

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

SINIF DENGESİZLİĞİ İÇEREN TIBBİ VERİLERDE AKCİĞER
KANSERİ EVRE SINIFLANDIRMASI

YÜKSEK LİSANS TEZİ

Bilal Talha AYVAZ

HAZİRAN 2025

TRABZON



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**SINIF DENGESİZLİĞİ İÇEREN TIBBİ VERİLERDE AKCİĞER
KANSERİ EVRE SINIFLANDIRMASI**

Bilal Talha AYVAZ

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
"ELEKTRONİK YÜKSEK MÜHENDİSİ"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

Tezin Enstitüye Verildiği Tarih : 30 / 05 / 2025

Tezin Savunma Tarihi : 25 / 06 / 2025

Tez Danışmanı : Prof. Dr. Ayten ATASOY

Trabzon 2025

ÖNSÖZ

“Sınıf Dengesizliği İçeren Tıbbi Verilerde Akciğer Kanseri Evre Sınıflandırması” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalı, Yüksek Lisans Programı’nda hazırlanmıştır.

Tez çalışmam süresince bilgi ve tecrübeleriyle yolumu aydınlatan, bilimsel yaklaşımı ve yönlendirmeleriyle her aşamada bana rehberlik eden danışman hocam Sayın Prof. Dr. Ayten ATASOY’a en içten teşekkürlerimi sunarım.

Bu tezde kullanılan veriler, danışman hocam Sayın Prof. Dr. Ayten ATASOY’un yürütücülüğünde, TÜBİTAK’ın 3001 programı kapsamında desteklenen “Taşınabilir Elektronik Burun ile Akciğer Kanseri Hastalığının Erken Tespiti” başlıklı proje (2016–2020, Proje No: 215E380) kapsamında toplanmıştır. Bu bağlamda, söz konusu projeye sağladığı destek için TÜBİTAK’a teşekkür ederim.

Tez çalışmam sürecinde bilgi ve birikimlerini benimle paylaşarak yardımlarını esirgemeyen Sayın Öğr. Gör. Dr. Ümit ÖZSANDIKCIOĞLU’na teşekkür ederim.

Tez çalışmam boyunca manevi desteğini ve katkılarını esirgemeyen değerli arkadaşım Ömer Faruk BEKİRYAZICI’ya içtenlikle teşekkür ederim.

Hayatım boyunca, koşulsuz sevgileri ve fedakârlıklarıyla her zaman yanımda olan, desteğini hiçbir zaman esirgemeyen aileme; onlara duyduğum derin minnettarlığı ifade etmek isterim.

Bu yorucu süreci daha anlamlı ve katlanabilir kılan, her zaman yanımda olup sabrıyla ve anlayışıyla bana güç veren sevgili kız arkadaşım Aleyna Begüm TAŞDEMİR’e teşekkür ederim.

Bu çalışmanın, konuya ilgi duyan araştırmacılar için faydalı olmasını ve ilerideki çalışmalara katkı sağlamasını temenni ederim.

Bilal Talha AYVAZ

Trabzon 2025

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Sınıf Dengesizliği İçeren Tıbbi Verilerde Akciğer Kanseri Evre Sınıflandırması” başlıklı bu çalışmayı baştan sona kadar danışmanım Prof. Dr. Ayten ATASOY’un sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 25/06/2025

Bilal Talha AYVAZ

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER	V
ÖZET	VIII
SUMMARY	IX
ŞEKİLLER DİZİNİ	X
TABLolar DİZİNİ	XII
SEMBOLLER VE KISALTMALAR DİZİNİ	XVI
1. GENEL BİLGİLER.....	1
1.1. Problemin Tanımı.....	1
1.2. Literatürdeki Boşluk ve Araştırma Gerekçesi	4
1.3. Araştırma Soruları.....	4
1.4. Özgün Katkılar	5
1.5. Tezin Yapısı	5
2. TEORİK ALTYAPI	7
2.1. Literatür Taraması	7
2.2. Akciğer Kanseri	20
2.2.1. Akciğer Kanseri Türleri	21
2.2.2. Akciğer Kanseri Risk Faktörleri	21
2.2.3. Akciğer Kanseri Belirtileri.....	24
2.2.4. Akciğer Kanseri Tanı Yöntemleri	24
2.2.5. Geleneksel Tanı Yöntemlerinin Kısıtlamaları ve Nefes Analizine Yönelik Gereksinim.....	26
2.2.6. Akciğer Kanseri Evreleme	27
2.2.7. Akciğer Kanseri Tedavi Yöntemleri	30
2.3. Nefes ve Uçucu Organik Bileşikler.....	32
2.3.1. Akciğer Kanseri Özelinde Uçucu Organik Bileşikler	33
2.3.2. Nefes Analizi Teknikleri.....	34
2.3.3. Nefes Analizi Tekniklerinin Sınırlamaları ve Elektronik Burun Yaklaşımı	36
2.4. Elektronik Burun.....	36
2.4.1. Algılayıcı Birim.....	38

2.4.2.	Dönüştürücü Birim	42
2.4.3.	Elektronik Birim.....	43
2.4.4.	Örüntü Tanıma Birimi	43
2.5.	Veri Önişleme Yöntemleri.....	44
2.5.1.	Referans Düzeltme İşlemi.....	44
2.5.2.	İletkenlik Dönüşümü	45
2.5.3.	Normalizasyon	46
2.6.	Öznitelik Mühendisliği.....	47
2.6.1.	Öznitelik Çıkarımı	47
2.6.2.	Boyut İndirgeme Yöntemleri	48
2.7.	Sınıf Dengesizliği Problemi.....	51
2.7.1.	Sentetik Azınlık Aşırı Örnekleme Tekniği	52
2.7.2.	Sınırdaki Sentetik Azınlık Aşırı Örnekleme Tekniği	54
2.7.3.	Sentetik Azınlık Aşırı Örnekleme ve Sınıf Sınırı Temizleme Tekniği.....	55
2.7.4.	Adaptif Sentetik Örnekleme	56
2.8.	Sınıflandırma Algoritmaları.....	57
2.8.1.	Lojistik Regresyon Sınıflandırma Algoritması	57
2.8.2.	Destek Vektör Makineleri Sınıflandırma Algoritması	60
2.8.3.	k-En Yakın Komşuluk Sınıflandırma Algoritması.....	63
2.8.4.	Rastgele Orman Sınıflandırma Algoritması.....	65
2.9.	Çapraz Doğrulama.....	67
2.10.	Performans Metrikleri	68
3.	YAPILAN ÇALIŞMALAR.....	71
3.1.	Veri Seti.....	71
3.2.	Elektronik Burun Deney Düzenneği ve Deney Aşaması	72
3.3.	Çalışma Sürecinin Genel Akış Diyagramı.....	75
3.4.	Veri Önişleme	76
3.4.1.	Referans Düzeltme	76
3.4.2.	İletkenlik Dönüşümü	77
3.5.	Öznitelik Çıkarımı.....	78
3.5.1.	Temel İstatistiksel Öznitelikler	79
3.5.2.	Zaman Temelli Öznitelikler	80
3.5.3.	Türetilmiş Öznitelikler	81
3.6.	Z-Skor Normalizasyonu	82

3.7. Sınıf Dengesizliğinin Giderilmesi.....	82
3.8. Temel Bileşenler Analizi ile Boyut İndirgeme	85
3.9. Sınıflandırma Modelleri ve Hiperparametre Ayarları.....	87
3.10. Çapraz Doğrulama.....	88
3.11. Performans Metrikleri	89
4. BULGULAR VE TARTIŞMA	90
4.1. Dengesiz MOS Sensör Verilerine Ait Sınıflandırma Sonuçları	91
4.2. SMOTE ile Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları	93
4.3. Borderline SMOTE İle Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları	95
4.4. SMOTE-Tomek ile Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları	97
4.5. ADASYN ile Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları	99
4.6. Dengesiz QCM Sensör Verilerine Ait Sınıflandırma Sonuçları	101
4.7. SMOTE ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları	103
4.8. Borderline SMOTE ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları	105
4.9. SMOTE-Tomek ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları	107
4.10. ADASYN ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları	109
4.11. Dengesiz Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları	111
4.12. SMOTE ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları	113
4.13. Borderline SMOTE ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları	115
4.14. SMOTE-Tomek ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları	117
4.15. ADASYN ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları	119
5. SONUÇLAR	123
6. ÖNERİLER.....	125
7. KAYNAKLAR	126
ÖZGEÇMİŞ.....	136

Yüksek Lisans Tezi

ÖZET

SINIF DENGESİZLİĞİ İÇEREN TIBBİ VERİLERDE AKCİĞER KANSERİ EVRE SINIFLANDIRMASI

Bilal Talha AYVAZ

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Ayten ATASOY
2025, 153 Sayfa

Bu tez kapsamında, 8 metal oksit yarı iletken ve 14 kuvars kristal mikro terazi sensörden oluşan hibrit bir elektronik burun sistemi kullanılarak, akciğer kanseri tanısı almış bireylerin nefes örnekleri analiz edilmiş ve hastalık evrelerine göre sınıflandırılmıştır. Analiz edilen 200 nefes örneği, hastalık evrelerine göre sırasıyla I. Evre: 14, II. Evre: 12, III. Evre: 23 ve IV. Evre: 151 örnekten oluşmaktadır. Veri setine öncelikle referans düzeltme ve iletkenlik dönüşümü gibi ön işleme adımları uygulanmıştır. Bu adımların ardından öznelik çıkarımı gerçekleştirilmiş; elde edilen öznelikler Z-Skor normalizasyonu ile ölçeklendirilmiş ve sonrasında Temel Bileşenler Analizi kullanılarak boyut indirgeme işlemi yapılmıştır. Sınıflandırma işlemi hem dengesiz veri seti ile hem de Sentetik Azınlık Aşırı Örnekleme Tekniği ve türevleriyle dengelenmiş veri setleri üzerinde gerçekleştirilmiştir. Lojistik Regresyon, Destek Vektör Makineleri, k-En Yakın Komşu ve Rastgele Orman algoritmaları farklı hiperparametre kombinasyonlarıyla eğitilmiş ve test edilmiştir. Modellerin performansı doğruluk, makro kesinlik, makro duyarlılık ve makro F1-skoru gibi metriklerle değerlendirilmiştir. Sonuçlar, dengesiz veri setlerinde hem veri dengeleme algoritmalarının hem de hibrit sensör sistemlerinin makine öğrenmesi tabanlı sınıflandırma performansını belirgin biçimde artırdığını göstermektedir.

Anahtar Kelimeler: Akciğer kanseri, Nefes analizi, Elektronik burun, Sınıf dengesizliği, Hibrit sensör analizi

Master Thesis

SUMMARY

LUNG CANCER STAGE CLASSIFICATION ON MEDICAL DATA WITH CLASS IMBALANCE

Bilal Talha AYVAZ

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Electrical and Electronics Engineering Graduate Program
Supervisor: Prof. Dr. Ayten ATASOY
2025, 153 Pages

In this thesis, a hybrid electronic nose system comprising 8 metal oxide semiconductor sensors and 14 quartz crystal microbalance sensors was used to analyze the breath samples of individuals diagnosed with lung cancer and classify them according to disease stages. A total of 200 breath samples were analyzed, with the distribution as follows: 14 from Stage I, 12 from Stage II, 23 from Stage III, and 151 from Stage IV. Preprocessing steps such as baseline correction and conductivity transformation were applied, followed by feature extraction, Z-score normalization, and dimensionality reduction via Principal Component Analysis. Classification was performed using both the original imbalanced dataset and versions balanced with the Synthetic Minority Oversampling Technique and its variants. Logistic Regression, Support Vector Machines, k-Nearest Neighbors, and Random Forest algorithms were evaluated under different hyperparameter combinations. The performance of the models was evaluated using metrics such as accuracy, macro precision, macro recall, and macro F1-Score. The results demonstrate that both data balancing algorithms and hybrid sensor systems significantly enhance machine learning-based classification performance on imbalanced data sets.

Key Words: Lung cancer, Breath analysis, Electronic nose, Class imbalance, Hybrid sensor analysis

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Kanser türlerine göre küresel vaka ve ölüm sayıları (2022)	1
Şekil 2. Akciğer kanserinde evre ve yaşa göre 3 yıllık net sağkalım oranları (%)	2
Şekil 3. Akciğer kanseri tanılarının evrelere göre yüzdesel dağılımı.....	3
Şekil 4. Akciğer kanseri gelişim süreci	20
Şekil 5. Akciğer kanseri alt tiplerinin sınıflandırılması.....	21
Şekil 6. Akciğer kanseri risk faktörleri.....	22
Şekil 7. Akciğer kanseri klinik evreleri	30
Şekil 8. Biyolojik ve yapay koku alma sistemlerinin karşılaştırılması	37
Şekil 9. SMOTE yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması	52
Şekil 10. SMOTE algoritması akış diyagramı	53
Şekil 11. Borderline SMOTE yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması	55
Şekil 12. SMOTE-Tomek yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması	56
Şekil 13. ADASYN yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması	57
Şekil 14. Sigmoid fonksiyonu	58
Şekil 15. SVM karar sınırı, marjın ve w vektörünün gösterimi	61
Şekil 16. a) Lineer olarak ayrılabilen b) Lineer olarak ayrılamayan veri örnekleri.....	62
Şekil 17. Kernel dönüşümü ile doğrusal olmayan verilerin ayırt edilmesi.....	62
Şekil 18. k-NN algoritmasının sınıflandırma kararına etkisi	64
Şekil 19. Rastgele Orman algoritmasında karar ağaçlarının oylama süreci	66
Şekil 20. İki sınıflı bir problem için örnek bir konfüzyon matrisi	69
Şekil 21. Akciğer kanseri bireylerin evrelere göre dağılımı	72
Şekil 22. Elektronik burun sisteminin genel yapısını gösteren blok diyagram	72
Şekil 23. Elektronik burun sisteminin laboratuvar ortamındaki görünümü	73
Şekil 24. Deney aşamaları	74
Şekil 25. Çalışma sürecine ait iş akış şeması	75
Şekil 26. (a) Ham MOS sensör verisi ve (b) önışlem uygulanmış MOS sensör verisi	78
Şekil 27. (a) Ham QCM sensör verisi ve (b) önışlem uygulanmış QCM sensör verisi ...	78
Şekil 28. Dengeleme öncesi ve sonrasında eğitim seti sınıf dağılımı	83
Şekil 29. Dengeleme öncesi ve sonrasında validasyon ve test setleri sınıf dağılımı	84

Şekil 30.	MOS sensör verilerine ait dengesiz eğitim kümesinin ilk üç özniteliğe göre üç boyutlu dağılımı	84
Şekil 31.	MOS sensör verilerine SMOTE uygulanması sonrası elde edilen sentetik ve orijinal örneklerin üç boyutlu dağılımı (ilk üç öznitelik)	85
Şekil 32.	Dengesiz MOS sensör verisine uygulanan PCA ile elde edilen ilk üç bileşen düzleminde sınıf (evre) dağılımı.....	86
Şekil 33.	Çapraz doğrulama süreci.....	88



TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. TNM sınıflamasına göre akciğer kanseri evreleme kriterleri.....	27
Tablo 2. TNM kriterlerine göre akciğer kanseri klinik evreleri.....	28
Tablo 3. Akciğer kanserine ilişkin VOC grupları, örnekleri ve olası biyolojik kaynakları.....	34
Tablo 4. Gaz sensörlerinin sınıflandırılması ve karşılaştırılması.....	42
Tablo 5. Lojistik Regresyon optimizasyon algoritmaları	60
Tablo 6. Elektronik burun sisteminin temel bileşenleri ve görevleri	73
Tablo 7. MOS ve QCM sensör verilerine ait öznelilikler ve boyutları	81
Tablo 8. Sınıflandırma modelleri ve çalışmada kullanılan hiperparametreler.....	88
Tablo 9. Dengesiz MOS sensör verileri ile LR performans değerlendirmesi.....	91
Tablo 10. Dengesiz MOS sensör verileri ile SVM performans değerlendirmesi	91
Tablo 11. Dengesiz MOS sensör verileri ile k-NN performans değerlendirmesi.....	92
Tablo 12. Dengesiz MOS sensör verileri ile RF performans değerlendirmesi	92
Tablo 13. SMOTE ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi	93
Tablo 14. SMOTE ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi	93
Tablo 15. SMOTE ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi	94
Tablo 16. SMOTE ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi	94
Tablo 17. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi.....	95
Tablo 18. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi.....	95
Tablo 19. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi.....	96
Tablo 20. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi.....	96
Tablo 21. SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi.....	97
Tablo 22. SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi.....	97

Tablo 23.	SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi.....	98
Tablo 24.	SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi.....	98
Tablo 25.	ADASYN ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi	99
Tablo 26.	ADASYN ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi	99
Tablo 27.	ADASYN ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi	100
Tablo 28.	ADASYN ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi	100
Tablo 29.	Dengesiz QCM sensör verileri ile LR performans değerlendirmesi	101
Tablo 30.	Dengesiz QCM sensör verileri ile SVM performans değerlendirmesi	101
Tablo 31.	Dengesiz QCM sensör verileri ile k-NN performans değerlendirmesi.....	102
Tablo 32.	Dengesiz QCM sensör verileri ile RF performans değerlendirmesi.....	102
Tablo 33.	SMOTE ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi	103
Tablo 34.	SMOTE ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi	103
Tablo 35.	SMOTE ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi	104
Tablo 36.	SMOTE ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi	104
Tablo 37.	Borderline SMOTE ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi.....	105
Tablo 38.	Borderline SMOTE ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi.....	105
Tablo 39.	Borderline SMOTE ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi.....	106
Tablo 40.	Borderline SMOTE ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi.....	106
Tablo 41.	SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi.....	107
Tablo 42.	SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi.....	107
Tablo 43.	SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi.....	108
Tablo 44.	SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi.....	108

Tablo 45.	ADASYN ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi	109
Tablo 46.	ADASYN ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi	109
Tablo 47.	ADASYN ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi	110
Tablo 48.	ADASYN ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi	110
Tablo 49.	Dengesiz hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirmesi	111
Tablo 50.	Dengesiz hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirmesi	111
Tablo 51.	Dengesiz hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirmesi	112
Tablo 52.	Dengesiz hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirmesi	112
Tablo 53.	SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirilmesi.....	113
Tablo 54.	SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirilmesi.....	113
Tablo 55.	SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirilmesi.....	114
Tablo 56.	SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirilmesi.....	114
Tablo 57.	Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirilmesi	115
Tablo 58.	Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirilmesi.....	115
Tablo 59.	Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirilmesi	116
Tablo 60.	Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirilmesi	116
Tablo 61.	SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirilmesi	117
Tablo 62.	SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirilmesi	117
Tablo 63.	SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirilmesi	118
Tablo 64.	SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirilmesi	118

Tablo 65.	ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performansının değerlendirilmesi.....	119
Tablo 66.	ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performansının değerlendirilmesi.....	119
Tablo 67.	ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performansının değerlendirilmesi.....	120
Tablo 68.	ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performansının değerlendirilmesi.....	120



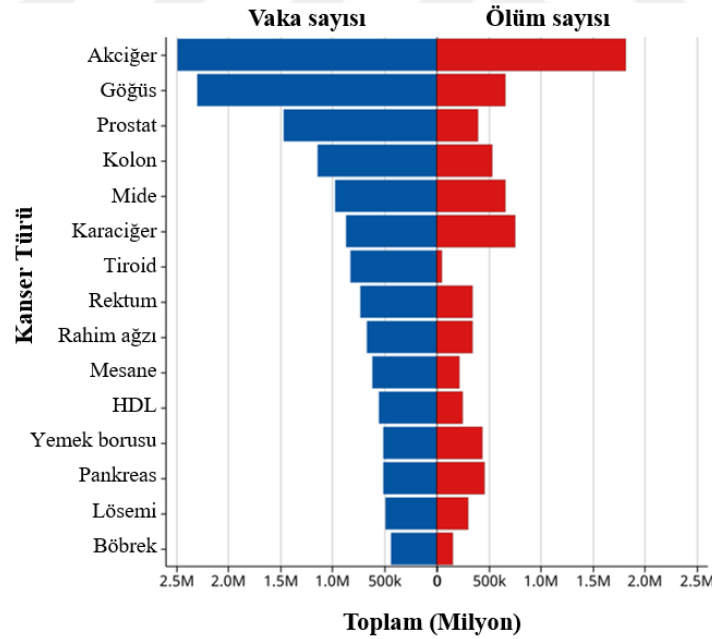
SEMBOLLER VE KISALTMALAR DİZİNİ

VOC	: Uçucu Organik Bileşikler
MOS	: Metal Oksit Yarı İletken
QCM	: Kuvars Kristal Mikro Terazî
SMOTE	: Sentetik Azınlık Aşırı Örnekleme Tekniğı
Borderline SMOTE	: Sınırdaki Sentetik Azınlık Aşırı Örnekleme Tekniğı
SMOTE-Tomek	: Sentetik Azınlık Aşırı Örnekleme ve Sınıf Sınırı Temizleme Tekniğı
ADASYN	: Adaptif Sentetik Örnekleme
LR	: Lojistik Regresyon
SVM	: Destek Vektör Makineleri
k-NN	: k-En Yakın Komşuluk
RF	: Rastgele Orman
DT	: Karar Ağaçları
ANN	: Yapay Sinir Ağları
PCA	: Temel Bileşenler Analizi
LDA	: Lineer Diskriminant Analizi
KHDAK	: Küçük Hücreli Dış Akciğer Kanseri
KHAK	: Küçük Hücreli Akciğer Kanseri
KOAH	: Kronik Obstrüktif Akciğer Hastalığı
SAW	: Yüzey Dalga Akustik
CP	: İletken Polimer

1. GENEL BİLGİLER

1.1. Problemin Tanımı

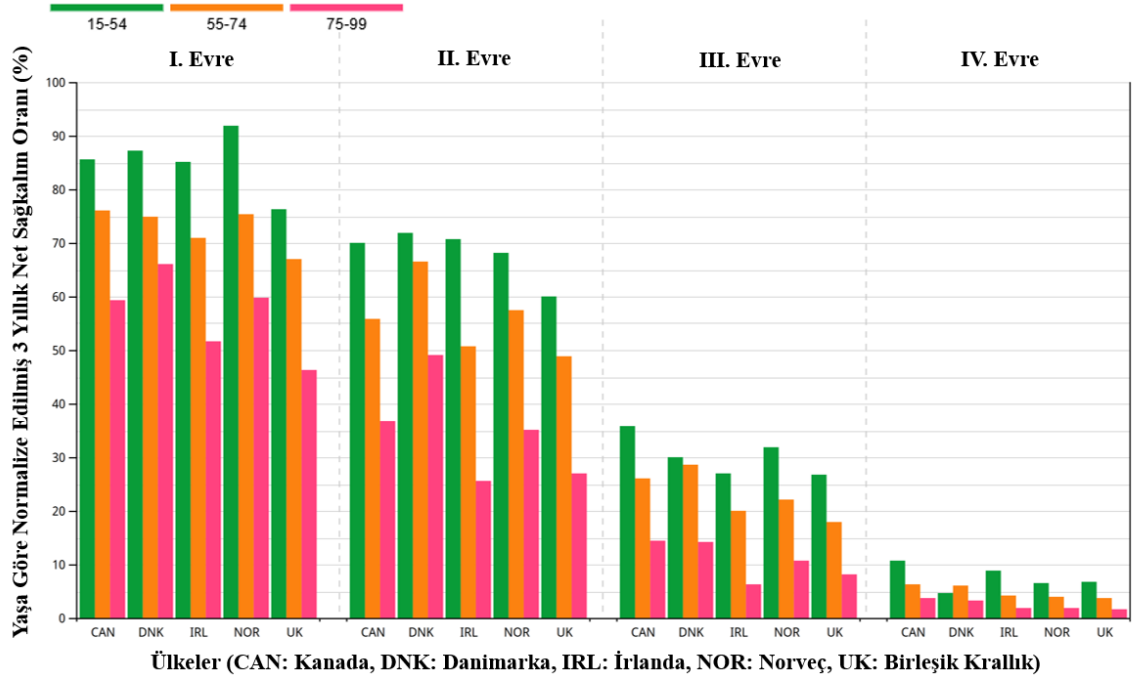
Modern tıp, birçok farklı kanser türünün tedavisinde önemli ilerlemeler kaydetmiş olsa da kanser hâlâ kesin ve evrensel bir tedavi sunulamayan en karmaşık sağlık sorunlarından biridir. Bu bağlamda, akciğer kanseri hem görülme sıklığı hem de yüksek ölüm oranları nedeniyle küresel ölçekte önemli bir sağlık sorunu olarak öne çıkmaktadır (Barta, Powell, & Wisnivesky, 2019). 2022 yılı Küresel Kanser Gözlemevi verilerine göre akciğer kanseri, dünya genelinde hem en fazla tanı konulan kanser türü hem de kansere bağlı ölümlerde en yüksek paya sahip hastalıktır. Bu da hastalığın, küresel sağlık gündeminde öncelikli bir yer tutmasına neden olmaktadır. Şekil 1’de, 2022 yılına ait farklı kanser türlerine göre küresel vaka ve ölüm sayıları karşılaştırmalı olarak verilmiştir.



Şekil 1. Kanser türlerine göre küresel vaka ve ölüm sayıları (2022)

Akciğer kanseri genellikle erken evrelerde belirti vermediğinden, tanı çoğu zaman ileri evrelerde konulmaktadır (Ellis & Vandermeer, 2011). Bu gecikme, hastalığın tedavi edilebilir evrelerinin atlanmasına neden olmakta ve sağkalım oranlarının azalmasına yol açmaktadır. Tanı anındaki evre kadar hastanın yaşı da sağkalım üzerinde belirleyici bir

faktördür (Nesbitt, Putnam, Walsh, Roth, & Mountain, 1995) (Taş, Çiftçi, Kılıç, & Karabulut, 2013). Genç yaş grubundaki hastalarda (15–54 yaş), erken evrede teşhis edilen vakalarda sağkalım oranları oldukça yüksektir. Ancak ileri evrelerde ya da ileri yaş grubundaki hastalarda (75 yaş ve üzeri) bu oranlar kayda değer ölçüde azalmaktadır. Özellikle IV. Evrede sağkalım oranları %10'un altına kadar düşebilmektedir. Şekil 2'de Küresel Kanser Gözlemevi veri tabanından alınan Kanada, Danimarka, İrlanda, Norveç ve Birleşik Krallık'a ait veriler, söz konusu ülkelerde yaşa göre normalize edilmiş 3 yıllık net sağkalım oranlarının; hastalığın evresine ve yaş gruplarına göre nasıl değiştiğini karşılaştırmalı olarak göstermektedir.

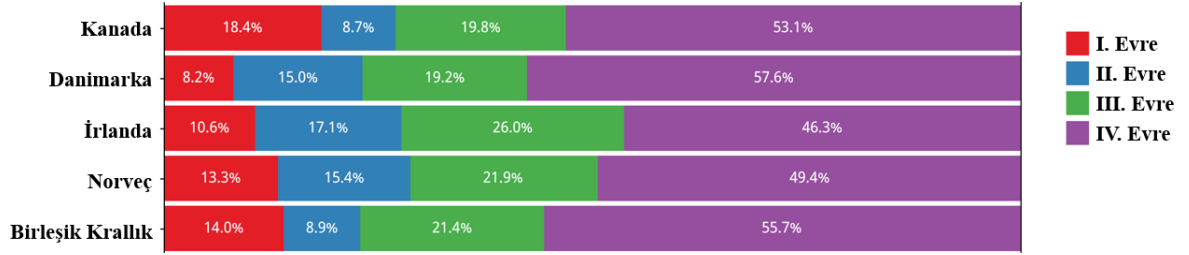


Şekil 2. Akciğer kanserinde evre ve yaşa göre 3 yıllık net sağkalım oranları (%)

Erken ve doğru tanı, tedavi sürecinin başarısı açısından kritik öneme sahiptir. Nitekim Şekil 2, akciğer kanserinde erken evrede teşhis edilen vakaların sağkalım oranlarının belirgin şekilde daha yüksek olduğunu; tanının ileri evrelere kalmasının ise ölüm oranlarını ciddi biçimde artırdığını ortaya koymaktadır. Ancak günümüzde yaygın olarak kullanılan tanı yöntemleri genellikle invaziv, zaman alıcı ve yüksek maliyetlidir (Nooreldeen & Bach, 2021). Bu durum, özellikle tarama amaçlı kullanımda önemli bir engel oluşturarak, erken tanı oranlarının düşük kalmasına neden olmaktadır.

Elektronik burun, yalnızca bir nefes örneği ile uçucu organik bileşikleri (VOC) analiz edebilme kapasitesi sayesinde, hızlı, invaziv olmayan ve düşük maliyetli bir tanı alternatifi olarak dikkat çekmektedir. Bu teknoloji hem erken hem de ileri evre akciğer kanserinin sınıflandırılmasında kayda değer bir potansiyele sahiptir.

Ancak, bu tür sınıflandırma çalışmalarında karşılaşılan en temel zorluklardan biri, tıbbi verilerde sıkça rastlanan sınıf dengesizliği problemidir (Rahman & Davis, 2013). Klinik uygulamalarda ileri evre hastalara ait örnekler daha fazla sayıda elde edilirken, erken evre hastalıklar çoğu zaman belirti göstermediği için tanı koymak zorlaşmakta ve dolayısıyla bu gruba ait veriler sınırlı kalmaktadır. Ayrıca, ileri evre hastalar daha sık klinik takibe alındıkları için veri kayıtları açısından baskın hale gelmektedir. Şekil 3, Küresel Kanser Gözlemevi veri tabanından alınan veriler doğrultusunda, Kanada, Danimarka, İrlanda, Norveç ve Birleşik Krallık'ta akciğer kanseri tanılarının hastalığın evrelerine göre yüzde dağılımını göstermektedir.



Şekil 3. Akciğer kanseri tanılarının evrelere göre yüzdesel dağılımı

Şekil 3'ten görüleceği üzere tüm ülkelerde en yüksek oran, IV. Evre'de teşhis edilen vakalara aittir. Buna karşılık, erken evrelerde tanı konan hasta oranı oldukça düşüktür. Bu dengesizlik, elektronik burunla elde edilen verilerin makine öğrenmesiyle analiz edilmesi sürecinde sınıflandırma performansını olumsuz yönde etkileyebilir. Makine öğrenmesi algoritmaları, çoğunluk sınıfına ait örneklerden daha fazla bilgi öğrenmeye eğilimli olduğundan, azınlık sınıfını temsil eden, yani erken evre hastalara ait örnekler yeterince temsil edilememektedir. Sonuç olarak, model tüm sınıfları dengeli bir şekilde öğrenmekte zorlanır ve özellikle erken evre hastaları doğru bir şekilde sınıflandıramaz.

1.2. Literatürdeki Boşluk ve Araştırma Gerekçesi

Son yıllarda elektronik burun sistemlerinin sağlık alanında, özellikle akciğer kanseri gibi solunum yolu hastalıklarının tanısında kullanımına yönelik çok sayıda çalışma gerçekleştirilmiştir. Buna karşın çalışmaların çoğu hastalığın varlığını tespit etmeye odaklanırken, hastalığın evrelerinin ayırt edilmesine yönelik çok sınıflı sınıflandırma problemleri literatürde oldukça sınırlı şekilde ele alınmaktadır. Ayrıca, mevcut çalışmalar genellikle yalnızca tek bir sensör teknolojisine dayalı olarak yürütülmekte ve farklı sensör yapılarını karşılaştırmalı veya entegre biçimde değerlendiren çalışmaların sayısı oldukça azdır.

Bununla birlikte, çok sınıflı tıbbi veri kümelerinde yaygın olarak karşılaşılan sınıf dengesizliği problemi, çoğu çalışmada göz ardı edilmekte ya da yalnızca temel yaklaşımlarla ele alınmaktadır. Özellikle erken evre hastalara ait verilerin azlığı, sınıflandırma algoritmalarının performansını olumsuz yönde etkileyerek modelin erken evreleri ayırt etmede yetersiz kalmasına neden olabilmektedir.

Bu nedenlerle, farklı sensör teknolojilerini entegre eden ve sınıf dengesizliği gibi yapısal sorunları dikkate alan modellere duyulan ihtiyaç açıktır. Bu tez çalışması, literatürdeki bu eksiklikleri gidermeye yönelik bütüncül bir yaklaşım sunmayı amaçlamaktadır.

1.3. Araştırma Soruları

Bu tez kapsamında aşağıdaki temel araştırma sorularına yanıt aranacaktır:

- I. Elektronik burun sistemlerinden elde edilen sensör verileriyle akciğer kanserinin farklı evreleri ne ölçüde doğru şekilde sınıflandırılabilir?
- II. Metal Oksit Yarı İletken (MOS) sensörler ve Kuartz Kristal Mikro Teraziler (QCM) sensörlerinin ayrı ayrı ve birlikte kullanımı sınıflandırma performansını nasıl etkilemektedir?
- III. Tıbbi verilerdeki sınıf dengesizliği, modellerin erken evreleri tanıma yeteneğini nasıl etkiler?
- IV. Sentetik Azınlık Aşırı Örneklemme Tekniği (SMOTE), Sınırdaki Sentetik Azınlık Aşırı Örneklemme Tekniği (Borderline-SMOTE), Sentetik Azınlık Aşırı Örneklemme ve Sınıf Sınırı Temizleme Tekniği (SMOTE-Tomek) ve Adaptif Sentetik

Örnekleme (ADASYN) gibi veri dengeleme yöntemleri sınıflandırma başarımına ne ölçüde katkı sağlar?

- V. Lojistik Regresyon (LR), Destek Vektör Makineleri (SVM), k-En Yakın Komşu (k-NN) ve Rastgele Orman (RF) gibi sınıflandırma algoritmaları, farklı parametre ayarlarıyla birlikte değerlendirildiğinde, akciğer kanseri evrelerinin sınıflandırılmasında en yüksek başarıyı hangi kombinasyonlarla sağlar?

1.4. Özgün Katkılar

Bu tez çalışmasında elektronik burun kullanılarak akciğer kanseri evrelerinin sınıflandırılmasına yönelik literatürdeki boşluklara katkı sağlamayı amaçlamaktadır. Çalışmanın özgün yönleri aşağıda sunulmuştur:

- I. Farklı sensör teknolojilerinin beraber kullanımı: MOS ve QCM tabanlı sensörlerin aynı elektronik burun sisteminde entegre edilerek ayrı ayrı ve hibrit olarak değerlendirilmesi, sınıflandırma performansına etkisinin sistematik biçimde incelenmesi açısından literatürde sınırlı çalışılan bir konudur.
- II. Tıbbi veri dengesizliğinin ele alınması: Hastalığın klinik dağılımından kaynaklanan belirgin sınıf dengesizliği problemi, çeşitli sentetik örnekleme teknikleri (SMOTE, Borderline-SMOTE, SMOTE-Tomek, ADASYN) uygulanarak ele alınmış ve bu yöntemlerin çok sınıflı medikal veri üzerindeki etkileri karşılaştırılmıştır. Tıbbi sınıflandırma problemlerinde veri dengesizliğini ele alan çok sınıflı ve sistematik karşılaştırma çalışmaları literatürde sınırlı sayıdadır; bu yönüyle çalışma, önemli bir boşluğu doldurmayı hedeflemektedir.

1.5. Tezin Yapısı

Bu tez, akciğer kanseri evrelerinin elektronik burun sistemleri kullanılarak sınıflandırılmasına yönelik yürütülen çalışmaları sistematik bir biçimde sunmak amacıyla altı ana bölümden oluşmaktadır:

Birinci bölümde, çalışmanın problemi tanımlanmış; literatürdeki boşluklar, araştırmanın gerekçesi, araştırma soruları ve çalışmanın özgün katkıları detaylı şekilde sunulmuştur.

İkinci bölüm, çalışmanın teorik altyapısını oluşturmaktadır. Bu bölümde, akciğer kanseri, geleneksel tanı yöntemleri ve bu yöntemlerin sınırlılıkları ile nefes analizine olan gereksinim ele alınmıştır. Ayrıca uçucu organik bileşikler, nefes analizi teknikleri ve elektronik burun teknolojileri hakkında literatürde yer alan güncel çalışmalar incelenmiştir. Bunun yanı sıra, veri ön işleme yöntemleri, öznitelik mühendisliği, sınıf dengesizliği problemleri ve çeşitli sınıflandırma algoritmalarına yönelik teorik bilgiler de bu bölümde sunulmuştur.

Üçüncü bölümde, tez kapsamında gerçekleştirilen uygulama detaylandırılmıştır. Kullanılan veri seti, elektronik burun düzeneği, veri ön işleme adımları, öznitelik çıkarımı, normalizasyon, sınıf dengesizliğinin giderilmesi, boyut indirgeme yöntemleri, sınıflandırma algoritmaları ve hiperparametre ayarları alt başlıklar hâlinde açıklanmıştır.

Dördüncü bölümde, elde edilen deneysel bulgular sunulmuş ve farklı sensör kombinasyonları, çeşitli veri dengeleme stratejileri ile sınıflandırma modellerinin performansları karşılaştırmalı olarak değerlendirilmiştir.

Beşinci bölümde, çalışmanın genel sonuçları özetlenmiş ve bulgular literatür ışığında tartışılmıştır.

Altıncı ve son bölümde ise, çalışmanın sınırlılıkları belirtilmiş ve gelecekte yapılabilecek araştırmalara yönelik öneriler sunulmuştur.

2. TEORİK ALTYAPI

2.1. Literatür Taraması

(Gordon, Szidon, Krotoszynski, Gibbons, & O'Neill, 1985), gaz kromatografisi ve kütle spektrometrisi yöntemlerini kullanarak akciğer kanseri teşhisi için nefes analizinin potansiyelini incelemişlerdir. Araştırmada 12 akciğer kanseri hastası ve 17 sağlıklı bireye ait nefes örnekleri toplanmıştır. Nefes analizinde hastaların nefesinde çeşitli organik bileşikler tespit edilmiş ve bu bileşiklerin kanserle bağlantılı olabileceği bulunmuştur. Tespit edilen organik bileşikler arasında alkanlar (heksan, metilpentan, trimetilheptan), aromatik hidrokarbonlar (benzen, toluen, etilbenzen, trimetilbenzen) ve oksijen içeren bileşikler (aseton, 2-bütanon, benzaldehit, oktanal, asetofenon) bulunmaktadır. Analizler, belirli bileşiklerin yalnızca akciğer kanseri hastalarında bulunduğunu ortaya koymaktadır. Örneğin, benzaldehit ve asetofon gibi bileşikler, bu hastalarda yüksek konsantrasyonlarda tespit edilmiştir.

(O'Neill, Gordon, O'Neill, Gibbons, & Szidon, 1988), akciğer kanseri hastalarının nefeslerinde birçok bileşik tespit etmiş ve bu bileşiklerin hastalıkla ilişkili olabileceğini saptamışlardır. Tespit edilen organik bileşikler arasında alkanlar (heksan, metilpentan, trimetilheptan), aromatik hidrokarbonlar (benzen, toluen, etilbenzen) ve oksijen içeren bileşikler (aseton, 2-bütanon, benzaldehit, asetofenon, nonanal) bulunmaktadır.

(Phillips, ve diğerleri, 1996), yaptıkları araştırmada 60 akciğer kanseri hastası ve 48 sağlıklı bireyden nefes örnekleri toplayarak, akciğer kanseri teşhisinde VOC'lerin potansiyel biyobelirteç olarak kullanımını araştırmıştır. Ağırlıklı olarak alkanlar, alkan türevleri ve benzen türevlerinden oluşan 22 farklı VOC kombinasyonu, hastalığın evresinden bağımsız olarak, akciğer kanseri hastaları ile sağlıklı bireyler arasında anlamlı bir ayrım yapılmasını mümkün kılmıştır.

(Philips, ve diğerleri, 2003), nefes örneklerinde bulunan VOC'lerin akciğer kanseri tespiti için biyobelirteçler olarak kullanılabilirliğini araştırmayı amaçlamaktadır. Çalışma, ABD ve Birleşik Krallık'taki beş akademik merkezden toplam 178 hasta (67 tanesi akciğer kanseri, 15 tanesi metastatik akciğer kanseri) ve 41 sağlıklı gönüllü üzerinde gerçekleştirilmiştir. Nefes örnekleri, gaz kromatografisi ve kütle spektrometrisi kullanılarak analiz edilmiştir. Tahmin modeli, akciğer kanserini %89,6 duyarlılık ve %82,9 özgüllükle

tespit etmiştir. Çapraz doğrulama sonuçlarına göre, modelin duyarlılığı %85,1, özgüllüğü ise %80,5 olarak belirlenmiştir. Model, metastatik akciğer kanserini %66,7 duyarlılık ile tespit ederken, negatif bronkoskopi bulgularına sahip hastaları %37,4 özgüllükle sınıflandırmıştır.

(Poli, ve diğerleri, 2005), yaptıkları çalışmada Küçük Hücreli Dışı Akciğer Kanseri (KHDAK) hastalarının nefeslerindeki VOC'lerin biyobelirteç olarak kullanımını araştırmıştır. Nefes örnekleri gaz kromatografisi ve kütle spektrometrisi yöntemiyle analiz edilmiştir. 13 VOC'nin kombinasyonu kullanılarak yapılan Çok Değişkenli Lojistik Regresyon Analizi, hastaların %80'ini doğru şekilde sınıflandırmıştır. Sonuçlar, VOC analizinin akciğer kanseri tanısında yardımcı bir test olarak kullanılabileceğini göstermektedir.

Bu çalışmalar, nefes örneklerinde tespit edilen VOC'lerin akciğer kanserinin teşhisinde anlamlı biyobelirteçler olabileceğini ortaya koymaktadır. Elde edilen bulgular, nefes analizine dayalı tanı yaklaşımlarının klinik açıdan taşıdığı potansiyeli desteklerken, aynı zamanda elektronik burun teknolojisinin geliştirilmesi için bilimsel bir temel sunmaktadır.

(Di Natale, ve diğerleri, 2003), akciğer kanserinin teşhisinde nefes örneklerindeki VOC'lerin elektronik burun kullanılarak analiz edilmesiyle hastalığın tespitinde kullanılabilirliğini incelemiştir. Toplamda 60 birey incelenmiş; bunların 35'i akciğer kanseri hastası, 18'i sağlıklı kontrol grubu ve 9'u cerrahi tedavi görmüş hastalardır. Elektronik burun teknolojisi kullanılarak nefes örnekleri analiz edilmiş, bu süreçte 8 adet QCM sensörü kullanılmıştır. Kısmi En Küçük Kareler Diskriminant Analizi ile yapılan veri analizinde, akciğer kanseri hastaları %100 doğrulukla, kontrol grubu %94 doğrulukla sınıflandırılmıştır. Cerrahi sonrası hastaların %44'ü doğru şekilde tanımlanırken, geri kalanları sağlıklı bireyler olarak sınıflandırılmıştır. Sonuçlar, akciğer kanseri hastalarının nefes bileşimindeki değişikliklerin hastalığın teşhisinde önemli bir rol oynadığını göstermektedir.

(Machado, ve diğerleri, 2005), elektronik burun teknolojisinin, akciğer kanseri hastalarının nefeslerinde bulunan VOC'leri tespit ederek hastalığı teşhis etme potansiyelini araştırmıştır. İlk aşamada, 14 akciğer kanseri hasta ve 45 sağlıklı bireyin nefes örnekleri analiz edilmiştir, VOC'leri sınıflandırmak için Temel Bileşenler Analizi (PCA) ve Kanonik Diskriminant Analizi kullanılarak bir tahmin modeli geliştirilmiştir. İkinci aşamada ise, model doğrulama amacıyla farklı bir gruptan 14 akciğer kanseri hastası ve 62 sağlıklı bireyin nefes örnekleri analiz edilmiştir. Elde edilen veriler SVM tabanlı bir tahmin modeli

oluşturulmuştur. Elektronik burun doğrulama aşamasında %71.4 duyarlılık ve %91.9 özgüllükle akciğer kanserini doğru şekilde tespit etmiştir.

(Chen, ve diğerleri, 2005), akciğer kanserinin teşhisi için Yüzey Akustik Dalga (SAW) sensör dizisi ve görüntü tanıma yöntemine dayalı bir elektronik burun sistemi geliştirmiştir. Toplamda 42 gönüllüden alınan nefes örnekleri analiz edilmiştir. Elektronik burun, stiren, dekan, izopren gibi 11 VOC'yi tespit etmek için kullanılmıştır. Nefes örnekleri, iki torbalı bir sistemle toplanmış, Katı Faz Mikroekstraksiyonu ile önceden konsantre edilmiş ve SAW sensörleri tarafından analiz edilmiştir. Model, akciğer kanseri hastalarının VOC profillerini başarılı şekilde tanımlarken, akciğer kanseri olan bireyler ile sağlıklı bireyler arasında anlamlı farklar tespit edilmiştir.

(Mazzone, ve diğerleri, 2007), akciğer kanseri olan hastaların nefeslerinde bulunan VOC'lerin tespit edilmesi amacıyla renk değiştirici sensör dizisi kullanarak akciğer kanserini teşhis etmeyi hedeflemiştir. Toplam 143 bireyden nefes örnekleri toplanmıştır. Bu bireyler arasında 49 küçük hücreli dışı akciğer kanseri (KHDAK) hastası, 18 kronik obstrüktif akciğer hastalığı (KOAH) hastası, çeşitli diğer solunum yolu hastalıklarına sahip bireyler ve 21 sağlıklı kontrol bulunmaktadır. Nefes örnekleri, renk değiştirici 36 sensörden oluşan bir diziden geçirilmiş ve bu sensörlerin verdiği renk değişimleri analiz edilmiştir. RF algoritması kullanılarak geliştirilen model, akciğer kanseri teşhisinde %73,3 duyarlılık ve %72,4 özgüllük sağlamıştır. Akciğer nodüllerine sahip bireylerde modelin duyarlılığı %100, özgüllüğü ise %60 olarak bulunmuştur. Sonuç olarak, renk değiştirici sensör dizisi, akciğer kanseri tanısında bir araç olarak umut vaat etmektedir, ancak daha geniş örneklemelerle doğrulanması gerekmektedir.

(Blatt, ve diğerleri, 2007), MOS sensör dizisi tabanlı bir elektronik burun kullanarak akciğer kanserini teşhis etmeyi amaçlamaktadır. Toplamda 101 birey (43 akciğer kanseri hastası ve 58 sağlıklı birey) üzerinde gerçekleştirilen çalışmada, altı adet MOS sensörü kullanılarak nefes örneklerindeki VOC'lerin tespit edilmesi hedeflenmiştir. Toplanan veriler PCA ve Parametrik Olmayan Lineer Diskriminant Analizi ile işlenmiş; k-NN, Yapay Sinir Ağları (ANN) gibi çeşitli sınıflandırma algoritmalarıyla analiz edilmiştir. Elde edilen en yüksek sonuçlar incelendiğinde, %92,6 doğruluk, %95,3 duyarlılık ve %90,5 özgüllük oranlarına ulaşılmıştır.

(Wang, ve diğerleri, 2008), erken evre akciğer kanseri teşhisi için SAW sensör dizisi ve görüntü tanıma yöntemine dayalı bir elektronik burun sistemi geliştirmeyi amaçlamaktadır. Toplamda 32 gönüllü (15 akciğer kanseri hastası, 10 sağlıklı birey ve 7

KOAH) üzerinde yapılan çalışmada, nefes örnekleri kontrollü koşullarda toplanmış ve SAW sensörleri kullanılarak VOC'ler tespit edilmiştir. ANN algoritması ve görüntü tanıma yöntemleri ile VOC desenleri analiz edilerek, akciğer kanseri hastalarının nefes profilleri sağlıklı bireylerden yüksek doğrulukla ayırt edilmiştir. Sonuçlar, SAW sensör dizisi ve nefes analizine dayalı bu elektronik burun teknolojisinin erken evre akciğer kanseri teşhisinde etkili bir yöntem olabileceğini göstermiştir. Ancak, sistemin daha geniş örneklerle doğrulanması ve klinik testlerle geliştirilmesi gerektiği vurgulanmıştır.

(Dragonieri, ve diğerleri, 2009), KHDAK hastalarını, KOAH hastaları ve sağlıklı bireylerden ayırt etmek amacıyla elektronik burun teknolojisi kullanılarak nefes örneklerindeki VOC'lerin tespit edilmesi hedeflenmiştir. Toplamda 30 katılımcıdan oluşan örnekleme, 10 KHDAK hastası, 10 KOAH hastası ve 10 sağlıklı kontrol yer almıştır. Katılımcılar, VOC filtreli bir valf aracılığıyla nefes vererek nefes örneklerini Tedlar torbalarına üflemiş ve bu örnekler 32 adet polimer sensörden oluşan Cyranose 320 adlı elektronik burunla analiz edilmiştir. Elde edilen verilerin boyutları PCA ve Kanonik Diskriminant Analizi kullanılarak indirgenmiştir. KHDAK hastalarının nefes profilleri, KOAH hastalarından %85 doğrulukla (M-mesafesi: 3.73) ve sağlıklı bireylerden %90 doğrulukla (M-mesafesi: 2.96) ayırt edilebilmiştir. Çalışma, elektronik burun teknolojisinin akciğer kanseri teşhisinde potansiyel olarak etkili bir araç olduğunu göstermektedir. Ancak, daha büyük örneklem gruplarıyla yapılacak çalışmalar ve farklı akciğer hastalıklarının teşhisi için daha fazla doğrulama gerekmektedir.

(D'Amico, ve diğerleri, 2010), akciğer kanseri teşhisinde elektronik burun teknolojisinin etkinliğini araştırmayı amaçlamaktadır. Toplamda 92 katılımcı (28 akciğer kanseri hastası, 28 diğer akciğer hastalığı olan birey ve 36 sağlıklı kontrol) üzerinde yürütülen çalışmada, sabah saatlerinde ve aç karna alınan nefes örnekleri özel bir çift torbalı sistem aracılığıyla toplanmıştır. Nefes örneklerindeki VOC'lerin tespiti için 8 adet QCM sensöründen oluşan bir elektronik burun kullanılmıştır. Veriler, Kısmi En Küçük Kareler Diskriminant Analizi yöntemi ile analiz edilmiş ve çapraz doğrulama yöntemi ile doğrulanmıştır. Analiz sonucunda, akciğer kanseri hastaları sağlıklı bireylerden %85 duyarlılık ve %100 özgüllükle, akciğer kanseri hastaları diğer akciğer hastalıklarından %92.8 duyarlılık ve %78.6 özgüllükle ayırt edilmiştir. Tüm gruplar (akciğer kanseri, diğer akciğer hastalıkları ve sağlıklı bireyler) arasında yapılan üçlü sınıflandırmada ise %79.3 genel doğruluk ve %89.3 duyarlılık elde edilmiştir.

(Wang, ve diğerleri, 2012), akciğer kanserini teşhis etmek amacıyla MOS ve SAW sensörlerini birleştiren hibrit bir elektronik burun sistemi geliştirmeyi amaçlamaktadır. Toplamda 47 akciğer kanseri hastası ve 42 sağlıklı bireyden oluşan 89 katılımcı ile yapılan çalışmada, katılımcıların nefes örnekleri 12 saatlik açlık ve sigara içmeme süresinden sonra Tedlar torbalarına toplanmıştır. Veri analizi, PCA, Lineer Diskriminant Analizi (LDA), Kısmi En Küçük Kareler ve ANN algoritmaları ile yapılmış ve ANN modeli %93.62 duyarlılık ve %83.37 özgüllük ile en yüksek başarı oranını sağlamıştır. MOS ve SAW sensörlerinin birleşimi, akciğer kanseri teşhisinde daha doğru bir ayırım sağlarken, bu hibrit sistemin tek tip sensörlere göre daha kapsamlı ve yüksek doğruluğa sahip bir tarama aracı olabileceği ortaya konmuştur. Test süresini azaltmak ve cihaz boyutlarını küçültmek için ise ileri çalışmalar önerilmektedir.

(Capuano, ve diğerleri, 2015), akciğer kanseri ve sağlıklı bireylerin nefes örneklerindeki VOC analizini yapmak için gaz kromatografisi ve kütle spektrometrisi ve elektronik burun teknolojisi yöntemlerini karşılaştırmıştır. Çalışmada 20 akciğer kanseri hastası ve 10 sağlıklı bireyden alınan nefes örnekleri her iki yöntemle analiz edilmiştir. Gaz kromatografisi ve kütle spektrometrisi yöntemi, VOC'leri ayrıntılı bir kimyasal analizle tek tek tanımlamaya odaklanmış, %76 doğruluk oranı ile VOC seviyelerindeki farklılıkları tespit etmiştir; ancak VOC profillerinin kanserli ve kansersiz bireyler arasında istatistiksel olarak anlamlı farklar göstermesi sınırlı kalmıştır. Elektronik burun yöntemi ise %93 doğrulukla kanserli ve sağlıklı bireyleri ayırt etmede daha başarılı olmuştur. Sonuç olarak, elektronik burun teknolojisi akciğer kanseri teşhisinde önemli bir tarama aracı olarak öne çıkmaktadır.

(Tirzite, Bukovskis, Strazda, Jurka, & Taivans, 2017), akciğer kanseri hastalarının nefes analizini elektronik burun teknolojisi kullanarak teşhis etmeyi amaçlamaktadır. Akciğer kanseri tanısı almış, diğer akciğer hastalıkları bulunan ve sağlıklı bireylerden oluşan örneklemede katılımcıların nefes örnekleri toplamda 32 polimer sensöre sahip Cyranose 320 cihazı ile analiz edilmiştir. Veriler SVM yöntemiyle analiz edilmiş ve kanser hastaları %87.3 duyarlılık ve %71.2 özgüllükle sağlıklı bireylerden ayrılmıştır; kanser ve sağlıklı gönüllüler %98.8 duyarlılık ve %81 özgüllükle doğru şekilde sınıflandırılmıştır. Akciğer kanseri ve KOAH hastaları %100 doğru ayrılabilirken KOAH tanısı konan bireylerde doğruluk oranı %62 olmuştur. Sonuçlar, akciğer kanseri teşhisinde elektronik burun ve SVM modeliyle yüksek doğruluk oranlarına ulaşılabileceğini, ancak daha geniş gruplarla yapılacak doğrulama çalışmalarının gerekli olduğunu göstermektedir.

(Tirzite, Bukovskis, Strazda, Jurka, & Taivans, 2018), akciğer kanseri hastalarını diğer akciğer hastalıklarına sahip bireylerden ve sağlıklı bireylerden ayırt etmek amacıyla elektronik burun teknolojisinin etkinliğini araştırmıştır. Toplamda 252 akciğer kanseri hastası ve 223 kanser dışı bireyden oluşan örneklem 32 sensöre sahip Cyranose 320 elektronik burun cihazı ile analiz edilmiştir. Veriler, yaş, sigara içme durumu, sigara geçmişi ve ortam sıcaklığı gibi değişkenlerle birlikte LR kullanılarak işlenmiştir. Sigara içmeyenlerde akciğer kanseri tespiti %96.2 duyarlılık ve %90.6 özgüllük ile gerçekleştirilirken, sigara içenlerde %95.8 duyarlılık ve %92.3 özgüllük elde edilmiştir. Sonuç olarak elektronik burun, akciğer kanseri tanısında yüksek doğruluk oranlarıyla güvenilir bir yöntem olarak öne çıkmakta olup, yöntemin daha geniş örneklem gruplarında doğrulanması önerilmektedir.

(Wong, ve diğerleri, 2018), elektronik burun tabanlı bir nefes analiz sistemi ile akciğer kanseri teşhisini incelemektedir. Toplamda 45 nefes örneği üzerinde yapılan analizlerde, 33 akciğer kanseri hastası ve 12 sağlıklı birey yer almıştır. Elektronik burun sistemi, 14 MOS sensörü, veri toplama kartı ve bir gaz akış düzenleme biriminden oluşmaktadır. Boyut indirgeme için PCA ve LDA sınıflandırma için k-NN ve SVM gibi algoritmalar kullanılmış ve çapraz doğrulama ile test edilmiştir. PCA+k-NN analizinde %84.4 doğruluk oranıyla en yüksek başarı elde edilmiş, LDA+KNN ile doğruluk %75.5 olarak bulunmuştur; SVM ile %71.1 ila %73.3 arasında doğruluk sağlanmıştır. Sonuçlar, elektronik burun sisteminin akciğer kanseri taramasında umut vadettiğini göstermekte olup, sistemin güvenilir bir tarama aracı olarak daha geniş bir örneklemle doğrulanması önerilmektedir.

