

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

SES DOSYALARINA YÖNELİK YAPILAN KOPYALA YAPIŞTIR SES
SAHTECİLİKLERİNİN TESPİTİ

YÜKSEK LİSANS TEZİ

Muhammed KILIÇ

AĞUSTOS 2024
TRABZON



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

SES DOSYALARINA YÖNELİK YAPILAN KOPYALA YAPIŞTIR SES
SAHTECİLİKLERİNİN TESPİTİ

Muhammed KILIÇ

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

Tezin Enstitüye Verildiği Tarih : 17 / 07 / 2024

Tezin Savunma Tarihi : 05 / 08 / 2024

Tez Danışmanı : Prof. Dr. Güzin ULUTAŞ

Trabzon 2024

ÖNSÖZ

“Ses Dosyalarına Yönelik Yapılan Kopyala Yapıştır Ses Sahteciliklerinin Tespiti” başlıklı bu tez, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Programı kapsamında hazırlanmıştır.

Tez çalışmam süresince hiçbir desteğini esirgemeyen, bilim insanı olma yolundaki ilk adımlarımda kıymetli bilgileri ve tecrübeleriyle yolumu aydınlatan değerli danışman hocam Sayın Prof. Dr. Güzin ULUTAŞ’a en içten teşekkürlerimi sunarım. Akademik hayatımda büyük katkıları olan Dr. Öğr. Üyesi Gül TAHAOĞLU, Dr. Öğr. Üyesi Beste ÜSTÜBİOĞLU ve Dr. Öğr. Üyesi Arda ÜSTÜBİOĞLU’na da minnettarım. Tez çalışması sırasında gerçekleştirilen 122E013 numaralı 1001 projesi ile maddi olarak beni destekleyen Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK)’a teşekkürlerimi sunarım.

Doğduğum günden beri desteğini esirgemeyen, hayatım boyunca aldığım tüm kararlarda bilgi ve tecrübelerinden faydalandığım, maddi ve manevi desteklerini hiçbir zaman eksik etmeyen ve bugünlere gelmemde en büyük paya sahip olan aileme sonsuz teşekkürlerimi ve minnettarlığımı ifade etmek isterim.

Bu tez çalışmasının, ileride yapılacak çalışmalara katkı sağlamasını temenni ederim.

Muhammed KILIÇ

Trabzon 2024

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Ses Dosyalarına Yönelik Yapılan Kopyala Yapıştır Ses Sahteciliklerinin Tespiti” başlıklı bu çalışmayı baştan sona kadar danışmanım Prof. Dr. Güzin ULUTAŞ’ın sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 17/07/2024

Muhammed KILIÇ

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	IX
ŞEKİLLER DİZİNİ.....	XI
TABLolar DİZİNİ.....	XIV
SEMBOLLER VE KISALTMALAR DİZİNİ	XVIII
1. GENEL BİLGİLER	1
1.1. Giriş.....	1
1.2. Ses Sahtecilikleri	2
1.2.1. Derin Sahte Ses Sahteciliği	2
1.2.1.1. Literatür Araştırması	3
1.2.2. Ses Montaj Sahteciliği	5
1.2.2.1. Literatür Araştırması	6
1.2.2.1.1. Geleneksel Yöntemler.....	6
1.2.2.1.2. Derin Öğrenme Tabanlı Yöntemler	7
1.2.3. Kopyala Yapıştır Ses Sahteciliği.....	9
1.2.3.1. Literatür Araştırması	10
1.2.3.1.1. Pencere Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti.....	10
1.2.3.1.2. VAD Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti.....	11
1.2.3.1.3. Spektrogram Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti. 14	
1.3. Kopyala Yapıştır Sahte Ses Tespitinde Kullanılan Yöntemler	16
1.3.1. Yüksek Çözünürlüklü Spektrogram.....	16
1.3.2. Anahtar Nokta ve Tanımlayıcıların Çıkarılması	20
1.3.2.1. Ölçekten Bağımsız Öznitelik Dönüşümü (SIFT)	20
1.3.2.2. SURF Yöntemi	24
1.3.2.3. BRIEF Anahtar Nokta ve Tanımlayıcıları	26
1.3.2.4. AKAZE	29
1.3.3. Özellik Tanımlayıcıların Kümelenendirilmesi Yaklaşımları.....	34
1.3.3.1. K-means kümeleme.....	34

