



T.C.

BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS PROGRAMI

**FOTOVOLTAİK SİSTEMLERDE HATA TESPİTİ VE SINIFLANDIRILMASI İÇİN  
DOĞRUSAL VE DOĞRUSAL OLMAYAN BOYUT AZALTMA TEKNİKLERİNİN  
KARŞILAŞTIRMALI BİR DEĞERLENDİRMESİ**

YÜKSEK LİSANS TEZİ

GÖKHAN KARACA

TEZ DANIŞMANI

DOÇ. DR. Ü. ÇİĞDEM TURHAL

BİLECİK, 2026

10812859

T.C.  
BİLECİK ŐEYH EDEBALI ÜNİVERSİTESİ  
LİSANSÜSTÜ EĐİTİM ENSTİTÜSÜ  
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĐİ ANABİLİM DALI  
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĐİ TEZLİ YÜKSEK LİSANS PROGRAMI

**FOTOVOLTAİK SİSTEMLERDE HATA TESPİTİ VE SINIFLANDIRILMASI İÇİN  
DOĐRUSAL VE DOĐRUSAL OLMAYAN BOYUT AZALTMA TEKNİKLERİNİN  
KARŐILAŐTIRMALI BİR DEĐERLENDİRMESİ**

YÜKSEK LİSANS TEZİ

GÖKHAN KARACA

TEZ DANIŐMANI  
DOĐ. DR. Ü. ÇİĐDEM TURHAL

BİLECİK, 2026

10812859

## BEYAN

'Fotovoltaik Sistemlerde Hata Tespiti Ve Sınıflandırılması İçin Doğrusal Ve Doğrusal Olmayan Boyut Azaltma Tekniklerinin Karşılaştırmalı Bir Değerlendirmesi' başlıklı yüksek lisans tezinin hazırlık ve yazım aşamasında bilimsel araştırma ve etik kurallarına uyduğumu, [Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehberine](#) uygun olarak tez/dönem projemi hazırladığımı, başkalarının eserlerinden yararlandığım bölümlerde bilimsel etik kurallarına uygun olarak atıfta bulunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, çalışmamın herhangi bir kısmının başka bir tez/dönem projesi olarak sunulmadığımı, aksinin tespit edilmesi durumunda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

**Gökhan Karaca**

.././20..

**İmza:**

## ÖN SÖZ

Güneş enerjisi, küresel enerji dönüşümünün en kritik bileşenlerinden biri haline gelmiştir. Bu çalışma, FV sistemlerde meydana gelen arızaların akıllı yöntemlerle tespit edilmesi sürecinde veri ön işleme aşamasının önemine odaklanmaktadır. Özellikle yüksek boyutlu elektriksel verilerin anlamlı bir şekilde indirgenmesi, sınıflandırma algoritmalarının başarısını doğrudan etkilemektedir.

Lisansüstü eğitim ve tez çalışmam süresince kıymetli yönlendirmelerini esirgemeyen danışman hocam Sayın Doç. Dr. Ü. Çiğdem TURHAL'a en içten teşekkürlerimi sunarım.

Çalışma hayatım ve lisansüstü eğitim sürecim boyunca desteklerini esirgemeyen ETİ Makine ailesine ve değerli mesai arkadaşlarıma teşekkür ederim.

Ayrıca bu zorlu yolculuğun her anında yanımda olan, desteklerini hiçbir zaman esirgemeyen eşim Canan GÜVEN KARACA'ya, kızım Zümranur KARACA'ya, annem Nurhan KARACA'ya, özellikle de maddi-manevi her zorlukta desteğini hep yanımda hissettiğim dedem Zekeriya ÇAKICI ve anneannem Emine ÇAKICI'ya en içten dileklerle şükranlarımı sunarım.

## ÖZET

Fotovoltaik (FV) sistemlerde arızaların doğru ve zamanında tespit edilmesi; operasyonel güvenilirliğin artırılması, enerji üretim kayıplarının azaltılması ve bakım süreçlerinin daha etkin yürütülmesi açısından kritik öneme sahiptir. Modern FV sistemlerden elde edilen akım, gerilim, güç, ışınım ve sıcaklık tabanlı ölçüm verileri yüksek boyutlu, korelasyonlu ve doğrusal olmayan özellikler gösterebilmektedir. Bu nedenle arıza sınıflandırma başarısı yalnızca kullanılan sınıflandırıcı algoritmaya değil, aynı zamanda ölçüm verilerinin nasıl temsil edildiğine ve boyut azaltma sürecinin etkinliğine de bağlıdır.

Bu çalışmada, FV arıza sınıflandırması için doğrusal ve çekirdek tabanlı doğrusal olmayan boyut azaltma tekniklerini karşılaştırmalı olarak değerlendiren bir analiz çerçevesi sunulmuştur. Bu kapsamda doğrusal yöntemler olarak Temel Bileşen Analizi (PCA), Doğrusal Ayırma Analizi (LDA) ve Bağımsız Bileşen Analizi (ICA); çekirdek tabanlı doğrusal olmayan yöntemler olarak ise Çekirdek Temel Bileşen Analizi (KPCA), Çekirdek Doğrusal Ayırma Analizi (KLDA) ve KPCA-ICA yaklaşımı incelenmiştir. KPCA-ICA yaklaşımı, bu çalışmada KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde gerçekleştirilmiştir.

Analizlerde, MATLAB/Simulink ortamında modellenen 250 kW gücündeki şebeke bağlantılı FV sistemden elde edilen simülasyon tabanlı veri seti kullanılmıştır. Veri setinde normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört farklı sistem durumu ele alınmıştır. FV sistemden elde edilen akım, gerilim, güç, ışınım, sıcaklık ve türetilmiş istatistiksel özellikler kullanılarak veri ön işleme, standardizasyon, boyut azaltma ve sınıflandırma adımları uygulanmıştır. Performans değerlendirmesinde doğruluk, AUC, kesinlik, duyarlılık, F1 skoru, özgüllük ve işlem süresi ölçütleri dikkate alınmıştır.

Elde edilen sonuçlar, boyut azaltma yöntemleri arasında KPCA'nın bu veri seti ve değerlendirme protokolü kapsamında AUC ve F1 skoru bakımından en başarılı temsil yaklaşımı olduğunu göstermiştir. KPCA ve PCA aynı doğruluk düzeyine ulaşmış olsa da KPCA'nın daha yüksek AUC ve F1 skoru üretmesi, doğrusal olmayan temsil uzayının arıza sınıfları arasındaki ayırt ediciliği artırdığını ortaya koymuştur. PCA ise doğrusal, sade ve yorumlanabilir yapısı nedeniyle mühendislik uygulamaları açısından güçlü bir alternatif olarak değerlendirilmiştir. LDA, ICA ve KLDA yöntemleri genel sınıflandırma performansı açısından orta düzeyde sonuçlar üretirken, KPCA-ICA yaklaşımı bu veri setinde sınırlı sınıflandırma başarısı göstermiştir.

Ayrıca yöntemlerin saha benzeri koşullar altındaki dayanıklılığını değerlendirmek amacıyla test

verisi üzerinde gürültü, rastgele eksik veri, blok eksik veri, sensör sürüklenmesi, tek sensör kaybı ve ışınım bandı temelli senaryolar oluşturulmuştur. Dayanıklılık analizleri, genel performans açısından KPCA'nın öne çıkmasına karşın, çevresel değişkenlik altında özellikle düşük ışınım bandında KLDA tabanlı yapının en güçlü sonucu verdiğini göstermiştir. Bu bulgu, boyut azaltma yöntemlerinin performansının yalnızca genel sınıflandırma başarısı üzerinden değil, farklı işletme ve bozulum koşulları altındaki kararlılıkları bakımından da değerlendirilmesi gerektiğini ortaya koymaktadır.

Sonuç olarak bu çalışma, FV sistemlerde arıza sınıflandırma başarısının yalnızca sınıflandırıcı seçimine değil, veri temsilinin niteliğine ve bozulum koşulları altındaki dayanıklılık performansına da bağlı olduğunu göstermektedir. Elde edilen bulgular, doğrusal ve doğrusal olmayan boyut azaltma tekniklerinin FV arıza teşhisinde farklı avantajlar sunduğunu ve uygun temsil yönteminin seçiminin sınıflandırma performansı üzerinde belirleyici bir etkiye sahip olduğunu ortaya koymaktadır.

**Anahtar Kelimeler:** Boyut azaltma, çekirdek yöntemleri, fotovoltaik arıza tespiti, makine öğrenmesi, sınıflandırma, dayanıklılık analizi.

## ABSTRACT

Accurate and timely fault detection in photovoltaic (PV) systems is critical for improving operational reliability, reducing energy production losses, and enabling more effective maintenance processes. Measurement data obtained from modern PV systems, including current, voltage, power, irradiance, and temperature-based variables, may exhibit high-dimensional, correlated, and nonlinear characteristics. Therefore, fault classification performance depends not only on the selected classifier algorithm but also on how measurement data are represented and how effectively dimensionality reduction is performed.

In this study, a comparative analysis framework is presented to evaluate linear and kernel-based nonlinear dimensionality reduction techniques for PV fault classification. Within this scope, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) were examined as linear methods, while Kernel Principal Component Analysis (KPCA), Kernel Linear Discriminant Analysis (KLDA), and the KPCA-ICA approach were investigated as kernel-based nonlinear methods. In this study, the KPCA-ICA approach was implemented by applying ICA to the nonlinear representation space obtained through KPCA.

The analyses were carried out using a simulation-based dataset obtained from a 250 kW grid-connected PV system modeled in the MATLAB/Simulink environment. The dataset included four different system states: normal operation, string fault, string-to-ground fault, and string-to-string fault. Data preprocessing, standardization, dimensionality reduction, and classification steps were applied using current, voltage, power, irradiance, temperature, and derived statistical features obtained from the PV system. Accuracy, AUC, precision, recall, F1-score, specificity, and processing time were considered as performance evaluation metrics.

The results showed that, among the dimensionality reduction methods, KPCA provided the most successful representation approach in terms of AUC and F1-score within the scope of this dataset and evaluation protocol. Although KPCA and PCA achieved the same accuracy level, the higher AUC and F1-score values obtained by KPCA indicated that the nonlinear representation space improved the discriminative capability among fault classes. PCA, on the other hand, was evaluated as a strong alternative for engineering applications due to its linear, simple, and interpretable structure. While LDA, ICA, and KLDA produced moderate results in terms of overall classification performance, the KPCA-ICA approach showed limited classification success for this dataset.

In addition, robustness scenarios were generated on the test data to evaluate the performance of the methods under field-like conditions, including noise, random missing data, block missing data, sensor drift, single-sensor loss, and irradiance-band-based operating conditions. The robustness analyses showed that although KPCA stood out in terms of overall performance, the KLDA-based structure provided the strongest result under environmental variability, particularly in the low-irradiance band. This finding indicates that dimensionality reduction methods should be evaluated not only in terms of general classification performance but also with respect to their stability under different operating and degradation conditions.

In conclusion, this study demonstrates that fault classification performance in PV systems depends not only on classifier selection but also on the quality of data representation and robustness performance under degraded conditions. The obtained findings reveal that linear and nonlinear dimensionality reduction techniques offer different advantages in PV fault diagnosis and that the selection of an appropriate representation method has a decisive effect on classification performance.

**Keywords:** Dimensionality reduction, kernel methods, photovoltaic fault detection, machine learning, classification, robustness analysis.

# İÇİNDEKİLER

	Sayfa
ÖN SÖZ .....	i
ÖZET .....	ii
ABSTRACT .....	iv
İÇİNDEKİLER.....	vi
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ.....	ix
KISALTMALAR VE SİMGELER LİSTESİ.....	x
<b>1. GİRİŞ.....</b>	<b>1</b>
<b>1.1. Fotovoltaik Sistemlerde Enerji Üretimi ve Arıza Problemi .....</b>	<b>2</b>
<b>1.2. Veri Odaklı Arıza Tespiti Yaklaşımları .....</b>	<b>3</b>
<b>1.3. Boyut Azaltma Tekniklerinin Önemi .....</b>	<b>5</b>
<b>1.4. Problemin Matematiksel Tanımı .....</b>	<b>6</b>
<b>1.5. Çalışmanın Amacı ve Katkıları.....</b>	<b>7</b>
<b>1.6. Tezin Organizasyonu.....</b>	<b>9</b>
<b>2. KURAMSAL VE KAVRAMSAL TEMELLER .....</b>	<b>10</b>
<b>2.1. Fotovoltaik Sistem Yapısı .....</b>	<b>10</b>
<b>2.2. Fotovoltaik Sistem Dinamikleri.....</b>	<b>12</b>
<b>2.3. Fotovoltaik Sistemlerde Arıza Türleri.....</b>	<b>16</b>
<b>2.3.1. String arızaları .....</b>	<b>19</b>
<b>2.3.2. String-toprak arızaları .....</b>	<b>20</b>
<b>2.3.3. String-string arızaları.....</b>	<b>21</b>
<b>2.4. Boyut Azaltma ve Özellik Çıkarma Teorisi .....</b>	<b>23</b>
<b>2.5. Doğrusal ve Çekirdek Tabanlı Boyut Azaltma Teknikleri.....</b>	<b>27</b>
<b>2.5.1. Temel bileşen analizi (PCA).....</b>	<b>30</b>
<b>2.5.2. Doğrusal ayırma analizi (LDA).....</b>	<b>32</b>
<b>2.5.3. Bağımsız bileşen analizi (ICA).....</b>	<b>33</b>
<b>2.5.4. Çekirdek tabanlı boyut azaltma yaklaşımı.....</b>	<b>34</b>
<b>2.5.5. Çekirdek temel bileşen analizi (KPCA).....</b>	<b>36</b>

2.5.6. Çekirdek doğrusal ayırma analizi (KLDA).....	37
2.5.7. KPCA-ICA yaklaşımı.....	38
2.5.8. Yöntemlerin FV arıza sınıflandırması açısından değerlendirilmesi .....	39
2.6. Destek vektör makineleri ve sınıflandırma mantığı.....	40
2.6.1. Temel SVM mantığı.....	41
2.6.2. Optimizasyon problemi .....	43
2.6.3. Doğrusal olmayan sınıflandırma ve kernel mantığı .....	44
2.6.4. Çok sınıflı SVM yaklaşımı .....	46
2.6.5. SVM'nin fotovoltaik arıza sınıflandırması açısından uygunluğu .....	47
2.6.6. Performans ölçütleri ile ilişkisi .....	49
<b>3. FOTOVOLTAİK SİSTEM MODELİ VE VERİ SETİ OLUŞTURMA SÜRECİ .....</b>	<b>52</b>
3.1. Simülasyon Ortamı ve Fotovoltaik Sistem Modeli.....	52
3.2. Arıza Senaryoları ve Veri Seti Oluşturma Süreci .....	54
3.3. Özellik Mühendisliği ve Veri Ön İşleme.....	57
3.4. Boyut Azaltma ve Model Eğitimi Süreci .....	61
3.5. Performans Değerlendirme Ölçütleri .....	64
3.6. Bozulmuş Senaryoları Altında Deney Protokolü .....	66
<b>4. BULGULAR .....</b>	<b>68</b>
4.1. Veri Ön İşleme ve Özellik Uzayı Analizi .....	68
4.2. Boyut Azaltma Yöntemlerinin Karşılaştırılması.....	72
4.3. Sınıflandırma Performans Analizi.....	74
4.4. Bozulmuş Senaryoları Altında Dayanıklılık Analizi.....	77
4.5. Çevresel Koşullara Göre Performans Değişimi.....	80
4.6. Bulguların Genel Değerlendirilmesi .....	81
<b>5. TARTIŞMA.....</b>	<b>83</b>
5.1. Çalışmanın Sınırlılıkları .....	86
<b>6. SONUÇLAR VE ÖNERİLER.....</b>	<b>88</b>
<b>KAYNAKÇA.....</b>	<b>93</b>
<b>EK-1: MATLAB SİMULASYON DEVRESİ.....</b>	<b>95</b>

## TABLULAR LİSTESİ

	Sayfa
<b>Tablo 2.1.</b> Fotovoltaik sistem dinamiklerini etkileyen temel değişkenler ve sinyallere etkileri .....	15
<b>Tablo 2.2.</b> Kararlı durum ve geçici durum sinyal davranışlarının karşılaştırılması .....	16
<b>Tablo 2.3.</b> Fotovoltaik sistemlerde temel arıza türleri ve ölçüm sinyallerine etkileri .....	18
<b>Tablo 2.4.</b> Arıza türlerinin güvenlik, verim ve teşhis zorluğu açısından karşılaştırılması .....	18
<b>Tablo 2.5.</b> Özellik çıkarma ve boyut azaltma kavramlarının karşılaştırılması.....	24
<b>Tablo 2.6.</b> İyi bir özellik temsilinin taşınması gereken temel nitelikler .....	25
<b>Tablo 2.7.</b> Doğrusal boyut azaltma tekniklerinin karşılaştırılması .....	28
<b>Tablo 2.8.</b> Çekirdek tabanlı boyut azaltma tekniklerinin karşılaştırılması .....	29
<b>Tablo 2.9.</b> Tez kapsamında kullanılan boyut azaltma tekniklerinin genel değerlendirmesi .....	30
<b>Tablo 2.10.</b> SVM sınıflandırma yapısının temel bileşenleri .....	48
<b>Tablo 2.11.</b> SVM'nin fotovoltaik arıza sınıflandırmasındaki avantaj ve sınırlılıkları .....	49
<b>Tablo 3.1.</b> Simülasyonda kullanılan değişkenler ve çalışma aralıkları .....	56
<b>Tablo 3.2.</b> Eğitim veri setinin arıza sınıflarına göre dağılımı .....	56
<b>Tablo 3.3.</b> Fotovoltaik sistemden elde edilen özellikler .....	58
<b>Tablo 3.4.</b> Fotovoltaik sistemden elde edilen değer aralıkları .....	60
<b>Tablo 4.1.</b> Boyut azaltma yöntemlerine göre en iyi sınıflandırma performansları.....	72
<b>Tablo 4.2.</b> Bozulmuş senaryolarına göre en başarılı yöntemler .....	79

## ŞEKİLLER LİSTESİ

Sayfa

Şekil 2.1. Tek diyot modeline göre fotovoltaik hücre eşdeğer devresi .....	10
Şekil 2.2. Şebeke bağlantılı fotovoltaik sistemin blok diyagramı.....	11
Şekil 2.3. Işınım ve sıcaklık değişimlerinin FV çıkış karakteristiklerine etkisi .....	13
Şekil 2.4. Fotovoltaik sistemde dinamik etki zinciri .....	14
Şekil 2.5. Fotovoltaik sistemlerde temel arıza türlerinin sınıflandırılması.....	17
Şekil 2.6. String arızasının FV dizi akım dağılımı üzerindeki etkisinin şematik gösterimi .....	20
Şekil 2.7. Topraklama arızasının fotovoltaik sistemde oluşum mekanizması.....	21
Şekil 2.8. String-string arızasında paralel kollar arasındaki istenmeyen akım yolu .....	22
Şekil 2.9. Ham ölçüm verisinden boyut azaltılmış özellik uzayına geçiş süreci .....	26
Şekil 2.10. Yüksek boyutlu özellik uzayı ile boyut azaltılmış temsil uzayının kavramsal karşılaştırması.....	26
Şekil 2.11. PCA’da maksimum varyans doğrultularına projeksiyonun şematik gösterimi .....	32
Şekil 2.12. Çekirdek yaklaşımında verinin doğrusal olmayan ayırmadan doğrusal ayrılabilir örtük uzaya taşınmasının kavramsal gösterimi .....	36
Şekil 2.13. SVM’de maksimum marjlı karar sınırı ve destek vektörlerinin şematik gösterimi .....	42
Şekil 2.14. Doğrusal ayrılmayan verinin kernel yaklaşımı ile ayrıştırılmasının kavramsal gösterimi .....	45
Şekil 2.15. Boyut azaltma sonrası SVM tabanlı sınıflandırma akışının şematik gösterimi .....	50
Şekil 3.1. Simülasyonda kullanılan fotovoltaik dizi yapısı.....	53
Şekil 4.1. Normalize edilmiş özellik uzayında arıza sınıflarının dağılımı .....	69
Şekil 4.2. Range-1 özelliğinin arıza sınıflarına göre frekans dağılımı.....	70
Şekil 4.3. Seçilmiş özellikler arası korelasyon ısı haritası .....	71
Şekil 4.4. PCA tabanlı sınıflandırma modeli için karışıklık matrisi .....	75
Şekil 4.5. PCA tabanlı sınıflandırma modeli için ROC eğrileri .....	76
Şekil 4.6. Gauss gürültüsü seviyesine göre yöntemlerin F1-macro değişimi .....	77
Şekil 4.7. Eksik veri oranına göre yöntemlerin F1-macro değişimi .....	78

## KISALTMALAR VE SİMGELER LİSTESİ

$\alpha$	: Alfa
$\sigma$	: Standart Sapma
$\lambda$	: Özdeğer
$\gamma$	: RBF çekirdek parametresi
$\Sigma$	: Kovaryans matrisi
$\varphi(\cdot)$	: Doğrusal Olmayan Eşleme Fonksiyonu
$K(x_i, x_j)$	: Çekirdek fonksiyonu
$\xi_i$	: SVM gevşeklik değişkeni
$x_i$	: i'inci gözleme ait özellik vektörü
$y_i$	: i'inci gözleme ait sınıf etiketi
$z_i$	: Boyut azaltılmış özellik vektörü
$\hat{y}_i$	: Tahmin edilen sınıf etiketi
$A$	: ICA karışım matrisi
$C$	: Bağlama göre sınıf kümesi veya SVM ceza parametresi
$d$	: Ham özellik uzayındaki boyut sayısı
$F / H$	: Örtük özellik uzayı
$N$	: Toplam örnek sayısı
$m$	: Boyut azaltma sonrası özellik sayısı
$s$	: Bağımsız kaynak bileşenleri
$\hat{s}$	: Tahmin edilen bağımsız bileşenler
$W$	: Dönüşüm matrisi
$w$	: Özvektör veya SVM karar hiper düzleminin normal vektörü
$b$	: SVM bias terimi
$AC$	: Alternating Current (Alternatif Akım)

<b>AUC</b>	: Area Under the Curve (Eđri Altında Kalan Alan)
<b>CA</b>	: Classification Accuracy (Sınıflandırma Doğruluđu)
<b>CN2</b>	: Kural Tabanlı Sınıflandırma Algoritması
<b>CNN</b>	: Convolutional Neural Network (Evrışimli Sinir Ađı)
<b>DC</b>	: Direct Current (Dođru Akım)
<b>F1</b>	: Kesinlik ve duyarlılık deđerlerinin harmonik ortalaması
<b>FP/FN</b>	: Yanlış pozitif ve yanlış negatif
<b>FV / PV</b>	: Fotovoltaik
<b>I</b>	: Akım
<b>ICA</b>	: Independent Component Analysis (Bađımsız Bileşen Analizi)
<b>I<sub>fault</sub></b>	: Stringler arası arıza akımı
<b>I<sub>g</sub></b>	: Topraklama arızası akımı
<b>IR</b>	: Irradiance (Güneş ışınımı)
<b>I<sub>total</sub></b>	: Toplam akım
<b>KLDA</b>	: Kernel Linear Discriminant Analysis (Çekirdek Doğrusal Ayırma Analizi)
<b>KPCA</b>	: Kernel Principal Component Analysis (Çekirdek Temel Bileşen Analizi)
<b>KPCA-ICA</b>	: KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanmasına dayalı yaklaşım
<b>LDA</b>	: Linear Discriminant Analysis (Dođrusal Ayırma Analizi)
<b>LR</b>	: Logistic Regression (Lojistik Regresyon)
<b>LSTM</b>	: Long Short-Term Memory (Uzun Kısa Süreli Bellek)
<b>ML</b>	: Machine Learning (Makine Öğrenmesi)
<b>MPPT</b>	: Maximum Power Point Tracking (Maksimum Güç Noktası Takibi)
<b>PCA</b>	: Principal Component Analysis (Temel Bileşen Analizi)

<b>P<sub>dc</sub></b>	: DC güç
<b>Range1–4</b>	: Akım veya sinyal grupları arasındaki deęişim aralığını temsil eden türetilmiş özellikler
<b>RBF</b>	: Radial Basis Function (Radyal Tabanlı Fonksiyon)
<b>ROC</b>	: Receiver Operating Characteristic (Alıcı İşletim Karakteristięi)
<b>R<sub>s</sub></b>	: Seri direnç etkisi
<b>R<sub>f</sub></b>	: Arıza direnci
<b>R<sub>g</sub></b>	: Topraklama arızası direnci
<b>S<sub>B</sub></b>	: Sınıflar arası saçılım matrisi
<b>S<sub>w</sub></b>	: Sınıf içi saçılım matrisi
<b>SVM</b>	: Support Vector Machine (Destek Vektör Makinesi)
<b>T</b>	: Temperature (Sıcaklık)
<b>TP/TN</b>	: Doğru pozitif, doğru negatif
<b>V</b>	: Gerilim
<b>V<sub>1</sub>, V<sub>2</sub></b>	: Arızaya dahil olan string gerilimleri
<b>V<sub>dc</sub></b>	: DC bağlantı gerilimi
<b>V<sub>string</sub></b>	: String terminal gerilimi
<b>V<sub>pv</sub></b>	: FV kaynak gerilimi

## 1. GİRİŞ

Enerji üretiminde sürdürülebilirliğin sağlanması, çevresel etkilerin azaltılması ve enerji arz güvenliğinin artırılması, günümüz enerji sistemlerinin temel öncelikleri arasında yer almaktadır. Bu kapsamda yenilenebilir enerji kaynakları içerisinde fotovoltaik (FV) sistemler; modüler yapıları, düşük işletme gereksinimleri, farklı ölçeklerde uygulanabilmeleri ve küresel kurulu güç içerisindeki paylarının hızla artması nedeniyle yaygın biçimde kullanılmaktadır (IEA-PVPS, 2024; IRENA, 2024). Şebeke bağlantılı FV santrallerin yaygınlaşması, bu sistemlerin yalnızca enerji üretim kapasitesi açısından değil, güvenilirlik, süreklilik, izlenebilirlik ve işletme verimliliği açısından da değerlendirilmesini gerekli kılmaktadır (Mellit vd., 2018; Sepúlveda-Oviedo vd., 2023; El-Banby vd., 2023).

FV sistemlerin uzun süreli ve güvenilir işletimi, nominal çalışma koşullarında enerji üretiminin sürdürülmesinin yanı sıra sistem içerisinde oluşabilecek elektriksel arızaların zamanında ve doğru biçimde tespit edilmesine de bağlıdır. Dizi arızaları, string-toprak arızaları, stringler arası kısa devreler, bağlantı problemleri ve izolasyon kaynaklı hatalar; enerji üretiminde azalma, ekipman zorlanması, güvenlik riski ve bakım maliyetlerinde artış gibi olumsuz sonuçlara yol açabilmektedir (Mellit vd., 2018; Ghoneim vd., 2021; El-Banby vd., 2023). Bu nedenle FV sistem davranışının doğru analiz edilmesi ve arızaların erken aşamada belirlenmesi, modern enerji sistemleri açısından önemli bir mühendislik gerekliliğidir.

Bununla birlikte FV sistemlerde arıza teşhisi, yalnızca klasik elektriksel gözlem veya sabit sınır değer denetimi ile güvenilir biçimde çözülebilecek basit bir problem değildir. FV sistem davranışı; güneş ışınımı, hücre sıcaklığı, yük değişimi, güç elektroniği çalışma koşulları, sistem konfigürasyonu ve çevresel değişkenler gibi çok sayıda etkene bağlı olarak değişmektedir. Bu nedenle aynı arıza türü farklı işletme koşullarında farklı elektriksel izler bırakabilmekte; buna karşılık normal çalışma koşullarındaki doğal değişkenlik bazı durumlarda arıza benzeri sinyal davranışları oluşturabilmektedir. Bu durum, FV arıza teşhis probleminin çok değişkenli veri analizi ve örüntü tanıma yaklaşımıyla ele alınmasını gerekli kılmaktadır (Mellit vd., 2018; Madeti ve Singh, 2017; Triki-Lahiani vd., 2018).

FV sistemlerden elde edilen akım, gerilim, güç, ışınım ve sıcaklık tabanlı ölçüm verileri, sistem durumuna ilişkin anlamlı bilgiler taşıyan istatistiksel ve zamansal yapılar içermektedir. Bu verilerin uygun biçimde işlenmesi, temsil edilmesi ve sınıflandırma algoritmalarına uygun hale getirilmesi, arıza teşhis performansını doğrudan etkilemektedir. Bu doğrultuda veri odaklı

yaklaşımlar, FV sistemlerde arıza tespiti ve sınıflandırılması için güçlü bir çözüm çerçevesi sunmaktadır (Chen vd., 2018; Eskandari vd., 2021; Sepúlveda-Oviedo vd., 2023).

Literatürde FV sistemlerde arıza tespiti için farklı makine öğrenmesi tabanlı sınıflandırıcıların kullanıldığı çok sayıda çalışma bulunmaktadır. Chine vd. (2016) yapay sinir ağları kullanarak FV sistem arızalarını tespit etmiş; Dhimish ve Holmes (2016) oransız fark yöntemine dayalı bir yaklaşım önermiştir. Garoudja vd. (2017) istatistiksel süreç kontrolü temelli bir izleme sistemi geliştirirken; Chen vd. (2018) ve Eskandari vd. (2021) destek vektör makineleri ve çeşitli denetimli öğrenme algoritmalarını değerlendirmiştir. Ghoneim vd. (2021) ise 250 kW gücündeki şebeke bağlantılı bir FV sistem üzerinde Random Forest, AdaBoost, Naive Bayes ve Logistic Regression gibi farklı sınıflandırıcıları karşılaştırmıştır. El-Banby vd. (2023) derin öğrenme tabanlı mimarileri arıza tespitine uygulamıştır. Bu çalışmaların büyük bölümü sınıflandırıcı seçimine odaklanmış olmakla birlikte, sınıflandırma başarısını doğrudan etkileyen veri temsil aşamasının özellikle doğrusal ve çekirdek tabanlı doğrusal olmayan boyut azaltma tekniklerinin karşılaştırmalı olarak ele alındığı çalışmalar daha sınırlıdır. Bu çalışma, söz konusu eksikliği gidermek amacıyla yöntemleri aynı veri seti, aynı sınıflandırma altyapısı ve aynı performans ölçütleri altında değerlendirerek literatürdeki sınıflandırıcı odaklı yaklaşımları tamamlayıcı bir veri temsil analizi sunmaktadır.

### **1.1. Fotovoltaik Sistemlerde Enerji Üretimi ve Arıza Problemi**

FV sistemlerin enerji üretim kapasitesi, temel olarak hücre bazlı fotovoltaik dönüşüm verimine, modül-string-dizi hiyerarşisine ve güç elektroniği dönüşüm zincirinin etkinliğine bağlıdır. Tek bir FV hücrenin ürettiği güç, standart test koşullarında (1000 W/m<sup>2</sup> ışınım, 25°C hücre sıcaklığı) genellikle 1–5 W aralığında kalmaktadır. Bu nedenle pratikte faydalı güç seviyeleri elde edebilmek için hücreler seri bağlanarak modüller, modüller seri bağlanarak stringler ve stringler paralel bağlanarak FV diziler oluşturulmaktadır. Bu çok katmanlı yapı, sistemin enerji üretim kapasitesini belirlerken aynı zamanda arıza davranışlarının hangi elektriksel büyüklükler üzerinden gözlemlenebileceğini de doğrudan etkilemektedir. Ancak FV sistemlerin enerji üretim performansı; güneş ışınımı, hücre sıcaklığı, modül karakteristikleri, dizi konfigürasyonu, güç elektroniği bileşenleri ve şebeke bağlantı koşulları gibi çok sayıda değişkene bağlı olarak değişmektedir (Villalva vd., 2009; Madeti ve Singh, 2017).

FV sistemlerde enerji üretimini ve güvenilirliği tehdit eden başlıca arıza türleri DC tarafında ortaya çıkmaktadır. String arızaları, bir veya birden fazla FV stringin kısmen veya tamamen devre dışı kalmasına yol açarak toplam dizi akımında ölçülebilir bir azalmaya neden

olmaktadır. String-toprak arızaları, DC iletkenlerin yalıtım bütünlüğünün bozulması sonucunda toprak potansiyeline karşı istenmeyen kaçak akım yolları oluşturmakta ve hem enerji kaybına hem de ciddi güvenlik riskine yol açmaktadır. String-string arızaları ise farklı potansiyel seviyelerindeki iki string arasında harici istenmeyen bir iletim yolu oluşması durumunu ifade etmektedir. Daha ileri durumlarda ise aşırı akım, gerilim dengesizliği, yerel ısınma, ekipman zorlanması ve güvenlik problemleri gibi etkiler ortaya çıkabilmektedir (Mellit vd., 2018; Triki-Lahiani vd., 2018; Hong ve Pula, 2022). Bu nedenle arızaların erken aşamada doğru biçimde belirlenmesi, sistem sürekliliğinin korunması ve güvenlik risklerinin azaltılması açısından kritik öneme sahiptir.

FV arıza teşhisini zorlaştıran temel unsurlardan biri, sistem davranışının çevresel koşullara yüksek duyarlılık göstermesidir. Aynı arıza türü, farklı ışınım ve sıcaklık seviyelerinde farklı elektriksel izler bırakabilmekte; buna karşılık normal çalışma koşullarındaki değişkenlik bazı durumlarda arıza ile benzer sinyal davranışları üretebilmektedir. Bu durum, arıza etkileri ile işletme koşullarına bağlı doğal değişimlerin birbirinden ayrıştırılmasını güçleştirmektedir (Madeti ve Singh, 2017; Triki-Lahiani vd., 2018).

Öte yandan FV sistemlerden elde edilen ölçüm sinyalleri, arıza durumlarında yalnızca tek bir büyüklükte değil, çok sayıda değişkenin birlikte oluşturduğu yapıda değişim göstermektedir. Akım, gerilim, güç, ışınım ve sıcaklık gibi temel ölçümlerin yanı sıra bunlardan türetilen istatistiksel göstergeler de sistem durumuna ilişkin tamamlayıcı bilgi sunmaktadır. Dolayısıyla güvenilir bir arıza teşhis yaklaşımı, yalnızca tekil sinyal seviyelerine değil, çok değişkenli ölçüm yapısının bütüncül biçimde değerlendirilmesine dayanmalıdır (Mellit vd., 2018; Hong ve Pula, 2022).

Bu çerçevede FV sistemlerde arıza problemi, ölçüm verilerinden anlamlı bilgi çıkarılmasını ve bu bilginin sınıflandırma amacıyla kullanılmasını gerektiren veri odaklı bir problem olarak ele alınmalıdır. Bu doğrultuda bir sonraki alt bölümde veri odaklı arıza tespiti yaklaşımları incelenmektedir.

## **1.2. Veri Odaklı Arıza Tespiti Yaklaşımları**

Modern FV izleme sistemleri, sistem davranışını değerlendirmek amacıyla akım, gerilim, güç, ışınım, sıcaklık ve bunlardan türetilen çok sayıda değişken üzerinden ölçüm verisi üretmektedir. Bu veriler çoğu durumda yüksek boyutlu, doğrusal olmayan, gürültü içeren ve işletme koşullarına bağlı olarak değişkenlik gösteren bir yapıya sahiptir. Bu nedenle FV sistemlerde arıza teşhisi problemi, yalnızca belirli eşik değerlerinin aşılmasına dayalı klasik

izleme mantığıyla güvenilir biçimde çözülememektedir (Madeti ve Singh, 2017; Triki-Lahiani vd., 2018).

Geleneksel arıza tespit yaklaşımları çoğunlukla önceden tanımlanmış limit değerler, uzman bilgisine dayalı kurallar veya belirli sinyal değişkenlerinin tekil incelenmesi üzerine kuruludur. Bu yöntemler basit ve yorumlanabilir olmalarına rağmen, çok değişkenli ölçüm yapılarında arıza örüntülerinin karmaşık biçimde ortaya çıktığı durumlarda yetersiz kalabilmektedir. Aynı arıza türünün farklı işletme koşullarında farklı elektriksel davranış göstermesi veya farklı arızaların benzer sinyal değişimleri üretmesi, klasik yaklaşımların ayırım gücünü azaltmaktadır. Bu bağlamda arıza teşhisi probleminin, çok değişkenli veri analizi ve örüntü tanıma temelli bir çerçevede ele alınması gerekmektedir (Mellit vd., 2018; Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

Veri odaklı arıza tespiti yaklaşımları, ölçüm verilerinden öğrenilen istatistiksel ve yapısal ilişkileri kullanarak sistem durumunun belirlenmesini amaçlamaktadır. Bu yaklaşımlarda genel işlem zinciri; veri toplama, ön işleme, özellik çıkarma, temsil iyileştirme, model eğitimi ve sınıflandırma adımlarından oluşmaktadır. Özellikle makine öğrenmesi tabanlı yöntemler, yüksek boyutlu ölçüm uzayında doğrudan gözlenemeyen ayırt edici örüntüleri ortaya çıkarabilmeleri nedeniyle FV sistemlerde arıza teşhisi için güçlü bir alternatif sunmaktadır (Chine vd., 2016; Dhimish ve Holmes, 2016; Garoudja vd., 2017; Chen vd., 2018; Eskandari vd., 2021; Sepúlveda-Oviedo vd., 2023).

Makine öğrenmesi tabanlı arıza teşhisinde temel amaç, normal çalışma durumu ile farklı arıza türlerine ait ölçüm örüntülerini birbirinden ayırt edebilen bir karar yapısı oluşturmaktır. Bu kapsamda destek vektör makineleri, karar ağaçları, rastgele orman, lojistik regresyon, yapay sinir ağları ve çeşitli derin öğrenme yaklaşımları literatürde yaygın biçimde kullanılmaktadır (Chine vd., 2016; Ghoneim vd., 2021; Eskandari vd., 2021; El-Banby vd., 2023; Sepúlveda-Oviedo vd., 2023). Ancak bu yöntemlerin başarısı, çoğu zaman sınıflandırıcıdan önce gerçekleştirilen veri ön işleme ve özellik temsil aşamalarına bağlıdır.

FV sistemlerden elde edilen ham veriler çoğu zaman tekrarlı bilgi, yüksek korelasyon, gürültü ve sınıflar arası örtüşme içermektedir. Bu durum model karmaşıklığını artırmakta ve sınıflandırma performansını sınırlayabilmektedir. Dolayısıyla veri odaklı arıza teşhisinde temel mesele yalnızca uygun bir öğrenme algoritması seçmek değil, veriyi daha ayırt edici ve daha düşük boyutlu bir temsil uzayına dönüştürmektir (Jolliffe, 2002; Schölkopf vd., 1998; Baudat ve Anouar, 2000).

