50	91.43	15.00	-3.75	91.43	88.57
	Ortalama			77.14			80.95			80.48	79.52
Sensör Referans değeri - 3	d3 & std	17.00	3.00	74.29	17.00	-1.25	82.86	17.00	-4.50	82.86	80.00
	d3 & pca	17.00	4.75	88.57	17.00	1.25	94.29	17.00	-3.00	94.29	92.38
	fd3 & std	17.00	5.00	80.00	17.00	-1.00	82.86	17.00	-4.25	82.86	81.90
	fd3 & pca	17.00	5.00	85.71	17.00	1.25	97.14	17.00	-3.00	94.29	92.38
	nfd3 & std	14.75	-5.00	88.57	17.00	-11.00	91.43	17.00	-13.25	85.71	88.57
	nfd3 & pca	17.00	-5.75	88.57	17.00	-9.25	91.43	17.00	-12.50	91.43	90.48
	Ortalama			84.29			90.00			88.57	87.62

(b)

(Tablo 2.13'ün devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ³
		log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	5.00	71.43	15.75	-5.00	68.57	15.75	-9.00	68.57	69.52
	d1 & pca	17.00	4.50	71.43	17.00	0.50	77.14	17.00	-3.50	74.29	74.29
	fd1 & std	17.00	4.75	68.57	17.00	-3.00	71.43	17.00	-7.00	71.43	70.48
	fd1 & pca	17.00	5.00	68.57	17.00	-12.00	71.43	17.00	-15.75	71.43	70.48
	nfd1 & std	17.00	0.25	65.71	17.00	-10.25	65.71	17.00	-14.00	65.71	65.71
	nfd1 & pca	14.25	-1.50	65.71	4.00	-1.00	74.29	4.75	-5.00	71.43	70.48
	Ortalama			68.57			71.43			70.48	70.16
Sensör Referans değeri - 2	d2 & std	16.50	5.00	68.57	16.00	-4.75	65.71	14.25	-7.00	65.71	66.67
	d2 & pca	17.00	3.75	57.14	17.00	1.75	57.14	17.00	-2.25	60.00	58.10
	fd2 & std	17.00	5.00	57.14	13.00	-2.75	60.00	16.00	-7.00	60.00	59.05
	fd2 & pca	17.00	3.00	57.14	9.50	-3.50	71.43	17.00	-9.50	68.57	65.71
	nfd2 & std	15.00	-0.25	71.43	14.25	-5.50	71.43	14.25	-9.50	71.43	71.43
	nfd2 & pca	14.75	-1.00	77.14	17.00	1.25	71.43	17.00	-3.00	74.29	74.29
	Ortalama			64.76			66.19			66.67	65.87
Sensör Referans değeri - 3	d3 & std	17.00	2.75	74.29	16.75	-8.00	77.14	17.00	-12.00	74.29	75.24
	d3 & pca	17.00	1.75	74.29	9.00	-1.75	74.29	8.50	-5.75	77.14	75.24
	fd3 & std	17.00	5.00	77.14	17.00	-2.25	77.14	17.00	-6.25	77.14	77.14
	fd3 & pca	17.00	2.75	74.29	11.00	-4.75	77.14	13.00	-11.25	77.14	76.19
	nfd3 & std	14.50	-1.25	74.29	13.00	-4.50	74.29	13.00	-8.50	74.29	74.29
	nfd3 & pca	14.50	-2.50	77.14	17.00	-13.00	80.00	17.00	-16.50	82.86	80.00
	Ortalama			75.24			76.67			77.14	76.35

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			12x10 ¹
		log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	log ₂ C	log ₂ γ	% Başarı	
Sensör Referans değeri - 1	d1 & std	17.00	5.00	68.57	14.75	-2.75	74.29	14.50	-5.75	74.29	72.38
	d1 & pca	17.00	5.00	77.14	17.00	0.50	82.86	17.00	-1.00	77.14	79.05
	fd1 & std	17.00	5.00	77.14	15.00	-3.25	80.00	16.50	-7.00	80.00	79.05
	fd1 & pca	17.00	5.00	77.14	17.00	-8.00	82.86	3.00	2.00	82.86	80.95
	nfd1 & std	17.00	-2.00	85.71	17.00	-6.50	85.71	17.00	-10.50	85.71	85.71
	nfd1 & pca	17.00	-6.00	82.86	-1.00	3.75	85.71	1.00	-1.25	80.00	82.86
	Ortalama			78.10			81.90			80.00	80.00
Sensör Referans değeri - 2	d2 & std	17.00	3.25	48.57	17.00	-1.50	57.14	17.00	-4.75	57.14	54.29
	d2 & pca	17.00	5.00	57.14	17.00	4.50	54.29	7.00	-1.25	62.86	58.10
	fd2 & std	17.00	4.75	71.43	16.75	-2.75	74.29	17.00	-5.00	71.43	72.38
	fd2 & pca	17.00	5.00	62.86	2.75	3.25	68.57	17.00	1.25	68.57	66.67
	nfd2 & std	17.00	2.50	77.14	14.25	-3.25	80.00	16.00	-9.00	80.00	79.05
	nfd2 & pca	17.00	-4.25	80.00	5.75	-1.25	85.71	17.00	-14.75	85.71	83.81
	Ortalama			66.19			70.00			70.95	69.05
Sensör Referans değeri - 3	d3 & std	16.25	4.25	62.86	17.00	-0.75	57.14	17.00	-5.00	57.14	59.05
	d3 & pca	16.25	2.75	62.86	13.50	-1.50	68.57	17.00	-5.25	68.57	66.67
	fd3 & std	17.00	5.00	74.29	17.00	-3.75	85.71	17.00	-2.50	82.86	80.95
	fd3 & pca	17.00	5.00	71.43	17.00	-3.00	91.43	17.00	-6.75	91.43	84.76
	nfd3 & std	17.00	-4.25	85.71	14.50	-7.00	85.71	17.00	-13.50	82.86	84.76
	nfd3 & pca	17.00	-3.25	85.71	17.00	-11.75	91.43	17.00	-14.50	91.43	89.52
	Ortalama			73.81			80.00			79.05	77.62

(d)

2.3.4. Sınıflandırma Yöntemlerinin Sonuçlarının Karşılaştırılması

Elde edilen sınıflandırma sonuçları incelendiğinde; 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarına ait başarıların, diğer konsantrasyonlara göre daha yüksek olduğu görülmüştür. Bu nedenle Tablo 2.14’de, her iki konsantrasyon için kullanılan sınıflandırma yöntemlerine göre 5-kat CV metodu ile hesaplanan en yüksek başarı oranları ve bu oranları sağlayan alt metot, ölçekleme, ön-işleme ve boyut azaltım yöntemleri verilmiştir.

Tablo 2.14. Mikroorganizma seti-2’nin 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarına ait örnekleri için, sınıflandırma yöntemlerine göre 5-kat CV metodu ile elde edilen en yüksek başarı oranları ve bu oranları sağlayan çeşitli yöntemler.

	Yöntem	k-NN	DA	MATLAB SVM	LIBSVM
12×10^8 CFU/mL	Ön-işleme *	<i>fd1</i>	<i>nfd1, d3, nfd3</i>	<i>d1</i>	<i>fd1, d3</i>
	Boyut azaltım **	PCA	PCA, <i>Std</i>	PCA	PCA
	Ölçekleme	[0-1] ölçekleme z-skör ölçekleme	-	-	[0-1] ölçekleme z-skör ölçekleme
	Metot	$k = 1$	Doğrusal, ikinci dereceden, diyagonal ikinci dereceden, Mahalanobis	SVM (1-1) ve doğrusal fonksiyon	İkinci ve üçüncü dereceden polinom, RBF
	Sınıflandırma başarıları (%)	91.43	85.71	94.29	97.14
12×10^5 CFU/mL	Ön-işleme *	<i>nfd3</i>	<i>nfd1</i>	<i>fd1, nfd1, d3, fd3</i>	<i>fd1, nfd1, d3, fd3</i>
	Boyut azaltım	PCA	PCA	PCA	PCA
	Ölçekleme	[0-1] ölçekleme z-skör ölçekleme	-	-	[0-1] ölçekleme z-skör ölçekleme
	Metot	$k = 3$	Doğrusal fonksiyon	SVM (1-1) ve doğrusal fonksiyon	İkinci ve üçüncü dereceden polinom, RBF
	Sınıflandırma başarıları (%)	94.29	100	94.29	97.14

(*) Ön-işleme kısaltmaları; sensör cevap modeli ve sensör referans değeri noktasını tanımlamaktadır. Eşitlik ifadeleri için Tablo 1.6’ya bakınız.