(Huang, ve diğerleri, 2018), kimyasal sensör dizisi ve makine öğrenme tekniklerini kullanarak akciğer kanseri teşhisi için nefes testi geliştirmeyi amaçlamaktadır. Toplamda 117 akciğer kanseri hastası ve 199 sağlıklı kontrol grubundan toplanan nefes örnekleri, 32 iletken polimer (CP) sensöre sahip Cyranose 320 elektronik burun cihazı ile analiz edilmiştir. Katılımcılar, VOC seviyelerini sabitlemek amacıyla 12 saat açlık sonrası bir tüp ile nefes örnekleri vermiştir. Veriler LR, SVM ve LDA ile analiz edilmiştir. LDA %90.2, SVM ise %92.7 doğruluk sağlamıştır. Sonuç olarak, elektronik burun teknolojisinin akciğer kanseri teşhisinde etkili bir tarama aracı olarak kullanılabileceği ancak daha geniş topluluklarda doğrulama çalışmalarına ihtiyaç duyulduğu vurgulanmıştır.

(van de Goor, van Hooren, Dingemans, Kremer, & Kross, 2018), taşınabilir bir elektronik burun (Aeonose) kullanarak akciğer kanseri teşhisinde nefes analizinin etkinliğini test etmeyi amaçlamaktadır. Toplamda 167 katılımcıdan alınan nefes örnekleri, 60 akciğer

kanseri hastası ve 107 sağlıklı bireyi içermektedir. Katılımcılar, Aeonose cihazıyla 5 dakika boyunca nefes vererek VOC verisi sağlamıştır ve veriler ANN modeli ile analiz edilmiştir. Eğitim setinde %83 duyarlılık, %84 özgüllük ve %83 genel doğruluk elde edilmiş; doğrulama setinde %88 duyarlılık, %86 özgüllük ve %86 genel doğruluk sağlanmıştır. Sonuçlar, taşınabilir elektronik burun cihazının akciğer kanseri taramasında umut vadeden bir araç olduğunu göstermektedir. Ancak yöntemin daha geniş hasta gruplarında ve erken evre tümörlerin teşhisinde doğrulanması gerektiği vurgulanmaktadır.

(Chang, ve diğerleri, 2018), akciğer kanserini erken evrede teşhis edebilmek amacıyla MOS sensörlerden oluşan bir elektronik burun sistemi geliştirilmiştir. Çalışmada, 37 KHDAK hastası ve 48 sağlıklı bireyden alınan nefes örnekleri incelenmiştir. Elektronik burun sistemi, 7 adet MOS sensörün yanı sıra gaz akış kontrol modülü, ısıtma modülü ve gaz adsorpsiyon-desorpsiyon modüllerinden oluşmaktadır. Veriler, Çok Katmanlı Algılayıcı ve SVM gibi sınıflandırma algoritmaları kullanılarak analiz edilmiştir. İlk örnek analizinde Çok Katmanlı Algılayıcı kullanılarak %75 doğruluk, %79 duyarlılık ve %72 özgüllük elde edilmiştir. İkinci ve üçüncü örnek analizlerinde doğruluk oranlarında kademeli bir düşüş gözlemlenmiştir. Çalışma sonucunda, elektronik burun sisteminin nefes analizi ile akciğer kanseri teşhisinde umut vaat eden bir araç olduğu belirtilmiş ve sistemin daha geniş bir örneklem grubunda test edilmesi önerilmiştir.

(Kononov, ve diğerleri, 2019), akciğer kanserinin teşhisi için MOS sensörlerden oluşan bir elektronik burun sistemi geliştirmeyi ve etkinliğini doğrulama yöntemleriyle değerlendirmeyi amaçlamaktadır. Toplamda 65 akciğer kanseri hastası ve 53 sağlıklı bireyden oluşan örneklem grubundan elde edilen veriler, üç farklı sıcaklık rejiminde çalışan altı farklı MOS sensörü ile toplanmıştır. LR, RF, k-NN, LDA, SVM gibi çeşitli sınıflandırma algoritmaları ile analiz edilen verilerde en yüksek başarı, %95 duyarlılık, %100 özgüllük ve %97.2 doğruluk oranı ile LR ve k-NN modellerinde elde edilmiştir. Sonuçlar, MOS tabanlı elektronik burun sisteminin akciğer kanseri teşhisinde yüksek doğruluk sağlayan etkili bir tanı aracı olduğunu göstermektedir; ancak, geniş çaplı klinik çalışmalarla daha fazla doğrulanması gerektiği vurgulanmaktadır.

(Lu, Fu, Nie, Peng, & Liu, 2019), elektronik burun teknolojisinin akciğer kanserini teşhis etme potansiyelini araştırmaktadır. Toplamda 98 akciğer kanseri hastası ve 116 sağlıklı bireyden oluşan 214 katılımcının nefes örnekleri, TGS2620 ve TGS2602 gibi 13 farklı sensör içeren bir elektronik burun cihazı ile toplanmıştır. Çapraz doğrulama sonucunda %93.55 doğruluk, %94.22 duyarlılık ve %92.80 özgüllük oranları elde edilmiştir.

(Marzorati D. , ve diğerleri, 2019), MOS sensörlerden oluşan bir elektronik burun sistemi ile akciğer kanseri teşhisi için bir nefes analiz yöntemi geliştirmeyi amaçlamaktadır. Çalışmada 10 sağlıklı birey ve 6 akciğer kanseri hastasından alınan nefes örnekleri analiz edilmiştir. TGS2600, TGS2602, TGS822 ve MQ3 olmak üzere dört farklı MOS sensöründen oluşan sistem, her sensörün farklı bileşiklere duyarlılığı sayesinde VOC profillerini tespit etmiştir. Veriler, ANN modeli ile analiz edilmiş ve %85.7 duyarlılık, %100 özgüllük ve %93.8 doğruluk oranı sağlanmıştır. Sağlıklı ve kanserli bireylerin nefes profilleri arasında belirgin VOC farklılıkları tespit edilmiştir.

(Li, ve diğerleri, 2020), akciğer kanserini teşhis etmek için geliştirilmiş ev yapımı bir elektronik burun sisteminin etkinliğini araştırmaktadır. Toplam 268 nefes örneğinden 153'ü sağlıklı bireylerden, 115'i akciğer kanseri hastalarından toplanmıştır. Sistem, dört farklı tipte toplam 10 gaz sensörü ve birer adet nem ve sıcaklık sensörü içermektedir. Veriler, LDA ve Hızlı Bağımsız Bileşenler Analizi ile özellik çıkarımı yapılarak, Gürültüye Duyarlı Yoğunluk Tabanlı Kümeleme ve Benzerlik Yayılımı ile Kümeleme algoritmalarıyla işlenmiştir; sınıflandırmada ise LR, RF ve SVM stratejileri kullanılmıştır. Kümeleme algoritmaları, %95'in üzerinde Düzeltilmiş Rand İndeksi değerleri ile yüksek başarı göstermiştir. Üç katlı çapraz doğrulama ile değerlendirilen sınıflandırma algoritmaları arasında, RF algoritması %86,42 doğruluk oranı ile en iyi performansı sergilemiştir. Bu sonuçlar, çeşitlendirilmiş sensörlerden oluşan elektronik burun sisteminin akciğer kanseri teşhisinde umut vadeden bir yöntem olduğunu göstermekte olup, geniş çaplı klinik doğrulama çalışmalarına ihtiyaç duyulduğu belirtilmiştir.

(Zhan, ve diğerleri, 2020), akciğer kanserinin teşhisinde elektronik burun sisteminin güvenilirlik ve doğruluğunu artırmak amacıyla konformal tahmin yöntemini kullanmayı amaçlamaktadır. Çalışmaya, son altı ayda kemoterapi veya radyoterapi görmemiş 31 akciğer kanseri hastası ve 41 sağlıklı birey dahil edilmiştir. Elektronik burun sistemi 16 adet MOS sensör içermekte olup, veriler en yakın komşu (1-NN ve 3-NN) algoritmalarına dayalı konformal tahmin yöntemi ile analiz edilmiştir. Çapraz doğrulama yöntemi kullanılarak %87.5 doğruluk (1NN) ve %83.33 doğruluk (3NN) elde edilmiştir. Konformal tahmin yöntemi, her tahmin için güven ve inanılabilirlik bilgisi sağlayarak teşhis sürecinde belirsizliği azaltmıştır. Bu sonuçlar, elektronik burun sisteminin bir teşhis aracı olarak güvenilir bir potansiyel sunduğunu ve konformal tahminin teşhiste güvenilirliği artırmada önemli bir yöntem olduğunu göstermektedir.

(Binson & Subramoniam, 2021), akciğer kanserini teşhis etmek amacıyla elektronik burun sistemi geliştirmeyi amaçlamaktadır. Toplamda 24 akciğer kanseri hastası ve 40 sağlıklı bireyden oluşan 64 katılımcının nefes örnekleri, beş adet MOS sensörden oluşan bir sensör dizisi ile analiz edilmiştir. Nefes örnekleri, Tedlar torbalarına üfleme yöntemiyle toplanmış ve veri toplama Arduino Uno aracılığıyla yapılmıştır. Python'da geliştirilen makine öğrenme algoritmaları ile veri analizi gerçekleştirilmiş; RF algoritması %85.38 doğruluk ile en iyi performansı göstermiştir. LR algoritması ise %83.59 doğruluğa ulaşmıştır. Çalışma, düşük maliyetli, taşınabilir ve hızlı bir tanı aracı olarak geliştirilen elektronik burun sisteminin, akciğer kanseri ve diğer solunum yolu hastalıklarının teşhisinde geniş çapta kullanılabileceğini öne sürmektedir.

(Marzorati, ve diğerleri, 2021), akciğer kanserini erken evrede teşhis etmek için MOS sensörlerden oluşan bir elektronik burun sistemi kullanarak nefes analizi yapmayı amaçlamaktadır. Çalışmaya, akciğer kanseri olan ve kanser riski taşıyan toplam 80 kişi katılmıştır. Her bir sensörün tepkisinden dokuz öznitelik çıkarılarak bireysel "nefes izi" profilleri oluşturulmuştur. Elde edilen veriler, klinik özelliklerle birlikte bir sınıflandırma modelinde kullanılmıştır ve 4 kat çapraz doğrulama yöntemiyle %77 doğruluk, %80 kesinlik ve %78 duyarlılık oranları elde edilmiştir. Sonuç olarak, sistemin performansını artırmak için sensör sayısının artırılması ve özel bileşiklere duyarlı sensörlerin eklenmesi gerektiği önerilmiştir.

(Binson V. A., Subramoniam, Ragesh, & Kumar, 2021), akciğer kanserinin erken teşhisi için elektronik burun sisteminin etkinliğini değerlendirmiştir. Erken evre akciğer kanseri tanısı konmuş 10 hasta ve 15 sağlıklı kontrol bireyinden oluşan 25 katılımcının nefes örnekleri, beş adet MOS sensörlü bir elektronik burun sistemi kullanılarak analiz edilmiştir. VOC profillerini oluşturan veriler, Arduino Uno ile toplanmış ve boyut indirgeme işlemi uygulanmıştır. Sınıflandırma için AdaBoost algoritması kullanılarak %76 doğruluk, %80 özgüllük ve %70 duyarlılık elde edilmiştir. Boyut indirgeme yapılmadığında doğruluk oranının %72'ye düştüğü gözlemlenmiştir. Sonuç olarak, elektronik burun sistemi akciğer kanseri için umut vadeden bir tarama aracı olarak değerlendirilmiş olup, performansın artırılması için daha geniş örneklem grupları ile doğrulama çalışmaları önerilmektedir.

(Binson, Subramoniam, & Luke, 2021), KOAH ve akciğer kanserini teşhis etmek için MOS sensörlerinden oluşan bir elektronik burun sistemi geliştirmeyi ve bu hastalıkları sağlıklı bireylerden ayırt etme yeteneğini değerlendirmeyi amaçlamaktadır. Araştırmaya, 93 sağlıklı birey, 55 KOAH hastası ve 51 akciğer kanseri hastası olmak üzere toplam 199

katılımcı dahil edilmiştir. Elektronik burun sistemi TGS2600, TGS2610, TGS2620, TGS822 ve TGS826 sensörlerinden oluşmaktadır. Örnekler Tedlar torbalarında toplanmış ve veri toplama Arduino Uno ile gerçekleştirilmiştir. Veriler, Kernel-PCA ve çeşitli sınıflandırma algoritmaları ile analiz edilmiştir. Topluluk öğrenmesi yöntemleri arasında XGBoost, AdaBoost ve RF algoritmaları kullanılmıştır. XGBoost algoritması, akciğer kanseri teşhisinde %93.04 doğruluk, %88.37 duyarlılık ve %95.83 özgüllük sağlamış; KOAH teşhisinde ise %90.68 doğruluk, %88.37 duyarlılık ve %92.00 özgüllük elde etmiştir. Model doğrulama aşamasında XGBoost'un akciğer kanseri teşhisi için %79.31 doğruluk ve %84.21 özgüllük; KOAH teşhisi için %76.67 doğruluk ve %83.33 özgüllük sağladığı gözlemlenmiştir. Çalışma, elektronik burun sisteminin KOAH ve akciğer kanserinin teşhisinde umut vaat ettiğini göstermektedir.

(Chen, ve diğerleri, 2021), elektronik burun sisteminin akciğer kanseri teşhisi ve evrelemesindeki etkinliğini değerlendirmektedir. Toplam 101 akciğer kanseri hastası ve 134 sağlıklı bireyden oluşan 235 nefes örneği analiz edilmiştir. Elektronik burun sistemi, 13 farklı sensörden oluşan bir dizi kullanarak nefes örneklerindeki VOC'leri tespit etmiştir. Veriler Kernel-PCA ile işlenmiş ve SVM ile XGBoost algoritmalarıyla sınıflandırılmıştır. Kernel-PCA-XGBoost kombinasyonu, %93.59 doğruluk, %95.60 duyarlılık ve %91.09 özgüllükle en yüksek performansı sağlamıştır. Kanser evrelemesinde (III. Evre ve IV. Evre) %82.42 doğruluk oranına ulaşılmıştır. Çalışma, elektronik burun sisteminin akciğer kanserinin teşhisinde etkili bir araç olduğunu ve özellikle ileri evre kanserlerin sınıflandırılmasında umut vaat ettiğini göstermektedir. Ancak, I. Evre ve II. Evre kanserlerin teşhisinde daha fazla örneklem ile doğrulama çalışmaları yapılması gerektiği belirtilmiştir.

(Liu, ve diğerleri, 2021), akciğer kanseri teşhisinde VOC'lerin nefes analizi ile tespit edilmesi için geliştirilmiş bir elektronik burun sisteminin etkinliğini değerlendirmektedir. Çalışmaya 46 akciğer kanseri hastası, 36 sağlıklı birey ve 5 iyi huylu akciğer hastalığı olan birey olmak üzere toplam 87 katılımcı dahil edilmiştir. Elektronik burun sistemi VOC algılama için 19 farklı gaz sensörü (MOS ve elektrokimyasal) ile donatılmış ve termal desorpsiyon ön yoğunlaştırma ünitesi kullanılarak VOC yoğunluğu artırılmıştır. Öznitelik seçimi yapılmış ve SVM, k-NN ve Naive Bayes gibi sınıflandırıcılar kullanılarak veriler analiz edilmiştir. SVM, %94.25 doğruluk, %97.83 duyarlılık, %90.24 özgüllük değeri ile en iyi performansı göstermiştir. Ayrıca, demografik faktörlerin sınıflandırma performansını anlamlı şekilde etkilemediği bulunmuştur.

(Liu, ve diğerleri, 2021), yaptıkları çalışmada 98 akciğer kanseri hastası ve 116 sağlıklı birey olmak üzere toplam 214 katılımcının nefes örneklerini analiz etmiştir. Elektronik burun sistemi, farklı tiplerde 11 gaz sensöründen oluşmaktadır ve nefes örneklerinde bulunan VOC'lerin tespit edilmesi için tasarlanmıştır. SVM, Karar Ağaçları (DT), RF, LR ve k-NN algoritmalarını içermektedir. Çalışmada ulaşılan en başarılı sonuç; %96,11 doğruluk, %95,27 duyarlılık, %97,17 özgüllük ve %96,31 F1 skoru ile, modelin hem genel başarı hem de sınıf ayırt etme açısından yüksek performans sergilediğini göstermektedir. Ayrıca, bu sistemin, uzun süreli sigara içenler ve KOAH hastaları gibi yüksek risk gruplarını ayırt etmede de etkili olduğu görülmüştür. Bu bulgular, elektronik burun sisteminin akciğer kanserinin erken teşhisi için umut vadeden bir araç olduğunu ve topluluk öğrenme yöntemlerinin tanısal doğruluğu artırmada önemli bir rol oynadığını göstermektedir.

(Binson V. A., Subramoniam, Sunny, & Mathew, 2021), akciğer kanseri, KOAH ve astım gibi akciğer hastalıklarının nefes analizi yoluyla tespitini amaçlamaktadır. Çalışmaya toplam 218 katılımcı dahil edilmiştir; bunların 48'i akciğer kanseri, 52'si KOAH, 55'i astım hastası ve 63'ü sağlıklı kontrol bireyidir. Çalışmada kullanılan elektronik burun sistemi, dört farklı sensör türünden oluşan sekiz gaz sensörü içermektedir. Örneklemeye için katılımcıların nefes örnekleri Tedlar torbalarında toplanmış ve sensörlerin yanıtları analiz edilmiştir. Veriler, farklı boyut indirgeme yöntemleri (PCA, LDA, Kernel-PCA) ve iki makine öğrenme algoritması (SVM ve XGBoost) kullanılarak işlenmiştir. Akciğer kanseri tespiti için Kernel-PCA+XGBoost modeli %91.74 doğruluk, %90.57 duyarlılık ve %92.65 özgüllük ile en iyi performansı göstermiştir. KOAH tespiti için aynı model %89.84 doğruluk, %88.14 duyarlılık ve %91.30 özgüllük değerlerine ulaşmıştır. Astım tespiti için başarı oranları daha düşük kalmış (%70.66 doğruluk, %68.75 duyarlılık, %72.41 özgüllük) ve bu, sistemin astım teşhisinde geliştirilmesi gerektiğini ortaya koymuştur. Ayrıca, akciğer kanseri ve KOAH hastaları da %84.75 doğruluk, %81.36 duyarlılık ve %88.14 özgüllük ile ayırt edilmiştir. Sonuç olarak bu çalışma, elektronik burun teknolojisinin akciğer hastalıklarının erken teşhisinde etkili ve taşınabilir bir yöntem olarak kullanılabileceğini göstermektedir. Ancak, astım teşhisindeki sınırlamalar ve erken evre akciğer kanseri tanısında daha fazla veri ihtiyacı gibi eksikliklerin giderilmesi gerektiği belirtilmiştir.

(Bharathy, Pavithra, & Akshaya, 2022), makine öğrenimi algoritmalarını kullanarak akciğer kanserinin erken teşhisi için bir elektronik burun sisteminin etkinliğini incelemektedir. Çalışmada Kaggle platformundan elde edilen 303 örnekte oluşan bir veri seti kullanılmıştır. Veriler yaş, cinsiyet, kan basıncı, kolesterol seviyeleri, kan şekeri,

elektrokardiyogram verileri ve maksimum kalp atış hızı gibi klinik parametreleri içermektedir. Eğitim (%75) ve test (%25) setleri olarak ayrılan veriler, LR, RF, DT, Naive Bayes ve SVM gibi algoritmalarla analiz edilmiştir. LR ve RF algoritmaları %88.5 doğruluk oranı ile en iyi performansı göstermiştir. Sonuçlar erken evre akciğer kanseri teşhisinde doğruluğun artırılması için sistemin daha büyük veri setleriyle test edilmesi gerektiği vurgulanmıştır.

(Hao & Huang, 2023), akciğer kanserinin teşhisinde kullanılmak üzere geliştirilen elektronik burun sistemiyle, AdaBoost algoritmasının etkinliği değerlendirilmiştir. Çalışmada, 91 akciğer kanseri hastası ve 51 sağlıklı bireyden oluşan bir eğitim seti ve 12 akciğer kanseri hastası ile 10 sağlıklı bireyden oluşan bir test seti kullanılmıştır. Elektronik burun sistemi, üç adet elektrokimyasal sensör, bir adet MOS sensörü ve kontrol edilebilir sıcaklık sensörü ile donatılmıştır. Sensör verilerinden zaman, frekans ve istatistiksel özelliklere dayalı toplam 1557 öznitelik çıkarılarak analiz edilmiştir. Boyut indirgeme için PCA kullanılmıştır. Geliştirilen AdaBoost algoritması, 5 katlı çapraz doğrulama ve oylama yöntemi kullanılarak oluşturulan farklı türdeki alt sınıflandırıcıları bir araya getirerek nihai modeli oluşturmuştur. Testlerde, %97.85 doğruluk, %98.33 duyarlılık, %97 özgüllük değeri elde edilmiştir. Algoritma, yüksek duyarlılığı ile akciğer kanserini doğru bir şekilde teşhis ederken, yanlış negatif oranını da düşük tutmayı başarmıştır. Sonuçlar, elektronik burun sisteminin erken akciğer kanseri teşhisinde etkili bir yöntem olduğunu göstermiştir. Ancak, sistemin daha geniş popülasyonlarda test edilmesi ve örnek çeşitliliğinin artırılması gerektiği belirtilmiştir.

(Kort, ve diğerleri, 2023), elektronik burun teknolojisi kullanılarak KHDAAK hastalarını nefes profili analizi yoluyla teşhis etmeyi ve bu yöntemin klinik uygulanabilirliğini değerlendirmeyi amaçlamaktadır. Yedi merkezde toplam 575 katılımcıyla yapılan çalışmada, 376 kişi eğitim setinde, 199 kişi ise doğrulama setinde yer almıştır. Katılımcılar, taşınabilir bir elektronik burun cihazı olan "aeoNose" ile 5 dakika boyunca nefes analizi yapmış ve cihazın VOC'lere duyarlı MOS sensörler ile nefes profilleri oluşturmuştur. Makine öğrenimi algoritmaları kullanılarak tahmin modelleri geliştirilmiş ve doğrulama setinde %88 duyarlılık, %52 özgüllük değeri elde edilmiştir. Ayrıca, yaş, cinsiyet, sigara içme geçmişi ve diğer klinik parametrelerin eklenmesiyle modelin doğruluğu artırılmış, %95 duyarlılık değeri sağlanmıştır. Sonuçlar, nefes analizi ve elektronik burun teknolojisinin akciğer kanserinin erken teşhisinde umut vaat ettiğini göstermektedir.

(Kao, Chiu, Lee, Sun, & Tang, 2023), elektronik burun sistemi ve derin öğrenme yöntemleri kullanarak akciğer kanserinin teşhisini geliştirmeyi amaçlamaktadır. Hsinchu ve Taipei'deki iki hastaneden toplanan verilerde, Hsinchu veri setinde 102 akciğer kanseri hastası ve 35 sağlıklı birey, Taipei veri setinde ise 35 akciğer kanseri hastası ve 11 sağlıklı birey yer almıştır. Elektronik burun sistemi, 14 sensörden oluşan bir dizi kullanılarak tasarlanmıştır. Her bir nefes örneği, 768 saniye boyunca toplanan verilerle analiz edilmiştir. Bu veriler, 256 veri noktasına indirgenmiş ve ardından her biri 14 kanallı 16x16 boyutunda görüntülere dönüştürülmüştür. Sınıflandırma işlemi için Evrişimsel Sinir Ağı modeli kullanılmıştır. Hsinchu veri setinde %93 doğruluk elde edilirken, transfer öğrenme yöntemi uygulanmadan önce Taipei veri setinde doğruluk %69.41 olmuştur; transfer öğrenme sonrası bu oran %86.41'e yükselmiştir. Sonuçlar, elektronik burun sistemleri ve transfer öğrenme ile derin öğrenme tekniklerinin akciğer kanserinin teşhisinde yüksek doğruluk oranları sağlayabileceğini göstermekte, ancak daha geniş veri setleri ve çapraz cihaz testlerine ihtiyaç duyulduğu vurgulanmaktadır.

(Moninuola, ve diğerleri, 2023), akciğer kanserinin erken teşhisine yönelik olarak, elektronik burun sistemi ve makine öğrenimi yöntemlerini bir araya getiren bir tanı yaklaşımını araştırmıştır. Toplamda 11 akciğer kanseri hastası, 11 KOAH hastası ve 15 sağlıklı bireyden oluşan 37 katılımcının nefes örnekleri, 5 adet MOS sensörden oluşan bir elektronik burun sistemi ile analiz edilmiştir. Veriler, Tedlar torbalarına alınan nefes örneklerinden toplanmış ve LDA ile boyut indirgeme yapılmıştır. Sınıflandırma, SVM, AdaBoost ve Çok Katmanlı Algılayıcı algoritmalarıyla gerçekleştirilmiş; Çok Katmanlı Algılayıcı modeli %89.05 doğruluk, %95.12 özgüllük ve %80 duyarlılık ile en iyi performansı göstermiştir. Çalışma, yöntemin genelleştirilebilirliğini artırmak için daha geniş popülasyonlarla ek çalışmalar yapılmasını önermektedir.

Son yıllarda yapılan çok sayıda çalışma, elektronik burun sistemlerinin akciğer kanseri teşhisinde umut vadeden bir araç olarak kullanılabileceğini ortaya koymuştur. Gerek klinik gerekse deneysel bağlamda yürütülen bu araştırmalar, akciğer kanseri hastalarının nefeslerinde bulunan VOC'lerin özgün profiller oluşturduğunu ve bu profillerin çeşitli sensör dizileri ile tespit edilebildiğini göstermektedir.

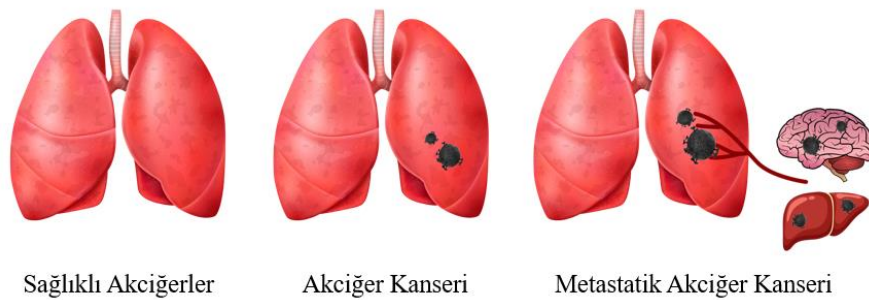
Bu sistemlerin en dikkat çekici avantajları; invaziv olmamaları, hızlı sonuç verebilmeleri, düşük maliyetli olmaları ve taşınabilirlik özellikleridir. Bu özellikleri sayesinde, akciğer kanseri gibi ölümcül hastalıkların erken teşhisinde tarama aracı olarak kullanılmaları mümkün görünmektedir. Ayrıca, birçok çalışmada elektronik burun

sistemlerinin yalnızca hastalık varlığını değil, farklı evreleri ve diğer solunum yolu hastalıklarıyla olan ayrımını da makul doğruluk oranlarıyla gerçekleştirebildiği rapor edilmiştir.

Bununla birlikte, çalışmalar arasında örneklem büyüklüğü, sensör tipi, veri toplama yöntemi ve sınıflandırma algoritmaları bakımından önemli farklılıklar bulunmaktadır. Ayrıca, bazı araştırmalarda elde edilen yüksek performans oranları, küçük ve homojen örneklemelerde test edilmiş sistemlerle sınırlı kalmıştır. Bu nedenle, geniş ölçekli, çok merkezli ve uzun dönemli klinik doğrulama çalışmaları, bu teknolojinin rutin klinik kullanıma geçebilmesi için kritik önem taşımaktadır.

2.2. Akciğer Kanseri

Akciğer kanseri, akciğerlerdeki normal hücresel büyüme ve çoğalma mekanizmalarının bozulması sonucu ortaya çıkan kötü huylu bir tümördür. Bu tümör, akciğerlerin bronş epitelinde, alveolar epitelde veya diğer akciğer dokularında anormal hücresel büyüme ve kontrolsüz çoğalması sonucu meydana gelir. Hücreler belirli bir düzen içinde bölünür ve ölü hücrelerin yerini almak için yenilenir. Ancak kanserli hücreler bu düzeni bozar ve sürekli olarak bölünerek ve çoğalarak kontrolsüz tümörler oluşturur. Bu tümörler, zamanla çevre dokulara ve akciğerlerin dışına yayılabilir. Metastaz (kanserin diğer organlara yayılması) süreci, tümör hücrelerinin kan veya lenf yoluyla vücudun farklı bölgelerine taşınmasıyla gerçekleşir, bu da hastalığın ilerlemesine ve tedavisinin zorlaşmasına neden olur (Hansen, 2008). Sonuç olarak, kötü huylu tümörler hem akciğerlerde hem de diğer organlarda yapısal ve fonksiyonel bozulmalara yol açarak ciddi sağlık sorunlarına yol açabilir. Akciğer kanserine ait tümör oluşum ve metastaz süreci Şekil 4'te verilmiştir.

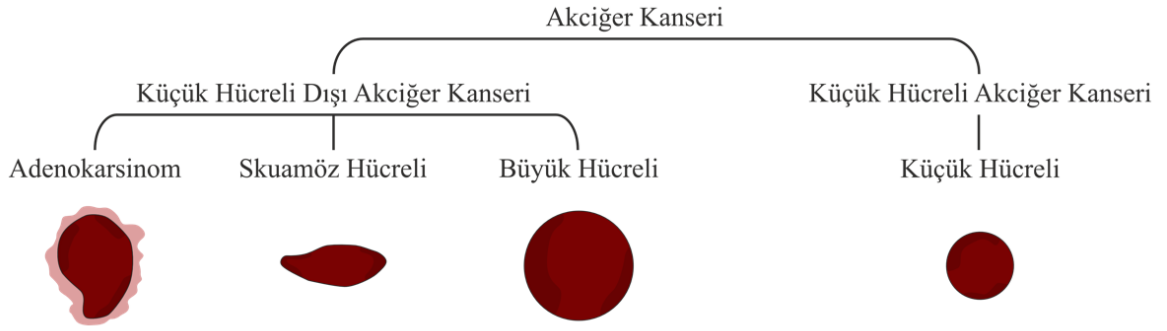


Şekil 4. Akciğer kanseri gelişim süreci

2.2.1. Akciğer Kanseri Türleri

Akciğer kanseri, küçük hücreli akciğer kanseri (KHAK) ve küçük hücreli dışı akciğer kanseri (KHDAK) olarak iki ana kategoriye ayrılır. KHAK hücreleri akciğerlerin merkezi bölgelerinde oluşur şekilleri ise küçük ve yuvarlaktır, çok hızlı bir şekilde bölünerek çoğalır ve yayılırlar (Rudin, Brambilla, Faivre-Finn, & Sage, 2021). Erken evrelerde bile lenf düğümlerine ve uzak organlara metastaz yapma eğilimi gösterirler.

KHDAK, tüm akciğer kanseri vakalarının yaklaşık %80'ini kapsayan daha yaygın bir türdür. Bu türün hücreleri daha büyük iken bölünme ve yayılma hızları ise daha yavaştır (Gridelli, ve diğerleri, 2015). KHDAK kendi arasında adenokarsinom, skuamöz hücreli karsinom ve büyük hücreli karsinom olmak üzere üçe ayrılır. Adenokarsinom, akciğerin dış bölgelerinde bulunan bir küçük hücreli olmayan akciğer kanseri türüdür. Vücudun iç ve dış yüzeylerini kaplayan, organların ve diğer yapılarının yüzeyini örten epitel doku hücrelerinde gelişir. Skuamöz hücreli karsinom, akciğerin merkezinde, bronş adı verilen hava yollarının yakınında ortaya çıkan bir türdür. Büyük hücreli karsinom, akciğerin her bölgesinde oluşabilen ve adenokarsinom veya skuamöz hücreli karsinomdan daha hızlı büyüyen ve yayılan bir türdür. Şekil 5, akciğer kanseri alt tiplerinin sınıflandırmasını göstermektedir.



Şekil 5. Akciğer kanseri alt tiplerinin sınıflandırılması

2.2.2. Akciğer Kanseri Risk Faktörleri

Akciğer kanseri, çeşitli faktörlerin etkisiyle gelişen bir hastalıktır. Hastalığın ortaya çıkma olasılığı, genetik, bireysel ve çevresel etkenlerle şekillenir (Malhotra, Malvezzi, Negri, La Vecchia, & Boffetta, 2016). Risk faktörlerinin tanımlanması, akciğer kanserinin

önlenmesi ve yönetilmesi açısından kritik öneme sahiptir. Şekil 6'da bu risk faktörlerini bütüncül bir şekilde özetleyen temsili bir görsel sunulmaktadır.



Şekil 6. Akciğer kanseri risk faktörleri

Akciğer kanseri riskini artıran genetik faktörler arasında ailevi yatkınlık, p-53 tümör baskılayıcı geninde görülen mutasyonlar ve kromozom anomalileri (6q23–25 kromozom bölgesindeki genetik duyarlılık bölgeleri) bulunmaktadır (Chiba, ve diğerleri, 1990; Bailey-Wilson, ve diğerleri, 2004).

Sigara içmek ve tütün mamulleri kullanmak, akciğer kanserinin en büyük risk faktörüdür. Sigara dumanı, kanserojen maddeler içerir ve bu maddeler akciğer hücrelerine zarar vererek genetik değişikliklere neden olarak tümör oluşumunu tetikler. Bilimsel veriler, sigara içmenin akciğer kanseri riskini artırdığını ve tütün dumanına maruz kalanlarda hastalığın sıklığını belirgin şekilde yükselttiğini göstermektedir (Boyle & Maisonneuve, 1995).

Pasif içicilik, başkalarının içtiği sigara dumanına maruz kalmayı ifade eder ve akciğer kanseri riskini artırır. Sigara dumanı, kanserojen maddeler içerdiğinden, pasif içiciler de bu zararlı maddelere maruz kalır. Araştırmalar, pasif içiciliğin akciğer kanseri riskini yükselttiğini, özellikle kapalı alanlarda uzun süre maruz kalan bireylerde bu riskin belirgin şekilde arttığını göstermektedir (Taylor, Najafi, & Dobson, 2007).

KOAH, akciğerlerin hava yollarında kalıcı hasara yol açar ve bu durum akciğer kanseri riskini artırabilir. KOAH hastalığı, hava yollarında sürekli iltihap ve oksidatif strese neden olarak akciğer hücrelerinde DNA hasarına yol açar. Bu durum, tümör oluşumunu teşvik

edebilir ve bu nedenle akciğer kanseri için önemli bir risk faktörüdür (Durham & Adcock, 2015).

Yaş, akciğer kanseri için önemli bir risk faktörüdür. Risk, genellikle yaşla birlikte artar ve çoğu akciğer kanseri vakası 65 yaş ve üzerindeki bireylerde görülür. Bu artış, yaşla birlikte biriken genetik değişiklikler ve uzun süreli çevresel risk faktörlerine maruz kalmayla ilişkilidir. Yaş, akciğer kanseri riskinin belirlenmesinde önemli bir faktördür, ancak diğer risk etmenleriyle birlikte ele alınmalıdır (Brown, Eraut, Trask, & Davison, 1996).

Cinsiyet, akciğer kanseri riskini etkiler. Genel olarak, erkekler akciğer kanseri için daha yüksek bir risk taşır; bu, erkeklerde sigara içme oranlarının genellikle daha yüksek olmasından kaynaklanır. Ancak, sigara içen kadınlarda akciğer kanseri riski sigara içen erkeklerden daha yüksektir. Kadınlarda akciğer hücrelerinde DNA katım seviyeleri daha yüksektir ve bu nedenle p53 tümör genlerinde mutasyonlar daha sık görülür. Bu durum, kadınların akciğer kanserine karşı daha duyarlı olduğunu gösterir. Cinsiyet, akciğer kanseri riskinin belirlenmesinde önemli bir faktördür, ancak diğer risk etmenleriyle birlikte ele alınmalıdır (Stapelfeld, Dammann, & Maser, 2020).

Akciğer kanserinin ortaya çıkmasında önemli rol oynayan çeşitli çevresel faktörler bulunmaktadır. Radon gazı radyoaktif bir gazdır ve uzun süreli radon maruziyeti akciğer kanseri riskini önemli ölçüde artırabilir (Ruidavets, Garcia de Herreros, Besse, & Mezquita, 2022). Endüstriyel süreçlerde kullanılan krom ve nikel gibi metallerin solunması da riski artıran çevresel etmenler arasındadır (Beveridge, Pintos, Parent, Asselin, & Siemiatycki, 2010). Ayrıca, kömür isisi ve katran solunması akciğerlerde birikerek kanser riskini artırabilir. Asbest solunması da akciğerlerde ciddi sağlık sorunlarına yol açarak akciğer kanseri riskini artırır (Klebe, Leigh, Henderson, & Nurminen, 2020). Arsenik içeren su ve toprakla teması, bu zehirli elementin vücuda geçmesine ve kanser riskinin artmasına neden olabilir (Wei, Zhang, & Tao, 2019). Son olarak, ağır hava kirliliği, şehirlerde ve endüstriyel bölgelerde yaygın bir sorundur ve partikül madde, azot dioksit ve sülfür dioksit gibi kirleticilere uzun süreli maruziyet, akciğerlerde iltihaplanma ve DNA hasarına yol açarak kanser riskini artırabilir. Bu çevresel faktörlerin her biri, akciğer kanserinin gelişiminde kritik bir rol oynar ve bu nedenle detaylı bir şekilde incelenmeleri gerekmektedir (Pope III, ve diğerleri, 2002).

2.2.3. Akciğer Kanseri Belirtileri

Akciğer kanseri, genellikle ileri evrelere ulaşana kadar belirgin belirtiler göstermeyebilir, bu da erken teşhisi zorlaştırır. Hastalık ilerledikçe, çeşitli semptomlar ortaya çıkabilir. Bu belirtiler genellikle solunum sistemiyle ilgili sorunlarla ilişkilidir ve bazı durumlarda diğer sağlık sorunlarıyla karışabilir. Akciğer kanserinin tanısında semptomlar kritik bir rol oynar; bu nedenle, herhangi bir belirgin solunum problemi veya sürekli rahatsızlık hissi, sağlık uzmanına başvurmayı gerektirir. Erken teşhis, tedavi seçeneklerini artırabilir ve yaşam kalitesini iyileştirebilir. Yapılan geniş çaplı bir meta analizde akciğer kanseri hastalarında görülen en yaygın belirtiler aşağıda verilmiştir (Cooley, 2000).

- Öksürük
- Nefes darlığı
- Boğaz ağrısı
- Göğüs ağrısı
- Disfaji (yutkunma güçlüğü)
- Aşırı balgam üretimi ve balgamdan kan gelmesi (hemoptizi)
- Hırıltı
- Ses kısıklığı
- Halsizlik
- İştahsızlık
- Kilo kaybı

2.2.4. Akciğer Kanseri Tanı Yöntemleri

Akciğer kanseri tedavisinde erken teşhis, tedavi sürecinin başarısını önemli ölçüde artırabilir. Akciğer kanserinin doğru şekilde teşhis edilmesi için çeşitli yöntemler kullanılır. Bu yöntemler arasında görüntüleme teknikleri, biyopsi işlemleri, laboratuvar testleri ve genetik analizler bulunur. Bu kısımda, akciğer kanserinin tanısında kullanılan başlıca yöntemler verilmiştir.

Göğüs Röntgeni basit ve hızlı bir görüntüleme yöntemidir. Akciğerlerdeki büyük kitleleri veya anormallikleri gösterebilir, ancak küçük lezyonlar veya erken evre hastalıklar için yeterli olmayabilir (Gavelli & Giampalma, 2000).

Bilgisayarlı Tomografi akciğerlerin detaylı kesitsel görüntülerini sağlar ve tümörlerin boyutunu, şeklini ve yayılma durumunu daha iyi değerlendirebilir. Ayrıca, lenf düğümleri veya diğer organlarda metastaz olup olmadığını belirlemede yardımcı olur (Bach, ve diğerleri, 2007).

Pozitron Emitron Tomografisi vücutta hücrel aktiviteyi gösterir ve kanser hücrelerinin metabolik aktivitesini değerlendirebilir. Genellikle Bilgisayarlı Tomografi ile birlikte kullanılır ve böylece tümörlerin ve metastazların yerleri daha iyi belirlenir (Fischer & Mortensen, 2006).

Manyetik Rezonans Görüntüleme de Bilgisayarlı Tomografi taramaları gibi vücuttaki yumuşak dokuların ayrıntılı görüntülerini gösterir. Ancak, x-ışınları yerine radyo dalgalarını ve güçlü mıknatısları kullanır. Bu yöntem çoğunlukla akciğer kanserinin beyne, omuriliğe veya karaciğere olası yayılımını araştırmak için kullanılır (Koyama, ve diğerleri, 2013).

Biyopsi, ince bir iğne kullanılarak göğüs duvarı üzerinden akciğerlerdeki kitlelerden veya nodüllerden doku örneği alınır. Bu yöntem, genellikle yüzeysel veya erişilebilir lezyonlar için kullanılır (Mazzone, Jain, Arroliga, & Matthay, 2002).

Bronkoscopi akciğerlerin içini incelemek ve buradan doku örneği almak için bronkoskop adı verilen ince bir tüp kullanılan bir yöntemdir. Tüp, ağız veya burun yoluyla ilerletilir. Bronkoscopi sırasında, akciğerlerden veya bronşlardan anormal bölgelerden doku örneği alınabilir (Mazzone, Jain, Arroliga, & Matthay, 2002).

Endobronşiyal Ultrason, Bronkoscopi sırasında ultrason teknolojisi kullanarak, bronşların ve çevresindeki lenf düğümlerinin daha ayrıntılı bir görüntüsünü sağlar. Küçük iğneler kullanarak lenf düğümlerinden ve diğer bölgelerden biyopsi almak için kullanılabilir (Gomez & Silvestri, 2009).

Torakoscopi, göğüs boşluğunun içini doğrudan görüntülemek için kullanılan bir yöntemdir. İnce, uzun bir tüp olan torakoskop, küçük bir kesiden göğüs boşluğuna yerleştirilir. Torakoskop, bir kamera ve ışık kaynağı içerir, böylece göğüs boşluğunun içi detaylı bir şekilde görüntülenebilir. Bu yöntemle anormal kitleler ve nodüller doğrudan görülebilir ayrıca doku örneği (biyopsi) alınabilir (Landreneau, ve diğerleri, 1998).

Balgam sitolojisi, akciğer hastalıklarının tanısında kullanılan bir laboratuvar testidir. Bu yöntem, balgam örneklerinde anormal veya kanserli hücrelerin varlığını incelemek için kullanılır (Thunnissen, 2003).

Moleküler testler, hücrel ve genetik düzeyde değişiklikleri tespit etmek için kullanılan testlerdir. Bu testler, DNA, RNA veya protein düzeyindeki değişiklikleri analiz

ederek kanserin türünü, evresini ve en uygun tedavi yöntemini belirlemeye yardımcı olur (Imyanitov & Iyevleva, 2021).

Akciğer kanseri gibi hastalıklarda, kan testleri genellikle doğrudan tanı koymak için kullanılmaz, ancak hastalığın değerlendirilmesinde ve tedavi sürecinin izlenmesinde yardımcı olabilir. Kan testleri sonuçları genellikle doktorlar tarafından diğer klinik bulgular, görüntüleme testleri ve biyopsi sonuçları ile birlikte değerlendirilir (Montani, ve diğerleri, 2015).

2.2.5. Geleneksel Tanı Yöntemlerinin Kısıtlamaları ve Nefes Analizine Yönelik Gereksinim

Akciğer kanseri tanısında kullanılan mevcut yöntemler, yüksek doğruluk oranlarına sahip olmakla birlikte, genellikle ileri evrelerde etkili sonuçlar vermektedir. Ayrıca bu yöntemlerin önemli bir kısmı invaziv nitelikte olup, hasta konforunu olumsuz etkileyebilmekte, maliyet ve işlem süresi gibi pratik sınırlılıklarla karşılaşmaktadır (Pulumati, Pulumati, Dwarakanath, Verma, & Papineni, 2023). Bu nedenlerle, özellikle erken evrede tanı koyabilecek, invaziv olmayan, kolay uygulanabilir ve düşük maliyetli alternatif tanı yaklaşımlarına yönelik gereksinim giderek artmaktadır.

Bu noktada, insan nefesi üzerinden yapılan analizler dikkat çekici bir potansiyel sunar. Nefes, vücuttaki metabolik süreçlerin bir yansıması olarak birçok VOC içerir. Akciğer dokusunda meydana gelen kanser gibi hastalıklar, bu bileşiklerin yapısını ve yoğunluğunu değiştirebilir. Dolayısıyla, nefes içerisindeki bu kimyasal değişiklikleri tespit etmek, hastalığın erken evrelerde tanınmasına yardımcı olabilecek biyobelirteçlerin belirlenmesini mümkün kılabilir.

Nefes analizine olan ilgi sadece bilimsel doğruluğundan değil, aynı zamanda toplum temelli tarama programlarında uygulanabilirliğinden kaynaklanmaktadır. Bu yöntem invaziv olmadığı için tekrarlanabilirliği yüksektir ve hastalar açısından oldukça konforludur. Özellikle risk grubundaki bireylerin düzenli olarak takip edilmesini kolaylaştırarak erken tanı şansını artırabilir. Aynı zamanda, klasik tanı yöntemlerinin uygun olmadığı durumlarda da etkili bir alternatif sunabilir.

Bu nedenlerle, akciğer kanseri başta olmak üzere birçok solunum sistemi hastalığında, nefes analizine dayalı tanı yöntemlerinin teorik temelleri güçlendirilmekte ve bu alan, klinik uygulamaya geçiş açısından önemli bir araştırma sahası haline gelmektedir.

2.2.6. Akciğer Kanseri Evreleme

Akciğer kanserinin evrenmesi, tümörün boyutu ve hastalığın yayılım derecesini sistematik olarak tanımlar. Bu sınıflama, tedavi planlarının oluşturulmasında kritik bir rol oynar. Temel amacı, akciğer kanserinin klinik yönetimini ve araştırmalarını uluslararası düzeyde tutarlı ve standart bir terminolojiyle tanımlamaktır. Evreleme, genellikle T (tümör), N (lenf düğümü) ve M (metastaz) parametrelerine dayalı olarak yapılır. TNM sistemi olarak adlandırılan bu sınıflama (Detterbeck, Boffa, Kim, & Tanoue, 2017) belirli aralıklarla güncellenen uluslararası standartlara dayanmaktadır. TNM evreleme sistemine ilişkin detaylı sınıflama, Tablo 1’de verilmiştir.

- T (Tümör): Tümörün boyutunu ve yerini ifade eder. "T" harfi, tümörün akciğerdeki yerini ve büyüklüğünü belirten bir numara ile birlikte kullanılır.
- N (Lenf düğümü): Lenf düğümlerinin etkilenme durumunu belirtir. "N" harfi, kanserin lenf düğümlerine yayılıp yayılmadığını ve eğer yayıldıysa ne kadar yayıldığını gösterir. Örneğin, N0 lenf düğümlerine yayılma olmadığını, N3 ise kanserin uzak lenf düğümlerine yayıldığını ifade eder.
- M (Metastaz): Kanserın vücudun diğer bölgelerine yayılıp yayılmadığını belirtir. "M" harfi, metastaz olup olmadığını ve eğer varsa ne kadar yayılmış olduğunu gösterir. M0 metastaz olmadığını, M1 ise kanserin uzak organlara yayıldığını gösterir.

Tablo 1. TNM sınıflamasına göre akciğer kanseri evreleme kriterleri

T	T0	Tümör belirtisi yoktur.
	T1	Tümör 3 cm'den küçüktür ve akciğerin bir lobunda sınırlıdır.
	T1a	Tümör 1 cm'den küçüktür.
	T1b	Tümör 1-2 cm arasındadır.
	T1c	Tümör 2-3 cm arasındadır.
	T2	Tümör 3-5 cm arasında olabilir ya da bronşlara, göğüs duvarına, diyaframa veya göğüs orta boşluğuna yayılmış olabilir.
	T2a	Tümör 3-4 cm arasındadır.
	T2b	Tümör 4-5 cm arasındadır.

Tablo 1'in devamı

	T3	Tümör 5-7 cm arasında veya akciğer dışındaki yapılara yayılım göstermiştir.
	T4	Tümör herhangi bir büyüklükte olabilir ve göğüs kafesinin dışına yayılım göstermiş olabilir.
	N0	Bölgesel lenf düğümü metastazı yoktur.
	N1	Tümör, akciğerin aynı tarafındaki (ipsilateral) bronş veya ilk düzey lenf düğümlerine metastaz yapmıştır.
N	N2	Tümör, akciğerin aynı tarafındaki daha uzak lenf düğümlerine metastaz yapmıştır.
	N3	Tümör, karşı taraf (kontralateral) lenf düğümlerine veya boyun bölgesindeki lenf düğümlerine metastaz yapmıştır.
M	M0	Uzak metastaz yoktur.
	M1a	Tümör, akciğer dışı yapılara sınırlı metastaz yapmıştır.
	M1b	Tümör, tek bir uzak organ veya dokuya metastaz yapmıştır.
	M1c	Tümör, birden fazla uzak organ veya dokuya metastaz yapmıştır.

Akciğer kanserinde TNM sistemine göre belirlenen tümör boyutu (T), lenf düğümü (N) ve metastaz (M) durumlarının çeşitli kombinasyonları, hastalığın klinik evresini oluşturmaktadır. Bu kombinasyonlara göre tanımlanan klinik evre gruplamaları, Uluslararası Akciğer Kanseri Çalışmaları Derneği tarafından önerilen ve TNM sınıflamasının 8. baskısında yer alan düzenlemelere göre aşağıdaki Tablo 2'de sunulmuştur (Goldstraw, ve diğerleri, 2016).

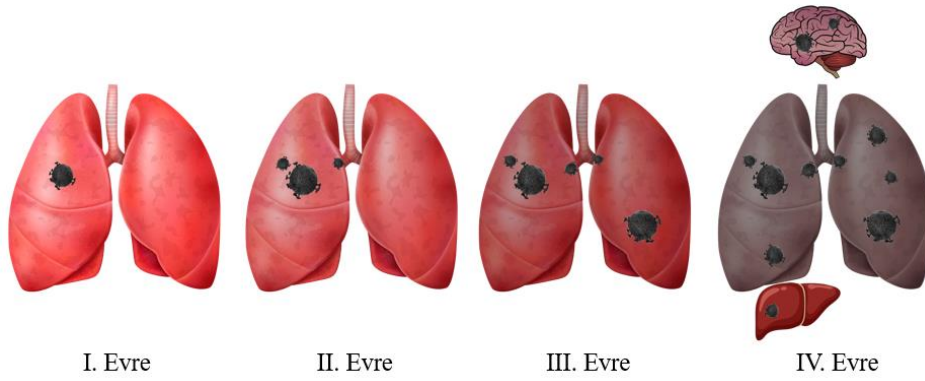
Tablo 2. TNM kriterlerine göre akciğer kanseri klinik evreleri

Evre I	IA1	T1a	N0	M0
	IA2	T1b	N0	M0
	IA3	T1c	N0	M0
	IB	T2a	N0	M0
Evre II	IIA	T2b	N0	M0
	IIB	T1a	N1	M0

Tablo 2'nin devamı

Evre III		T1b	N1	M0
		T1c	N1	M0
		T2a	N1	M0
		T2b	N1	M0
		T3	N0	M0
	IIIA	T1a	N2	M0
		T1b	N2	M0
		T1c	N2	M0
		T2a	N2	M0
		T2b	N2	M0
		T3	N1	M0
		T4	N0	M0
		T4	N1	M0
	IIIB	T1a	N3	M0
		T1b	N3	M0
		T1c	N3	M0
		T2a	N3	M0
		T2b	N3	M0
		T3	N2	M0
		T4	N2	M0
	IIIC	T3	N3	M0
		T4	N3	M0
Evre IV	IVA	Herhangi bir T	Herhangi bir N	M1a
		Herhangi bir T	Herhangi bir N	M1b
	IVB	Herhangi bir T	Herhangi bir N	M1c

TNM sistemine dayalı olarak yapılan sınıflama, akciğer kanserinin klinik seyri hakkında kapsamlı bilgi sunar. Bu sistemin sunduğu T, N ve M parametrelerinin çeşitli kombinasyonları, hastalığın farklı evrelerini tanımlayarak klinik karar süreçlerinde yol gösterici olur. Tanımlanan her evre, tümörün vücut içindeki yayılım derecesine göre hastalığın ciddiyetini ve tedavi stratejilerini belirlemede kritik öneme sahiptir. Şekil 7'de akciğer kanseri klinik evrelerinin görsel bir temsili verilmiştir.



Şekil 7. Akciğer kanseri klinik evreleri

- I. Evrede tümör akciğerin sınırlı bir bölgesindedir ve lenf düğümlerine yayılım göstermemiştir.
- II. Evrede tümör biraz daha büyümüş olup bazı bölgesel lenf düğümlerine ulaşmıştır.
- III. Evrede tümör her iki akciğere yayılmış olabilir ve göğüs ortasında yer alan lenf düğümleri etkilenmiştir.
- IV. Evre ise hastalığın en ileri düzeyini temsil etmekte olup kanser hücreleri uzak organlara metastaz yapmıştır.

2.2.7. Akciğer Kanseri Tedavi Yöntemleri

Tedavi yaklaşımları, hastalığın evresi, kanserin tipi ve hastanın genel sağlık durumu gibi faktörlere bağlı olarak değişkenlik göstermektedir. Akciğer kanseri tedavisi, cerrahi müdahale, radyoterapi, kemoterapi, hedefe yönelik tedavi ve immünoterapileri içeren çok disiplinli bir yaklaşımı gerektirir (Lemjabbar-Alaoui, Hassan, Yang, & Buchanan, 2015).

Her bir tedavi yöntemi, kanser hücrelerinin büyümesini kontrol altına almak, hastalığın yayılmasını durdurmak ve hastanın yaşam kalitesini artırmak amacıyla özel olarak tasarlanmıştır. Bu bölümde, akciğer kanserinin tedavisinde kullanılan başlıca yöntemlere değinilmiştir.

Cerrahi tedavi, akciğer kanserinin tedavisinde önemli bir seçenektir ve genellikle erken evrelerde tercih edilir. Bu yöntem, kanserli dokunun cerrahi olarak çıkarılmasını içerir ve kanserin tamamen ortadan kaldırılmasını hedefler. Cerrahi seçenekler arasında lobektomi (akciğer lobunun çıkarılması), pnömonektomi (bir akciğerin tamamının çıkarılması) ve segmentektomi (kanserli bölgenin çıkarılması) bulunur (Lang-Lazdunski, 2013).

Radyasyon tedavisi, yüksek enerjili ışınların kanser hücrelerini öldürmek için kullanıldığı bir yöntemdir. Genellikle cerrahiye uygun olmayan hastalar için tercih edilir veya cerrahi sonrası kalan kanser hücrelerini hedef alır. Ayrıca, radyasyon tedavisi metastatik hastalıklarda ve semptomları hafifletmek için de kullanılır (Timmerman, ve diğerleri, 2010).

Kemoterapi, kanser hücrelerini öldürmek veya büyümelerini durdurmak için kullanılan ilaç tedavisidir. Sistemik bir tedavi yöntemi olan kemoterapi, ilaçların vücuda genel olarak dağıtılması yoluyla tümörleri hedef alır. Kemoterapi genellikle ileri evre akciğer kanserlerinde, cerrahi öncesi tümörü küçültmek için veya cerrahi sonrası kalan kanser hücrelerini yok etmek için kullanılır (Ramalingam & Belani, 2008). Kemoterapi, diğer tedavi yöntemleriyle (cerrahi, radyoterapi, immünoterapiler) kombine edilebilir ve hastalığın yönetiminde önemli bir rol oynar. Yan etkileri, kullanılan ilaçlara bağlı olarak değişebilir ve kişiselleştirilmiş tedavi planları gerektirir.