### 1.3. Boyut Azaltma Tekniklerinin Önemi

FV sistemlerden elde edilen ölçüm verileri; akım, gerilim, güç, ışıınım ve sıcaklık gibi çok sayıda parametreden oluşan yüksek boyutlu veri yapıları üretmektedir. Bu veri yapılarında yer alan özellikler çoğu zaman birbirleriyle yüksek derecede korelasyon göstermekte, bazı değişkenler benzer bilgiyi tekrar etmekte ve bazıları ise sınıflandırma açısından sınırlı ayırt edici değer taşımaktadır. Bu durum, öğrenme algoritmalarının gereksiz bilgiyle beslenmesine, hesaplama yükünün artmasına ve sınıflar arası ayırımı zayıflamasına neden olabilmektedir. Dolayısıyla FV arıza sınıflandırmasında temel sorun yalnızca daha fazla veri toplamak değil, veriler içerisinde karar verme açısından anlamlı bir temsil elde edebilmektir (Jolliffe, 2002; Hyvärinen vd., 2001; Sepúlveda-Oviedo vd., 2023).

Boyut azaltma teknikleri, yüksek boyutlu ölçüm uzayını daha düşük boyutlu fakat daha bilgilendirici bir özellik uzayına dönüştürerek bu probleme çözüm sunmaktadır. Bu dönüşüm sayesinde veri içerisindeki baskın varyans yapıları, sınıf ayırımı destekleyen doğrultular veya gizli istatistiksel örüntüler daha belirgin hale getirilebilmektedir. Böylece sınıflandırma algoritmalarına daha düzenli, daha kararlı ve daha ayırt edici bir giriş uzayı sağlanmakta; aynı zamanda gürültü, tekrarlı bilgi ve korelasyon etkileri azaltılmaktadır. Bu nedenle boyut azaltma, yalnızca veri sıkıştırma işlemi değil, sınıflandırma performansını doğrudan etkileyen temel bir ön işleme ve temsil iyileştirme aşaması olarak değerlendirilmelidir (Jolliffe, 2002; Schölkopf vd., 1998; Baudat ve Anouar, 2000).

FV arıza teşhisinde kullanılan boyut azaltma yaklaşımları genel olarak doğrusal ve doğrusal olmayan yöntemler şeklinde iki ana grupta incelenebilir. Temel Bileşen Analizi, Doğrusal Ayırma Analizi ve Bağımsız Bileşen Analizi gibi doğrusal yöntemler; verideki varyans yapısı, sınıf ayırımı ve bağımsız bileşenler gibi farklı istatistiksel özellikleri öne çıkarmaktadır. Buna karşılık bu yöntemlerin çekirdek tabanlı uzantıları olan Çekirdek Temel Bileşen Analizi, Çekirdek Doğrusal Ayırma Analizi ve KPCA-ICA yaklaşımı, doğrusal uzayda yeterince ayrıştırılamayan karmaşık örüntülerin daha yüksek boyutlu örtük özellik uzaylarında temsil edilmesine olanak sağlamaktadır (Fisher, 1936; Schölkopf vd., 1998; Baudat ve Anouar, 2000; Bach ve Jordan, 2002).

Uygun boyut azaltma yönteminin seçimi, yalnızca yöntemin teorik yapısına değil, FV sinyallerin farklı işletme koşullarındaki değişken yapısını ne ölçüde temsil edebildiğine de bağlıdır. Bu çalışmada akım, gerilim, güç, ışıınım, sıcaklık ve türetilmiş istatistiksel göstergeler birlikte kullanıldığından özellik uzayı genişlemekte; bu nedenle boyut azaltma, hesaplama

yükünü azaltmanın yanında arıza sınıflarını daha ayırt edici bir temsil uzayında ele almak için temel bir modelleme adımı olarak kullanılmaktadır. Bu yaklaşımın daha açık biçimde ortaya konulabilmesi için bir sonraki alt bölümde, ölçüm vektöründen arıza sınıf etiketine giden sınıflandırma yapısı matematiksel olarak tanımlanmaktadır.

#### 1.4. Problemin Matematiksel Tanımı

FV sistemlerde arıza sınıflandırma problemi, çok değişkenli ölçüm verilerinden yararlanılarak her bir gözlemin ait olduğu işletme durumunun belirlenmesi şeklinde tanımlanabilir. Bu çalışmada sistemden elde edilen her bir örnek; akım, gerilim, güç, ısıtım, sıcaklık ve bunlardan türetilen istatistiksel göstergeleri içeren bir özellik vektörü ile temsil edilmektedir. Buna göre her bir gözlem,

$$\mathbf{X}_i = [X_{i1}, X_{i2}, \dots, X_{id}]^T \in \mathbb{R}^d \quad (\text{D.1})$$

şeklinde ifade edilir. Burada  $d$ , ham veri uzayındaki toplam özellik sayısını göstermektedir. Her bir örnek için amaç, ilgili gözlemin ait olduğu sınıf etiketini belirlemektir. Bu sınıf etiketi ise,

$$\mathbf{y}_i \in \mathbf{C} = \{\mathbf{c}^1, \mathbf{c}^2, \dots, \mathbf{c}^k\} \quad (\text{D.2})$$

şeklinde tanımlanır. Burada  $\mathbf{C}$ , normal çalışma durumu ile farklı arıza türlerinden oluşan sonlu sınıf kümesini ifade etmektedir. Bu tez kapsamında sınıf kümesi; normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört temel işletme durumunu içermektedir.

Bu durumda sınıflandırma problemi, ölçüm vektörünü sınıf etiketine eşleyen bir karar fonksiyonunun öğrenilmesi olarak yazılabilir:

$$\hat{\mathbf{y}}_i = \mathbf{f}(\mathbf{x}_i), \quad \mathbf{f}: \mathbb{R}^d \rightarrow \mathbf{C} \quad (\text{D.3})$$

Bu matematiksel gösterim, FV sistemlerden elde edilen çok boyutlu ölçüm verilerinin uygun bir karar mekanizması yardımıyla normal durum ve farklı arıza sınıflarına ayrılmasını ifade etmektedir. Ancak temel güçlük, ham özellik uzayının yüksek boyutlu, korelasyonlu ve kısmen gürültülü bir yapıya sahip olmasıdır. Bu nedenle karar fonksiyonunun doğrudan ham veri uzayı üzerinde kurulması, sınıflar arası ayrımı zorlaştırabilmekte ve sınıflandırma performansını sınırlayabilmektedir (Jolliffe, 2002; Hyvärinen vd., 2001; Bishop, 2006).

Bu problemi daha etkin biçimde çözebilmek amacıyla, ham özellik vektörü  $\mathbf{x}_i$  daha düşük boyutlu ve daha ayırt edici bir temsil uzayına dönüştürülmektedir. Doğrusal boyut azaltma durumunda bu dönüşüm genel olarak,

$$\mathbf{z}_i = \mathbf{W}^T \mathbf{x}_i, \mathbf{W} \in \mathbb{R}^{d \times m}, \mathbf{z}_i \in \mathbb{R}^m, m < d \quad (\text{D.4})$$

şeklinde ifade edilir. Burada  $\mathbf{W} \in \mathbb{R}^{d \times m}$ , ham veri uzayını  $m$  boyutlu yeni özellik uzayına eşleyen dönüşüm matrisini;  $\mathbf{z}_i \in \mathbb{R}^m$  ise boyut azaltılmış özellik vektörünü göstermektedir. Genellikle  $m < d$  olup amaç, bilgi kaybını mümkün olduğunca azaltırken sınıflandırma açısından ayırt edici yapıyı koruyan bir temsil elde etmektir.

Doğrusal olmayan boyut azaltma yöntemlerinde ise dönüşüm, örtük bir özellik uzayında tanımlanan doğrusal olmayan bir eşleme yardımıyla gerçekleştirilmektedir. Bu durumda giriş verisi önce,

$$\varphi: \mathbb{R}^d \rightarrow \mathbf{H}, \mathbf{x}_i \rightarrow \varphi(\mathbf{x}_i) \quad (\text{D.5})$$

şeklindeki bir eşleme ile daha yüksek boyutlu bir özellik uzayına taşınmakta, ardından ayrıştırıcı bileşenler bu uzayda elde edilmektedir. Böylece ham veri uzayında doğrusal olarak ayrıştırılamayan örüntüler, çekirdek tabanlı yöntemler yardımıyla daha belirgin hale getirilebilmektedir (Schölkopf vd., 1998; Baudat ve Anouar, 2000).

Sonuç olarak bu tezde ele alınan temel problem, FV sistemlerden elde edilen yüksek boyutlu ölçüm verilerinin uygun boyut azaltma teknikleri ile daha anlamlı bir temsil uzayına dönüştürülmesi ve bu yeni uzayda öğrenilen sınıflandırıcı yardımıyla arıza türlerinin doğru biçimde belirlenmesidir. Bu matematiksel çerçeve doğrultusunda bir sonraki alt bölümde çalışmanın amacı ve literatüre sunduğu katkılar açıklanmaktadır.

### 1.5. Çalışmanın Amacı ve Katkıları

Bu tez çalışmasının temel amacı, FV sistemlerde arıza sınıflandırma problemi için doğrusal ve doğrusal olmayan boyut azaltma tekniklerinin ayırt edici temsil üretme yeteneklerini karşılaştırmalı olarak incelemektir. Bu kapsamda, yüksek boyutlu FV ölçüm verilerinin daha düşük boyutlu ve sınıflandırma açısından daha anlamlı bir özellik uzayına dönüştürülmesinin arıza teşhis performansı üzerindeki etkisi değerlendirilmiştir.

Çalışmada bu amaç doğrultusunda 250 kW gücünde şebeke bağlantılı bir FV sistem MATLAB/Simulink ortamında modellenmiş, normal çalışma ve farklı arıza senaryoları altında sistem verileri üretilmiş ve elde edilen veri kümeleri üzerinde bütünsel bir analiz gerçekleştirilmiştir. Analiz kapsamında normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört farklı işletme durumu ele alınmıştır.

Bu çalışmanın bir diğer amacı, FV arıza teşhisinde yalnızca sınıflandırıcı seçimine odaklanmak yerine, sınıflandırıcı öncesindeki veri temsil aşamasının belirleyici rolünü ortaya

koymaktır. Bu çerçevede Temel Bileşen Analizi, Doğrusal Ayırma Analizi, Bağımsız Bileşen Analizi ile bunların çekirdek tabanlı karşılıkları olan Çekirdek Temel Bileşen Analizi, Çekirdek Doğrusal Ayırma Analizi ve KPCA-ICA yaklaşımı aynı deneysel yapı içerisinde değerlendirilmiştir. KPCA-ICA yaklaşımı, bu tez kapsamında KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde tanımlanmıştır. Böylece her bir yöntemin FV arıza örüntülerini ne ölçüde ayrıştırabildiği karşılaştırmalı olarak incelenmiştir.

Boyut azaltma sonrasında elde edilen temsil uzaylarının sınıflandırma performansı, makine öğrenmesi tabanlı sınıflandırıcılar kullanılarak değerlendirilmiştir. Bu değerlendirme sürecinde doğruluk, kesinlik, duyarlılık, F1 skoru, özgüllük, AUC değeri, karışıklık matrisi ve işlem süresi gibi performans ölçütleri dikkate alınmıştır (Fawcett, 2006; Sokolova ve Lapalme, 2009). Böylece yöntemlerin yalnızca genel doğruluk başarısı değil, aynı zamanda sınıflar arası ayırt edicilik, hata dağılımı ve mühendislik uygulamalarına uygunluk düzeyi de incelenmiştir.

Bu tez çalışmasının başlıca katkıları aşağıdaki gibi özetlenebilir:

- Şebeke bağlantılı 250 kW gücünde bir FV sistem için, normal çalışma durumu ile farklı arıza senaryolarını içeren simülasyon tabanlı bir veri üretim çerçevesi oluşturulmuştur.
- FV sistemden elde edilen akım, gerilim, güç, ışıınım ve sıcaklık tabanlı büyüklüklerden türetilen özelliklerin arıza sınıflandırma açısından taşıdığı ayırt edici bilgi sistematik olarak incelenmiştir.
- Doğrusal ve doğrusal olmayan boyut azaltma yöntemleri aynı veri seti ve aynı değerlendirme altyapısı altında karşılaştırılarak yöntemlerin görelî üstünlükleri ve sınırlılıkları ortaya konmuştur.
- Boyut azaltma yöntemlerinin FV arıza teşhis performansına etkisi; doğruluk, F1 skoru, AUC değeri, karışıklık matrisi, özgüllük ve işlem süresi gibi ölçütler üzerinden değerlendirilmiştir.
- Boyut azaltma yöntemlerinin yalnızca temiz test verisi üzerindeki sınıflandırma başarısı değil, aynı zamanda gürültü, rastgele eksik veri, blok eksik veri, sensör sürüklenmesi, tek sensör kaybı ve farklı ışıınım bandı koşulları altındaki dayanıklılık performansları da karşılaştırmalı olarak incelenmiştir.
- Elde edilen bulgular doğrultusunda, FV arıza sınıflandırmasında hangi tür boyut azaltma yaklaşımının daha uygun bir temsil uzayı sunduğuna ilişkin mühendislik bakış açısıyla bir değerlendirme yapılmıştır.

- Simülasyon tabanlı veri üretimi, özellik mühendisliği, boyut azaltma ve sınıflandırma adımlarını bir arada ele alan bütünlük bir analiz çerçevesi sunulmuştur.

## 1.6. Tezin Organizasyonu

Bu tez çalışması altı ana bölümden oluşmaktadır.

Birinci bölümde, FV sistemlerde arıza tespiti probleminin önemi ortaya konulmuş, veri odaklı arıza teşhis yaklaşımlarının genel çerçevesi sunulmuş, boyut azaltma tekniklerinin bu problem açısından taşıdığı önem açıklanmış ve çalışmanın matematiksel temeli, amacı ile katkıları verilmiştir.

İkinci bölümde, çalışmanın kuramsal ve kavramsal temelleri ele alınmıştır. Bu kapsamda FV sistem yapısı, FV sistem dinamikleri, temel arıza türleri, boyut azaltma ve özellik çıkarma teorisi, doğrusal ve çekirdek tabanlı boyut azaltma yöntemleri ile sınıflandırma yaklaşımı açıklanmıştır.

Üçüncü bölümde, simülasyon ortamı, FV sistem modeli, arıza senaryoları, veri seti oluşturma süreci, özellik mühendisliği, veri ön işleme, boyut azaltma ve model eğitimi süreci ile bozulmuş senaryoları altında kullanılan deney protokolü sunulmuştur.

Dördüncü bölümde, veri ön işleme sonuçları, özellik uzayı analizleri, boyut azaltma yöntemlerinin karşılaştırılması, sınıflandırma performansı ve dayanıklılık analizlerine ilişkin deneysel bulgular değerlendirilmiştir.

Beşinci bölümde, elde edilen bulgular literatür ve mühendislik uygulamaları çerçevesinde tartışılmıştır. Altıncı bölümde ise çalışmanın genel sonuçları, temel çıkarımları ve gelecek çalışmalara yönelik öneriler sunulmuştur.

Çalışmanın yeniden üretilebilirliğini ve yöntemsel şeffaflığını desteklemek amacıyla, MATLAB/Simulink ortamında oluşturulan simülasyon devresi EK-1'de, veri işleme, boyut azaltma ve sınıflandırma adımlarında Python tabanlı yazılım kullanılmıştır.

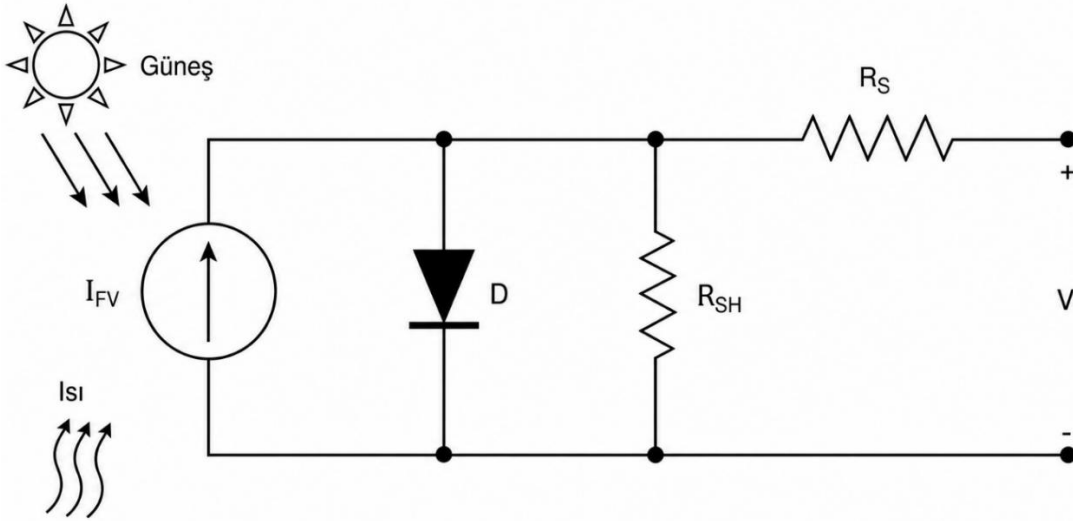
## 2. KURAMSAL VE KAVRAMSAL TEMELLER

Bu bölümde, tez çalışmasının dayandığı kuramsal ve kavramsal altyapı sunulmaktadır. Bu kapsamda öncelikle FV sistemlerin temel yapısı, çalışma dinamikleri ve arıza türleri ele alınmakta; ardından arıza sınıflandırma sürecinde kullanılan özellik çıkarma, boyut azaltma ve sınıflandırma yaklaşımları açıklanmaktadır. Böylece ilerleyen bölümlerde sunulan veri üretimi, özellik mühendisliği, boyut azaltma ve sınıflandırma adımlarının teorik temeli oluşturulmaktadır.

### 2.1. Fotovoltaik Sistem Yapısı

FV sistemler, güneş ışımını doğrudan elektrik enerjisine dönüştüren yarı iletken tabanlı enerji üretim sistemleridir. Bu sistemlerin temel yapı taşı FV hücredir. FV hücre, üzerine düşen ışınım seviyesi ve hücre sıcaklığına bağlı olarak doğrusal olmayan bir akım-gerilim karakteristiği sergilemektedir. Bu doğrusal olmayan yapı, FV sistemlerin modellenmesi ve arıza davranışlarının yorumlanması açısından önemlidir. Bu nedenle FV hücrelerin elektriksel davranışı çoğunlukla tek diyot modeli ile temsil edilmektedir (Villalva vd., 2009).

FV hücrenin eşdeğer elektriksel modeli Şekil 2.1’de gösterilmektedir.



Şekil 2.1. Tek diyot modeline göre FV hücre eşdeğer devresi

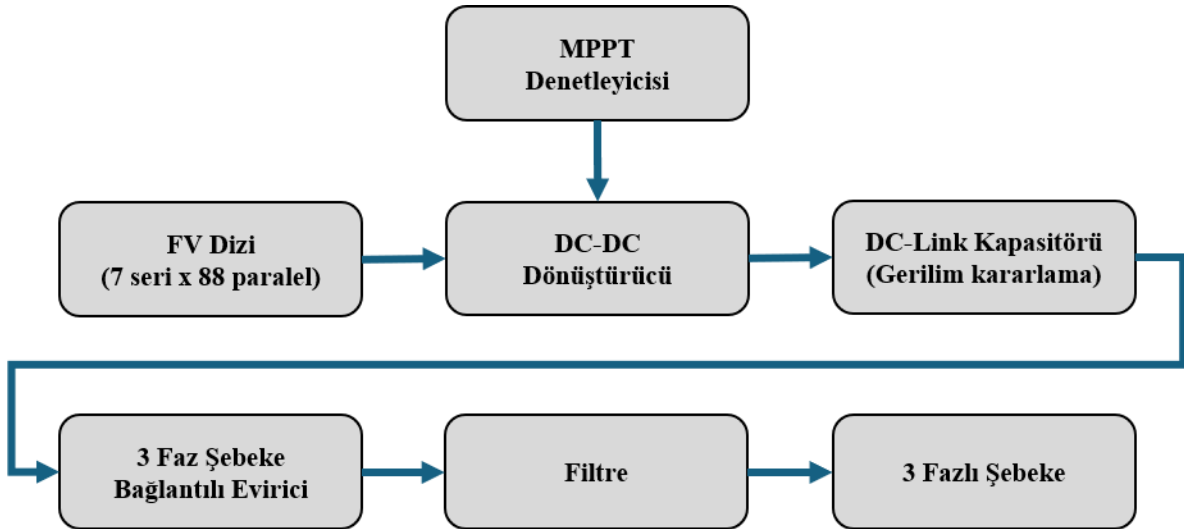
Şekil 2.1’de görüldüğü üzere, FV hücrenin eşdeğer elektriksel modeli; ışık kaynaklı akım üretimini temsil eden bir akım kaynağı, yarı iletken birleşimin doğrultucu davranışını ifade eden diyot ve ideal olmayan etkileri temsil eden seri ve paralel direnç elemanlarından oluşmaktadır. Seri direnç, hücre içi iletim yolları ve bağlantı elemanlarından kaynaklanan kayıpları temsil ederken; paralel direnç kaçak akım etkilerini göstermektedir. Bu yapı, FV

hücrenin gerçek çalışma koşullarındaki ideal olmayan davranışlarının daha doğru biçimde modellenmesine olanak sağlamaktadır (Villalva vd., 2009).

FV enerji üretim yapısı hiyerarşik bir organizasyona sahiptir. Hücrelerin seri ve paralel bağlanmasıyla modüller, modüllerin bir araya gelmesiyle stringler, stringlerin uygun biçimde bağlanmasıyla da FV diziler oluşturulmaktadır. Bu çok katmanlı yapı, sistemin istenen gerilim ve akım seviyelerinde çalışmasını sağlarken, meydana gelen arızaların etkilerinin farklı elektriksel büyüklükler üzerinden gözlenmesine de neden olmaktadır. Başka bir ifadeyle, hücre veya modül düzeyinde başlayan bir bozulma; string akımlarında, dizi çıkış geriliminde veya toplam DC güçte ölçülebilir değişimler oluşturabilmektedir.

Şebeke bağlantılı modern FV sistemler yalnızca FV diziden oluşmamaktadır. Bu sistemler genel olarak FV dizi, maksimum güç noktası takibi sağlayan denetim yapısı, DC-DC dönüştürücü, DC-link katmanlı, inverter, filtre ve şebeke bağlantısından meydana gelmektedir. FV sistemlerde izleme ve performans değerlendirme süreçleri de bu enerji dönüşüm zincirinde yer alan değişkenlerin uygun biçimde ölçülmesine ve analiz edilmesine dayanmaktadır (Madeti ve Singh, 2017; IEC, 2021).

Şebeke bağlantılı FV sistemin genel blok yapısı Şekil 2.2’de sunulmuştur.



Şekil 2.2. Şebeke bağlantılı FV sistemin blok diyagramı

Şekil 2.2 incelendiğinde, FV dizide üretilen DC gücün önce dönüştürücü ve denetim katmanlarından geçtiği, ardından inverter ve filtre yapısı aracılığıyla AC şebekeye aktarıldığı görülmektedir. Bu enerji dönüşüm zincirindeki her bir alt sistem, ölçülen akım, gerilim ve güç sinyallerinin oluşumunda belirleyici rol oynamaktadır. Dolayısıyla sistem üzerinde meydana gelen bir arıza yalnızca tek bir noktada sınırlı kalmamakta; DC tarafta akım dengesizliği, çıkış

gücünde azalma, gerilim seviyelerinde sapma veya şebeke tarafında güç kalitesi bozulması gibi farklı etkiler oluşturabilmektedir.

FV sistem yapısının bütüncül biçimde anlaşılması, arıza sınıflandırma çalışmalarında hangi değişkenlerin neden izlendiğini açıklamak açısından önemlidir. Özellikle string akımları, DC bağlantı gerilimi, toplam DC güç, sıcaklık ve ışıınım gibi büyüklükler, sistemin hem yapısal hem de işletme durumuna ilişkin ayırt edici bilgi taşımaktadır. Bu nedenle söz konusu sinyaller, yalnızca izleme amacıyla değil, aynı zamanda arıza örüntülerinin belirlenmesinde kullanılacak özelliklerin elde edilmesi için de temel veri kaynağı olarak değerlendirilmektedir.

Bu yapısal çerçeve, FV sistem dinamiklerinin ve arıza durumlarında ortaya çıkan sinyal değişimlerinin yorumlanabilmesi için gerekli altyapıyı oluşturmaktadır. Bu nedenle bir sonraki alt bölümde, FV sistem davranışını etkileyen temel dinamikler ele alınmaktadır.

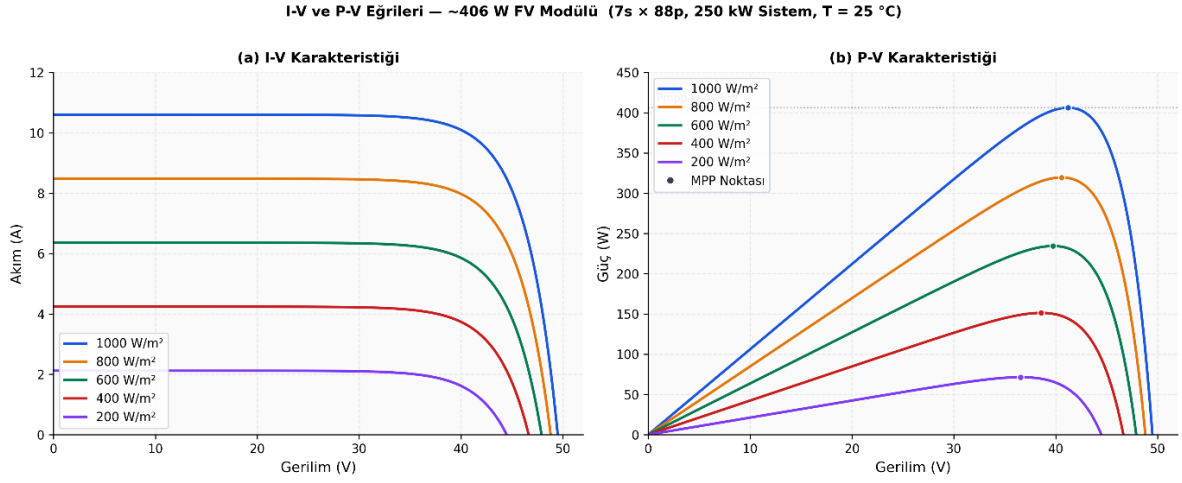
## **2.2. Fotovoltaik Sistem Dinamikleri**

FV sistemlerin dinamik davranışı, çevresel koşulların zamana bağlı değişimi ve güç elektroniği tabanlı denetim yapılarının bu değişimlere verdiği tepki nedeniyle çok değişkenli ve doğrusal olmayan bir karakter göstermektedir. Güneş ışıınımı, hücre sıcaklığı, yük durumu ve şebeke etkileşimi gibi dış etkenler, sistemin anlık akım, gerilim ve güç seviyelerini doğrudan etkilemektedir (Villalva vd., 2009; Madeti ve Singh, 2017). Bu nedenle FV sistem davranışı yalnızca statik çalışma eğrileri üzerinden değil, zamanla değişen dinamik bir süreç olarak değerlendirilmelidir.

Arıza teşhisi açısından ölçüm sinyallerindeki değişimlerin bir bölümü arıza etkisinden, bir bölümü ise normal işletme dinamiklerinden kaynaklanmaktadır. Bu iki etkinin birbirinden ayrılması, güvenilir sınıflandırma performansı açısından önemlidir. Özellikle ışıınım ve sıcaklık değişimlerinin sistem çıkışı üzerinde oluşturduğu doğal değişkenlik, bazı durumlarda arıza etkileriyle benzer sinyal davranışları oluşturabilmektedir. Bu durum, arıza teşhisinde yalnızca tekil ölçüm değerlerine değil, çok değişkenli ve örüntü temelli analiz yaklaşımlarına ihtiyaç duyulduğunu göstermektedir.

FV sistemlerde en belirgin dinamik etkilerden biri güneş ışıınımı ve hücre sıcaklığı değişimlerinden kaynaklanmaktadır. Güneş ışıınımındaki artış, hücre tarafından üretilen akımı artırarak toplam güç üretimini yükseltmektedir. Buna karşılık hücre sıcaklığındaki artış çoğu durumda açık devre gerilimini azaltmakta ve sistem verimini olumsuz etkilemektedir. Bu nedenle aynı sistem topolojisi, farklı çevresel koşullar altında farklı elektriksel davranışlar sergileyebilmektedir.

Işınım değişimlerinin FV çıkış karakteristiklerine etkisi Şekil 2.3'te gösterilmektedir.



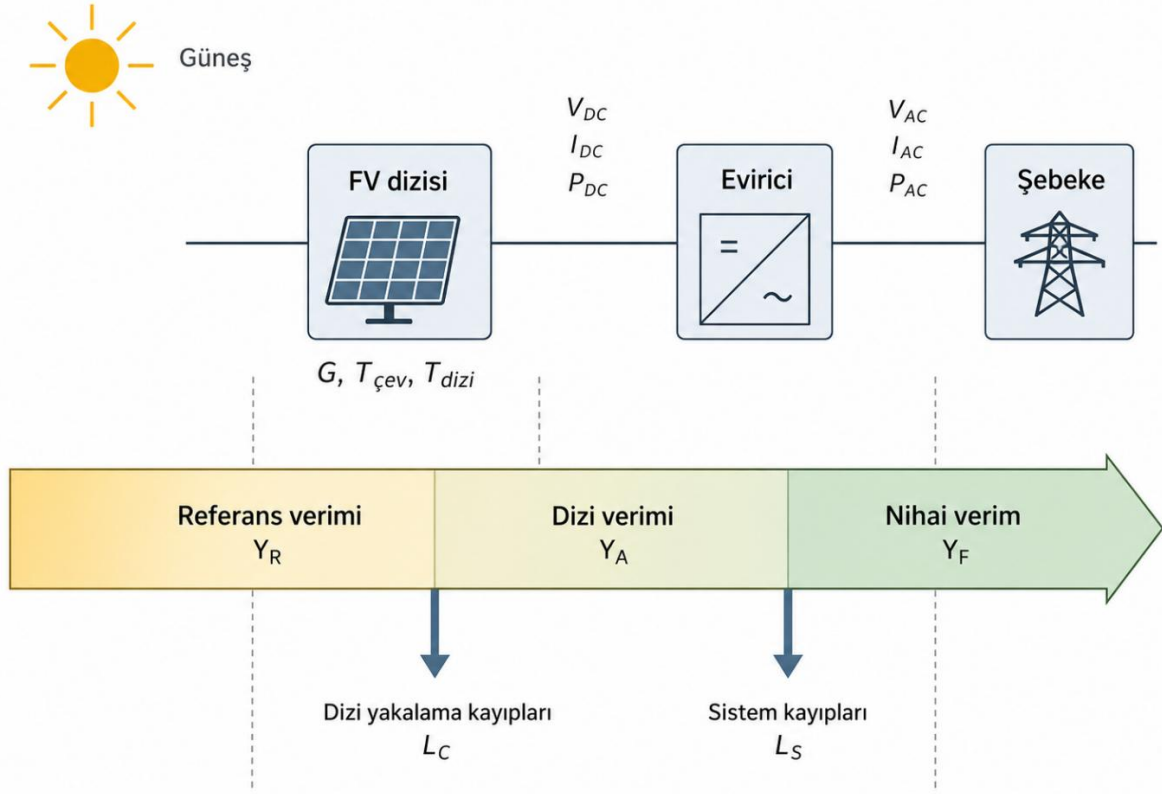
**Şekil 2.3.** Işınım değişimlerinin FV çıkış karakteristiklerine etkisi

Şekil 2.3 üzerinden, ışınımdaki artışın özellikle akım ve güç seviyelerinde yükselmeye neden olduğu görülmektedir. Bu durum, ölçülen sinyallerin yalnızca arıza bilgisini değil, çevresel işletme koşullarının etkisini de içerdiğini göstermektedir. Dolayısıyla arıza teşhisinde kullanılan özelliklerin, çevresel değişkenlerden kaynaklanan doğal varyasyonları dikkate alacak biçimde değerlendirilmesi gerekmektedir.

FV sistem dinamiklerinde maksimum güç noktası takibi de önemli bir rol üstlenmektedir. Maksimum güç noktası takibi algoritması, değişen çevresel koşullar altında FV dizinin mümkün olan en yüksek güçte çalışmasını sağlamayı amaçlamaktadır. Ancak ışınımın hızlı değiştiği veya bir arıza senaryosunun ortaya çıktığı durumlarda sistemin yeni çalışma noktasına geçişi geçici rejim davranışları oluşturabilmektedir. Bu süreçte akım, gerilim ve güç sinyallerinde salınımlar, gecikmeler veya sapmalar gözlenebilir. Arıza teşhisi açısından bu durum önemlidir; çünkü ölçülen bazı sinyal bozulmaları doğrudan arıza etkisini değil, denetim sisteminin yeni çalışma noktasına uyum sürecini yansıtabilir.

Benzer şekilde inverter ve DC-link dinamikleri de sistem davranışının belirlenmesinde önemli rol oynamaktadır. FV diziden gelen gücün şebekeye kararlı biçimde aktarılabilmesi için DC-link geriliminin belirli sınırlar içinde tutulması gerekmektedir. String arızası, topraklama arızası veya string-string arızası gibi durumlarda DC taraftaki akım dağılımı bozulmakta ve bu durum DC-link seviyesinde dalgalanmalara yol açabilmektedir. İnverter denetimi ise bu değişimlere karşı sistem çıkışını dengelemeye çalışmaktadır. Böylece ölçülen akım ve gerilim sinyalleri yalnızca kaynak tarafındaki arıza bilgisini değil, güç elektroniği katmanının dinamik tepkisini de içermektedir.

FV sistemde dinamik etki zinciri Şekil 2.4'te gösterilmektedir.



Şekil 2.4. FV sistemde dinamik etki zinciri

Şekil 2.4, sistemden elde edilen sinyallerin tek bir fiziksel kaynağa bağlı olmadığını; çevresel koşullar, denetim mekanizması ve güç dönüşüm zincirinin ortak etkisiyle oluştuğunu göstermektedir. Bu nedenle FV sistemlerden elde edilen ölçüm verileri yorumlanırken, sinyallerin hem fiziksel kaynak hem de dinamik sistem tepkisi açısından birlikte değerlendirilmesi gerekmektedir.

FV sistem dinamikleri, arıza teşhis problemi açısından iki yönlü öneme sahiptir. Birincisi, çevresel ve kontrol temelli değişimler ölçüm verilerinde doğal varyasyon oluşturarak sınıflar arası ayrımı zorlaştırmaktadır. İkincisi, bazı arıza türleri sistem dinamiğini yalnızca kararlı durumda değil, geçici rejimde de değiştirdiğinden ölçüm sinyalleri arıza hakkında ek bilgi taşımaktadır. Bu nedenle akım, gerilim, güç, sıcaklık ve ışıyım değişkenlerinin birlikte değerlendirilmesi, FV arıza sınıflandırması açısından anlamlıdır.

FV sistem dinamiklerini etkileyen temel değişkenler ve bu değişkenlerin ölçüm sinyallerine yansımaları Tablo 2.1'de özetlenmiştir.

**Tablo 2.1.** FV sistem dinamiklerini etkileyen temel deęişkenler ve sinyallere etkileri

<b>Dinamik etken</b>	<b>Etkilenen büyüklükler</b>	<b>Tipik etki</b>	<b>Arıza teşhisi açısından önemi</b>
Güneş ışınımı deęiřimi	Akım, güç	Genellikle artış/azalış	Normal işletme varyasyonu oluşturur
Hücre sıcaklığı deęiřimi	Gerilim, verim	Gerilim düşüşü, verim kaybı	Arıza ile karışabilecek performans azalması üretir
MPPT geçiş davranışı	Güç, gerilim, akım	Geçici salınım ve gecikme	Geçici rejim etkilerinin ayrıştırılmasını gerektirir
String arızası	String akımı, toplam güç	Dengesizlik ve güç kaybı	Sınıflandırmada kritik belirti oluşturur
Toprak arızası	Gerilim/akım dağılımı	Sapma ve dengesizlik	Güvenlik ve teşhis açısından önemlidir
İnverter/DC-link tepkisi	DC-link gerilimi, çıkış akımı	Dengeleyici veya bastırıcı tepki	Arıza etkisini ham sinyalde maskeleyebilir

Tablo 2.1’de görüldüğü üzere, güneş ışınımı ve hücre sıcaklığı gibi çevresel deęişkenler ile MPPT, inverter/DC-link tepkisi ve arıza türleri gibi sistem içi etkenler, ölçülen akım, gerilim ve güç sinyallerini farklı biçimlerde etkilemektedir. Bu durum, ölçüm sinyallerinin tek bir fiziksel nedene baęlı olmadığını; çevresel koşullar, denetim yapısı ve arıza türlerinin birlikte deęerlendirilmesi gerektiğini göstermektedir. Aynı zamanda Tablo 2.1, ilerleyen bölümlerde kullanılacak ölçüm deęişkenlerinin neden seçildiğini kuramsal olarak gerekçelendirmektedir.

FV sistemlerde kararlı durum ve geçici durum davranışlarının ayrıştırılması da arıza teşhisi açısından önem taşımaktadır. Kararlı durumda sinyal deęişimleri daha düzenli ve sınırlı iken, geçici durumlarda akım, gerilim ve güç sinyallerinde ani sapmalar veya salınımlar meydana gelebilmektedir. Ancak bu geçici davranışların tamamı arıza kaynaklı olmayabilir; bazıları çevresel deęişimlere veya denetim sisteminin yeni çalışma noktasına uyum sürecine baęlı olarak ortaya çıkabilir.

Kararlı durum ve geçici durum sinyal davranışlarının karşılaştırılması Tablo 2.2’de verilmiştir.

**Tablo 2.2.** Kararlı durum ve geçici durum sinyal davranışlarının karşılaştırılması

Durum	Akım davranışı	Gerilim davranışı	Güç davranışı	Teşhis zorluğu
Normal kararlı durum	Dengeli	Stabil	Stabil	Düşük
Çevresel değişim sonrası geçici durum	Değişken/salınımlı	Geçici sapmalı	MPPT arayışına bağlı değişken	Orta
Arıza sonrası geçici durum	Ani dengesizlik	Belirgin sapma	Hızlı güç kaybı	Yüksek
Arıza sonrası yeni kararlı durum	Kalıcı bozulma	Kaymış çalışma noktası	Sürekli düşük güç	Orta

Tablo 2.2, normal kararlı durum, çevresel değişim sonrası geçici durum, arıza sonrası geçici durum ve arıza sonrası yeni kararlı durum koşullarında akım, gerilim ve güç sinyallerinin farklı davranışlar sergileyebileceğini göstermektedir. Bu karşılaştırma, ölçülen sapmaların her zaman doğrudan arıza kaynaklı olmadığını; bazı değişimlerin sistemin doğal dinamik tepkisinden de kaynaklanabileceğini ortaya koymaktadır. Bu ayırım, veri odaklı sınıflandırma yaklaşımında kullanılan özelliklerin yorumlanması açısından önemlidir.

Sonuç olarak FV sistemlerin dinamik yapısı, arıza teşhis probleminin yalnızca anlık büyüklükler üzerinden değil, çevresel etkiler, kontrol tepkileri ve güç dönüşüm zinciriyle birlikte ele alınmasını gerekli kılmaktadır. Bu durum, arıza teşhisinde çok değişkenli veri analizi ve uygun özellik temsili kullanımını gerekli hale getirmektedir.