(**) Boyut azaltım kısmında; *Std*, standart sapmaya göre en fazla değişim gösteren 3 sensörün belirlenmesini ifade etmektedir.

Burada sınıflandırma yöntemleri arasında en yüksek başarıyı; 12×10^8 CFU/mL konsantrasyonda %97.14 ile LIBSVM yazılımının, 12×10^5 CFU/mL konsantrasyonda %100 ile DA yönteminin sağladığı görülmüştür. k -NN ve LIBSVM ile yapılan sınıflandırmalarda en yüksek başarılar, ölçekleme yöntemleri kullanılarak elde edilmiştir. Tablo genelindeki başarılı ön-işleme yöntemleri incelendiğinde; sensör referans değeri-1 ($R_i^o = R_i^{\min}$) ve 3 ($R_i^o = R_i^{\text{fbp}}$) ön plana çıkmıştır. Aynı zamanda tablo genelinde boyut azaltımında PCA yönteminin en başarılı sonuçları verdiği görülmüştür.

2.4. Mikroorganizma Seti-3 için Sınıflandırma Sonuçları

Bu mikroorganizma seti; ikinci setteki mikroorganizmalarla aynı olup, sabit konsantrasyondaki daha fazla örnek üzerinde elektronik burun cihazının kullanımı araştırılmıştır [62]. Daha fazla örnek üzerinde çalışma yapılmasının sebebi; elektronik burun cihazı ile elde edilen mikroorganizma koku verilerinin, sınıflandırma başarısının güvenilirliğini artırmaktır. Buradaki mikroorganizmalar spesifik besiyerlerine ekim yapılarak üretildikten sonra, her bir mikroorganizmanın, SF içeren ağzı kapaklı steril cam tüplerde, 1.5×10^8 CFU/mL konsantrasyonlarındaki süspansiyonları hazırlanmıştır. Gerçeğe yakın bir durum ortaya koymak için, ependorf tüplerde steril diş içeren örnekler oluşturulmuştur. Daha sonra her mikroorganizma süspansiyonu, ependorf tüplerde steril diş içeren 12 örneğe eklenerek toplam 84 örnek üzerinde çalışılmış ve her örnekten birer koku verisi alınmıştır. Dolayısıyla her mikroorganizmadan 12 koku verisi olmak üzere toplam 84 koku verisi üzerinde çalışılmıştır. Farklı yöntemlere göre elde edilen sınıflandırma sonuçları, aşağıdaki kısımlarda açıklanmıştır.

2.4.1. k -NN Algoritması ile Sınıflandırma

k -NN algoritması kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.15’de görülmektedir. Bu tablodaki koyu değerler, k değerlerine göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.15. Veri setlerinin k -NN algoritması kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			Ortalama
		$k = 1$	$k = 3$	$k = 5$	$k = 1$	$k = 3$	$k = 5$	$k = 1$	$k = 3$	$k = 5$	
Sensör Referans değeri - 1	<i>d1 & std</i>	55.95	64.29	64.29	58.33	65.48	67.86	58.33	63.10	67.86	62.83
	<i>d1 & pca</i>	59.52	65.48	67.86	67.86	72.62	72.62	70.24	69.05	73.81	68.78
	<i>fd1 & std</i>	60.71	61.90	65.48	63.10	66.67	66.67	63.10	66.67	67.86	64.68
	<i>fd1 & pca</i>	59.52	64.29	65.48	65.48	65.48	69.05	65.48	66.67	67.86	65.48
	<i>nfd1 & std</i>	64.29	63.10	61.90	61.90	58.33	61.90	64.29	61.90	64.29	62.43
	<i>nfd1 & pca</i>	67.86	70.24	71.43	66.67	65.48	65.48	67.86	65.48	65.48	67.33
	Ortalama	61.31	64.88	66.07	63.89	65.67	67.26	64.88	65.48	67.86	65.26
Sensör Referans değeri - 2	<i>d2 & std</i>	50.00	57.14	53.57	51.19	52.38	47.62	53.57	55.95	50.00	52.38
	<i>d2 & pca</i>	55.95	58.33	54.76	51.19	57.14	57.14	54.76	60.71	59.52	56.61
	<i>fd2 & std</i>	50.00	54.76	52.38	44.05	46.43	45.24	47.62	52.38	45.24	48.68
	<i>fd2 & pca</i>	55.95	55.95	53.57	51.19	55.95	53.57	50.00	55.95	53.57	53.97
	<i>nfd2 & std</i>	48.81	47.62	57.14	50.00	47.62	53.57	51.19	52.38	55.95	51.59
	<i>nfd2 & pca</i>	55.95	54.76	59.52	57.14	54.76	57.14	57.14	54.76	59.52	56.75
	Ortalama	52.78	54.76	55.16	50.79	52.38	52.38	52.38	55.36	53.97	53.33
Sensör Referans değeri - 3	<i>d3 & std</i>	61.90	66.67	66.67	60.71	67.86	65.48	63.10	67.86	65.48	65.08
	<i>d3 & pca</i>	63.10	66.67	67.86	64.29	67.86	70.24	63.10	66.67	67.86	66.40
	<i>fd3 & std</i>	55.95	66.67	60.71	60.71	67.86	70.24	61.90	69.05	70.24	64.81
	<i>fd3 & pca</i>	59.52	67.86	66.67	61.90	70.24	70.24	61.90	67.86	67.86	66.01
	<i>nfd3 & std</i>	63.10	65.48	60.71	59.52	61.90	59.52	61.90	64.29	65.48	62.43
	<i>nfd3 & pca</i>	64.29	65.48	61.90	63.10	64.29	63.10	59.52	61.90	64.29	63.10
	Ortalama	61.31	66.47	64.09	61.71	66.67	66.47	61.90	66.27	66.87	64.64

Bu tablo incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Veri setlerinin sınıflandırma sonuçlarının sensör referans değer noktalarına göre ortalamaları alındığında, sensör referans değeri-1'in daha başarılı olduğu görülmüştür.
- Sensör referans değeri-1 için en başarılı ön-işleme ve boyut azaltım yönteminin, ortalamalar dikkate alındığında, *d1&pca* olduğu görülmüştür.
- Ölçeklemenin sınıflandırma başarısını etkilediği görülmüştür.
- En yüksek sınıflandırma başarı oranı; z-skor ölçekleme yöntemine göre *d1&pca* yöntemi ve $k = 5$ değeri için %73.81 olarak elde edilmiştir.