1.3.3.2. Genelleştirilmiş 2 En Yakın Komşu.....	36
1.3.3.3. OPTICS	37
1.3.4. Hatalı Eşleşmelerin Ortadan Kaldırılması	41
1.3.4.1. RANSAC Metodu	41
1.3.4.2. Parametre Normalizasyon Metodu.....	42
1.3.5. Perde İzleme Yöntemi.....	43
1.3.5.1. Konuşma Sinyali Ön İşleme.....	43
1.3.5.2. Frekans Alanında Perde Dizisi Tahmini	43
1.3.5.3. Zaman Alanında Perde Dizisi Tahmini.....	44
1.3.5.4. Dinamik Programlama ile Perde Dizisi Belirleme	45
1.3.6. Graf Teorisi	45
2. YAPILAN ÇALIŞMALAR.....	48
2.1. Giriş.....	48
2.2. Anahtar Noktalarının Kümelendirilmesi Yaklaşımıyla Kopyala Yapıştır Ses Sahteciliği Tespiti	49
2.2.1. OPTICS Kümeleme Yaklaşımı Kullanımı.....	49
2.2.1.1. BRIEF Anahtar Noktası Metodu.....	50
2.2.1.2. AKAZE Anahtar Noktası Metodu.....	53
2.2.1.3. SIFT Anahtar Noktası Metodu.....	56
2.2.1.4. SURF Anahtar Noktası Metodu	60
2.2.2. BRIEF Anahtar Noktalarının g2NN Kümeleme Yaklaşımıyla Kullanılması	63
2.2.3. BRIEF Anahtar Noktalarının k-Means Kümeleme Yaklaşımıyla Kullanılması	65
2.3. Graf Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti	69
2.3.1. Yüksek Çözünürlüklü Spektrogram Görüntülerinden SIFT Anahtar Noktası Çıkarılması.....	70
2.3.2. g2NN ile Anahtar Noktalarının Kümelendirilmesi	71
2.3.3. Hatalı Eşleşmelerin Ortadan Kaldırılması.....	72
2.3.3. Yoğun Kümelenme Bölgelerinin Tespit Edilmesi	73
2.3.4. Band Geçiren Süzgeç İşlemi	74
2.3.5. Hızlı Fourier Transform Uygulaması.....	75
2.3.6. Frekans Uzayından Öznitelik Çıkarılması	76
2.3.7. Graf Görüntüsünün Çıkarılması.....	77
2.3.8. Graf Görüntüsünün Sınıflandırılarak Kopyala Yapıştır Sahte Ses Tespiti Yapılması	79

2.3.8.1. Veriseti Hazırlanması	80
2.3.8.2. Önerilen CNN Modeliyle Sınıflandırma	82
2.3.8.3. Transfer Öğrenme Modelleri.....	86
2.3.8.3.1. VGG16 Kullanımı.....	86
2.3.8.3.2. VGG19 Kullanımı.....	90
2.3.8.4. Füzyon Uygulanması.....	93
3. BULGULAR VE TARTIŞMA	94
3.1. Değerlendirme Metrikleri.....	94
3.2. Anahtar Noktalarının Kümelendirilmesi Yaklaşımı Deneysel Sonuçları	95
3.2.1. Kullanılan Verisetleri	95
3.2.1.1. TIMIT Kopyala-Yapıştır Veriseti	95
3.2.1.2. Arabic Kopyala Yapıştır Ses Veriseti	96
3.2.2. BRIEF Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi.....	97
3.2.3. AKAZE Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi.....	100
3.2.4. SIFT Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi.....	102
3.2.5. SURF Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi.....	105
3.2.6. BRIEF Anahtar Noktalarının g2NN Yaklaşımı ile Kümelendirilmesi.....	108
3.2.7. BRIEF Anahtar Noktalarının k-Means Yaklaşımı ile Kümelendirilmesi.....	111
3.2.8. Önerilen Yöntemlerin Literatür ile Karşılaştırılması	115
3.3. Graf Tabanlı Yöntemin Deneysel Sonuçları	119
3.3.1. Kullanılan Verisetleri	119
3.3.1.1. Arabic Artırımsız Eğitim Seti	119
3.3.1.2. Arabic Artırmalı Eğitim Seti.....	120
3.3.1.3. Arabic Test Seti	120
3.3.1.4. TIMIT Test Seti.....	121
3.3.1.5. Türkçe Test Seti	121
3.3.1.6. Librispeech Test Seti	122
3.3.1.7. Chinspeech Test Seti	123
3.3.2. Graf Tabanlı Yöntemin Performans Sonuçları	125