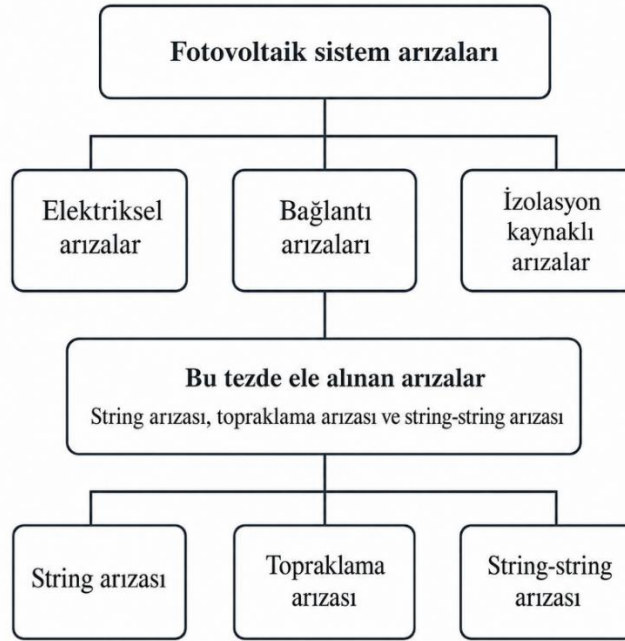
### 2.3. Fotovoltaik Sistemlerde Arıza Türleri

FV sistemlerde meydana gelen arızalar, enerji üretim performansını düşürmenin yanı sıra sistem güvenliği, ekipman ömrü ve işletme sürekliliği açısından da önemli riskler oluşturmaktadır. Bu arızalar; sistem topolojisine, elektriksel bağlantı yapısına, çevresel koşullara ve arızanın olduğu fiziksel noktaya bağlı olarak farklı biçimlerde ortaya çıkabilmektedir. Bu nedenle FV sistemlerde arıza teşhisi, tek bir bozulma tipine indirgenemeyecek kadar çok değişkenli bir problem olarak değerlendirilmelidir (Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

FV sistem arızalarının sınıflandırılmasında temel yaklaşım, her arıza türünün ölçüm sinyalleri üzerinde bıraktığı karakteristik etkilerin ayırt edilmesine dayanmaktadır. Bazı arızalar

akım dengesizliği oluştururken, bazıları gerilim seviyelerinde sapma veya güç üretiminde azalma şeklinde gözlenebilmektedir. Bu nedenle arızaların yalnızca fiziksel oluşum mekanizmalarıyla değil, ölçülebilir sinyal etkileri üzerinden de ele alınması gerekmektedir.

FV sistemlerde karşılaşılan temel arıza türlerinin genel sınıflandırması Şekil 2.5'te sunulmuştur.



**Şekil 2.5.** FV sistemlerde temel arıza türlerinin sınıflandırılması

Şekil 2.5'te görüldüğü üzere, FV sistem arızaları genel olarak elektriksel arızalar, bağlantı arızaları ve izolasyon kaynaklı arızalar şeklinde değerlendirilebilmektedir. Bu tez kapsamında ise normal çalışma durumu ile birlikte string arızası, string-toprak arızası ve string-string arızası ele alınmıştır. Bu arıza türleri, saha uygulamalarında karşılaşılabilmek olmaları ve ölçüm sinyallerine belirgin etkiler bırakmaları nedeniyle sınıflandırma açısından anlamlı bir problem yapısı sunmaktadır (Ghoneim vd., 2021; Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

String arızaları daha çok string bazlı akım dengesizliği ve toplam güç kaybı ile ilişkilidir. String-toprak arızaları gerilim dağılımı, kaçak akım oluşumu ve güvenlik riski açısından öne çıkarken, string-string arızaları paralel kollar arasındaki istenmeyen akım etkileşimi nedeniyle daha karmaşık bir akım paylaşım bozulmasına yol açabilmektedir (Mellit vd., 2018; Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

Tez kapsamında ele alınan temel arıza türlerinin oluşum nedenleri, etkiledikleri büyüklükler ve teşhis açısından ayırt edici yönleri Tablo 2.3'te özetlenmiştir.

**Tablo 2.3.** FV sistemlerde temel arıza türleri ve ölçüm sinyallerine etkileri

Arıza türü	Oluşum nedeni	Başlıca etkilenen büyüklükler	Tipik gözlenen etki	Teşhis açısından ayırt edici yön
String arızası	Açık devre, bağlantı bozulması, akım taşıma kapasitesi düşüşü	String akımları, toplam güç	Akım dengesizliği, güç kaybı	String bazlı akım farkları belirgindir
Topraklama arızası	İzolasyon bozulması, kablo hasarı, kaçak yol oluşumu	DC gerilim, kaçak akım, toplam güç	Gerilim sapması, güvenlik riski	Potansiyel dağılımı bozulur
String-string arızası	Paralel kollar arasında istenmeyen iletken yol	String akımları, toplam güç, akım paylaşımı	Karşılıklı akım etkileşimi, dengesizlik	İki kol arasındaki etkileşim belirgindir

Tablo 2.3, string arızası, string-toprak arızası ve string-string arızasının sistem sinyalleri üzerinde farklı etkiler oluşturduğunu göstermektedir. String arızasında string akımları ve toplam güç değişimleri öne çıkarken, string-toprak arızasında gerilim dağılımı ve kaçak akım etkileri daha belirgindir. String-string arızasında ise paralel kollar arasındaki istenmeyen akım yolu nedeniyle akım paylaşımı bozulmakta ve daha karmaşık bir ölçüm örüntüsü ortaya çıkmaktadır.

Arıza türlerinin enerji verimine etkisi, güvenlik riski, sinyal düzeyinde görünürlük ve sınıflandırma zorluğu bakımından karşılaştırılması Tablo 2.4'te verilmiştir.

**Tablo 2.4.** Arıza türlerinin güvenlik, verim ve teşhis zorluğu açısından karşılaştırılması

Arıza türü	Enerji verimine etkisi	Güvenlik riski	Sinyal düzeyinde görünürlük	Sınıflandırma zorluğu
String arızası	Orta-Yüksek	Orta	Yüksek	Orta
Topraklama arızası	Orta	Çok yüksek	Orta	Orta-Yüksek
String-string arızası	Yüksek	Yüksek	Orta	Yüksek

Tablo 2.4 incelendiğinde, arıza türlerinin yalnızca enerji üretimi üzerindeki etkileri bakımından değil, güvenlik riski ve teşhis zorluğu açısından da farklılaştığı görülmektedir. String arızaları çoğunlukla güç kaybı ve akım dengesizliği ile ilişkilendirilirken, string-toprak arızaları güvenlik riski bakımından daha kritik bir konumdadır. String-string arızaları ise paralel stringler arasındaki etkileşim nedeniyle daha karmaşık sinyal davranışları oluşturabilmekte ve sınıflandırma açısından daha zorlayıcı olabilmektedir.

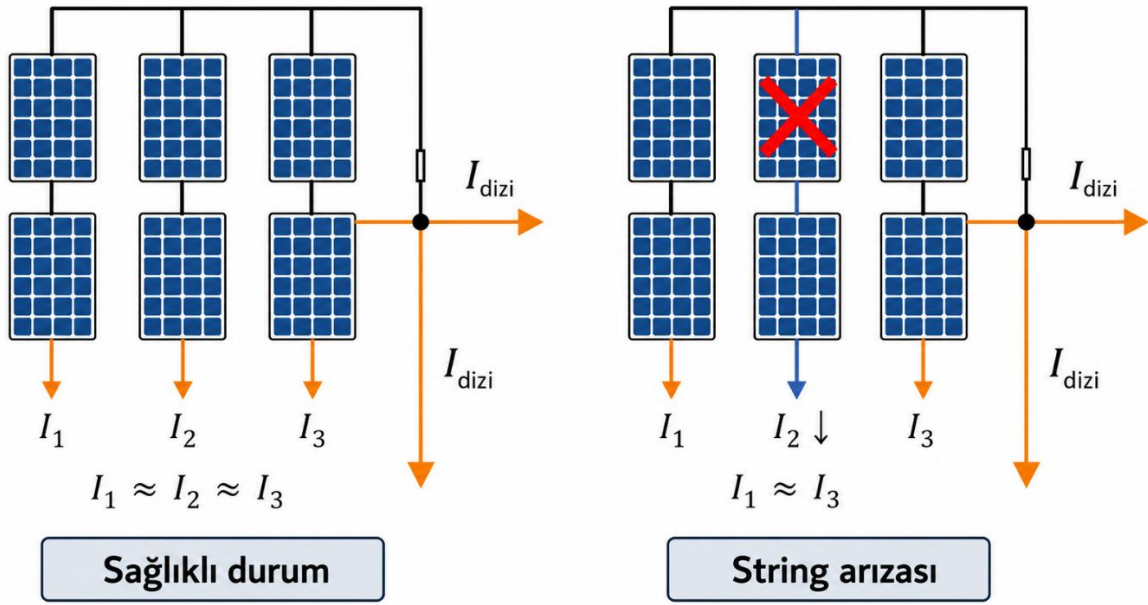
FV sistemlerde bu üç temel arıza türü, fiziksel oluşum mekanizmaları ve ölçüm sinyallerine yansımaları bakımından birbirinden ayrılmaktadır. Ancak çevresel değişimler, işletme koşulları ve kontrol dinamikleri arıza etkilerini kısmen maskeleyebilmektedir. Bu nedenle güvenilir bir arıza teşhis yaklaşımı, tek bir değişken üzerinden karar vermek yerine çoklu sinyal analizi ve veri odaklı sınıflandırma yapısına dayanmalıdır.

### **2.3.1. String arızaları**

String arızaları, FV sistemlerde seri bağlı modül grubunu oluşturan bir string üzerinde meydana gelen bağlantı bozulmaları, açık devre durumları, modül arızaları veya akım taşıma kapasitesindeki düşüşler sonucunda ortaya çıkabilmektedir. Bu arıza türünde ilgili stringin ürettiği akım azalmakta veya tamamen kesilebilmekte, buna bağlı olarak paralel bağlı diğer stringler ile arızalı string arasındaki akım dengesi bozulmaktadır. Bu durum, toplam dizi akımında ve sistemin DC güç üretiminde azalmaya neden olmaktadır (Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

String arızalarının en belirgin etkisi, string akımları arasındaki dengenin bozulmasıdır. Normal çalışma koşullarında benzer ışınım ve sıcaklık altında paralel stringlerden beklenen akım değerleri birbirine yakın olmalıdır. Ancak bağlantı kopması, kısmi devre dışı kalma veya yüksek dirençli temas problemi olduğunda ilgili stringin akımı diğer stringlere göre düşebilir. Bu nedenle string akımları arasındaki farklar, string arızalarının tespitinde önemli bir ayırt edici özellik olarak değerlendirilmektedir.

String arızasının FV dizi akım dağılımı üzerindeki etkisi Şekil 2.6'da gösterilmektedir.



**Şekil 2.6.** String arızasının FV dizi akım dağılımı üzerindeki etkisinin şematik gösterimi

Şekil 2.6’da normal çalışma koşullarında paralel stringler arasında dengeli bir akım paylaşımı bulunduğu, arıza durumunda ise arızalı stringin akım katkısının azaldığı görülmektedir. Bu azalma yalnızca arızalı kola ait akım değerini değil, sistemin toplam DC akımını ve üretilen toplam gücü de etkilemektedir. Dolayısıyla string arızaları, hem yerel string akımı hem de toplam güç değişimi üzerinden izlenebilir niteliktedir.

Bununla birlikte string arızalarının teşhisi her zaman doğrudan değildir. Düşük ışınım, kısmi gölgelenme, kirlenme veya modül yaşlanması gibi durumlar da string akımında düşüşe neden olabilir. Bu nedenle string arızalarının güvenilir biçimde sınıflandırılabilmesi için yalnızca tek bir akım değerinin değil; birden fazla string akımı, toplam akım, gerilim ve güç değişkenlerinin birlikte değerlendirilmesi gerekmektedir.

### 2.3.2. String-toprak arızaları

String-toprak arızaları, FV sistemin DC tarafında yer alan iletken bölümler ile toprak arasında istenmeyen bir elektriksel yol oluşması sonucunda meydana gelmektedir. Bu durum genellikle izolasyon bozulması, kablo hasarı, nem, bağlantı kutusu problemleri veya ekipman yalıtımındaki zayıflamalar nedeniyle ortaya çıkabilmektedir. String-toprak arızaları, enerji üretim kaybının yanı sıra kaçak akım ve elektriksel güvenlik riski oluşturabildiğinden kritik arıza türleri arasında değerlendirilmektedir (IEC, 2016; Triki-Lahiani vd., 2018).

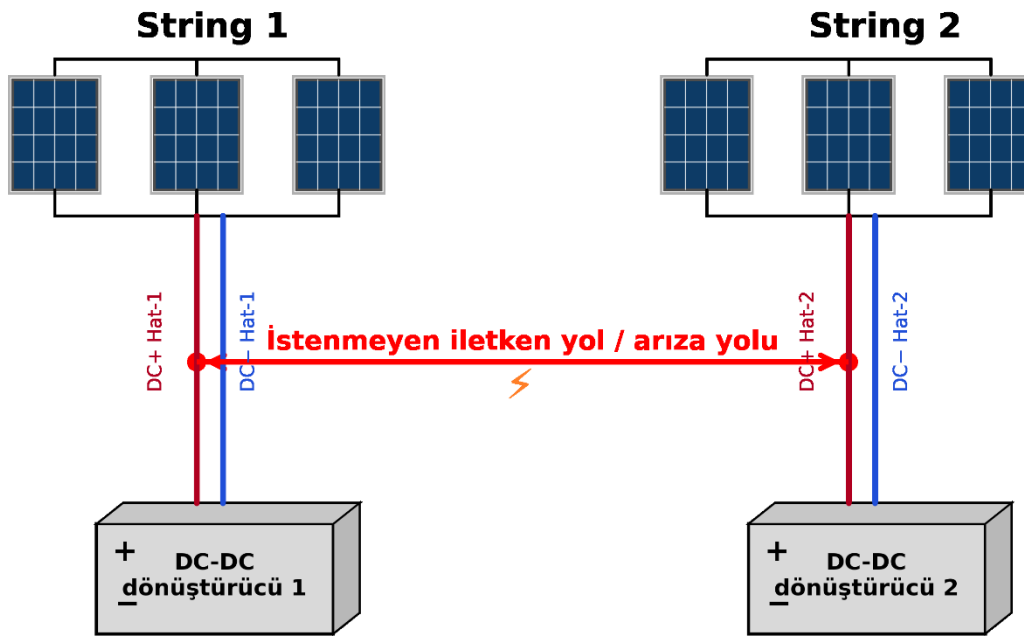
Bu arıza türünde, arızanın oluştuğu noktanın potansiyeline ve arıza direncine bağlı olarak sistemin gerilim dağılımı değişmektedir. Toprak ile oluşan istenmeyen iletken yol, DC



türünde farklı potansiyel seviyelerine sahip stringler arasında akım akışı oluşabilmekte ve sistemin normal akım paylaşımı bozulmaktadır. String-string arızaları, yalnızca tek bir stringin davranışını değil, birden fazla stringin akım dağılımını aynı anda etkileyebildiğinden karmaşık bir arıza yapısı oluşturmaktadır (Mellit vd., 2018; Hong ve Pula, 2022).

Normal çalışma koşullarında paralel stringlerin akım katkıları sistem topolojisi ve çevresel koşullar doğrultusunda dengeli biçimde dağılmaktadır. Ancak stringler arasında istenmeyen bir bağlantı oluştuğunda, daha yüksek potansiyele sahip koldan daha düşük potansiyele sahip kola doğru ek bir akım yolu meydana gelebilir. Bu durum bazı stringlerde akım artışı veya azalması, toplam dizi akımında bozulma ve güç üretiminde düşüş şeklinde gözlemlenebilir.

String-string arızasında paralel kollar arasındaki istenmeyen akım yolu Şekil 2.8’de gösterilmektedir.



*Not: String 1 ve String 2 normal çalışmada doğrudan paralel bağlı değildir; her string ayrı DC-DC dönüştürücü girişine bağlıdır.*

**Şekil 2.8.** String-string arızasında paralel kollar arasındaki istenmeyen akım yolu

Şekil 2.8’de, paralel stringler arasında oluşan istenmeyen iletken yolun akım paylaşımını nasıl bozduğu şematik olarak görülmektedir. Bu arıza türünde bozulma yalnızca bir stringin devre dışı kalması şeklinde değil, stringler arası etkileşimin değişmesi şeklinde ortaya çıkmaktadır. Bu nedenle string-string arızaları, string arızalarına kıyasla daha karmaşık sinyal davranışları üretebilmektedir.

String-string arızalarının teşhisinde temel güçlük, sinyal değişimlerinin farklı stringlerde eş zamanlı ve birbirine bağlı biçimde gözlenmesidir. Arıza direnci, arızanın oluştuğu string çifti, ışınım seviyesi ve sistemin çalışma noktası ölçülen sinyallerin karakterini doğrudan etkileyebilmektedir. Bu nedenle string-string arızalarının güvenilir biçimde sınıflandırılabilmesi için çok değişkenli özelliklerin birlikte değerlendirilmesi ve bu özelliklerin uygun boyut azaltma yöntemleriyle daha ayırt edici bir temsil uzayına taşınması önem taşımaktadır.

Sonuç olarak string, string-toprak ve string-string arızaları; oluşum mekanizmaları, etkiledikleri elektriksel büyüklükler ve sınıflandırma zorlukları bakımından farklı karakteristiklere sahiptir. Bu farklılıklar, FV arıza sınıflandırmasında kullanılacak özelliklerin yalnızca ham ölçüm değerlerinden değil, ölçüm sinyallerinin istatistiksel ve ilişkisel yapısından da türetilmesini gerekli kılmaktadır. Bu nedenle bir sonraki bölümde, ölçüm verilerinden anlamlı temsil elde edilmesini sağlayan boyut azaltma ve özellik çıkarma teorisi ele alınmaktadır.

#### **2.4. Boyut Azaltma ve Özellik Çıkarma Teorisi**

FV sistemlerde arıza teşhis performansı, yalnızca kullanılan sınıflandırma algoritmasına değil, sınıflandırıcıya sunulan veri temsilinin niteliğine de bağlıdır. Ham ölçüm verileri çoğu zaman yüksek boyutlu, gürültülü, korelasyonlu ve tekrarlı bilgi içeren bir yapıya sahiptir. Bu nedenle ham verilerin doğrudan kullanılması, model karmaşıklığını artırabilmekte ve sınıflar arası ayrımı zorlaştırabilmektedir. Bu noktada özellik çıkarma ve boyut azaltma yöntemleri, ölçüm verilerinden daha anlamlı, daha düşük boyutlu ve daha ayırt edici bir temsil elde etmek amacıyla kullanılmaktadır (Jolliffe, 2002; Bishop, 2006; Sepúlveda-Oviedo vd., 2023).

Özellik çıkarma, ham ölçüm verilerinden sınıflandırma açısından anlamlı değişkenlerin elde edilmesi sürecidir. FV sistemlerde bu özellikler genellikle akım, gerilim, güç, sıcaklık ve ışınım gibi ölçümlerden veya bu ölçümlerin istatistiksel türevlerinden elde edilmektedir. Örneğin bir string akımının ortalama değeri, maksimum ve minimum değeri, varyansı veya belirli bir zaman aralığındaki değişim aralığı, sistem durumunu temsil eden özellikler olarak kullanılabilir. Bu tür özellikler, arıza etkilerinin ham sinyal düzeyinden daha belirgin biçimde gözlenmesine katkı sağlamaktadır.

Boyut azaltma ise elde edilen özellik uzayının daha düşük boyutlu bir temsil uzayına dönüştürülmesi işlemidir. Bu dönüşümde amaç, gereksiz veya tekrarlı bilgiyi azaltırken sınıflandırma açısından önemli olan yapıyı mümkün olduğunca korumaktır. Boyut azaltma

yöntemleri sayesinde yüksek boyutlu veri uzayında dağınık ve örtüşen sınıf yapıları daha düzenli bir temsil uzayında incelenebilmekte; bu durum hem hesaplama yükünü azaltmakta hem de sınıflandırma algoritmalarının genelleme performansını iyileştirebilmektedir (Jolliffe, 2002; Bishop, 2006).

Özellik çıkarma ve boyut azaltma kavramları birbiriyle ilişkili olmakla birlikte aynı anlama gelmemektedir. Özellik çıkarma, ham veriden yeni ve anlamlı değişkenlerin üretilmesine odaklanırken; boyut azaltma, mevcut özellik uzayının daha kompakt ve ayırt edici bir temsil uzayına dönüştürülmesini amaçlamaktadır. Bu iki aşama birlikte kullanıldığında, FV arıza teşhisinde hem fiziksel olarak anlamlı hem de makine öğrenmesi açısından daha uygun bir veri temsili elde edilebilmektedir.

Özellik çıkarma ve boyut azaltma kavramlarının temel farkları Tablo 2.5'te karşılaştırmalı olarak verilmiştir.

**Tablo 2.5.** Özellik çıkarma ve boyut azaltma kavramlarının karşılaştırılması

Kavram	Temel amaç	Girdi yapısı	Çıktı yapısı	FV arıza teşhisindeki rolü
Özellik çıkarma	Ham veriden anlamlı göstergeler üretmek	Ölçüm sinyalleri	İstatistiksel/fiziksel özellikler	Arıza belirtilerini görünür kılar
Boyut azaltma	Özellik uzayını daha kompakt hale getirmek	Çıkarılmış özellikler	Düşük boyutlu temsil	Sınıf ayrımını güçlendirir, karmaşıklığı azaltır

Tablo 2.5 incelendiğinde, özellik çıkarma aşamasının ham ölçüm sinyallerinden anlamlı istatistiksel veya fiziksel göstergeler üretmeye odaklandığı görülmektedir. Buna karşılık boyut azaltma, çıkarılmış özelliklerden oluşan yüksek boyutlu veri uzayını daha düşük boyutlu ve daha kompakt bir temsil uzayına dönüştürmektedir. Bu nedenle özellik çıkarma arıza belirtilerini görünür hale getirirken, boyut azaltma bu belirtilerin sınıflandırma algoritmaları tarafından daha etkin kullanılmasını sağlamaktadır.

FV arıza teşhisinde iyi bir özellik temsili, arıza sınıflarını birbirinden ayırabilecek bilgiyi taşımali ve çevresel değişimlerden kaynaklanan doğal varyasyonlara karşı mümkün olduğunca kararlı olmalıdır. Çünkü ışınım ve sıcaklık değişimleri, normal çalışma koşullarında dahi akım, gerilim ve güç sinyallerinde önemli değişiklikler oluşturabilmektedir. Bu nedenle

seçilen veya dönüştürülen özelliklerin yalnızca arıza etkilerine duyarlı olması değil, normal işletme değişkenlikleri karşısında da ayırt edici niteliğini koruması beklenmektedir.

İyi bir özellik temsilinin taşınması gereken temel nitelikler Tablo 2.6’da özetlenmiştir.

**Tablo 2.6.** İyi bir özellik temsilinin taşınması gereken temel nitelikler

<b>Nitelik</b>	<b>Açıklama</b>	<b>Arıza sınıflandırmasına etkisi</b>
Ayırt edicilik	Farklı sınıfların temsil uzayında ayrışması	Sınıflandırma doğruluğunu artırır
Gürültüye dayanıklılık	Ölçüm sapmalarına rağmen örüntünün korunması	Kararlılığı yükseltir
Düşük korelasyon	Tekrarlı bilginin azaltılması	Aşırı uyumu ve karmaşıklığı azaltır
Düşük boyut	Kompakt temsil	Hesaplama maliyetini düşürür
Yorumlanabilirlik	Temsilin fiziksel/istatistiksel anlam taşınması	Mühendislik yorumunu kolaylaştırır

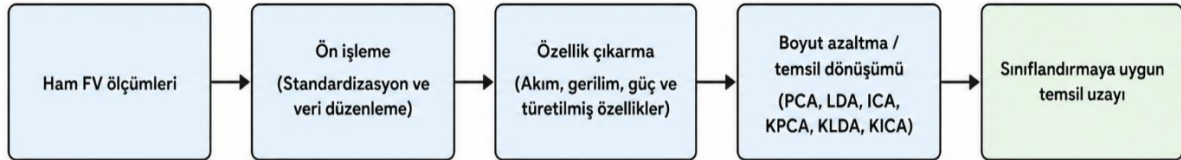
Tablo 2.6’da görüldüğü üzere, başarılı bir özellik temsilinin ayırt edicilik, gürültüye dayanıklılık, düşük korelasyon, düşük boyut ve yorumlanabilirlik gibi niteliklere sahip olması beklenmektedir. Ayırt edicilik, farklı arıza sınıflarının birbirinden ayrılmasını kolaylaştırırken; gürültüye dayanıklılık, ölçüm sapmaları veya çevresel değişimlere rağmen temsilin kararlı kalmasına katkı sağlamaktadır. Düşük korelasyon ve düşük boyut, modelin gereksiz bilgiyle beslenmesini önleyerek hesaplama maliyetini azaltmakta; yorumlanabilirlik ise elde edilen sonuçların mühendislik açısından açıklanabilir olmasını desteklemektedir.

Boyut azaltma yöntemleri, verideki yapıyı temsil etme biçimlerine göre genel olarak doğrusal ve doğrusal olmayan yöntemler şeklinde iki ana grupta incelenmektedir. Doğrusal yöntemler, ham veri uzayındaki ilişkilerin doğrusal dönüşümlerle yeterli ölçüde temsil edilebileceği varsayımına dayanır. Bu grupta Temel Bileşen Analizi, Doğrusal Ayırma Analizi ve Bağımsız Bileşen Analizi yer almaktadır. Temel Bileşen Analizi verideki maksimum varyans doğrultularını belirlemeye, Doğrusal Ayırma Analizi sınıflar arası ayrımı artırmaya, Bağımsız Bileşen Analizi ise veriyi istatistiksel olarak bağımsız bileşenlere ayırmaya odaklanmaktadır (Fisher, 1936; Jolliffe, 2002; Hyvärinen vd., 2001).

Doğrusal olmayan yöntemler ise ham veri uzayında doğrusal olarak ayrıştırılamayan karmaşık örüntülerin daha yüksek boyutlu veya örtük özellik uzaylarında daha belirgin hale

getirilebileceği varsayımına dayanır. Çekirdek tabanlı yöntemlerde veri, açıkça hesaplanması gerekmeyen bir özellik uzayına taşınmakta ve bu uzaydaki iç çarpımlar çekirdek fonksiyonları yardımıyla ifade edilmektedir. Böylece doğrusal olmayan sınıf sınırları daha ayırt edici biçimde temsil edilebilmektedir (Schölkopf vd., 1998; Baudat ve Anouar, 2000; Bach ve Jordan, 2002).

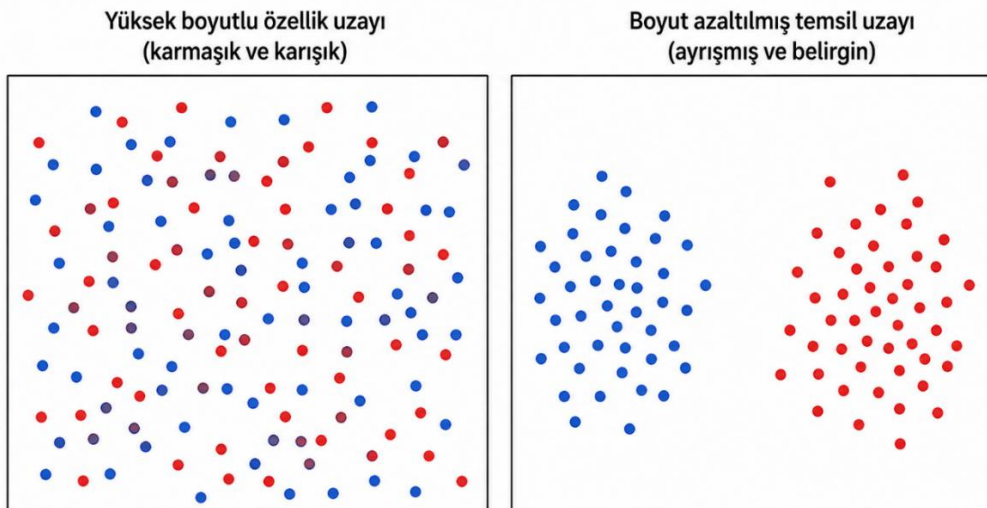
Ham ölçüm verisinden boyut azaltılmış özellik uzayına geçiş süreci Şekil 2.9’da gösterilmektedir.



**Şekil 2.9.** Ham ölçüm verisinden boyut azaltılmış özellik uzayına geçiş süreci

Şekil 2.9’da görüldüğü üzere, FV sistemden elde edilen ölçüm verileri öncelikle özellik çıkarma ve veri ön işleme adımlarından geçirilmekte, ardından boyut azaltma yöntemleri kullanılarak daha düşük boyutlu bir temsil uzayına dönüştürülmektedir. Bu yeni temsil uzayı, sınıflandırıcıya giriş olarak verilmekte ve sistem durumunun normal çalışma veya arıza sınıflarından hangisine ait olduğu belirlenmektedir. Bu akış, veri odaklı arıza teşhisinde ham ölçümden sınıflandırma kararına kadar uzanan temel işlem zincirini ortaya koymaktadır.

Yüksek boyutlu özellik uzayı ile boyut azaltılmış temsil uzayı arasındaki kavramsal fark Şekil 2.10’da verilmiştir.



**Şekil 2.10.** Yüksek boyutlu özellik uzayı ile boyut azaltılmış temsil uzayının kavramsal karşılaştırması

Şekil 2.10, yüksek boyutlu özellik uzayında sınıflar arası ayrımın karmaşık ve örtüşmeli olabileceğini, boyut azaltma sonrasında ise bu yapının daha kompakt ve yorumlanabilir bir temsil uzayına taşınabileceğini göstermektedir. Bu durum, özellikle FV arıza sınıflandırmasında önemlidir; çünkü farklı arıza türleri benzer akım, gerilim veya güç değişimleri oluşturabilmekte ve bu benzerlikler ham veri uzayında sınıfların birbirine yaklaşmasına neden olabilmektedir.

Sonuç olarak özellik çıkarma ve boyut azaltma, FV arıza sınıflandırma sürecinin temel aşamalarındandır. Özellik çıkarma, ölçüm sinyallerinden sistem durumunu temsil eden anlamlı göstergelerin elde edilmesini sağlarken; boyut azaltma, bu göstergelerin daha düşük boyutlu, daha kararlı ve daha ayırt edici bir uzayda ifade edilmesine olanak tanımaktadır. Bu nedenle bir sonraki bölümde, tez kapsamında kullanılan doğrusal ve çekirdek tabanlı boyut azaltma teknikleri ele alınmaktadır.

## **2.5. Doğrusal ve Çekirdek Tabanlı Boyut Azaltma Teknikleri**

Boyut azaltma yöntemleri, yüksek boyutlu veri yapılarında yer alan bilgiyi daha düşük boyutlu ve daha anlamlı bir temsil uzayında ifade etmeyi amaçlamaktadır. FV sistemlerden elde edilen ölçüm verileri çoğu zaman yüksek korelasyon, tekrarlı bilgi, gürültü ve sınıflar arası örtüşme içerebildiğinden, bu verilerin doğrudan sınıflandırıcıya verilmesi her zaman en uygun yaklaşım değildir. Bu nedenle boyut azaltma, FV arıza teşhisinde yalnızca hesaplama yükünü azaltan bir işlem değil, sınıflandırma performansını etkileyen temel bir temsil iyileştirme aşaması olarak değerlendirilmelidir (Jolliffe, 2002; Bishop, 2006).

Boyut azaltma teknikleri genel olarak doğrusal ve doğrusal olmayan yöntemler şeklinde iki ana grupta incelenmektedir. Doğrusal yöntemler, verideki varyans yapısını, sınıf ayrımını veya istatistiksel bağımsız bileşenleri doğrudan gözlenen özellik uzayı içinde ortaya çıkarmayı hedeflemektedir. Buna karşılık çekirdek tabanlı yöntemler, doğrusal uzayda yeterince ayrıştırılamayan veri yapılarını daha yüksek boyutlu veya örtük bir özellik uzayında temsil etmeye olanak sağlamaktadır (Schölkopf vd., 1998; Baudat ve Anouar, 2000; Bach ve Jordan, 2002).

Bu tez kapsamında doğrusal yöntemler olarak Temel Bileşen Analizi, Doğrusal Ayırma Analizi ve Bağımsız Bileşen Analizi; çekirdek tabanlı yöntemler olarak ise Çekirdek Temel Bileşen Analizi, Çekirdek Doğrusal Ayırma Analizi ve KPCA-ICA yaklaşımı değerlendirilmiştir. KPCA-ICA yaklaşımı, bu tez kapsamında KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde tanımlanmıştır. Bu yöntemlerin her

biri FV ölçüm verilerini farklı bir temsil mantığıyla ele almakta ve sınıflandırma performansını farklı yönlerden etkileyebilmektedir.

Doğrusal boyut azaltma tekniklerinin temel özellikleri Tablo 2.7’de karşılaştırmalı olarak sunulmuştur.

**Tablo 2.7.** Doğrusal boyut azaltma tekniklerinin karşılaştırılması

<b>Yöntem</b>	<b>Temel amaç</b>	<b>Denetimli/ Denetimsiz</b>	<b>Ana matematiksel fikir</b>	<b>FV verisi açısından güçlü yön</b>
PCA	Varyans koruma	Denetimsiz	Özdeğer-özvektör ayrışımı	Korelasyonlu özellikleri sıkıştırır
LDA	Sınıf ayrımını maksimize etmek	Denetimli	Saçılım oranı optimizasyonu	Arıza sınıflarını ayırmada etkilidir
ICA	Bağımsız bileşenleri bulmak	Denetimsiz	İstatistiksel bağımsızlık	Gizli örüntüleri yakalayabilir

Tablo 2.7’de görüldüğü üzere PCA, LDA ve ICA farklı temsil amaçlarına sahiptir. PCA verideki varyans bilgisini korumaya odaklanırken, LDA sınıflar arası ayrımı artırmayı hedeflemektedir. ICA ise verideki istatistiksel olarak bağımsız bileşenleri ortaya çıkarmaya çalışmaktadır. Bu nedenle doğrusal yöntemler aynı veri seti üzerinde farklı temsil uzayları oluşturabilmektedir.

Çekirdek tabanlı boyut azaltma tekniklerinin temel özellikleri Tablo 2.8’de verilmiştir.

**Tablo 2.8.** Çekirdek tabanlı boyut azaltma tekniklerinin karşılaştırılması

Yöntem	Doğrusal karşılığı	Temel avantaj	Temel sınırlılık	FV verisi açısından kullanım amacı
KPCA	PCA	Doğrusal olmayan varyansı yakalar	Çekirdek ve parametre seçimine duyarlı	Karmaşık veri yapısını sıkıştırmak
KLDA	LDA	Doğrusal olmayan sınıf ayrımı sağlar	Aşırı uyum riski oluşabilir	Arıza sınıflarını daha net ayırmak
KPCA-ICA	ICA	Doğrusal olmayan gizli bileşenleri arar	Hesaplama ve kararlılık zorluğu	Gizli örüntüleri açığa çıkarmak

Tablo 2.8 incelendiğinde, KPCA, KLDA ve KPCA-ICA yaklaşımlarının doğrusal yöntemlerin temel mantığını korumakla birlikte veriyi çekirdek fonksiyonları aracılığıyla tanımlanan örtük bir özellik uzayında işlediği görülmektedir. Bu sayede doğrusal olmayan sınıf yapılarının daha belirgin hale getirilmesi mümkün olabilmektedir. Bununla birlikte çekirdek tabanlı yöntemler, çekirdek fonksiyonu ve parametre seçimine duyarlı olduklarından uygulama sürecinde dikkatli biçimde değerlendirilmelidir.

Tez kapsamında kullanılan boyut azaltma tekniklerinin FV arıza sınıflandırması açısından genel değerlendirmesi Tablo 2.9’da özetlenmiştir.

**Tablo 2.9.** Tez kapsamında kullanılan boyut azaltma tekniklerinin genel deęerlendirmesi

Yöntem	Temsil ettięi yapı	Beklenen avantaj	Tez özetindeki genel gözlem
PCA	Küresel varyans	Kararlı ve kompakt temsil	En yüksek genel kararlılık
LDA	Sınıf ayrımı	Denetimli ayrışma	Güçlü ayırım potansiyeli
ICA	Bağımsız bileşen	Gizli örüntü çıkarımı	Sınırlı ayırt edicilik
KPCA	Doğrusal olmayan varyans	Karmaşık yapıları temsil	Orta düzey esneklik
KLDA	Doğrusal olmayan sınıf ayrımı	Güçlü denetimli ayırım	Öne çıkan yöntemlerden biri
KPCA-ICA	Doğrusal olmayan bağımsız yapı	Gizli karmaşık örüntüler	Kararlılık sınırlı olabilir

Tablo 2.9, tez kapsamında kullanılan yöntemlerin farklı veri temsil mantıklarına sahip olduğunu göstermektedir. PCA küresel varyans yapısını, LDA sınıf ayrımını, ICA bağımsız bileşenleri, KPCA doğrusal olmayan varyans yapısını, KLDA doğrusal olmayan sınıf ayrımını ve KPCA-ICA doğrusal olmayan temsil uzayındaki bağımsız bileşenleri öne çıkarmaktadır. Bu farklılıklar, yöntemlerin aynı FV arıza veri seti üzerinde farklı başarı düzeyleri göstermesinin temel nedenlerinden biridir.

Sonuç olarak doğrusal ve çekirdek tabanlı boyut azaltma yöntemleri, FV arıza sınıflandırmasında farklı temsil olanakları sunmaktadır. Bu yöntemlerin karşılaştırmalı biçimde incelenmesi, yalnızca hangi yöntemin daha yüksek sınıflandırma başarısı sağladığını belirlemek açısından değil, aynı zamanda hangi veri temsil yaklaşımının hangi koşullarda daha uygun olduğunu anlamak açısından da önemlidir. Bu nedenle izleyen alt bölümlerde tez kapsamında kullanılan boyut azaltma yöntemleri ayrı ayrı ele alınmaktadır.

### 2.5.1. Temel bileşen analizi (PCA)

Temel Bileşen Analizi, yüksek boyutlu ve birbiriyle ilişkili deęişkenlerden oluşan veri yapısını, daha düşük boyutlu ve birbirine dik bileşenler aracılığıyla temsil etmeyi amaçlayan doğrusal bir boyut azaltma yöntemidir. Yöntemin temel amacı, verideki toplam varyansın mümkün olduğunca büyük bir bölümünü koruyarak daha az sayıda bileşenle temsil sağlamaktır.

Bu yönüyle PCA, özellikle korelasyonlu özellikler içeren veri setlerinde tekrarlı bilgiyi azaltmak ve daha kompakt bir temsil elde etmek için yaygın biçimde kullanılmaktadır (Jolliffe, 2002).