Hedefe yönelik tedavi, kanser hücrelerinin belirli moleküler hedeflerini hedef alarak bu hücrelerin büyümesini ve yayılmasını engelleyen ilaçları içerir. Bu tedavi yöntemi, genellikle kanser hücrelerinde bulunan belirli genetik veya moleküler değişikliklere karşı etkilidir (Chan & Hughes, 2015). Hedefe yönelik tedaviler, genellikle akciğer kanserinin belirli alt türlerinde kullanılır ve kemoterapiye göre daha az yan etki içerir.

İmmünoterapi, vücudun bağışıklık sistemini kanser hücreleriyle savaşması için teşvik eden bir tedavi yöntemidir. Bu tedavi, bağışıklık sisteminin kanser hücrelerini tanımasını ve yok etmesini sağlar. İmmünoterapiler arasında kontrol noktası inhibitörleri, kanser aşıları ve adaptif hücre tedavileri bulunur (Steven, Fisher, & Robinson, 2016). İmmünoterapinin etkinliği, hastanın bağışıklık sisteminin yanıt verebilme yeteneğine ve tümörün özelliklerine bağlıdır.

Palyatif tedavi, akciğer kanseri tedavisinde özellikle ileri evre hastalarda yaşam kalitesini artırmak ve semptomları hafifletmek amacıyla uygulanan destekleyici bir tedavi yaklaşımıdır (Ferrell, Koczywas, Grannis, & Harrington, 2011). Bu tedavi yöntemi, hastalığı tamamen ortadan kaldırmayı hedeflemekten ziyade, hastanın ağrı, nefes darlığı, öksürük, iştahsızlık gibi kanserle ilişkili semptomlarını kontrol altına almayı amaçlar. Palyatif bakım; radyoterapi, kemoterapi, ağrı yönetimi, solunum desteği, beslenme desteği ve psikososyal destek gibi çok yönlü uygulamaları içerebilir. Palyatif bakımın erken dönemde entegre edilmesi, hastaların hem yaşam süresini hem de yaşam kalitesini artırabileceği gösterilmiştir

(Temel & diğ erleri, 2010). Ayrıca hasta ve ailelerine duygusal, sosyal ve manevi destek saėlayarak b t nc l bir yaklařım sunar.

Akciėer kanserinde tedavi  oėu zaman tek bir y ntemle sınırlı kalmaz; cerrahi, kemoterapi, radyoterapi, hedefe y nelik tedaviler ve imm noterapiler bir arada veya ardışık řekilde kullanılarak kombine tedavi yaklařımları uygulanır (Livingston, 1997) (Schild, ve diğ erleri, 2003). Bu multidisipliner yaklařım,  zellikle t m r n yayılım derecesine, histolojik tipine ve hastanın genel durumuna g re řekillendirilir.  rneėin, ameliyat  ncesi t m r  k   ltmek amacıyla uygulanan neoadjuvan kemoterapi veya radyoterapi, cerrahiye daha etkili hale getirebilir. Cerrahi sonrası (adjuvan) kemoterapi veya radyoterapi, mikroskobik kalıntıların ortadan kaldırılmasını hedefler. Ayrıca imm noterapi ve hedefe y nelik tedaviler, ileri evre hastalarda kemoterapiyle kombine edilerek daha g  l  yanıtlar elde edilebilir. Kombine tedavi stratejileri, tedavi etkinliėini artırmakla birlikte yan etki riskini de artırabileceėi i in kiřiselleřtirilmiř tedavi planlaması b y k  nem tařır. Her hasta i in en uygun kombinasyon, t m r n biyolojik  zellikleri ve hastanın yařam kalitesi g z  n nde bulundurularak belirlenmelidir.

2.3. Nefes ve U ucu Organik Bileřikler

Nefes alıřveriři, organizmanın hayati iřlevlerini s rd rebilmesi i in kritik bir fizyolojik s re tir. Nefes alma (inspirasyon) sırasında hava burun veya aėız yoluyla akciėerlere ulařır; alveollerde ger ekleřen gaz deėiřimiyle oksijen kana ge er ve dokulara tařınır. H cresel metabolizma sonucu oluřan karbondioksit ve diėer atık gazlar ise kan yoluyla akciėerlere tařınarak nefes verme (ekspirasyon) yoluyla v cuttan uzaklařtırılır. Bu d ng  yalnızca gaz alıřveriřini deėil, aynı zamanda v cudun asit-baz dengesinin korunmasını, enerji  retimini s rd r lmesini ve metabolik atıkların atılmasını da saėlar.

 nsan nefesi; oksijen, karbondioksit, azot ve su buharı gibi temel gazların yanı sıra,  ok sayıda VOC i ermektedir. VOC'ler, h cresel metabolizma sırasında oluřan ve nefes yoluyla dıř ortama salınabilen organik molek llerdir. Bu bileřikler arasında alkanlar, aldehytler, ketonlar, alkoller ve aromatik hidrokarbonlar yer almaktadır (Phillips, ve diğ erleri, 1999). VOC'ler; h cresel metabolizma, oksidatif stres, inflamasyon gibi biyolojik s re lerin bir sonucu olarak oluřabileceėi gibi, sigara dumanı, hava kirliliėi, beslenme ve ila  kullanımı gibi  evresel fakt rlere baėlı olarak da ortaya  ıkabilir (Blanchet, ve diğ erleri, 2017).

Akciğer kanseri gibi hastalıklarda, tümör hücrelerinin farklı metabolik aktiviteleri sonucu ortaya çıkan VOC profili, sağlıklı bireylerden anlamlı şekilde farklılık göstermektedir. Bu farklılık sayesinde VOC'ler, akciğer kanserinin erken tanısı ve hasta grupları arasında ayırım yapılması açısından potansiyel biyobelirteçler olarak değerlendirilmektedir.

VOC'lerin sadece hastalık varlığına değil, aynı zamanda hastalığın ilerleme derecesine, yani evresine göre de değişiklik gösterebileceği düşünülmektedir. Akciğer kanserinin farklı evrelerinde (Evre I–IV), tümörün büyüklüğü, oksijenlenme durumu, hücresel yoğunluğu ve metastatik yayılımı gibi faktörler tümör mikrosisteminde farklı metabolik yolları aktive eder. Bu da solunum havasında ölçülebilen VOC bileşenlerinin türünü ve yoğunluğunu etkileyebilir. Yapılan çalışmalar, bazı alkan, aldehit ve aromatik türevlerin ileri evre akciğer kanseri hastalarında daha yüksek seviyelerde bulunduğunu ortaya koymuştur.

Bu bağlamda, VOC profili analizi yalnızca akciğer kanserinin erken tanısı için değil, aynı zamanda evre tayini ve hastalık takibi açısından da umut vadeden bir yöntem olarak değerlendirilmektedir.

2.3.1. Akciğer Kanseri Özelinde Uçucu Organik Bileşikler

Araştırmalar, akciğer kanseri hastalarının nefesinde belirli VOC'lerin sağlıklı bireylere kıyasla daha yüksek konsantrasyonlarda bulunduğunu ortaya koymuştur. Tümörün metabolik aktivitesi, hücresel düzeyde oksidatif stresin artmasına ve inflamatuvar yanıtların tetiklenmesine yol açmakta; bu da bazı uçucu organik bileşiklerin yoğunluklarında anlamlı bir artışa neden olmaktadır. Ayrıca, VOC düzeyleri sadece hastalık varlığını değil, aynı zamanda hastalığın ilerleme düzeyini de yansıtabilir. İleri evrelerde VOC konsantrasyonlarının arttığına dair bulgular, bu bileşiklerin tanıya ek olarak evrelemede de kullanılabileceğini göstermektedir.

Özellikle bazı alkanlar, aldehitler, ketonlar ve aromatik hidrokarbonlar, akciğer kanseri ile ilişkili olarak tekrarlayan şekilde rapor edilen bileşikler arasında yer almaktadır. Literatürde akciğer kanseri hastalarının nefesinde tanımlanan yaygın VOC grupları ve bunların biyolojik kökenleri Tablo 3'te sunulmaktadır.

Tablo 3. Akciğer kanserine ilişkin VOC grupları, örnekleri ve olası biyolojik kaynakları

VOC Grubu	Örnek Bileşenler	Olası Neden
Alkanlar	Pentan, Etan, Heptan, Heksan, Metilpentan, Trimetilheptan	Lipid peroksidasyonu, hücre zarı bozulması
Alkenler	1-Buten, İzopren	Oksidatif stres, hücresel yıkım
Aldehitler	Formaldehit, Asetaldehit, Heksanal, Nonanal	Tümör metabolizması, inflamasyon
Alkoller	Etanol, İzopropanol	Fermantasyon süreçleri, mikrobiyal flora etkisi
Ketonlar	2-Butanon, Aseton	Anormal hücresel metabolizma
Aromatik hidrokarbonlar	Benzen, Toluen, Stiren	Sigara dumanı, çevresel maruziyet
Terpenler	Limonen, Metilsiklopentan	Diyet, çevresel kaynaklar

2.3.2. Nefes Analizi Teknikleri

Nefeste yer alan uçucu VOC'lerin tanımlanması ve miktarlarının belirlenmesi, akciğer kanseri gibi hastalıkların tanı ve takibinde önemli bir adımdır. Bu bileşiklerin doğru ve güvenilir bir şekilde analiz edilebilmesi için gelişmiş analitik tekniklere ihtiyaç duyulmaktadır. Nefes örneklerinde VOC'leri tespit etmek amacıyla kullanılan yöntemler, genellikle gazların ayrıştırılması, iyonlaştırılması ve bileşenlerinin kimyasal özelliklerine göre tanımlanması prensibine dayanır. En yaygın olarak kullanılan nefes analizi teknikleri aşağıda sıralanmıştır.

Gaz Kromatografisi, gaz karışımlarının bileşenlerine ayrılması için kullanılan yaygın bir tekniktir. Bu teknik, gazların ayırıcı bir kolondan geçerken her bir bileşenin farklı hızlarda hareket etmesi ve dedektöre farklı zamanlarda ulaşması sayesinde gaz bileşenlerinin ayrılmasını sağlar (Littlewood, 2013).

Kütle Spektrometrisi, iyonlaştırılmış moleküllerin kütle/yük oranlarına göre ayrılmasını temel alan bir tekniktir. İyonlaştırılmış moleküller bir manyetik alandan geçerken daha ağır moleküller daha yavaş hareket ederler ve daha az saparlar, daha hafif moleküller ise daha hızlı hareket ederler ve daha çok saparlar (De Hoffmann & Stroobant, 2007). İyonlar, kütle/yük oranlarına göre dedektörde sıralanır. Dedektör, her bir iyonun

ulaşma zamanını veya sinyalinin kaydeder, böylece hangi gaz bileşeninin bulunduğu belirlenir.

Gaz Kromatografisi-Kütle Spektrometrisi, gaz karışımlarındaki bileşenleri ayırtmak ve analiz etmek için kullanılan bir tekniktir. İlk olarak, gaz karışımı bir ayırıcı kolondan geçirilir; burada her bileşen farklı hızlarda hareket eder ve dedektöre farklı zamanlarda ulaşır. Bu ayırma işlemi, bileşenlerin karışımdaki miktarını belirler. Ayrılan bileşenler daha sonra kütle spektrometrisine yönlendirilir, burada iyonlaştırılır ve kütle/yük (m/z) oranlarına göre ayrılır (Karasek & Clement, 2012). Dedektör, her bileşimin kütle/yük oranını belirleyerek moleküllerin kimyasal yapısını tanımlar. Bu yöntem gaz karışımlarındaki bileşenleri hem ayırma hem de detaylı analiz için çok etkilidir.

Proton Transfer Reaksiyon Kütle Spektrometrisi, protonların gaz moleküllerine aktarılmasıyla iyonlaşma sağlayan bir tekniktir. Hidronyum iyonları, gaz örneğiyle reaksiyona girerek proton transferi yoluyla molekülleri iyonlaştırır. İyonlaştırılan moleküller, kütle spektrometrisine yönlendirilir. Kütle spektrometresinde, farklı kütle/yük oranlarına sahip iyonlar ayrılır ve dedektörde sıralanır. Bu süreç sayesinde her bir molekülün kütle/yük oranı belirlenir ve gaz bileşenleri tanımlanır (Ellis & Mayhew, 2013).

Seçilmiş İyon Akış Tüpü Kütle Spektrometrisi, nefes gazı analizinde hızlı ve hassas bir şekilde tespit edilmesinde kullanılan bir kütle spektrometrisi tekniğidir (Smith & Španěl, 2005). Hidronyum gibi seçilmiş iyonlar, gaz örneğiyle reaksiyona girerek molekülleri iyonlaştırır. İyonlaştırılan moleküller, kütle spektrometrisine yönlendirilir. Kütle spektrometresinde, farklı kütle/yük oranlarına sahip iyonlar ayrılır ve dedektörde sıralanır. Bu işlem, her bir molekülün kütle/yük oranını belirler ve gaz bileşenlerini tanımlar.

İyon Hareketliliği Spektrometrisi, gaz fazındaki iyonların elektrik alanında hareket etmeleri sırasında hızlarının ölçülmesiyle çalışan bir analiz yöntemidir (Borsdorf & G.A, 2006). Gaz örneği iyonlaştırılır ve iyonlar bir elektrik alanı içinde hareket eder. İyonlar, kütlelerine, şekillerine ve yüklerine bağlı olarak farklı hızlarda hareket eder. Daha hafif ve daha küçük iyonlar daha hızlı hareket ederken, daha ağır ve daha büyük iyonlar daha yavaş hareket eder. Bu farklı hareketlilik, iyonların dedektörde sıralanmasını sağlar. Sonuç olarak, her iyonun hareketlilik özellikleri belirlenir ve gaz bileşenleri ayrılır.

2.3.3. Nefes Analizi Tekniklerinin Sınırlamaları ve Elektronik Burun Yaklaşımı

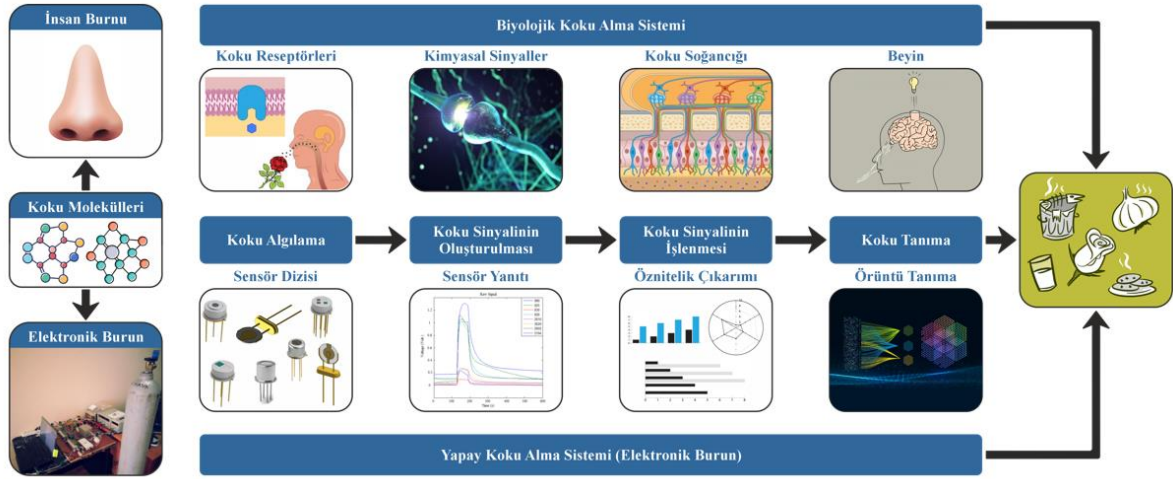
Nefes analizi teknikleri, insan sağlığına ilişkin önemli biyobelirteçlerin tespit edilmesini mümkün kılmaktadır. Ancak her yöntemin kendine özgü avantajları ve sınırlamaları bulunmaktadır. Gaz Kromatografisi-Kütle spektrometrisi gibi yöntemler, kimyasal çözünürlük açısından güçlüdür; ancak genellikle pahalı, taşınması zor ve uzmanlık gerektiren cihazlardır. Bu kombinasyon detaylı analiz sağlasa da klinik pratikte hızlı kullanım için uygun değildir. Diğer yöntemler ise taşınabilirlik ve hız avantajı sunmalarına karşın, sınırlı moleküler tanıma kapasitesi veya donanım maliyetleri gibi kısıtlarla karşı karşıyadır. Bu sınırlamalar, daha kolay uygulanabilir, hızlı ve kullanıcı dostu sistemlerin geliştirilmesi ihtiyacını doğurmuştur. Bu çerçevede elektronik burun sistemleri, hem VOC'leri tanıma hem de sınıflandırma potansiyeli taşıyan alternatif bir yaklaşım olarak öne çıkmaktadır.

Bu nedenlerle, daha hızlı, taşınabilir, düşük maliyetli ve kolay uygulanabilir alternatif sistemlerin geliştirilmesi gerekliliği doğmuştur. Bu ihtiyaca yanıt olarak geliştirilen elektronik burun sistemleri, uçucu organik bileşenleri doğrudan algılayan sensör teknolojileri ve örüntü tanıma algoritmalarıyla, nefes analizine dayalı tanı süreçlerinde umut vadeden bir yaklaşım sunmaktadır.

2.4. Elektronik Burun

Elektronik burun sistemleri, insan ve hayvanların koku alma mekanizmasından ilham alarak geliştirilmiş, karmaşık gaz karışımlarını tanıyabilen sensör tabanlı analiz cihazlarıdır (Karakaya, Ulucan, & Turkan, 2020). Bu sistemlerin temel amacı, çeşitli kimyasal bileşenleri algılamak, elde edilen sinyalleri işlemek ve bu verileri kullanarak örüntü tanıma yoluyla kokuları veya gazları sınıflandırmaktır. İnsan veya hayvan burnunda, koku molekülleri burun mukozasında bulunan koku reseptörlerine bağlanır, burada kimyasal etkileşim sonucunda sinirsel uyarılar oluşur. Bu uyarılar, sinir sistemi aracılığıyla beynin koku merkezine iletilir ve burada anlamlı bir koku algısına dönüştürülür. Elektronik burunda da gaz molekülleri sensör yüzeyleriyle etkileşime girer, bu etkileşimden sinyaller üretilir ve bu sinyaller yazılımsal sistemler aracılığıyla yorumlanır. Kısacası, elektronik burun sistemleri; doğal koku alma sisteminin sinyal algılama, dönüştürme ve yorumlama

aşamalarını teknolojik olarak yeniden üretir. Biyolojik koku alma sistemi ile yapay koku alma sisteminin karşılaştırmalı olarak gösterimi Şekil 8’de verilmiştir.



Şekil 8. Biyolojik ve yapay koku alma sistemlerinin karşılaştırılması

Elektronik burunların çalışma prensibi, gazın algılanmasından sınıflandırma sürecine kadar uzanan bir dizi işlem basamağından oluşur. İlk adımda, analiz edilmek üzere hedef gaz sisteme alınır. Ardından bu gaz örneği, sensör dizisiyle etkileşime girer. Her bir sensör, belirli gaz moleküllerine karşı farklı duyarlılığa sahiptir ve bu etkileşim sonucunda direnç değişimi veya frekans kayması gibi fiziksel ya da kimyasal değişimler meydana gelir. Sensörlerde oluşan bu değişimler, elektriksel sinyallere dönüştürülerek gerilim, akım veya frekans değişimleri şeklinde sistem tarafından algılanır. Üretilen bu sinyaller, dijital formata çevrilip çeşitli ön işleme adımlarından geçirilerek analiz için uygun hale getirilir. Son aşamada ise, işlenmiş veriler örüntü tanıma algoritmaları aracılığıyla değerlendirilir ve cihaz, tanımlanan gaz ya da bileşiğin varlığına ilişkin bir sınıflandırma sonucu üretir.

Elektronik burun sistemleri, çok sayıda uygulama alanında geleneksel analiz yöntemlerine alternatif olarak geliştirilmekte ve çeşitli araştırmalarda etkin biçimde test edilmektedir. Elektronik burunlar yiyecek ve içecek endüstrisinde, tarım ve ormancılıkta, tıp ve sağlıkta, iç ve dış mekân izleme uygulamalarında, askeri ve sivil güvenlikte, ürün paketleme, kozmetik ve daha birçok alanda yaygın bir cihaz haline gelmiştir (Karakaya, Ulucan, & Turkan, 2020).

Bir elektronik burun sistemi, gazları algılayıp tanıyabilme işlevini yerine getirebilmek için donanım ve yazılım temelli birden fazla bileşenin bir arada çalıştığı bütünleşik bir

yapıya sahiptir. Bu bileşenler, birbirini tamamlayan görevler üstlenir ve sistemin genel performansını doğrudan etkiler. Temel olarak bir elektronik burun; algılayıcı birim, dönüştürücü birim, elektronik birim ve örüntü tanıma birimi olmak üzere dört ana bileşenden oluşur. Aşağıda bu ana birimler detaylı olarak ele alınmaktadır.

2.4.1. Algılayıcı Birim

Elektronik burnun algılayıcı birimi, gazları algılamak için kullanılan sensörlerden oluşur. Kullanılan sensörler, gaz moleküllerinin fiziksel veya kimyasal özelliklerinde oluşturduğu değişimleri algılama prensibine dayanır. Bu sensörler, algılama mekanizmalarına göre sınıflandırıldığında, literatürde genel kabul gören beş temel grupta toplanmaktadır: rezistif sensörler, piezoelektrik sensörler, elektrokimyasal sensörler, kapasitif sensörler ve optik sensörler. Bu sınıflandırma, sensörlerin gazlarla etkileşim sonucu ortaya çıkan elektriksel, mekanik veya optik değişimleri nasıl ölçtüklerine göre yapılmaktadır. Aşağıda, elektronik burun teknolojilerinde yaygın olarak kullanılan bu beş ana sensör grubu sırasıyla tanıtılmakta ve her biri örneklerle açıklanmaktadır.

- 1) Rezistif sensörler, gaz algılama işlemini hedef gaz moleküllerinin sensör malzemesi ile etkileşimi sonucunda oluşan elektriksel direnç değişimlerini ölçerek gerçekleştirir. Bu tür sensörler, temel olarak iletken veya yarı iletken bir aktif malzemeden oluşur ve bu malzeme gaz ortamıyla temas ettiğinde yüzeyde fiziksel veya kimyasal reaksiyonlar meydana gelir. Bu reaksiyonlar, sensör malzemesinin iletkenliğinde (ya da tersinden, direncinde) ölçülebilir değişimlere neden olur. Rezistif sensörlerin başlıca avantajları, yapısal olarak basit olmaları, miniaturize edilebilmeleri ve düşük maliyetle üretilebilmeleridir. Ancak, nem, sıcaklık ve diğer çevresel koşullardan etkilenme gibi dezavantajları da vardır. Bu nedenle, elektronik burun sistemlerinde bu sensörlerin yanıtları genellikle örüntü tanıma algoritmalarıyla birlikte yorumlanır. Elektronik burun sistemlerinde kullanılan rezistif sensörler, genellikle MOS sensörler ve CP sensörler olmak üzere iki gruba ayrılır.

- MOS sensörleri, metal oksit malzemelerin elektriksel özelliklerindeki değişim prensibine dayanır. Bu sensörler, bir metal oksit tabakası ve bu tabakayı ısıtan bir ısıtıcı bobininden oluşur. Metal oksit tabakası genellikle titanyum dioksit (TiO_2), çinko oksit (ZnO) veya kalay oksit (SnO_2) gibi

materyallerden yapılır. Sensör, metal oksidin 200°C ila 500°C arasında ısıtılmasıyla çalışır; bu ısıtma, metal oksidin gazlarla kimyasal reaksiyonlara girmesini sağlar. Oksijenli gazlar metal oksit yüzeyinde oksidasyona neden olarak serbest elektronları bağlar ve elektriksel direnci artırırken iletkenliği azaltır. Oksijen içermeyen gazlar ise oksijeni yüzeyden uzaklaştırarak elektriksel direnci azaltırken iletkenliği artırır. Elektriksel iletkenlikteki bu değişiklikler, sensörün çıkış sinyali olarak ölçülür ve bu sinyal gazın konsantrasyonu hakkında bilgi sağlar.

- CP sensörlerde temel prensip, polimerlerin elektriksel özelliklerindeki değişikliklerdir. İletken polimerler, elektrik akımını iletebilen ve koku algılamasında kullanılan özel polimerlerdir. Koku molekülleri, polimer yüzeyine yapıştığında veya etkileşimde bulunduğunda, polimerin direnci değişir ve bu değişiklik bir voltmetre veya ohmmetre kullanılarak ölçülür. Elde edilen direnç değişimi verileri analiz edilir ve koku profilleri oluşturulur. Bu profiller, koku moleküllerinin türünü ve konsantrasyonunu belirlemek için kullanılır.

2) Piezoelektrik sensörler, gaz algılama işlemini gaz moleküllerinin sensör yüzeyinde oluşturduğu kütle veya mekanik yük değişikliklerinin, piezoelektrik malzeme üzerindeki mekanik titreşim frekansını etkilemesi prensibine dayanarak gerçekleştirir. Bu sensörlerde, gaz molekülleri sensör yüzeyine bağlandığında, sistemin toplam kütlesi artar ve bu kütle değişimi sensörün rezonans frekansında kaymaya neden olur. Frekanstaki bu değişim ölçülerek gazın türü ve konsantrasyonu belirlenebilir. Piezoelektrik sensörler, yüksek hassasiyet, frekans kararlılığı ve düşük tespit limitleri sayesinde elektronik burun sistemlerinde özellikle düşük konsantrasyonlu gazların algılanmasında yaygın olarak kullanılmaktadır. Ancak genellikle sıcaklık ve nem gibi çevresel faktörlere karşı duyarlıdırlar ve doğru çalışmaları için iyi bir kalibrasyon gerektirir. Elektronik burun sistemlerinde kullanılan piezoelektrik sensörler başlıca QCM sensörler, SAW sensörleri ve Hacim Akustik Dalga (BAW) sensörleridir.

- QCM sensörler, piezoelektrik özelliklere sahip kuvars kristallerinden yapılır ve bu kristaller belirli bir frekansta titreşir. Bu titreşim, kristalin doğal rezonans frekansına bağlıdır ve bu frekans oldukça stabil bir değerdir. Koku molekülleri, bu kuvars kristaller ile etkileşime girdiğinde,

yüzeyde birikerek veya kimyasal reaksiyonlar oluşturarak ek bir kütle ekler. Kütledeki değişiklik, kristalin titreşim frekansında bir kayma yaratır ve bu kayma, moleküllerin miktarına ve türüne bağlı olarak ölçülür. Elde edilen frekans verileri analiz edilir ve koku profilleri oluşturulur. Bu profiller, koku moleküllerinin türünü ve konsantrasyonunu belirlemek için kullanılır.

- SAW sensörleri, piezoelektrik özellik gösteren kuvars veya lityum niobat gibi malzemelerden üretilir. Bu sensörlerin temel prensibi, yüzeylerinde oluşturulan akustik dalgaların davranışlarını ölçmeye dayanır. Gaz molekülleri yüzeye tutunduğunda, sensör yüzeyindeki kütle artar. Bu kütle artışı, yüzeyde yayılan akustik dalgaların hızında ve genliğinde değişiklik meydana getirir. Bu değişiklik, sensör çıkışında frekans kayması şeklinde ölçülür. Elde edilen frekans verileri, analiz algoritmaları aracılığıyla yorumlanarak gazın türü ve konsantrasyonu belirlenir.
- BAW sensörleri, hacim içinde yayılan akustik dalgaları kullanarak gaz algılayan piezoelektrik tabanlı cihazlardır. SAW sensörlerinden farklı olarak, burada oluşturulan mekanik titreşimler sensörün yalnızca yüzeyinde değil, kristal yapının iç hacminde yayılır. Gaz molekülleri yüzeye bağlandığında, malzemenin kütle ve mekanik özelliklerinde değişiklik meydana gelir. Bu değişim, hacim içinde yayılan akustik dalgaların hızını ve frekansını etkiler. Elde edilen frekans değişimleri, sensör çıkışında ölçülerek gazın türü ve konsantrasyonu hakkında bilgi verir. BAW sensörleri, SAW sensörlerine kıyasla daha yüksek frekanslarda çalışabilir ve bu sayede daha yüksek hassasiyet sağlar.

3) Elektrokimyasal sensörler, gaz moleküllerinin elektrot yüzeyinde oluşturduğu kimyasal reaksiyonlar sonucu meydana gelen elektriksel sinyalleri ölçerek çalışan sensörlerdir. Bu sensörler, özellikle karbon monoksit (CO), hidrojen sülfür (H₂S), azot oksitler (NO_x), amonyak (NH₃) ve klor (Cl₂) gibi toksik ve reaktif gazların algılanmasında tercih edilir; çünkü düşük konsantrasyonlarda dahi yüksek hassasiyetle çalışabilirler. Elektrokimyasal sensörlerde gaz, redoks reaksiyonuna girerek akım (amperometrik) veya potansiyel (potansiyometrik) değişimine neden olur. Bu elektriksel değişim, gazın türü ve yoğunluğuna göre ölçülerek analiz edilir. Elektrokimyasal sensörlerin amperometrik, potansiyometrik, galvanik

hücreli gibi alt türleri mevcuttur. Düşük güç tüketimi, kompakt yapı ve yüksek seçicilik gibi avantajlara sahip olan elektrokimyasal sensörler; taşınabilir elektronik burun sistemleri için de uygundur. Bununla birlikte, sınırlı ömür, sıcaklık ve nem gibi çevresel koşullara duyarlılık, bu sensörlerin tasarımı ve kullanımını sırasında dikkate alınması gereken önemli faktörlerdir.

- 4) Kapasitif sensörler, gaz moleküllerinin bir sensör yapısındaki dielektrik özellikleri değiştirmesi prensibine dayalı olarak çalışan algılayıcılardır. Tipik olarak, bu sensörler iki elektrot arasında yer alan bir dielektrik tabakadan oluşur. Bu yapı, tıpkı bir kondansatör gibi davranır ve iki elektrot arasındaki kapasitans (C) oluşur. Gaz molekülleri sensörün dielektrik ortamına girdiğinde, ortamın dielektrik sabiti (ϵ) değişir. Bu değişiklik, doğrudan kapasitans değerinde bir farklılık yaratır. Kapasitans ölçümündeki bu değişim, gazın türüne ve konsantrasyonuna göre farklılık gösterir ve elektronik devreler tarafından ölçülerek sayısal veriye dönüştürülür. Kapasitif sensörler, düşük güç tüketimi, basit yapı ve hızlı yanıt süresi gibi avantajlara sahiptir. Ayrıca, çalışma sıcaklıkları genellikle düşüktür ve oda sıcaklığında rahatlıkla çalışabilirler. Bununla birlikte, nem ve sıcaklık gibi çevresel faktörlere karşı duyarlılık gösterebilirler; bu nedenle hassas uygulamalarda çevresel telafi mekanizmaları gerektirebilir.
- 5) Optik sensörler, gaz moleküllerinin ışıqla etkileşimi sonucu oluşan optik değişimleri algılayarak çalışan sensörlerdir. Bu etkileşim; ışığın soğurulması (absorpsiyon), kırılması (refraksiyon), yansıtılması (refleksiyon) veya yayılması (emisyon, floresans) gibi optik olaylar aracılığıyla gerçekleşir. Bu sensörler, gazların varlığını ve konsantrasyonunu, bu optik özelliklerde meydana gelen değişimleri ölçerek belirler. Optik sensörlerin başlıca avantajları arasında çok yüksek hassasiyet, doğrudan temas gerektirmeyen ölçüm, hızlı yanıt süresi ve kimyasal seçicilik yer alır. Bu sensörler, özellikle çok düşük konsantrasyonlardaki gazların tespitinde etkilidir. Bununla birlikte, yüksek donanım maliyeti, karmaşık yapı ve taşınabilir sistemlere entegrasyon zorlukları, optik sensörlerin bazı sınırlayıcı yönleridir. Elektronik burun sistemlerinde optik sensörler genellikle laboratuvar tipi sabit cihazlarda veya ileri düzey tıbbi tanı sistemlerinde kullanılır.

Farklı algılama prensiplerine dayanan gaz sensörleri, uygulama alanlarına ve performans gereksinimlerine göre çeşitli avantaj ve dezavantajlara sahiptir. Tablo 4'te

başlıca gaz sensör grupları, alt türleri, çalışma prensipleri ve karşılaştırmalı özellikleri özetlenmiştir.

Tablo 4. Gaz sensörlerinin sınıflandırılması ve karşılaştırılması

Sensör Türü	Algılama Prensibi	Avantajlar	Dezavantajlar
Rezistif sensörler (MOS, CP)	Direnç değişimi	Basit yapı, düşük maliyet	Nem ve sıcaklığa duyarlı, düşük seçicilik
Piezoelektrik sensörler (QCM, SAW, BAW)	Frekans değişimi	Yüksek hassasiyet,	Çevresel duyarlılık, kalibrasyon ihtiyacı
Elektrokimyasal sensörler (Amperometrik, Potansiyometrik, Galvanik vb.)	Redoks tepkimesi sonucu akım/potansiyel oluşumu	Yüksek seçicilik, düşük güç tüketimi	Sınırlı ömür, nem ve sıcaklığa duyarlı
Kapasitif sensörler	Kapasitans değişimi	Düşük güç tüketimi, hızlı yanıt	Nem/sıcaklık etkisi, stabilite gerektirir
Optik sensörler	Işıkla etkileşim	Çok yüksek hassasiyet, temassız algılama	Yüksek maliyet, karmaşık yapı

2.4.2. Dönüştürücü Birim

Elektronik burun sistemlerinde dönüştürücü birim, algılayıcı birimden gelen fiziksel veya kimyasal sinyalleri elektriksel sinyallere çevirir. Gaz molekülleriyle etkileşime giren sensörler, genellikle direnç, frekans, kapasitans veya akım gibi parametrelerde değişim oluşturur. Algılayıcı birim tarafından üretilen sinyaller genellikle analog formdadır. analog-dijital dönüştürücüler kullanarak, algılayıcı birimden gelen analog sinyalleri dijital verilere dönüştürür. Dönüştürücü birim analog-dijital dönüştürücülerün yanı sıra gerilim-yükselticiler ve filtre devreleri gibi elektronik bileşenlerden oluşur. Gerilim yükselticiler, sinyallerin genliğini artırarak zayıf sinyallerin daha belirgin hale gelmesini sağlar, bu da

sinyallerin daha doğru bir şekilde analiz edilmesini mümkün kılar. Filtreler, sinyallerdeki istenmeyen gürültüleri temizler ve belirli frekans aralıklarını seçer, böylece yalnızca anlamlı veriler işlenir.

Bu birim sayesinde sensör dizisinden elde edilen veriler, sayısal analiz için uygun formata getirilir. Dönüştürücü birimin doğru tasarlanması, sistemin genel hassasiyetini ve doğruluğunu doğrudan etkiler. Gürültüden arındırılmış, kararlı ve güvenilir sinyal üretimi, özellikle düşük konsantrasyonlu gazların algılanmasında kritik öneme sahiptir.

2.4.3. Elektronik Birim

Elektronik burun sistemlerinde elektronik birim, sistemin temel işlevselliğini sağlayan kritik bir bileşendir. Bu birimin başlıca görevi, sensörlerden gelen elektriksel sinyalleri almak, işlemek, düzenlemek ve örüntü tanıma birimine iletmektir. Böylece sensör verileri güvenilir, anlamlı ve analiz edilebilir bir formata dönüştürülür.

Elektronik birim; dönüştürücü birimden sonra gelen ve sinyal işleme sürecini üstlenen devreler, mikrodenetleyiciler, veri toplama kartları, sinyal işleyiciler ve iletişim arabirimlerinden oluşur. Bu yapı sayesinde sistem, sensörlerden gelen verileri toplar, gerekli ön işlemleri (örneğin filtreleme veya amplifikasyon) uygular ve verileri örüntü tanıma birimine aktarır. Ayrıca elektronik birim, haberleşme protokollerini yönetir, sistem bileşenlerinin koordinasyonunu sağlar ve enerji dağıtımını düzenler.

Doğru tasarlanmış bir elektronik birim, sistemin genel doğruluğu ve kararlılığı üzerinde doğrudan belirleyici bir etkiye sahiptir. Özellikle düşük konsantrasyonlu gazların tespitinde, gürültüsüz ve yüksek kaliteli sinyallerin sağlanması açısından kritik bir rol oynar.

2.4.4. Örüntü Tanıma Birimi

Örüntü tanıma birimi, elektronik burun sistemlerinde algılanan gaz karışımlarının sınıflandırılması ve tanımlanması görevini üstlenen yapay zekâ ve veri analizi temelli birimdir. Bu birim, sensör dizisinden gelen ve dönüştürücü-elektronik birim aracılığıyla sayısallaştırılmış verileri alarak, bu veriler üzerinde istatistiksel ve hesaplamalı analizler gerçekleştirir. Amaç, farklı gaz bileşenlerinin oluşturduğu sinyal desenlerini öğrenmek, ayırt etmek ve sınıflandırmaktır.

Elektronik burun sistemlerinde örüntü tanıma süreci genellikle dört temel adımdan oluşur. İlk adım olan veri ön işleme, sensör verilerinin analiz öncesinde uygun hale

getirilmesini amaçlar. Bu aşamada gürültü azaltma, normalizasyon ve boyut indirgeme gibi işlemler gerçekleştirilir. İkinci aşama olan öznitelik çıkarımı ise ham veriden istatistiksel veya fiziksel olarak anlamlı bilgilerin elde edilmesini sağlar. Üçüncü adım olan modelleme ve sınıflandırma sürecinde, çıkarılan öznitelikler kullanılarak verinin belirli bir sınıfa ait olup olmadığı makine öğrenmesi algoritmalarıyla belirlenir. Son olarak, doğrulama ve değerlendirme aşamasında modelin başarımı çeşitli performans metrikleriyle ölçülür ve modelin güvenilirliği test edilir.

Bu birim, geleneksel sınıflandırma algoritmalarının yanı sıra, modern elektronik burun sistemlerinde makine öğrenmesi ve hatta derin öğrenme tabanlı modellerle de desteklenebilmektedir. Bu sayede, karmaşık gaz karışımlarının ayrıştırılması ve çok boyutlu sensör verilerinin analizi daha doğru ve otomatik hale gelir. Sonuç olarak, örüntü tanıma birimi, elektronik burnun “yorumlayan zekâsı” olarak görev yapar ve sistemin tanı, teşhis veya karar verme yeteneğini doğrudan belirleyen en kritik yazılımsal bileşendir.

2.5. Veri Önileme Yöntemleri

Makine öğrenmesi uygulamalarında, sensörlerden elde edilen ham verilerin doğrudan analiz edilmesi, genellikle güvenilir sonuçlar üretmez. Özellikle çoklu sensör kullanılan sistemlerde, her sensörün yapısal ve çevresel farklılıklarından kaynaklı olarak sinyal karakteristiklerinde sapmalar, gürültüler ve karşılaştırma zorlukları ortaya çıkabilir. Bu nedenle, analiz öncesinde uygulanan önileme adımları, veri kalitesinin artırılması ve anlamlı özniteliklerin çıkarılabilmesi açısından kritik bir rol oynar. Bu kapsamda, veri önileme süreci; referans düzeltme, fiziksel dönüşümler (örneğin iletkenlik dönüşümü) ve sonrasında gerçekleştirilen normalizasyon gibi adımları içerebilir.

2.5.1. Referans Düzeltme İşlemi

Sensör sistemlerinden elde edilen sinyaller, yalnızca ölçülen bileşenlerin etkisiyle değil; aynı zamanda ortam sıcaklığı, nem, basınç, sensör yaşı ve kalibrasyon farklılıkları gibi çeşitli iç ve dış faktörlerden etkilenebilir. Bu durum, sensörlerin pasif halde (yani ölçüm öncesi) verdiği temel sinyal değerlerinde dahi tutarsızlıklara yol açar. Özellikle farklı sensör türlerinin bir arada kullanıldığı hibrit sistemlerde, bu farklar daha belirgin hâle gelir ve doğrudan karşılaştırmayı zorlaştırır.

Referans düzeltme, her sensör için ölçüm öncesinde elde edilen sabit bir referans değer (örneğin ilk birkaç saniyelik ortalama değer) ilgili sinyalin tamamından çıkarılması işlemidir. Bu sayede, her bir sensörün yalnızca ölçüme özgü tepkisi ön plana çıkarılır (Pearce, Schiffmann, Nagle, & Gardner, 2003).

Literatürde referans düzeltme yöntemleri genellikle fark, bağıl ve oransal olmak üzere üç temel grupta ele alınmaktadır. Bu yöntemler, her sensörün yanıtından belirli bir referans değer çıkarılması veya bu değere oranlanması yoluyla uygulanır. Fark referans düzeltme,

$$v_{n,s,r}(t) = v_{n,s}(t) - v_{n,s}(0) \quad (1)$$

biçiminde tanımlanır. Bağıl referans düzeltme,

$$v_{n,s,r}(t) = \frac{v_{n,s}(t)}{v_{n,s}(0)} \quad (2)$$

biçiminde tanımlanır. Oransal referans düzeltme,

$$v_{n,s,r}(t) = \frac{v_{n,s}(t) - v_{n,s}(0)}{v_{n,s}(0)} \quad (3)$$

biçiminde tanımlanır. Burada $v_{n,s}(t)$ ve n-inci örnek ve s-inci sensör için t anındaki gerilim değerini, $v_{n,s}(0)$ aynı sensörlerin t=0 anındaki referans değerlerini, $v_{n,s,r}(t)$ ise referans düzeltilmiş sinyalleri ifade eder. Bu yöntemler arasından hangisinin tercih edileceği, veri setinin yapısına ve kullanılacak analiz yöntemine bağlıdır. Ancak yapılan çalışmalarda, fark referans düzeltme yönteminin özellikle çoklu sensörlü ve kokuya dayalı uygulamalarda daha başarılı sonuçlar verdiği belirtilmektedir. Bu nedenle birçok uygulamada öncelikli tercih olarak kullanılmaktadır.

2.5.2. İletkenlik Dönüşümü

Referans düzeltme uygulandıktan sonra, özellikle MOS tabanlı gaz sensörlerinden elde edilen direnç verilerinin, kimyasal bağlamda daha anlamlı bir forma dönüştürülmesi gereklidir. Bu bağlamda, sensörlerin direnç bazlı çıkışları, çoğu analizde doğrudan

kullanılmak yerine, elektriksel iletkenlik formuna çevrilerek daha doğru yorumlanabilir hâle getirilir (Gutierrez-Osuna, 2002).

Bu dönüşüm, kimyasal bileşenler ile iletkenlik arasındaki ilişkinin daha tutarlı olmasını sağlar ve bazı makine öğrenmesi algoritmalarında daha dengeli veri dağılımları elde edilmesine olanak tanır. Özellikle düşük konsantrasyonlu gazların varlığı durumunda sensör tepkilerini daha hassas hâle getirebilir.

2.5.3. Normalizasyon

Veri setindeki değişkenler birbirinden farklı büyüklüklerde ölçüldüğünden, modelin öğrenme sürecinde bazı değişkenlerin diğerlerine göre daha baskın hale gelmesi söz konusu olabilir. Gerçek dünya verilerinde farklı değişkenler farklı ölçeklerde olabilir. Örneğin bir değişken 0 ile 1 arasında değişirken, bir diğeri 0 ile 1000 arasında olabilir. Bu durum özellikle mesafe tabanlı algoritmalar (örneğin k-NN ve SVM) ve gradyan temelli yöntemlerde modelin öğrenmesini etkiler. Özelliklerin farklı ölçü birimlerinde olması, modelin öğrenme sürecini olumsuz etkileyebilir. Normlizasyon işlemi için en yaygın kullanılan iki yöntem şunlardır:

- a) Minimum-Maksimum Normalizasyonu: Veri setindeki sayısal değişkenleri belirli bir aralığa (genellikle 0 ile 1) ölçeklendirme yöntemidir. Yorumlaması kolaydır ancak aykırı değerlere karşı hassastır çünkü uç değerler tüm ölçeklemeyi bozabilir. Minimum-Maksimum normalizasyonu,

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

biçiminde hesaplanır. Burada X orijinal değeri, X_{min} değişkenin minimum değerini, X_{max} değişkenin maksimum değeri X' normalize edilmiş değeri temsil etmektedir.

- b) Z-Skor Normalizasyonu: Veri setindeki sayısal değişkenlerin ortalamasını sıfır, standart sapmasını bir hale getirerek ölçeklendiren bir yöntemdir. Her gözlem, kendi değişkeninin ortalamasından ne kadar sapma gösterdiğine göre yeniden hesaplanır. Aykırı değerlere karşı daha dayanıklıdır ancak verilerin ortalaması ve standart sapması anlamlı değilse (örneğin yoğun biçimde çarpık dağılımlarda), bu yöntem etkili olmayabilir. Z-Skor normalizasyonu,

$$Z = \frac{X - \mu}{\sigma} \quad (5)$$

biçiminde hesaplanır. Formülde Z normalize edilmiş değeri, X orijinal veriyi, μ özniteliğin ortalamasını ve σ özniteliğin standart sapmasını temsil eder.

2.6. Öznitelik Mühendisliği

Makine öğrenmesi tabanlı uygulamalarda model performansını doğrudan etkileyen en kritik adımlardan biri, veri setinden anlamlı özniteliklerin elde edilmesidir. Bu süreç genel olarak öznitelik mühendisliği olarak adlandırılmakta ve iki temel aşamadan oluşmaktadır: öznitelik çıkarımı ve boyut indirgeme.

Öznitelik çıkarımı, ham verilerden istatistiksel, zamansal veya dönüşüme dayalı yöntemlerle bilgi içeriği yüksek yeni değişkenlerin elde edilmesini kapsar. Bu adım, özellikle sensör verilerinde gizli yapıları açığa çıkarmak ve sınıflandırma başarımını artırmak açısından büyük önem taşır.

Boyut indirgeme ise, elde edilen öznitelikler arasında bilgi tekrarını azaltmak, gürültüyü ortadan kaldırmak ve modelin daha genel geçer öğrenmesini sağlamak amacıyla uygulanır. Bu aşamada amaç, minimum bilgi kaybıyla veri uzayının boyutunu azaltmaktır.

2.6.1. Öznitelik Çıkarımı

Öznitelik çıkarımı, ham verilerden anlamlı, ayırt edici ve ölçülebilir niteliklerin elde edilmesi sürecidir. Özellikle sensör tabanlı sistemlerde, bu adım makine öğrenmesi algoritmalarının başarılı çalışabilmesi açısından kritik öneme sahiptir. Ham sensör sinyalleri çoğu zaman yüksek boyutlu, gürültülü ve sınıflandırma için doğrudan uygun olmayan bir yapı sergiler. Bu nedenle, verideki gizli örüntülerin açığa çıkarılması ve anlamlı bilginin sadeleştirilerek sunulması gerekir.

Öznitelik çıkarımı hem modelin doğruluğunu artırır hem de hesaplama maliyetini düşürerek öğrenme sürecinin daha verimli olmasına katkı sağlar. Ayrıca, aşırı öğrenme riskini azaltarak modelin genellenebilirliğini artırır. Öznitelik çıkarımı aşamasında elde edilen veriler, yapısal ve istatistiksel özelliklerine göre çeşitli kategoriler altında sınıflandırılabilir.

- Temel istatistiksel öznitelikler: Sensör verilerinin temel dağılım özelliklerini özetleyen tanımlayıcı ölçümlerdir. (maksimum, minimum, medyan, ortalama, standart sapma, varyans, kurtosis (çarpıklık), skewness (basıklık))
- Zaman temelli öznitelikler: Zaman serisi şeklinde kaydedilen sinyallerin zamana bağlı değişimlerini ve dinamik yapılarını yansıtır. (örneğin: sinyal eğimi, sinyalin altında kalan alan, pik sayısı, tepki süresi, toparlanma süresi, yükselme süresi, geçiş süresi)
- Frekans-temelli öznitelikler: Sinyalin frekans alanındaki karakteristiklerini ifade eder ve genellikle Fourier ya da Wavelet dönüşümleri ile elde edilir.
- Türetilmiş öznitelikler: İki veya daha fazla sensör çıktısının birleşiminden elde edilen yeni niteliklerdir. (örneğin: sensör farkları, oranlar, korelasyon katsayısı)

Öznitelik çıkarımı, elektronik burun gibi çok sensörlü sistemlerde oldukça yaygındır. Çünkü sensör yanıtları çoğu zaman doğrudan gaz türü veya durum hakkında yeterli ayrım gücüne sahip değildir; bu nedenle çeşitli sinyal karakteristikleri çıkarılarak sınıflandırma başarımları artırılır.

2.6.2. Boyut İndirgeme Yöntemleri

Boyut indirgeme, çok sayıda öznitelik içeren veri setlerini daha az sayıda öznitelikle ifade etmeyi sağlayan bir yöntemdir. Özellikle yüzlerce veya binlerce özniteliğe sahip verilerde, bazı öznitelikler tekrarlı olabilir veya modeli gereksiz yere karmaşıktırabilir. Bu gibi durumlarda boyut indirgeme kullanılarak, verinin özünü kaybetmeden daha sade bir yapı elde edilir.

Bu işlem sayesinde modelin eğitimi hızlanır, hesaplama maliyeti düşer ve modelin daha iyi genelleme yapması sağlanır. Ayrıca verinin daha kolay görselleştirilmesi ve yorumlanması da mümkün hale gelir. Boyut indirgeme, mevcut öznitelikleri doğrudan seçmek yerine, bu özniteliklerden yeni bileşenler oluşturarak çalışır. Ortaya çıkan yeni öznitelikler dönüştürülmüş yapılar olduğu için, yorumlanmaları genellikle daha zordur. Boyut indirgeme yöntemleri genellikle PCA ve LDA olmak üzere iki ana başlık altında incelenir.

PCA, bir veri kümesindeki toplam varyansı en iyi şekilde temsil eden yeni bir öznitelik uzayı oluşturmak amacıyla kullanılan, verideki bilgiyi kaybetmeden boyutsal indirgeme

sağlayan, doğrusal ve denetimsiz bir boyut indirgeme tekniğidir. Yüksek boyutlu verilerdeki korelasyonlu öznelilikleri daha az sayıda ve birbiriyle ilişkisiz (ortogonal) yeni bileşenler şeklinde ifade ederek boyut indirgeme işlemi yapan istatistiksel bir yöntemdir (Wold, Esbensen, & Geladi, 1987).

PCA matematiksel olarak birkaç adımdan oluşur. Birinci adım normalizasyon işlemidir. Bu adımın amacı, sürekli başlangıç değişkenlerinin aralığını, her birinin analize eşit katkıda bulunmasını sağlayacak şekilde standartlaştırmaktır. Bunun başlıca nedeni PCA'nın başlangıç değişkenlerinin varyansları konusunda oldukça hassas olmasıdır. Yani, başlangıç değişkenlerinin aralıkları arasında büyük farklar varsa, daha büyük aralıklara sahip değişkenler, küçük aralıklara sahip değişkenlere baskın gelecektir (örneğin, 0 ile 100 arasında değişen bir değişken, 0 ile 1 arasında değişen bir değişkene baskın gelecektir). Bu işlem, genellikle Z-Skor normalizasyonu uygulanarak, her bir özneliliğin ortalaması sıfır ve standart sapması bir olacak şekilde yeniden ölçeklendirilmesiyle gerçekleştirilir.

PCA'nın ikinci adımı kovaryans matrisinin hesaplanmasıdır. Bu adımın amacı, veri kümesindeki değişkenlerin aralarında herhangi bir ilişki olup olmadığını görmektir. Çünkü bazen değişkenler, gereksiz bilgiler içerecek şekilde yüksek oranda ilişkilidir. Bir değişkenin kendisiyle olan kovaryansı, o değişkenin varyansına eşittir; yani matematiksel olarak $Cov(X, X) = Var(X)$ bağıntısı geçerlidir. Bu nedenle, kovaryans matrisinin ana köşegeninde (sol üstten sağ alta uzanan diyagonal) her bir değişkenin varyans değeri yer alır. Ayrıca kovaryans simetrik bir ölçüttür; başka bir deyişle, $Cov(X, Y) = Cov(Y, X)$ eşitliği her zaman sağlanır. Bu özellik nedeniyle, kovaryans matrisi simetrik bir yapıya sahiptir; dolayısıyla matrisin üst ve alt üçgen bölgeleri birbiriyle aynıdır. Kovaryans matrisi Σ ya da C ile gösterilir. Kovaryans matrisinin her bir elemanı,

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (6)$$

biçiminde hesaplanır. Burada x_i ve y_i x ve y değişkenlerini, $\bar{x}$ ve $\bar{y}$ bu değişkenlerin ortalamasını, n ise örnek sayısını ifade eder. 3x3 lük bir kovaryans matrisi,

$$C = \begin{bmatrix} Var(X) & Cov(X, Y) & Cov(X, Z) \\ Cov(X, Y) & Var(Y) & Cov(Y, Z) \\ Cov(X, Z) & Cov(Z, Y) & Var(Z) \end{bmatrix} \quad (7)$$

biçiminde hesaplanır. PCA'nın üçüncü aşaması, özdeğerlerin ve özvektörlerin hesaplanmasıdır. Bu işlem, verilerin temel bileşenlerini belirlemek amacıyla kovaryans matrisinden gerçekleştirilir. Elde edilen özvektörler, veri kümesinde varyansın en yoğun olduğu yönleri gösterir; yani, verideki değişimlerin hangi doğrultularda daha fazla gerçekleştiğini ortaya koyar. Bu yönler, yeni bir koordinat sisteminin eksenleri hâline gelir ve orijinal veri boyutlarının doğrusal birleşimlerinden oluşur. Her bir özvektöre karşılık gelen özdeğer ise, ilgili yönün toplam varyans içerisindeki payını sayısal olarak ifade eder. Böylece, hangi bileşenlerin veride daha fazla bilgi taşıdığı belirlenebilir. Özdeğerlerin büyükten küçüğe sıralanmasıyla, en bilgilendirici yönler önceliklendirilir ve yalnızca en yüksek varyansı açıklayan bileşenler seçilerek boyut indirgeme gerçekleştirilir. Bu işlem, hem verinin daha sade bir biçimde temsil edilmesini sağlar hem de hesaplama yükünü azaltarak model performansını artırma potansiyeli sunar. Özdeğerler ve özvektörler her zaman çiftler hâlinde bulunur; her özvektör kendisine karşılık gelen bir özdeğere sahiptir. Özvektör ve özdeğer sayısı, veri kümesindeki boyut (değişken) sayısına eşittir. Örneğin, üç boyutlu bir veri kümesinde üç değişken bulunduğundan, üç özvektör ve bu özvektörlere karşılık gelen üç özdeğer hesaplanır. Bu ilişki aşağıdaki özdeğer-özvektör denklemi ile ifade edilir.

$$Cv = \lambda v \quad (8)$$

Burada C dxd boyutlu kovaryans matrisini, λ özdeğeri, v ise λ 'ya karşılık gelen özvektörü temsil eder. Özdeğerleri elde etmek için öncelikle aşağıdaki karakteristik denklem çözülür:

$$\det(C - \lambda I) = 0 \quad (9)$$

Bu denklemde I dxd boyutlu birim matrisi temsil eder. Bu denklemin çözümü ile birlikte her bir λ_i özdeğeri için aşağıdaki denklem çözülerek ilgili özvektör v_i bulunur.

$$(C - \lambda_i I)v_i = 0 \quad (10)$$

Bu özvektörler genellikle birim uzunlukta olacak şekilde normalize edilir. Özvektörler, özdeğerlerine göre büyükten küçüğe sıralanarak, temel bileşenlerin önem derecesine göre

sıralanması sağlanır. Böylece, verideki en yüksek bilgi içeriğine sahip yönlerden başlayarak, daha az bilgi taşıyan yönler doğru bir boyut indirgeme yapılabilir.