3.3.2.1. Artırımsız Veriseti Kullanılarak Eğitilen Mimarilerin Deneysel Sonuçları.....	127
3.3.2.2. Artırmalı Veriseti Kullanılarak Eğitilen Mimarilerin Deneysel Sonuçları.....	131
3.3.2.3. Yöntemlerin Birbirleriyle ve Literatür Karşılaştırılması.....	136
4. SONUÇLAR.....	149
5. ÖNERİLER.....	150
6. KAYNAKLAR	151
ÖZGEÇMİŞ.....	160



ÖZET

**SES DOSYALARINA YÖNELİK YAPILAN KOPYALA YAPIŞTIR SES
SAHTECİLİKLERİNİN TESPİTİ**

Muhammed KILIÇ

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Güzin ULUTAŞ
2024, 148 Sayfa

Günümüzde haberleşme, adli süreçler ve sesli komutla çalışan sistemlerde kullanılan ses kayıtlarının doğruluğu ve güvenilirliği büyük önem taşımaktadır. Ses kayıtları, kişilerin izni dahilinde veya rızası olmadan elde edilebilmektedir ve bu kayıtlar sahte ses üretimi amacıyla kullanılabilir. Ses üzerinde yapılan sahtecilik işlemlerinden en yaygın ve gerçekleştirilmesi kolay olanı kopyala yapıştır işlemidir.

Kopyala yapıştır ses sahteciliği tespiti için iki yöntem önerilmiştir: anahtar noktalarının kümelendirilmesi ve sesten elde edilen graf görüntülerinin sınıflandırılması. İlk yöntemde, BRIEF, AKAZE, SIFT ve SURF teknikleri kullanılarak anahtar noktalar ve tanımlayıcılar çıkarılmış, ardından bu tanımlayıcılar OPTICS, k-Means ve g2NN algoritmaları ile kümelendirilmiştir. Bu yöntem, TIMIT verisetinde ataksız sahte seslerde ve çeşitli saldırı senaryolarında ortalama 0.90 duyarlılık elde etmiştir. İkinci yöntemde, yüksek çözünürlüklü spektrogramdan elde edilen SIFT anahtar noktaları g2NN kullanılarak kümelendirilmiş, yoğun eşleşme bölgeleri tespit edildikten sonra bu bölgeye ait frekans aralığı belirlenmiş ve ses, band geçiren süzgeçten geçirilmiştir. Hızlı Fourier Transform ile frekans uzayına geçilerek, önerilen sarmal yapı sayesinde çerçeve bazında özellikler çıkarılmış ve bu özelliklerden graf görüntüleri oluşturulmuştur. Graf görüntülerinin sınıflandırılmasıyla, sesler sahte veya orijinal olarak etiketlenmiş ve bu yöntem Arabic, TIMIT ve Türkçe veri setlerinde sırasıyla 0.95, 0.87 ve 0.92 F-skoru ile başarı göstermiştir.

Anahtar Kelimeler: Kopyala Yapıştır Ses Sahteciliği, Yüksek Çözünürlüklü Spektrogram, Anahtar Noktası, Graf Görüntüsü

SUMMARY

Detection of Copy Move Audio Forgeries Targeting Audio Files

Muhammed KILIÇ

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Computer Engineering Graduate Program
Supervisor: Prof. Dr. Güzin ULUTAŞ
2024, 148 Pages

In today's world, the accuracy and reliability of audio recordings used in communication, judicial processes, and voice-activated systems are of great importance. Audio recordings can be obtained with or without the consent of individuals, and these recordings can be used to create fake audio. The most common and easily performed type of audio forgery is the copy-paste operation.

Two methods have been proposed for detecting copy move audio forgery: clustering of key points and classification of graph images derived from audio. In the first method, key points and descriptors were extracted using BRIEF, AKAZE, SIFT, and SURF techniques, and these descriptors were then clustered using the OPTICS, k-Means, and g2NN algorithms. This method achieved an average recall of 0.90 on the TIMIT dataset for no post-processing fake audios and various attack scenarios. In the second method, SIFT key points extracted from a high-resolution spectrogram were clustered using g2NN, and after identifying regions with dense matches, the corresponding frequency range was determined, and the audio was processed with a band-pass filter. By applying the Fast Fourier Transform, the frequency domain was accessed, and frame-by-frame features were extracted using the proposed spiral structure. These features were used to generate graph images, which were then classified to label the audio as either fake or original. This methods show F-scores of 0.95, 0.87, and 0.92 on the Arabic, TIMIT, and Turkish datasets, respectively.