2.4.2. DA ile Sınıflandırma

DA yöntemi kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.16'da görülmektedir. Ölçekleme yapılmış ve yapılmamış veri setlerinin sınıflandırma sonuçları aynı olduğu için, burada ayrıca verilmemiştir. Bu tablodaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.16. Veri setlerinin DA yöntemi kullanılarak 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Doğrusal	Diyagonal doğrusal	İkinci dereceden	Diyagonal ikinci dereceden	Mahalanobis	Ortalama
Sensör Referans değeri - 1	<i>d1 & std</i>	66.67	58.33	58.33	55.95	53.57	58.57
	<i>d1 & pca</i>	69.05	69.05	57.14	70.24	54.76	64.05
	<i>fd1 & std</i>	64.29	57.14	64.29	57.14	59.52	60.48
	<i>fd1 & pca</i>	69.05	71.43	66.67	64.29	61.90	66.67
	<i>nfd1 & std</i>	67.86	60.71	59.52	57.14	54.76	60.00
	<i>nfd1 & pca</i>	57.14	55.95	64.29	59.52	61.90	59.76
	Ortalama	65.67	62.10	61.71	60.71	57.74	61.59
Sensör Referans değeri - 2	<i>d2 & std</i>	52.38	42.86	59.52	41.67	58.33	50.95
	<i>d2 & pca</i>	57.14	54.76	58.33	60.71	59.52	58.10
	<i>fd2 & std</i>	51.19	40.48	59.52	39.29	48.81	47.86
	<i>fd2 & pca</i>	55.95	53.57	59.52	58.33	51.19	55.71
	<i>nfd2 & std</i>	57.14	55.95	51.19	58.33	55.95	55.71
	<i>nfd2 & pca</i>	51.19	57.14	61.90	58.33	59.52	57.62
	Ortalama	54.17	50.79	58.33	52.78	55.56	54.33
Sensör Referans değeri - 3	<i>d3 & std</i>	66.67	57.14	65.48	52.38	58.33	60.00
	<i>d3 & pca</i>	67.86	64.29	63.10	69.05	58.33	64.52
	<i>fd3 & std</i>	63.10	60.71	64.29	57.14	59.52	60.95
	<i>fd3 & pca</i>	63.10	63.10	63.10	63.10	59.52	62.38
	<i>nfd3 & std</i>	61.90	61.90	53.57	54.76	51.19	56.67
	<i>nfd3 & pca</i>	66.67	65.48	59.52	60.71	57.14	61.90
	Ortalama	64.88	62.10	61.51	59.52	57.34	61.07

Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Veri setlerinin sınıflandırma sonuçlarının sensör referans değer noktalarına göre ortalamaları alındığında, sensör referans değeri-1'in daha başarılı olduğu görülmüştür.
- Sensör referans değeri-1 için en başarılı ön-işleme ve boyut azaltım yönteminin, ortalamalar, dikkate alındığında *fd1&pca* olduğu görülmüştür.
- En yüksek sınıflandırma başarı oranı; diyagonal doğrusal fonksiyon kullanılarak, *fd1&pca* yöntemi ile %71.43 olarak elde edilmiştir.

2.4.3. SVM ile Sınıflandırma

MATLAB'deki yerleşik SVM fonksiyonları kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırılması ile ilgili sonuçlar Tablo 2.17'de görülmektedir. Ölçekleme yapılmış ve yapılmamış veri setlerinin sınıflandırma sonuçları aynı olduğu için, burada ayrıca verilmemiştir. Bu tablodaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.17. Veri setlerinin MATLAB'deki yerleşik SVM fonksiyonları ile 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%).

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	SVM (1-r)				SVM (1-1)				Ortalama
		Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	Doğrusal	İkinci dereceden polinom	Üçüncü dereceden polinom	RBF	
Sensör Referans değeri - 1	<i>d1 & std</i>	41.93	47.40	48.80	46.69	61.67	60.94	65.31	64.12	54.61
	<i>d1 & pca</i>	50.51	57.76	62.10	58.30	66.56	69.62	75.36	73.22	64.18
	<i>fd1 & std</i>	41.58	40.93	46.56	47.27	63.07	61.89	66.87	67.47	54.45
	<i>fd1 & pca</i>	50.12	62.31	62.36	59.83	64.70	65.68	72.34	70.01	63.42
	<i>nfd1 & std</i>	33.78	48.57	50.35	53.06	64.61	61.18	60.09	62.27	54.24
	<i>nfd1 & pca</i>	42.89	58.40	62.93	53.70	61.12	67.07	68.01	68.08	60.27
	Ortalama	43.47	52.56	55.52	53.14	63.62	64.40	67.99	67.53	58.53
Sensör Referans değeri - 2	<i>d2 & std</i>	29.90	37.08	38.79	29.07	56.89	54.28	54.43	53.85	44.29
	<i>d2 & pca</i>	38.97	50.30	58.30	47.57	61.99	56.29	57.25	59.79	53.81
	<i>fd2 & std</i>	28.12	30.22	39.05	27.91	51.25	51.11	54.38	47.91	41.24
	<i>fd2 & pca</i>	38.02	49.02	56.64	48.32	62.19	59.90	61.14	56.76	54.00
	<i>nfd2 & std</i>	29.25	38.66	46.06	38.19	57.74	58.44	55.98	56.07	47.55
	<i>nfd2 & pca</i>	27.79	45.27	44.57	36.98	63.25	63.88	57.29	57.96	49.62
	Ortalama	32.01	41.76	47.24	38.01	58.89	57.32	56.74	55.39	48.42
Sensör Referans değeri - 3	<i>d3 & std</i>	42.80	48.77	50.54	47.27	63.30	61.97	66.40	59.28	55.04
	<i>d3 & pca</i>	51.92	56.22	55.56	55.09	59.60	60.79	64.61	63.75	58.44
	<i>fd3 & std</i>	41.58	42.15	45.09	45.32	59.59	61.03	65.85	64.59	53.15
	<i>fd3 & pca</i>	52.19	58.17	58.03	56.88	64.43	61.81	63.18	62.88	59.70
	<i>nfd3 & std</i>	31.49	49.09	54.53	52.05	60.85	65.86	63.84	60.28	54.75
	<i>nfd3 & pca</i>	40.57	63.82	63.24	59.12	73.21	69.17	69.04	65.07	62.90
	Ortalama	43.43	53.04	54.50	52.62	63.49	63.44	65.49	62.64	57.33

Bu tablo incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- SVM (1-1), SVM (1-r)'ye göre daha başarılı sonuçlar vermiştir.
- En başarılı sınıflandırma fonksiyonu, ortalamalar dikkate alındığında, üçüncü dereceden polinom olarak elde edilmiştir.
- Veri setlerinin sınıflandırma sonuçlarının sensör referans değer noktalarına göre ortalamaları alındığında, sensör referans değeri-1'in daha başarılı olduğu görülmüştür.
- En yüksek sınıflandırma başarı oranı; SVM (1-1)'de üçüncü dereceden polinom kullanılarak *d1&pca* yöntemi ile %75.36 olarak elde edilmiştir.

LIBSVM kütüphanesindeki yerleşik SVM fonksiyonları kullanılarak, her bir veri setinin 5-kat CV metoduna göre sınıflandırma başarı oranları hesaplanmıştır. Kullanılan 4 farklı çekirdek fonksiyonu için elde edilen sonuçlar, Tablo 2.18'de görülmektedir. Bu tablolardaki koyu değerler, kullanılan fonksiyonlara göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.18. Veri setlerinin LIBSVM kütüphanesindeki çekirdek fonksiyonları ile 7 gruba sınıflandırılmasındaki 5-kat CV metoduna göre başarı oranları (%). Çekirdek fonksiyonları: (a) doğrusal, (b) ikinci dereceden polinom, (c) üçüncü dereceden polinom, (d) RBF.

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz		Ölçekleme [0-1]		Ölçekleme [z-skor]		Doğrusal
		$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	$\log_2 C$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	16.50	66.67	14.00	66.67	17.00	65.48	66.27
	<i>d1 & pca</i>	16.25	66.67	11.50	65.48	-0.50	67.86	66.67
	<i>fd1 & std</i>	14.50	66.67	9.25	69.05	5.00	69.05	68.25
	<i>fd1 & pca</i>	16.25	65.48	7.50	71.43	7.25	70.24	69.05
	<i>nfd1 & std</i>	13.75	66.67	10.00	69.05	6.50	67.86	67.86
	<i>nfd1 & pca</i>	16.25	70.24	11.25	71.43	6.75	70.24	70.63
	Ortalama		67.06		68.85		68.45	68.12
Sensör Referans değeri - 2	<i>d2 & std</i>	16.25	58.33	15.50	63.10	11.00	63.10	61.51
	<i>d2 & pca</i>	15.75	63.10	6.25	64.29	7.75	60.71	62.70
	<i>fd2 & std</i>	16.50	58.33	13.00	65.48	10.25	64.29	62.70
	<i>fd2 & pca</i>	16.75	61.90	15.00	63.10	10.25	63.10	62.70
	<i>nfd2 & std</i>	9.50	63.10	9.00	63.10	4.25	61.90	62.70
	<i>nfd2 & pca</i>	9.50	63.10	9.75	63.10	3.75	63.10	63.10
	Ortalama		61.31		63.69		62.70	62.57
Sensör Referans değeri - 3	<i>d3 & std</i>	16.25	66.67	6.50	66.67	2.25	66.67	66.67
	<i>d3 & pca</i>	17.00	67.86	5.00	72.62	0.00	71.43	70.63
	<i>fd3 & std</i>	16.75	65.48	9.75	66.67	11.75	64.29	65.48
	<i>fd3 & pca</i>	16.50	65.48	2.75	69.05	0.25	65.48	66.67
	<i>nfd3 & std</i>	15.50	64.29	5.00	65.48	0.75	64.29	64.68
	<i>nfd3 & pca</i>	17.00	69.05	17.00	67.86	3.00	69.05	68.65
	Ortalama		66.47		68.06		66.87	67.13