LDA, sınıflandırma problemlerinde kullanılan denetimli bir boyut indirgeme yöntemidir. LDA'nın temel amacı, farklı sınıfları birbirinden en iyi şekilde ayıran yeni bir alt uzay oluşturmaktır. Bunu yaparken sınıflar arasındaki varyansı maksimize ederken, sınıf içi varyansı minimize etmeye çalışır. LDA, sınıf etiketlerini kullanarak veri kümesini, sınıflar arasındaki farkların daha belirgin hale geldiği bir projeksiyon düzlemine indirger. Bu özellik, özellikle sınıflandırma görevlerinde ayırım gücünü artırır. Ancak LDA, sınıf bilgisi gerektirdiği için denetimli bir yöntemdir ve sınıf sayısına bağlı olarak çıkarılabilecek bileşen sayısı sınırlıdır (en fazla sınıf sayısı – 1 kadar).

2.7. Sınıf Dengesizliği Problemi

Makine öğrenmesi uygulamalarında sıkça karşılaşılan önemli sorunlardan biri, sınıf dağılımlarının dengesiz olduğu veri kümeleridir. Özellikle azınlık sınıfa ait örneklerin, çoğunluk sınıfa kıyasla çok daha az sayıda olduğu durumlarda, sınıflandırma algoritmaları genellikle çoğunluk sınıfa eğilim göstererek dengesiz sonuçlar üretir. Bu durum, modelin genel doğruluğunun yüksek görünmesine karşın, azınlık sınıfının düşük tespit başarımı nedeniyle uygulama açısından yetersiz performans sergilemesine yol açabilir.

Bu probleme çözüm bulmak amacıyla literatürde çeşitli veri düzeyli dengeleme yöntemleri geliştirilmiştir. Bu yöntemler, genellikle azınlık sınıfın örnek sayısını artırarak (aşırı örnekleme) ya da çoğunluk sınıfın örnek sayısını azaltarak (eksik örnekleme) sınıflar arasındaki dağılım farkını azaltmayı hedefler. Klasik aşırı örnekleme yaklaşımları, azınlık sınıfa ait verileri doğrudan kopyalayarak örnek sayısını artırır; ancak bu yöntemler, veri kümesinin çeşitliliğini artırmadığı için öğrenme algoritmalarında aşırı uyum riskini beraberinde getirir.

Bu sınırlamaları aşmak için geliştirilen SMOTE ve bu yönteme dayalı Borderline-SMOTE, SMOTE-Tomek ve ADASYN gibi gelişmiş teknikler, azınlık sınıfa ait daha anlamlı ve temsil gücü yüksek sentetik örnekler üretmeyi amaçlamaktadır. Bu yöntemler, azınlık sınıfına ait gözlemler etrafında yeni yapay (sentetik) veri noktaları oluşturarak sınıf dengesizliğini azaltır. Böylece sınıflar arası denge sağlanırken, modelin azınlık sınıf üzerindeki öğrenme kapasitesi ve genelleme gücü artırılmış olur. Aşağıda, söz konusu yöntemlerin temel prensipleri ve işleyiş mekanizmaları ayrıntılı olarak açıklanmıştır.

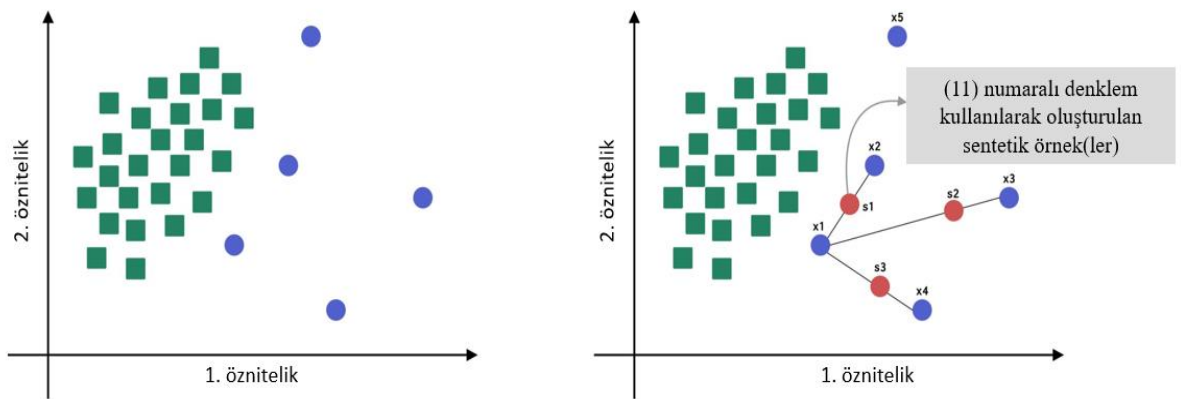
2.7.1. Sentetik Azınlık Aşırı Örneklemeye Tekniği

SMOTE, sınıf dengesizliği sorununu azaltmak amacıyla geliştirilmiş, veri düzeyinde bir aşırı örneklemeye yöntemidir. Bu yöntem klasik tekrar temelli örneklemeye yaklaşımlarına alternatif olarak azınlık sınıfa ait örneklerin etrafında yeni ve sentetik veri noktaları üretmeyi amaçlamaktadır (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Bu sayede modelin yalnızca çoğunluk sınıfa odaklanması engellenir ve azınlık sınıfa ait örneklerin karar sınırlarında daha iyi temsil edilmesi sağlanır. SMOTE, özellikle dengesiz veri kümelerinde sınıflandırma algoritmalarının genel doğruluğunu değil, azınlık sınıfa ait duyarlılığını artırmada etkili bir yöntem olarak öne çıkmaktadır.

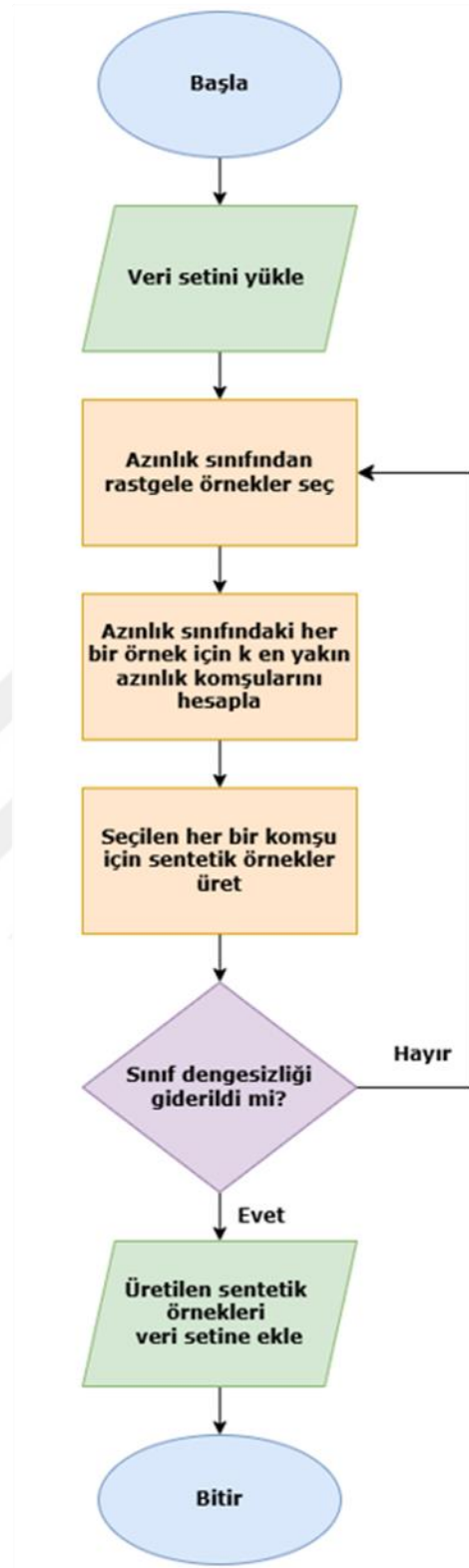
SMOTE azınlık sınıfa ait örneklerle onların yakın komşuları arasındaki doğrusal interpolasyon yoluyla sentetik örnekler üretir. Oluşturulan sentetik örnekler,

$$x_n = x_i + \delta(\hat{x}_i - x_i) \quad (11)$$

biçiminde tanımlanır. Burada x_n oluşturulan sentetik örneği, δ rastgele çarpanı, x_i azınlık sınıfına ait bir örneği, $\hat{x}_i$ ise bu örneğin en yakın azınlık komşusunu temsil eder. Rastgele çarpan 0 ile 1 arasında seçilen bir değerdir. Bu değer sentetik örneğin orijinal örneğe ne kadar yakın olacağını belirler. SMOTE yöntemiyle sentetik örneklerin oluşturulma süreci Şekil 9'da SMOTE yöntemi ile sentetik örneklerin oluşturulma süreci Şekil 10'da ise SMOTE algoritmasının akış diyagramı verilmiştir (López, 2024).



Şekil 9. SMOTE yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması



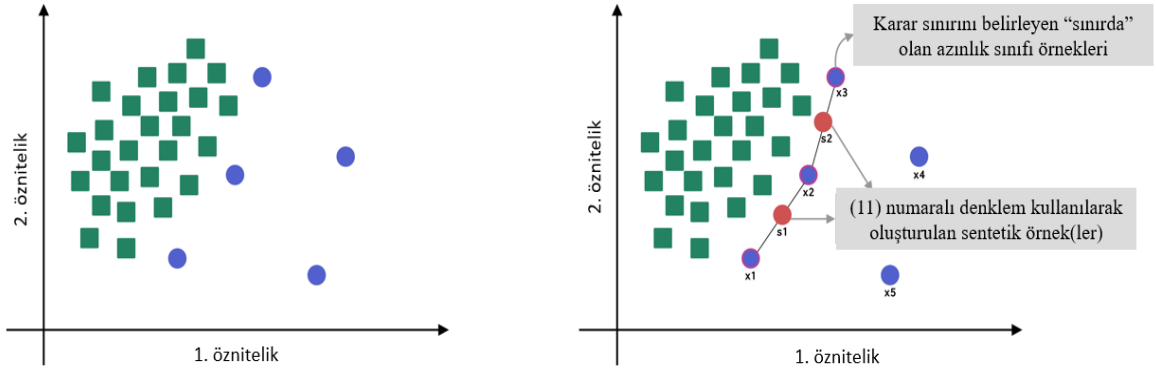
Şekil 10. SMOTE algoritması akış diyagramı

2.7.2. Sınırdaki Sentetik Azınlık Aşırı Örneklemeye Tekniği

Borderline-SMOTE, klasik SMOTE algoritmasının bir varyasyonudur ve özellikle sınıf ayırım çizgisine yakın olan azınlık sınıfı örneklerine odaklanarak sentetik veriler üretir (Han, Wang, & Mao, 2005). Bu yaklaşım, modelin karar sınırlarını daha hassas öğrenmesini sağlayarak sınıflar arasındaki ayrımı güçlendirir. Sınıf ayırım çizgisi (decision boundary), modelin farklı sınıflar arasında ayırt yapmasını sağlayan, karar mekanizmasını belirleyen sınırdır. Bu çizgiye yakın örnekler, sınıflar arası belirsizliğin en fazla olduğu noktalardır. Borderline-SMOTE, azınlık sınıfa ait ancak öğrenilmesi zor olan bu örnekleri çoğaltarak modelin genelleme yeteneğini artırmayı hedefler. Bu sayede, özellikle karmaşık ve örtüşen sınıflara sahip veri kümelerinde sınıflandırma başarımları daha dengeli hale getirilebilir. Borderline-SMOTE algoritmasının temel adımları aşağıda özetlenmiştir:

1. Veri kümesindeki azınlık sınıfı örnekleri tespit edilir.
2. Her bir azınlık sınıfı örneği için k-NN algoritması uygulanır. Ancak klasik SMOTE'tan farklı olarak burada, komşuların sınıflarına dikkat edilir.
 - Eğer bir azınlık örneğinin çoğunluk sınıfına ait komşularının oranı belirli bir eşiği (genellikle %50) aşıyorsa, bu örnek "borderline" (sınırdaki) olarak değerlendirilir.
 - Eğer çoğunluk komşu sayısı düşükse, örnek güvenli (safe) kabul edilir.
 - Çoğunluk komşu oranı çok yüksekse, örnek gürültü (noise) olarak değerlendirilir ve SMOTE uygulanmaz.
3. Yalnızca "borderline" olarak işaretlenen azınlık sınıfı örnekleri için sentetik veri üretilir. Bu sentetik örnekler, (11) numaralı denklem kullanılarak oluşturulur.
4. Üretilen sentetik örnekler veri kümesine eklenerek daha dengeli bir veri seti oluşturulur ve model eğitimi için kullanılır.

Borderline SMOTE yöntemi ile sentetik örneklerin oluşturulma süreci Şekil 11'de verilmiştir (López, 2024).



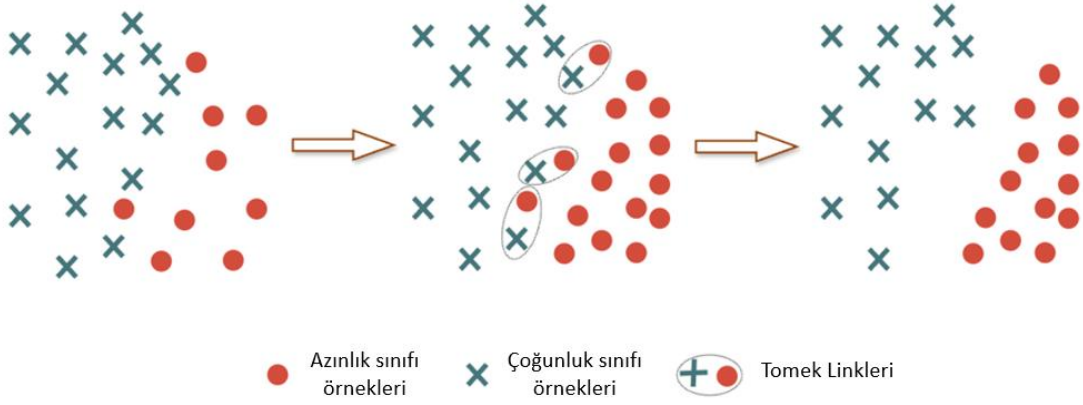
Şekil 11. Borderline SMOTE yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması

2.7.3. Sentetik Azınlık Aşırı Örnekleme ve Sınıf Sınırı Temizleme Tekniği

SMOTE-Tomek, dengesiz veri kümelerini dengelemek amacıyla SMOTE ve Tomek Link yöntemlerini birleştiren hibrit bir yaklaşımdır. Bu yöntem, bir yandan azınlık sınıfı için sentetik veriler üreterek örnek sayısını artırırken, diğer yandan sınıflar arasındaki belirsiz bölgeleri temizleyerek sınıflandırma performansını artırmayı hedefler (Batista, Prati, & Monard, 2004). Belirsiz bölgeler, farklı sınıflara ait örneklerin birbirine çok yakın bulunduğu, modelin karar vermekte zorlandığı sınıf sınırlarına yakın alanlardır; bu bölgeler modelde hatalı sınıflandırmalara yol açabilir. SMOTE-Tomek algoritmasının adımları aşağıda verilmiştir.

1. Öncelikle SMOTE uygulanarak azınlık sınıfına ait sentetik örnekler üretilir. Bu adım, azınlık sınıfının örnek sayısını artırarak dengesizliği azaltır.
2. Ardından veri setindeki Tomek Link Çiftleri belirlenir. Farklı sınıflara ait iki örnek birbirinin en yakın komşusuysa ve aralarında başka bir örnek yoksa oluşan çift Tomek Link Çifti olarak tanımlanır. Bu çiftler genellikle sınıf ayırım çizgisine yakın, karışıklığa neden olabilecek örneklerdir.
3. Tomek Link çiftlerinden çoğunluk sınıfına ait örnekler veri setinden silinir ve böylece sınıflar arasındaki sınır netleştirilir ve gürültülü veriler temizlenmiş olur.

SMOTE-Tomek yöntemi ile sentetik örneklerin oluşturulma süreci Şekil 12’de verilmiştir (Zhang, Deng, & Wei, 2024).



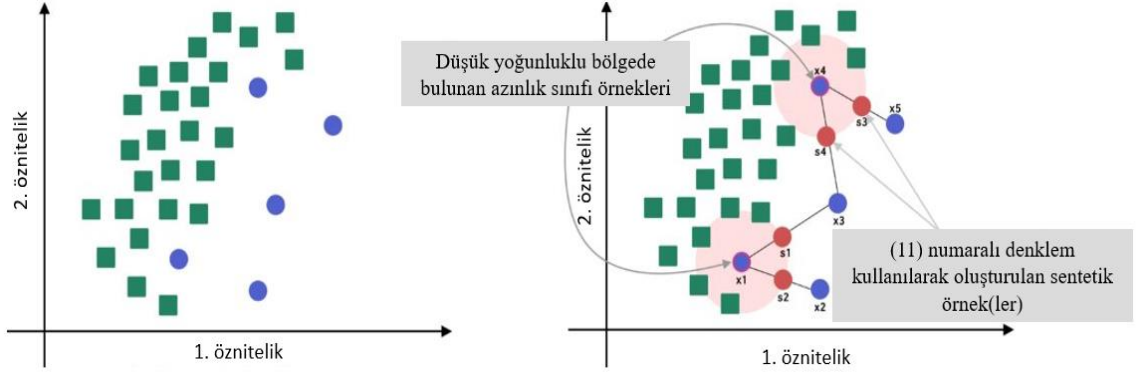
Şekil 12. SMOTE-Tomek yöntemi ile azınlık sınıfa ait sentetik örnek oluşturulması

2.7.4. Adaptif Sentetik Örneklemeye

ADASYN, dengesiz veri kümelerini dengelemek amacıyla geliştirilen, SMOTE temelli ancak daha adaptif bir aşırı örneklemeye yöntemidir. SMOTE gibi azınlık sınıfına sentetik örnekler ekleyerek sınıf dengesini sağlar; ancak farklı olarak, örnek üretim sürecini veri noktalarının öğrenilme zorluğuna göre uyarlayarak gerçekleştirir (He, Bai, Garcia, & Li, 2008). Bu sayede modelin sınıf sınırlarını daha doğru öğrenmesi amaçlanır. ADASYN algoritmasının temel adımları aşağıda verilmiştir.

- 1) Azınlık ve çoğunluk sınıfına ait örnekler belirlenir.
- 2) Her bir azınlık sınıfı örneği için k-en yakın komşuları tespit edilir. Bu komşuların kaç tanesinin çoğunluk sınıfına ait olduğu belirlenir. Çoğunluk sınıfı komşusu fazla olan azınlık örnekleri, model için daha zor öğrenilen bölgelerde yer alır. ADASYN, sentetik veri üretimini bu "zor" örnekler etrafında yoğunlaştırır.
- 3) Her örneğin zorluk derecesine göre üretilecek sentetik örnek sayısı belirlenir. Böylece daha az temsil edilen, belirsiz ve kritik bölgeler daha fazla örnekle desteklenmiş olur. Bu adaptif yaklaşım, öğrenme sürecini daha dengeli hale getirir.
- 4) Sentetik örnekler, tıpkı SMOTE’da olduğu gibi (11) denklemi kullanılarak üretilir.
- 5) Üretilen örnekler veri setine eklenerek model eğitimi için kullanılır.

ADASYN yöntemi ile sentetik örneklerin oluşturulma süreci Şekil 13’te verilmiştir (López, 2024).



Şekil 13. ADASYN yöntemi ile azınlık sınıfı ait sentetik örnek oluşturulması

2.8. Sınıflandırma Algoritmaları

Sınıflandırma süreci, ham verilerden çıkarılan anlamlı özelliklerin kullanılmasıyla, verilerin önceden tanımlanmış sınıflara doğru bir şekilde atanmasını sağlayan makine öğrenmesi tekniklerini içerir. Bu algoritmalar, etiketli eğitim verisi üzerinde öğrenme gerçekleştirir; yani, her bir eğitim örneği hem özellik bilgilerini hem de ait olduğu sınıf etiketini içerir. Model, bu etiketli verilerden sınıflar arasındaki ayırt edici örüntüleri öğrenerek, daha sonra karşılaştığı yeni veya bilinmeyen verilerin sınıfını tahmin edebilir.

2.8.1. Lojistik Regresyon Sınıflandırma Algoritması

Lojistik regresyon, iki veya çok sınıflı sınıflandırma problemlerinde yaygın olarak kullanılan, denetimli bir makine öğrenmesi algoritmasıdır. Temel amacı, bir gözlemin belirli bir sınıfa ait olma olasılığını tahmin etmektir (Hosmer Jr, Lemeshow, & Sturdivant, 2013). Düşük hesaplama maliyeti ve doğrusal karar sınırları üretme yeteneği sayesinde pek çok uygulamada tercih edilmektedir.

Modelin temel amacı, bir gözlemin belirli bir sınıfa ait olma olasılığını tahmin etmektir. Bu nedenle lojistik regresyon çıktısı 0 ile 1 arasında bir olasılık değeridir. Bu değer belirli bir eşik (genellikle 0.5) ile karşılaştırılarak sınıf tahmini yapılır. Örneğin, $\sigma(z) = 0.85$ çıktısı, gözlemin %85 ihtimalle pozitif sınıfa ait olduğunu ifade eder.

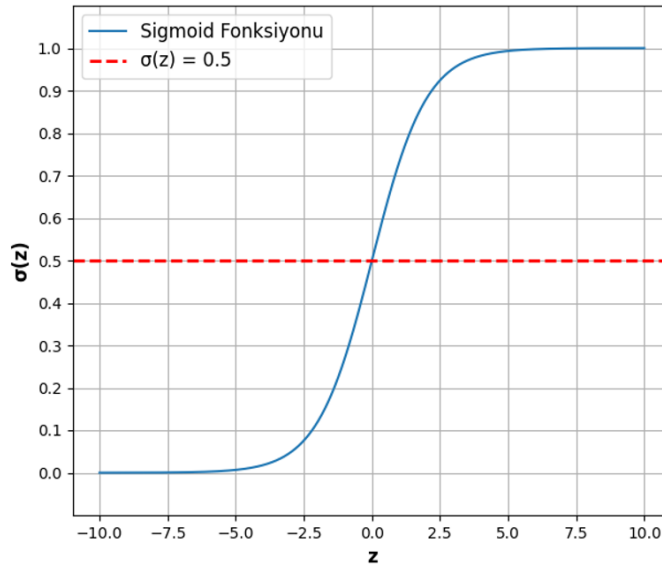
Lojistik regresyon, doğrusal bir fonksiyon kullanarak özellikler ile sınıf arasında ilişki kurar:

$$z = w^T x + b \quad (12)$$

Bu denklemde x öznitelik vektörünü, x modelin öğrenmeye çalıştığı ağırlıkları, b sabit terimi (bias) ve z doğrusal çıktıyı temsil eder. Bu doğrusal birleşim, sigmoid fonksiyon ile 0-1 aralığında değere dönüştürülür. Sigmoid fonksiyonu,

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (13)$$

biçiminde tanımlanır. Sigmoid fonksiyonuna ait grafik Şekil 14'te verilmiştir.



Şekil 14. Sigmoid fonksiyonu

Lojistik regresyon, doğrusal bir karar sınırı belirleyerek veri noktalarını iki sınıfa ayırmaya çalışır. Modelin temel amacı, her bir örneğin ait olduğu sınıfı doğru tahmin edebilmek için olasılık temelli bir tahmin fonksiyonu (sigmoid) ile çalışan bir karar yüzeyi öğrenmektir. Ancak yüksek boyutlu veya çok sayıda öznitelik içeren veri setlerinde model, eğitim verisine aşırı uyum sağlayarak aşırı öğrenme riski taşıyabilir. Bu riski azaltmak için regülerizasyon adı verilen bir yaklaşım kullanılır. Lojistik regresyonda kullanılan regülerizasyonlu maliyet fonksiyonu iki terimden oluşur. En genel haliyle, regülerize edilmiş bir maliyet fonksiyonu,

$$J(w) = \mathcal{L}(w) + \lambda \Omega(w) \quad (14)$$

biçiminde yazılır. Burada z modelin öğrenilen parametrelerini (ağırlıklar), $\mathcal{L}(w)$ tahmin edilen değerlerle gerçek değerler arasındaki farkı ölçen kayıp fonksiyonunu (genellikle log-kayıp veya çapraz-entropy), $\Omega(w)$ modelin ağırlıklarının büyüklüğüne dayalı olan ceza terimini, yani modelin karmaşıklığını ölçer (örneğin L1 veya L2 normları), λ bu iki hedef arasında bir denge sağlayan regülerizasyon katsayısıdır.

Buradaki hedef, hem modelin veriye iyi uymasını (doğru sınıflandırma yapmasını) sağlamak hem de modelin ağırlıklarını kontrol altında tutarak genelleme kabiliyetini korumaktır. Ancak bu iki hedef birbiriyle çelişebilir.

- 1) Ceza terimi çok azaltmak istenirse ağırlıklar büyüyebilir ve model eğitim verisine çok iyi uyum sağlar ancak genelleme kabiliyeti düşer ve aşırı öğrenmeye yol açar
- 2) Veri hatası çok azaltmak istenirse model karmaşılaşabilir ve bazı küçük detaylara gereğinden fazla uyum sağlar.

Regülerizasyon parametresi olan λ bu iki hedef arasında denge kurar. λ büyük seçilirse, ceza etkisi artar, ağırlıklar bastırılır, model sadeleşir ve overfitting riski azalır. λ küçük seçilirse, model veriye daha iyi uyum sağlar ancak genelleme kabiliyeti düşebilir. Bu denge sayesinde LR hem eğitim verisine uyum sağlamaya hem de aşırı karmaşılaşmadan kaçınmaya çalışır.

Python'daki Scikit-learn kütüphanesinde, doğrudan λ yerine onun tersi olan C parametresi kullanılır. Küçük C değerleri, güçlü regülerizasyon etkisi yaratır; model daha sade olur, aşırı öğrenme riski azalır. Büyük C değerleri, ceza etkisini azaltır; model veriye daha iyi uyum sağlar, ancak aşırı öğrenme riski artar.

Lojistik regresyon modelleri, kullanılan veri yapısına ve regülerizasyon türüne bağlı olarak farklı optimizasyon algoritmaları (solver) ile eğitilebilir. Solver'lar, modelin maliyet fonksiyonunu minimize etmek amacıyla kullanılan sayısal optimizasyon yöntemleridir. Bu algoritmalar, modelin parametrelerini (w) belirlemek için maliyet fonksiyonunun minimum noktasını sayısal olarak arar. Farklı solver'lar, çeşitli matematiksel yaklaşımlar ve teknik sınırlamalar içerir. Hangi solver'ın kullanılacağı; veri setinin büyüklüğü, kullanılan regülerizasyon türü (L1, L2, Elastic Net) ve veri yapısının seyrek (sparse) olup olmaması, sınıf sayısı gibi etkenlere göre belirlenir. Kullanılabilecek farklı solver seçenekleri, yöntemleri, destekledikleri regülerizasyon türleri ve öne çıkan kullanım özellikleri Tablo 5'te özetlenmiştir.

Tablo 5. Lojistik Regresyon optimizasyon algoritmaları

Optimizasyon algoritmaları	Yöntem	Desteklediği Ceza	Öne Çıkan Özellikler
liblinear	Koordinat inışı	L1, L2	Küçük veri kümeleri için uygundur. L1 destekler.
lbfgs	Quasi-Newton	L2	Orta-büyük veri için hızlıdır, çok sınıflı sınıflandırmayı destekler.
saga	Stokastik Ortalama Gradyan	L1, L2, ElasticNet	L1 ve Elastic Net destekler, çok büyük veriler için uygundur.

2.8.2. Destek Vektör Makineleri Sınıflandırma Algoritması

SVM, sınıflandırma problemleri için kullanılan güçlü bir denetimli öğrenme algoritmasıdır. Temel amacı, farklı sınıflara ait veri noktalarını N boyutlu öznitelik uzayında birbirinden en iyi şekilde ayıran bir $(N-1)$ boyutlu hiper düzlem bulmaktır (Corinna & Vapnik, 1995). Öznitelik uzayında, sınıfları ayırabilen birçok olası hiper düzlem olabilir. SVM algoritması bu düzlemler arasından, sınıflar arasındaki marjini maksimum yapan optimal hiper düzlemi bulmayı hedefler. Marjin, iki sınıfa ait en yakın veri noktaları (yani destek vektörleri) arasındaki mesafedir. Bu noktalar, karar sınırına en yakın olan ve marjini belirleyen örneklerdir. Optimal hiper düzleme paralel olan ve bu destek vektörlerden geçen çizgiler, karar sınırını belirlemede kullanılır. Hiper düzlemin matematiksel ifadesi şu şekildedir:

$$w^T \times x - b = 0 \quad (15)$$

Formülde w hiper düzlemin normal vektörünü (yönünü), b öğrenme sabiti, x sınıflandırılacak veri noktasını temsil eder. Marjin, hiper düzleme olan en kısa dik mesafe olup, aşağıdaki şekilde hesaplanır:

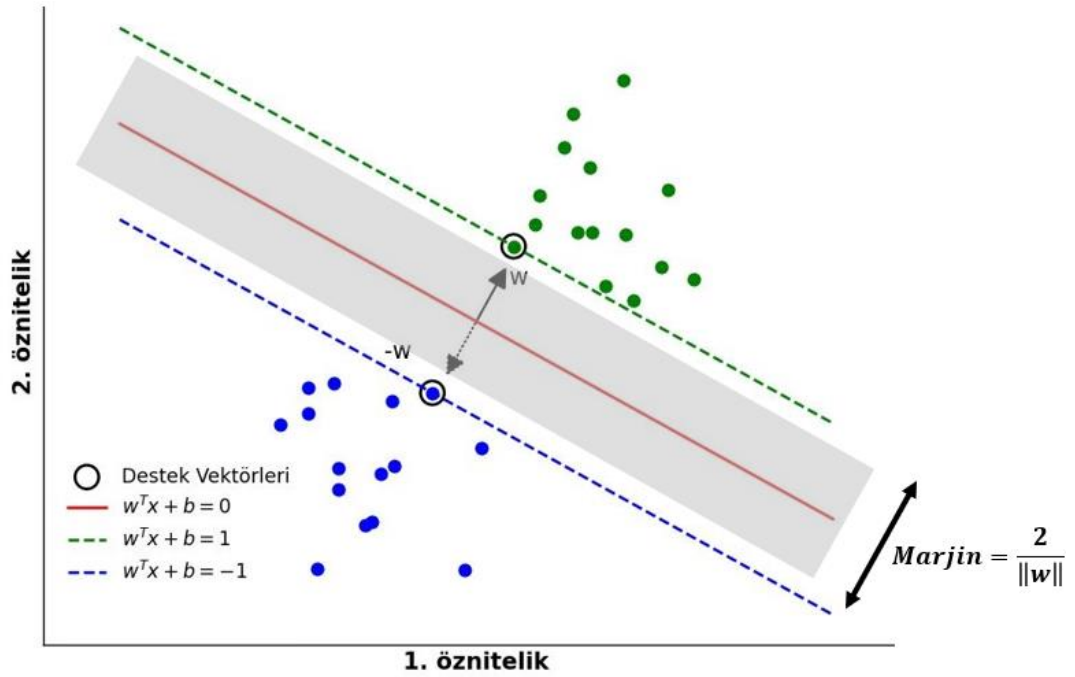
$$Marjin = \frac{2}{\|w\|} \quad (16)$$

Karar sınırı, bu marjini maksimum yapacak şekilde optimize edilir. Hiper düzlemin her iki yanında yer alan:

$$w^T \times x - b = 1 \quad (17)$$

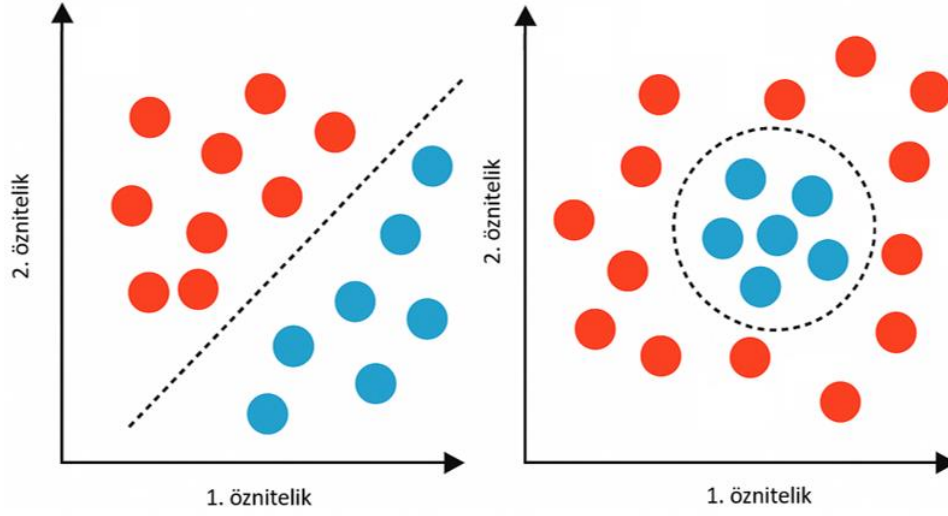
$$w^T \times x - b = -1 \quad (18)$$

çizgileri, destek vektörlerin üzerinden geçer ve marjin sınırlarını oluşturur. SVM algoritmasında karar sınırı, destek vektörlerine en yakın olacak şekilde yerleştirilir ve bu sınırın iki yanında kalan marjinler, sınıflar arasındaki ayrımı en üst düzeye çıkarmayı hedefler. Bu yapının geometrik görünümü, destek vektörleri ve w vektörünün yönüyle birlikte Şekil 15'te sunulmuştur.



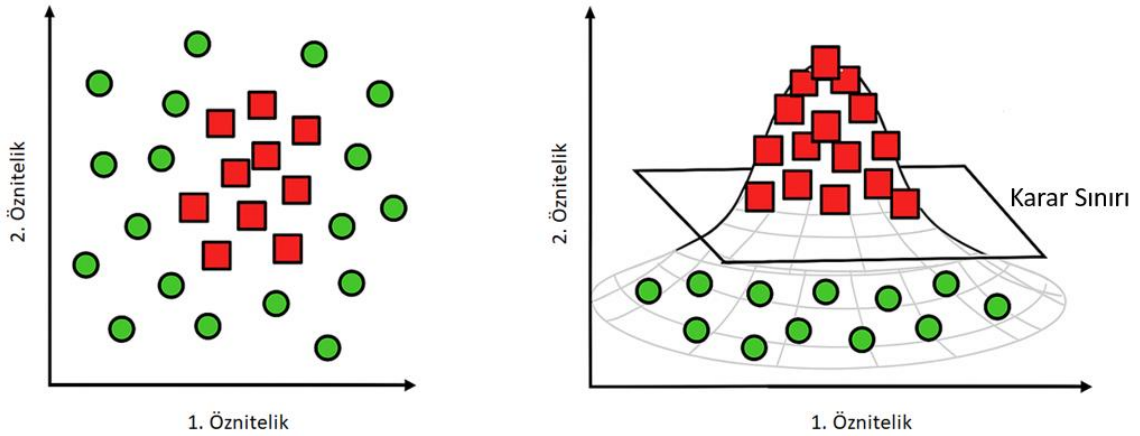
Şekil 15. SVM karar sınırı, marjin ve w vektörünün gösterimi

Veri setindeki sınıflar doğrusal olarak ayrılabiliriyorsa, tek bir hiper düzlem yeterlidir. Ancak birçok gerçek dünya problemlerinde sınıflar doğrusal ayrılabilir değildir. Bu durum, Şekil 16'da lineer olarak ayrılabilen ve ayrılabilen sınıflandırma problemlerinin karşılaştırılmasıyla görselleştirilmiştir. Lineer olarak ayrılabilen veri kümeleri için, verilerin daha yüksek boyutlu bir uzaya dönüştürülerek bu uzayda ayrılabilir hale getirilmesi gerekmektedir. Bu dönüşüm, çekirdek (kernel) fonksiyonları aracılığıyla gerçekleştirilir. SVM algoritmasında bu amaçla yaygın olarak kullanılan çekirdek fonksiyonları arasında lineer kernel, poly kernel ve rbf kerneldir.



Şekil 16. a) Lineer olarak ayrılabilen b) Lineer olarak ayrılamayan veri örnekleri

Doğrusal olarak ayrılabilir olmayan veriler, kernel fonksiyonları aracılığıyla daha yüksek boyutlu uzaylara dönüştürülerek ayrılabilir hale getirilebilir. Bu dönüşümün temel mantığı Şekil 17’de görsel olarak sunulmuştur.



Şekil 17. Kernel dönüşümü ile doğrusal olmayan verilerin ayırt edilmesi

SVM, sınıflar arası en geniş marjini bulmaya çalışırken aynı zamanda da mümkünse hiçbir noktayı yanlış sınıflandırmamak ister. Yani, SVM’nin birinci amacı olan marjini $\left(\frac{2}{\|w\|}\right)$ geniş tutmak için ağırlık vektörünün büyüklüğü w küçük tutmak istenir. SVM’nin ikinci amacı olan toplam sınıflandırma hatasını ölçen $C \sum \delta_i$ terimi de mümkün olduğunca

küçük tutulmaya çalışılır. Fakat bu ikisi her zaman aynı anda mümkün değildir. Bunun nedeni:

- 1) w çok küçültülmeye çalışılırsa marjın çok fazla genişler ve bu da bazı noktaların yanlış sınıflandırılmasına yol açarak $C \sum \delta_i$ terimini aşırı büyütür.
- 2) $C \sum \delta_i$ terimi çok küçültülmeye çalışılırsa da w büyümek zorunda kalır marjın daralır ve model aşırı öğrenme yapabilir.

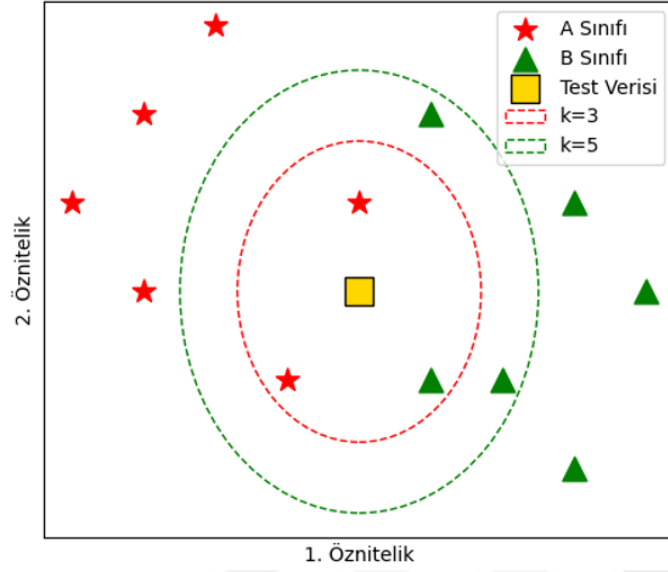
İdealde her iki terim de ayrı ayrı küçültülmek istenir ancak biri küçültülürken diğeri ise genellikle büyür. Bu yüzden aşağıdaki toplam ifadesi kullanılarak ikisi birden optimize edilmeye çalışılır. SVM bu iki terimi aynı amaç fonksiyonunda birleştirerek dengeyi optimize eder. SVM optimizasyon problemi, kısıtlı bir optimizasyon problemidir ve aşağıda ifadeyi minimum yapmak ister:

$$\min \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \delta_i \right) \quad (19)$$

Burada w marjın genişliğini kontrol ederken δ_i örnek başına hata miktarını C ise yapılan her bir hatalı sınıflandırma için verilecek cezayı belirler. C parametresi, bu iki hedef arasında bir denge unsuru olarak devreye girer. C değeri büyük olduğunda, model sınıflandırma hatalarına daha az tolerans gösterir ve her örneği doğru sınıflandırmaya çalışır. Bu durum, marjın daralmasına ve modelin eğitim verisine aşırı uyum sağlamasına (overfitting) neden olabilir. C değeri küçük olduğunda ise model bazı sınıflandırma hatalarına izin vererek marjını geniş tutmayı tercih eder. Bu da daha sade ve genellenebilir bir model elde edilmesini sağlar.

2.8.3. k-En Yakın Komşuluk Sınıflandırma Algoritması

k-NN algoritması, denetimli öğrenmeye dayalı bir sınıflandırma yöntemidir. Bu algoritma, bir veri noktasının sınıfını belirlerken, o noktaya en yakın k komşunun sınıflarını dikkate alır (Cover & Hart, 1967). Karar verme sürecinde çoğunluk oyu kuralı uygulanır. En yakın komşular arasında en sık görülen sınıf, yeni veri noktasına atanır. Örneğin, iki sınıflı bir problemde $k=5$ olarak belirlendiğinde, test verisine en yakın 5 komşudan 3'ü "A" sınıfına ve 2'si "B" sınıfına aitse, yeni veri noktası çoğunluk olan "A" sınıfına atanır. İki veya daha fazla sınıf eşit sayıda temsil ediliyorsa, algoritma genellikle rastgele seçim yapar. k-NN algoritmasında kullanılan k değerinin sınıflandırma kararına etkisi Şekil 18'de verilmiştir.



Şekil 18. k-NN algoritmasının sınıflandırma kararına etkisi

k-NN algoritması, eğitim aşamasında herhangi bir genelleştirici model oluşturmaz; bunun yerine tüm eğitim verilerini saklar ve sınıflandırmayı tahmin sırasında yapar. Bu nedenle “tembel öğrenme” yöntemleri arasında yer alır. Tembel öğrenme, eğitim sırasında bilgi öğrenmek yerine tahmin sırasında bütün öğrenme sürecini gerçekleştiren algoritmalar için kullanılan bir terimdir. Bu durum, k-NN’nın eğitim sürecini hızlı hale getirirken, tahmin sürecinde daha fazla hesaplama yapılmasına neden olur. k-NN algoritması, sınıflandırma performansını doğrudan etkileyen iki temel hiperparametreye sahiptir:

- a) Komşu sayısı (k): Bu parametre, yeni bir veri noktasının sınıflandırılmasında kaç adet komşunun dikkate alınacağını belirler. Küçük k değerleri, modeli gürültüye karşı daha hassas hale getirir; karar sınırları keskinleşir ve aşırı öğrenme riski artabilir. Büyük k değerleri ise sınıf sınırlarını yumuşatır, ancak sınıflar arasındaki ayrım gücünü azaltabilir.
- b) Uzaklık metriği: Veri noktaları arasındaki “yakınlık”, seçilen mesafe ölçütüne göre belirlenir. Uzaklık ölçütü, hangi komşuların “en yakın” olduğunu doğrudan etkiler ve dolayısıyla modelin kararlarını şekillendirir. Yaygın kullanılan mesafe ölçütleri şunlardır:
 - Öklid mesafesi: İki nokta arasındaki doğrusal (kuş uçuşu) mesafeyi ölçer. İki boyutta Pisagor teoremi gibi düşünülebilir. Gerçek dünya mesafelerine en yakın olanıdır. Öklid mesafesi aşağıdaki biçimde hesaplanır.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (20)$$

- Manhattan mesafesi: İki nokta arasındaki mesafeyi yalnızca yatay ve dikey yönlerde giderek ölçer. Örnek olarak şehir bloklarındaki yürüyüş mesafesi verilebilir. Manhattan mesafesi aşağıdaki biçimde hesaplanır.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (21)$$

- Chebyshev mesafesi: İki nokta arasındaki maksimum eksen farkı dikkate alınır. Chebyshev mesafesi aşağıdaki biçimde hesaplanır.

$$d(x, y) = \max(|x_i - y_i|) \quad (22)$$

2.8.4. Rastgele Orman Sınıflandırma Algoritması

RF, makine öğrenmesinde kullanılan güçlü bir topluluk öğrenme yöntemidir. Yüksek doğruluk oranı ve aşırı öğrenmeye karşı dayanıklılığı nedeniyle sıkça tercih edilmektedir. Ardındaki temel düşünce birden fazla zayıf modelin birleştirilmesi ile tek bir güçlü model oluşturmaktır (Breiman, 2001). Tek bir karar ağacı aşırı öğrenmeye yatkın olabilir. RF, tek bir karar ağacının zayıf yönlerini ortadan kaldırmak için birden fazla ağacın birlikte çalışmasını sağlar. Bu yöntemde:

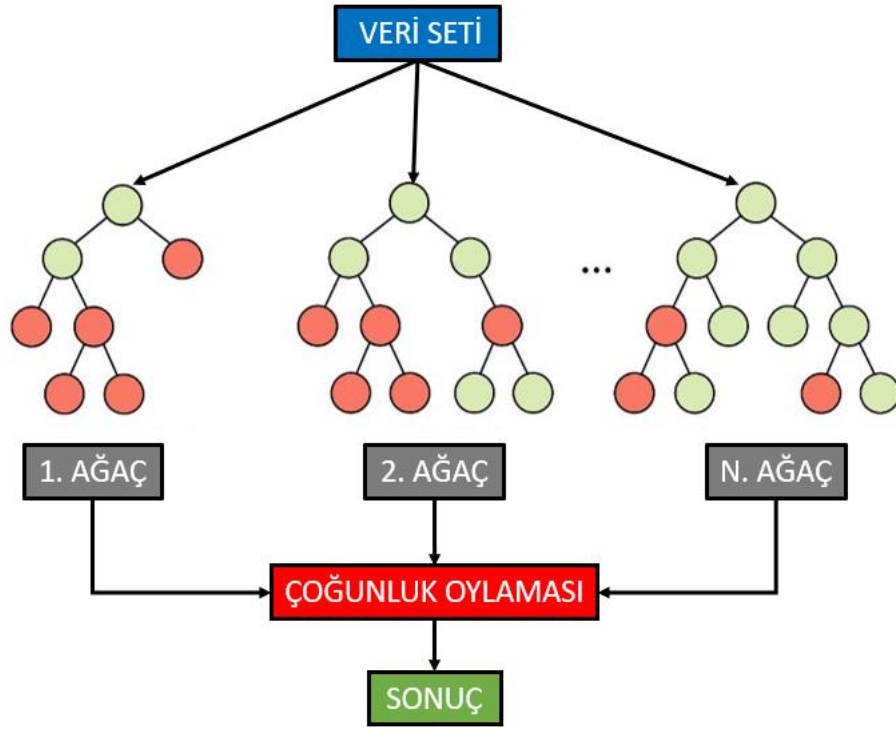
- Her bir ağaç veri setinin farklı bir rastgele örneklemeyle (bootstrap örnekleme) eğitilir.
- Ayrıca, her ağaç özelliklerin tamamı yerine, rastgele seçilen bir alt kümesiyle karar verir.

Bu iki tür rastgelelik sayesinde, ağaçlar arasında çeşitlilik sağlanır ve modelin genellenebilirliği artar. RF algoritmasının çalışma prensibi şu şekildedir:

1. Veri Hazırlığı (Bootstrap örnekleme): Ana eğitim veri setinden bootstrap örnekleme yöntemiyle rastgele ve tekrar etmeye izin vererek alt veri kümeleri oluşturulur. Her alt küme, orijinal veri kümesinin bir alt kümesi olur ve bazı örnekler birden fazla kez seçilebilirken bazıları hiç seçilmeyebilir.

2. Rastgele öznitelik seçimi: Her ağaçtaki her düğümde, tüm öznitelikler arasından rastgele bir alt küme seçilir ve bu kümeden en iyi bölme yapılır.
3. Ağaçların Eğitilmesi: Oluşturulan her alt veri kümesi için bağımsız karar ağaçları eğitilir. Rastgele Orman özelinde, her ağacın her düğümünde, özniteliklerin rastgele bir alt kümesi seçilir ve en iyi bölme noktası belirlenir.
4. Tahmin aşaması (Aggregating): Sınıflandırma problemlerinde: Her ağaç kendi tahminini yapar; sonuçlar oylama (majority voting) yöntemiyle birleştirilir. Regresyon problemlerinde: Ağaçların çıktılarının aritmetik ortalaması alınır.

Bu işlemlerin bütünsel akışı Şekil 19'da verilmiştir.



Şekil 19. Rastgele Orman algoritmasında karar ağaçlarının oylama süreci

En iyi bölme noktası, karar ağacı oluşturma sürecinde kritik bir rol oynar ve Gini safsızlık ölçütü kullanılarak belirlenir. Başlangıçta, tüm alt küme verisi tek bir düğümde bulunur ve bu düğümün başlangıç Gini safsızlığı hesaplanır. Daha sonra, farklı öznitelikler kullanılarak veri çeşitli senaryolarla bölünmeye çalışılır. Her bölme sonrası, sol ve sağ düğümlerin Gini safsızlıkları hesaplanarak bölmenin genel Gini azalması değerlendirilir. En düşük Gini'yi sağlayan bölme noktası seçilir ve bu bölme uygulanarak düğümler daha saf hale getirilir. Aynı işlem her yeni düğümde tekrar edilir ve bu süreç, belirlenen durdurma

koşulları (örneğin maksimum derinlik, minimum örnek sayısı vb...) sağlanana kadar devam eder. Gini safsızlığı,

$$Gini = 1 - \sum_{i=1}^c p_i^2 \quad (23)$$

biçiminde hesaplanır. Bu denklemde C sınıf sayısı, p_i ise sınıfa ait örneklerin oranıdır (sınıfın düğümdeki sıklığı). Gini safsızlığına alternatif olarak, entropi ölçütü de kullanılabilir. Rastgele Orman algoritmasında en önemli iki hiperparametre şunlardır:

- Ağaç sayısı: Ormanda oluşturulacak karar ağacı sayısını belirler. Daha fazla ağaç genellikle daha iyi genelleştirme sağlar ve modelin doğruluğunu artırır. Ancak, çok fazla ağaç kullanımı hesaplama maliyetini artırabilir. Genellikle 100-500 arasında bir değer uygundur, ancak veri kümesine ve işlem gücüne bağlı olarak optimize edilmelidir.
- Maksimum derinlik: Her bir karar ağacının maksimum derinliğini belirler. Çok büyük bir değer seçilirse, ağaçlar aşırı öğrenebilir. Çok küçük bir değer seçilirse, model yeterince öğrenemez. Varsayılan olarak None, yani ağaçlar tamamen büyüyebilir. Optimum değer genellikle 5-20 arasında değişir, ancak veri kümesine göre optimize edilmelidir.

RF algoritması, aşırı öğrenmeye karşı dayanıklı olup yüksek doğruluk oranına sahiptir. Ayrıca, eksik veriler ve veri kayıpları karşısında sağlam bir performans sergileyerek güvenilir tahminler üretebilir. Ancak, büyük veri kümeleri üzerinde çalışırken yüksek hesaplama maliyeti ve bellek kullanımı gerektirebilir.

2.9. Çapraz Doğrulama

Çapraz Doğrulama Yöntemi: Çapraz doğrulama (cross-validation), bir makine öğrenmesi modelinin genellenebilirliğini değerlendirmek amacıyla kullanılan bir doğrulama yöntemidir. Bu yöntem, modelin yalnızca belirli bir eğitim ve test verisi üzerinde değil, tüm veri kümesi genelinde nasıl performans gösterdiğini nesnel bir biçimde incelemeyi sağlar. En yaygın kullanılan türü k-katlı çapraz doğrulama (k-fold cross-validation) olup, veri kümesi eşit büyüklükte k alt gruba (kat) bölünür. Her iterasyonda, bir kat test verisi olarak ayrılırken geri kalan k-1 kat eğitim amacıyla kullanılır. Bu işlem k kez tekrarlanır ve her kat

bir kez test verisi olarak kullanılmış olur. Elde edilen performans metrikleri ortalanarak modelin genel başarımı hakkında daha güvenilir bir değerlendirme yapılır.

Standart çapraz doğrulama yönteminde, veriler rastgele ve eşit büyüklükteki katlara bölünür. Ancak bu yaklaşım, özellikle dengesiz sınıf dağılımlarının söz konusu olduğu durumlarda, her katın sınıf oranlarını korumaması nedeniyle yanıltıcı sonuçlar üretebilir. Bu sorunu gidermek amacıyla geliştirilen katmanlı k-kat çapraz doğrulama (Stratified k-fold cross-validation) yönteminde ise, her bir kat oluşturulurken veri kümesindeki sınıf dağılımı korunur (Szeghalmy & Fazekas, 2023). Yani her kat, orijinal veri kümesindeki sınıf oranlarına uygun şekilde örnekler içerir. Bu yöntem, özellikle sınıflar arasında ciddi dengesizliklerin olduğu sınıflandırma problemlerinde modelin daha tutarlı ve güvenilir bir şekilde değerlendirilmesini sağlar.

2.10. Performans Metrikleri

Makine öğrenmesi modellerinin başarısını değerlendirmek, geliştirilen sistemlerin etkinliğini ve güvenilirliğini belirlemek açısından kritik bir adımdır. Bu değerlendirme, modelin yaptığı tahminlerin doğruluğunu ve güvenilirliğini ölçen çeşitli performans metrikleri aracılığıyla gerçekleştirilir.

Doğruluk, kesinlik, duyarlılık ve F1 skoru gibi temel performans metrikleri özellikle sınıflandırma problemlerinde modelin başarımını değerlendirmek için yaygın olarak kullanılmaktadır (Hossin & Sulaiman, 2015). Bu metrikler, modelin tahmin performansını sayısal olarak ifade eder ve çoğunlukla konfüzyon matrisi üzerinde tanımlanan bileşenler — doğru pozitif (DP), yanlış pozitif (YP), doğru negatif (DN) ve yanlış negatif (YN) — üzerinden hesaplanır. Konfüzyon matrisi, modelin tahmin ettiği sınıflar ile gerçek sınıflar arasındaki ilişkiyi görsel ve sistematik bir biçimde özetleyen bir tablodur. Genellikle ikili sınıflandırma problemleri için 2x2 boyutunda bir yapıdadır; ancak sınıf sayısının artmasıyla birlikte matrisin boyutu da genişler. Şekil 20'de, iki sınıflı bir sınıflandırma problemi için kullanılan örnek bir konfüzyon matrisi yer almaktadır.

		Gerçek Sınıflar	
		Pozitif	Negatif
Tahmin Edilen Sınıflar	Pozitif	DP	YP
	Negatif	YN	DN

Şekil 20. İki sınıflı bir problem için örnek bir konfüzyon matrisi

Burada:

Doğru Pozitif (DP): Modelin doğru bir şekilde pozitif olarak tahmin ettiği örnekleri,

Yanlış Pozitif (YP): Modelin yanlış bir şekilde pozitif olarak tahmin ettiği örnekleri,

Doğru Negatif (DN): Modelin doğru bir şekilde negatif olarak tahmin ettiği örnekleri,

Yanlış Negatif (YN): Modelin yanlış bir şekilde negatif olarak tahmin ettiği örnekleri temsil etmektedir.

Doğruluk ölçütü, doğru olarak sınıflandırılmış örnek sayısının tüm örnek sayısına oranı olarak ifade edilir. Doğruluk ölçütü ile performans ölçülürken veri setinin dengeli dağılması önemli bir kriterdir. Veri setinde bir sınıf diğer sınıflara göre baskınsa, doğruluk yanıltıcı olabilir. Çoğunluk sınıfına ait tahminler yüksek doğruluk verebilir, ancak azınlık sınıfındaki hataları gizleyebilir. Doğruluk ölçütü,

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (25)$$

biçiminde hesaplanır. Kesinlik ölçütü, pozitif olarak sınıflandırılmış tüm örnekler arasında gerçekten kaç tanesinin doğru sınıflandırıldığını veren ölçüttür. Veri setinde sınıflar arasında büyük bir dengesizlik varsa, sadece kesinlik ölçütüne bakarak model performansını değerlendirmek yanıltıcı olabilir. Kesinlik, yanlış pozitiflere odaklandığı için modelin yanlış negatif performansı hakkında bilgi vermez. Kesinlik ölçütü,

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (26)$$

biçiminde hesaplanır. Duyarlılık ölçütü, modelin pozitif sınıfa ait olan gerçek örneklerden ne kadarını doğru bir şekilde tanıdığını gösteren ölçüttür; yani, toplam gerçek pozitifler

arasında doğru tahmin edilenlerin oranını ifade eder. Duyarlılık, özellikle pozitif sınıfın az olduğu dengesiz veri setlerinde önemli olabilir, ancak tek başına kullanıldığında modelin yanlış pozitif oranını göz ardı eder. Duyarlılık ölçütü,

$$Duyarlılık = \frac{DP}{DP + YN} \quad (27)$$

biçiminde hesaplanır. F1 skoru, bir sınıflandırma modelinin kesinlik ve duyarlılık ölçütleri arasında bir denge kurarak, modelin genel performansını ölçen bir metrik olarak tanımlanır. F1 skoru, özellikle dengesiz veri setlerinde, modelin pozitif sınıfı ne kadar iyi tanıdığını ve aynı zamanda bu tahminlerin ne kadar doğru olduğunu değerlendirmede önemli bir rol oynar. F1 Skoru,

$$F1 - Skoru = \frac{2 \times Duyarlilik \times Kesinlik}{Duyarlilik + Kesinlik} \quad (28)$$

biçiminde hesaplanır.