FV sistemlerden elde edilen ölçüm verileri çoğu zaman akım, gerilim, güç ve bunlardan türetilen istatistiksel değişkenlerden oluşmaktadır. Bu değişkenlerin bir kısmı benzer fiziksel etkileri temsil ettiği için aralarında yüksek korelasyon bulunabilmektedir. PCA, bu korelasyonlu özellikleri daha az sayıda bileşen altında toplayarak veri uzayını sadeleştirmektedir.

PCA'da ilk adım, veri setinin uygun biçimde merkezlenmesi ve gerekiyorsa ölçeklendirilmesidir. Daha sonra veri setinin kovaryans matrisi elde edilir ve bu matrisin özdeğer-özvektör ayrışımı gerçekleştirilir. Özdeğerler ilgili temel bileşenin açıkladığı varyans miktarını, özvektörler ise maksimum varyans doğrultularını göstermektedir. Bu yapı genel olarak aşağıdaki biçimde ifade edilebilir:

$$\Sigma \mathbf{w}_i = \lambda_i \mathbf{w}_i \quad (\text{D.6})$$

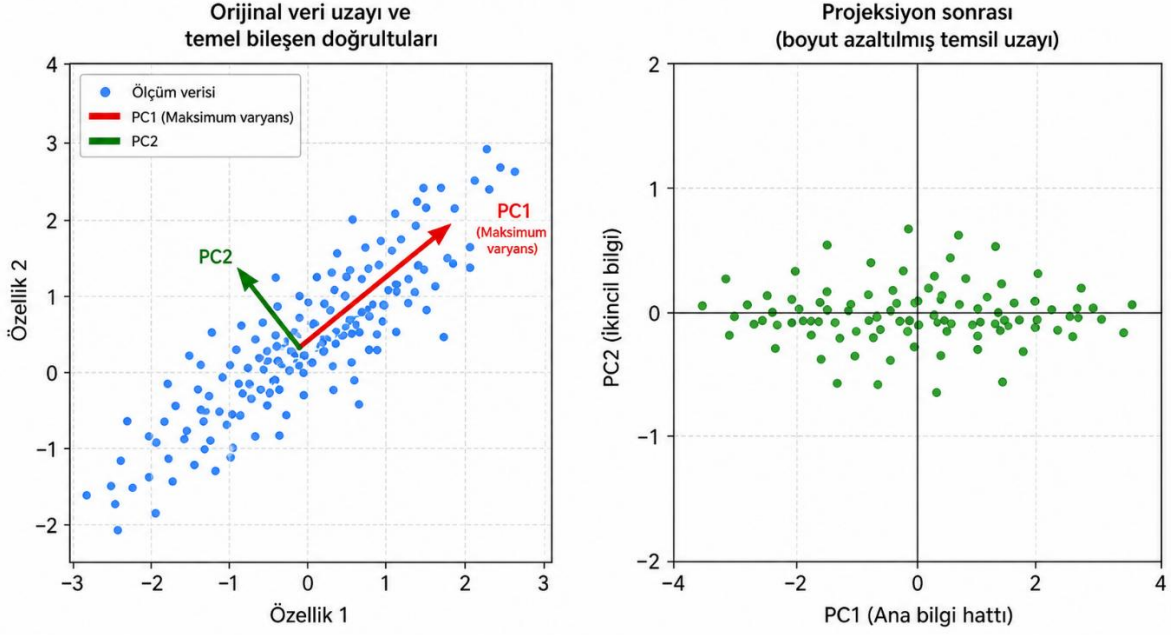
Burada  $\Sigma$  kovaryans matrisini,  $\mathbf{w}_i$  i'inci özvektörü,  $\lambda_i$  ise bu özvektöre karşılık gelen özdeğeri göstermektedir. Büyük özdeğerlere sahip bileşenler, verideki varyansın daha büyük kısmını açıkladığından boyut azaltma sürecinde öncelikli olarak seçilmektedir.

Seçilen temel bileşenler kullanılarak ham özellik vektörü daha düşük boyutlu bir uzaya aktarılır. Bu dönüşüm genel olarak şu şekilde ifade edilmektedir:

$$\mathbf{z} = \mathbf{W}^T \mathbf{x} \quad (\text{D.7})$$

Burada  $\mathbf{x}$  ham özellik vektörünü,  $\mathbf{W}$  seçilen temel bileşenlerden oluşan dönüşüm matrisini,  $\mathbf{z}$  ise boyut azaltılmış temsil vektörünü ifade etmektedir. Böylece yüksek boyutlu veri uzayı, daha az sayıda bileşenle temsil edilen yeni bir özellik uzayına dönüştürülmektedir.

PCA'da maksimum varyans doğrultularına projeksiyonun kavramsal gösterimi Şekil 2.11'de verilmiştir.



**Şekil 2.11.** PCA’da maksimum varyans doğrultularına projeksiyonun şematik gösterimi

Şekil 2.11’de görüldüğü üzere PCA, verideki en yüksek varyans doğrultularını belirlemekte ve gözlemleri bu doğrultular üzerine projekte etmektedir. Bu işlem sonucunda, verideki temel değişim yapısı korunurken boyut sayısı azaltılmaktadır. FV arıza sınıflandırması açısından bu durum, yüksek boyutlu ve korelasyonlu ölçüm özelliklerinin daha kompakt bir temsil uzayına taşınmasını sağlamaktadır.

PCA’nın önemli avantajlarından biri, hesaplama açısından sade ve yorumlanabilir bir yapı sunmasıdır. Özellikle mühendislik uygulamalarında PCA ile elde edilen bileşenlerin açıkladığı varyans oranları incelenerek, veri setindeki baskın değişim yönleri hakkında bilgi edinilebilir. Bununla birlikte PCA denetimsiz bir yöntem olduğundan, sınıf etiketlerini doğrudan dikkate almaz. Bu nedenle PCA bileşenlerinin her durumda en iyi sınıf ayrımını sağlaması garanti değildir.

Sonuç olarak PCA, FV arıza sınıflandırmasında yüksek boyutlu ölçüm özelliklerini daha düşük boyutlu ve yorumlanabilir bir temsil uzayına dönüştürmek için kullanılacak etkili bir doğrusal boyut azaltma yöntemidir. Bu tez kapsamında PCA, diğer doğrusal ve çekirdek tabanlı yöntemlerle aynı değerlendirme altyapısı altında karşılaştırılmıştır.

### 2.5.2. Doğrusal ayırma analizi (LDA)

Doğrusal Ayırma Analizi, sınıf etiketlerini dikkate alarak veri örneklerini daha düşük boyutlu bir uzaya dönüştürmeyi amaçlayan denetimli bir boyut azaltma yöntemidir. PCA

toplam varyansı korumaya odaklanırken, LDA sınıflar arası ayrımı artırmayı ve sınıf içi dağılımı azaltmayı hedeflemektedir. Bu yönüyle LDA, etiketli veri setlerinde ayırt edici bir temsil uzayı oluşturmak amacıyla kullanılmaktadır (Fisher, 1936; Bishop, 2006).

LDA'nın temel yaklaşımı, aynı sınıfa ait örneklerin dönüştürülmüş uzayda birbirine yakın, farklı sınıflara ait örneklerin ise birbirinden uzak konumlanmasını sağlamaktır. Bu amaçla yöntem, sınıf içi saçılım ve sınıflar arası saçılım kavramlarına dayanmaktadır.  $S_B$  sınıflar arası saçılımı,  $S_W$  ise sınıf içi saçılımı ifade etmektedir. LDA'da amaç, sınıflar arası saçılımı artırırken sınıf içi saçılımı azaltan bir dönüşüm matrisi elde etmektir. Bu yapı genel olarak aşağıdaki optimizasyon oranı ile ifade edilebilir:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (\text{D.8})$$

Burada  $\mathbf{W}$ , veriyi düşük boyutlu temsil uzayına dönüştüren dönüşüm matrisini ifade etmektedir. Bu oranın maksimize edilmesi, dönüştürülmüş uzayda sınıflar arası ayrımın artırılması ve aynı sınıfa ait örneklerin daha kompakt hale getirilmesi anlamına gelmektedir.

FV arıza sınıflandırması açısından LDA'nın önemli avantajı, normal çalışma ve farklı arıza sınıflarına ait etiket bilgisini doğrudan kullanmasıdır. Bu sayede LDA, ham ölçüm uzayında örtüşme gösterebilen arıza örüntülerini daha ayırt edici bir doğrusal temsil uzayına taşıyabilir.

Bununla birlikte LDA doğrusal bir yöntemdir ve sınıflar arasındaki ayrımın doğrusal dönüşümlerle yeterli ölçüde temsil edilebileceği varsayımına dayanır. Eğer veri yapısı belirgin doğrusal olmayan ilişkiler içeriyorsa, LDA'nın temsil kapasitesi sınırlı kalabilir. Ayrıca LDA'da elde edilebilecek en fazla bileşen sayısı sınıf sayısının bir eksiği ile sınırlıdır. Bu tez kapsamında dört sınıf ele alındığından, LDA ile elde edilebilecek en fazla ayırt edici bileşen sayısı üçtür.

Sonuç olarak LDA, sınıf etiketlerini dikkate alan denetimli yapısı sayesinde FV arıza sınıflandırmasında sınıflar arası ayrımı güçlendirebilecek önemli bir doğrusal boyut azaltma yöntemidir. Ancak doğrusal yapısı nedeniyle performansı veri setindeki sınıf dağılımına ve arıza örüntülerinin doğrusal ayrılabilirliğine bağlıdır.

### 2.5.3. Bağımsız bileşen analizi (ICA)

Bağımsız Bileşen Analizi, çok değişkenli gözlem verilerini istatistiksel olarak mümkün olduğunca bağımsız bileşenlere ayırmayı amaçlayan doğrusal bir özellik çıkarma ve boyut azaltma yöntemidir. PCA verideki maksimum varyans doğrultularını belirlemeye odaklanırken,

ICA verinin altında yatan gizli kaynakları veya bağımsız bileşenleri ortaya çıkarmaya çalışmaktadır (Hyvärinen vd., 2001; Bishop, 2006).

ICA'nın temel varsayımı, gözlenen çok değişkenli verinin istatistiksel olarak bağımsız bazı gizli kaynakların doğrusal karışımı şeklinde oluştuğudur. Bu yapı genel olarak aşağıdaki biçimde ifade edilebilir:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (\text{D.9})$$

Burada  $\mathbf{x}$  gözlenen veri vektörünü,  $\mathbf{s}$  bağımsız kaynak bileşenlerini ve  $\mathbf{A}$  karışım matrisini göstermektedir. ICA'nın amacı, yalnızca gözlenen  $\mathbf{x}$  verisinden hareketle bağımsız bileşenleri yaklaşık olarak elde edebilecek bir ayrıştırma matrisi bulmaktır. Bu durumda bağımsız bileşen tahmini aşağıdaki şekilde yazılabilir:

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \quad (\text{D.10})$$

Burada  $\mathbf{W}$  ayrıştırma matrisini,  $\hat{\mathbf{s}}$  ise tahmin edilen bağımsız bileşenleri ifade etmektedir.

FV sistemlerden elde edilen ölçüm verileri, arıza etkileri, çevresel koşullar ve güç elektroniği tepkilerinin birlikte gözlemlendiği çok değişkenli bir yapıya sahiptir. ICA, bu karmaşık ölçüm yapısı içerisinde doğrudan gözlenemeyen gizli bileşenleri ortaya çıkarma potansiyeline sahip olduğundan, FV arıza sınıflandırması açısından anlamlı bir temsil yöntemi olarak değerlendirilebilir.

Bununla birlikte ICA denetimsiz bir yöntemdir ve sınıf etiketlerini doğrudan dikkate almaz. Bu nedenle elde edilen bağımsız bileşenlerin arıza sınıflarını en iyi şekilde ayıracağı garanti değildir. Ayrıca ICA sonuçları, veri ölçeklendirme biçimine, bileşen sayısına, kullanılan algoritmanın yakınsama özelliklerine ve veri setindeki gürültü düzeyine duyarlı olabilir.

Sonuç olarak ICA, FV sistemlerden elde edilen çok değişkenli ölçüm verilerindeki gizli istatistiksel bileşenleri ortaya çıkarmak için kullanılacak önemli bir doğrusal temsil yöntemidir. Ancak sınıf etiketlerini dikkate almaması ve kararlılığının veri özelliklerine bağlı olması nedeniyle, ICA'nın arıza sınıflandırmasındaki etkinliği deneysel olarak değerlendirilmelidir.

#### **2.5.4. Çekirdek tabanlı boyut azaltma yaklaşımı**

Çekirdek tabanlı boyut azaltma yaklaşımı, doğrusal yöntemlerin sınırlı kaldığı veri yapılarında doğrusal olmayan ilişkileri daha etkili biçimde temsil etmek amacıyla kullanılmaktadır. PCA, LDA ve ICA gibi doğrusal yöntemler gözlenen özellik uzayındaki doğrusal dönüşümlere dayanırken, FV sistemlerden elde edilen ölçüm verileri çoğu zaman

çevresel koşullar, güç elektroniği tepkileri ve arıza etkilerinin birlikte oluşturduğu doğrusal olmayan ilişkiler içerebilmektedir.

Çekirdek yaklaşımının temel fikri, veriyi doğrudan gözlenen özellik uzayında işlemek yerine, doğrusal olmayan bir eşleme yardımıyla daha yüksek boyutlu veya örtük bir özellik uzayına taşımaktır. Bu eşleme genel olarak aşağıdaki biçimde ifade edilebilir:

$$\boldsymbol{\varphi}: \mathbb{R}^d \rightarrow \mathbf{F}, \mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x}) \quad (\mathbf{D.11})$$

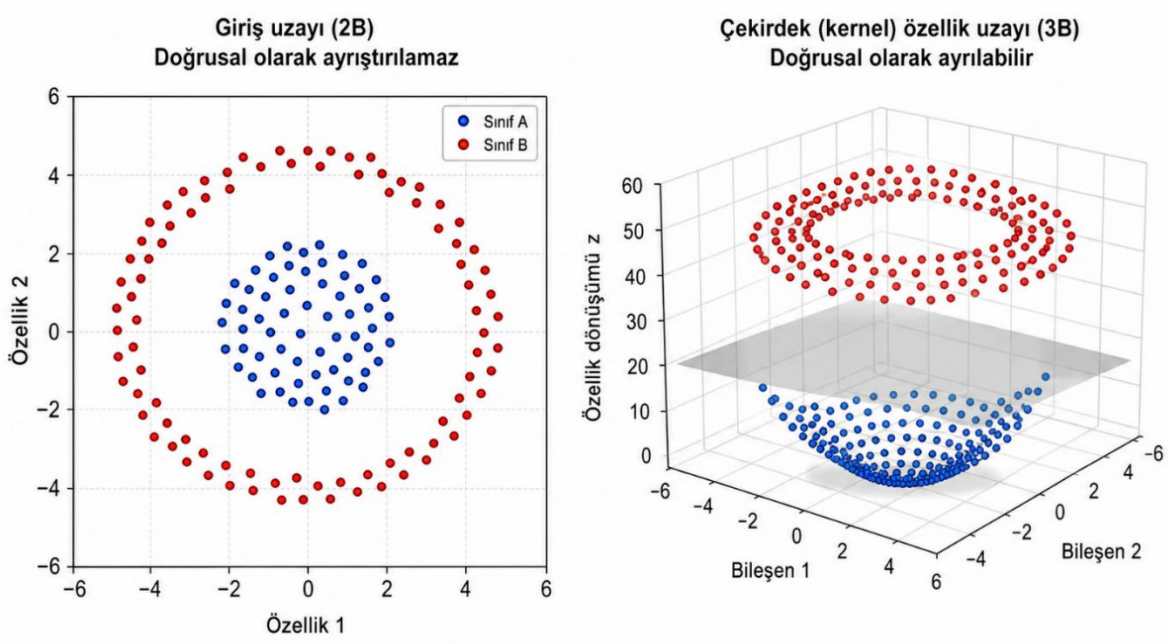
Burada  $\boldsymbol{\varphi}(\cdot)$ , giriş verisini örtük özellik uzayına taşıyan doğrusal olmayan eşleme fonksiyonunu;  $\mathbb{R}^d$  gözlenen özellik uzayını;  $\mathbf{F}$  ise örtük özellik uzayını göstermektedir.

Çekirdek tabanlı yöntemlerde bu örtük uzaydaki dönüşüm çoğu zaman açık biçimde hesaplanmaz. Bunun yerine, iki örnek arasındaki iç çarpım ilişkisi çekirdek fonksiyonu yardımıyla ifade edilir:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j) \quad (\mathbf{D.12})$$

Burada  $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ ,  $\mathbf{x}_i$  ve  $\mathbf{x}_j$  örnekleri arasındaki benzerliği örtük özellik uzayında tanımlayan çekirdek fonksiyonunu göstermektedir. Böylece yüksek boyutlu özellik uzayındaki işlemler, verinin açık dönüşümü hesaplanmadan çekirdek matrisi üzerinden gerçekleştirilebilmektedir (Schölkopf vd., 1998; Baudat ve Anouar, 2000).

Çekirdek yaklaşımında verinin doğrusal olmayan ayırmadan doğrusal ayrılabilir örtük uzaya taşınmasının kavramsal gösterimi Şekil 2.12’de sunulmuştur.



**Şekil 2.12.** Çekirdek yaklaşımında verinin doğrusal olmayan ayırmadan doğrusal ayrılabilir örtük uzaya taşınmasının kavramsal gösterimi

Şekil 2.12’de görüldüğü üzere, başlangıçtaki özellik uzayında doğrusal olarak ayrıştırılması güç olan veri örnekleri, çekirdek yaklaşımı yardımıyla daha ayrılabilir bir temsil uzayına taşınabilmektedir. Bu durum, özellikle FV arıza verilerinde çevresel değişkenlik ve arıza etkilerinin birlikte oluşturduğu karmaşık sınıf dağılımlarının temsil edilmesi açısından önemlidir.

Çekirdek tabanlı yöntemlerin başarısı, kullanılan çekirdek fonksiyonunun türüne ve ilgili parametrelerin seçimine bağlıdır. Uygun çekirdek seçimi sınıflar arası ayrımı güçlendirebilirken, uygun olmayan çekirdek veya parametre seçimi modelin genelleme yeteneğini sınırlayabilir. Bu nedenle bu tezde KPCA, KLDA ve KPCA-ICA yaklaşımı, doğrusal yöntemlerle aynı veri seti ve aynı sınıflandırma altyapısı altında karşılaştırılmıştır.

### 2.5.5. Çekirdek temel bileşen analizi (KPCA)

Çekirdek Temel Bileşen Analizi, PCA’nın doğrusal olmayan veri yapıları için genişletilmiş bir biçimi olarak değerlendirilebilir. PCA, gözlenen özellik uzayında maksimum varyans doğrultularını belirleyen doğrusal bir yöntemdir. Ancak veri yapısı doğrusal olmayan ilişkiler içerdiğinde, PCA’nın elde ettiği bileşenler sınıflar arasındaki karmaşık ayrımı yeterli düzeyde temsil edemeyebilir. KPCA, bu sınırlılığı aşmak amacıyla veriyi çekirdek fonksiyonu

aracılığıyla örtük bir özellik uzayına taşımakta ve temel bileşen analizini bu uzayda gerçekleştirmektedir (Schölkopf vd., 1998).

KPCA'da temel amaç, ham özellik uzayında doğrusal olarak ortaya çıkarılamayan varyans yapılarının, çekirdek fonksiyonları yardımıyla tanımlanan örtük özellik uzayında temsil edilmesidir. Bu sayede doğrusal olmayan örüntüler daha düzenli ve ayrıştırılabilir bir yapıya kavuşabilir. FV sistemlerden elde edilen ölçüm verilerinde arıza etkileri, çevresel koşullar ve sistem dinamikleri birlikte gözlemlendiğinden, bu tür doğrusal olmayan temsiller sınıflandırma açısından avantaj sağlayabilir.

KPCA'nın PCA'dan temel farkı, temel bileşenlerin doğrudan ham veri uzayındaki kovaryans yapısından değil, çekirdek matrisi üzerinden elde edilmesidir. Bu özellik, KPCA'yı doğrusal olmayan sınıf dağılımlarına sahip veri setleri için uygun bir boyut azaltma yöntemi haline getirmektedir.

Bununla birlikte KPCA'nın başarısı kullanılan çekirdek fonksiyonuna ve çekirdek parametrelerine duyarlıdır. Uygun parametre seçimi sınıflar arası ayrımı güçlendirebilirken, uygun olmayan parametreler veri yapısının yetersiz veya aşırı karmaşık temsil edilmesine neden olabilir. Ayrıca KPCA'da bileşenler örtük özellik uzayında tanımlandığı için fiziksel yorumlanabilirlik PCA'ya göre daha sınırlı olabilir.

Sonuç olarak KPCA, FV arıza sınıflandırmasında doğrusal olmayan ölçüm örüntülerini daha ayırt edici bir temsil uzayına taşımak için kullanılabilir. güçlü bir çekirdek tabanlı boyut azaltma yöntemidir. Bu tez kapsamında KPCA, doğrusal yöntemlerle ve diğer çekirdek tabanlı yöntemlerle aynı deneysel yapı altında karşılaştırılmıştır.

### **2.5.6. Çekirdek doğrusal ayırma analizi (KLDA)**

Çekirdek Doğrusal Ayırma Analizi, LDA'nın doğrusal olmayan veri yapıları için genişletilmiş bir biçimi olarak değerlendirilebilir. Klasik LDA, sınıf etiketlerini dikkate alarak sınıflar arası ayrımı artıran ve sınıf içi dağılımı azaltan doğrusal bir temsil uzayı oluşturmayı amaçlamaktadır. Ancak veri yapısında doğrusal olmayan ilişkiler bulunduğunda, klasik LDA'nın doğrusal dönüşüm yapısı sınıflar arasındaki karmaşık ayrımı yeterli düzeyde temsil edemeyebilir. KLDA, bu sınırlılığı gidermek amacıyla veriyi çekirdek fonksiyonu yardımıyla örtük bir özellik uzayına taşımakta ve sınıf ayrımını bu uzayda gerçekleştirmektedir (Baudat ve Anouar, 2000; Bishop, 2006).

KLDA'nın temel avantajı, LDA'nın denetimli sınıf ayrımı mantığını korurken çekirdek yaklaşımı sayesinde doğrusal olmayan sınıf sınırlarını modelleyebilmesidir. FV arıza sınıflandırması açısından bu özellik, özellikle çevresel değişkenlik veya arıza etkilerinin ham ölçüm uzayında sınıflar arası örtüşmeye neden olduğu durumlarda önemlidir.

KLDA, PCA ve KPCA'dan farklı olarak sınıf etiketlerini doğrudan kullanmaktadır. Bu nedenle arıza sınıflarının belirgin etiket yapısına sahip olduğu veri setlerinde sınıflar arası ayrımı güçlendirebilir. Bununla birlikte KLDA'nın başarısı çekirdek fonksiyonu, parametre seçimi, sınıf dağılımı ve etiket kalitesine duyarlıdır. Uygun olmayan parametre seçimi aşırı uyum riskini artırabilir veya sınıf ayrımını zayıflatabilir.

Sonuç olarak KLDA, LDA'nın denetimli yapısını çekirdek yaklaşımıyla birleştiren güçlü bir doğrusal olmayan boyut azaltma yöntemidir. Bu tez kapsamında KLDA, doğrusal LDA ve diğer çekirdek tabanlı yöntemlerle aynı deneysel yapı altında karşılaştırılmıştır.

### **2.5.7. KPCA-ICA yaklaşımı**

Bu tez kapsamında ele alınan KPCA-ICA yaklaşımı, çekirdek tabanlı doğrusal olmayan temsil ile bağımsız bileşen analizini ardışık biçimde birleştiren bir yöntem olarak değerlendirilmiştir. Bu yaklaşımda öncelikle KPCA kullanılarak ham FV ölçüm verileri doğrusal olmayan bir temsil uzayına dönüştürülmekte, ardından bu temsil uzayı üzerinde ICA uygulanarak istatistiksel olarak daha bağımsız bileşenler elde edilmeye çalışılmaktadır.

KPCA-ICA yaklaşımının temel amacı, ham ölçüm uzayında doğrudan gözlenemeyen karmaşık ilişkileri önce çekirdek tabanlı bir temsil uzayında belirgin hale getirmek, daha sonra bu uzaydaki bileşenleri istatistiksel bağımsızlık açısından ayırtmaktır. Böylece KPCA'nın doğrusal olmayan örüntüleri yakalama yeteneği ile ICA'nın gizli ve bağımsız bileşenleri ortaya çıkarma yaklaşımı birlikte kullanılmaktadır (Schölkopf vd., 1998; Hyvärinen vd., 2001).

FV arıza sınıflandırması açısından KPCA-ICA yaklaşımının beklenen katkısı, farklı arıza türlerine ait gizli ve doğrusal olmayan ölçüm örüntülerini daha belirgin hale getirebilmesidir. Ancak yöntem iki aşamalı bir yapıya sahip olduğu için hem KPCA aşamasındaki çekirdek fonksiyonu ve parametre seçimine hem de ICA aşamasındaki bileşen sayısı, yakınsama davranışı ve veri ölçeklendirme tercihine duyarlıdır.

Ayrıca KPCA-ICA yaklaşımı denetimsiz bir temsil yapısı sunduğundan, sınıf etiketlerini doğrudan kullanmamaktadır. Bu nedenle elde edilen bileşenlerin arıza sınıflarını en

iyi şekilde ayıracağı garanti değildir. Buna rağmen, doğrusal olmayan temsil ile bağımsız bileşen analizini birleştirmesi nedeniyle karşılaştırmalı analizde yer alması anlamlıdır.

Bu tez kapsamında KPCA-ICA yaklaşımı, KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde gerçekleştirilmiştir. Yöntem, PCA, LDA, ICA, KPCA ve KLDA ile aynı veri seti ve sınıflandırma altyapısı altında değerlendirilmiştir.

### **2.5.8. Yöntemlerin FV arıza sınıflandırması açısından değerlendirilmesi**

FV sistemlerde arıza sınıflandırma başarısı, yalnızca kullanılan sınıflandırıcı algoritmanın performansına değil, sınıflandırıcıya sunulan veri temsilinin niteliğine de bağlıdır. Akım, gerilim, güç, ısıtım, sıcaklık ve bunlardan türetilen istatistiksel özellikler, arıza durumlarına ilişkin önemli bilgiler taşımakla birlikte; bu özellikler yüksek korelasyon, doğrusal olmayan ilişki, gürültü ve sınıflar arası örtüşme içerebilmektedir. Bu nedenle boyut azaltma yöntemleri, sınıflar arasındaki ayırt edici yapıyı güçlendiren temel bir temsil oluşturma aşaması olarak değerlendirilmelidir.

Bu tez kapsamında ele alınan PCA, LDA, ICA, KPCA, KLDA ve KPCA-ICA yaklaşımları, FV ölçüm verilerini farklı matematiksel ve istatistiksel bakış açılarıyla temsil etmektedir. PCA sade, kararlı ve yorumlanabilir bir doğrusal temsil sunarken, LDA sınıf etiketlerini dikkate alarak ayırt edici bir doğrusal dönüşüm sağlamaktadır. ICA gizli ve bağımsız bileşenlere odaklanırken, KPCA doğrusal olmayan varyans yapısını, KLDA ise doğrusal olmayan sınıf ayrımını öne çıkarmaktadır. KPCA-ICA yaklaşımı ise doğrusal olmayan temsil ile bağımsız bileşen analizini birleştiren alternatif bir temsil yapısı sunmaktadır.

Yöntem seçimi yapılırken yalnızca genel doğruluk başarısına bakmak yeterli değildir. AUC, F1 skoru, duyarlılık, özgüllük, işlem süresi, yorumlanabilirlik ve bozulmuş koşulları altındaki dayanıklılık gibi ölçütler de dikkate alınmalıdır. Özellikle saha benzeri koşullarda gürültü, eksik veri, sensör sürüklenmesi, tek sensör kaybı veya farklı ısıtım bantı koşulları altında yöntemlerin kararlılıkları değişebilmektedir.

Sonuç olarak bu tezde doğrusal ve çekirdek tabanlı boyut azaltma yöntemleri, FV arıza sınıflandırmasında veri temsilinin sınıflandırma performansı üzerindeki etkisini karşılaştırmalı olarak incelemek amacıyla ele alınmıştır. Her bir yöntemin farklı avantajları ve sınırlılıkları bulunduğu için, aynı veri seti, aynı sınıflandırma altyapısı ve aynı performans ölçütleri altında değerlendirilmesi önem taşımaktadır.

Bu kuramsal değerlendirme doğrultusunda, bir sonraki bölümde FV arıza sınıflandırmasında kullanılan sınıflandırıcı yaklaşımı ele alınmakta ve destek vektör makinelerinin temel çalışma mantığı açıklanmaktadır.

## **2.6. Destek vektör makineleri ve sınıflandırma mantığı**

FV sistemlerde arıza sınıflandırma problemi, ölçüm verilerinden elde edilen özelliklerin veya boyut azaltma yöntemleriyle oluşturulan temsil vektörlerinin uygun bir karar mekanizması yardımıyla sınıflara ayrılması şeklinde ele alınmaktadır. Bu tez kapsamında boyut azaltma yöntemleriyle elde edilen temsil uzaylarının sınıflandırma başarısını değerlendirmek amacıyla makine öğrenmesi tabanlı sınıflandırıcılar kullanılmıştır. Bu sınıflandırıcılar içerisinde Destek Vektör Makineleri, yüksek boyutlu, sınırlı örnek sayısına sahip ve sınıflar arası ayrımın karmaşık olduğu veri setlerinde güçlü genelleme yeteneği sunması nedeniyle yaygın biçimde tercih edilmektedir (Cortes ve Vapnik, 1995; Vapnik, 1998).

Destek Vektör Makineleri, farklı sınıflara ait örnekleri birbirinden ayıran en uygun karar sınırını belirlemeyi amaçlayan denetimli bir öğrenme yöntemidir. Yöntemin temel ilkesi, sınıfları yalnızca ayıran herhangi bir karar sınırı bulmak değil, sınıflar arasındaki marjini en büyük yapan karar sınırını elde etmektir. Bu yaklaşım, modelin eğitim verisini doğru sınıflandırmasının yanında, görülmemiş test örnekleri üzerinde de daha iyi genelleme yapmasını hedeflemektedir (Cortes ve Vapnik, 1995; Vapnik, 1998).

SVM'nin bu tezdeki temel rolü, boyut azaltma yöntemleriyle elde edilen temsil uzaylarının ayırt ediciliğini karşılaştırmalı olarak değerlendirmektir. PCA, LDA, ICA, KPCA, KLDA ve KPCA-ICA yöntemleriyle elde edilen düşük boyutlu özellik uzayları aynı sınıflandırma altyapısına sunulduğunda, her bir temsil yönteminin arıza sınıflarını ne ölçüde ayırabileceği daha tutarlı biçimde incelenebilmektedir.

FV arıza verileri normal çalışma, string arızası, string-toprak arızası ve string-string arızası gibi birden fazla sınıf içermektedir. Bu sınıflar ölçüm uzayında her zaman doğrusal olarak ayrılabilir bir yapı göstermeyebilir. Işınım değişimleri, sıcaklık etkisi, ölçüm gürültüsü ve güç elektroniği dinamikleri sınıflar arasındaki ayrımı zorlaştırabilmektedir. Bu nedenle SVM'nin doğrusal ve doğrusal olmayan sınıflandırma yetenekleri, FV arıza sınıflandırması açısından uygun bir değerlendirme çerçevesi sunmaktadır.

Bu bölümde SVM'nin temel çalışma mantığı, optimizasyon problemi, doğrusal olmayan sınıflandırma için kernel yaklaşımı, çok sınıflı SVM yapısı ve performans ölçütleriyle

ilişkisi açıklanmaktadır. Böylece sonraki bölümlerde sunulan boyut azaltma ve sınıflandırma sonuçlarının hangi sınıflandırma mantığına dayandığı kuramsal olarak ortaya konulmaktadır.

### 2.6.1. Temel SVM mantığı

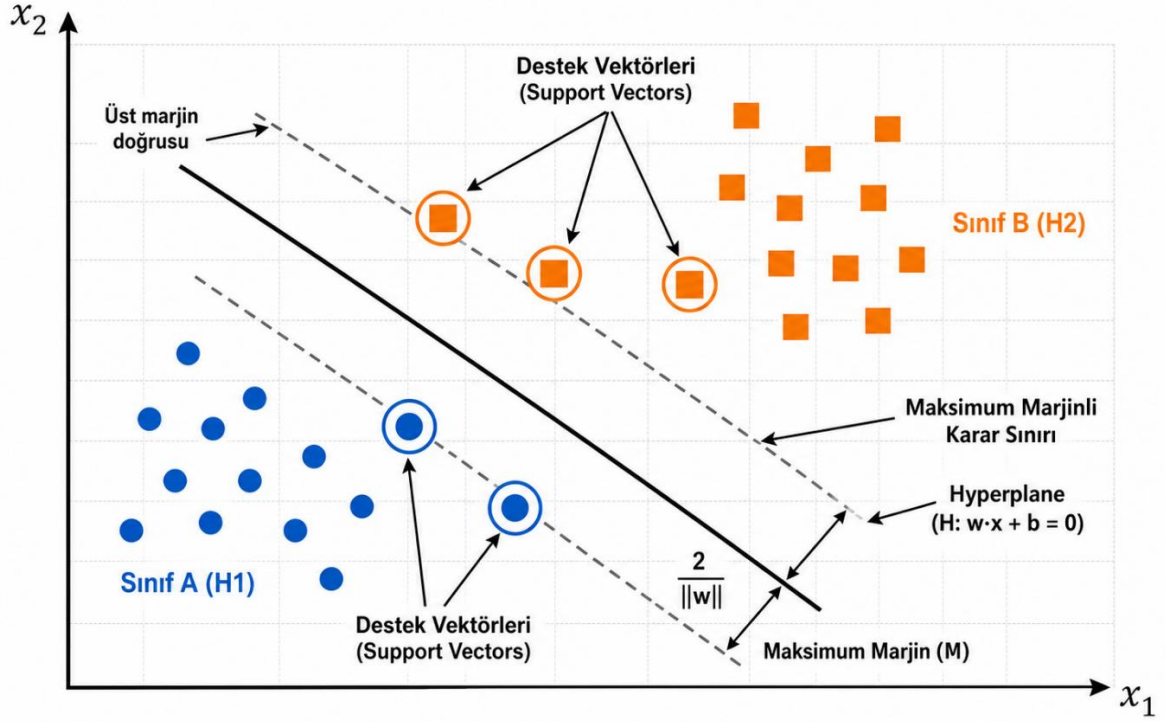
Destek Vektör Makineleri, iki sınıflı sınıflandırma problemleri için geliştirilen ve daha sonra çok sınıflı problemlere genişletilebilen denetimli bir öğrenme yöntemidir. Temel SVM yaklaşımında amaç, farklı sınıflara ait örnekleri birbirinden ayıran bir karar sınırı oluşturmaktır. Doğrusal olarak ayrılabilir bir veri setinde bu karar sınırı, özellik uzayında bir hiper düzlem olarak tanımlanır (Cortes ve Vapnik, 1995; Vapnik, 1998). Bu hiper düzlem genel olarak aşağıdaki biçimde ifade edilebilir:

$$\mathbf{w}^T \mathbf{x} + \mathbf{b} = 0 \quad (\text{D.13})$$

Burada  $\mathbf{w}$ , karar hiper düzleminin normal vektörünü;  $\mathbf{x}$ , giriş özellik vektörünü;  $\mathbf{b}$  ise bias terimini ifade etmektedir. Denklemden verilen hiper düzlem, iki sınıf arasındaki karar sınırını temsil etmektedir. SVM'nin amacı, yalnızca sınıfları ayıran herhangi bir hiper düzlem bulmak değil, sınıflar arasındaki marjini en büyük yapan hiper düzlemi seçmektir.

SVM'de marjin, karar sınırı ile bu sınıra en yakın eğitim örnekleri arasındaki uzaklığı ifade etmektedir. Bu en yakın örnekler **destek vektörleri** olarak adlandırılır. Destek vektörleri, karar sınırının konumunu doğrudan belirleyen kritik örneklerdir. Bu nedenle SVM, tüm veri setinden çok sınıflar arasındaki ayrımı belirleyen örneklere odaklanan bir sınıflandırma yaklaşımı sunmaktadır (Cortes ve Vapnik, 1995; Vapnik, 1998).

SVM'de maksimum marjinli karar sınırı ve destek vektörlerinin kavramsal gösterimi Şekil 2.13'te verilmiştir.



**Şekil 2.13.** SVM'de maksimum marjlinli karar sınırı ve destek vektörlerinin şematik gösterimi

Şekil 2.13'te görüldüğü üzere, farklı sınıflara ait örnekleri ayıran birden fazla karar sınırı oluşturulabilir. Ancak SVM, bu sınırlar içerisinde sınıflar arasındaki marjini en büyük yapan hiper düzlemi tercih etmektedir. Karar sınırına en yakın örnekler destek vektörleri olarak görev yapmakta ve sınıflandırma modelinin temel yapısını belirlemektedir.

FV arıza sınıflandırması açısından destek vektörleri, normal çalışma ve arıza sınıfları arasındaki sınır bölgelerinde yer alan kritik örnekleri temsil edebilir. Örneğin düşük ışımın koşullarında normal çalışma örnekleri ile bazı arıza örnekleri benzer akım veya güç değerleri gösterebilir. Bu tür sınır bölgelerinde yer alan örnekler, sınıflandırma kararının belirlenmesinde önemli rol oynar.

Bununla birlikte SVM'nin başarısı, kullanılan giriş özelliklerinin niteliğine doğrudan bağlıdır. Eğer sınıflandırıcıya sunulan özellik uzayı arıza sınıflarını yeterince ayırt edici biçimde temsil etmiyorsa, SVM'nin karar sınırı da sınıfları etkin biçimde ayıramayabilir. Bu nedenle bu tezde SVM sınıflandırması, boyut azaltma yöntemleriyle elde edilen temsil uzayları üzerinde gerçekleştirilmiştir.

Sonuç olarak SVM, maksimum marjin ilkesine dayalı karar yapısı ve destek vektörleri üzerinden kurduğu sınıflandırma mantığı sayesinde FV arıza sınıflandırması için uygun bir

yöntemdir. Ancak yöntemin başarısı, kullanılan temsil uzayının ayırt ediciliğine, veri dağılımına ve model parametrelerine bağlıdır.

### 2.6.2. Optimizasyon problemi

Destek Vektör Makineleri'nin temelinde, sınıflar arasındaki marjini en büyük yapan karar hiper düzleminin belirlenmesi yer almaktadır. Doğrusal olarak ayrılabilir iki sınıflı bir problemde amaç, eğitim örneklerini doğru sınıflandıran ve karar sınırına en yakın örneklerle arasındaki uzaklığı en büyükleyen hiper düzlemi elde etmektir. Bu yaklaşım, SVM'nin genelleme başarısını artırmayı hedeflemektedir (Cortes ve Vapnik, 1995; Vapnik, 1998).