(a)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			İkinci dereceden polinom
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	17.00	-1.50	66.67	15.00	4.00	77.38	17.00	-2.25	76.19	73.41
	<i>d1 & pca</i>	17.00	-1.50	66.67	13.25	2.25	80.95	17.00	-3.50	80.95	76.19
	<i>fd1 & std</i>	16.50	-3.00	66.67	17.00	0.00	72.62	6.50	1.25	78.57	72.62
	<i>fd1 & pca</i>	16.50	-2.75	66.67	16.75	4.00	79.76	15.25	-1.50	79.76	75.40
	<i>nfd1 & std</i>	11.00	3.75	70.24	11.50	-2.50	70.24	8.75	-0.50	70.24	70.24
	<i>nfd1 & pca</i>	15.00	1.75	72.62	17.00	3.00	76.19	15.00	-4.00	72.62	73.81
	Ortalama			68.25			76.19			76.39	73.61
Sensör Referans değeri - 2	<i>d2 & std</i>	17.00	4.75	61.90	14.00	-3.75	63.10	9.50	1.25	65.48	63.49
	<i>d2 & pca</i>	16.75	3.75	65.48	15.00	-4.75	66.67	8.00	-3.00	69.05	67.06
	<i>fd2 & std</i>	17.00	5.00	61.90	12.25	-3.75	69.05	13.00	-8.75	65.48	65.48
	<i>fd2 & pca</i>	17.00	5.00	67.86	14.50	-2.25	69.05	9.00	-4.50	69.05	68.65
	<i>nfd2 & std</i>	7.00	1.50	66.67	4.50	2.00	66.67	4.50	-2.00	66.67	66.67
	<i>nfd2 & pca</i>	12.25	2.50	65.48	14.75	1.50	63.10	10.75	-2.00	65.48	64.68
	Ortalama			64.88			66.27			66.87	66.01
Sensör Referans değeri - 3	<i>d3 & std</i>	17.00	-0.75	66.67	17.00	1.25	67.86	15.00	-2.00	69.05	67.86
	<i>d3 & pca</i>	17.00	1.00	70.24	17.00	-12.75	72.62	15.00	-16.00	71.43	71.43
	<i>fd3 & std</i>	15.00	0.00	66.67	9.75	2.00	76.19	15.00	0.00	75.00	72.62
	<i>fd3 & pca</i>	17.00	1.00	66.67	15.00	2.50	73.81	17.00	-2.00	75.00	71.83
	<i>nfd3 & std</i>	11.50	3.75	71.43	11.00	2.00	73.81	2.50	1.75	73.81	73.02
	<i>nfd3 & pca</i>	15.00	2.75	78.57	7.00	2.50	79.76	3.00	-2.25	78.57	78.97
	Ortalama			70.04			74.01			73.81	72.62

(b)

(Tablo 2.18'in devamı)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			Üçüncü dereceden polinom
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	17.00	-3.00	66.67	17.00	1.50	75.00	9.00	0.50	73.81	71.83
	<i>d1 & pca</i>	17.00	-2.25	66.67	13.75	1.50	83.33	16.75	-4.00	84.52	78.17
	<i>fd1 & std</i>	15.25	-2.50	67.86	12.50	1.25	75.00	11.50	-3.25	76.19	73.02
	<i>fd1 & pca</i>	17.00	0.75	66.67	17.00	2.25	78.57	8.00	0.25	79.76	75.00
	<i>nfd1 & std</i>	15.50	-3.25	67.86	13.00	3.00	69.05	12.75	-8.25	70.24	69.05
	<i>nfd1 & pca</i>	17.00	0.00	72.62	17.00	-3.75	76.19	7.75	-3.75	76.19	75.00
	Ortalama			68.06			76.19			76.79	73.68
Sensör Referans değeri - 2	<i>d2 & std</i>	16.75	-2.00	59.52	7.00	5.00	65.48	4.75	1.00	65.48	63.49
	<i>d2 & pca</i>	17.00	1.75	65.48	17.00	3.25	67.86	17.00	1.00	70.24	67.86
	<i>fd2 & std</i>	15.00	3.50	61.90	12.75	-4.75	67.86	13.50	-9.75	65.48	65.08
	<i>fd2 & pca</i>	17.00	4.50	67.86	14.50	-2.75	69.05	13.25	-6.25	69.05	68.65
	<i>nfd2 & std</i>	7.00	0.50	67.86	0.00	2.75	66.67	5.25	-3.25	66.67	67.06
	<i>nfd2 & pca</i>	11.75	2.50	66.67	2.50	1.00	64.29	7.50	-0.50	66.67	65.87
	Ortalama			64.88			66.87			67.26	66.34
Sensör Referans değeri - 3	<i>d3 & std</i>	17.00	-1.00	66.67	10.25	4.00	71.43	17.00	-2.75	69.05	69.05
	<i>d3 & pca</i>	17.00	0.25	70.24	17.00	4.75	72.62	13.00	-0.25	73.81	72.22
	<i>fd3 & std</i>	16.75	-2.00	66.67	11.75	2.75	73.81	12.25	-4.50	72.62	71.03
	<i>fd3 & pca</i>	17.00	0.25	66.67	17.00	1.00	77.38	17.00	-4.00	76.19	73.41
	<i>nfd3 & std</i>	9.75	3.25	70.24	5.00	2.75	72.62	8.75	-1.50	72.62	71.83
	<i>nfd3 & pca</i>	17.00	0.50	78.57	16.00	-3.50	79.76	3.75	-3.25	79.76	79.37
	Ortalama			69.84			74.60			74.01	72.82

(c)

5-kat CV	Veri seti (ön-işleme & boyut azaltım)	Ölçeklemesiz			Ölçekleme [0-1]			Ölçekleme [z-skor]			RBF
		$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	$\log_2 C$	$\log_2 \gamma$	% Başarı	
Sensör Referans değeri - 1	<i>d1 & std</i>	17.00	-2.25	66.67	13.75	3.25	73.81	14.00	-1.25	73.81	71.43
	<i>d1 & pca</i>	17.00	-1.75	66.67	17.00	0.50	85.71	17.00	-4.25	85.71	79.37
	<i>fd1 & std</i>	16.25	4.75	66.67	14.25	1.75	75.00	14.75	-2.75	75.00	72.22
	<i>fd1 & pca</i>	17.00	-1.75	65.48	14.50	3.75	80.95	12.50	-1.25	80.95	75.79
	<i>nfd1 & std</i>	12.25	0.50	70.24	13.25	-3.25	70.24	14.75	-9.25	69.05	69.84
	<i>nfd1 & pca</i>	17.00	5.00	76.19	8.75	0.25	75.00	4.25	-0.25	77.38	76.19
	Ortalama			68.65			76.79			76.98	74.14
Sensör Referans değeri - 2	<i>d2 & std</i>	17.00	3.75	60.71	15.25	-5.25	61.90	12.75	-2.50	64.29	62.30
	<i>d2 & pca</i>	16.75	3.50	66.67	17.00	3.00	67.86	6.50	0.75	69.05	67.86
	<i>fd2 & std</i>	14.00	5.00	63.10	14.25	-5.50	66.67	13.50	-9.25	66.67	65.48
	<i>fd2 & pca</i>	17.00	4.50	66.67	17.00	-4.75	67.86	3.00	-2.00	69.05	67.86
	<i>nfd2 & std</i>	9.75	-1.00	65.48	12.00	-4.00	64.29	7.00	-4.25	63.10	64.29
	<i>nfd2 & pca</i>	17.00	0.50	66.67	15.00	3.50	64.29	17.00	-4.75	66.67	65.87
	Ortalama			64.88			65.48			66.47	65.61
Sensör Referans değeri - 3	<i>d3 & std</i>	17.00	-0.50	66.67	13.50	-0.75	69.05	17.00	2.75	67.86	67.86
	<i>d3 & pca</i>	17.00	1.00	70.24	17.00	3.00	72.62	17.00	-0.25	73.81	72.22
	<i>fd3 & std</i>	15.50	-0.50	66.67	15.00	-1.75	71.43	15.00	-6.25	71.43	69.84
	<i>fd3 & pca</i>	17.00	1.00	66.67	9.00	3.00	75.00	17.00	-1.00	70.24	70.63
	<i>nfd3 & std</i>	11.00	4.50	73.81	8.75	2.25	72.62	9.00	-1.50	72.62	73.02
	<i>nfd3 & pca</i>	14.50	2.25	76.19	16.75	-3.50	79.76	11.00	-4.25	76.19	77.38
	Ortalama			70.04			73.41			72.02	71.83