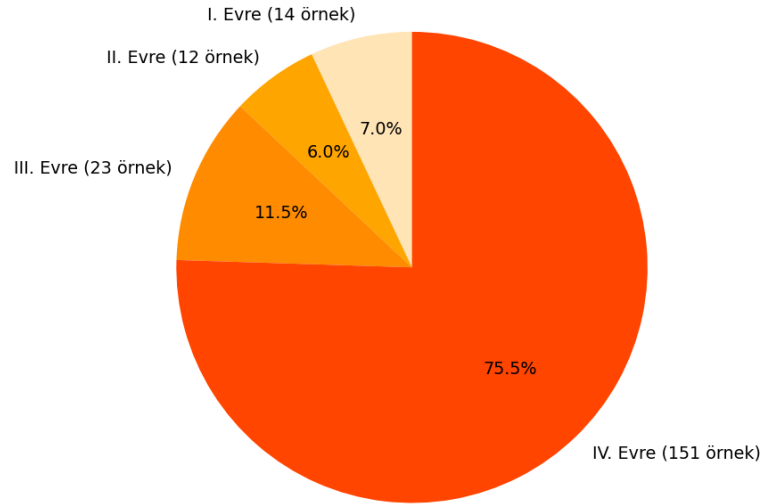
3. YAPILAN ÇALIŞMALAR

3.1. Veri Seti

Bu çalışmada kullanılan nefes örnekleri, Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi'nde tedavi gören akciğer kanseri bireylerden alınmıştır. Veriler, danışman hocam Sayın Prof. Dr. Ayten ATASOY'un yürütücülüğünü yaptığı Öğr. Gör. Dr. Ümit ÖZSANDIKCIOĞLU'nun doktora çalışmaları kapsamında, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenen "Taşınabilir Elektronik Burun ile Akciğer Kanseri Hastalığının Erken Tespiti" başlıklı, TÜBİTAK 3001 Projesi (2016–2020) kapsamında toplanmıştır (Özsandıkçıoğlu, 2023). Bu yüksek lisans çalışmasında ise farklı bir sınıflandırma amacıyla yeniden analiz edilmiştir. İlgili proje çalışmasında kapsamında tüm etik onaylar alınmış (Karadeniz Teknik Üniversitesi Etik Kurulu, Etik Onay No: 24237859-517), gönüllü bireylerin katılımı tamamen rıza esasına dayanmıştır. Tüm akciğer kanseri hastalarının tanısı BT, PET, biyopsi veya histoloji ile doğrulanmıştır. Akciğer kanseri hastaların tümörlerinin hangi evrede oldukları yukarıda da bahsedilen TNM yöntemi ile belirlenmiştir.

Hastalardan nefes örneği alınmadan önceki birkaç saat boyunca, katılımcıların yemek yememesine, su dışında herhangi bir içecek tüketmemesine, sigara içmemesine, ağız bakım suyu kullanmamasına ve ilaç almamış olmasına dikkat edilmiştir. Hastaların akciğer kanseri dışında herhangi bir akciğer hastalığına sahip değildir ve kısa bir süre önce cerrahi bir müdahale geçirmemiştir. Bütün nefes örnekleri kapalı bir odada ve oda sıcaklığında toplanmıştır. Deneylere katılan bireylerin nefes örnekleri Tedlar torbalar ile toplanmıştır. Her bir Tedlar torba yalnızca bir kez kullanılmıştır. Bu torbalar nefes örneklerini etkili bir şekilde koruyabilmesi ve kimyasal bileşiklerin sızıntısını önleyebilmesi nedeniyle tercih edilmiştir. Bu sayede, örneklerdeki uçucu bileşenlerin güvenilir bir şekilde analizi sağlanır.

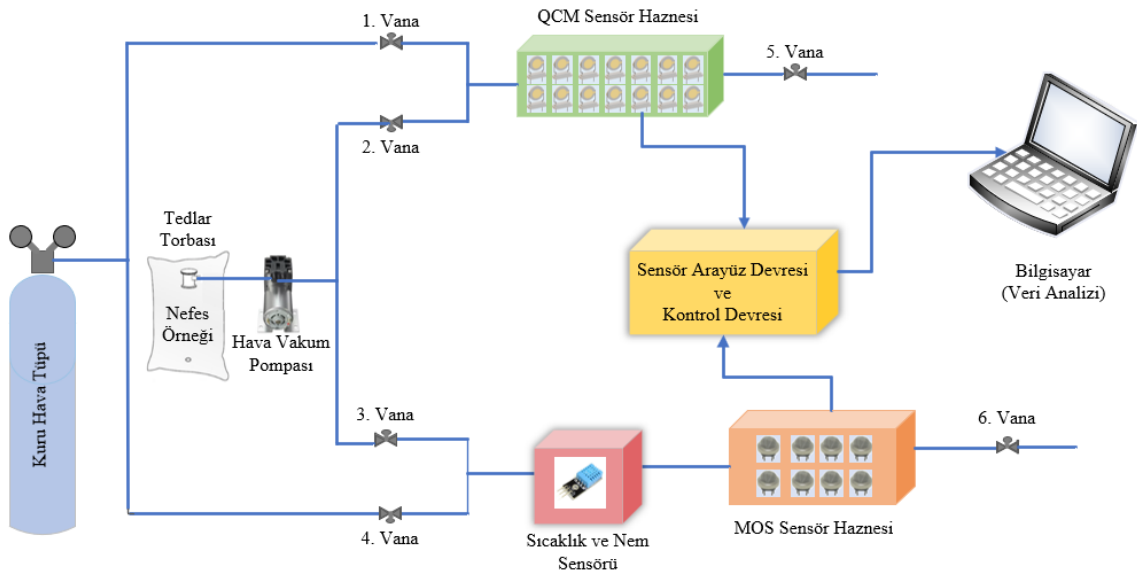
Toplamda 200 adet deney gerçekleştirilmiştir. Çalışmaya katılan akciğer kanseri bireylerin hastalık evrelerine göre dağılımları Şekil 21'de sunulmuştur.



Şekil 21. Akciğer kanseri bireylerin evrelere göre dağılımı

3.2. Elektronik Burun Deney Düzeneği ve Deney Aşaması

Bu çalışma kapsamında kullanılan elektronik burun sistemi, akciğer kanseri hastalarından alınan nefes örneklerinin analizinde görev almıştır. Şekil 22 ve Şekil 23'te sırasıyla elektronik burun sistemine ait şematik diyagram ve sistemin gerçek donanımsal görünümü sunulmuştur.



Şekil 22. Elektronik burun sisteminin genel yapısını gösteren blok diyagram



Şekil 23. Elektronik burun sisteminin laboratuvar ortamındaki görünümü

Elektronik burun sisteminin yapısal bileşenlerini ve bu bileşenlerin işlevlerini daha ayrıntılı bir biçimde ortaya koymak amacıyla, Şekil 22’de sunulan şematik blok diyagramda yer alan donanım birimleri, Tablo 6’da görev ve işlevleriyle birlikte özetlenmiştir.

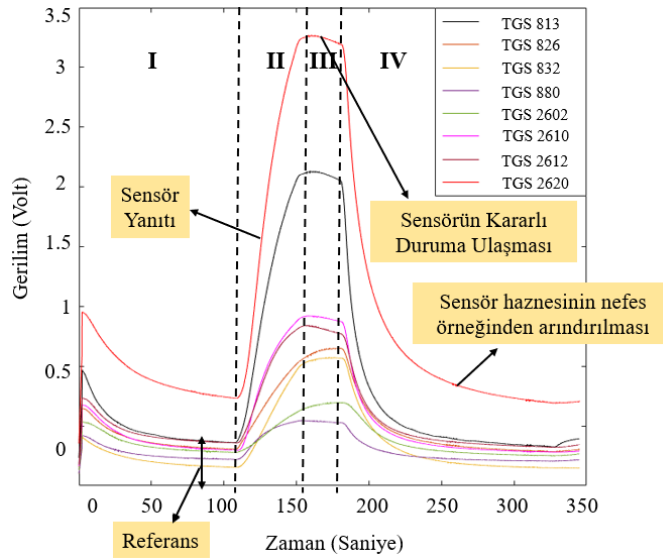
Tablo 6. Elektronik burun sisteminin temel bileşenleri ve görevleri

Bileşen	Görevi / Açıklama
Kuru Hava Tüpü	Sensörlerin kararlı çalışabilmesi için nemden arındırılmış referans gazı sağlar.
Tedlar Torbası	Nefes örneklerinin güvenli bir şekilde depolandığı ve sisteme aktarıldığı gaz torbasıdır.
Vakum Pompası	Gaz akışını sağlamak ve örnekleri sensör hücrelerine iletmek için gerekli basıncı oluşturur.
Valfler (Valf 1–6)	Gaz akış yönünü kontrol eder; örnek ya da referans gazını yönlendirir.
Sıcaklık ve Nem Sensörü	Ortam koşullarını izleyerek ölçümlerde çevresel etkileri kontrol eder.

Tablo 6'nın devamı

QCM Sensör Hücreleri	Kuartz Kristal Mikroterazi sensörleri frekans değişimi ile gazları algılar.
MOS Sensör Hücresi	Metal Oksit Yarıiletken sensörler direnç değişimi ile uçucu bileşikleri algılar.
Sensör Arayüz Devresi ve Veri Elde Etme Kartı	Verileri toplar, sayısallaştırır ve bilgisayara aktarır.
Bilgisayar	Veri işleme, analiz ve sınıflandırma işlemlerinin yapıldığı birimdir.

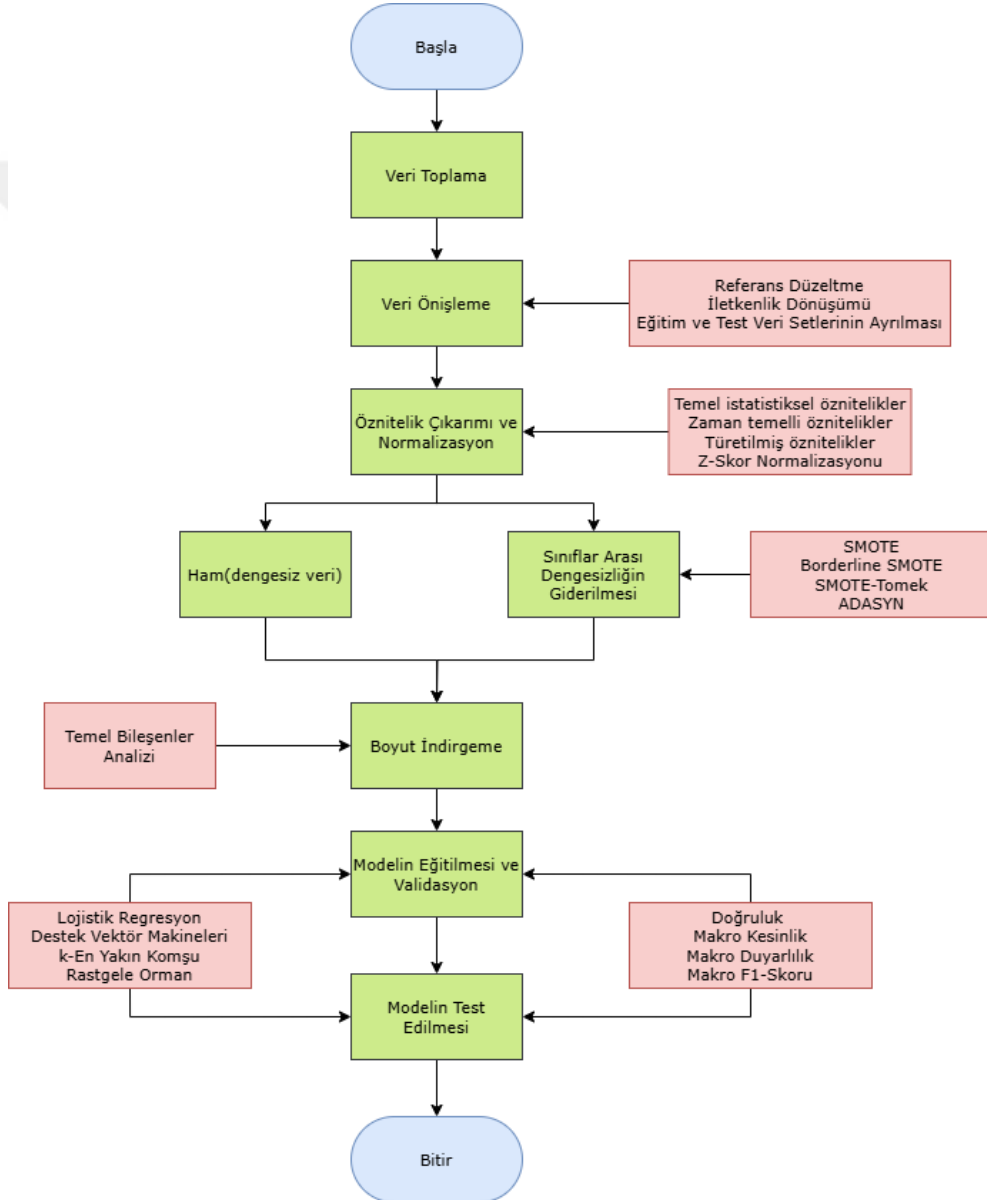
Geliştirilen elektronik burun devresiyle gerçekleştirilen deneyler dört temel aşamada yürütülmektedir. İlk aşamada, sistem 100 saniye boyunca kuru hava ile temizlenmekte, ardından tüm vanalar kapatılmaktadır. Temizleme işleminin ardından, vakum pompası aracılığıyla Tedlar torbasındaki nefes örneği 40 saniye süresince MOS ve QCM sensör haznelerine iletilmektedir. Bu sürecin sonunda vanalar kapatılarak, nefes örneklerinin 30 saniye boyunca sensör haznelerinde hapsedilmesi sağlanmakta ve sensörlerin tepki vermesi beklenmektedir. Son aşamada ise sistem, 140 saniye boyunca yeniden kuru hava ile temizlenerek bir sonraki deneye hazır duruma getirilir. Deney sürecinde MOS sensörlerin zamana bağlı gerilim tepkileri ve dört aşamalı deney döngüsü Şekil 24'te ayrıntılı olarak gösterilmiştir.



Şekil 24. Deney aşamaları

3.3. Çalışma Sürecinin Genel Akış Diyagramı

Bu bölümde, çalışmada uygulanan veri işleme ve modelleme sürecine ait genel iş akışı sunulmaktadır. Şekil 25, veri toplama aşamasından başlayarak sırasıyla veri ön işleme, özellik çıkarımı, sınıf dengesizliği giderme, boyut indirgeme ve model oluşturma gibi adımları kapsamaktadır. Her bir adım, tezde ilgili alt başlıklarda ayrıntılı biçimde ele alınmıştır.



Şekil 25. Çalışma sürecine ait iş akış şeması

3.4. Veri Önışleme

Makine öğrenmesi modellerinin etkin bir şekilde eğitilebilmesi için, ham sensör verilerinin belirli önışleme adımlarından geçirilmesi gerekmektedir. Bu çalışmada elektronik burun sistemiyle elde edilen sensör verileri üzerinde aşağıdaki önışleme adımları uygulanmıştır:

3.4.1. Referans Düzeltme

Elektronik burun sistemlerinde kullanılan sensörler, çevresel koşullara (örneğin sıcaklık, nem, basınç) ve sensör yaşlanmasına bağlı olarak zamanla sapma (drift) gösterebilir. Bu durum, ölçümler arasında tutarsızlıklara ve modelleme sürecinde güvenilirliğin azalmasına yol açar. Bu nedenle, bu çalışmada sensör yanıtlarının güvenilirliğini ve karşılaştırılabilirliğini artırmak amacıyla referans düzeltme (baseline correction) işlemi uygulanmıştır. Bu işlem sırasında, her sensör kanalından alınan sinyallerin başlangıç (referans) değeri çıkarılarak sinyaller ortak bir başlangıç noktasına göre normalize edilmiştir. Böylece hem çevresel etkiler hem de sensörler arası farklılıkların etkisi minimize edilmiştir. Kullanılan düzeltme yöntemi, sensör tipine göre aşağıdaki şekilde uygulanmıştır:

- MOS sensörleri direnç temelli çalıştıkları için, genellikle gerilim bölücü devre ile ölçülen gerilim değeri kullanılmış ve referans düzeltme işlemi şu şekilde tanımlanmıştır:

$$v_{n,s,r}(t) = v_{n,s}(t) - v_{n,s}(0) \quad (29)$$

- QCM sensörleri ise rezonans frekansındaki değişim prensibiyle çalıştığından, ölçülen frekans sinyalleri üzerinde referans düzeltme şu şekilde gerçekleştirilmiştir:

$$f_{n,s,r}(t) = f_{n,s}(t) - f_{n,s}(0) \quad (30)$$

Burada $v_{n,s}(t)$ ve $f_{n,s}(t)$ n-inci örnek ve s-inci sensör için t anındaki gerilim ya da frekans değerini, $v_{n,s}(0)$ ve $f_{n,s}(0)$ aynı sensörlerin t=0 anındaki referans değerlerini, $v_{n,s,r}(t)$ ve $f_{n,s,r}(t)$ referans düzeltilmiş sinyalleri ifade eder. Bu yöntem, literatürde fark

referans düzeltme (differential baseline correction) olarak geçmektedir ve çoklu sensör sistemlerinde karşılaştırmalı analizler yapılmasını mümkün kılmaktadır.

3.4.2. İletkenlik Dönüşümü

MOS sensörler, gazlarla etkileşime girdiklerinde iç dirençlerini değiştirirler. Bu direnç değişimi, genellikle bir gerilim bölücü devre aracılığıyla gerilim (voltaj) olarak ölçülür. Ancak bu gerilim değerleri:

- Doğrudan fiziksel anlam taşımaz,
- Ortam koşullarına duyarlıdır,
- Sensörden sensöre değişkenlik gösterebilir.

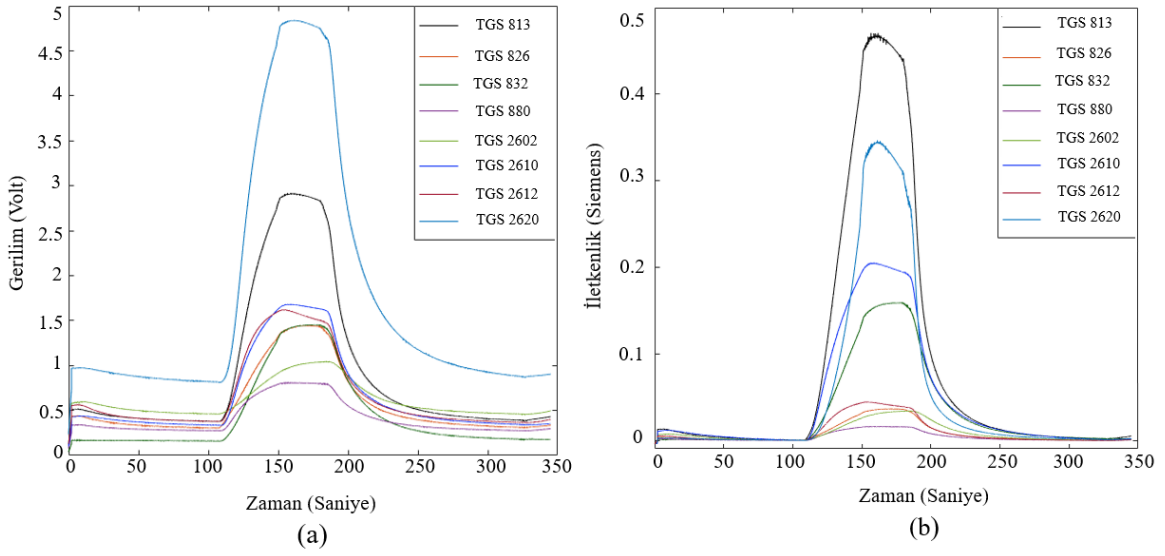
Bu nedenle, bu gerilimler aşağıdaki formül ile iletkenlik (G) değerine dönüştürülür:

$$G_{n,s}^R(t) = \frac{v_{n,s}^R(t)}{(v_c - v_{n,s}^R(t)) \times R_{L,s}} \quad (31)$$

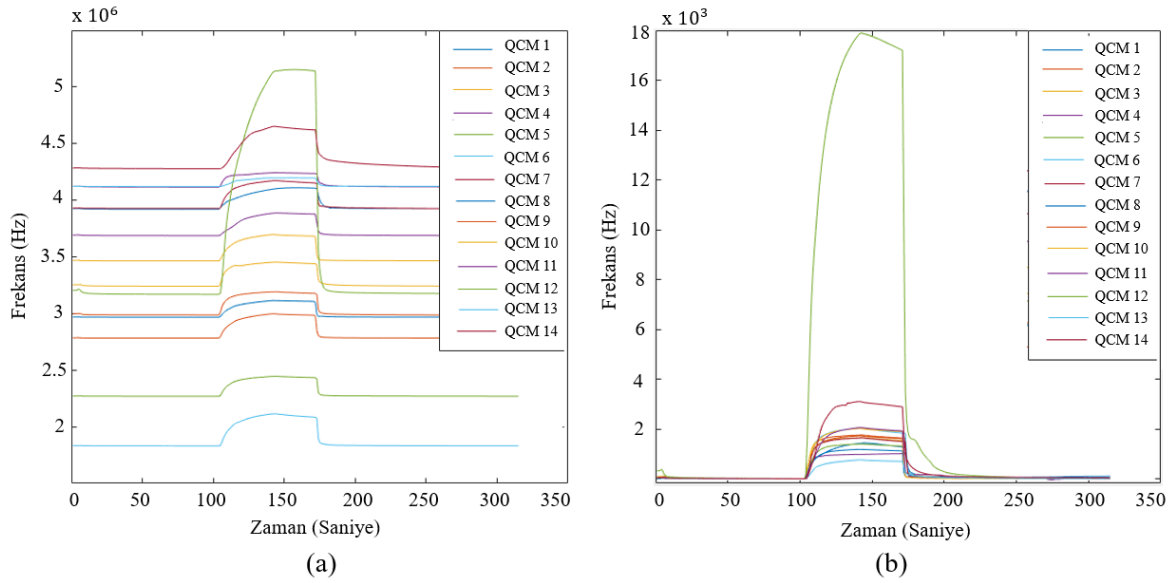
Bu denklemde $v_{n,s}^R(t)$ n-inci örnek ve s-inci sensör için t anında referans düzeltilmiş gerilim değerini, v_c besleme gerilimini, $R_{L,s}$ s-inci sensörün yük direncini, $G_{n,s}^R(t)$ ise n-inci örnek ve s-inci sensör için t anında referans düzeltilmiş veriye karşılık gelen iletkenlik değerini temsil eder.

Öte yandan, QCM sensörler, gaz moleküllerinin sensör yüzeyine tutunmasıyla rezonans frekansında değişim üretir. QCM sensörlerin çıkışı doğrudan fiziksel bir parametre olan frekans olduğundan, ek bir dönüşüme ihtiyaç duyulmadan doğrudan analiz için kullanılabilir.

MOS ve QCM sensörlerine ait ham veriler ile bu verilere uygulanan ön işleme işlemleri sonucunda elde edilen örnek sinyaller, sırasıyla Şekil 26 ve Şekil 27’de gösterilmektedir.



Şekil 26. (a) Ham MOS sensör verisi ve (b) önışlem uygulanmış MOS sensör verisi



Şekil 27. (a) Ham QCM sensör verisi ve (b) önışlem uygulanmış QCM sensör verisi

3.5. Öznitelik Çıkarımı

Elektronik burun sistemleri, nefes örneklerinde bulunan VOC duyarlı sensörler aracılığıyla sinyaller üreterek, akciğer kanseri evrelerinin tespitine yönelik anlamlı veri sağlar. Öznitelik çıkarımı, bu ham sensör verilerinin istatistiksel olarak anlamlı ve

sınıflandırılabilir hale getirilmesini sağlayarak, farklı kanser evrelerinin doğru şekilde ayrıştırılmasına olanak tanır. Bu aşama, modelleme ve sınıflandırma algoritmalarının başarısında belirleyici bir rol oynamaktadır.

Verilerden öznitelik çıkarımı yapılırken, deney süreci bütüncül olarak ele alınmış; ayrıca, nefes örneğinin iletilmesi, kararlı durumda tutulması (hapsedilmesi) ve temizlenmesi aşamaları ayrı ayrı değerlendirilerek analiz edilmiştir. Böylece her bir fazdan elde edilen sinyal karakteristikleri üzerinden daha zengin ve anlamlı öznitelikler oluşturulmuştur. Bu kapsamda MOS ve QCM sensörleri için ayrı ayrı hem temel istatistiksel öznitelikler hem de zaman temelli türetilmiş öznitelikler elde edilerek sınıflandırma modellerinde kullanılmak üzere veri kümesi oluşturulmuştur.

3.5.1. Temel İstatistiksel Öznitelikler

Her bir sensör için aşağıdaki temel istatistiksel öznitelikler çıkarılmıştır.

- 1) Maksimum değer: Bir sensörün tüm ölçüm değerleri arasındaki en yüksek değeri temsil eder.

$$\text{Maksimum}(s) = \max\{s_1, s_2, \dots, s_n\} \quad (32)$$

- 2) Medyan: Ölçüm değerleri küçükten büyüğe sıralandığında ortada kalan değeri ifade eder.

$$\text{Medyan}(s) = \begin{cases} s_{(\frac{n+1}{2})} & , n = \text{tek} \\ \frac{s_{(\frac{n}{2})} + s_{(\frac{n}{2}+1)}}{2} & , n = \text{çift} \end{cases} \quad (33)$$

- 3) Ortalama: Tüm ölçüm değerlerinin toplamının gözlem sayısına bölünmesiyle elde edilen değerdir.

$$\text{Ortalama}(s) = \frac{1}{n} \sum_{i=1}^n s_i \quad (34)$$

- 4) Standart Sapma: Verilerin ortalama etrafında ne kadar dağıldığını gösteren bir ölçüdür; yüksek standart sapma, değerlerin ortalamadan uzaklaştığını, düşük standart sapma ise ortalamaya yakın olduğunu ifade eder.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (35)$$

- 5) Skewness (Çarpıklık): Veri dağılımının simetrik olup olmadığını ve hangi yöne yatkın olduğunu gösterir.

$$Skewness = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 \quad (36)$$

- 6) Kurtosis (Basıklık): Veri dağılımının sivrilğini veya düzlüğünü, uç değer yoğunluğunu gösterir.

$$Kurtosis = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 \quad (37)$$

Bu öznitelikler, sensör tepkisinin genel davranışını nicel olarak ifade etmekte ve sınıflandırma performansına katkı sağlamaktadır.

3.5.2. Zaman Temelli Öznitelikler

Sensör sinyallerinin zaman içinde gösterdiği değişim ve eğilimlerin değerlendirilmesi amacıyla aşağıdaki zaman temelli öznitelikler çıkarılmıştır:

- 1) Zaman Aralıklarına Göre Eğim: Her bir sensör için, nefesin iletilmesi, hapsedilmesi ve temizlenmesi aşamalarına karşılık gelen zaman aralıklarında sensör tepkisinin eğim değeri hesaplanmıştır.

$$Eğim_{[t_1, t_2]} = \frac{s(t_2) - s(t_1)}{t_2 - t_1} \quad (38)$$

- 2) Zaman Ekseni ile Sensör Yanıtı Arasındaki Alan: Her bir sensörün zamana bağlı tepkisi ile x ekseni (zaman ekseni) arasında kalan alan tek seferde hesaplanmıştır. Bu öznitelik sinyalin zamana yayılmış toplam kuvvetini (integralini) temsil eder.

$$Alan_x = \sum_{i=1}^{n-1} \frac{(x_i - x_{i+1})}{2} \times (t_{i+1} - t_i) \quad (39)$$

3.5.3. Türetilmiş Öznitelikler

- 1) İki Sensör Arasında Kalan Alanlar: İki sensörün aynı fazdaki (iletme, hapsetme, temizleme) ölçümleri arasındaki farkın oluşturduğu alan, her bir faz için ayrı ayrı hesaplanmıştır. Bu öznitelik, sensör yanıtlarının birbiriyle olan benzerlik veya ayrışma düzeyini ortaya koyar.

$$Alan_{i,j}^{(f)} = \sum_{k=1}^{n_f-1} \left| \frac{(x_{i,k} - x_{j,k}) + (x_{i,k+1} - x_{j,k+1})}{2} \right| \times (t_{k+1} - t_k) \quad (40)$$

MOS ve QCM sensörlerinden elde edilen öznitelikler ve her bir öznitelik için karşılık gelen boyutlar Tablo 7’de verilmiştir.

Tablo 7. MOS ve QCM sensör verilerine ait öznitelikler ve boyutları

Öznitelik	Boyut (MOS)	Boyut (QCM)
Maksimum değer (her sensör)	200×8	200×14
Ortalama	200×8	200×14
Medyan	200×8	200×14
Standart Sapma	200×8	200×14
Skewness	200×8	200×14
Kurtosis	200×8	200×14
Eğim (nefes iletimi)	200×8	200×14
Eğim (kararlı durum)	200×8	200×14
Eğim (temizlenme)	200×8	200×14
Sensörler ile zaman eksenini arasında kalan alan	200×8	200×14
Sensörler arasında kalan alan (nefes iletimi)	200×28	200×91
Sensörler arasında kalan alan (kararlı durum)	200×28	200×91
Sensörler arasında kalan alan (temizlenme)	200×28	200×91
Toplam Boyut (Öznitelik Matrisi)	200x164	200x413

Bu çalışmada kullanılan öznitelik çıkarım süreci sonucunda, her biri 200 nefes örneğinden oluşan veri kümesi için; MOS sensörlerinden 164, QCM sensörlerinden ise 413 olmak üzere toplamda oldukça yüksek boyutlu öznitelik matrisleri elde edilmiştir. Bu durum, sınırlı sayıda örnekle temsil edilen veri kümesinde, yüksek boyutlu özellik uzayının yol açtığı boyutsallık laneti (curse of dimensionality) problemiyle karşı karşıya kalınmasına neden olmuştur. Özellikle örnek sayısının öznitelik sayısına oranla düşük olduğu bu gibi senaryolarda, modelin genelleme yeteneği zayıflayabilir ve aşırı öğrenme riski artabilir. Bu nedenle, hem hesaplama yükünü azaltmak hem de sınıflandırma başarımını iyileştirmek amacıyla, elde edilen özniteliklerin daha anlamlı alt kümelerine indirgenmesi için boyut indirgeme tekniklerine başvurulması gerekli hale gelmektedir. (Bknz. 3.8)

3.6. Z-Skor Normalizasyonu

Öznitelik çıkarımı sonrasında elde edilen özellikler farklı ölçek ve dağılımlara sahip olabileceğinden, sınıflandırma algoritmalarının daha sağlıklı öğrenebilmesi amacıyla normalizasyon işlemi uygulanmıştır. Bu çalışmada min-max normalizasyon yerine z-score (standart skor) yöntemi tercih edilmiştir. Bunun temel nedeni, min-max normalizasyonun verinin minimum ve maksimum değerlerine duyarlı olması ve aykırı değerlerin (outlier) etkisiyle ölçeklendirmeyi bozma riskidir. Z-score yöntemi ise öznitelikleri ortalama ve standart sapmaya göre dönüştürdüğü için daha sağlam ve istatistiksel olarak daha kararlı bir normalizasyon sağlar.

Ayrıca, normalizasyon işlemi doğrudan ham sensör verilerine değil, öznitelik çıkarımı sonrasında uygulanmıştır. Bunun nedeni, zaman serisi halindeki sensör sinyallerinin yapısal özelliklerinin (eğim, varyans, alan gibi) doğrudan normalizasyonla bozulma ihtimalidir. Önce öznitelikler çıkarılarak her sensör sinyali matematiksel olarak özetlenmiş, ardından bu öznitelik vektörlerine normalizasyon uygulanarak veri anlamı korunmuş ve modelin dengeli öğrenmesi sağlanmıştır.

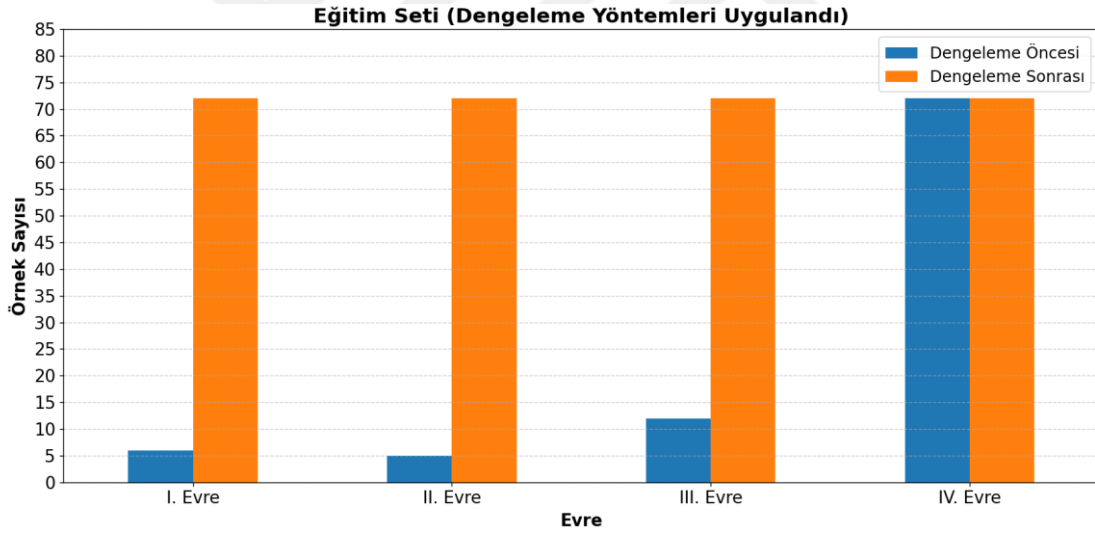
3.7. Sınıf Dengesizliğinin Giderilmesi

Bu çalışmada, sınıflandırma algoritmalarının daha dengeli ve başarılı bir şekilde öğrenebilmesi amacıyla çeşitli veri artırma (örnekleme) yöntemleri uygulanmıştır. Özellikle veri setinde sınıflar arasında gözlenen dengesizlik nedeniyle, azınlık sınıfa ait örnek sayısını artırmak ve modelin bu sınıfı daha iyi tanıyabilmesini sağlamak hedeflenmiştir. Çalışma

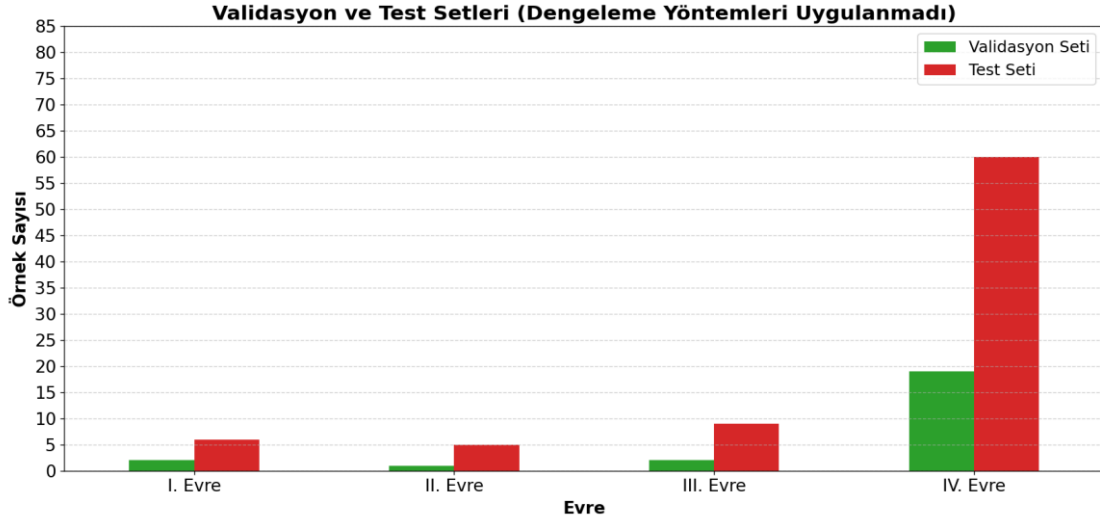
kapsamında SMOTE, Borderline SMOTE, SMOTE-Tomek ve ADASYN yöntemleri uygulanmıştır.

Uygulanan bu yöntemlerin her biri yalnızca eğitim verisi üzerinde uygulanmış; validasyon ve test verilerine kesinlikle müdahale edilmemiştir. Bu strateji, hem modelin gerçek veri üzerindeki performansını yapay olarak yükseltmemek hem de veri sızıntısını (data leakage) önlemek amacıyla tercih edilmiştir. Eğer veri artırma tüm veri setine uygulanmış olsaydı, model validasyon veya test aşamalarında eğitim sürecinde üretilmiş sentetik verilere maruz kalabilir, bu da modelin performansını olduğundan daha yüksek gösterebilirdi.

Deneysel süreçte her veri artırma yöntemi ayrı ayrı uygulanmış ve tüm sınıflandırma algoritmaları, bu yöntemlerle elde edilen dengeli eğitim veri setleri üzerinde eğitilmiştir. Veri dengeleme öncesi ve sonrası eğitim, validasyon ve test seti örnek dağılımlar Şekil 28 ve Şekil 29’da verilmiştir.

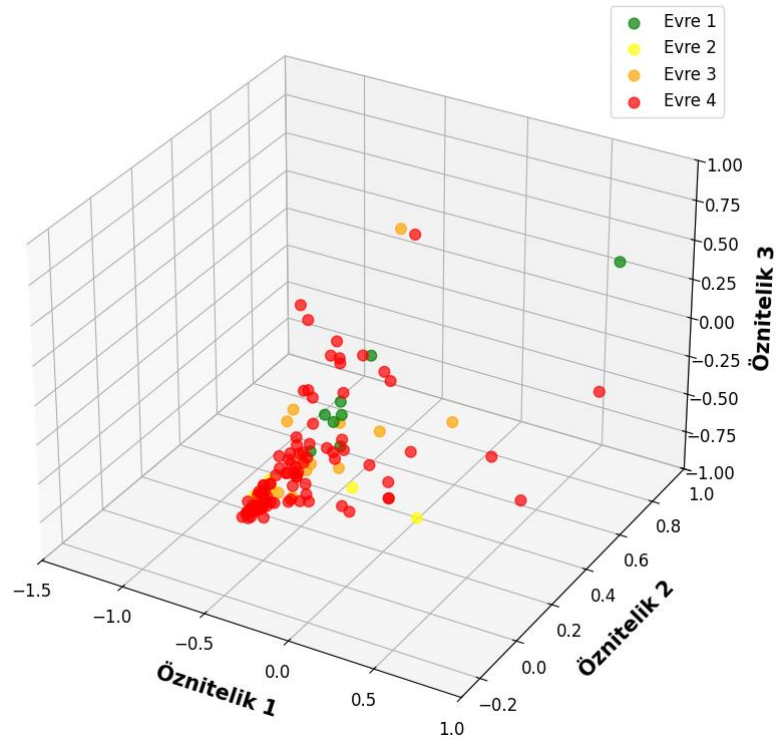


Şekil 28. Dengeleme öncesi ve sonrasında eğitim seti sınıf dağılımı

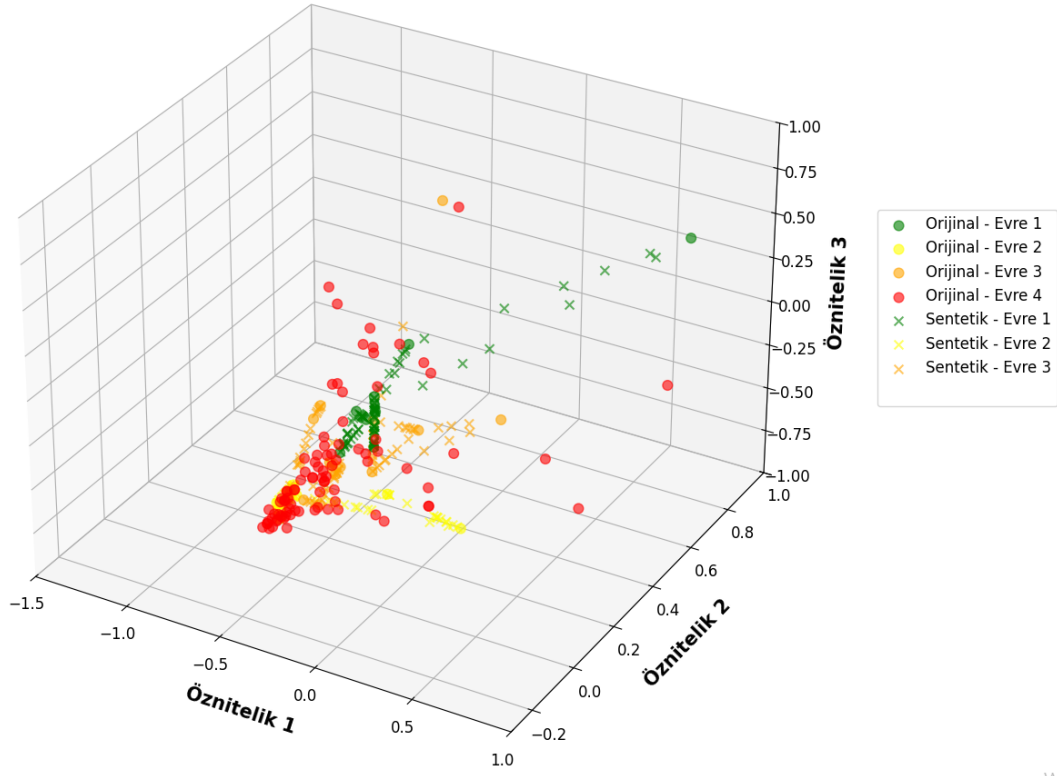


Şekil 29. Dengeleme öncesi ve sonrasında validasyon ve test setleri sınıf dağılımı

Dengesiz MOS sensör verilerinin ilk üç öz niteliği kullanılarak oluşturulan üç boyutlu dağılım, Şekil 30'da sunulmuştur. Aynı veri kümesine uygulanan SMOTE yöntemi sonrasında, üretilen sentetik örneklerle birlikte hem orijinal hem de dengelenmiş verilerin görsel temsili ise Şekil 31'de gösterilmiştir.



Şekil 30. MOS sensör verilerine ait dengesiz eğitim kümesinin ilk üç öz niteliğe göre üç boyutlu dağılımı



Şekil 31. MOS sensör verilerine SMOTE uygulanması sonrası elde edilen sentetik ve orijinal örneklerin üç boyutlu dağılımı (ilk üç öznitelik)

3.8. Temel Bileşenler Analizi ile Boyut İndirgeme

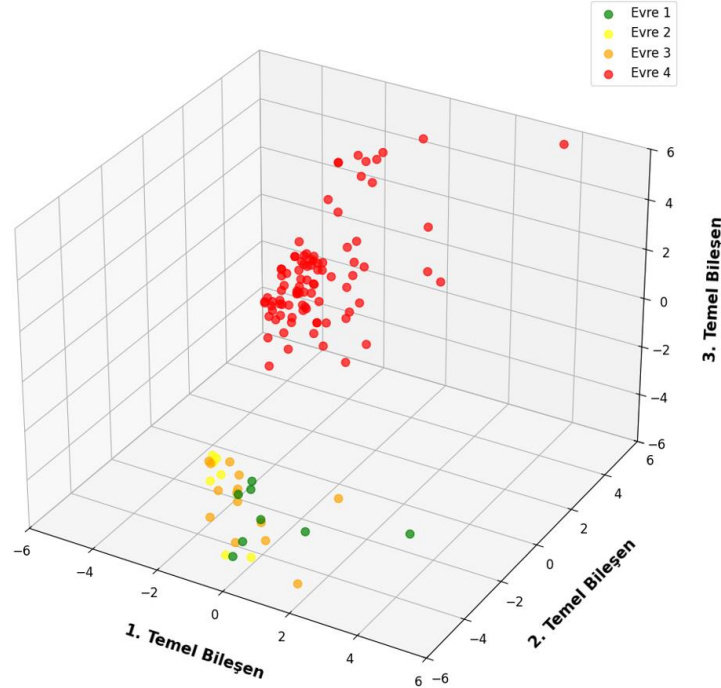
Çalışmada, öznitelik çıkarımı sonrası elde edilen yüksek boyutlu veri setinde bilgi tekrarını ve boyutsal yükü azaltmak amacıyla PCA yöntemi ile boyut indirgeme işlemi gerçekleştirilmiştir. PCA, verideki toplam varyansı koruyarak daha az sayıda yeni bileşen üzerinden verinin temsil edilmesini sağlar ve bu sayede hem işlem süresi kısaltılır hem de modelin aşırı öğrenme riski azaltılır. Model eğitimi öncesinde, sınıflar arası dengesizlikleri gidermek amacıyla SMOTE, Borderline SMOTE, SMOTE-Tomek ve ADASYN gibi farklı örnekleme yöntemleri denenmiştir. Tüm bu yöntemlerde ortak olarak, veri artırma işlemi PCA'dan önce gerçekleştirilmiştir. Bunun iki temel nedeni vardır:

- 1) PCA veriyi yeni ve doğrusal bileşenlerden oluşan farklı bir uzaya dönüştürür. Eğer veri dengeleme işlemleri bu dönüşümden sonra uygulanırsa, sentetik örnekler artık orijinal anlam taşıyan öznitelikler üzerinden değil, soyut ve sıkıştırılmış bileşenler üzerinden oluşturulur. Bu durum, özellikle azınlık sınıf için yapay ama anlamsal olarak zayıf örneklerin üretilmesine neden olabilir. Bu nedenle, veri dengeleme

algoritmalarının orijinal öznitelik uzayında çalışması, daha tutarlı örnekler üretmesi açısından önemlidir.

- 2) PCA varyansa dayalı bir yöntem olduğu için, sınıf dengesizliği olan veri setlerinde ağırlıklı olarak çoğunluk sınıfın yapısına göre bileşenler üretir. Bu da azınlık sınıfın PCA bileşenleri içinde yeterince temsil edilememesine ve bilgi kaybına yol açabilir. Dengeleme işleminin daha önce yapılması, PCA'nın tüm sınıfları daha adil şekilde temsil etmesini sağlar.

Boyut indirgeme sürecinde uygulanan PCA ile, veri setinde yer alan toplam bilginin (varyansın) en az %95'ini açıklayan bileşenler seçilmiştir. Bu yaklaşım, bilgi kaybını en aza indirerek sınıflandırma modellerinin doğruluğunu korumayı hedeflemiş ve aynı zamanda modelin eğitim süresini ve hesaplama maliyetini azaltmada etkili olmuştur. MOS verileri için PCA uygulanmadan önce veri seti 164 öznitelikten oluşmaktayken, PCA sonrası %95 oranında birikimli varyansı koruyacak şekilde 11 bileşen seçilmiştir. Benzer şekilde, QCM verileri için PCA öncesinde 413 öznitelik bulunmakta olup, uygulanan boyut indirgeme işlemi sonucunda 34 bileşen ile temsil edilmiştir. Bu dönüşümün görselleştirilmesine yönelik olarak, dengesiz MOS verilerine ait ilk üç PCA bileşenine göre oluşturulan dağılım grafiği Şekil 32'de sunulmuştur.



Şekil 32. Dengesiz MOS sensör verisine uygulanan PCA ile elde edilen ilk üç bileşen düzleminde sınıf (evre) dağılımı

3.9. Sınıflandırma Modelleri ve Hiperparametre Ayarları

Bu çalışmada, öznetelik çıkarımı ve boyut indirgeme işlemleri sonrasında elde edilen veri seti üzerinde çeşitli denetimli sınıflandırma algoritmaları kullanılmıştır. Amaç, farklı algoritmaların ve hiperparametre ayarlarının sınıflandırma başarımı üzerindeki etkisini karşılaştırmalı olarak incelemek ve en yüksek performansı veren yapılandırmaları belirlemektir.

Kullanılan sınıflandırma algoritmaları; Lojistik Regresyon (LR), k-En Yakın Komşu (k-NN), Destek Vektör Makineleri (SVM) ve Rastgele Orman (RF) modelleridir. Bu algoritmaların tercih edilme gerekçeleri aşağıda belirtilmiştir:

Lojistik Regresyon (LR): Doğrusal sınıflandırma problemleri için sıkça tercih edilen temel bir yöntemdir. Özellikle düşük boyutlu, iyi ölçeklendirilmiş veri setlerinde etkili çalışır. Modelin çıktıları doğrudan olasılıklar şeklinde yorumlanabildiği için açıklanabilirlik açısından avantajlıdır.

Destek Vektör Makineleri (SVM): Doğrusal ve doğrusal olmayan sınıflandırma görevleri için uygun, esnek bir modeldir. Farklı çekirdek fonksiyonları (örneğin RBF, polinom, lineer) sayesinde karmaşık veri dağılımlarına uyum sağlayabilir. Özellikle sınırlı sayıda örnekle çalışıldığında bile yüksek başarı gösterebilir.

k-En Yakın Komşu (k-NN): Öğrenme aşaması olmayan, örnek tabanlı bir sınıflandırma algoritmasıdır. Bir örneğin sınıfı, en yakın komşularının sınıflarına göre belirlenir. Boyutu düşük ve benzerlik ölçümlerine duyarlı veri setlerinde yüksek doğruluk sağlayabilir.

Rastgele Orman (RF): Birden çok karar ağacının birleşiminden oluşan bir topluluk öğrenme yöntemidir. Her ağaç, farklı veri alt kümeleriyle eğitilir ve bu çeşitlilik modelin genelleme kabiliyetini artırır. Hem lineer hem de karmaşık ilişkileri modelleme yeteneğine sahiptir.

Her algoritma için farklı hiperparametre kombinasyonları denenmiş ve bu kombinasyonların performansları hem çapraz doğrulama hem de test veri seti üzerinde değerlendirilmiştir. Kullanılan hiperparametreler ve denenilen değerler aşağıdaki Tablo 8’de özetlenmiştir.

Tablo 8. Sınıflandırma modelleri ve çalışmada kullanılan hiperparametreler

Sınıflandırma Algoritması	Hiperparametreler
LR	Regülerizasyon parametresi (C): 0.1, 1, 10, 100 Optimizasyon algoritması: linear, saga, lbfgs
k-NN	Komşu Sayısı (k): 1, 3, 5, 7 Mesafe Metrikleri: öklid, manhattan, chebyshev
SVM	Regülerizasyon parametresi (C): 0.1, 1, 10, 100 Kernel Fonksiyonu: rbf, linear, poly
RF	Ağaç sayısı (n): 20, 50, 100 Maksimum derinlik: 5, 10, 20

Her bir hiperparametre kombinasyonu, farklı örnekleme yöntemleriyle birlikte yeniden eğitilmiş ve test verisi üzerinde ayrı ayrı değerlendirilmiştir. Böylece, örnekleme tekniği ve hiperparametre ayarlarının sınıflandırma başarımı üzerindeki etkisi analiz edilmiştir

3.10. Çapraz Doğrulama

Tablo 8’de yer alan her bir hiperparametre kombinasyonu, 5 katlı katmanlı (stratified) çapraz doğrulama yöntemiyle değerlendirilmiştir. Bu işlem öncesinde, veri seti sınıf dağılımı korunarak %60 eğitim ve %40 test olacak şekilde ikiye ayrılmıştır. Nihai olarak eğitilen modeller, ayrı bir test seti üzerinde değerlendirilmiş ve test performans ölçütleri bu veri üzerinden raporlanmıştır. Bu sürecin şematik gösterimi Şekil 31’de sunulmaktadır.

VERİ SETİ					
EĞİTİM VERİSİ					TEST VERİSİ
1. KAT (VALİDASYON)	2. KAT (EĞİTİM)	3. KAT (EĞİTİM)	4. KAT (EĞİTİM)	5. KAT (EĞİTİM)	PERFORMANS DEĞERLENDİRMESİ
1. KAT (EĞİTİM)	2. KAT (VALİDASYON)	3. KAT (EĞİTİM)	4. KAT (EĞİTİM)	5. KAT (EĞİTİM)	
1. KAT (EĞİTİM)	2. KAT (EĞİTİM)	3. KAT (VALİDASYON)	4. KAT (EĞİTİM)	5. KAT (EĞİTİM)	
1. KAT (EĞİTİM)	2. KAT (EĞİTİM)	3. KAT (EĞİTİM)	4. KAT (VALİDASYON)	5. KAT (EĞİTİM)	
1. KAT (EĞİTİM)	2. KAT (EĞİTİM)	3. KAT (EĞİTİM)	4. KAT (EĞİTİM)	5. KAT (VALİDASYON)	

Şekil 33. Çapraz doğrulama süreci

3.11. Performans Metrikleri

Sınıf dengesizliği nedeniyle doğruluk metriği yetersiz kaldığından, değerlendirme için kesinlik, duyarlılık ve F1-skoru gibi sınıf bazlı metrikler kullanılmıştır. Ancak bu metriklerin hesaplanmasında da sınıf örneklem büyüklüklerinden etkilenmemek adına, ağırlıklı ortalama yerine her sınıfa eşit ağırlık veren makro ortalama tercih edilmiştir. Böylece, her sınıfın performansı eşit önemde değerlendirilerek daha dengeli bir analiz yapılması sağlanmıştır. Makro kesinlik, makro duyarlılık ve makro F1 – skoru sırasıyla;

$$Makro\ Kesinlik = \frac{1}{N} \sum_{i=1}^N Kesinlik_i \quad (42)$$

$$Makro\ Duyarlilik = \frac{1}{N} \sum_{i=1}^N Duyarlilik_i \quad (43)$$

$$Makro\ F1 - Skoru = \frac{1}{N} \sum_{i=1}^N F1 - Skoru_i \quad (44)$$

biçiminde hesaplanır. Burada N sınıf sayısını, $Kesinlik_i$, $Duyarlilik_i$, $F1 - Skoru_i$ i-inci sınıfın kesinlik, duyarlılık ve F1-Skor değerini ifade eder. Elde edilen sonuçlar her bir metrik için ortalama $\pm$ standart sapma biçiminde raporlanmış ve karşılaştırmalı tablolar halinde sunulmuştur. Bu yaklaşım, hem modelin genel başarısını hem de farklı hiperparametre ve örnekleme kombinasyonlarının etkisini ortaya koymak açısından daha güvenilir ve kapsamlı bir analiz imkânı sunmuştur.

4. BULGULAR VE TARTIŞMA

Bu bölümde, elektronik burun sisteminden elde edilen sensör verileri kullanılarak gerçekleştirilen sınıflandırma çalışmalarının deneysel bulgularına ayrıntılı olarak yer verilmektedir. Çalışmada, MOS ve QCM sensörlerinden elde edilen veriler hem ayrı ayrı hem de birlikte değerlendirilmiş; böylece sensör kaynaklı bilgi çeşitliliğinin sınıflandırma performansına etkisi incelenmiştir. MOS+QCM kombinasyonu, hibrit sensör kullanımının potansiyel katkılarını ortaya koymak üzere ayrıca ele alınmıştır.

Veri setinde mevcut olan sınıf dengesizliği, modelin azınlık sınıflarda gösterdiği performansı olumsuz etkileyebileceği için dikkate alınmış ve bu doğrultuda dört farklı sentetik örnekleme yöntemi (SMOTE, Borderline-SMOTE, SMOTE-Tomek ve ADASYN) uygulanmıştır.