İki sınıflı bir sınıflandırma probleminde her bir eğitim örneği  $\mathbf{x}_i$  özellik vektörü ve  $y_i$  sınıf etiketi ile temsil edilebilir. Burada  $y_i \in \{-1, +1\}$  olmak üzere, karar hiper düzlemi D.13'te verilen  $\mathbf{w}^T \mathbf{x} + \mathbf{b} = 0$  ifadesi ile tanımlanmaktadır. Doğrusal olarak ayrılabilir durumda SVM optimizasyon problemi genel olarak aşağıdaki biçimde ifade edilebilir:

$$\min_{\mathbf{w}, \mathbf{b}} \frac{1}{2} \|\mathbf{w}\|^2 \quad (\text{D.14})$$

Bu optimizasyon problemi aşağıdaki kısıt altında çözülmektedir:

$$y_i(\mathbf{w}^T \mathbf{x}_i + \mathbf{b}) \geq 1, i = 1, 2, \dots, N \quad (\text{D.15})$$

Burada  $\mathbf{w}$  karar hiper düzleminin normal vektörünü,  $\mathbf{b}$  bias terimini,  $\mathbf{x}_i$  i'inci eğitim örneğini,  $y_i$  ilgili sınıf etiketini ve  $N$  toplam eğitim örneği sayısını göstermektedir. Amaç fonksiyonunda yer alan  $\|\mathbf{w}\|^2$  teriminin minimize edilmesi, karar hiper düzlemi ile destek vektörleri arasındaki marjinin büyütülmesi anlamına gelmektedir. Kısıt koşulu ise her bir eğitim örneğinin doğru sınıfta ve karar sınırının uygun tarafında yer almasını sağlamaktadır.

Pratik uygulamalarda veri setleri çoğu zaman tam olarak doğrusal ayrılabilir değildir. FV arıza sınıflandırmasında da normal çalışma ve arıza durumlarına ait ölçüm örnekleri, çevresel değişkenlik, gürültü veya benzer sinyal davranışları nedeniyle kısmen örtüşebilir. Bu nedenle SVM'de yumuşak marjin yaklaşımı kullanılarak bazı örneklerin karar sınırını ihlal etmesine kontrollü biçimde izin verilmektedir. Bu yapı aşağıdaki şekilde ifade edilebilir:

$$\min_{\mathbf{w}, \mathbf{b}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (\text{D.16})$$

Bu problem aşağıdaki kısıtlar altında çözülür:

$$\mathbf{y}_i(\mathbf{w}^T \mathbf{x}_i + \mathbf{b}) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (\text{D.17})$$

Burada  $\xi_i$ , i'inci eğitim örneği için marjin ihlalini ifade eden gevşeklik değişkenini;  $C$  ise sınıflandırma hatası ile marjin genişliği arasındaki dengeyi belirleyen ceza parametresini göstermektedir.  $C$  parametresinin büyük seçilmesi sınıflandırma hatalarını daha fazla cezalandırırken, küçük seçilmesi daha geniş marjinli fakat hatalara daha toleranslı bir karar sınırı oluşturabilir.

Yumuşak marjin yaklaşımı, FV arıza sınıflandırması açısından önemlidir. Çünkü FV sistemlerden elde edilen ölçüm verileri yalnızca arıza etkilerini değil; ışınım, sıcaklık, sistem dinamikleri ve ölçüm gürültüsü gibi faktörleri de içermektedir. Bu durum, sınıflar arasında tam ayrılabilir bir yapı oluşmasını zorlaştırabilir. SVM'nin yumuşak marjin yapısı, bu tür örtüşmeli veri durumlarında daha dengeli bir sınıflandırma performansı elde edilmesine katkı sağlayabilir.

Bu tez kapsamında SVM optimizasyon yapısı, boyut azaltma yöntemleriyle elde edilen temsil uzaylarının sınıflandırma başarısını değerlendirmek amacıyla kullanılmıştır. PCA, LDA, ICA, KPCA, KLDA ve KPCA-ICA yöntemleriyle elde edilen özellik uzayları aynı sınıflandırma altyapısına sunulurken, her bir temsil yaklaşımının SVM karar sınırı tarafından ne ölçüde ayrıştırılabildiği karşılaştırmalı olarak incelenmiştir.

Sonuç olarak SVM optimizasyon problemi, sınıflar arasındaki marjini en büyükleyen ve sınıflandırma hatalarını kontrol eden bir karar yapısı oluşturmaya dayanmaktadır. FV arıza verilerinde sınıflar her zaman doğrusal olarak ayrılabilir olmadığından, bir sonraki alt bölümde doğrusal olmayan sınıflandırma için kernel yaklaşımı ele alınmaktadır.

### 2.6.3. Doğrusal olmayan sınıflandırma ve kernel mantığı

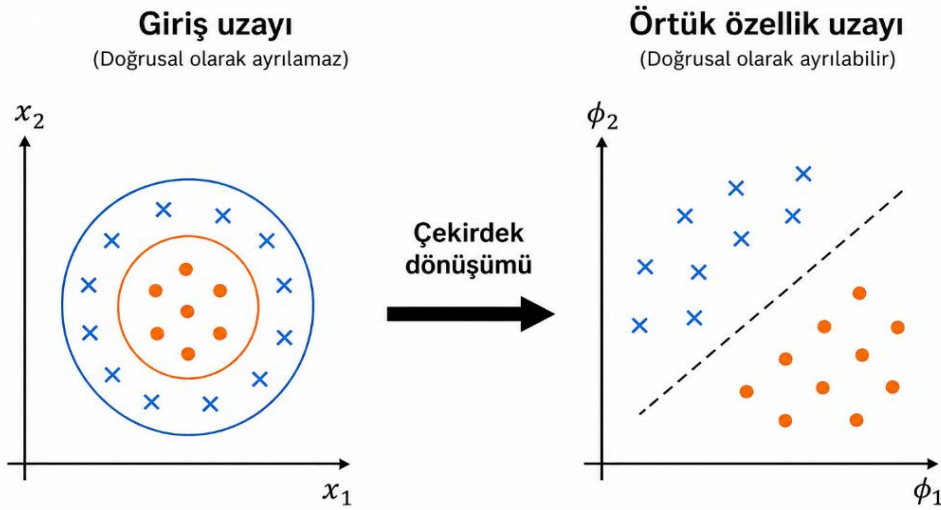
Doğrusal SVM yaklaşımı, sınıfların özellik uzayında doğrusal bir hiper düzlem ile ayrılabilirdiği durumlarda etkili bir sınıflandırma yapısı sunmaktadır. Ancak pratik veri setlerinde sınıflar çoğu zaman doğrusal olarak ayrılabilir bir dağılım göstermemektedir. FV arıza sınıflandırması açısından da normal çalışma, string arızası, string-toprak arızası ve string-string arızası gibi sistem durumları ölçüm uzayında karmaşık ve kısmen örtüşen örüntüler oluşturabilmektedir.

Bu nedenle SVM'de doğrusal olmayan sınıflandırma problemlerini çözebilmek için kernel yaklaşımı kullanılmaktadır. Kernel yaklaşımının temel fikri, giriş verilerini doğrusal olarak ayrıştırılması güç olan ham özellik uzayından daha yüksek boyutlu veya örtük bir özellik

uzayına taşımaktır. Bu örtük uzayda, başlangıçta doğrusal olarak ayırlamayan örnekler daha ayrılabilir bir yapı kazanabilir (Cortes ve Vapnik, 1995; Vapnik, 1998).

Kernel yaklaşımı, verinin yüksek boyutlu özellik uzayındaki açık dönüşümünü doğrudan hesaplamak yerine, örnekler arasındaki benzerlik ilişkilerini kernel fonksiyonu aracılığıyla ifade eder. Böylece SVM’de doğrusal olmayan sınıf sınırları, yüksek boyutlu örtük uzayda doğrusal ayırım mantığıyla modellenenmektedir. Bu yapı, özellikle ham ölçüm uzayında iç içe geçmiş veya doğrusal olmayan dağılım gösteren FV arıza sınıflarının ayrıştırılması açısından önemlidir.

Doğrusal ayırlamayan verinin kernel yaklaşımı ile ayrıştırılmasının kavramsal gösterimi Şekil 2.14’te sunulmuştur.



**Şekil 2.14.** Doğrusal ayırlamayan verinin kernel yaklaşımı ile ayrıştırılmasının kavramsal gösterimi

Şekil 2.14’te görüldüğü üzere, ham özellik uzayında doğrusal bir karar sınırı ile ayrılması güç olan veri örnekleri, kernel fonksiyonu yardımıyla daha yüksek boyutlu bir temsil uzayına taşındığında doğrusal olarak ayrılabilir hale gelebilmektedir. Bu dönüşüm, SVM’nin doğrusal olmayan sınıflandırma problemlerinde daha esnek karar sınırları oluşturmasına olanak sağlamaktadır.

SVM’de yaygın olarak kullanılan kernel fonksiyonları arasında doğrusal kernel, polinom kernel ve radyal tabanlı fonksiyon kernel yer almaktadır. Doğrusal kernel, veri yapısının doğrusal ayırım için yeterli olduğu durumlarda tercih edilebilirken; polinom ve RBF kernel daha karmaşık doğrusal olmayan ilişkileri modelleyebilmektedir. Özellikle RBF kernel,

doğrusal olmayan ve karmaşık dağılım gösteren sınıflandırma problemlerinde yaygın olarak kullanılmaktadır.

Kernel seçimi ve kernel parametreleri, SVM sınıflandırma başarısı üzerinde doğrudan etkiye sahiptir. Örneğin RBF kernel kullanıldığında gamma parametresi, tek bir eğitim örneğinin karar sınırı üzerindeki etki alanını belirlemektedir. Gamma değerinin çok büyük seçilmesi aşırı uyum riskini artırabilirken, çok küçük seçilmesi sınıflar arasındaki ayrımı yeterince temsil edemeyebilir. Benzer şekilde ceza parametresi  $C$ , sınıflandırma hataları ile marjin genişliği arasındaki dengeyi belirlemektedir.

FV arıza sınıflandırmasında kernel tabanlı SVM kullanımının temel avantajı, doğrusal olmayan sınıf yapılarının daha esnek biçimde modellenebilmesidir. FV sistemlerde farklı arıza türleri bazı ölçüm değişkenlerinde benzer değişimler oluşturabilir. Kernel tabanlı SVM, bu tür karmaşık ve örtüşmeli veri yapılarında doğrusal sınıflandırıcıya göre daha ayırt edici karar sınırları oluşturma potansiyeline sahiptir.

Sonuç olarak kernel yaklaşımı, SVM'nin doğrusal olmayan sınıflandırma problemlerinde kullanılabilmesini sağlayan temel mekanizmadır. Bununla birlikte uygun kernel ve parametre seçimi, modelin genelleme başarısı açısından kritik öneme sahiptir. Bu nedenle bir sonraki alt bölümde, birden fazla arıza sınıfının bulunduğu durumlar için çok sınıflı SVM yaklaşımı ele alınmaktadır.

#### **2.6.4. Çok sınıflı SVM yaklaşımı**

Destek Vektör Makineleri başlangıçta iki sınıflı sınıflandırma problemleri için geliştirilmiş bir yöntemdir. Ancak FV arıza sınıflandırması problemlerinde sistem durumu yalnızca iki sınıfla ifade edilememektedir. Bu tez kapsamında normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört farklı sistem durumu ele alınmıştır. Bu nedenle SVM yapısının çok sınıflı sınıflandırma problemlerine uygun biçimde genişletilmesi gerekmektedir.

Çok sınıflı SVM yaklaşımında temel amaç, birden fazla sınıfa ait örnekleri uygun karar sınırları yardımıyla birbirinden ayırmaktır. Bunun için iki sınıflı SVM yapısı farklı stratejiler kullanılarak çok sınıflı problemlere uyarlanabilmektedir. Yaygın yaklaşımlar arasında bire-karşı-hepsi ve bire-karşı-bir stratejileri yer almaktadır.

Bire-karşı-hepsi yaklaşımında sınıf sayısı kadar ikili SVM modeli oluşturulur. Her bir modelde ilgili sınıfa ait örnekler pozitif sınıf, diğer tüm sınıflara ait örnekler ise negatif sınıf

olarak değerlendirilir. Bire-karşı-bir yaklaşımında ise her sınıf çifti için ayrı bir SVM sınıflandırıcısı eğitilir ve nihai sınıf etiketi genellikle kararların birleştirilmesiyle belirlenir.

FV arıza sınıflandırması açısından çok sınıflı SVM yaklaşımı gereklidir. Çünkü arıza teşhisi yalnızca sistemin normal veya arızalı olduğunu belirlemekle sınırlı değildir; arızanın türünün de doğru biçimde sınıflandırılması gerekir. String arızası, string-toprak arızası ve string-string arızası sistem üzerinde farklı fiziksel etkiler oluşturmakta ve bakım stratejisi açısından farklı müdahaleler gerektirmektedir.

Çok sınıflı SVM yapısının başarısı, kullanılan ikili sınıflandırıcıların ayırt ediciliğine ve karar birleştirme stratejisinin etkinliğine bağlıdır. Eğer bazı sınıflar ölçüm uzayında birbirine yakın konumlanıyorsa veya benzer elektriksel etkiler oluşturuyorsa, ilgili sınıf çiftleri arasında hatalı sınıflandırmalar ortaya çıkabilir. Bu nedenle çok sınıflı SVM performansı, yalnızca sınıflandırıcının yapısına değil, boyut azaltma sonrasında elde edilen temsil uzayının sınıfları ne ölçüde ayrıştırılabildiğine de bağlıdır.

Bu tez kapsamında çok sınıflı SVM yaklaşımı, boyut azaltma yöntemleriyle elde edilen temsil uzaylarının sınıflandırma başarısını değerlendirmek için kullanılmıştır. PCA, LDA, ICA, KPCA, KLDA ve KPCA-ICA yöntemleri sonucunda elde edilen düşük boyutlu temsil vektörleri, aynı çok sınıflı sınıflandırma yapısına sunulurken yöntemlerin arıza sınıfları üzerindeki ayırt ediciliği karşılaştırılmıştır.

Sonuç olarak çok sınıflı SVM yaklaşımı, FV sistemlerde birden fazla arıza türünün aynı sınıflandırma çerçevesi içerisinde değerlendirilmesine olanak sağlamaktadır. Bu tezde ele alınan dört sınıflı problem yapısı, SVM'nin çok sınıflı sınıflandırma stratejileriyle uyumlu biçimde modellenmiştir.

#### **2.6.5. SVM'nin fotovoltaik arıza sınıflandırması açısından uygunluğu**

SVM'nin FV arıza sınıflandırması açısından uygunluğu, yöntemin yüksek boyutlu ve kısmen örtüşmeli veri yapılarında güçlü genelleme yeteneği sunmasından kaynaklanmaktadır. FV sistemlerden elde edilen ölçüm verileri; akım, gerilim, güç, ışınım, sıcaklık ve türetilmiş istatistiksel özellikleri içermekte, bu değişkenler çevresel koşullar ve arıza etkileri nedeniyle karmaşık ilişkiler gösterebilmektedir. SVM, bu tür veri yapılarında sınıflar arasında maksimum marjlinli karar sınırları oluşturarak etkili bir sınıflandırma çerçevesi sunmaktadır.

SVM sınıflandırma yapısının temel bileşenleri Tablo 2.10'da özetlenmiştir.

**Tablo 2.10.** SVM sınıflandırma yapısının temel bileşenleri

Bileşen	Açıklama	Tezdeki Rolü
Hiper düzlem	Sınıfları ayıran karar sınırı	Boyut azaltılmış uzayda sınıf ayrımı yapar
Marjin	Karar sınırı ile en yakın örnekler arası mesafe	Genelleme yeteneğini artırır
Destek vektörleri	Karar sınırını belirleyen kritik örnekler	Sınıf ayrımını belirleyen temel noktalar
Ceza parametresi	Hata-ceza dengesi	Aşırı uyum ile esneklik arasında denge kurar
Kernel fonksiyonu	Doğrusal olmayan ayrımı mümkün kılar	Karmaşık PV veri yapılarında etkili olabilir

Tablo 2.10’da görüldüğü üzere, SVM yapısı karar hiper düzlemi, destek vektörleri, marjin, kernel fonksiyonu ve model parametreleri gibi temel bileşenlerden oluşmaktadır. Bu bileşenler birlikte değerlendirildiğinde, SVM’nin hem doğrusal hem de doğrusal olmayan sınıflandırma problemlerine uyarlanabilir bir yapı sunduğu görülmektedir.

FV arıza sınıflandırmasında SVM’nin önemli avantajlarından biri, boyut azaltma yöntemleriyle birlikte kullanılabilmesidir. Boyut azaltma sonucunda elde edilen daha kompakt ve ayırt edici temsil uzayları, SVM’nin karar sınırını daha kararlı biçimde oluşturmasına katkı sağlayabilir. Bu nedenle bu tezde SVM, farklı boyut azaltma yöntemlerinin temsil başarısını karşılaştırmak için ortak sınıflandırıcı yapısı olarak kullanılmıştır.

SVM’nin FV arıza sınıflandırmasındaki avantaj ve sınırlılıkları Tablo 2.11’de verilmiştir.

**Tablo 2.11.** SVM'nin FV arıza sınıflandırmasındaki avantaj ve sınırlılıkları

<b>Kriter</b>	<b>Avantaj</b>	<b>Sınırlılık</b>
Yüksek boyutlu veri	Etkili karar sınırı üretir	Parametre ayarı gerekebilir
Doğrusal olmayan yapı	Kernel ile esnek genişleme sağlar	Kernel seçimine duyarlıdır
Çok sınıflı problem	Genişletilmiş stratejilerle uygulanabilir	Yorumlanabilirlik bazen sınırlıdır
Kararlılık	Marjın maksimizasyonu avantaj sağlar	Gürültülü veride parametre hassasiyeti olabilir

Tablo 2.11 incelendiğinde, SVM'nin yüksek boyutlu veriyle çalışabilme, maksimum marjlinli karar yapısı oluşturma ve kernel fonksiyonlarıyla doğrusal olmayan sınıflandırmaya uygun olma gibi avantajlara sahip olduğu görülmektedir. Buna karşılık model başarısı kernel tipi, ceza parametresi, gamma değeri ve kullanılan özellik temsilinin niteliğine duyarlıdır. Bu nedenle SVM performansı, uygun veri temsili ve parametre seçimiyle birlikte değerlendirilmelidir.

Sonuç olarak SVM, FV arıza sınıflandırmasında etkili ve esnek bir sınıflandırıcı olarak değerlendirilebilir. Ancak bu tezde asıl odak SVM'nin kendisinden çok, farklı boyut azaltma yöntemlerinin SVM sınıflandırıcısı üzerindeki etkisini karşılaştırmalı olarak incelemektir.

### **2.6.6. Performans ölçütleri ile ilişkisi**

FV arıza sınıflandırmasında kullanılan yöntemlerin değerlendirilmesi, yalnızca genel doğruluk değerine dayandırılmamalıdır. Çünkü bazı arıza sınıfları diğerlerine göre daha kritik güvenlik ve işletme sonuçları doğurabilir. Bu nedenle sınıflandırma performansının doğruluk, kesinlik, duyarlılık, F1 skoru, özgüllük, AUC değeri, karışıklık matrisi ve işlem süresi gibi farklı ölçütlerle birlikte değerlendirilmesi gerekmektedir (Fawcett, 2006; Sokolova ve Lapalme, 2009). Bu ölçütlerin matematiksel tanımları ve tez kapsamında kullanım biçimleri Bölüm 3.5'te verilmiştir.

Doğruluk, tüm örnekler içinde doğru sınıflandırılan örneklerin oranını göstermektedir. Ancak sınıf dağılımının dengesiz olduğu durumlarda tek başına yeterli bir ölçüt olmayabilir. Kesinlik, modelin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten pozitif olduğunu gösterirken; duyarlılık, ilgili sınıfa ait gerçek örneklerin ne kadarının doğru tespit

edildiğini ifade etmektedir. F1 skoru ise kesinlik ve duyarlılık değerlerini birlikte dikkate alan dengeli bir ölçüttür.

Özgüllük, negatif sınıfa ait örneklerin doğru biçimde ayırt edilmesini değerlendirmektedir. AUC değeri ise ROC eğrisi altında kalan alanı ifade etmekte ve sınıflandırıcının farklı eşik değerleri altında sınıflar arasında ne ölçüde ayırım yapabildiğini göstermektedir. Karışıklık matrisi ise hangi sınıfların doğru sınıflandırıldığını ve hangi sınıfların birbirine karıştırıldığını ayrıntılı biçimde ortaya koymaktadır.

Boyut azaltma sonrası SVM tabanlı sınıflandırma akışı Şekil 2.15'te gösterilmektedir.



**Şekil 2.15.** Boyut azaltma sonrası SVM tabanlı sınıflandırma akışının şematik gösterimi

Şekil 2.15'te görüldüğü üzere, FV sistemden elde edilen ölçüm verileri öncelikle özellik çıkarma ve ön işleme adımlarından geçirilmekte, ardından boyut azaltma yöntemleriyle daha düşük boyutlu bir temsil uzayına dönüştürülmektedir. Elde edilen temsil vektörleri SVM sınıflandırıcısına verilmekte ve sınıflandırma sonuçları farklı performans ölçütleriyle değerlendirilmektedir. Bu akış, tez kapsamında uygulanan genel sınıflandırma sürecini özetlemektedir.

Bu tez kapsamında performans ölçütleri, boyut azaltma yöntemlerinin yalnızca genel doğruluk açısından değil, sınıflar arası ayırt edicilik, hata dağılımı, işlem süresi ve dayanıklılık açısından da karşılaştırılmasına olanak sağlamaktadır. Özellikle AUC ve F1 skoru, sınıflandırma başarısının daha dengeli biçimde yorumlanmasına katkı sunarken; karışıklık matrisi arıza sınıfları arasındaki olası karışmaları ayrıntılı biçimde göstermektedir.

Sonuç olarak performans ölçütleri, SVM sınıflandırıcısının ve boyut azaltma yöntemlerinin FV arıza sınıflandırmasındaki etkinliğini çok yönlü biçimde değerlendirmek için kullanılmaktadır. Bu nedenle deneysel bulgular bölümünde yöntemler yalnızca doğruluk değerleri üzerinden değil, farklı performans ölçütleri ve dayanıklılık senaryoları üzerinden karşılaştırılmıştır.

### 3. FOTOVOLTAİK SİSTEM MODELİ VE VERİ SETİ OLUŞTURMA SÜRECİ

Bu bölümde, tez kapsamında kullanılan FV sistem modeli, arıza senaryoları, veri seti oluşturma süreci, özellik mühendisliği yaklaşımı ve model eğitime hazırlık adımları açıklanmaktadır. Çalışmanın deneysel altyapısı, MATLAB/Simulink ortamında oluşturulan şebeke bağlantılı FV sistem modeline dayanmaktadır. Bu model üzerinden normal çalışma ve farklı arıza durumları simüle edilmiş; elde edilen akım, gerilim, güç ve çevresel değişkenler sınıflandırma sürecinde kullanılmak üzere veri setine dönüştürülmüştür.

Bu bölümün temel amacı, kuramsal olarak açıklanan FV arıza türlerinin simülasyon ortamında nasıl temsil edildiğini ve bu simülasyonlardan elde edilen verilerin makine öğrenmesi tabanlı sınıflandırmaya nasıl hazırlandığını ortaya koymaktır. Bu kapsamda sistem modeli, arıza senaryoları, ölçülen değişkenler, özellik çıkarma yaklaşımı ve boyut azaltma öncesi veri düzenleme süreci ele alınmaktadır.

#### 3.1. Simülasyon Ortamı ve Fotovoltaik Sistem Modeli

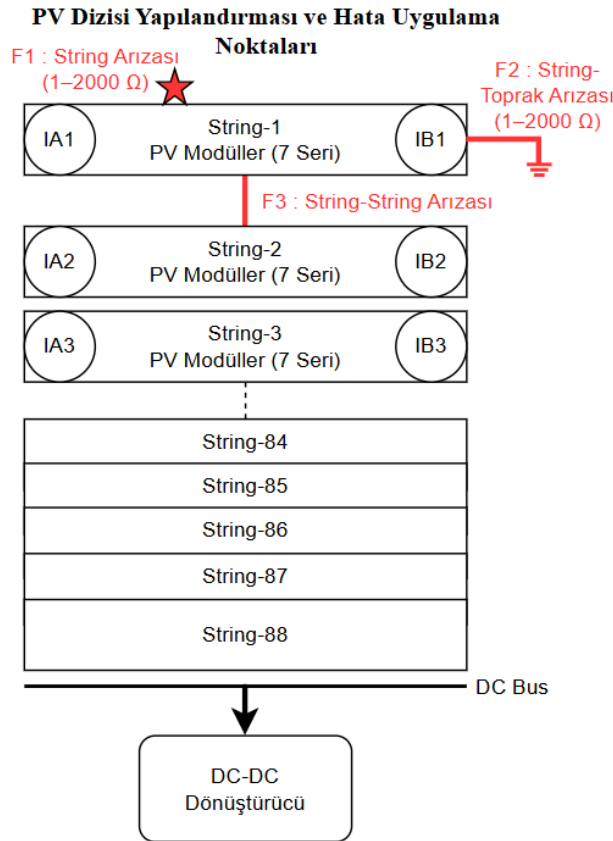
Bu tez çalışmasında, FV sistemlerde arıza tespiti ve sınıflandırılması amacıyla MATLAB/Simulink ortamında modellenen şebeke bağlantılı bir FV sistem yapısı kullanılmıştır. Simülasyon modelinin temel yapısı, Ghoneim vd. (2021) tarafından sunulan 250 kW gücündeki FV santral arıza tespit çalışması ile Rashed tarafından GitHub ortamında paylaşılan “Fault-Detection-Dataset-in-Photovoltaic-Farms” deposundaki MATLAB kodları ve veri seti esas alınarak oluşturulmuştur. Kaynak çalışmada, 250 kW gücünde şebeke bağlantılı bir FV santral modeli kullanılmış; normal çalışma durumu ile string arızası, string-toprak arızası ve string-string arızası olmak üzere farklı sistem durumları ele alınmıştır (Ghoneim vd., 2021; Rashed, 2022).

Ghoneim vd. (2021) tarafından verilen çalışmada, FV sistemden elde edilen elektriksel ve çevresel ölçüm büyüklükleri kullanılarak arıza tespitine yönelik veri seti oluşturulmuştur. İlgili çalışmada akım, gerilim, güç, sıcaklık, ışıma ve arıza direnci gibi değişkenler dikkate alınmış; normal çalışma, string arızası, string-toprak arızası ve string-string arızası durumları sınıflandırma problemi kapsamında değerlendirilmiştir. Ayrıca çalışmada lojistik regresyon, rastgele orman, Naive Bayes, AdaBoost ve CN2 gibi farklı sınıflandırıcıların performansları doğruluk, kesinlik, duyarlılık, F-ölçütü, özgüllük ve AUC gibi ölçütler üzerinden karşılaştırılmıştır (Ghoneim vd., 2021).

Bu tez çalışmasında kullanılan FV dizi yapısı, SunPower SPR-415E-WHT-D PV panel modeli esas alınarak düzenlenmiştir. Bu panel modelinin STC koşullarındaki maksimum güç

değeri yaklaşık 414,8 W, açık devre gerilimi 85,3 V, kısa devre akımı 6,09 A, maksimum güç noktası gerilimi 72,9 V ve maksimum güç noktası akımı 5,69 A'dır. Panel monokristal silikon hücre yapısına sahip olup referans çalışma koşulları 1000 W/m<sup>2</sup> ışınım ve 25 °C hücre sıcaklığı olarak alınmıştır. Modelde her bir string 7 seri PV modülden oluşmakta ve toplam 88 paralel string bulunmaktadır. Buna göre simülasyon modelinde kullanılan toplam PV panel/modül sayısı  $7 \times 88 = 616$ 'dır. Kullanılan panel gücü dikkate alındığında FV dizinin nominal DC kurulu gücü yaklaşık 255,5 kW olup sistem bu tez kapsamında 250 kW sınıfında şebeke bağlantılı FV sistem olarak değerlendirilmiştir.

Tez kapsamında kullanılan güncellenmiş FV dizi yapısı ve hata uygulama noktaları Şekil 3.1'de verilmiştir. Şekilde, String-1, String-2 ve String-3 ayrıntılı olarak gösterilmiş; String-4 ile String-83 arası sadeleştirilmiş; son kısım ise String-84, String-85, String-86, String-87 ve String-88 olarak verilmiştir. Bu gösterim, toplam 88 paralel string yapısını sade ve okunabilir biçimde ifade etmek amacıyla tercih edilmiştir. Şekilde F1 string arızasını, F2 string-toprak arızasını ve F3 string-string arızasını temsil etmektedir. Simülasyonda kullanılan FV dizi yapısı Şekil 3.1'de verilmiştir.



**Şekil 3.1.** Simülasyonda kullanılan FV dizi yapısı

Şekil 3.1’de verilen yapı üzerinden normal çalışma durumu ve arıza durumlarına ait simülasyon verileri elde edilmiştir. FV sistemden alınan ölçüm verileri; string akımları, toplam akım, DC gerilim, DC güç, ışıınım, sıcaklık ve bunlardan türetilen istatistiksel özelliklerden oluşmaktadır. Elde edilen veri yapısı, daha sonraki aşamalarda veri ön işleme, özellik mühendisliği, boyut azaltma ve sınıflandırma adımlarında kullanılmıştır. Bu yönüyle simülasyon modeli, tez çalışmasında kullanılan veri setinin oluşturulması ve arıza sınıflandırma sürecinin temelini oluşturmaktadır.

Bu çalışmada kullanılan modelleme yaklaşımı, kaynak çalışmadaki FV arıza senaryoları ve veri üretim mantığı korunarak; tez kapsamında incelenen 7 seri–88 paralel string yapısı, panel elektriksel özellikleri, arıza uygulama noktaları, boyut azaltma yöntemleri ve çoklu sınıflandırıcı değerlendirme süreci doğrultusunda güncellenmiştir. Bu nedenle tezde kullanılan simülasyon yapısı, Ghoneim vd. (2021) ve Rashed (2022) kaynaklarından uyarlanarak yazar tarafından güncellenmiş bir FV arıza sınıflandırma modeli olarak değerlendirilmiştir.

### **3.2. Arıza Senaryoları ve Veri Seti Oluşturma Süreci**

Bu tez çalışmasında FV sistemin normal çalışma durumu ile üç farklı arıza durumu ele alınmıştır. Arıza durumları; **string arızası**, **string-toprak arızası** ve **string-string arızası** olarak belirlenmiştir. Böylece sınıflandırma problemi; normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört sınıflı bir yapı altında tanımlanmıştır (Ghoneim vd., 2021). Bu sınıflar, FV sistemlerde karşılaşılabilecek temel DC tarafı arıza davranışlarını temsil edecek şekilde seçilmiştir.

Normal çalışma durumu, FV sistemde herhangi bir elektriksel arıza bulunmadığı ve sistemin çevresel koşullara bağlı olarak olağan üretim davranışı gösterdiği durumu ifade etmektedir. Bu durumda string akımları arasında belirgin bir dengesizlik beklenmemekte; DC gerilim ve güç değerleri sistemin normal işletme karakteristiği doğrultusunda değişmektedir. Normal çalışma verileri, arıza durumlarının ayırt edilebilmesi için referans sınıf niteliği taşımaktadır.

String arızası senaryosunda, FV dizide yer alan bir stringin akım katkısının azalması veya ilgili stringin devre dışı kalmasına benzer bir durum modellenmiştir. Bu arıza türü, string akımları arasında dengesizlik oluşturmakta ve toplam dizi akımında azalmaya neden olabilmektedir. String arızasının elektriksel etkisi, ilgili string üzerindeki gerilim düşümü ve arıza direnci etkisiyle ilişkilendirilebilir. Bu durum aşağıdaki ifade ile temsil edilebilir:

$$V_{string} = V_{pv} - I(R_s + R_f) \quad (D.18)$$

Burada **Vstring** string terminal gerilimini, **Vpv** FV kaynak gerilimini, **I** akımı, **Rs** seri direnç etkisini ve **Rf** arıza direncini ifade etmektedir. Bu ifade, string arızasının sistem üzerinde oluşturduğu ek direnç etkisini ve buna bağlı gerilim değişimini açıklamak amacıyla kullanılmıştır.

String-toprak arızası senaryosunda, FV sistemin DC tarafı ile toprak arasında istenmeyen bir elektriksel yol oluşması modellenmiştir. Bu arıza türü, özellikle izolasyon bozulması veya iletken bölümlerin toprakla teması sonucunda ortaya çıkabilecek kaçak akım davranışını temsil etmektedir. Topraklama arızasında oluşabilecek kaçak akım genel olarak şu şekilde ifade edilebilir:

$$I_g = \frac{V_{string}}{R_g} \quad (D.19)$$

Burada **Ig** topraklama arızası akımını, **Vstring** string gerilimini ve **Rg** topraklama arızası direncini göstermektedir. Arıza direncinin azalması durumunda kaçak akımın artması beklenmektedir. Bu nedenle string-toprak arızaları, enerji üretim performansının yanında güvenlik açısından da kritik arıza türleri arasında değerlendirilmektedir.

String-string arızası senaryosunda ise paralel stringler arasında istenmeyen bir iletim yolu oluşması modellenmiştir. Bu durumda farklı potansiyel seviyelerine sahip stringler arasında akım akışı meydana gelebilmekte ve normal akım paylaşımı bozulmaktadır. String-string arızasında oluşan arıza akımı genel olarak aşağıdaki ifade ile gösterilebilir:

$$I_{fault} = \frac{V_1 - V_2}{R_f} \quad (D.20)$$

Burada **Ifault** stringler arası arıza akımını, **V1** ve **V2** arızaya dahil olan stringlerin gerilimlerini, **Rf** ise arıza direncini ifade etmektedir. Bu senaryoda arıza etkisi yalnızca tek bir string üzerinde değil, paralel bağlı stringler arasındaki akım paylaşımı üzerinde de gözlenmektedir. Bu nedenle string-string arızası, sınıflandırma açısından daha karmaşık bir sinyal örüntüsü oluşturabilmektedir.

Simülasyon sırasında sistemin farklı çevresel ve arıza parametreleri altındaki davranışı gözlemlenmiştir. Bu kapsamda sıcaklık, ışınlam ve arıza direnci gibi değişkenler belirli aralıklarda değiştirilmiş; sistemin farklı çalışma koşullarındaki tepkileri kaydedilmiştir. Sıcaklık değişimi FV hücre gerilimi ve verimi üzerinde etkili olurken, ışınlam değişimi özellikle

üretilem akım ve güç seviyelerini deęiřtirmektedir. Arıza direnci ise arıza řiddetini ve ölçüm sinyallerine yansıyan deęiřimin büyüklüğünü belirleyen temel parametrelerden biridir.

Simülasyonda kullanılan deęiřkenler ve çalışma aralıkları Tablo 3.1’de verilmiştir.

**Tablo 3.1.** Simülasyonda kullanılan deęiřkenler ve çalışma aralıkları

Deęiřken	Sembol	Aralık	Acıklama
Sıcaklık	T	10–35 °C	Panel çalışma sıcaklığı
Iřınım	IR / G	100–1000 W/m <sup>2</sup>	Çevresel giriş deęiřkeni
Arıza direnci	Rf	1–2000 Ω	Arıza şiddetini temsil eder

Tablo 3.1’de görüldüğü üzere, veri üretim sürecinde FV sistemin yalnızca sabit bir işletme noktasındaki davranışı deęil, farklı çevresel koşullar ve arıza parametreleri altındaki tepkisi de dikkate alınmıştır. Bu yaklaşım, elde edilen veri setinin daha geniş bir çalışma aralığını temsil etmesine katkı sağlamaktadır.

Oluřturulan veri setinde her bir örnek, belirli bir işletme koşulu ve arıza durumuna karşılık gelecek şekilde düzenlenmiştir. Her örnek için sistemden elde edilen akım, gerilim, güç, sıcaklık ve ışınım deęiřkenleri kaydedilmiş; daha sonra bu veriler sınıf etiketi ile birlikte denetimli öğrenmeye uygun hale getirilmiştir. Böylece her bir veri örneęi, hem fiziksel ölçüm bilgilerini hem de ait olduđu işletme durumunu içeren bir gözlem olarak tanımlanmıştır.

Eđitim veri setinin arıza sınıflarına göre dağılımı Tablo 3.2’de verilmiştir.

**Tablo 3.2.** Eđitim veri setinin arıza sınıflarına göre dağılımı

Arıza Türü	Tanım	Kayıt Yüzdesi	Örnek Sayısı
Normal	Hata yok, sağlıklı çalışma	16,67%	100
String Arızası	Açık devre veya modül kaybı	25,50%	153
String Toprak Arızası	Pozitif/Negatif terminal toprak teması	24,83%	149
Stringler Arası Arıza	Diziler arası kısa devre	33,00%	198

Tablo 3.2’ye göre veri seti dört sınıftan oluşmaktadır. Veri setinde 100 normal çalışma örneęi, 153 string arızası örneęi, 149 string-toprak arızası örneęi ve 198 string-string arızası örneęi yer almaktadır. Bu dağılım, literatürde 250 kW řebeke bağlantılı FV sistem için oluşturulan benzer simülasyon tabanlı veri seti yapısıyla uyumludur (Ghoneim vd., 2021). Bununla birlikte sınıflar arasında tam dengeli olmayan bir dağılım bulunduğundan, model

performansının yalnızca genel doğruluk ile değil, F1 skoru, AUC ve karışıklık matrisi gibi ölçütlerle de değerlendirilmesi gerekmektedir.

Veri seti oluşturma sürecinde dikkat edilen temel noktalardan biri, arıza sınıflarının yalnızca etiket düzeyinde değil, fiziksel sinyal davranışı bakımından da anlamlı biçimde temsil edilmesidir. Bu nedenle her arıza senaryosunda yalnızca toplam güç değişimi değil; string akımları, toplam akım, DC gerilim, DC güç ve çevresel değişkenler birlikte kaydedilmiştir. Bu yaklaşım, arıza sınıflarının çok değişkenli ölçüm uzayında daha güvenilir biçimde ayrıştırılmasına olanak sağlamaktadır.

Sonuç olarak bu bölümde, FV sistem modelinde ele alınan normal çalışma ve arıza senaryoları açıklanmış; veri setinin hangi değişkenler ve sınıf etiketleri üzerinden oluşturulduğu ortaya konulmuştur. Bir sonraki alt bölümde, elde edilen ham simülasyon verilerinden sınıflandırmaya uygun özelliklerin nasıl türetildiği ve veri ön işleme sürecinin nasıl yürütüldüğü açıklanmaktadır.