(d)

Bu tablolar incelenerek, aşağıdaki sonuçlar çıkarılmıştır:

- Tüm fonksiyonlarda, sensör referans değeri-1 ve sensör referans değeri-3'ün başarılı olduğu görülmüştür.

- Tüm fonksiyonlarda, en başarılı sonuçların PCA yöntemi ile sağlandığı görülmüştür.
- Ölçeklemenin sınıflandırma başarısını etkilediği görülmüştür.
- Fonksiyonlara göre en yüksek sınıflandırma başarı oranları;
 - Doğrusal fonksiyon için; $d3\&pca$ yöntemlerinde %72.62
 - İkinci dereceden polinom için; $d1\&pca$ yönteminde %80.95
 - Üçüncü dereceden polinom için; $d1\&pca$ yönteminde %84.52
 - RBF fonksiyonu için; $d1\&pca$ yönteminde %85.71
 olarak elde edilmiştir.

MATLAB ve LIBSVM kütüphanesi kullanılarak gerçekleştirilen 5-kat CV metoduna göre SVM ile sınıflandırmada, çekirdek fonksiyonları için hesaplanan en yüksek başarı oranları Tablo 2.19’da verilmiştir. Bu tablodaki koyu değerler, kullanılan yöntemlere göre en yüksek sınıflandırma başarı oranlarını belirtmektedir.

Tablo 2.19. Veri setlerinin 5-kat CV metoduna göre SVM ile sınıflandırılmasında, çekirdek fonksiyonları için hesaplanan en yüksek başarı oranları (%).

	MATLAB		LIBSVM		
	SVM (1-r)	SVM (1-1)	Ölçeklemesiz	Ölçekleme [0 1]	Ölçekleme [z-skor]
Doğrusal	52.19	73.21	70.24	72.62	71.43
İkinci dereceden polinom	63.82	69.62	78.57	80.95	80.95
Üçüncü dereceden polinom	63.24	75.36	78.57	83.33	84.52
RBF	59.83	73.22	76.19	85.71	85.71

Bu tablodan, LIBSVM ile gerçekleştirilen sınıflandırma uygulamalarının MATLAB ortamında hazırlananlardan daha başarılı olduğu görülmektedir. Bunun nedeni; LIBSVM ortamında hazırlanan programların, her veri seti için çekirdek fonksiyonlarına göre en başarılı sınıflandırma sonucunu sağlayan (C, γ) çiftini belirlemesinden dolayıdır.

2.4.4. Sınıflandırma Yöntemlerinin Sonuçlarının Karşılaştırılması

Kullanılan sınıflandırma yöntemlerine göre 5-kat CV metodu ile hesaplanan en yüksek başarı oranları ve bu oranları sağlayan alt metot, ölçekleme, ön-işleme ve boyut azaltım yöntemleri Tablo 2.20’de verilmiştir. Burada sınıflandırma yöntemleri arasında en yüksek başarıyı; RBF çekirdek fonksiyonu kullanılarak %85.71 ile LIBSVM yazılımının sağladığı görülmektedir. k -NN ve LIBSVM ile yapılan sınıflandırmalarda en yüksek başarılar, ölçekleme yöntemleri kullanılarak elde edilmiştir. Tablo genelinde, $d1$ ön-işleme yöntemi başarılı sonuçları sağlamıştır. Bu yöntem; sensör cevap modeli olarak fark modelini ($x_i = R_i^{\max} - R_i^0 = \Delta R_i$) ve sensör referans değeri-1’i ($R_i^0 = R_i^{\min}$) temsil etmektedir. Ayrıca tüm sınıflandırıcılarda, boyut azaltımında PCA yönteminin en başarılı sonuçları verdiği görülmüştür.

Tablo 2.20. Mikroorganizma seti-3 için sınıflandırma yöntemlerine göre 5-kat CV metodu ile elde edilen en yüksek başarı oranları ve bu oranları sağlayan çeşitli yöntemler.

Yöntem	<i>k</i> -NN	DA	MATLAB SVM	LIBSVM
Ön-işleme*	<i>d1</i>	<i>fd1</i>	<i>d1</i>	
Boyut azaltım	PCA			
Ölçekleme	z-skor ölçekleme	-		[0-1] ölçekleme z-skor ölçekleme
Metot	<i>k</i> = 5	Diyagonal doğrusal	SVM (1-1) ve üçüncü dereceden polinom	RBF
Sınıflandırma başarıları (%)	73.81	71.43	75.36	85.71

(*) Ön-işleme kısaltmaları; sensör cevap modeli ve sensör referans değer noktasını tanımlamaktadır. Eşitlik ifadeleri için Tablo 1.6’ya bakınız.

3. BÖLÜM

SONUÇLAR ve ÖNERİLER

3.1. Sonuçlar

Enfeksiyon hastalıklarının tedavisine hızlı bir şekilde başlanabilmesi, hastalık etkeni patojen mikroorganizmanın, en kısa sürede doğru olarak tanımlanmasına bağlıdır. Bazı durumlarda patojen mikroorganizmaların tanımlanması; zahmetli ve zaman alıcı olabilmektedir. Bundan dolayı bu tez çalışmasında, enfeksiyonlara yol açan mikroorganizmaların daha kısa sürede tanımlanması için elektronik burun cihazının kullanımı araştırılmıştır.

Bu tez kapsamında; üç ayrı mikroorganizma seti üzerinde çalışmalar yapılmıştır. Mikroorganizma setlerinden elektronik burun cihazı ile elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması gerçekleştirilmiştir. Mikroorganizma setlerinden elde edilen koku verilerine, sınıflandırma performansına etkilerini incelemek amacıyla, farklı veri ön-işleme, boyut azaltım ve ölçekleme yöntemleri uygulanmıştır. Bu yöntemlerden sonra sınıflandırma aşamasına geçilmiş ve k -NN, DA ve SVM yöntemleri kullanılmıştır. Ayrıca her sınıflandırma yönteminin alt metotları da araştırılmış ve performansları karşılaştırılmıştır.

Mikroorganizma setlerinden birincisi; çeşitli enfeksiyon hastalıklarına yol açan ve rutinde sıklıkla karşılaşılan 8 mikroorganizmadan oluşmaktadır. Bu çalışmada, her mikroorganizmanın petri kaplarında hazırlanan kültürlerinden elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması gerçekleştirilmiştir. Bu mikroorganizma seti için, kullanılan sınıflandırma yöntemleri arasında en yüksek başarıyı, %82.81 ile SVM'in sağladığı görülmüştür. Bu sonuç, LIBSVM yazılımında RBF çekirdek fonksiyonu kullanılarak elde edilmiştir.