Sınıflandırma işlemleri dört farklı denetimli öğrenme algoritması (Lojistik Regresyon, k-En Yakın Komşu, Destek Vektör Makineleri ve Rastgele Orman) ile gerçekleştirilmiş ve her algoritma için çeşitli hiperparametre ayarları sistematik olarak test edilmiştir. Model başarımı; doğruluk, makro kesinlik, makro duyarlılık ve makro F1-skoru gibi performans metrikleri üzerinden hem 5 katlı katmanlı çapraz doğrulama hem de ayrı test veri seti kullanılarak değerlendirilmiştir.

Aşağıda sunulan tablolar (Tablo 9 – Tablo 68), sensör tipine, veri dengeleme yöntemine ve algoritma hiperparametre yapılandırmasına göre elde edilen sınıflandırma başarımlarını karşılaştırmalı olarak özetlemektedir. Performans sonuçları ortalama $\pm$ standart sapma biçiminde raporlanarak, yalnızca ortalama başarımlar değil, aynı zamanda modellerin istikrar düzeyleri hakkında da yorum yapılmasına olanak sağlanmaktadır. Tablo 9'dan Tablo 68'e kadar olan tüm tablolarda kullanılan kısaltmalardan ÇD, Çapraz Doğrulamay; Kesinlik, Makro Kesinliği; Duyarlılık, Makro Duyarlılığı ve F1-Skoru ise Makro F1 - Skoru değerlerini ifade etmektedir.

4.1. Dengesiz MOS Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 9. Dengesiz MOS sensör verileri ile LR performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	85.00 ± 2.04	39.00 ± 5.61	51.11 ± 7.73	42.93 ± 6.32	83.75	34.72	44.44	37.96
(C=0.1, optimizasyon yöntemi=saga)	85.83 ± 3.33	48.67 ± 10.94	60.28 ± 12.47	52.40 ± 11.15	91.25	72.92	73.61	72.22
(C=0.1, optimizasyon yöntemi=lbgfs)	86.67 ± 3.12	49.50 ± 11.77	60.56 ± 12.69	53.05 ± 11.71	91.25	72.92	73.61	72.22
(C=1, optimizasyon yöntemi=liblinear)	85.00 ± 4.25	48.33 ± 11.37	60.00 ± 12.89	51.98 ± 11.73	88.75	69.44	71.81	67.65
(C=1, optimizasyon yöntemi=saga)	85.83 ± 4.25	51.67 ± 15.50	62.50 ± 15.44	53.82 ± 13.56	91.25	75.43	77.36	74.23
(C=1, optimizasyon yöntemi=lbgfs)	86.67 ± 3.12	58.83 ± 10.46	67.50 ± 7.33	59.37 ± 6.37	92.50	78.33	81.53	79.57
(C=10, optimizasyon yöntemi=liblinear)	86.67 ± 3.12	55.83 ± 7.26	66.67 ± 10.54	59.12 ± 8.16	88.75	67.01	71.81	67.35
(C=10, optimizasyon yöntemi=saga)	85.00 ± 3.33	50.83 ± 15.00	60.00 ± 14.17	52.48 ± 12.52	91.25	75.43	77.36	74.23
(C=10, optimizasyon yöntemi=lbgfs)	86.67 ± 5.53	60.92 ± 18.44	70.00 ± 9.77	62.80 ± 14.62	91.25	74.11	78.75	75.77
(C=100, optimizasyon yöntemi=liblinear)	85.83 ± 5.00	63.42 ± 11.18	74.17 ± 3.58	65.05 ± 9.02	91.25	74.11	78.75	75.77
(C=100, optimizasyon yöntemi=saga)	83.33 ± 5.27	49.83 ± 15.83	59.44 ± 17.27	51.26 ± 13.86	91.25	75.43	77.36	74.23
(C=100, optimizasyon yöntemi=lbgfs)	88.33 ± 7.17	66.25 ± 21.26	75.00 ± 15.44	68.71 ± 19.77	91.25	74.11	78.75	75.77

Tablo 10. Dengesiz MOS sensör verileri ile SVM performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	86.67 ± 3.12	52.00 ± 13.91	60.56 ± 15.48	54.77 ± 14.16	91.25	70.08	73.61	70.78
(C=0.1, çekirdek fonksiyonu=poly)	75.00 ± 2.64	18.90 ± 0.46	24.72 ± 0.56	21.42 ± 0.43	75.00	18.75	25.00	21.43
(C=0.1, çekirdek fonksiyonu=rbf)	75.83 ± 1.67	18.96 ± 0.42	25.00 ± 0.00	21.56 ± 0.27	75.00	18.75	25.00	21.43
(C=1, çekirdek fonksiyonu=linear)	84.17 ± 3.12	47.92 ± 11.64	56.39 ± 17.25	50.62 ± 12.83	91.25	73.48	76.39	73.49
(C=1, çekirdek fonksiyonu=poly)	77.50 ± 3.33	25.21 ± 12.40	33.06 ± 16.11	28.32 ± 13.46	76.25	31.98	27.78	26.44
(C=1, çekirdek fonksiyonu=rbf)	87.50 ± 0.00	37.00 ± 1.00	50.00 ± 0.00	41.19 ± 0.95	86.25	36.25	50.00	40.52
(C=10, çekirdek fonksiyonu=linear)	88.33 ± 4.86	56.25 ± 15.70	68.06 ± 12.11	59.81 ± 14.89	93.75	81.55	83.33	81.86
(C=10, çekirdek fonksiyonu=poly)	80.83 ± 5.65	39.21 ± 22.03	41.39 ± 24.73	38.39 ± 21.90	76.25	28.33	27.78	26.39
(C=10, çekirdek fonksiyonu=rbf)	86.67 ± 3.12	52.50 ± 15.05	57.50 ± 14.29	53.00 ± 12.41	91.25	81.53	73.61	68.87
(C=100, çekirdek fonksiyonu=linear)	88.33 ± 4.86	56.25 ± 15.70	68.06 ± 12.11	59.81 ± 14.89	91.25	76.94	77.78	75.35
(C=100, çekirdek fonksiyonu=poly)	89.17 ± 2.04	58.75 ± 14.81	63.89 ± 12.58	60.67 ± 14.03	92.50	83.12	77.78	75.15
(C=100, çekirdek fonksiyonu=rbf)	88.33 ± 6.12	59.17 ± 21.95	64.17 ± 19.29	60.69 ± 21.08	92.50	76.61	79.17	77.12

Tablo 11. Dengesiz MOS sensör verileri ile k-NN performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	87.50 ± 3.73	59.50 ± 17.67	65.56 ± 16.95	61.05 ± 16.92	90.00	71.88	73.61	71.77
(k=1, uzaklık metriği=manhattan)	89.17 ± 5.65	64.92 ± 23.87	67.50 ± 18.89	64.64 ± 21.03	90.00	71.88	73.61	71.77
(k=1, uzaklık metriği=chebyshev)	87.50 ± 5.27	60.75 ± 21.08	62.22 ± 16.18	60.33 ± 18.37	88.75	65.48	69.44	66.67
(k=3, uzaklık metriği=euclidean)	89.17 ± 4.25	67.50 ± 17.16	67.50 ± 17.36	65.17 ± 15.80	91.25	74.40	76.39	74.89
(k=3, uzaklık metriği=manhattan)	90.00 ± 4.25	69.17 ± 16.58	70.83 ± 16.67	67.67 ± 15.86	95.00	86.11	86.11	86.11
(k=3, uzaklık metriği=chebyshev)	90.00 ± 4.25	69.17 ± 15.94	69.17 ± 16.37	66.36 ± 14.58	90.00	71.88	73.61	71.77
(k=5, uzaklık metriği=euclidean)	85.00 ± 2.04	54.74 ± 6.91	56.39 ± 9.32	53.89 ± 6.83	87.50	62.72	66.67	63.12
(k=5, uzaklık metriği=manhattan)	86.67 ± 3.12	52.08 ± 7.22	58.33 ± 10.54	53.67 ± 8.04	88.75	67.00	69.44	66.52
(k=5, uzaklık metriği=chebyshev)	85.83 ± 3.33	56.25 ± 14.31	57.22 ± 14.36	55.00 ± 12.92	88.75	64.91	69.44	65.82
(k=7, uzaklık metriği=euclidean)	85.83 ± 2.04	52.24 ± 6.91	55.00 ± 8.50	52.08 ± 5.69	88.75	65.62	68.06	63.64
(k=7, uzaklık metriği=manhattan)	85.83 ± 2.04	52.24 ± 6.91	55.00 ± 8.50	52.08 ± 5.69	90.00	72.51	72.22	69.84
(k=7, uzaklık metriği=chebyshev)	85.83 ± 4.25	50.08 ± 11.52	54.17 ± 12.64	50.87 ± 10.74	87.50	61.90	66.67	63.54

Tablo 12. Dengesiz MOS sensör verileri ile RF performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	86.67 ± 6.67	61.07 ± 25.62	63.61 ± 20.63	60.54 ± 21.84	91.25	74.59	75.00	74.19
(ağaç sayısı=20, maksimum derinlik=10)	88.33 ± 6.12	65.24 ± 23.41	68.61 ± 18.08	64.88 ± 19.72	91.25	72.50	75.00	73.48
(ağaç sayısı=20, maksimum derinlik=20)	88.33 ± 6.12	65.24 ± 23.41	68.61 ± 18.08	64.88 ± 19.72	91.25	72.50	75.00	73.48
(ağaç sayısı=50, maksimum derinlik=5)	87.50 ± 5.89	58.47 ± 20.90	61.67 ± 18.33	59.14 ± 18.67	88.75	60.03	66.67	63.11
(ağaç sayısı=50, maksimum derinlik=10)	90.00 ± 6.24	66.39 ± 24.15	69.17 ± 19.65	66.68 ± 21.81	88.75	60.03	66.67	63.11
(ağaç sayısı=50, maksimum derinlik=20)	90.00 ± 6.24	66.39 ± 24.15	69.17 ± 19.65	66.68 ± 21.81	88.75	60.03	66.67	63.11
(ağaç sayısı=100, maksimum derinlik=5)	88.33 ± 4.86	66.04 ± 21.75	64.17 ± 17.00	64.09 ± 18.22	88.75	60.03	66.67	63.11
(ağaç sayısı=100, maksimum derinlik=10)	90.83 ± 5.53	77.89 ± 18.23	71.67 ± 18.52	72.84 ± 18.94	88.75	60.03	66.67	63.11
(ağaç sayısı=100, maksimum derinlik=20)	90.83 ± 5.53	77.89 ± 18.23	71.67 ± 18.52	72.84 ± 18.94	88.75	60.03	66.67	63.11

4.2. SMOTE ile Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 13. SMOTE ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	87.50 ± 2.64	62.50 ± 14.43	70.28 ± 14.29	62.38 ± 11.74	86.25	59.72	65.28	57.40
(C=0.1, optimizasyon yöntemi=saga)	86.67 ± 1.67	59.17 ± 8.50	67.78 ± 6.77	59.88 ± 6.53	87.50	67.29	71.81	64.85
(C=0.1, optimizasyon yöntemi=lbgfs)	85.83 ± 2.04	56.67 ± 6.24	67.50 ± 6.67	59.24 ± 6.39	87.50	66.19	70.42	66.34
(C=1, optimizasyon yöntemi=liblinear)	87.50 ± 2.64	68.33 ± 13.07	72.78 ± 8.50	65.40 ± 8.40	90.00	77.86	75.00	70.83
(C=1, optimizasyon yöntemi=saga)	85.00 ± 2.04	55.00 ± 4.08	62.50 ± 6.45	56.60 ± 4.30	90.00	73.97	77.36	73.24
(C=1, optimizasyon yöntemi=lbgfs)	85.83 ± 3.33	56.67 ± 3.33	64.17 ± 6.24	58.60 ± 3.85	90.00	73.21	79.72	75.32
(C=10, optimizasyon yöntemi=liblinear)	86.67 ± 3.12	62.50 ± 16.46	66.94 ± 11.37	60.88 ± 12.13	91.25	78.97	77.78	74.55
(C=10, optimizasyon yöntemi=saga)	85.00 ± 2.04	52.50 ± 8.16	61.67 ± 7.79	54.57 ± 7.57	92.50	78.57	82.92	79.76
(C=10, optimizasyon yöntemi=lbgfs)	86.67 ± 5.53	61.67 ± 19.61	68.33 ± 14.89	62.07 ± 17.08	88.75	70.63	76.94	72.08
(C=100, optimizasyon yöntemi=liblinear)	87.50 ± 5.27	60.83 ± 19.11	67.22 ± 13.91	60.69 ± 16.46	91.25	78.97	77.78	74.55
(C=100, optimizasyon yöntemi=saga)	85.00 ± 4.25	56.67 ± 17.60	65.00 ± 11.37	57.24 ± 13.17	92.50	78.57	82.92	79.76
(C=100, optimizasyon yöntemi=lbgfs)	86.67 ± 5.53	61.67 ± 19.61	68.33 ± 14.89	62.07 ± 17.08	90.00	73.97	77.36	73.24

Tablo 14. SMOTE ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	88.33 ± 3.12	63.33 ± 13.28	71.39 ± 9.44	64.19 ± 11.04	93.75	84.72	84.72	82.25
(C=0.1, çekirdek fonksiyonu=poly)	35.00 ± 24.52	36.65 ± 11.96	38.80 ± 13.62	23.22 ± 16.72	41.25	40.05	39.03	24.69
(C=0.1, çekirdek fonksiyonu=rbf)	86.67 ± 6.12	55.99 ± 23.28	61.67 ± 21.63	55.70 ± 21.02	90.00	71.88	75.00	70.05
(C=1, çekirdek fonksiyonu=linear)	87.50 ± 2.64	60.00 ± 12.68	64.72 ± 7.38	58.30 ± 8.64	92.50	83.33	81.94	78.74
(C=1, çekirdek fonksiyonu=poly)	80.00 ± 7.17	53.26 ± 17.89	57.81 ± 16.09	51.73 ± 14.57	88.75	62.85	68.06	63.68
(C=1, çekirdek fonksiyonu=rbf)	90.00 ± 3.33	68.07 ± 16.83	70.83 ± 13.94	65.70 ± 14.35	88.75	66.34	70.83	66.48
(C=10, çekirdek fonksiyonu=linear)	88.33 ± 3.12	60.50 ± 12.72	66.39 ± 6.66	60.27 ± 8.04	92.50	83.33	81.94	78.74
(C=10, çekirdek fonksiyonu=poly)	86.67 ± 4.86	58.33 ± 16.67	67.50 ± 17.36	59.90 ± 16.18	92.50	75.83	77.78	76.15
(C=10, çekirdek fonksiyonu=rbf)	90.83 ± 4.86	70.57 ± 16.74	72.50 ± 15.28	69.53 ± 17.13	93.75	81.55	83.33	81.86
(C=100, çekirdek fonksiyonu=linear)	88.33 ± 3.12	60.50 ± 12.72	66.39 ± 6.66	60.27 ± 8.04	92.50	83.33	81.94	78.74
(C=100, çekirdek fonksiyonu=poly)	89.17 ± 2.04	70.00 ± 10.34	74.44 ± 8.36	68.21 ± 8.61	92.50	77.08	79.17	77.87
(C=100, çekirdek fonksiyonu=rbf)	92.50 ± 4.08	73.90 ± 16.46	75.83 ± 14.04	73.20 ± 15.92	93.75	81.55	83.33	81.86

Tablo 15. SMOTE ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	87.50 ± 3.73	59.50 ± 16.87	66.39 ± 15.30	59.38 ± 14.14	90.00	69.79	72.22	70.77
(k=1, uzaklık metriği=manhattan)	91.67 ± 7.45	68.67 ± 27.04	74.17 ± 23.33	70.36 ± 25.84	90.00	69.79	72.22	70.77
(k=1, uzaklık metriği=chebyshev)	86.67 ± 4.86	61.25 ± 18.52	62.22 ± 15.67	59.31 ± 15.91	90.00	69.79	72.22	70.77
(k=3, uzaklık metriği=euclidean)	85.00 ± 2.04	51.17 ± 12.30	60.56 ± 11.03	53.71 ± 10.72	90.00	73.07	75.97	73.86
(k=3, uzaklık metriği=manhattan)	89.17 ± 4.25	63.67 ± 21.01	68.33 ± 16.37	63.21 ± 17.72	91.25	74.40	76.39	74.89
(k=3, uzaklık metriği=chebyshev)	85.83 ± 5.00	54.92 ± 18.47	62.22 ± 13.19	56.71 ± 16.82	87.50	66.37	69.03	67.16
(k=5, uzaklık metriği=euclidean)	86.67 ± 4.08	57.83 ± 19.92	65.56 ± 14.84	58.55 ± 15.84	87.50	69.10	75.14	71.69
(k=5, uzaklık metriği=manhattan)	89.17 ± 3.33	65.33 ± 19.65	70.56 ± 15.89	65.24 ± 16.68	91.25	74.40	76.39	74.89
(k=5, uzaklık metriği=chebyshev)	85.00 ± 5.65	53.75 ± 19.29	61.94 ± 13.54	55.71 ± 17.64	85.00	63.89	68.19	65.52
(k=7, uzaklık metriği=euclidean)	86.67 ± 4.08	57.83 ± 19.92	65.56 ± 14.84	58.55 ± 15.84	87.50	69.10	75.14	71.69
(k=7, uzaklık metriği=manhattan)	88.33 ± 3.12	64.50 ± 19.13	68.89 ± 14.07	62.90 ± 15.05	91.25	74.40	76.39	74.89
(k=7, uzaklık metriği=chebyshev)	85.00 ± 5.65	53.75 ± 19.29	61.94 ± 13.54	55.71 ± 17.64	85.00	63.89	68.19	65.52

Tablo 16. SMOTE ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	86.67 ± 6.67	62.92 ± 20.51	62.79 ± 22.05	61.01 ± 19.94	90.00	72.88	72.22	72.03
(ağaç sayısı=20, maksimum derinlik=10)	90.00 ± 4.25	68.83 ± 15.57	71.40 ± 16.51	68.45 ± 15.73	91.25	73.33	74.17	73.48
(ağaç sayısı=20, maksimum derinlik=20)	90.00 ± 4.25	68.83 ± 15.57	71.40 ± 16.51	68.45 ± 15.73	91.25	73.33	74.17	73.48
(ağaç sayısı=50, maksimum derinlik=5)	88.33 ± 3.12	65.00 ± 14.81	63.33 ± 13.79	62.19 ± 12.73	91.25	74.59	75.00	74.19
(ağaç sayısı=50, maksimum derinlik=10)	87.50 ± 3.73	63.33 ± 16.33	60.83 ± 15.50	60.36 ± 14.21	91.25	74.59	75.00	74.19
(ağaç sayısı=50, maksimum derinlik=20)	88.33 ± 4.86	64.17 ± 17.60	62.50 ± 18.07	62.52 ± 17.38	91.25	74.59	75.00	74.19
(ağaç sayısı=100, maksimum derinlik=5)	89.17 ± 4.25	65.83 ± 16.12	65.00 ± 16.37	64.36 ± 15.95	93.75	88.14	79.17	75.32
(ağaç sayısı=100, maksimum derinlik=10)	88.33 ± 4.86	64.17 ± 17.60	62.50 ± 18.07	62.52 ± 17.38	95.00	89.58	83.33	81.66
(ağaç sayısı=100, maksimum derinlik=20)	88.33 ± 4.86	64.17 ± 17.60	62.50 ± 18.07	62.52 ± 17.38	95.00	89.58	83.33	81.66

4.3. Borderline SMOTE İle Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 17. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	84.17 ± 1.67	45.83 ± 4.56	49.72 ± 7.32	46.74 ± 4.16	92.50	79.91	80.56	77.47
(C=0.1, optimizasyon yöntemi=saga)	87.50 ± 3.73	53.75 ± 3.33	65.28 ± 7.45	57.36 ± 4.09	91.25	78.97	80.14	76.81
(C=0.1, optimizasyon yöntemi=lbgfs)	87.50 ± 3.73	53.75 ± 3.33	65.28 ± 7.45	57.36 ± 4.09	91.25	78.97	80.14	76.81
(C=1, optimizasyon yöntemi=liblinear)	87.50 ± 3.73	60.83 ± 17.20	66.11 ± 15.61	59.38 ± 13.66	90.00	73.61	75.14	72.97
(C=1, optimizasyon yöntemi=saga)	85.83 ± 3.33	53.33 ± 6.12	63.33 ± 8.63	56.29 ± 6.41	91.25	78.97	80.14	76.81
(C=1, optimizasyon yöntemi=lbgfs)	88.33 ± 4.86	60.42 ± 11.02	70.00 ± 10.12	62.07 ± 9.12	91.25	78.97	80.14	76.81
(C=10, optimizasyon yöntemi=liblinear)	87.50 ± 3.73	56.67 ± 15.72	66.94 ± 11.57	58.05 ± 11.46	91.25	78.97	80.14	76.81
(C=10, optimizasyon yöntemi=saga)	85.83 ± 3.33	52.92 ± 13.79	63.33 ± 14.21	54.43 ± 11.93	91.25	78.97	80.14	76.81
(C=10, optimizasyon yöntemi=lbgfs)	86.67 ± 3.12	55.00 ± 13.02	66.67 ± 11.28	57.24 ± 10.39	92.50	78.57	82.92	79.76
(C=100, optimizasyon yöntemi=liblinear)	87.50 ± 4.56	57.72 ± 15.40	64.44 ± 11.20	58.33 ± 11.53	93.75	82.14	83.33	81.22
(C=100, optimizasyon yöntemi=saga)	85.83 ± 3.33	52.92 ± 13.79	63.33 ± 14.21	54.43 ± 11.93	91.25	78.97	80.14	76.81
(C=100, optimizasyon yöntemi=lbgfs)	86.67 ± 1.67	57.17 ± 9.78	66.67 ± 7.61	58.82 ± 6.05	87.50	68.25	76.53	71.03

Tablo 18. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	89.17 ± 3.33	64.17 ± 11.96	72.22 ± 6.51	64.52 ± 8.34	93.75	84.72	84.72	82.25
(C=0.1, çekirdek fonksiyonu=poly)	35.83 ± 23.21	27.89 ± 2.23	30.75 ± 10.79	17.71 ± 9.62	48.75	27.84	39.17	23.19
(C=0.1, çekirdek fonksiyonu=rbf)	87.50 ± 4.56	52.65 ± 11.09	60.00 ± 10.74	53.67 ± 9.97	88.75	75.98	72.22	66.03
(C=1, çekirdek fonksiyonu=linear)	88.33 ± 3.12	56.67 ± 15.50	67.22 ± 11.74	58.19 ± 11.55	93.75	84.72	84.72	82.25
(C=1, çekirdek fonksiyonu=poly)	77.50 ± 5.65	38.67 ± 10.20	48.64 ± 13.17	41.00 ± 10.83	92.50	78.27	79.17	77.50
(C=1, çekirdek fonksiyonu=rbf)	89.17 ± 4.25	60.24 ± 14.89	66.67 ± 15.59	61.61 ± 14.50	93.75	83.52	83.33	82.94
(C=10, çekirdek fonksiyonu=linear)	88.33 ± 1.67	58.83 ± 12.62	67.22 ± 8.27	59.77 ± 7.69	92.50	83.33	81.94	78.74
(C=10, çekirdek fonksiyonu=poly)	88.33 ± 5.53	62.08 ± 17.60	72.50 ± 17.53	65.60 ± 17.58	93.75	80.28	81.94	80.81
(C=10, çekirdek fonksiyonu=rbf)	91.67 ± 4.56	73.90 ± 16.46	74.17 ± 15.46	72.53 ± 17.24	93.75	81.55	83.33	81.86
(C=100, çekirdek fonksiyonu=linear)	88.33 ± 1.67	58.83 ± 12.62	67.22 ± 8.27	59.77 ± 7.69	92.50	83.33	81.94	78.74
(C=100, çekirdek fonksiyonu=poly)	89.17 ± 2.04	66.58 ± 10.99	72.78 ± 9.53	67.21 ± 11.75	91.25	75.30	78.75	76.70
(C=100, çekirdek fonksiyonu=rbf)	92.50 ± 4.86	74.74 ± 20.18	75.83 ± 17.36	74.20 ± 18.88	92.50	77.08	79.17	77.87

Tablo 19. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=öklid)	87.50 ± 5.27	62.08 ± 19.83	65.56 ± 20.31	62.86 ± 19.51	92.50	79.17	80.56	78.68
(k=1, uzaklık metriği=manhattan)	88.33 ± 4.86	64.50 ± 20.87	68.89 ± 17.97	64.05 ± 18.84	91.25	74.40	76.39	74.89
(k=1, uzaklık metriği=chebyshev)	88.33 ± 4.08	61.25 ± 17.36	63.89 ± 13.55	61.48 ± 14.71	92.50	74.80	77.78	75.28
(k=3, uzaklık metriği=öklid)	85.83 ± 5.00	54.58 ± 19.02	57.50 ± 16.54	55.67 ± 18.08	91.25	76.94	77.78	75.35
(k=3, uzaklık metriği=manhattan)	88.33 ± 4.86	59.50 ± 20.75	67.50 ± 17.16	61.52 ± 18.71	91.25	74.40	76.39	74.89
(k=3, uzaklık metriği=chebyshev)	86.67 ± 4.08	60.00 ± 15.05	60.56 ± 14.15	60.19 ± 14.67	91.25	76.19	80.14	76.58
(k=5, uzaklık metriği=öklid)	85.83 ± 5.00	56.25 ± 18.16	62.22 ± 17.09	58.02 ± 17.29	91.25	75.36	77.78	74.40
(k=5, uzaklık metriği=manhattan)	87.50 ± 5.27	59.50 ± 20.75	67.22 ± 17.05	61.38 ± 18.74	91.25	73.21	76.39	73.96
(k=5, uzaklık metriği=chebyshev)	85.00 ± 5.65	57.08 ± 18.52	58.61 ± 16.37	57.55 ± 17.78	90.00	74.80	79.72	76.26
(k=7, uzaklık metriği=öklid)	86.67 ± 4.86	60.42 ± 18.63	63.89 ± 18.47	61.02 ± 18.10	90.00	72.02	77.36	73.24
(k=7, uzaklık metriği=manhattan)	87.50 ± 5.27	59.50 ± 20.75	67.22 ± 17.05	61.38 ± 18.74	91.25	73.21	76.39	73.96
(k=7, uzaklık metriği=chebyshev)	85.00 ± 5.65	57.08 ± 18.52	58.61 ± 16.37	57.55 ± 17.78	91.25	77.08	82.50	79.24

Tablo 20. Borderline SMOTE ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	87.50 ± 2.64	55.28 ± 6.30	60.28 ± 6.89	55.99 ± 4.75	88.75	67.72	66.39	67.00
(ağaç sayısı=20, maksimum derinlik=10)	86.67 ± 4.08	58.64 ± 12.24	56.39 ± 11.93	55.48 ± 10.64	91.25	75.78	76.39	74.79
(ağaç sayısı=20, maksimum derinlik=20)	86.67 ± 4.08	58.64 ± 12.24	56.39 ± 11.93	55.48 ± 10.64	91.25	75.78	76.39	74.79
(ağaç sayısı=50, maksimum derinlik=5)	86.67 ± 4.08	52.50 ± 6.77	56.39 ± 8.94	53.07 ± 6.30	91.25	74.11	78.75	75.77
(ağaç sayısı=50, maksimum derinlik=10)	89.17 ± 4.25	62.22 ± 15.60	65.56 ± 12.47	62.10 ± 12.80	93.75	80.28	81.94	80.81
(ağaç sayısı=50, maksimum derinlik=20)	89.17 ± 4.25	62.22 ± 15.60	65.56 ± 12.47	62.10 ± 12.80	93.75	80.28	81.94	80.81
(ağaç sayısı=100, maksimum derinlik=5)	87.50 ± 2.64	53.33 ± 6.12	58.06 ± 7.16	54.40 ± 4.84	90.00	70.98	75.97	72.83
(ağaç sayısı=100, maksimum derinlik=10)	89.17 ± 3.33	62.38 ± 15.22	65.56 ± 11.70	62.09 ± 12.33	91.25	72.02	75.00	72.50
(ağaç sayısı=100, maksimum derinlik=20)	89.17 ± 3.33	62.38 ± 15.22	65.56 ± 11.70	62.09 ± 12.33	91.25	72.02	75.00	72.50

4.4. SMOTE-Tomek ile Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 21. SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	87.50 ± 2.64	59.17 ± 13.79	68.06 ± 13.12	60.36 ± 11.81	86.25	59.72	65.28	57.40
(C=0.1, optimizasyon yöntemi=saga)	85.83 ± 3.33	56.67 ± 9.72	62.78 ± 10.07	58.07 ± 9.55	92.50	80.36	80.56	77.98
(C=0.1, optimizasyon yöntemi=lbgfs)	85.00 ± 2.04	55.00 ± 8.50	65.00 ± 10.74	57.57 ± 8.79	87.50	64.60	70.42	64.90
(C=1, optimizasyon yöntemi=liblinear)	87.50 ± 2.64	65.83 ± 9.28	69.44 ± 9.38	64.57 ± 7.29	91.25	78.97	77.78	74.55
(C=1, optimizasyon yöntemi=saga)	85.00 ± 4.25	50.00 ± 9.50	62.50 ± 9.13	53.07 ± 9.08	90.00	77.86	77.36	73.51
(C=1, optimizasyon yöntemi=lbgfs)	85.83 ± 3.33	54.17 ± 8.33	63.33 ± 7.79	56.57 ± 7.85	92.50	78.57	82.92	79.76
(C=10, optimizasyon yöntemi=liblinear)	88.33 ± 4.08	63.00 ± 9.55	68.89 ± 11.93	64.77 ± 10.33	93.75	82.14	83.33	81.22
(C=10, optimizasyon yöntemi=saga)	84.17 ± 3.12	49.17 ± 8.50	59.17 ± 7.17	51.76 ± 7.29	92.50	80.36	82.92	79.91
(C=10, optimizasyon yöntemi=lbgfs)	86.67 ± 5.53	61.67 ± 19.61	68.33 ± 14.89	62.07 ± 17.08	88.75	70.63	76.94	72.08
(C=100, optimizasyon yöntemi=liblinear)	86.67 ± 4.86	58.67 ± 16.17	66.39 ± 14.28	58.02 ± 13.06	91.25	78.97	77.78	74.55
(C=100, optimizasyon yöntemi=saga)	85.00 ± 4.25	52.17 ± 10.64	60.83 ± 6.77	53.49 ± 8.08	91.25	78.97	80.14	76.81
(C=100, optimizasyon yöntemi=lbgfs)	86.67 ± 5.53	61.67 ± 19.61	68.33 ± 14.89	62.07 ± 17.08	87.50	68.25	76.53	71.03

Tablo 22. SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	88.33 ± 4.08	65.00 ± 16.37	71.39 ± 10.49	64.19 ± 13.43	92.50	79.17	80.56	78.68
(C=0.1, çekirdek fonksiyonu=poly)	35.00 ± 24.52	36.66 ± 12.01	38.82 ± 13.73	23.31 ± 16.97	41.25	40.05	39.03	24.69
(C=0.1, çekirdek fonksiyonu=rbf)	85.83 ± 3.33	52.24 ± 13.65	56.67 ± 14.81	52.03 ± 11.27	88.75	64.91	69.44	65.82
(C=1, çekirdek fonksiyonu=linear)	86.67 ± 3.12	57.00 ± 15.55	62.22 ± 9.88	55.38 ± 11.17	95.00	86.46	87.50	85.58
(C=1, çekirdek fonksiyonu=poly)	78.33 ± 4.86	50.17 ± 10.41	55.03 ± 13.03	49.77 ± 10.21	88.75	54.26	66.67	59.23
(C=1, çekirdek fonksiyonu=rbf)	89.17 ± 3.33	63.07 ± 13.74	68.33 ± 12.53	62.70 ± 12.69	88.75	66.34	70.83	66.48
(C=10, çekirdek fonksiyonu=linear)	87.50 ± 2.64	60.00 ± 12.68	64.72 ± 7.38	58.30 ± 8.64	92.50	83.33	81.94	78.74
(C=10, çekirdek fonksiyonu=poly)	89.17 ± 3.33	65.33 ± 15.77	73.33 ± 15.58	66.79 ± 15.00	92.50	75.83	77.78	76.15
(C=10, çekirdek fonksiyonu=rbf)	91.67 ± 5.89	70.15 ± 23.19	72.50 ± 19.11	69.98 ± 21.08	93.75	81.55	83.33	81.86
(C=100, çekirdek fonksiyonu=linear)	87.50 ± 2.64	59.17 ± 11.17	64.72 ± 7.38	58.46 ± 8.95	91.25	76.94	77.78	75.35
(C=100, çekirdek fonksiyonu=poly)	89.17 ± 2.04	69.17 ± 10.41	74.44 ± 8.36	68.38 ± 8.64	91.25	75.30	78.75	76.70
(C=100, çekirdek fonksiyonu=rbf)	91.67 ± 2.64	72.65 ± 14.56	73.33 ± 10.07	70.82 ± 12.19	93.75	81.55	83.33	81.86

Tablo 23. SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	87.50 ± 3.73	59.50 ± 17.67	65.56 ± 16.95	60.21 ± 16.08	90.00	69.79	72.22	70.77
(k=1, uzaklık metriği=manhattan)	92.50 ± 7.64	74.50 ± 28.20	76.67 ± 23.95	74.36 ± 26.50	91.25	74.40	76.39	74.89
(k=1, uzaklık metriği=chebyshev)	87.50 ± 5.27	61.67 ± 20.52	64.72 ± 17.18	61.14 ± 17.92	88.75	63.48	68.06	65.08
(k=3, uzaklık metriği=euclidean)	89.17 ± 3.33	64.50 ± 17.42	70.56 ± 14.52	65.88 ± 16.20	88.75	68.60	71.81	69.92
(k=3, uzaklık metriği=manhattan)	89.17 ± 5.00	65.75 ± 21.87	69.72 ± 18.58	65.83 ± 19.56	91.25	74.40	76.39	74.89
(k=3, uzaklık metriği=chebyshev)	88.33 ± 4.86	67.08 ± 16.48	69.44 ± 11.28	65.00 ± 14.43	87.50	66.37	69.03	67.16
(k=5, uzaklık metriği=euclidean)	88.33 ± 4.08	62.83 ± 18.68	68.06 ± 17.24	64.05 ± 17.77	86.25	66.67	72.36	68.85
(k=5, uzaklık metriği=manhattan)	89.17 ± 5.00	66.17 ± 21.79	69.17 ± 18.18	65.19 ± 18.97	92.50	79.17	80.56	78.68
(k=5, uzaklık metriği=chebyshev)	85.83 ± 5.65	59.17 ± 21.27	62.78 ± 15.20	57.50 ± 17.70	87.50	66.67	71.39	68.79
(k=7, uzaklık metriği=euclidean)	88.33 ± 4.08	62.83 ± 18.68	68.06 ± 17.24	64.05 ± 17.77	87.50	68.45	72.78	69.80
(k=7, uzaklık metriği=manhattan)	88.33 ± 4.86	64.50 ± 20.87	68.89 ± 17.97	63.88 ± 18.79	90.00	72.02	78.33	74.58
(k=7, uzaklık metriği=chebyshev)	85.00 ± 4.25	56.58 ± 16.43	61.11 ± 12.64	56.07 ± 14.30	87.50	70.00	72.78	70.01

Tablo 24. SMOTE-Tomek ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	87.50 ± 4.56	64.40 ± 8.05	62.78 ± 9.15	62.04 ± 9.19	90.00	72.21	72.78	71.69
(ağaç sayısı=20, maksimum derinlik=10)	85.83 ± 4.25	57.22 ± 14.99	59.46 ± 14.35	58.11 ± 14.67	88.75	63.64	65.83	63.74
(ağaç sayısı=20, maksimum derinlik=20)	88.33 ± 1.67	55.00 ± 14.53	60.00 ± 10.41	55.71 ± 12.61	93.75	88.61	80.56	80.02
(ağaç sayısı=50, maksimum derinlik=5)	89.17 ± 3.33	63.33 ± 14.77	64.72 ± 11.64	63.02 ± 13.35	91.25	76.19	80.14	76.58
(ağaç sayısı=50, maksimum derinlik=10)	84.17 ± 4.86	51.17 ± 18.26	53.89 ± 15.75	51.95 ± 17.19	91.25	75.78	75.00	73.40
(ağaç sayısı=50, maksimum derinlik=20)	89.17 ± 3.33	63.75 ± 13.15	65.56 ± 13.10	63.83 ± 13.01	93.75	84.86	85.69	84.94
(ağaç sayısı=100, maksimum derinlik=5)	88.33 ± 4.86	63.75 ± 19.17	61.67 ± 16.75	61.64 ± 17.30	92.50	79.91	80.56	77.47
(ağaç sayısı=100, maksimum derinlik=10)	88.33 ± 3.12	57.50 ± 16.54	60.83 ± 13.59	58.69 ± 15.41	91.25	73.72	75.97	71.95
(ağaç sayısı=100, maksimum derinlik=20)	88.33 ± 5.53	60.00 ± 23.25	60.83 ± 17.99	58.81 ± 20.35	92.50	77.08	79.17	77.87

4.5. ADASYN ile Dengelenmiş MOS Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 25. ADASYN ile dengelenmiş MOS sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	86.67 ± 3.12	52.50 ± 14.09	59.72 ± 13.29	53.36 ± 11.38	87.50	73.89	69.44	60.36
(C=0.1, optimizasyon yöntemi=saga)	84.17 ± 4.08	49.50 ± 12.06	58.61 ± 10.07	50.38 ± 8.49	90.00	77.86	77.36	73.51
(C=0.1, optimizasyon yöntemi=lbgfs)	84.17 ± 4.08	49.50 ± 12.06	58.61 ± 10.07	50.38 ± 8.49	91.25	78.97	80.14	76.81
(C=1, optimizasyon yöntemi=liblinear)	83.33 ± 5.89	49.50 ± 15.13	51.11 ± 14.09	48.88 ± 13.27	88.75	76.95	74.58	69.89
(C=1, optimizasyon yöntemi=saga)	85.83 ± 4.25	53.33 ± 13.79	64.17 ± 10.84	55.24 ± 11.32	91.25	78.97	80.14	76.81
(C=1, optimizasyon yöntemi=lbgfs)	86.67 ± 5.53	56.25 ± 14.67	65.00 ± 10.52	57.86 ± 12.23	92.50	80.36	80.56	77.98
(C=10, optimizasyon yöntemi=liblinear)	85.83 ± 5.65	57.00 ± 16.34	62.78 ± 15.20	56.55 ± 13.43	90.00	77.86	77.36	73.51
(C=10, optimizasyon yöntemi=saga)	85.83 ± 4.25	53.33 ± 13.79	64.17 ± 10.84	55.24 ± 11.32	91.25	78.97	80.14	76.81
(C=10, optimizasyon yöntemi=lbgfs)	85.00 ± 3.33	53.33 ± 13.28	61.67 ± 9.40	54.57 ± 10.17	91.25	76.19	80.14	76.58
(C=100, optimizasyon yöntemi=liblinear)	88.33 ± 4.86	61.67 ± 17.95	69.72 ± 12.37	61.19 ± 13.34	91.25	78.97	77.78	74.55
(C=100, optimizasyon yöntemi=saga)	85.83 ± 4.25	53.33 ± 13.79	64.17 ± 10.84	55.24 ± 11.32	91.25	78.97	80.14	76.81
(C=100, optimizasyon yöntemi=lbgfs)	85.00 ± 3.33	53.33 ± 13.28	61.67 ± 9.40	54.57 ± 10.17	91.25	78.97	77.78	74.55

Tablo 26. ADASYN ile dengelenmiş MOS sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	90.00 ± 4.25	69.17 ± 18.18	74.72 ± 14.20	69.02 ± 16.51	95.00	86.46	87.50	85.58
(C=0.1, çekirdek fonksiyonu=poly)	40.00 ± 22.45	30.05 ± 3.93	34.93 ± 6.97	20.82 ± 10.59	51.25	36.31	43.75	29.30
(C=0.1, çekirdek fonksiyonu=rbf)	84.17 ± 3.12	50.00 ± 12.64	49.17 ± 7.17	47.67 ± 8.71	86.25	58.86	63.89	57.41
(C=1, çekirdek fonksiyonu=linear)	86.67 ± 3.12	55.00 ± 15.68	62.22 ± 9.88	55.36 ± 11.18	92.50	79.17	80.56	78.68
(C=1, çekirdek fonksiyonu=poly)	79.17 ± 5.27	46.58 ± 16.48	51.42 ± 14.90	47.19 ± 16.44	87.50	61.46	67.64	62.73
(C=1, çekirdek fonksiyonu=rbf)	88.33 ± 4.08	58.07 ± 16.31	65.83 ± 12.75	59.53 ± 13.49	91.25	75.36	77.78	74.40
(C=10, çekirdek fonksiyonu=linear)	86.67 ± 3.12	55.00 ± 15.68	62.22 ± 9.88	55.36 ± 11.18	92.50	83.33	81.94	78.74
(C=10, çekirdek fonksiyonu=poly)	87.50 ± 4.56	60.50 ± 9.90	63.61 ± 11.37	59.32 ± 8.85	91.25	82.20	79.17	74.93
(C=10, çekirdek fonksiyonu=rbf)	92.50 ± 5.53	70.99 ± 22.51	75.00 ± 18.45	71.48 ± 20.15	93.75	81.55	83.33	81.86
(C=100, çekirdek fonksiyonu=linear)	86.67 ± 3.12	55.00 ± 15.68	62.22 ± 9.88	55.36 ± 11.18	92.50	83.33	81.94	78.74
(C=100, çekirdek fonksiyonu=poly)	87.50 ± 2.64	58.33 ± 12.08	70.28 ± 9.65	60.38 ± 10.03	93.75	84.72	84.72	82.25
(C=100, çekirdek fonksiyonu=rbf)	93.33 ± 5.65	78.49 ± 21.90	77.50 ± 19.11	76.48 ± 20.19	93.75	81.55	83.33	81.86

Tablo 27. ADASYN ile dengelenmiş MOS sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	88.33 ± 4.08	62.83 ± 18.68	68.06 ± 17.24	62.38 ± 16.63	90.00	69.79	72.22	70.77
(k=1, uzaklık metriği=manhattan)	90.83 ± 6.67	69.92 ± 25.68	71.67 ± 20.98	68.81 ± 23.45	90.00	69.79	72.22	70.77
(k=1, uzaklık metriği=chebyshev)	88.33 ± 1.67	61.67 ± 11.61	67.78 ± 9.48	62.21 ± 8.84	88.75	67.26	69.44	67.92
(k=3, uzaklık metriği=euclidean)	88.33 ± 4.08	62.83 ± 18.68	68.06 ± 17.24	63.21 ± 17.13	90.00	75.83	77.36	74.30
(k=3, uzaklık metriği=manhattan)	90.83 ± 6.12	71.58 ± 25.69	73.89 ± 21.63	70.83 ± 23.32	90.00	76.74	77.36	72.74
(k=3, uzaklık metriği=chebyshev)	87.50 ± 3.73	59.17 ± 10.99	63.61 ± 9.10	59.05 ± 8.74	86.25	65.83	70.00	66.68
(k=5, uzaklık metriği=euclidean)	87.50 ± 3.73	61.17 ± 17.48	65.56 ± 14.52	61.71 ± 16.09	88.75	75.00	78.33	72.73
(k=5, uzaklık metriği=manhattan)	89.17 ± 5.00	65.75 ± 21.87	69.72 ± 18.58	65.83 ± 19.56	91.25	80.36	81.53	76.63
(k=5, uzaklık metriği=chebyshev)	88.33 ± 5.53	62.92 ± 19.11	68.61 ± 13.28	62.83 ± 15.72	86.25	69.17	73.75	68.91
(k=7, uzaklık metriği=euclidean)	88.33 ± 4.08	62.83 ± 18.68	68.06 ± 17.24	64.05 ± 17.77	88.75	75.00	78.33	72.73
(k=7, uzaklık metriği=manhattan)	89.17 ± 5.00	65.75 ± 21.87	69.72 ± 18.58	65.83 ± 19.56	91.25	80.36	81.53	76.63
(k=7, uzaklık metriği=chebyshev)	87.50 ± 5.27	62.42 ± 19.15	66.94 ± 13.93	60.86 ± 16.35	86.25	69.17	73.75	68.91

Tablo 28. ADASYN ile dengelenmiş MOS sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	89.17 ± 4.25	62.08 ± 12.39	63.33 ± 12.47	60.95 ± 10.85	92.50	79.17	79.17	79.17
(ağaç sayısı=20, maksimum derinlik=10)	89.17 ± 4.25	62.08 ± 12.39	63.33 ± 12.47	60.95 ± 10.85	91.25	76.34	76.39	76.24
(ağaç sayısı=20, maksimum derinlik=20)	89.17 ± 4.25	62.08 ± 12.39	63.33 ± 12.47	60.95 ± 10.85	91.25	76.34	76.39	76.24
(ağaç sayısı=50, maksimum derinlik=5)	87.50 ± 2.64	59.17 ± 13.28	62.50 ± 11.79	57.86 ± 10.10	90.00	70.83	73.61	71.77
(ağaç sayısı=50, maksimum derinlik=10)	90.00 ± 5.00	65.83 ± 17.11	68.33 ± 13.84	64.40 ± 14.59	92.50	77.98	80.56	78.74
(ağaç sayısı=50, maksimum derinlik=20)	89.17 ± 4.25	62.92 ± 12.80	66.67 ± 11.79	61.79 ± 10.78	92.50	77.98	80.56	78.74
(ağaç sayısı=100, maksimum derinlik=5)	88.33 ± 3.12	64.40 ± 15.92	64.17 ± 11.06	61.11 ± 12.23	91.25	77.08	79.17	74.89
(ağaç sayısı=100, maksimum derinlik=10)	90.00 ± 5.00	65.83 ± 17.11	68.33 ± 13.84	64.40 ± 14.59	90.00	71.94	75.00	71.77
(ağaç sayısı=100, maksimum derinlik=20)	89.17 ± 4.25	62.92 ± 12.80	66.67 ± 11.79	61.79 ± 10.78	90.00	71.94	75.00	71.77

4.6. Dengesiz QCM Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 29. Dengesiz QCM sensör verileri ile LR performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	75.00 ± 5.89	43.45 ± 13.06	43.95 ± 14.04	42.68 ± 11.98	72.50	39.00	49.17	42.58
(C=0.1, optimizasyon yöntemi=saga)	75.83 ± 6.12	29.01 ± 7.46	33.64 ± 9.84	31.03 ± 8.38	77.50	48.41	48.61	46.49
(C=0.1, optimizasyon yöntemi=lbgfs)	74.17 ± 6.67	33.08 ± 8.95	33.10 ± 6.66	32.57 ± 6.87	76.25	46.11	44.44	43.60
(C=1, optimizasyon yöntemi=liblinear)	71.67 ± 6.12	36.25 ± 14.71	39.23 ± 16.48	37.48 ± 15.25	73.75	42.84	47.36	44.26
(C=1, optimizasyon yöntemi=saga)	70.00 ± 7.17	37.65 ± 18.11	38.67 ± 16.92	38.05 ± 17.46	75.00	49.00	52.50	50.43
(C=1, optimizasyon yöntemi=lbgfs)	70.00 ± 5.53	36.46 ± 18.25	36.45 ± 11.74	35.62 ± 14.16	77.50	54.16	50.56	51.10
(C=10, optimizasyon yöntemi=liblinear)	70.00 ± 6.67	37.34 ± 13.59	40.06 ± 15.36	38.35 ± 14.22	75.00	49.96	45.00	46.50
(C=10, optimizasyon yöntemi=saga)	67.50 ± 8.08	36.98 ± 15.30	38.68 ± 16.85	37.37 ± 15.84	70.00	44.73	47.08	45.27
(C=10, optimizasyon yöntemi=lbgfs)	69.17 ± 8.58	43.18 ± 17.89	38.39 ± 12.80	39.72 ± 14.40	72.50	52.34	51.25	49.70
(C=100, optimizasyon yöntemi=liblinear)	69.17 ± 4.25	39.43 ± 13.98	41.18 ± 14.24	39.69 ± 13.80	72.50	59.53	46.53	49.05
(C=100, optimizasyon yöntemi=saga)	66.67 ± 7.45	36.53 ± 15.15	38.41 ± 16.79	36.98 ± 15.72	70.00	44.73	47.08	45.27
(C=100, optimizasyon yöntemi=lbgfs)	68.33 ± 7.73	42.12 ± 18.64	40.34 ± 16.93	40.62 ± 17.51	71.25	51.30	48.47	47.80

Tablo 30. Dengesiz QCM sensör verileri ile SVM performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	75.83 ± 7.17	38.74 ± 15.27	35.86 ± 10.22	36.41 ± 11.81	80.00	75.23	62.36	58.25
(C=0.1, çekirdek fonksiyonu=poly)	77.50 ± 2.04	29.28 ± 12.48	28.33 ± 4.08	26.77 ± 6.28	75.00	18.75	25.00	21.43
(C=0.1, çekirdek fonksiyonu=rbf)	75.83 ± 1.67	18.96 ± 0.42	25.00 ± 0.00	21.56 ± 0.27	75.00	18.75	25.00	21.43
(C=1, çekirdek fonksiyonu=linear)	65.00 ± 6.24	44.60 ± 15.19	46.18 ± 15.50	42.44 ± 13.05	75.00	58.08	70.14	62.02
(C=1, çekirdek fonksiyonu=poly)	78.33 ± 3.12	29.46 ± 12.70	30.00 ± 6.67	28.38 ± 8.63	75.00	18.75	25.00	21.43
(C=1, çekirdek fonksiyonu=rbf)	77.50 ± 2.04	29.28 ± 12.48	28.33 ± 4.08	26.77 ± 6.28	75.00	18.75	25.00	21.43
(C=10, çekirdek fonksiyonu=linear)	65.00 ± 8.16	47.24 ± 14.51	48.41 ± 15.79	45.00 ± 14.07	76.25	62.52	74.31	65.74
(C=10, çekirdek fonksiyonu=poly)	79.17 ± 2.64	39.78 ± 19.38	33.89 ± 10.15	34.29 ± 13.61	72.50	44.00	27.92	28.25
(C=10, çekirdek fonksiyonu=rbf)	75.83 ± 7.17	35.05 ± 19.01	32.78 ± 9.28	32.50 ± 11.53	80.00	67.22	57.22	54.28
(C=100, çekirdek fonksiyonu=linear)	65.00 ± 8.16	47.24 ± 14.51	48.41 ± 15.79	45.00 ± 14.07	76.25	59.48	71.94	63.02
(C=100, çekirdek fonksiyonu=poly)	74.17 ± 6.12	36.18 ± 12.92	37.51 ± 9.94	35.56 ± 10.27	80.00	59.27	58.33	57.99
(C=100, çekirdek fonksiyonu=rbf)	70.83 ± 9.50	38.42 ± 15.14	36.40 ± 12.15	35.66 ± 11.35	76.25	59.52	69.17	63.43

Tablo 31. Dengesiz QCM sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	73.33 ± 5.65	38.25 ± 11.45	37.78 ± 8.89	35.60 ± 6.91	72.50	51.10	56.11	47.68
(k=1, uzaklık metriği=manhattan)	74.17 ± 6.12	38.40 ± 12.21	38.06 ± 9.40	35.99 ± 7.90	71.25	45.55	55.69	47.47
(k=1, uzaklık metriği=chebyshev)	72.50 ± 4.25	33.20 ± 5.35	37.79 ± 7.39	34.48 ± 5.19	67.50	42.26	46.25	43.71
(k=3, uzaklık metriği=euclidean)	76.67 ± 3.33	35.97 ± 9.26	33.35 ± 5.05	33.29 ± 5.88	76.25	58.53	53.75	53.74
(k=3, uzaklık metriği=manhattan)	74.17 ± 4.08	33.19 ± 11.84	31.13 ± 6.20	31.22 ± 8.06	77.50	55.11	51.94	52.89
(k=3, uzaklık metriği=chebyshev)	76.67 ± 2.04	36.62 ± 8.43	34.18 ± 4.62	34.49 ± 5.36	71.25	40.51	43.61	40.72
(k=5, uzaklık metriği=euclidean)	75.83 ± 5.53	30.84 ± 15.44	34.18 ± 14.50	32.36 ± 15.03	75.00	40.87	50.83	44.57
(k=5, uzaklık metriği=manhattan)	78.33 ± 5.53	36.68 ± 18.98	39.44 ± 14.69	37.31 ± 15.95	76.25	42.32	51.25	45.90
(k=5, uzaklık metriği=chebyshev)	74.17 ± 3.12	27.75 ± 5.16	29.47 ± 5.11	28.45 ± 4.97	67.50	33.47	29.44	30.02
(k=7, uzaklık metriği=euclidean)	78.33 ± 4.08	35.27 ± 12.47	37.22 ± 11.19	35.61 ± 11.42	72.50	33.06	37.08	34.94
(k=7, uzaklık metriği=manhattan)	77.50 ± 5.00	33.11 ± 11.63	34.44 ± 9.60	33.19 ± 10.12	76.25	37.50	47.50	41.54
(k=7, uzaklık metriği=chebyshev)	77.50 ± 4.25	35.36 ± 10.46	32.78 ± 4.70	32.70 ± 6.11	68.75	27.34	27.50	27.33

Tablo 32. Dengesiz QCM sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	76.67 ± 3.33	26.91 ± 10.61	31.39 ± 9.69	28.74 ± 10.16	73.75	31.58	29.17	28.47
(ağaç sayısı=20, maksimum derinlik=10)	79.17 ± 2.64	37.18 ± 10.43	36.39 ± 8.35	34.68 ± 8.29	73.75	32.57	33.75	32.87
(ağaç sayısı=20, maksimum derinlik=20)	79.17 ± 2.64	37.18 ± 10.43	36.39 ± 8.35	34.68 ± 8.29	75.00	35.07	38.75	36.76
(ağaç sayısı=50, maksimum derinlik=5)	77.50 ± 2.04	29.62 ± 12.22	28.33 ± 4.08	26.98 ± 6.11	76.25	36.33	34.58	34.35
(ağaç sayısı=50, maksimum derinlik=10)	78.33 ± 1.67	31.66 ± 11.18	30.00 ± 4.08	28.87 ± 5.71	77.50	57.71	38.75	40.43
(ağaç sayısı=50, maksimum derinlik=20)	78.33 ± 1.67	31.66 ± 11.18	30.00 ± 4.08	28.87 ± 5.71	78.75	66.32	41.53	44.77
(ağaç sayısı=100, maksimum derinlik=5)	76.67 ± 3.33	24.13 ± 10.76	30.00 ± 10.00	26.67 ± 10.48	76.25	44.16	34.58	35.82
(ağaç sayısı=100, maksimum derinlik=10)	77.50 ± 2.04	27.30 ± 10.07	28.33 ± 4.08	26.59 ± 5.67	77.50	38.42	39.58	38.52
(ağaç sayısı=100, maksimum derinlik=20)	77.50 ± 2.04	27.30 ± 10.07	28.33 ± 4.08	26.59 ± 5.67	77.50	38.42	39.58	38.52