### **3.3. Özellik Mühendisliği ve Veri Ön İşleme**

FV sistemlerden elde edilen ham ölçüm verileri arıza sınıflandırması için doğrudan bilgi içermekle birlikte, bu verilerin sınıflandırıcıya uygun hale getirilebilmesi için özellik mühendisliği ve veri ön işleme adımlarından geçirilmesi gerekmektedir. Modern FV sistemler yüksek boyutlu, doğrusal olmayan ve gürültü içeren veri yapıları üretebildiğinden, anlamlı arıza özelliklerinin çıkarılması sınıflandırma başarısı açısından önemli bir aşamadır (Mellit vd., 2018; Sepúlveda-Oviedo vd., 2023; El-Banby vd., 2023).

Bu çalışmada özellik mühendisliği süreci, FV sistemin elektriksel davranışını temsil eden akım, gerilim, güç ve çevresel değişkenlerden ayırt edici özelliklerin elde edilmesi amacıyla yürütülmüştür. Özellik seçimi yalnızca sayısal çeşitliliğe göre değil, arıza mekanizmalarını temsil etme gücüne göre yapılmıştır. Bu kapsamda string akımları, toplam akım, DC gerilim, DC güç, güneş ışınımı ve panel sıcaklığı temel değişkenler olarak dikkate alınmıştır.

Veri setinde doğrudan ölçülen veya simülasyon çıktılarından elde edilen elektriksel büyüklüklerin yanında, bu büyüklüklerden türetilmiş istatistiksel özellikler de kullanılmıştır. Özellikle string akımlarına ilişkin ortalama, maksimum, minimum ve varyans temelli değişkenler, arıza etkilerinin akım dağılımı üzerindeki yansımalarını temsil etmektedir (Ghoneim vd., 2021). Toplam akım, ortalama DC gerilim, maksimum ve minimum DC gerilim,

ortalama DC güç, ışıınım ve sıcaklık deęişkenleri ise sistemin genel işleme durumunu açıklayan temel özellikler olarak deęerlendirilmiştir.

FV sistemden elde edilen başlıca özellikler Tablo 3.3'te verilmiştir.

**Tablo 3.3.** FV sistemden elde edilen özellikler

Özellik	Açıklama
I1 mean	String 1 ortalama akım
I1 max	String 1 maksimum akım
I1 min	String 1 minimum akım
I1 var	String 1 varyans
I2 mean	String 2 ortalama akım
Vdc mean	Ortalama DC gerilim
Pdc mean	Ortalama DC güç
Temperature	Panel sıcaklığı
Irradiance	Güneş ışıınımı
Range1	Akım aralık özellięi

Tablo 3.3.'te görüldüğü üzere, kullanılan özellikler yalnızca sayısal büyüklükler olarak deęil, sistem davranışını temsil eden mühendislik göstergeleri olarak ele alınmıştır. Örneęin string akımı temelli özellikler, paralel kollar arasındaki akım paylaşımını ve olası dengesizlikleri göstermektedir. DC gerilim ve DC güç deęişkenleri sistemin çalışma noktasındaki deęişimleri ifade ederken, ışıınım ve sıcaklık deęişkenleri çevresel koşulların sistem çıktısı üzerindeki etkisini temsil etmektedir.

Bu çalışmada kullanılan özellikler arasında **I1, I2, I3, I4, I5, I6, Itotal, Vdcmean, Vdcmax, Vdcmin, Pdcmean, IR** ve **T** gibi doğrudan ölçüm veya simülasyon çıktısı niteliğindeki deęişkenler bulunmaktadır. Ayrıca string akımlarına ve toplam akıma ait maksimum, minimum ve varyans temelli türev özellikler de veri setine dahil edilmiştir. Bu özellikler, arıza durumlarında yalnızca anlık deęişimleri deęil, belirli bir zaman aralığındaki dağılım ve deęişkenlik yapısını da temsil etmektedir.

Özellik mühendisliği sürecinde özellikle range tabanlı deęişkenler arıza sınıflarının ayrıştırılması açısından önemli görülmüştür. **Range1, Range2, Range3** ve **Range4** gibi deęişkenler, belirli akım veya sinyal grupları arasındaki farkları ve deęişim aralıklarını temsil etmektedir. Bu tür özelliklerin PV arıza sınıflandırmasında ayırt edici bilgi taşıyabildiği önceki çalışmalarda da gösterilmiştir (Ghoneim vd., 2021). Çünkü bazı arıza türleri toplam güç

üzerinde benzer etkiler oluştursa da, stringler arası akım farkları ve dağılım yapıları bakımından farklı davranışlar sergileyebilmektedir.

Veri ön işleme sürecinde ilk olarak veri setindeki özelliklerin sayısal bütünlüğü kontrol edilmiştir. Bu aşamada eksik değerler, veri tipi uyumsuzlukları ve sınıf etiketlerinin tutarlılığı değerlendirilmiştir. Makine öğrenmesi tabanlı sınıflandırma modellerinin güvenilir biçimde eğitilebilmesi için her bir örneğin aynı özellik yapısına sahip olması ve sınıf etiketlerinin doğru biçimde tanımlanması gerekmektedir. Bu nedenle veri seti, özellik matrisi ve hedef sınıf etiketi olacak şekilde düzenlenmiştir.

Özelliklerin farklı fiziksel birimlere ve farklı büyüklük ölçeklerine sahip olması, veri ön işleme sürecinde dikkate alınması gereken önemli bir konudur. Akım, gerilim, güç, sıcaklık ve ışınım değişkenleri farklı aralıklarda değer aldığından, bu değişkenlerin doğrudan aynı model içerisinde kullanılması bazı özelliklerin sınıflandırıcı üzerinde baskın hale gelmesine neden olabilir. Örneğin ışınım değeri yüzler mertebesinde değişirken, bazı range tabanlı özellikler daha düşük aralıklarda değer alabilmektedir. Bu nedenle özelliklerin karşılaştırılabilir hale getirilmesi için ölçekleme işlemi uygulanmıştır.

Bu çalışmada veri seti, boyut azaltma ve sınıflandırma modellerine verilmeden önce standardizasyon işleminden geçirilmiştir. Standardizasyon, her bir özelliğin ortalamasının sıfıra ve standart sapmasının bire yakın olacak şekilde dönüştürülmesini sağlar. Böylece farklı büyüklük aralıklarına sahip değişkenlerin model üzerindeki etkileri daha dengeli hale getirilir. Bu işlem özellikle PCA, LDA, ICA ve SVM gibi uzaklık, varyans veya projeksiyon temelli yöntemlerde önem taşımaktadır (Jolliffe, 2002; Bishop, 2006).

Veri ön işleme sürecinde eğitim ve test verilerinin ayrı ele alınması da önemlidir. Ölçekleme katsayılarının yalnızca eğitim verisi üzerinden öğrenilmesi, test verisinin ise bu katsayılar kullanılarak dönüştürülmesi gerekmektedir. Bu yaklaşım, test verisine ait bilginin eğitim sürecine sızmasını önleyerek model değerlendirmesinin daha güvenilir yapılmasına katkı sağlar. Böylece sınıflandırma performansı, modelin daha önce görmediği veriler üzerindeki başarısını daha doğru biçimde yansıtır.

FV sistemden elde edilen özelliklerin değer aralıkları Tablo 3.4.'te sunulmuştur.

**Tablo 3.4.** FV sistemden elde edilen deęer aralıkları

Özellik	Min	Max	Ortalama	Varyans	Sonsuz Kazanç	GINI
I1	-99.26	5.66	2.17	3.11	0.015	0.011
I2	-99.26	5.89	2.70	2.25	0.018	0.012
I1max	0.60	5.75	3.17	0.46	0.006	0.004
I1min	-106.20	5.65	2.07	3.46	0.015	0.011
I1VAR	0.00	2.79	0.01	14.52	0.028	0.017
I2max	0.60	5.91	3.20	0.46	0.007	0.004
I2min	-106.2	5.76	2.63	2.46	0.019	0.013
I2VAR	0.00	2.79	0.01	17.05	0.053	0.032
I3	0.60	5.84	3.17	0.46	0.007	0.004
I4	0.60	5.69	3.09	0.46	0.009	0.006
I3max	0.60	5.87	3.19	0.46	0.006	0.004
I3min	0.60	5.76	3.15	0.46	0.007	0.004
I3VAR	0,00	0,00	0,00	1,69	0,01	0.008
I4max	0.60	5.76	3.18	0.46	0.006	0.004
I4min	0.60	5.66	3.08	0.46	0.009	0.006
I5	0.60	5.69	3.15	0.46	0.007	0.004
I6	0.60	5.69	3.15	0.46	0.007	0.004
Itotal	425.97	519.04	436.04	0.05	0.005	0.003
Itotalmax	456.22	519.84	461.39	0.03	0.006	0.004
Itotalmin	393.08	517.67	410.31	0.08	0.004	0.002
Vdcmean	473.80	529.35	508.33	0.02	0.012	0.009
Vdcmax	475.94	533.49	510.43	0.02	0.005	0.003
Vdcmin	473.43	517.83	502.35	0.02	0.004	0.003
Pdcmean	25.49	263.57	141.32	0.46	0.007	0.004
IR	106.00	1000.00	552.20	0.46	0.007	0.004
T	10.00	35.00	22.12	0.34	0.005	0.004
Range1	0.00	111.47	1.11	6.48	0.349	0.204
Range2	0,00	111,47	0,57	11,22	0.399	0.269
Range3	-66.59	0.00	-0.53	-5.90	0.699	0.497
Range4	0,00	3,61	0,08	3,43	0.676	0.433

Tablo 3.4, kullanılan özelliklerin minimum, maksimum, ortalama ve deęişkenlik yapıları bakımından farklı ölçeklerde yer aldığını göstermektedir. Toplam akım, DC gerilim,

DC güç, ışı nım ve sıcaklık de ğiřkenleri farklı fiziksel büyüklüklere karşılık geldi ğinden, bu de ğiřkenlerin doğrudan karşılaştırılması uygun de ğildir. Bu nedenle standardizasyon işle mi, boyut azaltma ve sınıflandırma sürecinin dengeli yürütülmesi açısından gerekli bir ön işleme adı mıdır.

Tablo 3.4'te yer alan de ğer aralıkları, bazı özelliklerin arıza sınıfları açısından daha fazla ayırt edici bilgi taşıyabilece ğini de göstermektedir. Özellikle range tabanlı de ğiřkenlerin da ğılım yapısı, stringler arası akım dengesizli ği ve arıza kaynaklı de ğiřimlerin belirlenmesinde yararlı olabilir. Buna karşılık ışı nım ve sıcaklık gibi çevresel de ğiřkenler, sistem davranışını etkileyen ancak doğrudan arıza etiketi olarak yorumlanmaması gereken işletme koşulu bilgileri sunmaktadır.

Özellik mühendisli ği ve veri ön işleme aşamalarının birlikte yürütülmesi, modelleme sürecinin güvenilirli ği açısından önemlidir. Özellik mühendisli ği, ham simülasyon çıktılarından arıza mekanizmalarını temsil eden anlamlı de ğiřkenlerin elde edilmesini sağlarken; veri ön işleme, bu de ğiřkenlerin makine öğrenmesi modelleri tarafından dengeli biçimde kullanılmasına olanak tanımaktadır. Böylece oluşturulan özellik uzayı, hem fiziksel olarak yorumlanabilir hem de sınıflandırma algoritmaları açısından kullanılabilir hale getirilmektedir.

Sonuç olarak bu bölümde, FV sistemden elde edilen ham simülasyon verilerinin sınıflandırmaya uygun bir özellik uzayına dönüřtürülmesi için izlenen özellik mühendisli ği ve veri ön işleme yaklaşımı açıklanmıştır. Kullanılan özellikler, sistemin fiziksel davranışını temsil etmekte ve arıza sınıfları arasındaki ayırt edici farkların görünür hale getirilmesine katkı sağlamaktadır. Bir sonraki bölümde, bu özellik uzayı üzerinde boyut azaltma yöntemlerinin nasıl uygulandı ğı ve sınıflandırma modellerinin nasıl eğitildi ği açıklanmaktadır.

### 3.4. Boyut Azaltma ve Model Eğitimi Süreci

Bu çalışmada oluşturulan özellik uzayı, boyut azaltma yöntemleri ve sınıflandırma algoritmaları kullanılarak FV arıza sınıflandırma problemine uygulanmıştır. Modelleme sürecinin temel amacı, FV sistemden elde edilen çok de ğiřkenli ölçüm verilerini daha düşük boyutlu ve ayırt edici bir temsil uzayına dönüřtürmek, ardından bu temsil uzayı üzerinde arıza sınıflarını makine öğrenmesi tabanlı yöntemlerle sınıflandırmaktır.

Model eğitimi sürecinde veri seti öncelikle **özellik matrisi** ve **hedef sınıf etiketi** olacak şekilde düzenlenmiştir. Özellik matrisi; akım, gerilim, güç, ışı nım, sıcaklık ve bunlardan türetilmiş istatistiksel de ğiřkenlerden oluşmaktadır. Hedef de ğiřken ise her bir örne ğin ait oldu ğu sınıfı göstermektedir. Bu sınıflar; normal çalışma, string arızası, string-toprak arızası ve

string-string arızası olarak tanımlanmıştır. Bu sınıf yapısı, 250 kW gücünde şebeke bağlantılı FV sistem üzerinde normal durum ve üç temel DC tarafı arızasını ele alan simülasyon tabanlı çalışmalarla uyumludur (Ghoneim vd., 2021).

Veri seti, model başarısının güvenilir biçimde değerlendirilebilmesi için eğitim ve test verisi olarak ayrılmıştır. Eğitim verisi, boyut azaltma yöntemlerinin ve sınıflandırma modellerinin öğrenilmesi amacıyla kullanılmış; test verisi ise modelin daha önce görmediği örnekler üzerindeki başarısını değerlendirmek için ayrılmıştır. Bu yaklaşım, modelin yalnızca eğitim verisini ezberleyip ezberlemediğini anlamak ve genelleme performansını daha sağlıklı değerlendirmek açısından önemlidir.

Boyut azaltma ve sınıflandırma sürecinden önce veri setine standardizasyon işlemi uygulanmıştır. Akım, gerilim, güç, ısıtım ve sıcaklık gibi değişkenler farklı ölçüm birimlerine ve sayısal aralıklara sahip olduğundan, bu değişkenlerin doğrudan modele verilmesi bazı özelliklerin sınıflandırma sürecinde baskın hale gelmesine neden olabilir. Bu nedenle her bir özellik, ortalaması sıfıra ve standart sapması birine yakın olacak şekilde dönüştürülmüştür.

Bu aşamada veri sızıntısını önlemek amacıyla standardizasyon katsayıları yalnızca eğitim verisi üzerinden hesaplanmıştır. Test verisi ise eğitim verisinden elde edilen ölçekleme katsayıları kullanılarak dönüştürülmüştür. Böylece test verisine ait istatistiksel bilginin eğitim sürecine dahil edilmesi engellenmiş ve model değerlendirmesinin daha güvenilir yapılması sağlanmıştır.

Standardizasyon sonrasında veri setine doğrusal ve çekirdek tabanlı boyut azaltma yöntemleri uygulanmıştır. Bu kapsamda doğrusal yöntemler olarak Temel Bileşen Analizi, Doğrusal Ayırma Analizi ve Bağımsız Bileşen Analizi; çekirdek tabanlı doğrusal olmayan yöntemler olarak ise Çekirdek Temel Bileşen Analizi, Çekirdek Doğrusal Ayırma Analizi ve KPCA-ICA yaklaşımı değerlendirilmiştir. KPCA-ICA yaklaşımı, bu çalışmada KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde gerçekleştirilmiştir. Böylece PCA-KPCA, LDA-KLDA ve ICA-KPCA-ICA yöntem çiftleri aynı deneysel değerlendirme çerçevesi içerisinde karşılaştırılmıştır. Bu yöntemlerin her biri, veri setini farklı bir temsil mantığıyla daha düşük boyutlu bir uzaya taşımaktadır (Fisher, 1936; Hyvärinen vd., 2001; Jolliffe, 2002; Schölkopf vd., 1998; Baudat ve Anouar, 2000; Bach ve Jordan, 2002).

Boyut azaltma yöntemleri uygulanırken eğitim ve test ayrımı korunmuştur. Boyut azaltma dönüşümü yalnızca eğitim verisi üzerinde öğrenilmiş, ardından aynı dönüşüm test verisine uygulanmıştır. Bu yaklaşım, test verisine ait bilginin dönüşüm matrisi veya temsil

uzayı oluşturma sürecine dahil edilmesini önlemektedir. Aksi durumda tüm veri seti üzerinde öğrenilen bir dönüşüm, test verisine ait bilgiyi de içereceğinden performans sonuçlarının olduğundan daha yüksek görünmesine neden olabilir.

Boyut azaltma işlemi sonucunda elde edilen düşük boyutlu temsil uzayları, sınıflandırma algoritmalarına giriş olarak verilmiştir. Bu çalışmada sınıflandırma sürecinde farklı makine öğrenmesi algoritmaları denenmiş ve her boyut azaltma yöntemi için en iyi performansı sağlayan sınıflandırıcı sonuçları raporlanmıştır. Literatürde FV arıza sınıflandırması için Random Forest, Logistic Regression, Naive Bayes, AdaBoost ve CN2 gibi farklı sınıflandırıcıların karşılaştırıldığı çalışmalar bulunmaktadır (Ghoneim vd., 2021). Bu tezde ise sınıflandırıcı karşılaştırmasının ötesinde, asıl odak farklı boyut azaltma yöntemlerinin sınıflandırıcıya sunulan veri temsilini nasıl etkilediğini incelemektir.

Modelleme sürecinde kullanılan genel işlem zinciri; veri setinin özellik matrisi ve sınıf etiketi olarak düzenlenmesi, eğitim-test ayrımının yapılması, standardizasyon işleminin uygulanması, boyut azaltma yöntemlerinin eğitim verisi üzerinde öğrenilmesi, aynı dönüşümün test verisine uygulanması, sınıflandırma modellerinin eğitilmesi ve sonuçların performans ölçütleriyle değerlendirilmesi adımlarından oluşmaktadır. Bu işlem zinciri, istatistiksel öğrenme ve makine öğrenmesi uygulamalarında kullanılan genel modelleme akışıyla uyumludur (Hastie vd., 2009; Pedregosa vd., 2011).

Model performansının değerlendirilmesinde doğruluk, kesinlik, duyarlılık, F1 skoru, özgüllük, AUC değeri, karışıklık matrisi ve işlem süresi dikkate alınmıştır. Bu ölçütler, FV arıza sınıflandırması çalışmalarında sınıflandırıcı başarısını çok yönlü değerlendirmek amacıyla yaygın olarak kullanılmaktadır (Ghoneim vd., 2021; Fawcett, 2006; Sokolova ve Lapalme, 2009). Doğruluk genel sınıflandırma başarısını gösterirken, kesinlik ve duyarlılık sınıf bazlı hata davranışlarının incelenmesine olanak sağlamaktadır. F1 skoru, kesinlik ve duyarlılık arasındaki dengeyi gösterdiğinden özellikle sınıflar arasında tam dengeli olmayan veri dağılımlarında önemlidir.

Karışıklık matrisi, hangi sınıfların doğru sınıflandırıldığını ve hangi sınıfların birbirine karıştırıldığını göstermesi açısından önemlidir. FV arıza sınıflandırmasında bazı arıza türlerinin benzer elektriksel etkiler oluşturması mümkündür. Bu nedenle karışıklık matrisi, yalnızca sayısal performans değerlendirmesi değil, aynı zamanda mühendislik yorumu açısından da yararlı bir araçtır.

ROC eğrileri ve AUC değerleri, sınıflandırıcıların sınıflar arası ayırt etme kapasitesini değerlendirmek amacıyla kullanılmıştır. Çok sınıflı sınıflandırma probleminde bu değerlendirme, her sınıfın diğer sınıflara karşı ayrıştırılabilirliği üzerinden yapılabilmektedir. Böylece yalnızca genel model başarısı değil, her bir arıza sınıfının ayırt edilebilirlik düzeyi de incelenmiştir.

İşlem süresi ise modelin pratik uygulanabilirliği açısından dikkate alınmıştır. FV arıza teşhisi, özellikle çevrim içi izleme sistemlerinde hızlı karar üretmeyi gerektirebilir. Bu nedenle yüksek doğruluk sağlayan ancak işlem süresi çok yüksek olan bir yöntemin saha uygulamalarındaki kullanımı sınırlı olabilir. Bu çalışmada boyut azaltma yöntemleri değerlendirilirken sınıflandırma başarısının yanında hesaplama maliyeti de göz önünde bulundurulmuştur.

Sınıflandırıcıların karşılaştırılmasında aynı eğitim-test bölünmesi ve aynı performans ölçütleri kullanılmıştır. Kullanılan sınıflandırıcılar, hiperparametre seçimleri, çekirdek fonksiyonu türü, çekirdek parametreleri ve boyut azaltma sonrasında kullanılan bileşen sayıları tüm yöntemler için sabit bir değerlendirme protokolü altında belirlenmiştir. Bu yaklaşım, farklı boyut azaltma tekniklerinin sınıflandırma performansı üzerindeki etkisinin daha adil biçimde karşılaştırılmasını sağlamaktadır.

Sonuç olarak bu bölümde, FV arıza sınıflandırması için izlenen boyut azaltma ve model eğitimi süreci açıklanmıştır. Bu süreçte amaç, yalnızca sınıflandırma algoritmalarının başarısını karşılaştırmak değil, farklı boyut azaltma tekniklerinin veri temsilini nasıl etkilediğini ve bu etkinin arıza sınıflandırma performansına nasıl yansıdığını ortaya koymaktır. Bir sonraki bölümde, uygulanan analizler sonucunda elde edilen bulgular ve performans sonuçları değerlendirilmektedir.

Boyut azaltma yöntemleri ve sınıflandırıcı kombinasyonları sonucunda elde edilen modellerin karşılaştırılabilmesi için doğruluk, AUC, kesinlik, duyarlılık, F1 skoru, özgüllük, karışıklık matrisi ve işlem süresi gibi performans ölçütleri kullanılmıştır. Bu ölçütlerin tanımları ve çalışma kapsamında kullanım amaçları bir sonraki bölümde açıklanmaktadır.

### **3.5. Performans Değerlendirme Ölçütleri**

Bu çalışmada, boyut azaltma yöntemleri ile sınıflandırıcı algoritmaların oluşturduğu model kombinasyonlarının başarısını karşılaştırmalı olarak değerlendirmek amacıyla birden fazla performans ölçütü kullanılmıştır. FV arıza sınıflandırma problemi çok sınıflı bir yapıya sahip olduğundan, model performansının yalnızca genel doğruluk değeri ile değerlendirilmesi

yeterli görülmemiştir. Bu nedenle doğruluk, kesinlik, duyarlılık, F1 skoru, özgüllük, AUC değeri, karışıklık matrisi ve işlem süresi birlikte dikkate alınmıştır.

Sınıflandırma performansının hesaplanmasında doğru pozitif (TP), doğru negatif (TN), yanlış pozitif (FP) ve yanlış negatif (FN) değerleri temel alınmıştır. Çok sınıflı sınıflandırma probleminde bu değerler, her bir sınıf için ilgili sınıf pozitif, diğer sınıflar ise negatif kabul edilerek hesaplanmıştır. Böylece her arıza sınıfının model tarafından ne ölçüde doğru ayrıştırılabildiği ayrı ayrı değerlendirilebilmiştir.

Doğruluk, modelin tüm örnekler içerisindeki doğru tahmin oranını göstermektedir. Bu ölçüt, genel sınıflandırma başarısı hakkında temel bir bilgi sağlamakla birlikte, sınıf dağılımının dengesiz olduğu durumlarda tek başına yeterli olmayabilir. Doğruluk değeri D.21’de verilmiştir.

$$\mathbf{Doğruluk} = (TP + TN) / (TP + TN + FP + FN) \quad \mathbf{(D.21)}$$

Kesinlik, modelin pozitif olarak tahmin ettiği örnekler içerisinde gerçekten pozitif olanların oranını ifade etmektedir. Bu ölçüt, özellikle yanlış pozitif tahminlerin azaltılmasının önemli olduğu durumlarda dikkate alınmaktadır. Kesinlik değeri D.22’de verilmiştir.

$$\mathbf{Kesinlik} = TP / (TP + FP) \quad \mathbf{(D.22)}$$

Duyarlılık, gerçek pozitif örneklerin model tarafından ne kadarının doğru biçimde tespit edildiğini göstermektedir. Arıza tespit problemlerinde duyarlılık önemli bir ölçüttür; çünkü gerçek arıza durumlarının normal olarak sınıflandırılması bakım ve güvenlik açısından istenmeyen sonuçlar doğurabilir. Duyarlılık değeri D.23’te verilmiştir.

$$\mathbf{Duyarlılık} = TP / (TP + FN) \quad \mathbf{(D.23)}$$

F1 skoru, kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır. Bu ölçüt, kesinlik ve duyarlılık arasında denge kurarak modelin sınıf bazlı başarısını daha dengeli biçimde değerlendirmeye olanak sağlar. Özellikle sınıflar arası örnek sayılarının farklı olduğu problemlerde F1 skoru, doğruluk değerine göre daha açıklayıcı bir ölçüt olabilmektedir. F1 skoru D.24’te verilmiştir.

$$F1 \text{ Skoru} = 2 \times (Kesinlik \times Duyarluluk) / (Kesinlik + Duyarluluk) \quad (D.24)$$

Özgüllük, gerçek negatif örneklerin model tarafından doğru biçimde negatif olarak sınıflandırılma oranını ifade etmektedir. Bu ölçüt, modelin ilgili sınıfa ait olmayan örnekleri ayırt etme başarısını değerlendirmek için kullanılmıştır. Özgüllük değeri D.25'te verilmiştir.

$$\text{Özgüllük} = TN / (TN + FP) \quad (D.25)$$

AUC, ROC eğrisi altında kalan alanı ifade etmektedir ve modelin sınıflar arasındaki ayırt edicilik düzeyini değerlendirmek için kullanılmıştır. AUC değerinin yüksek olması, modelin ilgili sınıfa ait örnekleri diğer sınıflardan daha başarılı biçimde ayırabildiğini göstermektedir. Bu çalışmada AUC değeri, farklı boyut azaltma yöntemleri sonucunda elde edilen temsil uzaylarının sınıflar arası ayrımı ne ölçüde güçlendirdiğini değerlendirmek amacıyla dikkate alınmıştır.

Karışıklık matrisi, modelin hangi sınıfları doğru, hangi sınıfları ise hatalı tahmin ettiğini ayrıntılı biçimde göstermektedir. Bu nedenle karışıklık matrisi yalnızca genel başarıyı değil, sınıflar arası hata dağılımını değerlendirmek amacıyla da kullanılmıştır. FV arıza sınıflandırma probleminde özellikle string arızası, string-toprak arızası ve string-string arızası gibi elektriksel olarak kısmen benzer davranışlar gösterebilen sınıflar arasındaki karışmaların incelenmesi açısından karışıklık matrisi önemli bir değerlendirme aracıdır.

İşlem süresi ise boyut azaltma ve sınıflandırma adımlarının hesaplama maliyetini değerlendirmek amacıyla kullanılmıştır. Bu ölçüt, özellikle gerçek zamanlı veya çevrim içi izleme sistemlerine yönelik uygulamalarda önem taşımaktadır. Yüksek sınıflandırma başarısı sağlayan bir yöntemin işlem süresi bakımından da uygulanabilir olması, mühendislik uygulamaları açısından dikkate alınması gereken bir kriterdir.

Bu çalışmada performans değerlendirmesi tek bir ölçüte dayandırılmamış; doğruluk, AUC, F1 skoru, kesinlik, duyarlılık, özgüllük, karışıklık matrisi ve işlem süresi birlikte değerlendirilmiştir. Böylece boyut azaltma yöntemlerinin yalnızca genel sınıflandırma başarısı değil, aynı zamanda sınıflar arası ayırt edicilik, hata dağılımı, dengeli performans ve hesaplama maliyeti açısından etkileri karşılaştırmalı olarak incelenmiştir.

### 3.6. Bozulmuş Senaryoları Altında Deney Protokolü

Bu çalışmada boyut azaltma yöntemlerinin yalnızca temiz veri üzerindeki sınıflandırma başarısı değil, saha benzeri bozulmuş koşulları altındaki dayanıklılığı da değerlendirilmiştir. Bu

amaçla model eğitimi temiz eğitim verisi üzerinde gerçekleştirilmiş, bozulmuş senaryoları yalnızca test verisine uygulanmıştır. Böylece modelin eğitim aşamasında bozulmuş veriyi görmeden; gürültü, eksik veri, sensör kaybı ve çevresel değişkenlik gibi gerçek uygulamalarda karşılaşılabilecek durumlar altındaki genelleme performansı analiz edilmiştir. Bu yaklaşım, veri sızıntısını önlemek ve yöntemlerin gerçekçi test koşullarındaki dayanıklılığını ölçmek amacıyla tercih edilmiştir.

Deney protokolünde farklı bozulmuş ve çalışma koşulu senaryoları oluşturulmuştur. Gaussian gürültüsü senaryolarında test verisine %1, %3, %5 ve %10 düzeylerinde gürültü eklenmiştir. Rastgele eksik veri senaryolarında özelliklerin %5, %10, %20 ve %30 oranlarında eksik bırakılması değerlendirilmiştir. Blok eksik veri senaryolarında ise 5, 10 ve 20 uzunluklarında ardışık özellik bloklarının çıkarılması uygulanmıştır. Bu senaryolar, ölçüm gürültüsü, haberleşme kesintisi veya veri kayıt hatası gibi durumların sınıflandırma başarısı üzerindeki etkisini incelemek amacıyla oluşturulmuştur.

Sensör sürüklenmesi senaryolarında test verisi üzerinde 0.01, 0.02 ve 0.05 düzeylerinde sistematik kayma oluşturulmuştur. Bu senaryo, zamanla kalibrasyon bozulması veya ölçüm cihazı sapması gibi durumların model performansına etkisini değerlendirmek için kullanılmıştır. Tek sensör kaybı senaryosunda ise ışınım değişkeni veri setinden çıkarılarak, kritik bir çevresel ölçüm bilgisinin bulunmadığı durumda yöntemlerin sınıflandırma başarısı incelenmiştir.

Çevresel çalışma koşullarının etkisini değerlendirmek amacıyla test verisi ayrıca düşük, orta ve yüksek ışınım bantlarına ayrılmıştır. Bu sayede boyut azaltma yöntemlerinin yalnızca genel test başarımı değil, farklı ışınım seviyeleri altındaki kararlılığı da karşılaştırılmıştır. FV sistemlerde ışınım düzeyi akım ve güç üretimini doğrudan etkilediğinden, bu ayırım yöntemlerin çevresel değişkenlik karşısındaki dayanıklılığını değerlendirmek açısından önemlidir.

Bu deney protokolü, temiz veri başarımı ile bozulmuş altındaki dayanıklılık başarımının ayrı ayrı değerlendirilmesine olanak sağlamaktadır. Böylece PCA, LDA, ICA, KPCA, KLDA ve KPCA-ICA yöntemlerinin yalnızca ideal koşullardaki sınıflandırma performansı değil, saha benzeri bozulmuş test koşulları altındaki kararlılıkları da karşılaştırmalı olarak incelenmiştir. Bu yaklaşım, FV arıza sınıflandırmasında uygun boyut azaltma yönteminin seçimini yalnızca doğruluk başarısına değil, aynı zamanda yöntemlerin farklı işletme ve bozulmuş koşulları altındaki güvenilirliğine dayandırmayı mümkün kılmaktadır.

## 4. BULGULAR

Bu bölümde, FV arıza sınıflandırması kapsamında gerçekleştirilen veri ön işleme, özellik uzayı analizi, boyut azaltma uygulamaları ve sınıflandırma performansına ilişkin bulgular sunulmaktadır. Çalışmada normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört farklı sınıf ele alınmıştır. Bu sınıflara ait ölçüm verileri önce özellik mühendisliği ve veri ön işleme adımlarından geçirilmiş, ardından doğrusal ve çekirdek tabanlı boyut azaltma yöntemleriyle daha düşük boyutlu temsil uzaylarına dönüştürülmüştür.

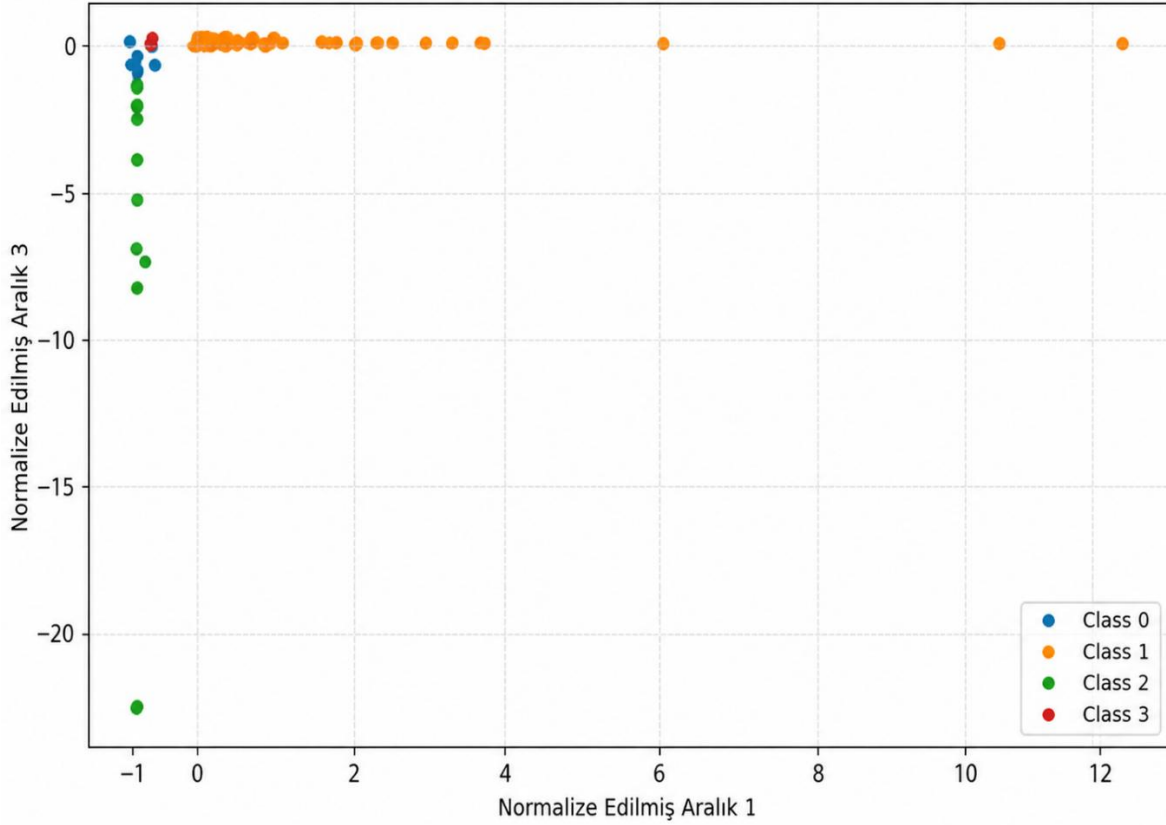
Bulguların değerlendirilmesinde yalnızca sınıflandırma doğruluğu değil; özellik uzayının yapısı, sınıflar arası örtüşme durumu, değişkenler arasındaki korelasyon ilişkileri, boyut azaltma yöntemlerinin temsil gücü, dayanıklılık performansı ve mühendislik açısından uygulanabilirlik de dikkate alınmıştır. Böylece elde edilen sonuçlar hem sayısal performans hem de FV sistem davranışının fiziksel yorumu açısından değerlendirilmiştir.

### 4.1. Veri Ön İşleme ve Özellik Uzayı Analizi

Veri ön işleme ve özellik uzayı analizi, sınıflandırma sürecinden önce veri setinin genel yapısını anlamak amacıyla gerçekleştirilmiştir. FV sistemden elde edilen veri seti; string akımları, toplam akım, DC gerilim, DC güç, ışıınım, sıcaklık ve bu büyüklüklerden türetilmiş istatistiksel özelliklerden oluşmaktadır. Bu özellikler farklı fiziksel büyüklüklere ve farklı sayısal aralıklara sahip olduğundan, doğrudan sınıflandırıcıya verilmeden önce ortak bir ölçeğe taşınmaları gerekmektedir.

Bu amaçla veri setine normalizasyon/standardizasyon işlemi uygulanmıştır. Ölçekleme işlemi, farklı ölçüm birimlerine sahip özelliklerin model üzerinde daha dengeli biçimde etkili olmasını sağlamaktadır. Özellikle akım, gerilim, güç, ışıınım ve sıcaklık gibi değişkenler farklı değer aralıklarında bulunduğundan, ölçekleme yapılmadığında bazı özellikler yalnızca sayısal büyüklükleri nedeniyle sınıflandırma veya boyut azaltma sürecinde baskın hale gelebilir.

Normalize edilmiş özellik uzayında arıza sınıflarının genel dağılımı Şekil 4.1'de verilmiştir.



**Şekil 4.1.** Normalize edilmiş özellik uzayında arıza sınıflarının dağılımı

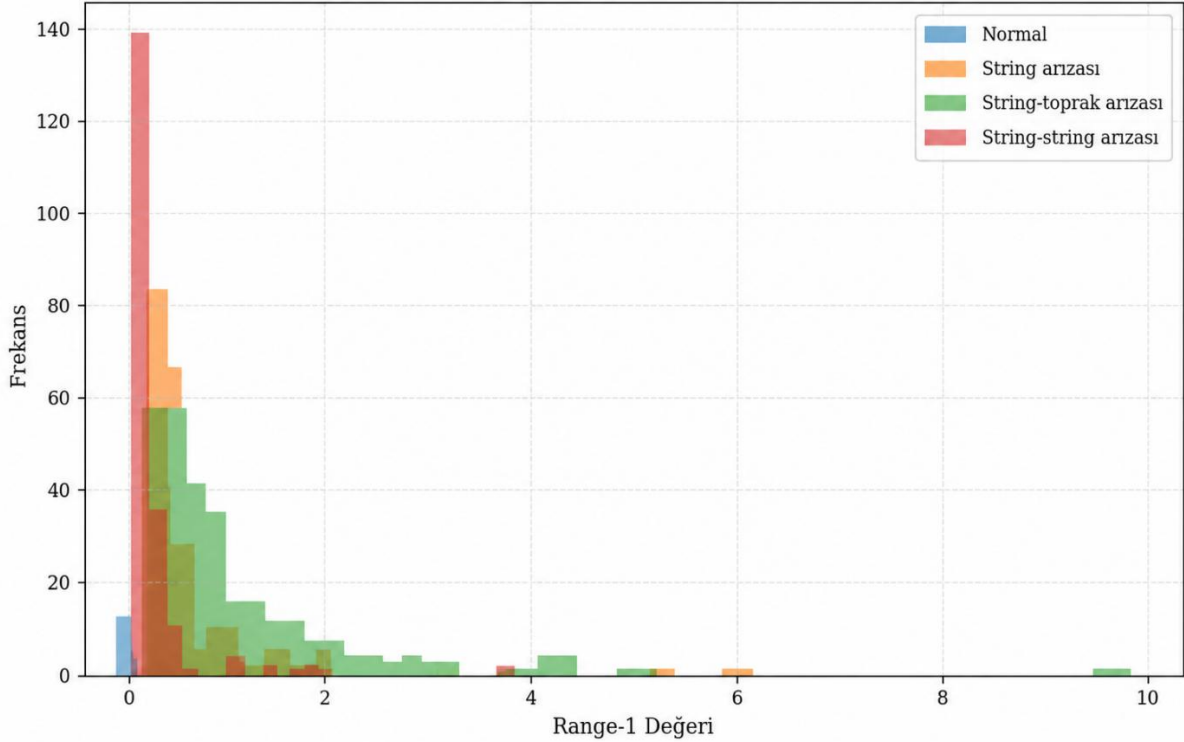
Şekil 4.1 incelendiğinde, normal çalışma, string arızası, string-toprak arızası ve string-string arızası sınıflarının normalize edilmiş özellik uzayında tamamen ayrık kümeler oluşturmadığı görülmektedir. Bazı sınıflar belirli bölgelerde kısmi ayrışma gösterse de genel dağılımda sınıflar arasında örtüşmeler bulunmaktadır. Bu durum, FV arıza sınıflandırma probleminin basit eşikleme veya tekil değişken karşılaştırması ile güvenilir biçimde çözülemeyeceğini göstermektedir.