Key Words: Copy-move Audio Forgery, High-Resolution Spectrogram, Key Point, Graph Image

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Ses Montaj Sahteciliği Oluşturulma Adımları	6
Şekil 2. (a) Orijinal Ses (b) Kopyala Yapıştır Sahte Ses	9
Şekil 3. Graf Örneği	46
Şekil 4. Renklendirilmemiş Graf Örneği	47
Şekil 5. Renklendirilmiş Graf Örneği	48
Şekil 6. Yüksek çözünürlüklü spektrogram görüntüsü (a) Orijinal ses (b) Sahte ses	49
Şekil 7. Merkez Çevre Ekstrema tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	51
Şekil 8. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses	52
Şekil 9. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses	52
Şekil 10. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması	53
Şekil 11. AKAZE tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	54
Şekil 12. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses	55
Şekil 13. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses	55
Şekil 14. RANSAC tarafından hatalı eşleşmelerin kaldırılması (a) Orijinal ses (b) Sahte ses	56
Şekil 15. SIFT tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	58
Şekil 16. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses	58
Şekil 17. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses	59
Şekil 18. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması	60
Şekil 19. SURF tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	61
Şekil 20. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses	61
Şekil 21. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses	62
Şekil 22. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması	63
Şekil 23. Merkez Çevre Ekstrema tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	64

Şekil 24. g2NN tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses.....	65
Şekil 25. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması	65
Şekil 26. Merkez Çevre Ekstrema tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	66
Şekil 27. k-Means tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses.....	67
Şekil 28. k-Means tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses.....	67
Şekil 29. Sahte sesin dalga formu ve Perde dizisi (a) Dalga formu (b) Perde dizisi	68
Şekil 30. Yüksek Çözünürlüklü Spektrogram (a) Orijinal ses (b) Sahte ses	70
Şekil 31. SIFT tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses	71
Şekil 32. g2NN tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses.....	72
Şekil 33. Parametre Normalizasyonu tarafından hatalı eşleşmelerin kaldırılması (a) Orijinal ses (b) Sahte ses.....	73
Şekil 34. Sahte spektrogram üzerinde yoğun eşleşme bölgelerinin tespit edilmesi	74
Şekil 35. (a) Sahte sese ait dalga formu (b) Sahte sese ait band geçiren süzgeç işlemi uygulanmış hali	74
Şekil 36. Sahte spektrogram üzerinde eşleşmeler (b). Band geçiren süzgeç işleminden sonra spektrogram görüntüsü	75
Şekil 37. Sesin frekans uzayı (a) Band geçiren süzgeç işlemi uygulanmamış (b) Band geçiren süzgeç işlemi uygulanmış	76
Şekil 38. Spiral örüntü yöntemiyle bilgi üretimi	77
Şekil 39. Örnek bir öznitelik	78
Şekil 40. Örnek öznitelige ait graf görüntüsü (a) Renksiz (b) Renklendirilmiş	78
Şekil 41. Graf görüntüsü (a) Orijinal ses (b) Sahte ses	79
Şekil 42. Önerilen mimari	82
Şekil 43. Eğitim ve geçerleme doğruluk grafikleri	84
Şekil 44. Eğitim ve geçerleme yitim grafikleri	84
Şekil 45. Eğitim ve geçerleme doğruluk grafikleri	85
Şekil 46. Eğitim ve geçerleme yitim grafikleri	85
Şekil 47. Önerilen VGG16 Modeli.....	86
Şekil 48. Eğitim ve geçerleme doğruluk grafikleri	87
Şekil 49. Eğitim ve geçerleme yitim grafikleri	88
Şekil 50. Eğitim ve geçerleme doğruluk grafikleri	89
Şekil 51. Eğitim ve geçerleme yitim grafikleri	89
Şekil 52. Önerilen VGG19 modeli	90

Şekil 53. Eğitim ve geçerleme doğruluk grafikleri	91
Şekil 54. Eğitim ve geçerleme yitim grafikleri	91
Şekil 55. Eğitim ve geçerleme doğruluk grafikleri	92
Şekil 56. Eğitim ve geçerleme yitim grafikleri	92
Şekil 57. Füzyon grafiği	94
Şekil 58. Türkçe verisetine ilişkin orijinal seslerin süre dağılımı	121



TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Orijinal graf görüntülerine uygulanan veri artırım yöntemleri	81
Tablo 2. Sahte graf görüntülerine uygulanan veri artırım yöntemleri	81
Tablo 3. Anahtar Noktaların Kümelenendirilmesi Yaklaşımı için oluşturulmuş TIMIT kopyala yapıştır veriseti özeti	96
Tablo 4. Anahtar Noktaların Kümelenendirilmesi Yaklaşımı için oluşturulmuş Arabic kopyala yapıştır veriseti özeti	97
Tablo 5. BRIEF ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler	97
Tablo 6. BRIEF ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları	98
Tablo 7. TIMIT kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları	99
Tablo 8. Arabic kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları	99
Tablo 9. AKAZE ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler	100
Tablo 10. AKAZE ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları	101
Tablo 11. TIMIT veriseti üzerinde yöntemin deneysel sonuçları	102
Tablo 12. Arabic veriseti üzerinde yöntemin deneysel sonuçları	102
Tablo 13. SIFT ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler	103
Tablo 14. SIFT ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları	104
Tablo 15. TIMIT veriseti üzerinde yöntemin deneysel sonuçları	105
Tablo 16. Arabic veriseti üzerinde yöntemin deneysel sonuçları	105
Tablo 17. SURF ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler	106
Tablo 18. SURF ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları	107
Tablo 19. TIMIT veriseti üzerinde yöntemin deneysel sonuçları	108
Tablo 20. Arabic veriseti üzerinde yöntemin deneysel sonuçları	108
Tablo 21. BRIEF ve g2NN'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler	109
Tablo 22. BRIEF ve g2NN'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları	110

Tablo 23. TIMIT veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları.....	111
Tablo 24. Arabic veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları.....	111
Tablo 25. BRIEF ve k-Means'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları.....	112
Tablo 26. BRIEF ve k-Means'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler	113
Tablo 27. TIMIT veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları.....	114
Tablo 28. Arabic veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları.....	115
Tablo 29. TIMIT veriseti üzerinde 6 farklı yöntemin deneysel sonuçları.....	115
Tablo 30. TIMIT veriseti üzerinde Duyarlılık metriği kullanılarak BRIEF ve OPTICS'in beraber kullanıldığı yöntemin literatür sonuçları	117
Tablo 31. Arabic veriseti üzerinde 6 farklı yöntemin deneysel sonuçları	118
Tablo 32. Graf tabanlı yöntem için oluşturulmuş Arabic kopyala yapıştır eğitim veriseti özeti.....	119
Tablo 33. Graf tabanlı yöntem için oluşturulmuş Artırımlı Arabic kopyala yapıştır eğitim veriseti özeti.....	120
Tablo 34. Graf tabanlı yöntem için oluşturulmuş Arabic kopyala yapıştır test veriseti özeti.....	120
Tablo 35. Graf tabanlı yöntem için oluşturulmuş TIMIT kopyala yapıştır test veriseti özeti.....	121
Tablo 36. Graf tabanlı yöntem için oluşturulmuş Türkçe kopyala yapıştır test veriseti özeti.....	122
Tablo 37. Graf tabanlı yöntemin test edilmesinde kullanılan Librispeech V1 test veriseti özeti.....	122
Tablo 38. Graf tabanlı yöntemin test edilmesinde kullanılan Librispeech V2 test veriseti özeti.....	123
Tablo 39. Graf tabanlı yöntemin test edilmesinde kullanılan Librispeech V3 test veriseti özeti.....	123
Tablo 40. Graf tabanlı yöntemin test edilmesinde kullanılan Chinspeech V1 test veriseti özeti.....	124
Tablo 41. Graf tabanlı yöntemin test edilmesinde kullanılan Chinspeech V2 test veriseti özeti.....	124
Tablo 42. Graf tabanlı yöntemin test edilmesinde kullanılan Chinspeech V3 test veriseti özeti.....	125
Tablo 43. Yüksek çözünürlüklü spektrogram görüntüleri üzerinde eşleşme doğruları ve graf görselleri	125