Mikroorganizma setlerinden ikincisi, diş-kök kanal enfeksiyonlarına sebep olan 7 mikroorganizmadan oluşmaktadır. Her mikroorganizmanın serum fizyolojik içeren 4 farklı konsantrasyonda (12×10^8 , 12×10^5 , 12×10^3 ve 12×10^1 CFU/mL) kültürleri hazırlanmıştır. Bu çalışma ile aynı konsantrasyondaki kültürlerden elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması gerçekleştirilmiştir. Bu sayede farklı konsantrasyonlardaki etkinlik test edilmiştir. Elde edilen sınıflandırma sonuçları incelendiğinde; 12×10^8 ve 12×10^5 CFU/mL konsantrasyonlarına ait başarıların, diğer konsantrasyonlara göre daha yüksek olduğu görülmüştür. Bu iki konsantrasyon için; kullanılan sınıflandırma yöntemleri arasında en yüksek başarıyı,

- 12×10^8 CFU/mL konsantrasyonda, LIBSVM yazılımının doğrusal hariç diğer çekirdek fonksiyonları ile %97.14,
- 12×10^5 CFU/mL konsantrasyonda, doğrusal fonksiyon ile DA yönteminin %100 olarak sağladığı görülmüştür.

Mikroorganizma setlerinden üçüncüsünde de, ikinci setteki mikroorganizmalar kullanılmıştır. Bu çalışmada ise, diş hekimliğindeki uygulamalara benzer bir durum oluşturabilmek için, her mikroorganizmanın steril diş kökleri ve serum fizyolojik içeren sabit konsantrasyonda kültürleri hazırlanmıştır. Bu kültürlerden elde edilen koku verilerinin, mikroorganizma türüne göre sınıflandırılması gerçekleştirilmiştir. Bu mikroorganizma seti için, kullanılan sınıflandırma yöntemleri arasında en yüksek başarıyı, %85.71 ile SVM'in sağladığı görülmüştür. Bu sonuç, LIBSVM yazılımında RBF çekirdek fonksiyonu kullanılarak elde edilmiştir.

Çalışılan mikroorganizma setlerinin sınıflandırma sonuçları incelendiğinde, ikinci setin başarı oranlarının diğer setlere nazaran daha yüksek olduğu görülmektedir. Bunun sebebi; ikinci sette sabit konsantrasyonda her bir mikroorganizma örneğinden (numunesinden), 5 koku verisi alınmış olmasından kaynaklanmaktadır. Diğer setlerde ise, her bir mikroorganizma için hazırlanan örneklerden birer koku verisi alınmıştır. Dolayısıyla, aynı mikroorganizmaya ait örneklerin koku verileri arasındaki küçük değişimler, bu setlerdeki sınıflandırma başarılarını negatif yönde etkilemiştir.

Mikroorganizma setleri; veri ön-işleme, boyut azaltım, ölçekleme ve sınıflandırma yöntemleri açılarından incelendiğinde, genel olarak aşağıdaki sonuçlar çıkarılabilir:

- Veri ön-işlemede $d1$ ve $d3$ yöntemlerinin başarılı olduğu görülmüştür. Bu yöntemler sensör cevap modeli ve sensör referans direnç değer noktası olarak;
 - $d1$ yöntemi: Sensör cevap modeli olarak fark modelini ($x_i = R_i^{\max} - R_i^o = \Delta R_i$) ve sensör referans değeri-1'i ($R_i^o = R_i^{\min}$)
 - $d3$ yöntemi: Sensör cevap modeli olarak fark modelini ($x_i = R_i^{\max} - R_i^o = \Delta R_i$) ve sensör referans değeri-3'ü ($R_i^o = R_i^{\text{fbp}}$) temsil etmektedir.
- Boyut azaltımında, PCA yöntemi uygulanan veri setlerinin sınıflandırılmasında daha başarılı sonuçlar elde edilmiştir.
- Ölçeklemenin, k -NN ve LIBSVM ile gerçekleştirilen sınıflandırma uygulamalarında daha başarılı sonuçlar sağladığı görülmüştür.
- Sınıflandırma yöntemleri açısından incelendiğinde; genellikle SVM yönteminin diğer sınıflandırıcılara nazaran daha başarılı sonuçlar verdiği görülmüştür. MATLAB ve LIBSVM ortamlarında hazırlanan SVM programlarının sonuçları ile ilgili aşağıdaki değerlendirmelere varılmıştır:
 - MATLAB ortamında yazılan SVM programlarının sonuçları incelendiğinde; SVM (1-1) yönteminin SVM (1-r) yöntemine göre daha başarılı sınıflandırma sağladığı anlaşılmaktadır.
 - LIBSVM yazılım paketi ile gerçekleştirilen sınıflandırma uygulamalarının ise MATLAB ortamında yazılan SVM programlarından daha başarılı olduğu gözlemlenmiştir. Bunun nedeni; LIBSVM ortamında hazırlanan programların, her veri seti için çekirdek fonksiyonlarına göre en başarılı sınıflandırma sonucunu sağlayan (C, γ) çiftini belirlemesinden dolayıdır.
 - LIBSVM ile yapılan sınıflandırma uygulamalarında en başarılı sonuçlar RBF çekirdek fonksiyonu kullanılarak elde edilmiştir.

Literatürde elektronik burun cihazı kullanılarak mikroorganizmaların sınıflandırılması üzerine yapılmış çeşitli çalışmalar bulunmaktadır. Bu çalışmalardan bazılarının; konuları, kullandıkları sınıflandırma yöntemleri ve elde ettikleri başarı oranları Tablo 3.1'de verilmiştir. Bu çalışmalar incelendiğinde, sınıflandırmada genellikle ANN tabanlı yöntemlerin tercih edildiği görülmektedir. Bu tez çalışmasında, üç ayrı mikroorganizma seti üzerinde k -NN, DA ve SVM yöntemlerinin sınıflandırma performansları incelenmiştir. Genellikle RBF çekirdek fonksiyonu kullanılan SVM

yöntemi, (C, γ) parametreleri uygun değerlerde seçildiğinde, %80'in üzerinde sınıflandırma başarısı ile k -NN ve DA sınıflandırıcılarına nazaran daha iyi sonuçlar vermiştir.

Tablo 3.1. Literatürdeki elektronik burun cihazı ile mikroorganizmaların sınıflandırılması üzerine yapılmış olan çalışmalardan bazılarının; konuları, kullandıkları sınıflandırma yöntemleri ve elde ettikleri başarı oranları.

Çalışmanın konusu	Kullanılan yöntem ve sınıflandırma başarısı
Kan ve idrar örneklerindeki bakterilerin elektronik burun cihazı kullanılarak belirlenmesi [8].	➤ Kan örneklerinde Sammon haritalama ile %100 ➤ İdrar örneklerinde zaman serileri modeli ile %80
Göz enfeksiyonlarına sebep olan 6 farklı bakteri türünün sınıflandırılması [10].	➤ MLP ile %75 ➤ PNN ile %94 ➤ SOM ile %96 ➤ RBF ile %98
Kulak-burun-boğaz ve göz enfeksiyonlarına sebep olan bakterilerin sınıflandırılması [9].	➤ Göz enfeksiyonları ile ilgili 6 bakteride; • MLP kullanılarak %97.3 • RBF kullanılarak %92.8 ➤ Kulak-burun-boğaz enfeksiyonları ile ilgili 4 bakteride MLP kullanılarak %97.6
Kulak-burun-boğaz enfeksiyonlarına sebep olan <i>Staphylococcus aureus</i> bakterisinin metisiline dirençli ve duyarlı (MRSA ve MSSA) suşları ile koagülaz negatif stafilokokların (C-NS) hastane ortamında tanımlanması [4].	➤ MLP ile %75 ➤ PNN ile %96 ➤ RBF ile %99.69
Elektronik burun cihazı kullanılarak 12 farklı bakteri türü ve 1 patojen mayanın yapay sinir ağları ile sınıflandırılması [3].	➤ MLP ile %93.4
Klinik olarak 10 önemli mikroorganizmanın elektronik burun cihazı kullanılarak tanımlanması [12].	➤ ANN ile %100 ➤ k-NN ile %100

Sonuç olarak; literatürdeki çalışmalar ile tez kapsamında gerçekleştirilen sınıflandırma çalışmaları karşılaştırıldığında, SVM yönteminin sonuçlarının tatmin edici olduğu görülmüştür. Ayrıca bu tez ile sınıflandırma performansına etkisi bulunan başarılı veri ön-işleme, boyut azaltım ve ölçekleme yöntemleri belirlenmiştir. Bu tez çalışmasındaki

sınıflandırma başarıları, hastalık etkeni patojen mikroorganizmaların tespitinde, elektronik burun cihazının rutin mikrobiyolojik tekniklere alternatif veya destekleyici bir yöntem olarak kullanılabileceğini göstermiştir. Özellikle, dış kök kanal enfeksiyonlarına sebep olan mikroorganizmaların sınıflandırılması üzerine yapılan çalışmalar, elektronik burun cihazının bu alanda kullanımının başarılı ve ümit verici olduğunu göstermiştir.