Sınıflar arasındaki örtüşmenin temel nedenlerinden biri, FV sistem davranışının yalnızca arıza türüne değil; ışınım, sıcaklık, arıza direnci ve sistem çalışma noktası gibi değişkenlere de bağlı olmasıdır. Örneğin düşük ışınım koşullarında normal çalışma durumunda da akım ve güç değerleri azalabilmekte, bu davranış bazı arıza durumlarıyla benzerlik gösterebilmektedir. Benzer şekilde string arızası ve string-string arızası toplam güç düşüşü açısından yakın sonuçlar oluşturabilir; ancak string akımları ve dağılım özellikleri bakımından farklılaşabilir.

Veri ön işleme ve özellik uzayı analizinin ikinci aşamasında, tekil özelliklerin sınıf ayırma gücü incelenmiştir. Bu kapsamda özellikle range tabanlı değişkenler dikkate alınmıştır. Range türü özellikler, belirli sinyal grupları arasındaki değişim aralığını veya maksimum-

minimum farkını temsil ettiğinden, arıza kaynaklı dengesizlikleri görünür hale getirebilmektedir.

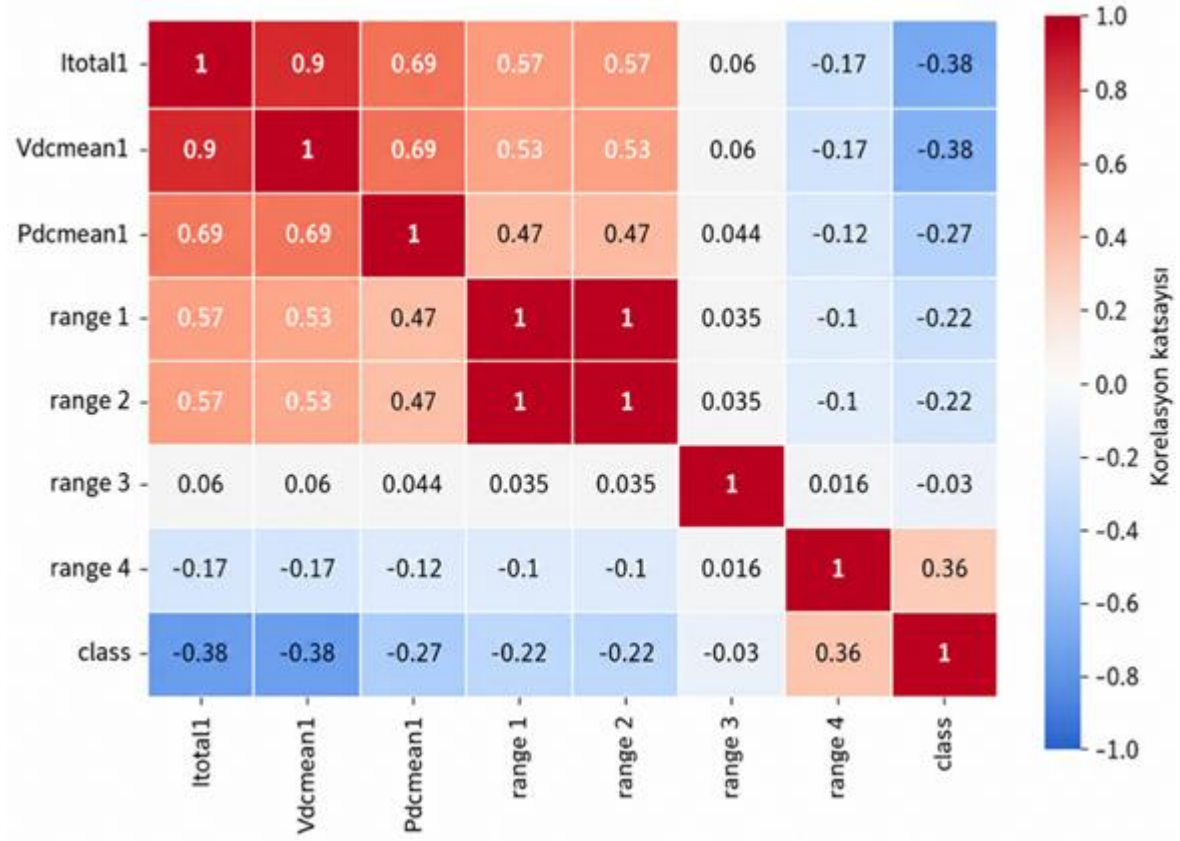
Range-1 özelliğinin arıza sınıflarına göre frekans dağılımı Şekil 4.2’de verilmiştir.



Şekil 4.2. Range-1 özelliğinin arıza sınıflarına göre frekans dağılımı

Şekil 4.2 incelendiğinde, Range-1 özelliğinin bazı arıza sınıfları için farklı dağılım karakteristikleri sergilediği görülmektedir. Bu durum, tekil bir özelliğin dahi arıza sınıfları hakkında belirli ölçüde bilgi taşıdığını göstermektedir. Ancak dağılımlar arasında tam bir ayrışma oluşmadığından, yalnızca Range-1 özelliğine dayalı bir sınıflandırma yaklaşımı yeterli değildir. Bu bulgu, çoklu özellik kullanımının ve bu özelliklerin birlikte değerlendirilmesinin gerekli olduğunu göstermektedir.

Seçilmiş özellikler arasındaki korelasyon ilişkileri Şekil 4.3’te verilmiştir.



**Şekil 4.3.** Seçilmiş özellikler arası korelasyon ısı haritası

Şekil 4.3'te verilen korelasyon ısı haritası, veri setindeki bazı özellik grupları arasında güçlü ilişkiler bulunduğunu göstermektedir. Özellikle aynı fiziksel büyüklükten türetilen özellikler veya doğrudan bağlantılı elektriksel değişkenler arasında yüksek korelasyon gözlenmesi beklenen bir durumdur. Örneğin string akımları ile toplam akım, DC gerilim ile güç veya maksimum-minimum tabanlı türetilmiş değişkenler arasında yapısal bağıntılar oluşabilmektedir.

Korelasyon analizinin iki önemli sonucu bulunmaktadır. Birincisi, veri setinde benzer bilgiyi taşıyan özelliklerin bulunması, boyut azaltma yöntemleriyle daha kompakt bir temsil elde edilebileceğini göstermektedir. İkincisi, korelasyon yapısı değişkenler arasındaki ilişkileri ortaya koysa da tek başına sınıf ayırımı sağlamamaktadır. Bu nedenle korelasyon analizi, boyut azaltma ihtiyacını destekleyen bir ön bulgu olarak değerlendirilmelidir.

Sonuç olarak normalize edilmiş dağılım grafiği, Range-1 özelliğinin sınıflara göre frekans dağılımı ve korelasyon ısı haritası birlikte değerlendirildiğinde, veri setinin çok boyutlu, kısmen örtüşen ve korelasyon içeren bir yapıya sahip olduğu görülmektedir. Bu nedenle bir sonraki bölümde, boyut azaltma yöntemlerinin bu yapıyı ne ölçüde daha ayırt edici bir temsil uzayına dönüştürebildiği karşılaştırmalı olarak incelenmektedir.

## 4.2. Boyut Azaltma Yöntemlerinin Karşılaştırılması

Bu bölümde, FV sistemden elde edilen özellik uzayı üzerinde uygulanan boyut azaltma yöntemleri karşılaştırmalı olarak değerlendirilmiştir. Çalışmada doğrusal yöntemler olarak Temel Bileşen Analizi, Doğrusal Ayırma Analizi ve Bağımsız Bileşen Analizi; çekirdek tabanlı doğrusal olmayan yöntemler olarak ise Çekirdek Temel Bileşen Analizi, Çekirdek Doğrusal Ayırma Analizi ve KPCA-ICA yaklaşımı ele alınmıştır. Karşılaştırma, boyut azaltma sonrasında elde edilen temsil uzayları üzerinden yapılmıştır.

Önceki bölümde sunulan özellik uzayı analizi, veri setinin çok boyutlu, kısmen örtüşen ve korelasyon içeren bir yapıya sahip olduğunu göstermiştir. Bu nedenle boyut azaltma yöntemlerinin başarısı yalnızca veri boyutunu düşürmeleriyle değil, sınıflar arası ayrımı ne ölçüde güçlendirdikleriyle değerlendirilmelidir. Bu amaçla her boyut azaltma yöntemi sonrasında farklı sınıflandırma algoritmaları uygulanmış ve en iyi performans üreten model sonuçları Tablo 4.1’de karşılaştırmalı olarak verilmiştir.

**Tablo 4.1.** Boyut azaltma yöntemlerine göre en iyi sınıflandırma performansları

Boyut azaltma yöntemi	En iyi sınıflandırıcı	Doğruluk	AUC	Kesinlik	Duyarlılık	F1 skoru	Özgüllük	İşlem süresi (s)
PCA	Lojistik Regresyon	0,7100	0,8160	0,6383	0,7100	0,6430	0,9033	0,0250
LDA	Naive Bayes	0,6700	0,7872	0,6952	0,6700	0,6377	0,8900	0,0100
ICA	Naive Bayes	0,6400	0,7509	0,6748	0,6400	0,6320	0,8800	0,0090
KPCA	L1-Lojistik	0,7100	0,9003	0,7272	0,7100	0,6960	0,9033	0,1762
KLDA	CN2 Kural	0,6500	0,7667	0,6284	0,6500	0,6247	0,8833	0,0100
KPCA-ICA	Rastgele Orman	0,5800	0,8471	0,6257	0,5800	0,5319	0,8600	0,2292

Tablo 4.1 incelendiğinde, en yüksek doğruluk değerinin PCA ve KPCA yöntemlerinde elde edildiği görülmektedir. Her iki yöntemde de doğruluk değeri **0,71** olarak bulunmuştur. Ancak bu iki yöntem F1 skoru ve AUC değeri bakımından farklılık göstermektedir. KPCA tabanlı L1-lojistik modelinde AUC değeri **0,9003** ve F1 skoru **0,6960** olarak elde edilirken, PCA tabanlı lojistik regresyon modelinde AUC değeri **0,8160** ve F1 skoru **0,6430** olarak

bulunmuştur. Bu sonuç, KPCA yönteminin sınıflar arası ayırt edicilik bakımından PCA'ya göre daha güçlü bir temsil uzayı sağladığını göstermektedir.

KPCA'nın daha yüksek AUC ve F1 skoru üretmesi, FV arıza verilerinde doğrusal olmayan ilişkilerin sınıflandırma performansı açısından önemli olduğunu göstermektedir. FV sistemlerde arıza etkileri; ışıınım, sıcaklık, arıza direnci, akım paylaşımı ve sistem çalışma noktası gibi değişkenlerle birlikte ortaya çıkmaktadır. Bu nedenle arıza sınıfları ham özellik uzayında veya doğrusal dönüşüm uzayında tamamen ayrılmayabilir. KPCA, çekirdek fonksiyonu yardımıyla doğrusal olmayan temsil uzayı oluşturduğu için bu tür karmaşık ilişkileri daha iyi temsil edebilmiştir.

PCA ise KPCA ile aynı doğruluk düzeyine ulaşmasına rağmen F1 skoru ve AUC değeri bakımından daha düşük sonuç üretmiştir. Buna karşılık PCA'nın doğrusal, sade ve yorumlanabilir yapısı mühendislik uygulamaları açısından önemli bir avantajdır. Bu nedenle PCA, sınıflandırma başarısı açısından KPCA'nın gerisinde kalsa da, yorumlanabilirlik ve uygulanabilirlik bakımından güçlü bir alternatif olarak değerlendirilebilir.

LDA yöntemi **0,67 doğruluk** ve **0,6377 F1 skoru** ile orta düzeyde bir performans göstermiştir. LDA'nın sınıf etiketlerini dikkate alan denetimli bir yöntem olması teorik olarak avantaj sağlasa da, bu veri setinde PCA ve KPCA kadar yüksek başarı üretmemiştir. Bu durum, FV arıza verilerinde sınıflar arası ayrımın yalnızca doğrusal sınıf ayrımı varsayımıyla tam olarak temsil edilemediğini göstermektedir.

ICA yöntemi **0,64 doğruluk** ve **0,6320 F1 skoru** ile LDA'ya yakın ancak daha sınırlı bir performans üretmiştir. ICA'nın temel amacı, veriyi istatistiksel olarak bağımsız bileşenlere ayırmaktır. Ancak FV sistemlerde arıza etkileri, çevresel koşullar ve güç elektroniği dinamikleri çoğu zaman birbirinden tamamen bağımsız değildir. Bu nedenle ICA ile elde edilen bileşenlerin sınıflandırma açısından en ayırt edici bileşenler olması garanti edilememektedir.

KLDA yöntemi **0,65 doğruluk** ve **0,6247 F1 skoru** ile orta düzeyde bir performans göstermiştir. KLDA, sınıf bilgisini çekirdek tabanlı temsil uzayında değerlendirmesi nedeniyle teorik olarak güçlü bir yöntemdir. Ancak bu çalışmada KPCA kadar başarılı bir temsil üretmemiştir. Bu durum, denetimli veya çekirdek tabanlı bir yöntemin her veri setinde otomatik olarak en yüksek performansı sağlamadığını göstermektedir.

KPCA-ICA yöntemi ise **0,58 doğruluk** ve **0,5319 F1 skoru** ile en düşük sınıflandırma performanslarından birini üretmiştir. Bununla birlikte KPCA-ICA'nın AUC değerinin **0,8471** olması, yöntemin bazı sınıflar için belirli bir ayırt edicilik taşıdığını; ancak bu ayırt ediciliğin

genel doğruluk ve F1 skoru bakımından dengeli bir sınıflandırma başarısına dönüşmediğini göstermektedir.

Tablo 4.1'deki sonuçlar genel olarak değerlendirildiğinde, KPCA'nın bu veri seti ve değerlendirme protokolü kapsamında F1 skoru ve AUC değeri bakımından en başarılı temsil yaklaşımı olduğu görülmektedir. PCA aynı doğruluk düzeyine ulaşmış, ancak KPCA'ya göre daha düşük F1 ve AUC değerleri üretmiştir. LDA, ICA ve KLDA orta düzeyde performans üretirken, KPCA-ICA bu veri setinde sınırlı sınıflandırma başarısı göstermiştir.

### 4.3. Sınıflandırma Performans Analizi

Bu bölümde, boyut azaltma yöntemleri sonrasında elde edilen temsil uzaylarının sınıflandırma başarısı ayrıntılı olarak değerlendirilmiştir. Analizde PCA, LDA, ICA, KPCA, KLDA ve KPCA-ICA yöntemleri kullanılmıştır. Her bir temsil uzayı üzerinde farklı sınıflandırıcılar denenmiş ve her yöntem için en iyi sonucu veren sınıflandırıcı Tablo 4.1'de sunulmuştur.

Sınıflandırma performansı değerlendirilirken yalnızca doğruluk değeri dikkate alınmamıştır. FV arıza sınıflandırması çok sınıflı bir problem olduğundan, model başarısının daha güvenilir biçimde yorumlanabilmesi için doğruluk, AUC, kesinlik, duyarlılık, F1 skoru, özgüllük ve işlem süresi birlikte değerlendirilmiştir (Fawcett, 2006; Sokolova ve Lapalme, 2009).

Tablo 4.1'deki sonuçlara göre KPCA ve PCA, **0,71 doğruluk değeri** ile en yüksek doğruluk düzeyine ulaşmıştır. Ancak KPCA'nın **0,9003 AUC** ve **0,6960 F1 skoru** üretmesi, bu yöntemin PCA'ya göre sınıflar arası ayırt edicilik ve dengeli tahmin başarısı açısından daha güçlü olduğunu göstermektedir. PCA'da ise **AUC 0,8160** ve **F1 skoru 0,6430** düzeyinde kalmıştır.

KPCA'nın yüksek AUC değeri, modelin farklı karar eşikleri altında arıza sınıflarını ayırt etme kapasitesinin güçlü olduğunu göstermektedir. Bu durum, FV sistem verilerinde doğrusal olmayan ilişkilerin sınıflandırma performansı açısından önemli olduğunu ortaya koymaktadır. Buna karşılık PCA, KPCA'ya göre daha düşük AUC ve F1 skoru üretmiş olsa da, doğrusal ve yorumlanabilir yapısı nedeniyle mühendislik açısından uygulanabilir bir alternatif sunmaktadır.

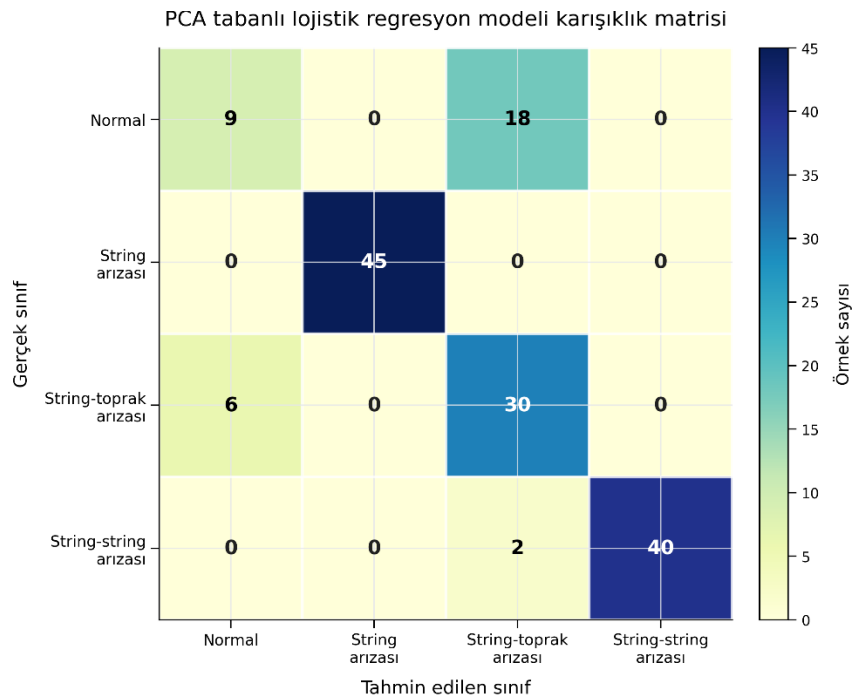
LDA, ICA ve KLDA yöntemleri orta düzeyde sonuçlar üretmiştir. LDA'nın denetimli yapısı sınıf ayrımını güçlendirmeyi hedeflese de, veri setindeki sınıf içi değişkenlik ve sınıflar

arası örtüşme bu yöntemin başarısını sınırlamış olabilir. ICA ise bağımsız bileşenlere odaklandığından, arıza etkileri ile çevresel değişkenlerin birlikte gözlemlendiği bu veri setinde sınıf ayrımını sınırlı düzeyde desteklemiştir. KLDA da çekirdek tabanlı denetimli bir yöntem olmasına rağmen, bu çalışmada KPCA kadar yüksek performans üretmemiştir.

KPCA-ICA yöntemi, yüksek sayılabilecek AUC değerine rağmen doğruluk ve F1 skoru bakımından daha sınırlı bir başarı göstermiştir. Bu durum, yöntemin bazı sınıflar için ayırt etme kapasitesi taşıdığını; ancak bu kapasitenin tüm sınıflara dengeli biçimde yansımadığını göstermektedir. Bu nedenle KPCA-ICA, mevcut veri setinde KPCA ve PCA yöntemlerine göre daha sınırlı bir temsil gücü sunmuştur.

İşlem süresi açısından değerlendirildiğinde, doğrusal yöntemlerin daha düşük hesaplama maliyetine sahip olduğu; KPCA ve KPCA-ICA gibi çekirdek tabanlı yöntemlerde ise işlem süresinin daha yüksek olduğu görülmektedir. Bununla birlikte KPCA'nın AUC ve F1 skoru bakımından sağladığı performans artışı, bazı uygulamalarda ek işlem maliyetini kabul edilebilir hale getirebilir. Gerçek zamanlı izleme sistemlerinde ise yöntem seçimi yapılırken performans ve işlem süresi birlikte değerlendirilmelidir.

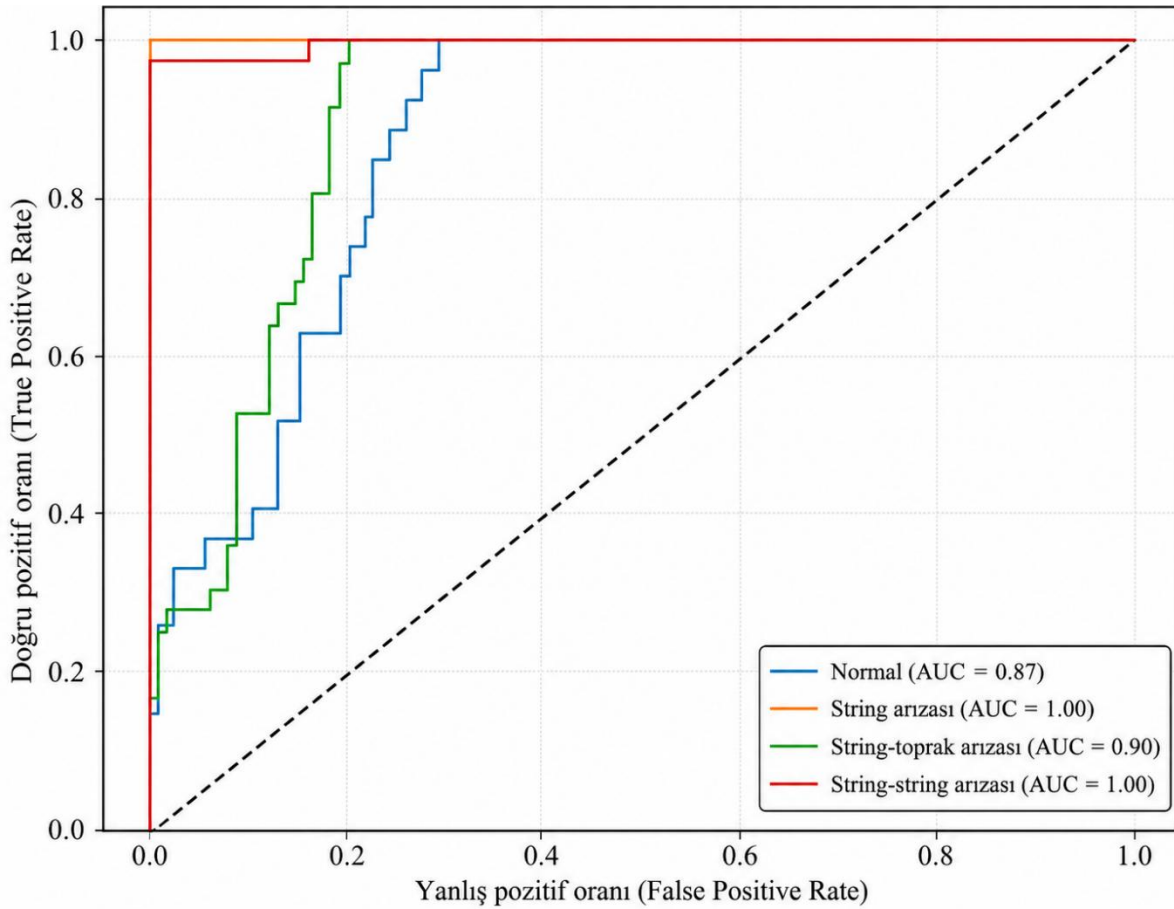
Sınıflandırma performansının sınıf bazlı hata dağılımı üzerinden incelenmesi için PCA tabanlı modelin karışıklık matrisi Şekil 4.4'te verilmiştir.



Şekil 4.4. PCA tabanlı sınıflandırma modeli için karışıklık matrisi

Şekil 4.4'te verilen karışıklık matrisi, modelin hangi sınıfları doğru sınıflandırdığını ve hangi sınıflar arasında karışmalar meydana geldiğini göstermektedir. FV arıza sınıflandırmasında özellikle benzer elektriksel etkiler oluşturan arıza türleri arasında karışmaların görülmesi beklenebilir. Bu nedenle karışıklık matrisi, model performansının yalnızca genel doğruluk değeriyle değil, hata dağılımı üzerinden de yorumlanmasını sağlamaktadır.

PCA tabanlı sınıflandırma modeli için ROC eğrileri Şekil 4.5'te verilmiştir.



**Şekil 4.5.** PCA tabanlı sınıflandırma modeli için ROC eğrileri

Şekil 4.5'te verilen ROC eğrileri, PCA tabanlı modelin farklı karar eşikleri altında sınıfları ayırt etme kapasitesini göstermektedir. ROC analizi, sınıflandırıcının yalnızca doğru tahmin oranını değil, duyarlılık ve yanlış pozitif oranı arasındaki ilişkiyi de değerlendirmeye imkân tanımaktadır. Bu nedenle ROC eğrileri, çok sınıflı FV arıza teşhisi problemlerinde modelin genel ayırt edicilik düzeyini yorumlamak açısından önemlidir.

Genel olarak değerlendirildiğinde, KPCA bu veri seti ve değerlendirme protokolü kapsamında en güçlü ayırt ediciliği sunarken, PCA daha sade ve yorumlanabilir bir temel temsil

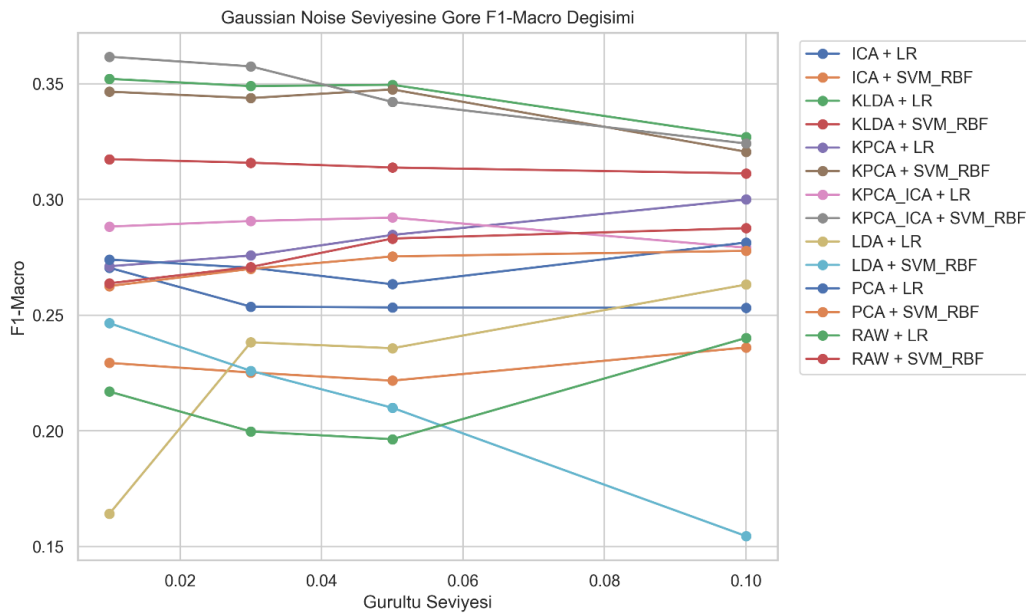
sağlamaktadır. Bu sonuçlar, FV arıza sınıflandırmasında yöntem seçiminin yalnızca doğruluk değerine değil; AUC, F1 skoru, işlem süresi ve yorumlanabilirlik gibi ölçütlere birlikte dayandırılması gerektiğini göstermektedir.

#### 4.4. Bozulmuş Senaryoları Altında Dayanıklılık Analizi

Bu bölümde, boyut azaltma yöntemlerinin saha benzeri bozulmuş koşulları altındaki dayanıklılığı değerlendirilmiştir. Temiz veri üzerindeki sınıflandırma performansı önemli olmakla birlikte, gerçek uygulamalarda ölçüm gürültüsü, eksik veri, sensör kaybı ve sistematik ölçüm sapmaları gibi durumlarla karşılaşılması mümkündür. Bu nedenle yöntemlerin yalnızca ideal koşullardaki başarıları değil, bozulmuş test koşulları altındaki kararlılıkları da incelenmiştir.

Bozulmuş senaryolarında elde edilen bulgular, boyut azaltma yöntemlerinin dayanıklılık özelliklerinin birbirinden farklı olduğunu göstermektedir. Gauss gürültüsü deneylerinde düşük ve orta düzey gürültü altında KPCA-ICA ve KLDA tabanlı yapıların daha yüksek F1-macro değerleri ürettiği gözlemlenmiştir. Gürültü düzeyi arttıkça bazı doğrusal yöntemlerde performans düşüşü belirginleşirken, doğrusal olmayan temsil yöntemlerinin görece daha kararlı kaldığı görülmüştür.

Gauss gürültüsü seviyesine göre yöntemlerin F1-macro değişimi Şekil 4.6'da verilmiştir.

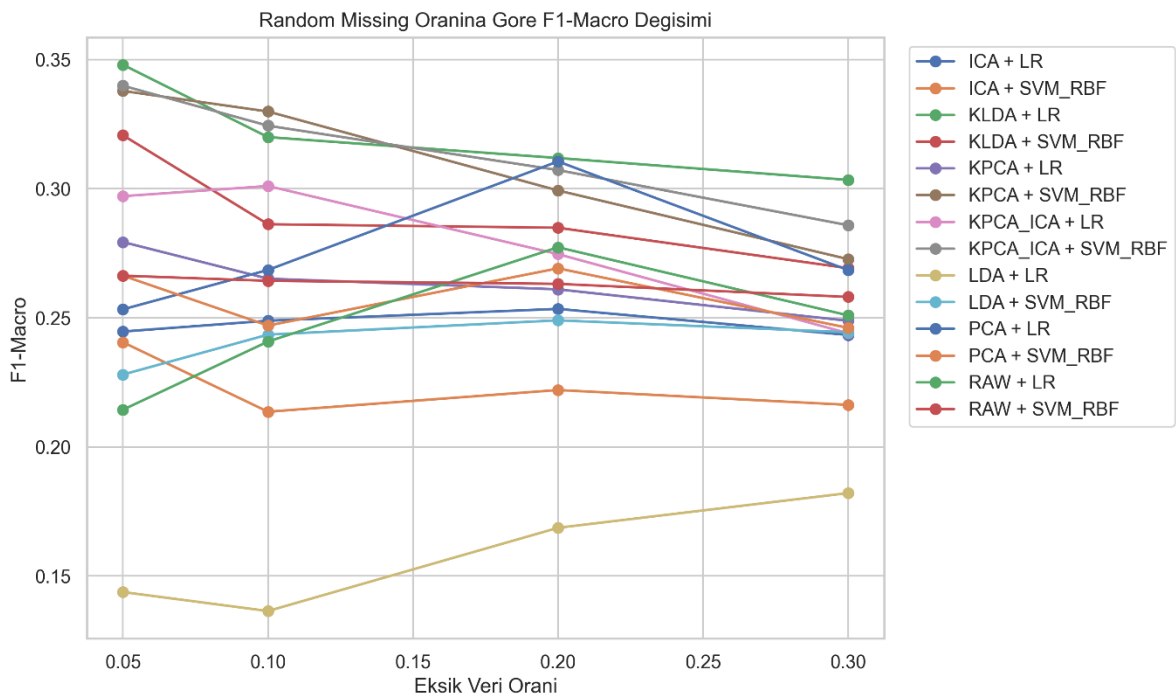


Şekil 4.6. Gauss gürültüsü seviyesine göre yöntemlerin F1-macro değişimi

Şekil 4.6, gürültü düzeyi arttıkça yöntemlerin sınıflandırma başarısında farklı oranlarda değişim meydana geldiğini göstermektedir. Bu sonuç, ölçüm gürültüsünün yalnızca sınıflandırıcı performansını değil, aynı zamanda kullanılan boyut azaltma yönteminin temsil kararlılığını da etkilediğini ortaya koymaktadır.

Eksik veri senaryolarında, özellikle %20 ve %30 düzeylerinde KLDA tabanlı yapının en başarılı yöntemlerden biri olduğu belirlenmiştir. Bu durum, doğrusal olmayan özellik uzayında sınıf ayırımı güçlendiren yaklaşımların veri kaybı altında daha dayanıklı olabildiğini göstermektedir.

Eksik veri oranına göre yöntemlerin F1-macro değişimi Şekil 4.7’de verilmiştir.



Şekil 4.7. Eksik veri oranına göre yöntemlerin F1-macro değişimi

Şekil 4.7, eksik veri oranı arttıkça yöntemler arasındaki dayanıklılık farklarının belirginleştiğini göstermektedir. Özellikle yüksek eksik veri oranlarında bazı yöntemlerin performansı daha hızlı düşerken, KLDA ve KPCA-ICA tabanlı yapıların görece daha kararlı sonuçlar ürettiği görülmektedir.

Tek sensör kaybı koşulunda tüm yöntemlerde performans düşüşü gözlenmiştir. Ancak KLDA ve KPCA-ICA tabanlı yapıların görece daha yüksek başarıyı koruduğu belirlenmiştir. Bu bulgu, çevresel veya elektriksel bir ölçüm değişkeninin kaybı durumunda yöntemlerin aynı

ölçüde etkilenmediğini ve bazı temsil yaklaşımlarının sensör kaybına karşı daha dayanıklı olabildiğini göstermektedir.

Bozulmuş senaryolarına göre en başarılı yöntemler Tablo 4.2’de özetlenmiştir.

**Tablo 4.2.** Bozulmuş senaryolarına göre en başarılı yöntemler

Senaryo	Bozulma seviyesi	Boyut azaltma yöntemi	Sınıflandırıcı	F1-macro	Accuracy	ROC-AUC OVR macro	Specificity macro	Robustness F1
block_missing	10	KLDA	LR	0.3531	0.4100	0.5526	0.8033	1.0030
block_missing	20	KPCA_ICA	SVM_RBF	0.3549	0.3900	0.5505	0.7967	1.0007
block_missing	5	KPCA_ICA	SVM_RBF	0.3522	0.3867	0.5575	0.7956	0.9928
gaussian_noise	0.01	KPCA_ICA	SVM_RBF	0.3617	0.3967	0.5686	0.7989	1.0196
gaussian_noise	0.03	KPCA_ICA	SVM_RBF	0.3575	0.3967	0.5646	0.7989	1.0079
gaussian_noise	0.05	KLDA	LR	0.3495	0.4067	0.5505	0.8022	0.9927
gaussian_noise	0.1	KLDA	LR	0.3271	0.3833	0.5485	0.7944	0.9289
irradiance_band	high	LDA	SVM_RBF	0.2209	0.3235	0.4768	0.7619	1.0353
irradiance_band	low	KLDA	LR	0.4439	0.5455	0.7185	0.8509	1.2607
irradiance_band	mid	KLDA	SVM_RBF	0.3139	0.4242	0.5892	0.8021	0.9772
random_missing	0.05	KLDA	LR	0.3480	0.4067	0.5565	0.8022	0.9883
random_missing	0.1	KPCA	SVM_RBF	0.3299	0.3767	0.6124	0.7922	0.9655
random_missing	0.2	KLDA	LR	0.3119	0.3700	0.5583	0.7900	0.8857
random_missing	0.3	KLDA	LR	0.3034	0.3567	0.5535	0.7856	0.8617
sensor_drift	0.01	KPCA_ICA	SVM_RBF	0.3672	0.4000	0.5680	0.8000	1.0352
sensor_drift	0.02	KPCA_ICA	SVM_RBF	0.3672	0.4000	0.5658	0.8000	1.0352

Senaryo	Bozulma seviyesi	Boyut azaltma yöntemi	Sınıflandırıcı	F1-macro	Accuracy	ROC-AUC OVR macro	Specificity macro	Robustness F1
sensor_drift	0.05	KLDA	LR	0.3521	0.4100	0.5501	0.8033	1.0000
single_sensor_dropout	1	KLDA	LR	0.2963	0.3600	0.5297	0.7867	0.8415

Tablo 4.2, temiz veri üzerinde en iyi sonuç veren yöntemin her bozulma senaryosunda aynı başarıyı göstermeyebileceğini ortaya koymaktadır. Bu nedenle yöntem seçimi yapılırken yalnızca temiz veri performansı değil, gürültü, eksik veri, sensör sürüklenmesi ve sensör kaybı gibi saha benzeri koşullar altındaki dayanıklılık da dikkate alınmalıdır.

#### 4.5. Çevresel Koşullara Göre Performans Değişimi

FV sistemlerde ölçüm verileri, arıza durumlarının yanı sıra çevresel koşullardan da etkilenmektedir. Bu nedenle yöntemlerin farklı ışınım seviyeleri altında nasıl davrandığının incelenmesi, saha uygulamaları açısından önem taşımaktadır. Bu kapsamda test verisi düşük, orta ve yüksek ışınım bantlarına ayrılmış ve yöntem performansları karşılaştırılmıştır.

Işınım bantlarına göre yapılan değerlendirme, yöntem performansının çevresel çalışma koşullarına duyarlı olduğunu göstermiştir. Düşük ışınım bandında sınıflandırma başarımının daha yüksek olduğu, yüksek ışınım bandında ise performansın belirgin biçimde düştüğü gözlenmiştir. Bu bulgu, FV arıza teşhisinde yalnızca genel doğruluk değerinin değil, farklı işletme bölgelerinde elde edilen başarımın da dikkate alınması gerektiğini göstermektedir.

Çevresel değişkenlik altında en güçlü sonuç düşük ışınım bandında KLDA tabanlı yapı ile elde edilmiştir. Buna karşılık yüksek ışınım bandında yöntemler arasındaki performans farkı azalmış ve genel başarı düzeyi düşmüştür. Bu durum, model kararlarının çevresel koşullara bağlı değişebildiğini ve saha uygulamalarında çevresel genelleme analizinin gerekli olduğunu ortaya koymaktadır.

Genel performans açısından KPCA'nın AUC ve F1 skoru bakımından öne çıkmasına karşın, düşük ışınım koşulundaki dayanıklılık analizlerinde KLDA tabanlı yapının güçlü sonuç vermesi, yöntem performansının değerlendirme koşuluna bağlı olarak değişebileceğini göstermektedir. Bu nedenle FV arıza sınıflandırmasında tek bir "en iyi yöntem" tanımı yerine, temiz veri başarımı, bozulma dayanıklılığı ve çevresel koşullara göre kararlılık birlikte değerlendirilmelidir.