Tablo 44. Arabic test seti üzerinde Füzyon işleminin performans sonuçları	127
Tablo 45. TIMIT test seti üzerinde Füzyon işleminin performans sonuçları	128
Tablo 46. Türkçe test seti üzerinde Füzyon işleminin performans sonuçları	128
Tablo 47. Librispeech V1 test seti üzerinde Füzyon işleminin performans sonuçları	129
Tablo 48. Librispeech V2 test seti üzerinde Füzyon işleminin performans sonuçları	129
Tablo 49. Librispeech V3 test seti üzerinde Füzyon işleminin performans sonuçları	130
Tablo 50. Chinspeech V1 test seti üzerinde Füzyon işleminin performans sonuçları	130
Tablo 51. Chinspeech V2 test seti üzerinde Füzyon işleminin performans sonuçları	131
Tablo 52. Chinspeech V3 test seti üzerinde Füzyon işleminin performans sonuçları	131
Tablo 53. Arabic test seti üzerinde Füzyon işleminin performans sonuçları	132
Tablo 54. TIMIT test seti üzerinde Füzyon işleminin performans sonuçları	133
Tablo 55. Türkçe test seti üzerinde Füzyon işleminin performans sonuçları	133
Tablo 56. Librispeech V1 test seti üzerinde Füzyon işleminin performans sonuçları	134
Tablo 57. Librispeech V2 test seti üzerinde Füzyon işleminin performans sonuçları	134
Tablo 58. Librispeech V3 test seti üzerinde Füzyon işleminin performans sonuçları	135
Tablo 59. Chinspeech V1 test seti üzerinde Füzyon işleminin performans sonuçları	135
Tablo 60. Chinspeech V2 test seti üzerinde Füzyon işleminin performans sonuçları	136
Tablo 61. Chinspeech V3 test seti üzerinde Füzyon işleminin performans sonuçları	136
Tablo 62. Artırımlı ve artırısız eğitim seti ile eğitilmiş mimarilerin Arabic test seti üzerinde F-skor metriği ile füzyon performans sonuçları	137
Tablo 63. Artırımlı eğitim seti ile eğitilmiş mimarinin Arabic test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri	138
Tablo 64. Artırımlı ve artırısız eğitim seti ile eğitilmiş mimarilerin TIMIT test seti üzerinde F-skor metriği ile füzyon performans sonuçları	138
Tablo 65. Artırımlı eğitim seti ile eğitilmiş mimarinin TIMIT test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri	139
Tablo 66. Artırımlı ve artırısız eğitim seti ile eğitilmiş mimarilerin Türkçe test seti üzerinde F-skor metriği ile füzyon performans sonuçları	140

Tablo 67. Artırımı eğitim seti ile eğitilmiş mimarinin Türkçe test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri.....	140
Tablo 68. Artırımı ve artırımı eğitim seti ile eğitilmiş mimarilerin Librispeech V1 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	141
Tablo 69. Artırımı eğitim seti ile eğitilmiş mimarinin Librispeech V1 test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri	142
Tablo 70. Artırımı ve artırımı eğitim seti ile eğitilmiş mimarilerin Librispeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	142
Tablo 71. Artırımı eğitim seti ile eğitilmiş mimarilerin Librispeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	143
Tablo 72. Artırımı ve artırımı eğitim seti ile eğitilmiş mimarilerin Librispeech V3 test seti üzerinde Duyarlılık metriği ile füzyon performans sonuçları	144
Tablo 73. Artırımı eğitim seti ile eğitilmiş mimarilerin Librispeech V3 test seti üzerinde Duyarlılık metriği ile füzyon performans sonuçları.....	144
Tablo 74. Artırımı ve artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V1 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	145
Tablo 75. Artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V1 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	145
Tablo 76. Artırımı ve artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	146
Tablo 77. Artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları.....	147
Tablo 78. Artırımı ve artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V3 test seti üzerinde Duyarlılık metriği ile füzyon performans sonuçları	147
Tablo 79. Artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V3 test seti üzerinde Duyarlılık metriği ile füzyon performans sonuçları.....	148

SEMBOLLER VE KISALTMALAR DİZİNİ

AD	: Ortalama Fark
ADFA	: Keyfi Ayrık Fourier Analizi
CQCC	: Sabit-Q Kepstral Katsayıları
CQT	: Sabit-Q Transform
CSR	: Sürekli Konuşma Tanıma
DCT	: Ayrık Kosinüs Dönüşümü
DFT	: Ayrık Fourier Transform
DTW	: Dinamik Zaman Atlama
ED	: Öklid Uzaklığı
ER	: Hata oranı
FIR	: Sonlu Dürtü Tepkisi
GMM	: Gaussian Karışım Modeli
GTCC	: Gammatone Kepstral Katsayıları
LBP	: Yerel İkili Örüntü
LFB	: Doğrusal Filtre Bankaları
LFCC	: Doğrusal Frekanslı Kepstral Katsayıları
LP	: Doğrusal Tahmin
LSTM	: Uzun Kısa Vadeli Bellek
LT	: Uzun Vadeli
MDCT	: Modifiye Ayrık Kosinüs Dönüşümü
MFCC	: Mel-frekans Kepstral Katsayıları
MSE	: Ortalama Kare Hatası
PCC	: Pearson Korelasyon Katsayısı
ST	: Kısa Vadeli
STFT	: Kısa Zamanlı Fourier Dönüşümü
SVD	: Tekil Değer Ayrıştırma
TTS	: Metinden Konuşmaya
VAD	: Konuşma Etkinliği Algılama
VC	: Ses Dönüşüm
ViT	: Görüntü Transformatörü