4.7. SMOTE ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 33. SMOTE ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	63.33 ± 7.17	35.92 ± 12.61	39.52 ± 17.44	34.86 ± 11.01	61.25	38.70	61.25	42.16
(C=0.1, optimizasyon yöntemi=saga)	63.33 ± 7.17	31.19 ± 9.53	37.28 ± 18.35	32.75 ± 12.03	62.50	39.97	52.78	42.34
(C=0.1, optimizasyon yöntemi=lbgfs)	62.50 ± 5.27	30.06 ± 8.34	32.56 ± 13.16	30.57 ± 9.73	66.25	43.94	56.39	46.74
(C=1, optimizasyon yöntemi=liblinear)	65.83 ± 7.17	31.45 ± 8.41	38.13 ± 17.35	32.94 ± 10.25	67.50	47.36	64.31	51.48
(C=1, optimizasyon yöntemi=saga)	66.67 ± 8.33	38.01 ± 14.17	38.38 ± 17.81	36.97 ± 14.32	68.75	48.52	58.89	51.89
(C=1, optimizasyon yöntemi=lbgfs)	65.83 ± 7.17	37.93 ± 13.97	35.88 ± 13.38	36.30 ± 13.02	68.75	47.50	55.14	49.96
(C=10, optimizasyon yöntemi=liblinear)	68.33 ± 4.25	33.03 ± 6.47	40.35 ± 15.22	34.82 ± 8.07	70.00	51.11	60.69	54.07
(C=10, optimizasyon yöntemi=saga)	65.83 ± 8.50	37.96 ± 14.12	38.10 ± 17.79	36.80 ± 14.25	70.00	48.74	55.56	51.01
(C=10, optimizasyon yöntemi=lbgfs)	67.50 ± 4.86	38.05 ± 13.80	36.43 ± 12.67	36.64 ± 12.56	71.25	50.87	48.47	47.77
(C=100, optimizasyon yöntemi=liblinear)	65.83 ± 4.86	32.66 ± 6.62	37.31 ± 15.01	33.69 ± 8.57	66.25	45.91	53.33	48.56
(C=100, optimizasyon yöntemi=saga)	65.83 ± 8.50	37.96 ± 14.12	38.10 ± 17.79	36.80 ± 14.25	71.25	50.23	58.33	52.78
(C=100, optimizasyon yöntemi=lbgfs)	65.83 ± 9.65	38.19 ± 14.49	38.10 ± 18.14	36.97 ± 14.64	70.00	46.66	48.06	46.13

Tablo 34. SMOTE ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	63.33 ± 7.17	39.83 ± 17.67	39.50 ± 17.66	38.26 ± 16.69	73.75	60.42	75.83	64.33
(C=0.1, çekirdek fonksiyonu=poly)	25.00 ± 5.89	43.61 ± 13.25	37.44 ± 7.35	22.13 ± 8.58	23.75	27.24	30.42	13.01
(C=0.1, çekirdek fonksiyonu=rbf)	56.67 ± 4.25	28.70 ± 6.63	34.55 ± 11.42	28.41 ± 6.25	62.50	32.71	45.00	34.66
(C=1, çekirdek fonksiyonu=linear)	69.17 ± 7.73	39.44 ± 15.29	39.20 ± 14.27	38.61 ± 14.33	77.50	64.55	74.72	67.28
(C=1, çekirdek fonksiyonu=poly)	39.17 ± 8.58	28.15 ± 5.68	38.49 ± 4.67	22.46 ± 4.71	41.25	26.34	36.25	20.62
(C=1, çekirdek fonksiyonu=rbf)	65.83 ± 4.08	45.67 ± 14.42	53.42 ± 18.63	44.63 ± 12.82	66.25	44.36	61.53	48.53
(C=10, çekirdek fonksiyonu=linear)	65.00 ± 8.58	41.47 ± 16.66	41.45 ± 18.44	40.28 ± 16.67	82.50	76.55	78.75	75.20
(C=10, çekirdek fonksiyonu=poly)	54.17 ± 8.33	39.39 ± 12.66	41.48 ± 15.79	35.64 ± 11.94	55.00	38.56	48.61	36.33
(C=10, çekirdek fonksiyonu=rbf)	72.50 ± 7.73	35.98 ± 11.29	36.96 ± 9.80	35.71 ± 9.38	73.75	60.00	72.08	64.03
(C=100, çekirdek fonksiyonu=linear)	65.00 ± 8.58	41.47 ± 16.66	41.45 ± 18.44	40.28 ± 16.67	81.25	69.22	75.97	70.77
(C=100, çekirdek fonksiyonu=poly)	59.17 ± 4.08	26.82 ± 4.66	29.50 ± 5.94	26.79 ± 4.38	63.75	51.10	68.75	55.23
(C=100, çekirdek fonksiyonu=rbf)	73.33 ± 8.16	35.48 ± 11.22	37.24 ± 9.99	35.44 ± 9.19	76.25	60.76	69.17	64.23

Tablo 35. SMOTE ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	69.17 ± 4.25	35.59 ± 8.87	37.25 ± 4.99	34.82 ± 5.49	67.50	45.39	54.58	47.98
(k=1, uzaklık metriği=manhattan)	67.50 ± 3.12	37.09 ± 9.38	41.43 ± 13.96	36.95 ± 7.56	68.75	46.92	55.83	48.66
(k=1, uzaklık metriği=chebyshev)	66.67 ± 3.73	36.38 ± 10.64	37.82 ± 7.02	35.12 ± 7.67	68.75	48.75	61.94	52.04
(k=3, uzaklık metriği=euclidean)	60.00 ± 8.58	38.55 ± 8.73	42.56 ± 9.74	36.24 ± 6.73	61.25	42.44	64.44	45.62
(k=3, uzaklık metriği=manhattan)	56.67 ± 4.25	35.24 ± 9.05	40.09 ± 14.02	33.70 ± 7.84	62.50	40.81	56.53	43.31
(k=3, uzaklık metriği=chebyshev)	58.33 ± 5.89	36.18 ± 7.18	42.56 ± 15.64	34.99 ± 7.73	53.75	38.46	51.39	38.95
(k=5, uzaklık metriği=euclidean)	52.50 ± 5.65	37.51 ± 8.70	44.85 ± 10.88	34.64 ± 6.50	56.25	40.80	62.78	42.17
(k=5, uzaklık metriği=manhattan)	50.83 ± 4.86	32.58 ± 5.87	38.16 ± 13.24	30.43 ± 5.61	57.50	41.36	63.19	43.10
(k=5, uzaklık metriği=chebyshev)	52.50 ± 6.24	32.80 ± 5.69	41.23 ± 15.83	31.46 ± 6.43	50.00	33.83	45.42	32.87
(k=7, uzaklık metriği=euclidean)	50.00 ± 4.56	36.09 ± 7.95	44.05 ± 10.73	32.79 ± 4.29	53.75	40.09	61.94	41.17
(k=7, uzaklık metriği=manhattan)	50.00 ± 7.45	33.75 ± 3.87	42.62 ± 10.04	31.64 ± 2.76	52.50	37.54	59.17	38.45
(k=7, uzaklık metriği=chebyshev)	47.50 ± 5.00	29.50 ± 3.95	34.87 ± 13.72	27.12 ± 4.41	50.00	33.83	45.42	32.87

Tablo 36. SMOTE ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	73.33 ± 7.73	42.58 ± 24.29	39.74 ± 18.25	40.08 ± 20.45	71.25	50.23	59.17	53.41
(ağaç sayısı=20, maksimum derinlik=10)	74.17 ± 4.86	37.87 ± 10.66	38.35 ± 7.35	37.35 ± 8.33	68.75	43.19	47.64	44.74
(ağaç sayısı=20, maksimum derinlik=20)	74.17 ± 4.86	37.87 ± 10.66	38.35 ± 7.35	37.35 ± 8.33	68.75	43.19	47.64	44.74
(ağaç sayısı=50, maksimum derinlik=5)	73.33 ± 7.26	35.82 ± 23.27	36.13 ± 19.09	35.38 ± 20.21	73.75	53.53	61.39	56.43
(ağaç sayısı=50, maksimum derinlik=10)	74.17 ± 3.12	42.38 ± 16.09	39.20 ± 10.54	39.30 ± 12.51	68.75	44.26	47.64	45.20
(ağaç sayısı=50, maksimum derinlik=20)	74.17 ± 3.12	42.38 ± 16.09	39.20 ± 10.54	39.30 ± 12.51	70.00	44.91	48.06	45.91
(ağaç sayısı=100, maksimum derinlik=5)	75.00 ± 5.89	36.28 ± 18.57	35.85 ± 13.10	35.43 ± 15.08	71.25	52.34	56.81	53.83
(ağaç sayısı=100, maksimum derinlik=10)	77.50 ± 3.33	41.48 ± 16.23	41.68 ± 10.71	40.80 ± 12.75	72.50	49.43	52.64	50.72
(ağaç sayısı=100, maksimum derinlik=20)	76.67 ± 4.25	38.72 ± 17.02	39.18 ± 11.53	38.18 ± 13.47	71.25	48.48	52.22	49.92

4.8. Borderline SMOTE ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 37. Borderline SMOTE ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	69.17 ± 9.35	44.90 ± 19.91	42.85 ± 13.82	40.15 ± 13.14	67.50	38.24	59.58	42.82
(C=0.1, optimizasyon yöntemi=saga)	66.67 ± 5.89	35.97 ± 17.90	31.45 ± 7.07	32.10 ± 9.91	70.00	45.83	55.28	45.80
(C=0.1, optimizasyon yöntemi=lbgfs)	69.17 ± 4.25	37.32 ± 17.61	34.50 ± 8.38	33.99 ± 9.89	73.75	50.23	58.89	51.59
(C=1, optimizasyon yöntemi=liblinear)	70.00 ± 8.08	34.93 ± 12.99	36.18 ± 9.23	34.08 ± 9.61	70.00	53.10	62.78	49.97
(C=1, optimizasyon yöntemi=saga)	66.67 ± 5.89	33.06 ± 13.32	31.45 ± 7.49	31.38 ± 9.37	72.50	57.47	66.94	58.34
(C=1, optimizasyon yöntemi=lbgfs)	66.67 ± 5.27	28.53 ± 5.45	31.45 ± 7.64	29.24 ± 5.81	75.00	51.95	58.89	54.75
(C=10, optimizasyon yöntemi=liblinear)	70.83 ± 7.91	35.46 ± 12.91	36.45 ± 9.45	34.73 ± 9.71	72.50	54.74	64.58	57.02
(C=10, optimizasyon yöntemi=saga)	65.83 ± 6.67	37.74 ± 16.06	35.89 ± 13.05	35.86 ± 13.80	71.25	49.15	59.03	52.23
(C=10, optimizasyon yöntemi=lbgfs)	64.17 ± 7.26	33.53 ± 16.54	35.34 ± 17.61	33.65 ± 16.52	73.75	60.65	51.67	51.49
(C=100, optimizasyon yöntemi=liblinear)	67.50 ± 7.17	29.95 ± 7.58	34.23 ± 12.73	31.31 ± 9.27	67.50	49.23	59.17	52.33
(C=100, optimizasyon yöntemi=saga)	65.83 ± 6.67	37.74 ± 16.06	35.89 ± 13.05	35.86 ± 13.80	70.00	48.00	58.61	51.02
(C=100, optimizasyon yöntemi=lbgfs)	64.17 ± 8.98	33.93 ± 15.43	35.34 ± 17.63	34.13 ± 16.26	70.00	50.89	48.06	47.21

Tablo 38. Borderline SMOTE ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	67.50 ± 4.86	29.15 ± 5.49	34.21 ± 12.19	30.30 ± 6.59	77.50	64.77	69.58	66.47
(C=0.1, çekirdek fonksiyonu=poly)	30.00 ± 5.53	33.09 ± 11.29	34.37 ± 2.62	20.39 ± 6.23	26.25	27.31	31.25	14.23
(C=0.1, çekirdek fonksiyonu=rbf)	65.00 ± 5.65	35.88 ± 14.42	43.41 ± 13.39	33.78 ± 8.59	68.75	36.85	56.25	41.17
(C=1, çekirdek fonksiyonu=linear)	66.67 ± 8.74	41.15 ± 13.35	42.82 ± 16.49	40.13 ± 13.72	80.00	67.73	74.17	70.03
(C=1, çekirdek fonksiyonu=poly)	42.50 ± 7.17	32.40 ± 8.55	35.15 ± 9.29	25.08 ± 2.19	50.00	27.02	39.17	23.71
(C=1, çekirdek fonksiyonu=rbf)	74.17 ± 4.86	38.10 ± 14.78	45.57 ± 15.68	38.43 ± 11.51	71.25	45.99	59.44	48.15
(C=10, çekirdek fonksiyonu=linear)	66.67 ± 7.91	42.23 ± 13.88	42.84 ± 16.40	40.34 ± 13.22	77.50	61.67	74.72	65.87
(C=10, çekirdek fonksiyonu=poly)	62.50 ± 5.89	36.23 ± 15.38	42.85 ± 14.61	34.72 ± 12.43	70.00	48.16	55.28	47.61
(C=10, çekirdek fonksiyonu=rbf)	72.50 ± 6.77	39.34 ± 16.38	36.96 ± 11.87	36.29 ± 11.85	77.50	62.90	65.83	63.86
(C=100, çekirdek fonksiyonu=linear)	65.83 ± 8.50	41.87 ± 14.13	42.56 ± 16.40	39.79 ± 13.50	77.50	61.67	74.72	65.87
(C=100, çekirdek fonksiyonu=poly)	67.50 ± 6.67	40.25 ± 20.03	35.34 ± 12.11	34.55 ± 12.19	68.75	54.89	50.00	51.10
(C=100, çekirdek fonksiyonu=rbf)	72.50 ± 10.74	39.95 ± 16.75	36.96 ± 13.22	36.92 ± 13.18	75.00	58.43	68.75	62.46

Tablo 39. Borderline SMOTE ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	71.67 ± 4.86	32.25 ± 5.40	40.57 ± 15.52	34.59 ± 7.60	71.25	46.92	54.31	47.61
(k=1, uzaklık metriği=manhattan)	72.50 ± 6.24	34.78 ± 7.96	40.83 ± 15.46	35.56 ± 7.89	71.25	45.67	54.31	48.21
(k=1, uzaklık metriği=chebyshev)	64.17 ± 4.25	29.09 ± 3.00	33.38 ± 6.09	29.77 ± 3.18	65.00	39.21	45.28	39.58
(k=3, uzaklık metriği=euclidean)	69.17 ± 5.00	32.38 ± 7.74	40.04 ± 16.15	34.12 ± 9.05	62.50	42.62	55.14	44.86
(k=3, uzaklık metriği=manhattan)	70.00 ± 4.08	34.72 ± 5.50	41.71 ± 14.55	36.44 ± 7.21	67.50	44.65	56.81	47.89
(k=3, uzaklık metriği=chebyshev)	65.83 ± 4.08	31.71 ± 6.70	38.11 ± 9.77	32.26 ± 6.17	60.00	39.37	45.97	39.51
(k=5, uzaklık metriği=euclidean)	60.83 ± 6.24	25.72 ± 5.60	28.14 ± 7.98	25.64 ± 5.67	61.25	42.98	54.72	44.56
(k=5, uzaklık metriği=manhattan)	65.83 ± 5.53	32.75 ± 1.93	42.85 ± 9.32	34.24 ± 2.65	60.00	39.05	51.94	40.78
(k=5, uzaklık metriği=chebyshev)	61.67 ± 6.67	29.27 ± 5.27	35.37 ± 10.10	29.48 ± 5.59	58.75	38.61	45.56	39.30
(k=7, uzaklık metriği=euclidean)	61.67 ± 4.86	28.11 ± 6.76	37.88 ± 14.14	28.90 ± 6.86	58.75	36.55	52.92	38.57
(k=7, uzaklık metriği=manhattan)	59.17 ± 9.65	31.71 ± 1.80	40.70 ± 9.33	31.57 ± 3.10	57.50	36.13	52.50	37.92
(k=7, uzaklık metriği=chebyshev)	60.83 ± 7.26	27.00 ± 6.92	32.88 ± 12.26	26.99 ± 6.86	58.75	37.37	46.94	38.01

Tablo 40. Borderline SMOTE ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	75.83 ± 3.12	37.01 ± 14.06	33.07 ± 7.78	33.64 ± 9.96	68.75	45.31	37.36	37.38
(ağaç sayısı=20, maksimum derinlik=10)	74.17 ± 5.53	36.18 ± 17.22	38.63 ± 16.43	35.97 ± 15.17	75.00	62.44	42.78	45.86
(ağaç sayısı=20, maksimum derinlik=20)	73.33 ± 4.25	34.30 ± 15.39	33.63 ± 11.45	33.35 ± 13.01	75.00	62.44	42.78	45.86
(ağaç sayısı=50, maksimum derinlik=5)	76.67 ± 2.04	39.89 ± 11.94	33.90 ± 5.87	35.02 ± 7.99	71.25	50.08	41.39	44.11
(ağaç sayısı=50, maksimum derinlik=10)	75.83 ± 4.86	39.81 ± 13.42	34.44 ± 5.92	35.68 ± 8.65	71.25	46.97	39.17	41.23
(ağaç sayısı=50, maksimum derinlik=20)	75.83 ± 4.86	39.81 ± 13.42	34.44 ± 5.92	35.68 ± 8.65	71.25	46.97	39.17	41.23
(ağaç sayısı=100, maksimum derinlik=5)	76.67 ± 2.04	34.75 ± 14.89	31.67 ± 6.66	31.72 ± 9.48	73.75	58.97	52.08	52.55
(ağaç sayısı=100, maksimum derinlik=10)	75.83 ± 1.67	29.73 ± 14.32	30.00 ± 7.06	29.21 ± 9.78	75.00	68.51	47.36	53.34
(ağaç sayısı=100, maksimum derinlik=20)	75.83 ± 1.67	29.73 ± 14.32	30.00 ± 7.06	29.21 ± 9.78	75.00	68.51	47.36	53.34

4.9. SMOTE-Tomek ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 41. SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	65.83 ± 10.34	36.89 ± 13.61	39.52 ± 17.12	35.50 ± 11.30	60.00	37.63	57.08	40.82
(C=0.1, optimizasyon yöntemi=saga)	61.67 ± 6.67	30.54 ± 8.58	35.92 ± 16.77	31.44 ± 10.62	62.50	43.20	58.89	46.26
(C=0.1, optimizasyon yöntemi=lbgfs)	61.67 ± 6.12	32.78 ± 13.97	32.30 ± 12.66	32.20 ± 13.35	66.25	48.80	64.86	52.95
(C=1, optimizasyon yöntemi=liblinear)	64.17 ± 8.58	30.07 ± 7.14	36.74 ± 16.93	31.42 ± 9.24	70.00	55.97	65.28	56.90
(C=1, optimizasyon yöntemi=saga)	66.67 ± 5.89	33.05 ± 11.41	33.95 ± 12.52	33.14 ± 11.54	68.75	52.02	51.39	49.27
(C=1, optimizasyon yöntemi=lbgfs)	68.33 ± 6.24	39.62 ± 13.53	38.10 ± 12.76	38.10 ± 12.50	67.50	43.99	50.97	46.06
(C=10, optimizasyon yöntemi=liblinear)	66.67 ± 7.45	37.31 ± 13.53	38.41 ± 17.04	36.18 ± 12.17	58.75	40.69	43.33	40.48
(C=10, optimizasyon yöntemi=saga)	67.50 ± 10.34	39.49 ± 16.65	38.65 ± 17.72	38.27 ± 16.54	68.75	44.90	50.00	45.87
(C=10, optimizasyon yöntemi=lbgfs)	65.00 ± 4.25	36.45 ± 14.40	35.61 ± 12.45	35.13 ± 12.90	71.25	51.80	50.83	48.81
(C=100, optimizasyon yöntemi=liblinear)	67.50 ± 6.67	32.36 ± 7.38	34.52 ± 9.37	32.82 ± 7.86	66.25	45.42	49.58	46.88
(C=100, optimizasyon yöntemi=saga)	64.17 ± 8.58	35.83 ± 15.01	33.10 ± 15.22	33.79 ± 14.45	67.50	47.09	42.50	43.32
(C=100, optimizasyon yöntemi=lbgfs)	66.67 ± 7.45	37.22 ± 13.98	36.17 ± 12.46	35.83 ± 12.75	70.00	46.66	48.06	46.13

Tablo 42. SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	65.00 ± 4.25	41.57 ± 12.06	42.28 ± 15.54	39.23 ± 10.38	73.75	60.42	75.83	64.33
(C=0.1, çekirdek fonksiyonu=poly)	26.67 ± 7.73	43.65 ± 13.30	39.39 ± 10.43	23.36 ± 10.22	23.75	27.24	30.42	13.01
(C=0.1, çekirdek fonksiyonu=rbf)	64.17 ± 5.65	27.19 ± 6.71	34.78 ± 13.93	28.34 ± 7.77	65.00	34.79	50.42	37.79
(C=1, çekirdek fonksiyonu=linear)	70.00 ± 6.67	40.15 ± 12.62	36.97 ± 9.64	37.52 ± 9.85	78.75	66.03	77.50	69.43
(C=1, çekirdek fonksiyonu=poly)	40.00 ± 6.77	34.06 ± 8.38	40.16 ± 2.96	24.88 ± 4.04	42.50	26.33	36.67	20.91
(C=1, çekirdek fonksiyonu=rbf)	69.17 ± 4.25	38.44 ± 11.43	53.70 ± 19.73	42.15 ± 13.66	66.25	44.36	61.53	48.53
(C=10, çekirdek fonksiyonu=linear)	69.17 ± 8.16	43.87 ± 16.45	46.73 ± 19.37	43.41 ± 16.52	81.25	70.07	82.08	73.30
(C=10, çekirdek fonksiyonu=poly)	58.33 ± 11.49	45.94 ± 7.85	45.10 ± 11.35	40.24 ± 6.92	62.50	43.07	60.28	44.92
(C=10, çekirdek fonksiyonu=rbf)	71.67 ± 8.90	34.98 ± 15.29	35.85 ± 12.18	34.08 ± 11.89	75.00	61.58	76.25	66.35
(C=100, çekirdek fonksiyonu=linear)	67.50 ± 6.67	38.66 ± 14.65	37.82 ± 13.43	37.41 ± 13.43	78.75	64.09	75.14	67.64
(C=100, çekirdek fonksiyonu=poly)	64.17 ± 7.73	30.39 ± 6.66	31.99 ± 10.33	29.70 ± 6.91	61.25	48.00	65.56	51.39
(C=100, çekirdek fonksiyonu=rbf)	71.67 ± 8.08	33.07 ± 9.66	35.85 ± 10.37	34.00 ± 9.69	75.00	59.75	68.75	63.33

Tablo 43. SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	68.33 ± 5.00	29.76 ± 2.84	34.75 ± 5.34	31.00 ± 2.80	73.75	57.25	72.08	62.24
(k=1, uzaklık metriği=manhattan)	69.17 ± 5.00	34.42 ± 8.17	39.74 ± 14.50	34.73 ± 7.49	73.75	56.64	69.72	60.66
(k=1, uzaklık metriği=chebyshev)	65.83 ± 4.86	32.10 ± 7.78	35.32 ± 9.23	32.47 ± 7.25	67.50	44.95	52.50	46.57
(k=3, uzaklık metriği=euclidean)	58.33 ± 7.91	34.26 ± 10.18	43.13 ± 16.87	34.67 ± 12.28	66.25	52.93	75.56	58.25
(k=3, uzaklık metriği=manhattan)	55.00 ± 3.12	29.43 ± 4.91	37.32 ± 12.21	29.44 ± 6.08	63.75	49.06	67.78	52.52
(k=3, uzaklık metriği=chebyshev)	58.33 ± 5.27	31.39 ± 4.44	35.06 ± 4.02	30.61 ± 3.98	58.75	42.65	61.39	44.61
(k=5, uzaklık metriği=euclidean)	51.67 ± 4.25	33.07 ± 7.96	44.58 ± 19.59	32.39 ± 10.04	60.00	46.66	68.75	49.09
(k=5, uzaklık metriği=manhattan)	58.33 ± 4.56	39.85 ± 6.91	47.62 ± 10.93	38.57 ± 4.39	61.25	49.07	67.78	51.18
(k=5, uzaklık metriği=chebyshev)	50.83 ± 5.53	28.90 ± 2.41	33.45 ± 6.85	27.64 ± 3.86	57.50	44.17	61.11	46.58
(k=7, uzaklık metriği=euclidean)	50.83 ± 3.12	32.77 ± 5.12	42.35 ± 12.71	30.64 ± 5.68	57.50	46.87	70.28	49.61
(k=7, uzaklık metriği=manhattan)	50.83 ± 4.86	38.06 ± 12.69	46.52 ± 20.89	34.55 ± 12.45	51.25	39.05	56.53	39.31
(k=7, uzaklık metriği=chebyshev)	49.17 ± 3.12	31.37 ± 4.08	35.12 ± 3.12	28.81 ± 3.41	55.00	41.35	60.14	42.55

Tablo 44. SMOTE-Tomek ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	73.33 ± 7.73	42.15 ± 22.67	41.99 ± 18.87	41.03 ± 19.64	65.00	44.12	52.36	46.84
(ağaç sayısı=20, maksimum derinlik=10)	72.50 ± 7.73	31.06 ± 12.58	32.54 ± 9.28	30.95 ± 9.84	76.25	52.62	52.36	52.12
(ağaç sayısı=20, maksimum derinlik=20)	71.67 ± 4.86	30.85 ± 7.35	38.10 ± 12.57	32.99 ± 8.38	75.00	52.56	56.53	54.03
(ağaç sayısı=50, maksimum derinlik=5)	73.33 ± 4.25	35.01 ± 13.74	36.15 ± 12.13	35.40 ± 12.98	71.25	47.23	52.08	49.21
(ağaç sayısı=50, maksimum derinlik=10)	74.17 ± 4.86	30.20 ± 7.61	35.58 ± 12.44	32.16 ± 9.02	75.00	52.82	52.08	50.78
(ağaç sayısı=50, maksimum derinlik=20)	74.17 ± 4.08	41.84 ± 19.08	33.90 ± 9.39	35.68 ± 12.03	75.00	53.15	60.28	55.80
(ağaç sayısı=100, maksimum derinlik=5)	75.83 ± 4.86	41.16 ± 13.07	39.74 ± 11.93	39.73 ± 12.16	75.00	57.82	60.28	58.12
(ağaç sayısı=100, maksimum derinlik=10)	74.17 ± 4.08	31.69 ± 11.73	29.46 ± 4.75	29.11 ± 6.82	77.50	57.82	62.08	59.75
(ağaç sayısı=100, maksimum derinlik=20)	75.83 ± 3.12	39.84 ± 14.10	39.21 ± 16.27	37.72 ± 13.86	73.75	51.74	53.89	52.70

4.10. ADASYN ile Dengelenmiş QCM Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 45. ADASYN ile dengelenmiş QCM sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	63.33 ± 7.17	36.82 ± 12.28	40.91 ± 16.57	35.93 ± 10.65	63.75	48.08	65.42	50.67
(C=0.1, optimizasyon yöntemi=saga)	65.83 ± 6.67	40.51 ± 14.51	39.50 ± 17.22	38.91 ± 14.69	65.00	46.11	58.47	49.32
(C=0.1, optimizasyon yöntemi=lbgfs)	65.00 ± 5.65	40.51 ± 14.51	37.00 ± 13.05	38.08 ± 13.39	65.00	46.11	58.47	49.32
(C=1, optimizasyon yöntemi=liblinear)	66.67 ± 7.45	32.20 ± 9.27	39.80 ± 18.26	33.83 ± 11.09	66.25	48.10	59.44	49.85
(C=1, optimizasyon yöntemi=saga)	68.33 ± 8.58	40.39 ± 14.67	40.32 ± 17.38	39.20 ± 14.81	66.25	46.40	58.06	49.59
(C=1, optimizasyon yöntemi=lbgfs)	68.33 ± 6.24	40.45 ± 14.57	38.10 ± 12.76	38.60 ± 13.26	68.75	46.61	51.39	47.86
(C=10, optimizasyon yöntemi=liblinear)	67.50 ± 4.86	34.24 ± 8.00	41.48 ± 15.82	36.09 ± 9.82	73.75	55.56	64.17	58.02
(C=10, optimizasyon yöntemi=saga)	69.17 ± 8.58	40.45 ± 14.52	40.60 ± 16.92	39.34 ± 14.45	66.25	44.96	54.31	47.58
(C=10, optimizasyon yöntemi=lbgfs)	69.17 ± 6.24	40.83 ± 14.59	38.39 ± 12.50	39.09 ± 13.16	75.00	56.97	55.83	54.30
(C=100, optimizasyon yöntemi=liblinear)	70.83 ± 7.91	39.09 ± 15.07	42.57 ± 16.88	38.83 ± 13.08	71.25	52.44	55.83	53.73
(C=100, optimizasyon yöntemi=saga)	68.33 ± 7.73	40.00 ± 14.88	40.32 ± 17.04	38.96 ± 14.67	65.00	44.13	53.89	46.65
(C=100, optimizasyon yöntemi=lbgfs)	69.17 ± 7.73	40.15 ± 15.01	40.61 ± 16.94	39.34 ± 14.81	75.00	63.91	53.47	54.60

Tablo 46. ADASYN ile dengelenmiş QCM sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	65.00 ± 10.07	40.54 ± 18.73	40.03 ± 18.57	39.14 ± 18.18	73.75	60.57	77.22	64.56
(C=0.1, çekirdek fonksiyonu=poly)	26.67 ± 6.77	38.99 ± 10.84	36.87 ± 5.53	20.63 ± 4.68	21.25	27.27	29.58	11.91
(C=0.1, çekirdek fonksiyonu=rbf)	55.83 ± 11.06	29.77 ± 7.04	35.67 ± 16.64	29.49 ± 9.31	62.50	32.32	45.00	34.26
(C=1, çekirdek fonksiyonu=linear)	68.33 ± 6.77	38.57 ± 14.20	38.93 ± 13.90	37.96 ± 13.39	77.50	65.60	78.47	68.87
(C=1, çekirdek fonksiyonu=poly)	39.17 ± 8.58	30.90 ± 5.87	39.90 ± 5.48	24.19 ± 6.32	40.00	27.27	35.83	20.60
(C=1, çekirdek fonksiyonu=rbf)	66.67 ± 4.56	42.59 ± 13.42	49.53 ± 20.29	43.11 ± 14.45	68.75	52.59	69.44	57.22
(C=10, çekirdek fonksiyonu=linear)	65.83 ± 9.28	40.17 ± 14.14	38.65 ± 13.46	38.51 ± 13.25	78.75	65.93	78.89	68.72
(C=10, çekirdek fonksiyonu=poly)	54.17 ± 10.54	40.21 ± 10.38	40.10 ± 13.75	36.28 ± 9.84	58.75	51.26	53.19	47.25
(C=10, çekirdek fonksiyonu=rbf)	71.67 ± 8.90	34.98 ± 15.29	35.85 ± 12.18	34.08 ± 11.89	73.75	58.76	72.08	63.23
(C=100, çekirdek fonksiyonu=linear)	65.83 ± 9.28	40.17 ± 14.14	38.65 ± 13.46	38.51 ± 13.25	80.00	66.79	79.31	69.76
(C=100, çekirdek fonksiyonu=poly)	58.33 ± 7.45	29.41 ± 7.46	30.06 ± 9.04	27.65 ± 5.91	66.25	53.56	69.58	57.30
(C=100, çekirdek fonksiyonu=rbf)	71.67 ± 10.34	33.95 ± 14.38	34.46 ± 12.74	32.89 ± 11.27	75.00	57.23	65.00	60.06

Tablo 47. ADASYN ile dengelenmiş QCM sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	68.33 ± 4.25	30.17 ± 2.58	34.75 ± 5.27	31.25 ± 2.55	72.50	55.48	69.31	59.44
(k=1, uzaklık metriği=manhattan)	67.50 ± 6.12	31.34 ± 6.05	39.20 ± 15.69	33.29 ± 8.13	72.50	54.99	65.56	57.58
(k=1, uzaklık metriği=chebyshev)	63.33 ± 4.08	32.18 ± 7.32	34.49 ± 5.29	31.14 ± 4.24	70.00	47.55	54.72	50.04
(k=3, uzaklık metriği=euclidean)	55.83 ± 5.00	29.90 ± 5.40	34.25 ± 9.56	29.20 ± 6.04	66.25	51.06	73.19	55.73
(k=3, uzaklık metriği=manhattan)	60.83 ± 4.25	36.40 ± 9.05	41.46 ± 9.37	35.25 ± 6.08	62.50	48.95	67.36	51.61
(k=3, uzaklık metriği=chebyshev)	55.83 ± 4.25	28.53 ± 1.69	32.02 ± 4.95	27.41 ± 2.30	65.00	47.07	61.39	50.07
(k=5, uzaklık metriği=euclidean)	52.50 ± 6.24	31.18 ± 5.93	36.78 ± 11.01	29.88 ± 6.99	58.75	47.84	70.69	50.14
(k=5, uzaklık metriği=manhattan)	50.00 ± 4.56	40.09 ± 11.22	41.49 ± 11.79	34.06 ± 7.51	57.50	47.43	70.28	49.69
(k=5, uzaklık metriği=chebyshev)	50.00 ± 3.73	32.78 ± 8.95	32.35 ± 6.40	28.89 ± 5.40	60.00	46.66	64.31	48.91
(k=7, uzaklık metriği=euclidean)	50.00 ± 5.89	32.20 ± 4.52	40.70 ± 11.20	29.78 ± 4.94	58.75	47.77	70.69	50.40
(k=7, uzaklık metriği=manhattan)	44.17 ± 4.25	34.65 ± 11.39	37.37 ± 14.65	28.56 ± 9.21	53.75	43.97	66.67	45.21
(k=7, uzaklık metriği=chebyshev)	49.17 ± 5.53	36.30 ± 9.04	36.52 ± 8.14	31.64 ± 7.44	57.50	45.46	63.47	47.37

Tablo 48. ADASYN ile dengelenmiş QCM sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	70.00 ± 7.64	34.20 ± 19.06	32.81 ± 14.37	32.81 ± 15.92	71.25	49.76	53.61	51.22
(ağaç sayısı=20, maksimum derinlik=10)	73.33 ± 3.33	34.78 ± 18.74	33.92 ± 13.42	33.76 ± 15.30	71.25	46.66	47.50	47.03
(ağaç sayısı=20, maksimum derinlik=20)	73.33 ± 3.33	34.78 ± 18.74	33.92 ± 13.42	33.76 ± 15.30	70.00	45.45	47.08	46.13
(ağaç sayısı=50, maksimum derinlik=5)	73.33 ± 8.58	33.90 ± 19.13	33.90 ± 14.58	33.38 ± 16.00	70.00	44.38	45.69	44.93
(ağaç sayısı=50, maksimum derinlik=10)	74.17 ± 4.86	29.66 ± 9.07	31.96 ± 9.45	30.60 ± 9.18	76.25	58.47	53.89	55.20
(ağaç sayısı=50, maksimum derinlik=20)	74.17 ± 4.86	29.66 ± 9.07	31.96 ± 9.45	30.60 ± 9.18	76.25	58.47	53.89	55.20
(ağaç sayısı=100, maksimum derinlik=5)	74.17 ± 6.12	34.60 ± 18.90	34.20 ± 13.62	33.76 ± 15.46	68.75	42.96	45.28	43.98
(ağaç sayısı=100, maksimum derinlik=10)	75.00 ± 3.73	29.48 ± 9.21	32.24 ± 9.17	30.63 ± 9.14	78.75	62.71	59.31	60.60
(ağaç sayısı=100, maksimum derinlik=20)	75.00 ± 3.73	29.48 ± 9.21	32.24 ± 9.17	30.63 ± 9.14	78.75	62.71	59.31	60.60

4.11. Dengesiz Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 49. Dengesiz hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	86.67 ± 6.12	67.92 ± 13.67	68.06 ± 17.30	65.71 ± 14.69	86.25	61.79	70.56	64.01
(C=0.1, optimizasyon yöntemi=saga)	85.83 ± 5.65	60.83 ± 14.58	65.56 ± 15.30	60.93 ± 13.26	86.25	67.08	72.92	67.14
(C=0.1, optimizasyon yöntemi=lbgfs)	89.17 ± 5.00	68.00 ± 13.65	68.06 ± 15.98	66.44 ± 13.88	92.50	83.64	81.11	79.15
(C=1, optimizasyon yöntemi=liblinear)	85.00 ± 5.65	63.67 ± 15.32	63.89 ± 14.78	60.93 ± 13.49	85.00	61.16	68.75	64.13
(C=1, optimizasyon yöntemi=saga)	85.00 ± 5.65	57.83 ± 15.44	63.89 ± 14.78	58.43 ± 13.26	86.25	68.75	75.28	69.54
(C=1, optimizasyon yöntemi=lbgfs)	90.83 ± 4.86	75.97 ± 14.16	72.22 ± 15.61	72.03 ± 14.38	93.75	82.29	84.72	82.25
(C=10, optimizasyon yöntemi=liblinear)	84.17 ± 6.12	59.92 ± 16.59	62.22 ± 16.33	58.86 ± 15.10	85.00	62.15	68.75	64.70
(C=10, optimizasyon yöntemi=saga)	84.17 ± 4.08	56.17 ± 12.28	63.63 ± 14.36	57.29 ± 11.15	86.25	67.32	75.28	68.84
(C=10, optimizasyon yöntemi=lbgfs)	90.00 ± 4.25	75.29 ± 13.35	70.56 ± 14.28	70.94 ± 13.29	93.75	82.29	84.72	82.25
(C=100, optimizasyon yöntemi=liblinear)	81.67 ± 3.33	51.25 ± 3.63	60.57 ± 13.30	52.75 ± 5.50	85.00	61.79	67.36	64.05
(C=100, optimizasyon yöntemi=saga)	84.17 ± 4.08	56.17 ± 12.28	63.63 ± 14.36	57.29 ± 11.15	86.25	67.32	75.28	68.84
(C=100, optimizasyon yöntemi=lbgfs)	86.67 ± 6.12	70.21 ± 15.40	66.67 ± 16.03	64.97 ± 15.62	92.50	80.56	84.31	80.85

Tablo 50. Dengesiz hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	90.00 ± 6.24	67.50 ± 17.56	72.78 ± 15.46	67.38 ± 16.96	96.25	91.67	91.67	90.00
(C=0.1, çekirdek fonksiyonu=poly)	77.50 ± 2.04	29.28 ± 12.48	28.33 ± 4.08	26.77 ± 6.28	75.00	18.75	25.00	21.43
(C=0.1, çekirdek fonksiyonu=rbf)	75.83 ± 1.67	18.96 ± 0.42	25.00 ± 0.00	21.56 ± 0.27	75.00	18.75	25.00	21.43
(C=1, çekirdek fonksiyonu=linear)	90.00 ± 6.24	67.50 ± 17.56	72.78 ± 15.46	67.38 ± 16.96	96.25	91.67	91.67	90.00
(C=1, çekirdek fonksiyonu=poly)	76.67 ± 3.33	29.23 ± 12.53	28.06 ± 4.34	26.63 ± 6.40	75.00	18.75	25.00	21.43
(C=1, çekirdek fonksiyonu=rbf)	77.50 ± 2.04	26.32 ± 10.02	28.33 ± 4.08	26.16 ± 5.61	85.00	62.06	50.00	51.73
(C=10, çekirdek fonksiyonu=linear)	90.00 ± 6.24	67.50 ± 17.56	72.78 ± 15.46	67.38 ± 16.96	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=poly)	80.83 ± 3.33	36.70 ± 11.33	37.22 ± 8.89	35.65 ± 8.62	78.75	30.80	37.50	33.79
(C=10, çekirdek fonksiyonu=rbf)	90.00 ± 2.04	63.83 ± 9.29	68.33 ± 10.07	64.25 ± 8.91	95.00	90.00	88.89	86.61
(C=100, çekirdek fonksiyonu=linear)	90.00 ± 6.24	67.50 ± 17.56	72.78 ± 15.46	67.38 ± 16.96	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=poly)	85.83 ± 3.33	59.22 ± 15.72	57.22 ± 8.21	55.60 ± 9.18	92.50	90.10	79.72	83.92
(C=100, çekirdek fonksiyonu=rbf)	90.00 ± 5.00	68.33 ± 17.60	69.17 ± 16.37	65.67 ± 17.44	96.25	91.67	91.67	90.00

Tablo 51. Dengesiz hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	84.17 ± 3.12	56.76 ± 13.27	56.67 ± 13.28	54.80 ± 11.35	92.50	84.10	88.06	84.99
(k=1, uzaklık metriği=manhattan)	85.83 ± 5.00	51.83 ± 13.85	58.61 ± 12.37	53.26 ± 11.92	93.75	87.28	88.47	86.85
(k=1, uzaklık metriği=chebyshev)	85.83 ± 4.25	56.70 ± 18.32	56.39 ± 14.87	55.29 ± 15.53	86.25	68.01	75.28	70.24
(k=3, uzaklık metriği=euclidean)	84.17 ± 4.08	57.83 ± 23.88	55.56 ± 12.20	53.88 ± 16.27	90.00	83.77	80.28	78.10
(k=3, uzaklık metriği=manhattan)	80.00 ± 4.86	40.28 ± 12.88	42.22 ± 12.87	40.56 ± 12.56	88.75	85.89	76.11	77.00
(k=3, uzaklık metriği=chebyshev)	78.33 ± 4.86	33.47 ± 9.19	41.68 ± 10.10	36.06 ± 9.10	87.50	72.49	75.56	69.32
(k=5, uzaklık metriği=euclidean)	81.67 ± 4.25	43.99 ± 12.41	45.56 ± 10.56	43.61 ± 10.15	82.50	50.17	60.83	54.54
(k=5, uzaklık metriği=manhattan)	80.83 ± 4.25	51.04 ± 12.74	49.44 ± 11.20	49.24 ± 11.32	86.25	54.37	69.58	60.65
(k=5, uzaklık metriği=chebyshev)	82.50 ± 6.12	52.93 ± 20.50	51.94 ± 15.53	50.36 ± 16.83	82.50	49.14	60.83	53.72
(k=7, uzaklık metriği=euclidean)	81.67 ± 5.65	38.00 ± 13.33	41.67 ± 13.18	39.47 ± 13.16	82.50	54.13	54.86	54.09
(k=7, uzaklık metriği=manhattan)	79.17 ± 4.56	31.08 ± 6.48	42.22 ± 10.49	34.71 ± 7.28	83.75	57.85	56.11	54.64
(k=7, uzaklık metriği=chebyshev)	82.50 ± 5.53	42.96 ± 10.64	48.06 ± 11.74	44.67 ± 10.84	86.25	67.60	60.00	61.69

Tablo 52. Dengesiz hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	78.33 ± 1.67	34.63 ± 12.55	31.67 ± 5.65	31.15 ± 7.84	86.25	88.10	57.22	66.33
(ağaç sayısı=20, maksimum derinlik=10)	80.83 ± 2.04	43.99 ± 3.04	42.22 ± 6.72	41.66 ± 5.04	87.50	87.19	63.75	71.26
(ağaç sayısı=20, maksimum derinlik=20)	80.83 ± 2.04	43.99 ± 3.04	42.22 ± 6.72	41.66 ± 5.04	87.50	87.19	63.75	71.26
(ağaç sayısı=50, maksimum derinlik=5)	78.33 ± 3.12	32.31 ± 11.99	34.17 ± 9.28	32.43 ± 10.18	85.00	67.56	51.67	55.94
(ağaç sayısı=50, maksimum derinlik=10)	79.17 ± 2.64	37.82 ± 10.34	36.67 ± 8.08	36.08 ± 8.54	87.50	88.41	58.61	64.77
(ağaç sayısı=50, maksimum derinlik=20)	79.17 ± 2.64	37.82 ± 10.34	36.67 ± 8.08	36.08 ± 8.54	87.50	88.41	58.61	64.77
(ağaç sayısı=100, maksimum derinlik=5)	77.50 ± 3.33	27.48 ± 10.47	29.17 ± 5.27	27.53 ± 7.16	82.50	65.55	46.11	51.13
(ağaç sayısı=100, maksimum derinlik=10)	79.17 ± 2.64	34.07 ± 10.57	35.83 ± 8.16	34.32 ± 8.82	87.50	96.43	60.00	69.86
(ağaç sayısı=100, maksimum derinlik=20)	79.17 ± 2.64	34.07 ± 10.57	35.83 ± 8.16	34.32 ± 8.82	87.50	96.43	60.00	69.86

4.12. SMOTE ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 53. SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	90.83 ± 4.86	73.33 ± 14.81	73.89 ± 15.60	71.19 ± 13.68	91.25	79.51	79.17	73.40
(C=0.1, optimizasyon yöntemi=saga)	90.83 ± 6.67	75.00 ± 18.45	73.06 ± 20.61	72.69 ± 19.43	92.50	84.47	85.69	80.55
(C=0.1, optimizasyon yöntemi=lbgfs)	92.50 ± 4.86	76.67 ± 16.58	78.06 ± 14.99	76.02 ± 15.71	93.75	83.33	88.47	84.18
(C=1, optimizasyon yöntemi=liblinear)	91.67 ± 5.27	79.58 ± 13.07	78.06 ± 16.75	75.81 ± 14.25	91.25	79.51	79.17	73.40
(C=1, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	69.72 ± 17.02	69.06 ± 15.23	93.75	83.33	88.47	84.18
(C=1, optimizasyon yöntemi=lbgfs)	90.83 ± 4.86	76.96 ± 15.40	73.89 ± 16.68	73.42 ± 16.06	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=liblinear)	90.83 ± 4.86	73.83 ± 11.84	75.56 ± 16.07	71.94 ± 12.96	91.25	79.51	79.17	73.40
(C=10, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	69.72 ± 17.02	69.06 ± 15.23	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=lbgfs)	90.00 ± 5.65	76.13 ± 16.32	71.39 ± 18.60	72.09 ± 17.32	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=liblinear)	90.83 ± 4.86	73.83 ± 11.84	75.56 ± 16.07	71.94 ± 12.96	88.75	68.12	74.58	69.02
(C=100, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	69.72 ± 17.02	69.06 ± 15.23	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=lbgfs)	90.00 ± 5.65	72.50 ± 16.16	71.39 ± 18.60	69.69 ± 16.48	96.25	91.67	91.67	90.00

Tablo 54. SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=0.1, çekirdek fonksiyonu=poly)	28.33 ± 7.64	47.13 ± 18.24	41.33 ± 13.58	26.01 ± 14.20	23.75	27.24	30.42	13.01
(C=0.1, çekirdek fonksiyonu=rbf)	80.83 ± 3.33	49.35 ± 10.26	55.85 ± 13.37	49.48 ± 9.84	81.25	71.90	70.28	58.46
(C=1, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=1, çekirdek fonksiyonu=poly)	60.00 ± 8.16	51.08 ± 15.41	54.55 ± 8.46	41.65 ± 7.99	63.75	54.29	57.50	46.66
(C=1, çekirdek fonksiyonu=rbf)	90.83 ± 1.67	69.74 ± 13.38	71.67 ± 8.08	68.53 ± 10.28	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=poly)	84.17 ± 4.86	58.10 ± 13.07	65.83 ± 10.71	57.99 ± 10.30	90.00	81.22	89.58	84.12
(C=10, çekirdek fonksiyonu=rbf)	89.17 ± 2.04	67.24 ± 14.58	66.67 ± 10.54	64.53 ± 11.71	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=poly)	88.33 ± 3.12	62.51 ± 6.52	66.94 ± 5.72	61.31 ± 2.96	95.00	87.05	87.50	86.61
(C=100, çekirdek fonksiyonu=rbf)	89.17 ± 2.04	67.24 ± 14.58	66.67 ± 10.54	64.53 ± 11.71	96.25	91.67	91.67	90.00

Tablo 55. SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=öklid)	87.50 ± 2.64	70.00 ± 7.77	65.83 ± 8.81	63.35 ± 6.05	91.25	81.24	85.28	81.45
(k=1, uzaklık metriği=manhattan)	87.50 ± 2.64	60.83 ± 7.98	65.83 ± 8.81	60.68 ± 5.64	92.50	84.10	88.06	84.99
(k=1, uzaklık metriği=chebyshev)	86.67 ± 3.12	61.94 ± 9.77	68.89 ± 7.11	62.01 ± 6.52	90.00	75.91	83.47	78.92
(k=3, uzaklık metriği=öklid)	87.50 ± 2.64	71.89 ± 7.84	71.94 ± 9.19	67.41 ± 7.17	87.50	73.31	84.03	76.58
(k=3, uzaklık metriği=manhattan)	84.17 ± 4.86	63.98 ± 13.28	64.74 ± 9.74	60.11 ± 8.52	85.00	70.51	85.56	75.21
(k=3, uzaklık metriği=chebyshev)	83.33 ± 2.64	60.22 ± 11.38	71.43 ± 10.19	60.68 ± 9.07	82.50	65.06	76.39	69.32
(k=5, uzaklık metriği=öklid)	81.67 ± 5.65	60.08 ± 9.78	70.03 ± 9.33	59.98 ± 8.26	80.00	68.44	81.53	70.54
(k=5, uzaklık metriği=manhattan)	80.00 ± 7.17	60.39 ± 10.86	64.75 ± 12.61	57.80 ± 9.78	81.25	66.46	84.31	71.21
(k=5, uzaklık metriği=chebyshev)	77.50 ± 3.33	55.54 ± 11.10	67.30 ± 8.69	54.90 ± 7.97	81.25	63.38	75.14	66.93
(k=7, uzaklık metriği=öklid)	79.17 ± 5.27	55.58 ± 12.38	66.99 ± 9.76	55.42 ± 10.07	78.75	66.78	83.47	70.34
(k=7, uzaklık metriği=manhattan)	77.50 ± 4.25	54.60 ± 9.11	65.04 ± 11.59	55.47 ± 8.32	80.00	66.24	83.89	69.79
(k=7, uzaklık metriği=chebyshev)	74.17 ± 3.12	49.10 ± 7.21	66.18 ± 9.58	51.74 ± 7.82	80.00	61.61	74.72	65.05

Tablo 56. SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	85.00 ± 5.65	64.67 ± 23.05	56.40 ± 18.49	57.68 ± 19.46	92.50	86.31	80.56	82.72
(ağaç sayısı=20, maksimum derinlik=10)	85.00 ± 4.25	54.02 ± 12.39	53.07 ± 10.46	51.79 ± 10.31	91.25	86.53	73.33	78.36
(ağaç sayısı=20, maksimum derinlik=20)	85.00 ± 4.25	54.02 ± 12.39	53.07 ± 10.46	51.79 ± 10.31	91.25	86.53	73.33	78.36
(ağaç sayısı=50, maksimum derinlik=5)	85.00 ± 5.65	62.17 ± 24.87	53.33 ± 17.95	54.14 ± 19.97	93.75	91.07	82.50	86.20
(ağaç sayısı=50, maksimum derinlik=10)	86.67 ± 5.53	62.60 ± 21.13	58.33 ± 18.63	57.72 ± 19.83	92.50	85.48	79.72	81.89
(ağaç sayısı=50, maksimum derinlik=20)	86.67 ± 5.53	62.60 ± 21.13	58.33 ± 18.63	57.72 ± 19.83	92.50	85.48	79.72	81.89
(ağaç sayısı=100, maksimum derinlik=5)	84.17 ± 3.12	57.54 ± 19.34	51.67 ± 12.80	51.89 ± 14.11	92.50	85.48	79.72	81.89
(ağaç sayısı=100, maksimum derinlik=10)	85.00 ± 4.25	58.61 ± 20.91	53.33 ± 15.46	54.19 ± 16.89	92.50	89.27	79.72	83.59
(ağaç sayısı=100, maksimum derinlik=20)	85.00 ± 4.25	58.61 ± 20.91	53.33 ± 15.46	54.19 ± 16.89	92.50	89.27	79.72	83.59

4.13. Borderline SMOTE ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 57. Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	89.17 ± 4.25	67.08 ± 14.93	68.89 ± 12.32	65.14 ± 12.35	90.00	71.92	74.17	69.87
(C=0.1, optimizasyon yöntemi=saga)	91.67 ± 5.89	77.22 ± 15.85	75.56 ± 17.71	75.22 ± 16.70	92.50	79.52	83.47	80.41
(C=0.1, optimizasyon yöntemi=lbfsgs)	92.50 ± 4.86	78.06 ± 14.84	78.06 ± 14.99	76.56 ± 15.11	92.50	79.52	83.47	80.41
(C=1, optimizasyon yöntemi=liblinear)	90.00 ± 4.25	73.00 ± 10.96	73.89 ± 14.68	69.61 ± 11.00	90.00	71.92	74.17	69.87
(C=1, optimizasyon yöntemi=saga)	90.00 ± 5.65	74.74 ± 18.09	71.39 ± 18.60	71.56 ± 17.91	93.75	83.33	88.47	84.18
(C=1, optimizasyon yöntemi=lbfsgs)	92.50 ± 4.86	78.06 ± 14.84	78.06 ± 14.99	76.56 ± 15.11	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=liblinear)	90.00 ± 4.25	73.00 ± 10.96	73.89 ± 14.68	69.61 ± 11.00	91.25	75.00	79.31	76.31
(C=10, optimizasyon yöntemi=saga)	90.00 ± 5.65	74.74 ± 18.09	71.39 ± 18.60	71.56 ± 17.91	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=lbfsgs)	90.00 ± 5.65	76.13 ± 16.32	71.39 ± 18.60	72.09 ± 17.32	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=liblinear)	90.00 ± 4.25	73.00 ± 10.96	73.89 ± 14.68	69.61 ± 11.00	91.25	75.15	80.14	76.31
(C=100, optimizasyon yöntemi=saga)	90.00 ± 5.65	74.74 ± 18.09	71.39 ± 18.60	71.56 ± 17.91	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=lbfsgs)	90.00 ± 6.77	71.67 ± 18.89	71.39 ± 20.39	69.69 ± 19.36	95.00	87.50	88.89	85.58

Tablo 58. Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=0.1, çekirdek fonksiyonu=poly)	28.33 ± 8.50	47.13 ± 18.21	41.32 ± 13.50	25.89 ± 14.07	25.00	27.31	30.83	13.68
(C=0.1, çekirdek fonksiyonu=rbf)	80.00 ± 4.08	51.34 ± 13.05	56.97 ± 10.94	50.12 ± 8.86	81.25	70.93	70.28	57.56
(C=1, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=1, çekirdek fonksiyonu=poly)	60.00 ± 7.73	53.25 ± 21.94	53.96 ± 12.50	42.23 ± 13.41	63.75	54.41	52.92	42.66
(C=1, çekirdek fonksiyonu=rbf)	88.33 ± 1.67	60.57 ± 9.69	61.67 ± 9.28	58.58 ± 8.01	95.00	88.10	86.67	85.97
(C=10, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=poly)	83.33 ± 8.74	67.23 ± 12.82	64.44 ± 14.13	60.66 ± 14.10	88.75	78.03	80.83	77.73
(C=10, çekirdek fonksiyonu=rbf)	90.00 ± 2.04	72.24 ± 11.84	69.17 ± 10.74	68.70 ± 10.96	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=poly)	88.33 ± 4.08	67.13 ± 16.57	63.89 ± 13.55	61.98 ± 13.13	92.50	82.29	77.50	78.27
(C=100, çekirdek fonksiyonu=rbf)	90.00 ± 2.04	72.24 ± 11.84	69.17 ± 10.74	68.70 ± 10.96	96.25	91.67	91.67	90.00

Tablo 59. Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=öklid)	87.50 ± 2.64	70.00 ± 7.77	65.83 ± 8.81	63.35 ± 6.05	90.00	77.07	84.86	78.96
(k=1, uzaklık metriği=manhattan)	86.67 ± 3.12	64.68 ± 15.26	65.56 ± 13.13	61.60 ± 10.99	91.25	79.93	87.64	82.50
(k=1, uzaklık metriği=chebyshev)	85.83 ± 2.04	59.86 ± 12.05	63.07 ± 8.40	57.86 ± 9.19	86.25	72.17	75.28	71.77
(k=3, uzaklık metriği=öklid)	87.50 ± 2.64	69.60 ± 9.22	69.44 ± 12.79	65.32 ± 8.65	81.25	65.72	77.36	68.25
(k=3, uzaklık metriği=manhattan)	85.83 ± 4.25	66.20 ± 11.00	64.17 ± 8.35	62.35 ± 7.59	82.50	67.09	80.14	72.00
(k=3, uzaklık metriği=chebyshev)	82.50 ± 4.86	54.08 ± 9.84	61.14 ± 7.40	54.51 ± 7.85	78.75	65.18	72.78	65.20
(k=5, uzaklık metriği=öklid)	80.83 ± 5.00	54.73 ± 10.85	65.01 ± 12.21	55.68 ± 9.66	76.25	63.27	75.69	64.74
(k=5, uzaklık metriği=manhattan)	81.67 ± 3.33	61.83 ± 8.77	65.29 ± 12.83	58.72 ± 8.90	73.75	61.60	74.86	64.21
(k=5, uzaklık metriği=chebyshev)	76.67 ± 5.65	49.62 ± 9.59	57.85 ± 5.36	49.54 ± 7.39	77.50	63.96	74.72	65.47
(k=7, uzaklık metriği=öklid)	80.00 ± 3.12	55.06 ± 9.88	64.75 ± 11.83	55.70 ± 8.69	73.75	61.51	77.22	64.29
(k=7, uzaklık metriği=manhattan)	80.83 ± 4.25	61.46 ± 9.36	65.01 ± 13.08	58.33 ± 9.12	71.25	60.47	74.03	62.64
(k=7, uzaklık metriği=chebyshev)	74.17 ± 4.86	46.60 ± 9.03	55.63 ± 7.61	46.74 ± 6.43	75.00	61.22	73.89	63.54