#### 4.6. Bulguların Genel Değerlendirilmesi

Bu çalışmada elde edilen bulgular, FV sistemlerde arıza sınıflandırma probleminin yalnızca sınıflandırıcı seçimiyle çözülebilecek bir problem olmadığını göstermektedir. Veri temsili, özellik mühendisliği, boyut azaltma ve performans değerlendirme adımları birlikte ele alındığında, arıza sınıfları arasındaki ayırım daha sağlıklı biçimde yorumlanabilmektedir. FV sistemlerden elde edilen ölçüm verileri; akım, gerilim, güç, ısıtım, sıcaklık ve arıza koşullarının birlikte etkilediği çok değişkenli bir yapı sunduğundan, yalnızca tekil ölçüm büyüklüklerine veya sabit eşik değerlerine dayalı yaklaşımlar sınırlı kalabilmektedir (Madeti ve Singh, 2017; Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

Tablo 4.1’de verilen sonuçlar, KPCA’nın bu veri seti ve değerlendirme protokolü kapsamında F1 skoru ve AUC değeri bakımından en başarılı temsil yaklaşımı olduğunu göstermektedir. KPCA ve PCA aynı doğruluk değerine ulaşmış olsa da, KPCA’nın daha yüksek AUC ve F1 skoru üretmesi, doğrusal olmayan temsil uzayının arıza sınıflarını daha dengeli ve ayırt edici biçimde ifade edebildiğini göstermektedir. Bu durum, FV sistem verilerinde arıza etkilerinin her zaman doğrusal ilişkilerle açıklanamayabileceğini ve çekirdek tabanlı dönüşümlerin bazı sınıflandırma problemlerinde önemli katkılar sağlayabileceğini ortaya koymaktadır (Schölkopf vd., 1998).

PCA’nın doğrusal ve yorumlanabilir yapısı ise mühendislik açısından önemli bir avantajdır. Eğer amaç daha sade, açıklanabilir ve düşük karmaşıklıkta bir yapı kurmak ise PCA uygulanabilir bir alternatif sunmaktadır. Buna karşılık en yüksek sınıflar arası ayırt ediciliğin hedeflendiği durumda KPCA bu veri seti için daha güçlü bir temsil yaklaşımı sağlamaktadır. Bu nedenle yöntem seçimi yapılırken doğruluk, F1 skoru, AUC değeri, işlem süresi, yorumlanabilirlik ve uygulanabilirlik birlikte değerlendirilmelidir (Fawcett, 2006; Sokolova ve Lapalme, 2009).

LDA ve KLDA sonuçları, sınıf bilgisine dayalı yöntemlerin bu veri setinde orta düzeyde performans ürettiğini göstermektedir. LDA’nın sınıf etiketlerini kullanarak sınıflar arası ayırımı artırmayı amaçlaması teorik olarak avantajlıdır. Ancak FV arıza verilerinde sınıf içi değişkenliğin yüksek olması, çevresel koşulların sinyal davranışlarını etkilemesi ve bazı arıza türlerinin birbirine benzer ölçüm örüntüleri üretmesi, doğrusal sınıf ayırımını sınırlayabilmektedir. KLDA ise çekirdek tabanlı bir ayırım yaklaşımı sunmasına rağmen, bu çalışmada KPCA kadar güçlü bir performans elde edememiştir (Fisher, 1936; Baudat ve Anouar, 2000).

ICA ve KPCA-ICA sonuçları, bağımsız bileşen temelli temsil yaklaşımlarının bu veri setinde sınırlı kaldığını göstermektedir. FV sistemlerde arıza etkileri, çevresel koşullar ve güç elektroniği dinamikleri çoğu zaman birbirinden tamamen bağımsız değildir. Bu nedenle ICA ile elde edilen bileşenlerin sınıflandırma açısından her zaman en ayırt edici temsil olması beklenmemelidir (Hyvärinen vd., 2001). KPCA-ICA yaklaşımında ise KPCA ile elde edilen doğrusal olmayan temsil üzerinde ICA uygulanmış, ancak bu yapı mevcut veri setinde sınıflandırma başarısını KPCA düzeyine taşıyamamıştır.

İşlem süresi açısından değerlendirildiğinde, çekirdek tabanlı yöntemlerin doğrusal yöntemlere göre daha fazla hesaplama maliyeti oluşturduğu görülmektedir. KPCA ve KPCA-ICA yöntemlerinde işlem süresinin daha yüksek olması, çekirdek matrislerinin oluşturulması ve doğrusal olmayan dönüşüm işlemlerinden kaynaklanmaktadır. Buna karşın KPCA'nın F1 skoru ve AUC değeri bakımından sağladığı avantaj, bazı uygulamalarda bu ek hesaplama maliyetini kabul edilebilir hale getirebilir. Ancak gerçek zamanlı izleme sistemlerinde kullanılacak yöntemlerin seçimi yapılırken işlem süresi ve donanım gereksinimleri de dikkate alınmalıdır.

Dayanıklılık analizleri, temiz veri üzerinde en iyi sonuç veren yöntemin her bozulmuş koşulunda aynı üstünlüğü göstermeyebileceğini ortaya koymuştur. Gürültü, eksik veri, sensör sürüklenmesi, tek sensör kaybı ve ışınım bandı temelli değerlendirmeler, yöntemlerin saha benzeri koşullardaki kararlılıklarının farklılaştığını göstermektedir. Özellikle düşük ışınım bandında KLDA tabanlı yapının güçlü sonuç vermesi, çevresel koşullara göre yöntem performansının değişebileceğini göstermektedir.

Genel olarak bulgular, FV arıza sınıflandırmasında uygun boyut azaltma yönteminin seçiminin sınıflandırma başarısı üzerinde belirleyici olduğunu ortaya koymaktadır. Bu çalışma kapsamında KPCA, genel AUC ve F1 skoru bakımından en güçlü temsil yaklaşımı olarak öne çıkarken; PCA sade ve yorumlanabilir yapısıyla mühendislik açısından güçlü bir alternatif sunmaktadır. KLDA ise özellikle bazı bozulmuş ve çevresel koşullar altında dayanıklılık açısından avantaj sağlayabilmektedir. Bu nedenle yöntem seçimi, yalnızca temiz veri üzerindeki başarıya değil, yorumlanabilirlik, işlem süresi ve saha benzeri koşullar altındaki kararlılığa da dayandırılmalıdır.

## 5. TARTIŞMA

Bu tez çalışmasında, FV sistemlerde arıza tespiti ve sınıflandırılması problemi, doğrusal ve çekirdek tabanlı doğrusal olmayan boyut azaltma tekniklerinin karşılaştırmalı değerlendirilmesi üzerinden ele alınmıştır. Çalışmada Temel Bileşen Analizi, Doğrusal Ayırma Analizi, Bağımsız Bileşen Analizi, Çekirdek Temel Bileşen Analizi, Çekirdek Doğrusal Ayırma Analizi ve KPCA-ICA yaklaşımı aynı deneysel çerçevede değerlendirilmiştir. Elde edilen bulgular, FV arıza sınıflandırma başarısının yalnızca kullanılan sınıflandırıcıya değil, sınıflandırıcıya sunulan veri temsiline de güçlü biçimde bağlı olduğunu göstermektedir.

FV sistemlerden elde edilen ölçüm verileri, çevresel koşullar ve sistem dinamikleriyle birlikte değişen çok boyutlu bir yapıya sahiptir. Güneş ışınımı, hücre sıcaklığı, arıza direnci, güç elektroniği davranışı ve sistem çalışma noktası gibi değişkenler; akım, gerilim ve güç sinyalleri üzerinde eş zamanlı etkiler oluşturabilmektedir. Bu nedenle FV sistemlerde arıza teşhisi, yalnızca tek bir sinyalin izlenmesine veya sabit eşik değerlerine dayalı olarak güvenilir biçimde gerçekleştirilemeyebilir. Literatürde de FV sistemlerde arıza tespiti ve izleme çalışmalarında çok değişkenli veri analizi, özellik çıkarma ve makine öğrenmesi tabanlı yöntemlerin önemi vurgulanmaktadır (Madeti ve Singh, 2017; Mellit vd., 2018; Triki-Lahiani vd., 2018; Hong ve Pula, 2022).

Bu çalışmada kullanılan özellikler; FV sistemin elektriksel davranışını temsil eden akım, gerilim, güç, ışınım, sıcaklık ve bu büyüklüklerden türetilmiş istatistiksel değişkenlerden oluşturulmuştur. Özellikle string akımları, toplam akım, DC gerilim, DC güç ve range tabanlı değişkenler, arıza etkilerinin çok değişkenli ölçüm uzayında temsil edilmesine katkı sağlamıştır. Ghoneim vd. (2021) tarafından yürütülen çalışmada da 250 kW şebeke bağlantılı FV sistem için 30 özellikten oluşan veri seti kullanılmış ve range tabanlı değişkenlerin arıza tespitinde önemli bilgi taşıdığı gösterilmiştir. Bu açıdan mevcut tezde kullanılan özellik yapısı, literatürdeki benzer FV arıza sınıflandırma çalışmalarıyla uyumludur.

Ghoneim vd. (2021), 250 kW gücündeki şebeke bağlantılı bir FV çiftliği için normal çalışma, string arızası, string-toprak arızası ve string-string arızası sınıflarını içeren simülasyon tabanlı bir veri seti üzerinde farklı makine öğrenmesi algoritmalarını karşılaştırmıştır. Söz konusu çalışmada Random Forest, Logistic Regression, Naive Bayes, AdaBoost ve CN2 rule induction gibi sınıflandırıcılar değerlendirilmiş ve AdaBoost'un test verisi üzerinde en yüksek performansı sağladığı raporlanmıştır. Bu tez çalışması ise aynı problem bağlamında

sınıflandırıcı seçiminden çok, sınıflandırıcıya sunulan veri temsilinin niteliğine ve doğrusal/doğrusal olmayan boyut azaltma yöntemlerinin etkisine odaklanmaktadır. Bu yönüyle mevcut çalışma, literatürdeki sınıflandırıcı odaklı yaklaşımları tamamlayıcı bir veri temsil analizi sunmaktadır.

Boyut azaltma yöntemleri arasında en dikkat çekici sonuç KPCA yöntemiyle elde edilmiştir. KPCA, PCA ile aynı doğruluk değerine ulaşmış olmasına rağmen AUC ve F1 skoru bakımından daha yüksek performans üretmiştir. Bu sonuç, FV arıza verilerinde doğrusal olmayan ilişkilerin sınıflandırma başarısı açısından önemli olduğunu göstermektedir. KPCA'nın çekirdek fonksiyonu yardımıyla veriyi örtük bir özellik uzayında temsil etmesi, doğrusal olarak ayrılması güç olan sınıf yapılarının daha belirgin hale gelmesine katkı sağlamış olabilir. Bu bulgu, çekirdek tabanlı yöntemlerin doğrusal olmayan veri yapılarında güçlü temsil kapasitesi sunabileceğini belirten temel çalışmalarla da uyumludur (Schölkopf vd., 1998).

KPCA'nın AUC ve F1 skoru bakımından öne çıkması, FV arıza sınıflandırmasında yalnızca genel doğruluk değerinin yeterli bir değerlendirme ölçütü olmadığını da göstermektedir. PCA ve KPCA aynı doğruluk düzeyine ulaşmış olsa da, KPCA'nın daha yüksek AUC değeri sınıflar arası ayırt etme kapasitesinin daha güçlü olduğunu; daha yüksek F1 skoru ise kesinlik ve duyarlılık dengesinin daha başarılı kurulduğunu göstermektedir. Bu nedenle çok sınıflı FV arıza teşhisinde doğruluk, F1 skoru ve AUC gibi ölçütlerin birlikte değerlendirilmesi gerekmektedir (Fawcett, 2006; Sokolova ve Lapalme, 2009).

PCA yöntemi ise doğrusal yapısına rağmen çalışma kapsamında önemli bir alternatif olarak öne çıkmıştır. PCA, KPCA ile aynı doğruluk düzeyine ulaşmış, ancak F1 skoru ve AUC değeri bakımından daha düşük sonuç üretmiştir. Buna rağmen PCA'nın sade, kararlı ve yorumlanabilir bir temsil sağlaması mühendislik uygulamaları açısından önemlidir. PCA, yüksek korelasyon içeren özellikleri daha az sayıda bileşenle temsil ederek veri uzayını sadeleştirmektedir. Bu nedenle PCA, özellikle gerçek zamanlı izleme veya düşük hesaplama maliyeti gerektiren uygulamalarda güçlü bir seçenek olarak değerlendirilebilir.

LDA ve KLDA sonuçları, sınıf bilgisine dayalı yöntemlerin bu veri setinde orta düzeyde performans ürettiğini göstermektedir. LDA'nın sınıf etiketlerini kullanarak sınıflar arası ayrımı artırmayı amaçlaması teorik olarak avantajlıdır. Ancak FV arıza verilerinde sınıf içi değişkenliğin yüksek olması, çevresel koşulların sinyal davranışlarını etkilemesi ve bazı arıza türlerinin birbirine benzer ölçüm örüntüleri üretmesi, doğrusal sınıf ayrımını sınırlandırmaktadır. KLDA ise çekirdek tabanlı bir ayırım yaklaşımı sunmasına rağmen, genel

performans açısından KPCA kadar güçlü sonuç üretmemiştir. Bu durum, denetimli veya çekirdek tabanlı bir yöntemin tek başına yüksek başarıyı garanti etmediğini; veri yapısı, sınıf dağılımı ve yöntem-parametre uyumunun birlikte değerlendirilmesi gerektiğini göstermektedir (Fisher, 1936; Baudat ve Anouar, 2000).

ICA ve KPCA-ICA sonuçları, bağımsız bileşen temelli temsil yaklaşımlarının bu veri setinde sınırlı kaldığını göstermektedir. ICA, gözlenen veriyi istatistiksel olarak bağımsız bileşenlere ayırmayı amaçlamaktadır. Ancak FV sistemlerde arıza etkileri, çevresel koşullar ve güç elektroniği dinamikleri çoğu zaman birbirinden tamamen bağımsız değildir. Bu nedenle ICA ile elde edilen bileşenlerin sınıflandırma açısından her zaman en ayırt edici temsili üretmesi beklenmemelidir (Hyvärinen vd., 2001).

KPCA-ICA yaklaşımı, bu çalışmada KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde gerçekleştirilmiştir. Bu yapı teorik olarak doğrusal olmayan temsil ile bağımsız bileşen ayrıştırmasını bir araya getirmektedir. Ancak deneysel sonuçlar, KPCA-ICA'nın doğruluk ve F1 skoru bakımından diğer yöntemlerin gerisinde kaldığını göstermiştir. KPCA-ICA'nın AUC değerinin görece yüksek olmasına rağmen F1 skorunun düşük kalması, yöntemin bazı sınıflar için ayırt edici bilgi taşıdığını; ancak bu bilginin tüm sınıflara dengeli biçimde yansımadığını düşündürmektedir. Bu durum, yalnızca AUC değerine göre yöntem seçimi yapılmasının yanıltıcı olabileceğini göstermektedir.

Bu bulgular, daha karmaşık veya gelişmiş görünen yöntemlerin her zaman daha yüksek sınıflandırma başarısı sağlamayacağını ortaya koymaktadır. FV arıza sınıflandırmasında yöntem başarısı; veri setinin yapısı, sınıflar arası örtüşme, kullanılan özellikler, dönüşüm parametreleri, sınıflandırıcı seçimi ve değerlendirme ölçütleriyle birlikte ele alınmalıdır. Bu nedenle çekirdek tabanlı veya bağımsız bileşen temelli yöntemlerin başarısı, yalnızca teorik üstünlükleriyle değil, deneysel performans ölçütleriyle birlikte yorumlanmalıdır.

Çalışmada elde edilen sonuçlar, performans ölçütlerinin birlikte değerlendirilmesinin önemini açık biçimde göstermektedir. Doğruluk değeri genel başarı düzeyini özetlese de, çok sınıflı arıza teşhisi problemlerinde sınıflar arasındaki hata dağılımını tek başına açıklamamaktadır. F1 skoru kesinlik ve duyarlılık dengesini, AUC değeri sınıflar arası ayırt ediciliği, karışıklık matrisi ise hangi sınıfların birbirine karıştırıldığını göstermektedir. Bu nedenle FV arıza sınıflandırmasında model değerlendirmesi yalnızca tek bir metrik üzerinden değil, çoklu performans ölçütleri üzerinden yapılmalıdır (Ghoneim vd., 2021; Fawcett, 2006; Sokolova ve Lapalme, 2009).

Çalışmanın güçlü yönlerinden biri, doğrusal ve çekirdek tabanlı doğrusal olmayan boyut azaltma yöntemlerinin aynı veri seti ve aynı değerlendirme çerçevesi altında karşılaştırılmış olmasıdır. Bu yaklaşım, yöntemlerin yalnızca teorik özelliklerinin değil, aynı FV arıza sınıflandırma problemindeki pratik etkilerinin de değerlendirilmesine olanak sağlamıştır. Ayrıca çalışmanın normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört sınıflı bir yapı içermesi, arıza teşhisinin yalnızca arıza var/yok düzeyinde değil, arıza türü sınıflandırması düzeyinde ele alınmasını sağlamıştır.

Dayanıklılık deneyleri, en iyi yöntemin değerlendirme bağlamına göre değişebildiğini göstermiştir. Temiz veri üzerindeki genel değerlendirmede KPCA, AUC ve F1 skoru bakımından en başarılı temsil yaklaşımı olarak öne çıkmıştır. Buna karşılık bozulmuş senaryoları ve çevresel çalışma koşulları altında yöntemlerin performans sıralaması değişebilmiştir. Gürültü, eksik veri, sensör sürüklenmesi, tek sensör kaybı ve farklı ışınım bantları altında elde edilen sonuçlar, yöntemlerin saha benzeri koşullardaki kararlılıklarının farklılaştığını göstermektedir.

Özellikle düşük ışınım bandında KLDA tabanlı yapının güçlü sonuç vermesi, yöntem seçiminin yalnızca temiz veri başarımına göre yapılmaması gerektiğini ortaya koymaktadır. Bu bulgu, FV arıza teşhisinde uygun yöntemin kullanım amacına ve çalışma koşullarına göre değişebileceğini göstermektedir. Eğer amaç genel ayırt ediciliği artırmak ise KPCA güçlü bir temsil yaklaşımı sunarken; düşük ışınım gibi belirli çevresel koşullarda KLDA tabanlı yapı daha dayanıklı sonuçlar üretebilmektedir. Bu nedenle FV arıza sınıflandırmasında tek bir “en iyi yöntem” tanımı yerine, performans, yorumlanabilirlik, işlem süresi ve dayanıklılık ölçütlerinin birlikte ele alınması daha doğru bir yaklaşımdır.

### **5.1. Çalışmanın Sınırlılıkları**

Bu tez çalışması, MATLAB/Simulink ortamında oluşturulan 250 kW gücündeki şebeke bağlantılı FV sistemden elde edilen simülasyon tabanlı veri seti üzerinde gerçekleştirilmiştir. Simülasyon verileri, arıza senaryolarının kontrollü biçimde oluşturulmasına ve yöntemlerin karşılaştırmalı olarak değerlendirilmesine olanak sağlamaktadır. Bununla birlikte gerçek saha koşullarında karşılaşılabilecek sensör gürültüsü, ölçüm belirsizliği, kısmi gölgelenme, panel yaşlanması, inverter davranışı, farklı çevresel koşullar ve veri eksikliği gibi etkiler simülasyon ortamında tam olarak temsil edilemeyebilir. Bu nedenle elde edilen sonuçlar, önerilen boyut azaltma ve sınıflandırma yaklaşımının potansiyelini göstermekte; ancak yöntemin saha verileri üzerinde ayrıca doğrulanması gerekmektedir.

Çalışmanın bir diğer sınırlılığı, ele alınan arıza türleriyle ilgilidir. Bu tez kapsamında normal çalışma, string arızası, string-toprak arızası ve string-string arızası sınıfları değerlendirilmiştir. Bu arıza türleri FV sistemlerin DC tarafındaki temel arıza davranışlarını temsil etmekle birlikte, gerçek saha uygulamalarında kısmi gölgelenme, modül yaşlanması, bağlantı gevşekliği, sıcak nokta oluşumu, inverter kaynaklı arızalar veya sensör kaynaklı hatalar gibi daha geniş arıza durumlarıyla karşılaşılabılır. Bu nedenle gelecek çalışmalarda daha geniş arıza türlerinin ve farklı sistem topolojilerinin değerlendirilmesi yararlı olacaktır.

Bir diğer sınırlılık, yöntemlerin hiperparametre seçimiyle ilişkilidir. Özellikle KPCA, KLDA ve KPCA-ICA gibi çekirdek tabanlı yaklaşımlarda çekirdek fonksiyonu, gamma değeri, bileşen sayısı ve sınıflandırıcı parametreleri sonuçları doğrudan etkileyebilmektedir. Bu çalışmada yöntemler aynı karşılaştırma çerçevesinde değerlendirilmiş olmakla birlikte, gelecekte sistematik hiperparametre optimizasyonu ve çapraz doğrulama yaklaşımları kullanılarak yöntemlerin daha geniş bir parametre uzayında incelenmesi yararlı olacaktır.

Genel olarak tartışma sonuçları, FV arıza sınıflandırmasında veri temsilinin belirleyici bir rol oynadığını göstermektedir. KPCA, AUC ve F1 skoru bakımından bu veri seti ve değerlendirme protokolü kapsamında en başarılı temsil yaklaşımı olarak öne çıkmıştır. PCA ise aynı doğruluk seviyesine ulaşması ve daha sade yapısı nedeniyle mühendislik uygulamaları açısından önemli bir alternatif sunmuştur. LDA, ICA ve KLDA orta düzeyde sonuçlar üretirken, KPCA-ICA yaklaşımı bu veri setinde sınırlı kalmıştır. Dayanıklılık analizleri ise yöntem performansının bozulma türüne ve çevresel koşullara bağlı olarak değişebileceğini göstermiştir.

Sonuç olarak bu çalışma, FV sistemlerde arıza sınıflandırmasında veri temsilinin belirleyici rolünü ortaya koymaktadır. Elde edilen bulgular, yalnızca daha karmaşık sınıflandırıcılar kullanmanın yeterli olmadığını; veri yapısına uygun boyut azaltma ve temsil yöntemlerinin seçilmesi gerektiğini göstermektedir. Bu nedenle FV sistemlerde güvenilir ve uygulanabilir arıza teşhis sistemleri geliştirilirken veri toplama, özellik mühendisliği, boyut azaltma, sınıflandırma, dayanıklılık analizi ve performans değerlendirme adımları bütüncül bir mühendislik süreci olarak ele alınmalıdır.

## 6. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında, şebeke bağlantılı FV sistemlerde meydana gelen arıza türlerinin veri odaklı yöntemlerle sınıflandırılması ve bu süreçte doğrusal ile çekirdek tabanlı doğrusal olmayan boyut azaltma tekniklerinin sınıflandırma performansı üzerindeki etkilerinin karşılaştırmalı olarak incelenmesi amaçlanmıştır. Bu kapsamda normal çalışma, string arızası, string-toprak arızası ve string-string arızası olmak üzere dört farklı sistem durumu ele alınmıştır.

Çalışmada MATLAB/Simulink ortamında modellenen 250 kW gücündeki şebeke bağlantılı FV sistem üzerinden simülasyon tabanlı bir veri seti oluşturulmuştur. Elde edilen akım, gerilim, güç, ısıtım, sıcaklık ve türetilmiş istatistiksel özellikler; veri ön işleme, standardizasyon, boyut azaltma ve sınıflandırma adımlarından geçirilmiştir. Böylece FV arıza sınıflandırmasında veri temsilinin sınıflandırma başarısı üzerindeki etkisi sistematik biçimde değerlendirilmiştir.

Elde edilen sonuçlar, FV arıza sınıflandırmasında yalnızca sınıflandırıcı seçiminin değil, sınıflandırıcıya sunulan veri temsilinin niteliğinin de belirleyici olduğunu göstermektedir. Normalize edilmiş özellik uzayı analizleri, arıza sınıflarının ham veya yalnızca ölçeklenmiş veri uzayında tamamen ayrışmadığını ortaya koymuştur. Bu durum, FV sistemlerde arıza etkilerinin çevresel koşullar, arıza direnci, sistem çalışma noktası ve güç elektroniği dinamikleriyle birlikte değişmesinden kaynaklanmaktadır. Dolayısıyla çok değişkenli veri analizi ve boyut azaltma temelli temsil yöntemleri, FV arıza teşhisinde önemli bir gereklilik olarak öne çıkmaktadır.

Bu tez çalışmasından elde edilen temel sonuçlar aşağıdaki şekilde özetlenebilir:

- PCA ve KPCA yöntemleri, **0,71 doğruluk değeri** ile en yüksek genel doğruluk düzeyine ulaşmıştır.
- KPCA, **0,9003 AUC** ve **0,6960 F1 skoru** ile sınıflar arası ayırt edicilik ve dengeli sınıflandırma başarısı bakımından en başarılı temsil yaklaşımı olmuştur.
- PCA, KPCA'ya göre daha düşük AUC ve F1 skoru üretmesine rağmen, doğrusal, sade ve yorumlanabilir yapısı nedeniyle mühendislik uygulamaları açısından güçlü bir alternatif sunmaktadır.
- LDA, ICA ve KLDA yöntemleri genel sınıflandırma performansı bakımından orta düzeyde sonuçlar üretmiştir.

- KPCA-ICA yaklaşımı, temiz veri üzerindeki ana karşılaştırmada sınırlı sınıflandırma başarısı göstermiştir. Bununla birlikte bazı bozulmuş senaryolarında görece kararlı sonuçlar üretebilmesi, yöntemin yalnızca genel doğruluk değeriyle değil, farklı koşullar altındaki davranışıyla da değerlendirilmesi gerektiğini göstermektedir.
- Bozulmuş ve çevresel çalışma koşulu analizleri, yöntem seçiminin yalnızca temiz veri üzerindeki başarıya göre yapılamayacağını ortaya koymuştur. Gürültü, eksik veri, sensör sürüklenmesi, tek sensör kaybı ve farklı ışınım bandı koşulları altında yöntemlerin performans sıralamaları değişebilmektedir.
- Düşük ışınım bandında KLDA tabanlı yapının güçlü sonuç vermesi, bazı çevresel koşullarda denetimli ve çekirdek tabanlı temsil yaklaşımlarının dayanıklılık açısından avantaj sağlayabileceğini göstermektedir.
- Elde edilen bulgular, FV arıza sınıflandırmasında boyut azaltma ve özellik temsilinin sınıflandırıcı seçimi kadar önemli olduğunu ortaya koymaktadır.

Çalışmada kullanılan özellik uzayı; string akımları, toplam akım, DC gerilim, DC güç, ışınım, sıcaklık ve bu büyüklüklerden türetilmiş istatistiksel değişkenlerden oluşturulmuştur. Bulgular, tekil ölçüm değişkenlerinin arıza sınıflarını güvenilir biçimde ayırmak için çoğu durumda yeterli olmadığını göstermektedir. Özellikle string akımları, toplam akım ve range tabanlı türetilmiş değişkenler, arıza kaynaklı akım dengesizliklerinin temsil edilmesinde önemli rol oynamıştır. Bununla birlikte, bu özelliklerin sınıflandırma açısından daha etkili kullanılabilmesi için uygun ölçekleme ve boyut azaltma adımlarının uygulanması gerekmektedir.

Boyut azaltma yöntemleri arasında KPCA'nın öne çıkması, FV arıza verilerinde doğrusal olmayan ilişkilerin önemli olduğunu göstermektedir. Arıza etkileri yalnızca tek bir elektriksel büyüklükte değil; akım, gerilim, güç, ışınım ve sıcaklık değişkenlerinin birlikte oluşturduğu örüntülerde ortaya çıkmaktadır. KPCA'nın çekirdek tabanlı yapısı, bu karmaşık ilişkilerin daha ayırt edici bir temsil uzayında ifade edilmesine katkı sağlamıştır. Bu nedenle KPCA, genel performans açısından bu veri seti için güçlü bir temsil yaklaşımı olarak değerlendirilmiştir.

Buna karşılık PCA'nın aynı doğruluk düzeyine ulaşması ve daha sade bir yapıya sahip olması, mühendislik uygulamaları açısından önemlidir. PCA, yüksek korelasyon içeren özellikleri daha az sayıda bileşenle temsil edebilmekte ve hesaplama açısından daha düşük

karmaşıklık sunmaktadır. Bu nedenle gerçek zamanlı izleme, düşük işlem yükü ve yorumlanabilirlik gerektiren uygulamalarda PCA güçlü bir alternatif olarak değerlendirilebilir.

LDA ve KLDA sonuçları, sınıf bilgisine dayalı yöntemlerin her durumda en yüksek performansı garanti etmediğini göstermiştir. LDA'nın doğrusal sınıf ayrımı varsayımı, FV arıza verilerindeki doğrusal olmayan örüntüleri temsil etmede sınırlı kalabilmektedir. KLDA ise çekirdek tabanlı ve denetimli bir yapı sunmasına rağmen, genel temiz veri performansında KPCA'nın gerisinde kalmıştır. Ancak düşük ışınım bandı ve bazı bozulmuş koşullarında KLDA'nın güçlü sonuçlar üretmesi, bu yöntemin çevresel değişkenlik altında dikkate değer bir dayanıklılık potansiyeline sahip olduğunu göstermektedir.

ICA ve KPCA-ICA yaklaşımları ise bu veri setinde sınırlı sınıflandırma başarısı üretmiştir. ICA'nın bağımsız bileşen varsayımı, FV sistemlerdeki arıza etkileri, çevresel koşullar ve güç elektroniği dinamiklerinin birbirinden tamamen bağımsız olmaması nedeniyle sınırlı kalmış olabilir. KPCA-ICA yaklaşımı, KPCA ile elde edilen doğrusal olmayan temsil uzayı üzerinde ICA uygulanması şeklinde gerçekleştirilmiştir; ancak bu yapı mevcut veri setinde sınıf ayrımını KPCA veya PCA düzeyinde güçlendirememiştir.

Sınıflandırma performansı değerlendirilirken yalnızca doğruluk değerinin yeterli olmadığı görülmüştür. KPCA ve PCA yöntemlerinin aynı doğruluk düzeyine ulaşmasına rağmen AUC ve F1 skoru bakımından farklı sonuçlar üretmesi, çoklu performans ölçütlerinin birlikte kullanılmasının önemini ortaya koymaktadır. FV arıza teşhisinde yanlış alarm ve kaçırılan arıza kararlarının mühendislik karşılıkları farklı olduğundan; doğruluk, AUC, F1 skoru, kesinlik, duyarlılık, özgüllük, karışıklık matrisi ve işlem süresi gibi ölçütlerin birlikte değerlendirilmesi gerekmektedir.

Bu çalışmanın genel sonucu, FV arıza sınıflandırmasında veri temsilinin sınıflandırma başarısı üzerinde belirleyici bir etkiye sahip olduğudur. Uygun özellik mühendisliği ve boyut azaltma adımları uygulanmadan, yalnızca daha karmaşık sınıflandırıcıların kullanılması yeterli olmayabilir. Elde edilen bulgular, FV sistemlerde akıllı arıza teşhisi geliştirilirken ölçüm verilerinin uygun biçimde ölçeklendirilmesi, özellikler arasındaki korelasyonların dikkate alınması, doğrusal olmayan ilişkilerin temsil edilmesi ve bu verilerin daha ayırt edici temsil uzaylarına dönüştürülmesi gerektiğini göstermektedir.

Bu sonuçlar doğrultusunda aşağıdaki öneriler sunulabilir:

- FV arıza sınıflandırma çalışmalarında yalnızca sınıflandırıcı seçimine odaklanılmamalı; veri ön işleme, özellik mühendisliği ve boyut azaltma adımları da modelleme sürecinin temel bileşenleri olarak değerlendirilmelidir.
- Yöntem karşılaştırmalarında yalnızca doğruluk değeri kullanılmamalı; AUC, F1 skoru, duyarlılık, özgüllük, karışıklık matrisi, işlem süresi ve dayanıklılık ölçütleri birlikte ele alınmalıdır.
- KPCA, doğrusal olmayan arıza örüntülerinin belirgin olduğu durumlarda güçlü bir temsil yöntemi olarak değerlendirilebilir. Ancak çekirdek tabanlı yapısı nedeniyle işlem yükü, parametre seçimi ve hesaplama maliyeti dikkate alınmalıdır.
- PCA, sade ve yorumlanabilir yapısı nedeniyle gerçek zamanlı izleme sistemleri veya sınırlı hesaplama kaynağı bulunan uygulamalar için uygun bir alternatif olarak değerlendirilebilir.
- KLDA gibi denetimli çekirdek tabanlı yöntemler, özellikle düşük ışınım gibi belirli çevresel koşullarda dayanıklılık açısından avantaj sağlayabileceğinden, saha uygulamalarında yalnızca genel performansa göre değil, çalışma koşullarına göre de değerlendirilmelidir.
- Bozulmuş senaryoları, FV arıza teşhis sistemlerinin güvenilirliğini değerlendirmek için önemli bir test yaklaşımıdır. Bu nedenle gürültü, eksik veri, sensör sürüklenmesi, tek sensör kaybı ve farklı ışınım bantı koşulları, model değerlendirme sürecine dahil edilmelidir.
- Simülasyon tabanlı sonuçların saha uygulanabilirliğini artırmak için önerilen yaklaşımın gerçek FV santral verileri üzerinde ayrıca doğrulanması gerekmektedir.

Bu tez çalışması simülasyon tabanlı veri seti üzerinde gerçekleştirilmiştir. Simülasyon ortamı, arıza senaryolarının kontrollü, güvenli ve tekrarlanabilir biçimde incelenmesine olanak sağlamıştır. Bununla birlikte gerçek saha koşullarında sensör gürültüsü, ekipman yaşlanması, kısmi gölgelenme, kirlenme, bağlantı gevşemeleri, bakım geçmişi ve çevresel belirsizlikler gibi ek değişkenler bulunabilir. Bu nedenle elde edilen sonuçlar, önerilen yaklaşımın potansiyelini göstermekte; ancak yöntemin farklı saha verileri üzerinde doğrulanması, daha geniş arıza türleriyle test edilmesi ve çevrim içi izleme yapılarıyla bütünleştirilmesi gelecek çalışmalar açısından önem taşımaktadır.

Sonu olarak bu tez, FV sistemlerde arıza sınıflandırma başarısının yalnızca sınıflandırıcı seçimine deęil, veri temsilinin niteliğine ve bozulum koşulları altındaki dayanıklılık performansına da baęlı olduğunu göstermektedir. Doğrusal ve doğrusal olmayan boyut azaltma yöntemleri farklı avantajlar sunmakta; uygun yöntemin seçimi ise doğruluk, AUC, F1 skoru, işlem süresi, yorumlanabilirlik ve saha benzeri koşullar altındaki kararlılık birlikte deęerlendirilerek yapılmalıdır.

## KAYNAKÇA

- Bach, F. R., & Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research*, 3, 1–48.
- Baudat, G., & Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12(10), 2385–2404.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Chen, L. C., Lin, P. J., Zhang, J., Chen, Z. C., Lin, Y. H., Wu, L. J., & Cheng, S. Y. (2018). Fault diagnosis and classification for photovoltaic arrays based on principal component analysis and support vector machine. *IOP Conference Series: Earth and Environmental Science*, 188(1), 012089.
- Chine, W., Mellit, A., Lughi, V., Malek, A., Sulligoi, G., & Massi Pavan, A. (2016). A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renewable Energy*, 90, 501–512.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Dhimish, M., & Holmes, V. (2016). Fault detection algorithm for grid-connected photovoltaic plants. *Solar Energy*, 137, 236–245.
- El-Banby, G. M., Moawad, N. M., Abouzalm, B. A., & Ramadan, E. A. (2023). Photovoltaic system fault detection techniques: A review. *Neural Computing and Applications*, 36, 1451–1492.
- Eskandari, A., Milimonfared, J., & Aghaei, M. (2021). Fault detection and classification for photovoltaic systems based on hierarchical classification and machine learning technique. *IEEE Transactions on Industrial Electronics*, 68(12), 12750–12759.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179–188.
- Garoudja, E., Chouder, A., Kara, K., & Silvestre, S. (2017). An enhanced machine learning based approach for failures detection and diagnosis of PV systems. *Energy Conversion and Management*, 151, 496–513.
- Ghoneim, S. S. M., Rashed, A. E., & Elkalashy, N. I. (2021). Fault detection algorithms for achieving service continuity in photovoltaic farms. *Intelligent Automation & Soft Computing*, 30(2), 467–479. <https://doi.org/10.32604/iasc.2021.016681>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2. baskı). Springer.
- Hong, Y. Y., & Pula, R. A. (2022). Methods of photovoltaic fault detection and classification: A review. *Energy Reports*, 8, 5898–5929.
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. Wiley.
- IEA-PVPS. (2024). *Trends in photovoltaic applications 2024*. International Energy Agency Photovoltaic Power Systems Programme.
- IEC. (2016). *Photovoltaic systems — Requirements for testing, documentation and maintenance — Part 1: Grid connected systems — Documentation, commissioning tests and inspection* (IEC 62446-1:2016). International Electrotechnical Commission.
- IEC. (2021). *Photovoltaic system performance — Part 1: Monitoring* (IEC 61724-1:2021). International Electrotechnical Commission.
- IRENA. (2024). *Renewable capacity statistics 2024*. International Renewable Energy Agency.
- Jolliffe, I. T. (2002). *Principal component analysis* (2. baskı). Springer.
- Madeti, S. R., & Singh, S. N. (2017). Monitoring system for photovoltaic plants: A review. *Renewable and Sustainable Energy Reviews*, 67, 1180–1207.

- Mellit, A., Tina, G. M., & Kalogirou, S. A. (2018). Fault detection and diagnosis methods for photovoltaic systems: A review. *Renewable and Sustainable Energy Reviews*, 91, 1–17.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rashed, A. E. (2022). Fault-Detection-Dataset-in-Photovoltaic-Farms [MATLAB codes and dataset]. GitHub. <https://github.com/amrrashed/Fault-Detection-Dataset-in-Photovoltaic-Farms>
- Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299–1319.
- Sepúlveda-Oviedo, E. H., Travé-Massuyès, L., & Valdés-González, A. (2023). Fault diagnosis of photovoltaic systems using artificial intelligence techniques: A bibliometric approach. *Heliyon*, 9(11), e21491.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- Triki-Lahiani, A., Bennani-Ben Abdelghani, A., & Slama-Belkhodja, I. (2018). Fault detection and monitoring systems for photovoltaic installations: A review. *Renewable and Sustainable Energy Reviews*, 82, 2680–2692.
- Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.
- Villalva, M. G., Gazoli, J. R., & Filho, E. R. (2009). Comprehensive approach to modeling and simulation of photovoltaic arrays. *IEEE Transactions on Power Electronics*, 24(5), 1198–1208.