3.2. Öneriler

İleriki çalışmalarda, bu çalışma ile belirlenen başarılı ön-işleme teknikleri uygulandıktan sonra, boyut azaltımı için öznitelik çıkarımı (*feature extraction*) yerine öznitelik alt kümesi seçme (*feature subset selection*) yöntemleri kullanılabilir. Böylece gerçekleştirilen uygulamalara özgü önemli sensörler belirlenebilir. Belirlenen bu sensörler temin edilerek, kullanılan cihazdan bağımsız, daha az sensör içeren, uygulamaya özgü bir sistem oluşturulabilir.

İleride daha fazla veri üzerinde çalışma yapılarak test için ayrı veri setleri oluşturulabilir. Çapraz geçerlilik ile belirlenen sınıflandırıcıya özgü parametreler, test veri setinde kullanılarak sınıflandırma başarıları hesaplanabilir. Bu sayede sınıflandırma sonuçlarının güvenilirliği artırılabilir.

Bu tez çalışmasındaki mikroorganizmaların sınıflandırılması üzerine yürütülen başarılı uygulamalardan sonra, gelecekte klinik ortamda veriler alınması planlanmaktadır. Özellikle diş hekimliği alanında, doğrudan hastanın dış kök kanal kokusuna göre patojen mikroorganizmanın tespiti üzerine çalışılması hedeflenmektedir.

KAYNAKLAR

1. Akın, L., 2001. İnfeksiyon hastalıklarının tanısında laboratuvarın rolü, s. 37. Güncel Bilgiler Işığında İnfeksiyon Hastalıkları, Bilimsel Tıp Yayınevi, Ankara.
2. Pavlou, A. K., Magan, N., Sharp, D., Brown, J., Barr, H., Turner, A. P. F., 2000. An intelligent rapid odour recognition model in discrimination of helicobacter pylori and other gastroesophageal isolates in vitro. **Elsevier Biosensors & Bioelectronics**, **15**: 333-342.
3. Gibson, T. D., Prosser, O., Hulbert, J. N., Marshall, R. W., Corcoran, P., 1997. Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose. **Elsevier Sensors and Actuators**, **44**: 413-422.
4. Dutta, R., Morgan, D., Baker, N., Gardner, J. W., Hines, E. L., 2005. Identification of Staphylococcus aureus infections in hospital environment: electronic nose based approach. **Elsevier Sensors and Actuators B**, **109**: 355-362.
5. Stetter, J. R., Penrose, W. R., 2002. Understanding chemical sensors and chemical sensor arrays (electronic noses): past, present and future. **Sensors Update**, **10**: 189-229.
6. Lai, S. Y., et al., 2002. Identification of upper respiratory bacterial pathogens with the electronic nose, **Laryngoscope**, **112** (6): 975-979.
7. Pavlou, A. K., Magan, N., McNulty, C., Jones, J. M., Sharp, D., Brown, J., Turner, A. P. F., 2002. Use of an electronic nose system for diagnoses of urinary tract infections. **Elsevier Biosensors and Bioelectronics**, **17**: 893-899.
8. Yates, J. W. T., et al., 2005. Data reduction in headspace analysis of blood and urine samples for robust bacterial identification. **Elsevier Computer Methods and Programs in Biomedicine**, **79**: 259-271.
9. Boilot, P., Hines, E. L., Gardner, J. W., Pitt, R., John, S., Mitchell, J., Morgan, D. W., 2002. Classification of bacteria responsible for ENT and eye infections using the Cyranose system. **IEEE Sensors Journal**, **2** (3): 247-253.
10. Dutta, R., Hines, E. L., Gardner, J. W., Boilot, P., 2002. Bacteria classification using Cyranose 320 electronic nose. **BioMedical Engineering OnLine**, 1-4.

11. McEntegart, C. M., Penrose, W. R., Strathmann, S., Stetter, J. R., 2000. Detection and discrimination of coliform bacteria with gas sensor arrays. **Elsevier Sensors and Actuators B**, **70**: 170-176.
12. Moens, M., Smet, A., Naudts, B., et. al., 2006. Fast identification of ten clinically important micro-organisms using an electronic nose. **Letters in Applied Microbiology**, **42**: 121-126.
13. Pavlou, A., Turner, A. P. F., Magan, N., 2002. Recognition of anaerobic bacterial isolates in vitro using electronic nose technology. **Letters in Applied Microbiology**, **35**: 366-369.
14. Yamada, Y., Takahashi, Y., Konishi, K., Katsuumi, I., 2007. Association of odor from infected root canal analyzed by an electronic nose with isolated bacteria. **Journal of Endodontics**, **33** (9): 1106-1109.
15. Nagle H. T., Schiffman S. S., Gutierrez-Osuna R., 1998. The how and why of electronic noses. **IEEE Spectrum September**, **35** (9), p. 22-34.
16. Schiffman, S. S., Pearce, T. C., 2004. Introduction to olfaction: perception, anatomy, physiology, and molecular biology, ch1. *In*: Handbook of Machine Olfaction: Electronic Nose Technology, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG.
17. Cyranose 320: <http://www.ideo.com/work/cyranose-320/> (Date accessed: June 2011).
18. Persaud, K., Dodd, G., 1982. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. **Nature**, **299**: 352–355.
19. Stetter, J. R., Zaromb, S., Penrose, W. R., Findlay, M. W., Otagawa, T., Sincali, A. J., 1984. Portable device for detecting and identifying hazardous vapors, pp. 183–194. *In*: Hazardous Material Spills Conference Proceedings, Nashville, TN.
20. Ikegami, A., Kaneyasu, M., 1985. Olfactory detection using integrated sensors, pp. 136–139. *Proceeding of the 3rd International Conference on Solid-State Sensors and Actuators (Transducers' 85) Philadelphia PA, USA*.
21. James, D., Scott, S. M., Zulfiquir, A., O'Hare, W. T., 2005. Chemical sensors for electronic nose systems. **Microchimica Acta**, **149**: 1-17.