1. GENEL BİLGİLER

1.1. Giriş

Günümüzde, dijital teknolojilerin hızla gelişmesiyle birlikte, ses sahtecilikleri büyük bir sorun haline gelmektedir. Ses sahtecilikleri, insan sesini manipüle etme veya taklit etme sürecidir ve genellikle kötü niyetli amaçlarla kullanılmaktadır. Bu kötü niyetli amaçlar arasında sahte haberlerin yayılması, kişisel verilerin dolandırıcılık amaçlarıyla kullanılması ve manipüle edilmiş ses kayıtlarının yargı süreçlerinde delil olarak sunulması gibi durumlar bulunmaktadır.

Geleneksel ses sahtecilikleri teknikleri, ses kayıtları içerisindeki kelime veya kelime gruplarını kopyalayıp yapıştırma, farklı sesleri birleştirme yöntemleriyle gerçekçi olmayan sesler oluşturmayı içermektedir. Ancak, son yıllarda yapay zeka ve derin öğrenme tekniklerinin ilerlemesiyle birlikte, ses sahtecilikleri daha sofistike ve gerçekçi hale gelmiştir. Günümüzde sahte sesler üretmek için kullanılan algoritmalar, bir kişinin sesini gerçekçi şekilde taklit etmek veya tamamen yapay sesler oluşturmak için oldukça gelişmiştir.

Ses sahteciliklerine karşı sistemler geliştirmek için, bu türdeki sesleri tespit etmek konusunda yeni araçlar ve yöntemler geliştirilmektedir. Bu yöntemler arasında ses biyometrisi, konuşma analitiği ve derin öğrenme tabanlı ses sahtecilik tespit sistemleri bulunmaktadır. Ancak, sahte ses üreten teknolojilerin hızla gelişmesi, sahte ses tespit sistemlerini sürekli olarak güncellemeyi ve iyileştirmeyi zorunlu kılmaktadır.

Teknolojinin günlük hayatımızın her alanına yayılması, birçok konuda kolaylığı ve yeniliği getirirken, aynı zamanda bazı riskleri de beraberinde taşımaktadır. Özellikle haberleşme alanında sağladığı hız ve kolaylık, yaşamımıza büyük avantajlar sunmaktadır. Örneğin, ses kayıt cihazlarının yaygınlaşması, mahkemelerde delil olarak kullanılabilen güvenilir teknoloji olarak öne çıkmaktadır. Ancak, bu teknoloji aynı zamanda kötü niyetli kişilerin suistimaline açık alan da oluşturmaktadır. Kişilerin izni ile veya izni dışında elde edilen ses kayıtları mevcut teknolojiler vasıtasıyla sahte ses üretmek amacıyla kullanılmaktadır. Günümüzde, ses sahtecilikleri genellikle Derin Sahte Ses (Deep Fake Audio), Ses Montajı (Audio Splicing) ve Kopyala Yapıştır Ses Sahteciliği (Copy Move Audio Forgery) gibi üç ana başlık altında incelenmektedir.

Dijital ses doğrulama yöntemleri, aktif doğrulama ve pasif doğrulama olmak üzere iki ana kategoriye ayrılır. Aktif doğrulama teknikleri, ses verilerine gizlice yerleştirilen dijital filigran (watermark) ve imza (signature) yöntemlerini içerir. Bu teknikler, filigran veya imzanın ses dosyasından yeniden üretilmesini sağlar ve ardından elde edilen filigran veya imza, kimlik doğrulama sürecinde kullanılır. Eğer yeniden üretilen filigran veya imza orijinal veriyle tutarlı ise, ses kaydının orijinal olduğu kabul edilir; aksi durumda ise sahte olduğu düşünülür. Ancak, günümüzde çoğu ses kaydında filigran veya imza bulunmaması nedeniyle aktif doğrulama yöntemlerinin kullanımı sınırlıdır.