Tablo 60. Borderline SMOTE ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	82.50 ± 4.86	53.11 ± 14.05	48.33 ± 8.62	49.34 ± 10.39	85.00	69.89	66.53	68.00
(ağaç sayısı=20, maksimum derinlik=10)	81.67 ± 4.25	54.08 ± 16.38	51.39 ± 13.23	50.98 ± 13.29	88.75	89.02	72.36	78.97
(ağaç sayısı=20, maksimum derinlik=20)	81.67 ± 4.25	54.08 ± 16.38	51.39 ± 13.23	50.98 ± 13.29	88.75	89.02	72.36	78.97
(ağaç sayısı=50, maksimum derinlik=5)	83.33 ± 3.73	58.70 ± 15.83	50.29 ± 12.92	52.68 ± 13.82	88.75	71.47	73.75	71.13
(ağaç sayısı=50, maksimum derinlik=10)	82.50 ± 6.12	55.39 ± 26.23	46.39 ± 17.38	48.98 ± 20.32	88.75	76.75	73.75	74.82
(ağaç sayısı=50, maksimum derinlik=20)	82.50 ± 6.12	55.39 ± 26.23	46.39 ± 17.38	48.98 ± 20.32	88.75	76.75	73.75	74.82
(ağaç sayısı=100, maksimum derinlik=5)	84.17 ± 5.53	56.26 ± 25.63	48.33 ± 16.37	50.45 ± 19.47	91.25	79.61	78.33	76.13
(ağaç sayısı=100, maksimum derinlik=10)	83.33 ± 5.89	55.23 ± 26.49	46.67 ± 17.36	49.01 ± 20.48	91.25	81.31	78.33	77.36
(ağaç sayısı=100, maksimum derinlik=20)	83.33 ± 5.89	55.23 ± 26.49	46.67 ± 17.36	49.01 ± 20.48	91.25	81.31	78.33	77.36

4.14. SMOTE-Tomek ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 61. SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	88.33 ± 4.08	65.00 ± 11.96	66.39 ± 14.20	62.69 ± 10.97	91.25	79.51	79.17	73.40
(C=0.1, optimizasyon yöntemi=saga)	90.83 ± 6.67	75.00 ± 18.45	73.06 ± 20.61	72.69 ± 19.43	92.50	80.83	85.69	80.55
(C=0.1, optimizasyon yöntemi=lbgfs)	92.50 ± 4.86	76.67 ± 16.58	78.06 ± 14.99	76.02 ± 15.71	92.50	80.83	85.69	80.55
(C=1, optimizasyon yöntemi=liblinear)	90.83 ± 4.86	73.83 ± 11.84	75.56 ± 16.07	71.94 ± 12.96	90.00	73.26	78.75	72.23
(C=1, optimizasyon yöntemi=saga)	89.17 ± 5.65	71.40 ± 18.50	69.72 ± 18.20	68.56 ± 17.64	95.00	87.50	88.89	85.58
(C=1, optimizasyon yöntemi=lbgfs)	91.67 ± 5.89	77.22 ± 15.78	75.56 ± 18.67	74.56 ± 17.62	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=liblinear)	90.00 ± 4.25	72.17 ± 10.28	73.89 ± 14.68	69.44 ± 10.92	91.25	76.25	81.53	76.16
(C=10, optimizasyon yöntemi=saga)	88.33 ± 4.86	69.74 ± 16.20	67.22 ± 15.70	66.22 ± 14.58	93.75	85.83	88.47	84.12
(C=10, optimizasyon yöntemi=lbgfs)	90.83 ± 4.86	76.96 ± 15.48	73.06 ± 16.49	73.28 ± 15.91	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=liblinear)	90.83 ± 4.86	73.83 ± 11.84	75.56 ± 16.07	71.94 ± 12.96	88.75	68.97	74.58	69.67
(C=100, optimizasyon yöntemi=saga)	90.00 ± 6.77	71.67 ± 18.89	71.39 ± 20.39	69.69 ± 19.36	93.75	85.83	88.47	84.12
(C=100, optimizasyon yöntemi=lbgfs)	89.17 ± 5.65	70.42 ± 17.10	68.89 ± 17.25	67.31 ± 15.91	93.75	85.83	88.47	84.12

Tablo 62. SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=0.1, çekirdek fonksiyonu=poly)	27.50 ± 7.73	47.10 ± 18.24	41.04 ± 13.65	25.54 ± 14.26	23.75	27.24	30.42	13.01
(C=0.1, çekirdek fonksiyonu=rbf)	79.17 ± 2.64	49.09 ± 9.51	56.70 ± 10.92	49.81 ± 7.71	77.50	69.10	69.03	55.29
(C=1, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=1, çekirdek fonksiyonu=poly)	55.00 ± 5.53	57.83 ± 18.00	53.71 ± 11.82	42.53 ± 12.22	52.50	53.41	44.58	32.75
(C=1, çekirdek fonksiyonu=rbf)	90.83 ± 1.67	69.74 ± 13.38	71.67 ± 8.08	68.53 ± 10.28	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=poly)	85.00 ± 7.26	62.00 ± 11.92	68.33 ± 11.50	61.61 ± 10.49	91.25	82.39	90.00	84.43
(C=10, çekirdek fonksiyonu=rbf)	89.17 ± 2.04	67.24 ± 14.58	66.67 ± 10.54	64.53 ± 11.71	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=linear)	91.67 ± 5.89	75.00 ± 17.28	75.56 ± 16.49	72.36 ± 17.29	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=poly)	85.83 ± 2.04	60.43 ± 6.95	61.67 ± 5.17	57.02 ± 2.00	95.00	87.05	87.50	86.61
(C=100, çekirdek fonksiyonu=rbf)	90.00 ± 3.33	70.57 ± 17.57	69.17 ± 14.09	67.70 ± 15.44	96.25	91.67	91.67	90.00

Tablo 63. SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	87.50 ± 2.64	70.00 ± 7.77	65.83 ± 8.81	63.35 ± 6.05	90.00	77.07	84.86	78.96
(k=1, uzaklık metriği=manhattan)	88.33 ± 3.12	69.13 ± 10.58	67.50 ± 11.44	64.21 ± 8.36	90.00	79.87	84.86	80.13
(k=1, uzaklık metriği=chebyshev)	82.50 ± 3.12	52.81 ± 7.20	55.85 ± 7.33	52.01 ± 5.72	88.75	73.87	77.64	74.71
(k=3, uzaklık metriği=euclidean)	87.50 ± 2.64	70.22 ± 8.72	71.94 ± 9.19	66.22 ± 6.65	85.00	71.19	83.19	74.26
(k=3, uzaklık metriği=manhattan)	85.00 ± 7.73	65.92 ± 12.20	73.61 ± 6.21	64.53 ± 7.54	86.25	72.90	85.97	77.51
(k=3, uzaklık metriği=chebyshev)	79.17 ± 5.89	51.55 ± 14.52	58.65 ± 8.16	51.30 ± 11.58	87.50	73.31	84.03	76.58
(k=5, uzaklık metriği=euclidean)	82.50 ± 4.86	55.08 ± 11.98	65.85 ± 13.39	56.60 ± 11.08	80.00	69.77	83.89	72.49
(k=5, uzaklık metriği=manhattan)	81.67 ± 5.65	62.00 ± 8.48	70.03 ± 8.73	61.53 ± 8.28	81.25	66.96	84.31	70.75
(k=5, uzaklık metriği=chebyshev)	81.67 ± 5.65	60.79 ± 13.72	68.64 ± 12.36	59.27 ± 12.50	81.25	64.44	80.56	69.53
(k=7, uzaklık metriği=euclidean)	79.17 ± 3.73	55.85 ± 10.68	65.60 ± 7.48	55.61 ± 8.17	75.00	62.29	82.22	66.57
(k=7, uzaklık metriği=manhattan)	79.17 ± 5.27	59.48 ± 7.22	69.21 ± 8.68	58.58 ± 6.49	76.25	59.64	77.92	63.41
(k=7, uzaklık metriği=chebyshev)	73.33 ± 3.33	52.42 ± 11.14	65.91 ± 11.10	52.67 ± 9.41	78.75	64.22	81.11	68.25

Tablo 64. SMOTE-Tomek ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performans değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	80.83 ± 6.24	48.91 ± 25.95	44.44 ± 18.76	44.41 ± 20.23	86.25	69.58	71.53	69.55
(ağaç sayısı=20, maksimum derinlik=10)	86.67 ± 7.17	61.02 ± 20.51	65.56 ± 17.82	60.61 ± 17.97	95.00	91.07	86.11	88.36
(ağaç sayısı=20, maksimum derinlik=20)	87.50 ± 5.89	69.36 ± 20.71	64.72 ± 21.17	63.93 ± 19.88	93.75	89.64	83.33	85.30
(ağaç sayısı=50, maksimum derinlik=5)	87.50 ± 5.27	60.68 ± 21.60	61.67 ± 17.36	58.66 ± 18.99	88.75	78.49	68.61	72.33
(ağaç sayısı=50, maksimum derinlik=10)	85.83 ± 7.26	66.32 ± 27.25	60.56 ± 22.70	62.00 ± 24.11	90.00	77.14	72.78	73.52
(ağaç sayısı=50, maksimum derinlik=20)	85.83 ± 5.65	58.14 ± 22.26	58.33 ± 18.07	55.52 ± 18.03	91.25	88.91	76.94	81.34
(ağaç sayısı=100, maksimum derinlik=5)	85.83 ± 5.65	60.49 ± 21.01	59.74 ± 18.91	57.56 ± 18.52	91.25	77.94	74.17	75.87
(ağaç sayısı=100, maksimum derinlik=10)	85.00 ± 4.25	63.08 ± 14.20	54.17 ± 14.19	56.33 ± 13.42	93.75	90.24	83.33	86.32
(ağaç sayısı=100, maksimum derinlik=20)	82.50 ± 4.86	45.12 ± 15.25	45.56 ± 12.65	43.94 ± 12.26	92.50	93.91	81.94	84.84

4.15. ADASYN ile Dengelenmiş Hibrit (MOS+QCM) Sensör Verilerine Ait Sınıflandırma Sonuçları

Tablo 65. ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile LR performansının değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, optimizasyon yöntemi=liblinear)	89.17 ± 5.00	70.83 ± 14.19	68.89 ± 16.42	66.86 ± 13.72	91.25	79.51	79.17	73.40
(C=0.1, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	68.89 ± 16.42	69.22 ± 15.37	91.25	79.47	85.28	79.24
(C=0.1, optimizasyon yöntemi=lbgfs)	91.67 ± 3.73	75.42 ± 14.93	75.56 ± 11.84	73.64 ± 12.46	92.50	81.67	88.06	82.72
(C=1, optimizasyon yöntemi=liblinear)	90.00 ± 5.65	73.42 ± 11.97	73.89 ± 16.18	70.11 ± 13.60	90.00	73.26	78.75	72.23
(C=1, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	68.89 ± 16.42	69.22 ± 15.37	93.75	83.33	88.47	84.18
(C=1, optimizasyon yöntemi=lbgfs)	90.00 ± 4.25	75.29 ± 13.26	71.39 ± 14.64	71.09 ± 13.50	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=liblinear)	90.00 ± 5.65	73.42 ± 11.97	73.89 ± 16.18	70.11 ± 13.60	91.25	76.25	81.53	76.16
(C=10, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	68.89 ± 16.42	69.22 ± 15.37	93.75	83.33	88.47	84.18
(C=10, optimizasyon yöntemi=lbgfs)	89.17 ± 5.00	74.46 ± 14.22	68.89 ± 16.42	69.75 ± 14.77	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=liblinear)	90.00 ± 5.65	73.42 ± 11.97	73.89 ± 16.18	70.11 ± 13.60	90.00	71.58	77.36	72.74
(C=100, optimizasyon yöntemi=saga)	89.17 ± 5.00	73.07 ± 16.08	68.89 ± 16.42	69.22 ± 15.37	93.75	83.33	88.47	84.18
(C=100, optimizasyon yöntemi=lbgfs)	89.17 ± 5.00	66.25 ± 16.16	68.06 ± 15.98	64.98 ± 14.50	93.75	83.33	88.47	84.18

Tablo 66. ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile SVM performansının değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(C=0.1, çekirdek fonksiyonu=linear)	90.83 ± 5.53	68.33 ± 17.60	73.06 ± 15.41	68.02 ± 16.70	96.25	91.67	91.67	90.00
(C=0.1, çekirdek fonksiyonu=poly)	31.67 ± 5.00	47.24 ± 18.24	42.43 ± 13.14	27.73 ± 13.55	26.25	39.88	35.83	20.95
(C=0.1, çekirdek fonksiyonu=rbf)	79.17 ± 5.89	42.12 ± 14.91	50.29 ± 14.05	43.36 ± 11.92	85.00	74.89	71.53	61.78
(C=1, çekirdek fonksiyonu=linear)	90.83 ± 5.53	68.33 ± 17.60	73.06 ± 15.41	68.02 ± 16.70	96.25	91.67	91.67	90.00
(C=1, çekirdek fonksiyonu=poly)	63.33 ± 13.02	45.90 ± 9.31	52.00 ± 6.12	39.03 ± 6.46	62.50	53.69	57.08	46.17
(C=1, çekirdek fonksiyonu=rbf)	89.17 ± 2.04	68.07 ± 15.11	66.67 ± 10.54	65.53 ± 12.67	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=linear)	90.83 ± 5.53	68.33 ± 17.60	73.06 ± 15.41	68.02 ± 16.70	96.25	91.67	91.67	90.00
(C=10, çekirdek fonksiyonu=poly)	81.67 ± 7.26	61.58 ± 13.31	67.24 ± 11.28	60.53 ± 11.68	88.75	77.53	85.42	80.58
(C=10, çekirdek fonksiyonu=rbf)	89.17 ± 2.04	67.24 ± 14.58	66.67 ± 10.54	64.53 ± 11.71	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=linear)	90.83 ± 5.53	68.33 ± 17.60	73.06 ± 15.41	68.02 ± 16.70	96.25	91.67	91.67	90.00
(C=100, çekirdek fonksiyonu=poly)	87.50 ± 6.97	72.62 ± 18.32	71.39 ± 14.02	67.74 ± 16.49	92.50	82.29	86.67	84.10
(C=100, çekirdek fonksiyonu=rbf)	89.17 ± 2.04	67.24 ± 14.58	66.67 ± 10.54	64.53 ± 11.71	96.25	91.67	91.67	90.00

Tablo 67. ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile k-NN performansının değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(k=1, uzaklık metriği=euclidean)	87.50 ± 2.64	67.50 ± 10.42	65.83 ± 8.81	62.52 ± 6.99	91.25	81.24	85.28	81.45
(k=1, uzaklık metriği=manhattan)	85.83 ± 3.33	60.78 ± 10.98	62.22 ± 11.86	58.39 ± 9.06	88.75	76.64	84.44	78.54
(k=1, uzaklık metriği=chebyshev)	83.33 ± 4.56	54.73 ± 7.83	58.61 ± 6.66	53.61 ± 3.06	92.50	85.85	85.69	83.67
(k=3, uzaklık metriği=euclidean)	88.33 ± 1.67	69.67 ± 10.06	72.22 ± 9.34	66.34 ± 7.41	86.25	72.16	83.61	75.37
(k=3, uzaklık metriği=manhattan)	83.33 ± 4.56	56.09 ± 8.67	59.18 ± 11.56	55.20 ± 6.66	82.50	65.97	82.36	70.29
(k=3, uzaklık metriği=chebyshev)	80.83 ± 4.25	52.16 ± 9.91	62.53 ± 3.81	53.57 ± 5.83	85.00	74.54	83.19	75.75
(k=5, uzaklık metriği=euclidean)	81.67 ± 5.65	62.17 ± 10.28	70.03 ± 9.33	60.32 ± 8.33	80.00	68.44	81.53	70.54
(k=5, uzaklık metriği=manhattan)	78.33 ± 6.67	56.50 ± 10.99	58.38 ± 12.44	54.26 ± 9.12	78.75	63.32	83.47	69.15
(k=5, uzaklık metriği=chebyshev)	75.83 ± 5.53	50.44 ± 9.51	62.30 ± 5.28	50.77 ± 5.16	81.25	66.65	81.94	70.14
(k=7, uzaklık metriği=euclidean)	80.83 ± 3.33	60.83 ± 8.27	71.99 ± 10.80	60.56 ± 7.17	77.50	67.52	83.06	70.61
(k=7, uzaklık metriği=manhattan)	75.83 ± 4.86	55.31 ± 10.68	57.54 ± 12.30	53.04 ± 8.39	76.25	61.45	82.64	67.24
(k=7, uzaklık metriği=chebyshev)	70.00 ± 7.17	48.11 ± 10.73	61.18 ± 8.42	48.05 ± 8.49	77.50	66.28	83.06	69.53

Tablo 68. ADASYN ile dengelenmiş hibrit (MOS+QCM) sensör verileri ile RF performansının değerlendirilmesi

Hiperparametreler	ÇD Doğruluk (%)	ÇD Kesinlik (%)	ÇD Duyarlılık (%)	ÇD F1-Skoru (%)	Test Doğruluk (%)	Test Kesinlik (%)	Test Duyarlılık (%)	Test F1-Skoru (%)
(ağaç sayısı=20, maksimum derinlik=5)	85.00 ± 4.25	64.59 ± 11.34	58.06 ± 9.52	58.65 ± 8.65	90.00	78.10	72.78	75.07
(ağaç sayısı=20, maksimum derinlik=10)	84.17 ± 3.12	51.49 ± 14.38	54.72 ± 10.71	52.14 ± 12.15	90.00	80.04	70.56	74.67
(ağaç sayısı=20, maksimum derinlik=20)	84.17 ± 3.12	51.49 ± 14.38	54.72 ± 10.71	52.14 ± 12.15	90.00	80.04	70.56	74.67
(ağaç sayısı=50, maksimum derinlik=5)	83.33 ± 6.45	50.81 ± 16.16	54.44 ± 16.72	51.34 ± 15.50	90.00	80.10	72.78	75.71
(ağaç sayısı=50, maksimum derinlik=10)	85.00 ± 3.33	56.15 ± 7.58	56.39 ± 12.32	54.42 ± 10.59	90.00	84.74	72.78	76.47
(ağaç sayısı=50, maksimum derinlik=20)	85.00 ± 3.33	56.15 ± 7.58	56.39 ± 12.32	54.42 ± 10.59	90.00	84.74	72.78	76.47
(ağaç sayısı=100, maksimum derinlik=5)	86.67 ± 4.08	68.08 ± 15.54	60.00 ± 15.05	60.85 ± 14.78	90.00	77.14	72.78	74.39
(ağaç sayısı=100, maksimum derinlik=10)	85.83 ± 5.65	59.33 ± 21.74	57.50 ± 17.56	56.59 ± 18.27	90.00	84.74	72.78	76.47
(ağaç sayısı=100, maksimum derinlik=20)	85.83 ± 5.65	59.33 ± 21.74	57.50 ± 17.56	56.59 ± 18.27	90.00	84.74	72.78	76.47

Yalnızca MOS veya yalnızca QCM sensörleriyle gerçekleştirilen deneyler, bu sensörlerin farklı hassasiyet ve tepki karakteristikleri nedeniyle sınıflar arasında ayırt edici performanslarının sınırlı kaldığını göstermiştir. Buna karşın, MOS ve QCM sensörlerinden elde edilen verilerin birleştirilmesiyle oluşturulan hibrit öznitelik uzayı, sınıfların daha başarılı biçimde ayrılmasını sağlamıştır.

MOS sensörüyle elde edilen en yüksek başarı, $k=3$ ve manhattan uzaklığı kullanılarak; %95 doğruluk, %86,11 makro kesinlik, %86,11 makro duyarlılık ve %86,11 makro F1 skoru olarak kaydedilmiştir. QCM sensörüyle elde edilen en iyi sonuç ise, $C=10$ ve lineer çekirdek fonksiyonu kullanılarak, %76,25 doğruluk, %62,52 makro kesinlik, %74,31 makro duyarlılık ve %65,74 makro F1 skoru şeklindedir. Her iki sensör tipinin birlikte kullanıldığı hibrit modelde ise, farklı hiperparametre kombinasyonlarıyla benzer başarılar elde edilmiş; en yüksek performans olarak %96,25 doğruluk, %91,67 makro kesinlik, %91,67 makro duyarlılık ve %90 makro F1 skoru gözlemlenmiştir. Bu değerlendirmede yalnızca dengesiz veriler karşılaştırılmıştır. Böylece SMOTE ve türevlerinin modeli nasıl etkilediğine dair yanlılık oluşturulmadan, sensör tiplerinin kendi ayırt edici kapasiteleri karşılaştırılmıştır.

Dengesiz veri seti ile elde edilen sonuçlar ile sınıf dengesizliğini gidermeye yönelik yöntemler uygulandıktan sonra elde edilen sonuçlar karşılaştırıldığında, performans metriklerinde genel bir iyileşme gözlemlenmiştir. Özellikle makro kesinlik, makro duyarlılık ve makro F1-skoru gibi sınıflar arası dengeyi gözetken metriklerdeki artış, modelin tüm sınıflara karşı daha adil ve dengeli bir performans sergilediğini göstermektedir. Dengesiz veri ile yapılan sınıflandırmalarda IV. Evre örnekleri baskın olduğu için bu sınıf büyük oranda doğru tahmin edilmekte; bu durum genel doğruluk oranını yükseltirken, azınlık sınıflar için yetersiz performansa yol açmakta ve makro bazlı metriklerin düşük kalmasına neden olmaktadır.

SMOTE, Borderline-SMOTE, SMOTE-Tomek ve ADASYN gibi dengeleme algoritmaları kullanılarak yapılan deneylerde, farklı hiperparametre kombinasyonlarıyla benzer düzeyde başarılı sonuçlar elde edilmiştir. Bu deneyler sonucunda ulaşılan en yüksek performans değerleri; %96,25 doğruluk, %91,67 makro kesinlik, %91,67 makro duyarlılık ve %90 makro F1 skoru olarak gözlemlenmiştir.

Uygulanan dört dengeleme yöntemi arasında SMOTE, genel olarak en istikrarlı ve güvenilir performansı göstermiştir. Borderline-SMOTE, SMOTE-Tomek ve ADASYN algoritmaları ise, özellikle sınıf sınırlarına odaklanarak daha rafine sentetik örnekler üretmiş;

bu durum bazı senaryolarda performansı artırırken, bazı durumlarda ise aşırı örnekleme veya gürültü etkisi nedeniyle sınıflandırma başarımını olumsuz yönde etkilemiştir.

Kullanılan sınıflandırma algoritmaları arasında SVM ve RF modelleri, sınıf ayırım gücü ve genelleme yeteneği açısından öne çıkmıştır. SVM, farklı kernel fonksiyonları ve C değerleriyle test edildiğinde, sınıf sınırlarını etkili biçimde öğrenerek tutarlı ve dengeli bir performans ortaya koymuştur. Bu durum, modelin hem azınlık hem de çoğunluk sınıfları ayırt edebilme yeteneğini desteklemektedir. RF algoritması ise, özellikle farklı ağaç sayıları ve derinliklerle yapılan denemelerde istikrarlı sonuçlar vermiş; bu da topluluk yapısının, aşırı öğrenme riskini dengeleyerek genelleme başarısını artırdığını göstermiştir. Buna karşılık, k-NN algoritması komşu sayısı ve mesafe metriklerine oldukça duyarlı olduğu için bazı senaryolarda tutarsız sonuçlar üretmiştir. Özellikle sınıflar arasında net ayrımların olmadığı durumlarda performans düşüşü gözlemlenmiştir. LR algoritması ise, farklı optimizasyon algoritmaları ve C değerleri ile test edilmiş; genellikle daha sade ve yorumlanabilir sonuçlar üretmiş olsa da, karmaşık sınıf ayrımlarında sınırlı kalmıştır.

5. SONUÇLAR

Bu yüksek lisans çalışmasında, 8 adet MOS sensör ve 14 adet QCM sensör içeren bir elektronik burun sistemi kullanılarak, akciğer kanseri teşhisi almış bireylerden toplanan nefes örnekleri analiz edilmiş ve bu veriler aracılığıyla akciğer kanseri evre sınıflandırması gerçekleştirilmiştir. Toplamda 200 nefes örneği kullanılmış olup, örneklerin evrelere göre dağılımı I. Evre için 14, II. Evre için 12, III. Evre için 23 ve IV. Evre için 151 şeklindedir. Verilere ilk olarak referans düzeltme ve iletkenlik dönüşümü gibi ön işleme adımları uygulanmış, ardından istatistiksel, zaman temelli ve türetilmiş öznitelikler çıkarılmıştır. Öznitelikler, Z-skor normalizasyonu ile aynı ölçeğe getirilmiş ve ardından yüksek boyutluluğun neden olabileceği hesaplama zorluklarını ve gereksiz bilgileri azaltmak amacıyla, toplam varyansın %95'ini koruyacak şekilde başlıca bileşenleri seçen PCA yöntemi uygulanmıştır. Sınıflandırma süreci hem dengesiz veri seti ile hem de dengeleme yöntemleri uygulanmış veri setleri ile ayrı ayrı gerçekleştirilmiştir. Sınıf dengesizliğini gidermek amacıyla SMOTE, Borderline-SMOTE, SMOTE-Tomek ve ADASYN yöntemleri kullanılmıştır. LR, SVM, k-NN ve RF sınıflandırma algoritmaları çeşitli hiperparametre ayarları ile test edilmiş; modeller 5 katlı katmanlı çapraz doğrulama ve ayrı bir test veri seti kullanılarak değerlendirilmiştir. Elde edilen sonuçlar doğruluk, makro kesinlik, makro duyarlılık ve makro F1-skoru metrikleri üzerinden karşılaştırılmıştır.

Yalnızca MOS, yalnızca QCM ve her iki sensör verisinin (MOS+QCM) birlikte kullanıldığı senaryolar karşılaştırıldığında, en yüksek performansın MOS+QCM kombinasyonu ile elde edildiği görülmüştür. Bu yapı, öznitelik uzayının zenginleşmesini sağlayarak, modellerin sınıflar arası ayrımı daha etkin bir şekilde öğrenmesine imkân tanımıştır. Her iki verinin birlikte kullanılması, makro kesinlik, makro duyarlılık ve makro F1 metriklerinde en yüksek değerlere ulaşılmasını sağlamış, bu da sensör birleşimlerinin tamamlayıcı nitelikte olduğunu ortaya koymuştur.

Veri setindeki sınıf dengesizliğini gidermek amacıyla uygulanan örnekleme yöntemleri, sınıflandırma performansı üzerinde doğrudan etkili olmuştur. SMOTE, Borderline-SMOTE, SMOTE-Tomek ve ADASYN gibi dengeleme teknikleri, özellikle azınlık sınıflarda makro duyarlılık ve makro F1-skoru açısından dikkate değer iyileşmeler sağlamıştır. Bu sonuçlar, dengeli bir örnek dağılımının, modelin tüm sınıfları öğrenme kapasitesini artırarak sınıf temelli başarıyı olumlu yönde etkilediğini ortaya koymaktadır.

Dört yöntem arasında SMOTE, genel itibarıyla en istikrarlı ve güvenilir performansı sunarken; sınıf sınırlarına odaklanan Borderline-SMOTE, SMOTE-Tomek ve ADASYN teknikleri, bazı durumlarda daha rafine örnekler üreterek başarıyı artırmış, ancak bazı senaryolarda ise aşırı örnekleme ya da gürültü üretimi nedeniyle performans kayıplarına neden olmuştur. Bu bulgular, sınıf dengesizliğini gidermeye yönelik yöntemlerin dikkatli seçilmesi ve uygulama bağlamına uygun biçimde optimize edilmesi gerektiğini göstermektedir.

Kullanılan sınıflandırma algoritmaları arasında genel olarak en iyi performansı SVM ve RF modelleri sergilemiştir. SVM, farklı çekirdek fonksiyonları sayesinde hem doğrusal hem doğrusal olmayan ayrımları etkin şekilde yapabilmiş; RF ise topluluk yapısı sayesinde aşırı öğrenme riskini azaltarak daha tutarlı sonuçlar üretmiştir. k-NN algoritmasının performansı, seçilen k değeri ve mesafe metriğine oldukça duyarlı bulunmuştur. LR ise yorumlanabilirliği yüksek olmakla birlikte karmaşık sınıf ayrımlarında sınırlı başarı göstermiştir. Denenen hiperparametre kombinasyonlarının, model başarımını anlamlı düzeyde etkilediği görülmüştür; bu da model yapılandırmasının özenli şekilde yapılması gerektiğini ortaya koymaktadır.

Bu çalışmada, sınıf dengesizliği içeren bir veri yapısına sahip olunması nedeniyle model performans değerlendirmelerinde yalnızca doğruluk metriğiyle yetinilmemiş; bunun yanı sıra makro duyarlılık, makro kesinlik ve makro F1-skoru gibi metriklere odaklanılmıştır. Özellikle I. ve II. evre gibi düşük örnek sayısına sahip sınıfların doğru temsil edilebilmesi açısından bu yaklaşım, değerlendirme sürecine adillik kazandırmıştır. Ayrıca, sadece çapraz doğrulama sonuçlarına bağlı kalınmamış; ayrı bir test veri seti üzerinden yapılan değerlendirmelerle modelin genelleme yeteneği de test edilerek sonuçların güvenilirliği artırılmıştır.

6. ÖNERİLER

Bu çalışmada elde edilen bulgular, MOS ve QCM sensör verilerinin birlikte kullanımıyla oluşturulan hibrit elektronik burun sistemlerinin, makine öğrenmesi tabanlı yöntemlerle birlikte akciğer kanseri evrelerinin sınıflandırılmasında etkili bir şekilde kullanılabileceğini göstermektedir. Veri setinde gözlemlenen sınıf dengesizliği, SMOTE ve benzeri örnekleme yöntemleriyle dengelenmiş ve bu sayede azınlık sınıfların da model performansına katkısı artırılmıştır. Ancak çalışmanın belirli sınırlılıkları ve veri yapısının doğası, gelecekte yapılacak araştırmalar için çeşitli iyileştirme alanlarına işaret etmektedir. Bu bölümde, hem bu sınırlılıkların aşılmasına yönelik öneriler hem de ilerleyen çalışmalarda dikkate alınabilecek yöntemsel ve uygulamalı geliştirme alanları sunulmaktadır.

Veri Seti ve Örnek Sayısı: Verilerin yalnızca tek bir merkezden değil, farklı coğrafi bölgeleri ve sağlık kuruluşlarını kapsayacak şekilde, daha çeşitli ve geniş bir hasta popülasyonundan elde edilmesi, geliştirilen modelin genellenebilirliğini ve farklı klinik ortamlarda da geçerli olma potansiyelini önemli ölçüde artıracaktır.

Gerçek Zamanlı Uygulama ve Klinik Entegrasyon: Geliştirilen sistemin bir prototip cihaz haline getirilerek hastane ortamında gerçek zamanlı olarak test edilmesi, klinik geçerliliğini değerlendirmek açısından önemlidir. Sensör verisinin gerçek zamanlı analizi için gömülü sistemlerle entegrasyon çalışmaları yapılabilir. Modelin kullanıcı arayüzü ile desteklenmesi, hekimlerin karar verme süreçlerine yardımcı olabilir. Ayrıca sistemin taşınabilir hale getirilmesi, sahada veya evde erken tarama amaçlı kullanımını mümkün kılarak erişilebilirliği artırabilir.

7. KAYNAKLAR

- Bach, P. B., Jett, J. R., Pastorino, U., Tockman, M. S., Swensen, S. J., & Begg, C. B. (2007). Computed Tomography Screening and Lung Cancer Outcomes. *The Journal of the American Medical Association*, 297(9), 953-961.
- Bailey-Wilson, J. E., Amos, C. I., Pinney, S. M., Petersen, G. M., Wiest, J. S., Fain, P., . . . Liu, Q. (2004). A Major Lung Cancer Susceptibility Locus Maps to Chromosome 6q23–25. 75(3), 460-474.
- Barta, J. A., Powell, C. A., & Wisnivesky, J. P. (2019). Global Epidemiology of Lung Cancer. *Annals of global health*, 85(1), 8.
- Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 1, 20-29.
- Beveridge, R., Pintos, J., Parent, M.-É., Asselin, J., & Siemiatycki, J. (2010). Lung cancer risk associated with occupational exposure to nickel, chromium VI, and cadmium in two population-based case–control studies in Montreal. *American Journal of Industrial Medicine*, 53(5), 476-485.
- Bharathy, S., Pavithra, R., & Akshaya, B. (2022). Lung Cancer Detection using Machine Learning. *International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (539-543). Salem: IEEE.
- Binson, V. A., & Subramoniam, M. (2021). Artificial Intelligence Based Breath Analysis System for the Diagnosis of lung cancer. *Journal of Physics: Conference Series*, 1950(1), 012065.
- Binson, V. A., Subramoniam, M., Ragesh, G. K., & Kumar, A. (2021). Early Detection of Lung Cancer Through Breath Analysis Using AdaBoost Ensemble Learning Method. *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)* (183-187). Ernakulam, India: IEEE.
- Binson, V. A., Subramoniam, M., Sunny, Y., & Mathew, L. (2021). Prediction of Pulmonary Diseases With Electronic Nose Using SVM and XGBoost. *IEEE Sensors Journal*, 21(18), 20886-20895.
- Binson, V., Subramoniam, M., & Luke, M. (2021). Detection of COPD and Lung Cancer with electronic nose using ensemble learning methods. *Clinica Chimica Acta*, 231-238.
- Blanchet, L., Smolinska, A., Baranska, A., Tigchelaar, E., Swertz, M., Zhernakova, A., . . . van Schooten, F. J. (2017). Factors that influence the volatile organic compound content in human breath. *Journal of Breath Research*, 11(1), 016013.

- Blatt, R., Bonarini, A., Calabro, E., Torre, M. D., Matteucci, M., & Pastorini, U. (2007). Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors. *2007 International Joint Conference on Neural Networks* (1423-1428). Orlando, Florida: IEEE.
- Bona, E., Silva, R. S., Borsato, D., & Bassoli, D. G. (2011). Optimized Neural Network for Instant Coffee Classification through an Electronic Nose. *International Journal of Food Engineering*, 7(6).
- Borsdorf, H., & G.A, E. (2006). Ion mobility spectrometry: principles and applications. *Applied Spectroscopy Review*, 41(4), 323-375.
- Boyle, P., & Maisonneuve, P. (1995). Lung cancer and tobacco smoking. *Lung Cancer*, 12(3), 167-181.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- Brown, J. S., Eraut, D., Trask, C., & Davison, A. G. (1996). Age and the treatment of lung cancer. *Thorax*, 51(6), 564-568.
- Canhoto, O., & Magan, N. (2005). Electronic nose technology for the detection of microbial and chemical contamination of potable water. *Sensors and Actuators B: Chemical*, 106(1), 3-6.
- Capuano, R., Santonico, M., Pennazza, G., Ghezzi, S., Martinelli, E., Roscioni, C., . . . D'Amico, A. (2015). The lung cancer breath signature: a comparative analysis of exhaled breath and air sampled from inside the lungs. *Scientific Reports*, 5(1), 16491.
- Chan, B. A., & Hughes, B. G. (2015). Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Translational Lung Cancer Research*, 4(1), 36-54.
- Chang, J.-E., Lee, D.-S., Ban, S.-W., Oh, J., Jung, M. Y., Kim, S.-H., . . . Jheon, S. (2018). Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system. *Sensors and Actuators B: Chemical*, 255, 800-807.
- Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Chen, K., Liu, L., Nie, B., Lu, B., Fu, L., He, Z., . . . Liu, H. (2021). Recognizing lung cancer and stages using a self-developed electronic nose system. *Computers in Biology and Medicine*, 131, 104294.
- Chen, X., Cao, M., Li, Y., Hu, W., Wang, P., Ying, K., & Pan, H. (2005). A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method. *Measurement Science and Technology*, 16(8), 1535.
- Cheng, Z. J., Warwick, G., Yates, D. H., & Thomas, P. S. (2009). An electronic nose in the discrimination of breath from smokers and non-smokers: a model for toxin exposure. *Journal of Breath Research*, 036003.

- Chiba, I., Takahashi, T., Nau, M. M., D'Amico, D., Curiel, D. T., Mitsudomi, T., . . . Koga, H. (1990). Mutations in the p53 gene are frequent in primary, resected non-small cell lung cancer. Lung Cancer Study Group. *Oncogene*, 5(10), 1603-1610.
- Cooley, M. E. (2000). Symptoms in Adults with Lung Cancer: A Systematic Research Review. *Journal of Pain and Symptom Management*, 19(2), 137-153.
- Corinna, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.
- D'Amico, A., Pennazza, G., Santonico, M., Martinelli, E., Roscioni, C., Galluccio, G., . . . Di Natale, C. (2010). An investigation on electronic nose diagnosis of lung cancer. *Lung Cancer*, 68(2), 170-176.
- De Hoffmann, E., & Stroobant, V. (2007). *Mass spectrometry: principles and applications*. John Wiley & Sons.
- Detterbeck, F. C., Boffa, D. J., Kim, A. W., & Tanoue, L. T. (2017). The Eighth Edition Lung Cancer Stage Classification. *Chest*, 151(1), 193-203.
- Di Natale, C., Macagnano, A., Martinelli, E. P., D'Arcangelo, G., Roscioni, C., Finazzi-Agrò, A., & D'Amico, A. (2003). Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosensors and Bioelectronics*, 18(10), 1209-1218.
- Dragonieri, Silvano, Annema, J. T., Schot, R., van der Schee, M. P., Spanevello, A., . . . Rabe, K. F. (2009). An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. *Lung Cancer*, 64(2), 166-170.
- Durham, A. L., & Adcock, I. M. (2015). The relationship between COPD and lung cancer. *Lung Cancer*, 90(2).
- Ellis, A. M., & Mayhew, C. (2013). *Proton transfer reaction mass spectrometry: principles and applications*. John Wiley & Sons.
- Ellis, P. M., & Vandermeer, R. (2011). Delays in the diagnosis of lung cancer. *Journal of thoracic disease*, 3(3), 183.
- Ferrell, B., Koczywas, M., Grannis, F., & Harrington, A. (2011). Palliative care in lung cancer. *The Surgical clinics of North America*, 91(2), 403.
- Fischer, B. M., & Mortensen, J. (2006). The Future in Diagnosis and Staging of Lung Cancer: Positron Emission Tomography. *Respiration*, 73(3), 267-276.
- Gavelli, G., & Giampalma, E. (2000). Sensitivity and specificity of chest x-ray screening for lung cancer. *Cancer*, 89(S11), 2453-2456.
- Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., . . . Detterbeck, F. (2016). The IASLC Lung Cancer Staging Project: Proposals

- for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *Journal of Thoracic Oncology*, 11(1), 39-51.
- Gomez, M., & Silvestri, G. A. (2009). Endobronchial Ultrasound for the Diagnosis and Staging of Lung Cancer. *Proceedings of the American Thoracic Society*, 6(2), 180-186.
- Gordon, S. M., Szidon, J. P., Krotoszynski, B. K., Gibbons, R. D., & O'Neill, H. J. (1985, 8 1). Volatile organic compounds in exhaled air from patients with lung cancer. *Clinical Chemistry*, 31(8), 1278-1282.
- Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., . . . Rosell, R. (2015). Non-small-cell lung cancer. *Nature reviews Disease primers*, 1(1), 1-16.
- Gutierrez-Osuna, R. (2002). Pattern analysis for machine olfaction: A review. *IEEE Sensors journal*, 2(3), 189-202.
- Haddi, Z., Amari, A., Alami, H., El Bari, N., Llobet, E., & Bouchikhi, B. (2011). A portable electronic nose system for the identification of cannabis-based drugs. *Sensors and Actuators B: Chemical*, 155(2), 456-463.
- Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. *International conference on intelligent computing*, 878-887.
- Hansen, H. (2008). *Textbook of lung cancer*. CRC Press.
- Hao, L., & Huang, G. (2023). An improved AdaBoost algorithm for identification of lung cancer based on electronic nose. *Heliyon*, 9(3).
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (1322-1328). Hong Kong: IEEE.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression*. John Wiley and Sons.
- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining and knowledge management process*, 5(2), 1.
- Huang, C.-H., Zeng, C., Wang, Y.-C., Peng, H.-Y., Lin, C.-S., Chang, C.-J., & Yang, H.-Y. (2018). A Study of Diagnostic Accuracy Using a Chemical Sensor Array and a Machine Learning Technique to Detect Lung Cancer. *Sensors*, 18(9), 2845.
- Imyanitov, E. N., & Iyevleva, A. G. (2021). Molecular testing and targeted therapy for non-small cell lung cancer: Current status and perspectives. *Critical Reviews in Oncology/Hematology*, 157, 103194.

- Jasinski, G., Wozniak, L., Kalinowski, P., & Jasinski, P. (2018). Evaluation of the Electronic Nose Used for Monitoring Environmental Pollution. *2018 XV International Scientific Conference on Optoelectronic and Electronic Sensors (COE)* (1-4). Warsaw, Poland: IEEE.
- Kao, M.-H., Chiu, S.-W., Lee, M.-R., Sun, M., & Tang, K.-T. (2023). Deep neural network of E-nose sensor for lung cancer classification. *IEEE BioSensors Conference (BioSensors)* (1-4). London: IEEE.
- Karakaya, D., Ulucan, O., & Turkan, M. (2020). Electronic nose and its applications: A survey. *International journal of Automation and Computing*, 17(2), 172-209.
- Karasek, F. W., & Clement, R. E. (2012). *Karasek, F. W., & Clement, R. E. (2012). Basic gas chromatography-mass spectrometry: principles and techniques*. Elsevier.
- Klebe, S., Leigh, J., Henderson, D. W., & Nurminen, M. (2020). Asbestos, Smoking and Lung Cancer: An Update. *International Journal of Environmental Research and Public Health*, 17(1), 258.
- Kononov, A., Korotetsky, B., Jahatspanian, I., Gubal, A., Vasiliev, A., Arsenjev, A., . . . Ganeev, A. (2019). Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer. *Journal of Breath Research*, 14(1), 016004.
- Kort, S., Brusse-Keizer, M., Schouwink, H., Citgez, E., de Jongh, F. H., van Putten, J. W., . . . J. (2023). Diagnosing Non-Small Cell Lung Cancer by Exhaled Breath Profiling Using an Electronic Nose: A Multicenter Validation Study. *Chest*, 163(3), 697-706.
- Koyama, H., Ohno, Y., Seki, S., Nishio, M., Yoshikawa, T., Matsumoto, S., & Sugimura, K. (2013). Magnetic Resonance Imaging for Lung Cancer. *Journal of Thoracic Imaging*, 28(3), 138.
- Landreneau, R. J., Mack, M. J., Dowling, R. D., Luketich, J. D., Keenan, R. J., Ferson, P. F., & R, H. S. (1998). The Role of Thoracoscopy in Lung Cancer Management. *Chest*, 113(1, Supplement), 6S-12S.
- Lang-Lazdunski, L. (2013). Surgery for nonsmall cell lung cancer. *European Respiratory Review*, 22(129), 382-404.
- Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y.-W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1856(2), 189-210.
- Li, W., Jia, Z., Xie, D., Chen, K., Cui, J., & Liu, H. (2020). Recognizing lung cancer using a homemade e-nose: A comprehensive study. *Computers in Biology and Medicine*, 120, 103706.
- Lima, R. S., Danielski, G. C., & Pires, A. C. (2018). Mastitis Detection and Prediction of Milk Composition Using Gas Sensor and Electrical Conductivity. *Food and Bioprocess Technology*, 11(3), 551-560.

- Limbo, S., Torri, L., Sinelli, N., Franzetti, L., & Casiraghi, E. (2010). Evaluation and predictive modeling of shelf life of minced beef stored in high-oxygen modified atmosphere packaging at different temperatures. *Meat Science*, 84(1), 129-136.
- Littlewood, A. B. (2013). *Gas chromatography: principles, techniques, and applications*. Elsevier.
- Liu, B., Yu, H., Zeng, X., Zhang, D., Gong, J., Tian, L., . . . Liu, R. (2021). Lung cancer detection via breath by electronic nose enhanced with a sparse group feature selection approach. *Sensors and Actuators B: Chemical*, 339, 129896.
- Liu, L., Li, W., He, Z., Chen, W., Liu, H., Chen, K., & Pi, X. (2021). Detection of lung cancer with electronic nose using a novel ensemble learning framework. *Journal of Breath Research*, 15(2), 026014.
- Livingston, R. B. (1997). Combined modality therapy of lung cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 3(12), 2638-2647.
- López, F. (2024, 10 24). *SMOTE: Synthetic Data Augmentation for Tabular Data*. towards data science: <https://towardsdatascience.com/smote-synthetic-data-augmentation-for-tabular-data-1ce28090debc/> adresinden alındı
- Lu, B., Fu, L., Nie, B., Peng, Z., & Liu, H. (2019). A Novel Framework with High Diagnostic Sensitivity for Lung Cancer Detection by Electronic Nose. *Sensors*, 19(23), 5333.
- Machado, R. F., Laskowski, D., Deffenderfer, O., Burch, T., Zheng, S., Mazzone, P. J., . . . Erzurum, S. C. (2005). Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath. *American Journal of Respiratory and Critical Care Medicine*, 171(11), 1286-1291.
- Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., & Boffetta, P. (2016). Risk factors for lung cancer worldwide. *European Respiratory Journal*, 48(3), 889-902.
- Marzorati, D., Mainardi, L., Sedda, G., Gasparri, R., Spaggiari, L., & Cerveri, P. (2019). A Metal Oxide Gas Sensors Array for Lung Cancer Diagnosis Through Exhaled Breath Analysis. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (1584-1587). Berlin: IEEE.
- Marzorati, Mainardi, Sedda, Gasparri, Spaggiari, & Cerveri. (2021). MOS Sensors Array for the Discrimination of Lung Cancer and At-Risk Subjects with Exhaled Breath Analysis. *Chemosensors*, 9(8), 209.
- Mazzone, P. J., Hammel, J., Dweik, R., Na, J., Czich, C., Laskowski, D., & Mekhail, T. (2007). Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array. *Thorax*, 62(7), 565-568.
- Mazzone, P., Jain, P., Arroliga, A. C., & Matthay, R. A. (2002). Bronchoscopy and needle biopsy techniques for diagnosis and staging of lung cancer. *Clinics in Chest Medicine*, 23(1), 137-158.

- Moninuola, F. S., Adetiba, E., Atayero, A. A., Awelewa, A., Adeyeye, A., Oshin, O. I., . . . Ezekiel, V. (2023). Early Detection of Lung Cancer via Breath Analysis Utilising Electronic Nose. *2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (1-6). Durban: IEEE.
- Montani, F., Marzi, M., Dezi, F., Dama, E., Carletti, R. M., Bonizzi, G., . . . Bianchi, F. (2015). miR-Test: A Blood Test for Lung Cancer Early Detection. *JNCI: Journal of the National Cancer Institute*, 107(6), djv063.
- Mumyalmaz, B., & Karabacak, K. (2015). An E-Nose-based indoor air quality monitoring system: prediction of combustible and toxic gas concentrations. *Turkish Journal of Electrical Engineering and Computer Sciences*, 23(3), 729-740.
- Nesbitt, J. C., Putnam, J. B., Walsh, G. L., Roth, J. A., & Mountain, C. F. (1995). Survival in early-stage non-small cell lung cancer. *The Annals of Thoracic Surgery*, 60(2), 466-472.
- Nooreldeen, R., & Bach, H. (2021). Current and future development in lung cancer diagnosis. *International journal of molecular sciences*, 22(16), 8661.
- O'Neill, H. J., Gordon, S. M., O'Neill, M. H., Gibbons, R. D., & Szidon, J. P. (1988, 8 1). A computerized classification technique for screening for the presence of breath biomarkers in lung cancer. *Clinical Chemistry*, 34(8), 1613-1318.
- Ordukaya, E., & Karlik, B. (2017). Quality Control of Olive Oils Using Machine Learning and Electronic Nose. *Journal of Food Quality*, 2017(1), 9272404.
- Özsandıkçığlu, Ü. (2023). Karma sensör yapıları elektronik burun ile akciğer kanseri tespiti. Trabzon, Türkiye: Karadeniz Teknik Üniversitesi, Doktora Tezi.
- Pearce, T. C., Schiffmann, S. S., Nagle, H. T., & Gardner, J. W. (2003). *Handbook of machine olfaction*. Wiley Online Library.
- Philips, M., Cataneo, R. N., Cummin, A. R., Gagliardi, A. J., Gleeson, K., Greenberg, J., . . . Rom, W. N. (2003). Detection of Lung Cancer With Volatile Markers in the Breath. *Chest*, 123(6), 2115-2123.
- Phillips, M., Gleeson, K., Hughes, J. M., Greenberg, J., Cataneo, R. N., Baker, L., & McVay, W. P. (1996). Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *The Lancet*, 353(9168), 1930-1933.
- Phillips, M., Herrera, J., Krishnan, S., Zain, M., Greenberg, J., Cataneo, & N., R. (1999). Variation in volatile organic compounds in the breath of normal humans. *Journal of Chromatography B: Biomedical Sciences and Applications*, 729(1), 75-88.
- Poli, D., Carbognani, P., Corradi, M., Goldoni, M., Acampa, O., Balbi, B., . . . Mutti, A. (2005). Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. *Respiratory Research*, 6(1), 71.

- Pope III, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. *The Journal of the American Medical Association*, 287(9), 1132-1141.
- Pulumati, A., Pulumati, A., Dwarakanath, B. S., Verma, A., & Papineni, R. V. (2023). Technological advancements in cancer diagnostics: Improvements and limitations. *Cancer Reports*, 6(2), e1764.
- Qi, P.-F., Hao, M., & Zeng, M. (2017). A CNN-Based Simplified Data Processing Method for Electronic Nose. *2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*. Montreal, QC, Canada: IEEE.
- Rahman, M. M., & Davis, D. N. (2013). Addressing the class imbalance problem in medical datasets. *International Journal of Machine Learning and Computing*, 3(2), 224.
- Ramalingam, S., & Belani, C. (2008). Systemic Chemotherapy for Advanced Non-Small Cell Lung Cancer: Recent Advances and Future Directions. *The Oncologist*, 13(S1), 5-13.
- Rudin, C., Brambilla, E., Faivre-Finn, C., & Sage, J. (2021). Small-cell lung cancer. *Nature Reviews Disease Primers*, 7(1), 3.
- Ruidavets, M., Garcia de Herreros, M., Besse, B., & Mezquita, L. (2022). Radon and Lung Cancer: Current Trends and Future Perspectives. *Cancers*, 14(13), 3142.
- Schild, S. E., Stella, P. J., Geyer, S. M., Bonner, J. A., McGinnis, W. L., Mailliard, J. A., . . . Jatoti, A. R. (2003). The Outcome of Combined-Modality Therapy for Stage III Non-Small-Cell Lung Cancer in the Elderly. *Journal of clinical oncology*, 21(17), 3201-3206.
- Smith, D., & Španěl, P. (2005). Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass spectrometry reviews*, 24(5), 661-700.
- Stapelfeld, C., Dammann, C., & Maser, E. (2020). Sex-specificity in lung cancer risk. *International Journal of Cancer*, 146(9), 2376-2382.
- Steven, A., Fisher, S. A., & Robinson, B. W. (2016). Immunotherapy for lung cancer. *Respirology*, 21(5), 821-833.
- Szeghalmy, S., & Fazekas, A. (2023). A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sensors*, 23(4), 2333.
- Taş, F., Çiftçi, R., Kılıç, L., & Karabulut, S. (2013). Age is a prognostic factor affecting survival in lung cancer patients. *Oncology letters*, 6(5), 1507-1513.
- Taylor, R., Najafi, F., & Dobson, A. (2007). Meta-analysis of studies of passive smoking and lung cancer: effects of study type and continent. *International Journal of Epidemiology*, 36(5).

- Temel, J. S., & diğerleri, v. (2010). Early palliative care for patients with metastatic non--small-cell lung cancer. *New England Journal of Medicine*, 363(8), 733-742.
- Thunnissen, F. B. (2003). Sputum examination for early detection of lung cancer. *Journal of Clinical Pathology*, 56(11), 805-810.
- Timmerman, R., Paulus, R., Galvin, J., Michalski, J., Straube, W., Bradley, J., . . . Choy, H. (2010). Stereotactic Body Radiation Therapy for Inoperable Early Stage Lung Cancer. *The Journal of the American Medical Association*, 303(11), 1070-1076.
- Tirzīte, M., Bukovskis, M., Strazda, G., Jurka, N., & Taivans, I. (2017). Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis. *Journal of Breath Research*, 11(3), 036009.
- Tirzīte, M., Bukovskis, M., Strazda, G., Jurka, N., & Taivans, I. (2018). Detection of lung cancer with electronic nose and logistic regression analysis. *Journal of Breath Research*, 13(1), 016006.
- Valdez, L. F., & Gutiérrez, J. M. (2016). Chocolate Classification by an Electronic Nose with Pressure Controlled Generated Stimulation. *Sensors*, 16(10), 1745.
- van de Goor, R., van Hooren, M., Dingemans, A.-M., Kremer, B., & Kross, K. (2018). Training and Validating a Portable Electronic Nose for Lung Cancer Screening. *Journal of Thoracic Oncology*, 13(5), 676-681.
- W., C. N., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Wang, D., Yu, K., Wang, Y., Hu, Y., Zhao, C., Wang, L., . . . Wang, P. (2012). A hybrid electronic noses' system based on mos-saw detection units intended for lung cancer diagnosis. *Journal of Innovative Optical Health Sciences*, 05(01), 1150006.
- Wang, P., Chen, X., Xu, F., Lu, D., Cai, W., Ying, K., . . . Hu, Y. (2008). Development of electronic nose for diagnosis of lung cancer at early stage. *2008 International Conference on Information Technology and Applications in Biomedicine* (588-591). Shenzhen: IEEE.
- Wei, S., Zhang, H., & Tao, S. (2019). A review of arsenic exposure and lung cancer. *Toxicology Research*, 8(3), 319-327.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.
- Wong, D.-M., Fang, C.-Y., Chen, L., Chiu, C., Chou, T., Wu, C.-C., . . . Tang, K.-T. (2018). Development of a breath detection method based E-nose system for lung cancer identification. *IEEE International Conference on Applied System Invention (ICASI)* (1119-1120). Chiba, Japan: IEEE.

Zhan, X., Wang, Z., Yang, M., Luo, Z., You, W., & Li, G. (2020). An electronic nose-based assistive diagnostic prototype for lung cancer detection with conformal prediction. *Measurement*, 158, 107588.

Zhang, Y., Deng, L., & Wei, B. (2024). Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation. *Mathematics*, 11, 1709.



ÖZGEÇMİŞ

Bilal Talha AYVAZ, Karadeniz Teknik Üniversitesi Elektrik Elektronik Mühendisliği bölümünden mezun oldu. 2022-2023 eğitim-öğretim yılının bahar döneminde Karadeniz Teknik Üniversitesi Elektrik Elektronik Mühendisliği bölümünde tezli yüksek lisans eğitime başladı. Aynı yıl Karadeniz Teknik Üniversitesi Elektrik Elektronik Mühendisliği Bölümü Devreler ve Sistemler Anabilim Dalı'na araştırma görevlisi olarak atandı. Karadeniz Teknik Üniversitesi'nde araştırma görevlisi olarak görevine devam etmektedir.