22. O'Connell, M., Valdora, G., Peltzer, G., Negri, R. M., 2001. A practical approach for fish freshness determinations using a portable electronic nose. **Elsevier Sensors and Actuators**, **80**: 149–154.
23. Blixt, Y., Borch, E., 1999. Using an electronic nose for determining the spoilage of vacuum-packaged beef. **Elsevier International Journal of Food Microbiology**, **46**: 123–134.
24. Martin, Y. G., Pavon, J. L. P., Cordero, B. M., Pinto, C. G., 1999. Classification of vegetable oils by linear discriminant analysis of electronic nose data. **Elsevier Analytica Chimica Acta**, **384**: 83-94.
25. Lozano, J., Santos, J. P., Sayago, I., Gutierrez, J., Horrillo, M. C., 2004. Identification of typical wine aromas by means of an electronic nose. *IEEE Conference Proceedings*, p. 345-348.
26. Dutta, R., Hines, E. L., Gardner, J. W., Kashwan, K. R., Bhuyan, M., 2003. Determination of tea quality by using a neural network based electronic nose. *IEEE Conference Proceedings*, p. 404-409.
27. Bhattacharyya, N., Tudu, B., Bandyopadhyay, R., Bhuyan, M., Mudi, R., 2004. Aroma characterization of orthodox black tea with electronic nose. *IEEE Conference Proceedings*, p. 427-430.
28. Pardo, M., Niederjaufner, G., Benussi, G., Comini, E., Faglia, G., Sberveglieri, G., Holmberg, M., Lundstrom, I., 2000. Data preprocessing enhances the classification of different brands of Espresso coffee with an electronic nose. **Elsevier Sensors and Actuators**, **69**: 397-403.
29. Pardo, M., Sberveglieri, G., 2002. Coffee analysis with an electronic nose. **IEEE Transactions on Instrumentation and Measurement**, **51** (6): 1334-1339.
30. Pardo, M., Sberveglieri, G., 1997. Electronic nose for microbial quality classification of grains. **Elsevier International Journal of Food Microbiology**, **35**: 187-193.
31. Evans, P., Persaud, K. C., McNeish, A. S., Sneath, R. W., Hobson, N., Magan, N., 2000. Evaluation of a radial basis function neural network for the determination of wheat quality from electronic nose data. **Elsevier Sensors and Actuators B**, **69**: 348–358.

32. Trihaas, J., Vognsen, L., Nielsen, P.V., 2005. Electronic nose: New tool in modeling the ripening of Danish blue cheese. **Elsevier International Dairy Journal**, **15**: 679-691.
33. Brezmez, J., Fructose, M. L. L., Llobet, E., Vilanova, X., Recasens, I., Orts, J., Saiz, G., Correig, X., 2005. Evaluation of an electronic nose to assess fruit ripeness. **IEEE Sensors Journal**, **5** (1): 97-108.
34. Brezmes, J., Llobet, E., Vilanova, X., Saiz, G., Correig, X., 2000. Fruit ripeness monitoring using an Electronic Nose. **Elsevier Sensors and Actuators B**, **69**: 223-229.
35. Gardner, J. W., Shin, H. W., Hines, E. L., Dow, C. S., 2000. An electronic nose system for monitoring the quality of potable water. **Elsevier Sensors and Actuators B**, **69**: 336-341.
36. Stuetz, R. M., Fener, R. A., Engin, G., 1999. Characterisation of wastewater using an electronic nose. **Pergamon Water Resources**, **33** (2): 442-452.
37. Baby, R. E., Cabezas, M., Walsøe de Reca, E. N., 2000. Electronic nose a useful tool for monitoring environmental contamination. **Elsevier Sensors and Actuators B**, **69**: 214-218.
38. Bender, F., Barie, N., Romoudis, G., Voigt, A., Rapp, M., 2003. Development of a preconcentration unit for a SAW sensor micro array and its use for indoor air quality monitoring. **Elsevier Sensors and Actuators B**, **93**: 135-141.
39. Natale, C. D., Macagnano, A., Martinelli, E., Paolesse, R., D'Arcangelo, G., Roscioni, C. A., Agro`, F., D'Amico, A., 2003. Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. **Elsevier Biosensors and Bioelectronics**, **18**: 1209-1218.
40. Machado, F. R., Laskowski, D., Deffenderfer, O., 2005. Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath. **American Journal of Respiratory and Critical Care Medicine**, **171**: 1286-1291.
41. Lin, Y. J., Guo, H. R., Chang, Y. H., Kao, M. T., Wang, H. H., Hong, R. I., 2001. Application of the electronic nose for uremia diagnosis. **Elsevier Sensors and Actuators B**, **76**: 177-180.
42. Lundstrom I., Shivaraman S., Svensson C., Lundkvist L., 1975. A hydrogen-sensitive MOS field-effect transistor. **Applied Physics Letters**, **26** (2): 55.

43. The Cyranose 320 Electronic Nose - User's Manual. Edition 4, June 2001. Cyrano Sciences Inc.
44. PhoenixSpec türbidimetre:
<http://www.bd.com/ds/productCenter/IS-PhoenixInstrumentationAccessories.asp> (Date accessed: June 2011)
45. Asyali M. H., Colak D., Demirkaya O., Inan M. S., 2006. Gene expression profile classification: a review. **Current Bioinformatics**, **1**: 55–73.
46. Jain, A. K., Duin, R. P. W., Mao, J., 2000. Statistical pattern recognition: A review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, **22** (1): 4-37.
47. Polikar R., 2006. Pattern Recognition, *In*: Wiley Encyclopedia of Biomedical Engineering (Eds: Akay, M.) Wiley, New York.
48. Duda, R., Hart, P., Stork, D., 2001. Pattern Classification. 2nd Edition, Wiley, NY.
49. Burges, C. J. C., 1998. A tutorial on support vector machines for pattern recognition, **Data Mining and Knowledge Discovery**, **2**, 121–167, Kluwer Academic Publishers, Boston.
50. Hearst, M. A., et al., 1998. Support vector machines. **IEEE Intelligent Systems**, **13** (4): 18–28.
51. Schölkopf, B., Smola, A. J., 2002. Learning with Kernels. MIT Press, Cambridge, MA.
52. Alpaydm, E., 2010. Introduction to Machine Learning. 2nd Edition, MIT Press.
53. Hsu, C. W., Lin, C. J., 2002. A comparison of methods for multi-class support vector machines. **IEEE Transactions on Neural Networks**, **13** (2): 415–425.
54. Chang, C. C., Lin, C. J., 2001. LIBSVM: a library for support vector machines.
<http://www.csie.ntu.edu.tw/~cjlin/libsvm> (Date accessed: June 2011).
55. Vapnik, V., 2000. The Nature of Statistical Learning Theory. 2nd Edition, Springer, Berlin.
56. Karaç, E. I., 2005. Model Selection for Multi-Class Support Vector Machines. Boğaziçi University, Institute for Graduate Studies in Science and Engineering, Master of Science, İstanbul, 59 s.
57. Prescott, L. M., Harley, J. P., Klein, D. A., 1996. Microbiology. 3rd Edition. WBC, Dubuque, IA.
58. MATLAB (The MathWorks Inc., Natick, MA) (<http://www.mathworks.com>)

59. Hsu, C. W., Chang, C. C., Lin, C. J., 2008. A practical guide to support vector classification. Department of Computer Science and Information Engineering, National Taiwan University.
60. Kara S, Aksebzeci B. H, Kodaz H, Güneş S, Kaya E, Özbilge H, 2009. Medical application of information gain-based artificial immune recognition system (IG-AIRS): Classification of microorganism species. **Expert Systems with Applications**, **36**: 5168–5172.
61. Aksebzeci B. H., Asyalı M. H., Kahraman Y., Er Ö., Kaya E., Özbilge H., Kara S., 2010. Classification of root canal microorganisms using electronic-nose and discriminant analysis. **BioMedical Engineering OnLine**, **9**:77.
62. Aksebzeci B. H., Asyalı M. H., Kahraman Y., Er Ö., Kaya E., Özbilge H., Kara S., 2010. Classification of microorganism species using electronic nose and support vector machines, *The 1st International Symposium on Computing in Science & Engineering – ISCSE 2010, June 3-5, 2010, Kusadasi, Turkey.*

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı : Bekir Hakan AKSEBZECİ
Doğum Tarihi ve Yeri : 1980, KAYSERİ
Medeni Durumu : Evli
Tel : +90 352 437 49 01 - 32237
Fax : +90 352 437 57 84
e-mail : bekirhakan@erciyes.edu.tr
Yazışma Adresi : Erciyes Üniversitesi Mühendislik Fakültesi
Elektrik - Elektronik Mühendisliği Bölümü
38039 Melikgazi/KAYSERİ

EĞİTİM

Lisans : Niğde Üniversitesi Mühendislik - Mimarlık Fakültesi
Elektrik - Elektronik Mühendisliği Bölümü, 2001.
Yüksek Lisans : Niğde Üniversitesi Fen Bilimleri Enstitüsü
Elektrik - Elektronik Mühendisliği Anabilim Dalı, 2005.