Pasif ses doğrulama tekniği, herhangi bir filigran veya imza kullanmadan doğrudan ses dosyasının özelliklerini çıkararak çalışır. Pasif doğrulama yöntemleri, içerik (content) ve dosya kapsayıcı tabanlı (container-based) olmak üzere iki ana sınıfa ayrılır. Onaltılık veriler (hex data), dosya formatı (file format) ve ses dosyası başlığı (audio file header) gibi özellikler, dosya kapsayıcı tabanlı analizin kapsamına girer. Bununla birlikte, ses dosyasının hash ve zaman etiketi tabanlı (time stamp-based) özellikleri, bütünlük kapsayıcısı tabanlı (integrity container-based) analizle ilişkilidir. İçerik analizi (Content analysis), dijital ses adli analiz sürecinin temelini oluşturur ve sahtecilik ve saldırı izlerini tespit etmek için ses dosyasının gerçek bitlerini ve baytlarını kullanır.

1.2. Ses Sahtecilikleri

Ses sahteciliği, üç ana başlık altında incelenmektedir: derin sahte ses, ses montajı ve kopyala yapıştır ses sahteciliği. Bu bölümde, her bir sahtecilik türü detaylandırılarak, literatürde yer alan tespit yöntemleri açıklanacaktır. Ayrıca, kopyala yapıştır ses sahteciliğinin tespiti için kullanılan veri setleri de bu kısımda ayrıntılı olarak ele alınacaktır.

1.2.1. Derin Sahte Ses Sahteciliği

Derin Sahte Ses (Deep Fake Audio) sahteciliği, yapay zeka ve derin öğrenme tekniklerinin kullanılmasıyla gerçekçi ve manipüle edilmiş seslerin üretilmesi sürecidir (Lyu, 2020). Bu teknikler, kişinin sesini taklit etmek veya tamamen yapay sesler oluşturmak için kullanılabilir. Derin Sahte Ses Sahteciliği, özellikle ses tabanlı dolandırıcılık, sahte haberlerin yayılması ve manipüle edilmiş ses kayıtlarının yargı süreçlerinde delil olarak sunulması gibi kötü niyetli amaçlar için kullanılabilir. Bu teknoloji, gerçeklik algısını

zorlayan ve ses kaynaklarının güvenilirliğini sorgulatan yeni bir güvenlik tehdidi oluşturabilir.

Derin sahte ses manipölasyonu saldırıları, sahte ses dosyasının oluşturulmasındaki tekniklere göre gruplandırılabilir. En popüler saldırı türleri; Yeniden oynatma (Replay), Metinden Konuşmaya (Text-to-Speech, TTS) ve Ses Dönüşüm (Voice Conversion, VC) olup Aldatma Saldırıları sınıfında değerlendirilmektedirler. Aldatma saldırıları çeşitli yöntemlerle gerçekleştirilebilir. Ses dönüşüm saldırılarında, hedeflenen kişinin sesi analiz edilerek özellikleri belirlenir (Walczyna & Piotrowski, 2023). Saldırgan, sisteme herhangi ses kaydını verdiğinde, sistem bu sesi özelliklerine göre işler ve akustik modelleme yöntemleriyle hedef kişinin sesine benzetir.

İkinci saldırı yöntemi olan Metinden Konuşmaya saldırılarında hedef kişinin söylemesi istenen metin sisteme verilir (Wang, Skerry, Ryan, Stanton, Wu, Weiss, Jaitly, & Saurous, 2017). Sistem, metni analiz ederek kelimeleri ayırır ve dil bilgisi kurallarını inceleyerek cümle yapılarını belirler. Daha sonra, bu metin hedef kişinin sesine dönüştürülerek sahte ses oluşturulur. Bu saldırı yöntemi, genellikle Konuşma Sentezi saldırılarını içerir.

Yeniden Oynatma saldırılarında hedef kişinin sesi gizlice kayıt cihazı kullanılarak kaydedilir (Kinnunen, Sahidullah, Falcone, Costantini, Hautamäki, Thomsen, & Lee, 2017). Daha sonra, bu kaydedilen sesin yeniden oynatılmasıyla sahte ses oluşturulur. Bu şekilde, hedef kişinin sesi taklit edilerek aldatma amaçlı kullanılabilir.

1.2.1.1. Literatür Araştırması

Derin sahte ses tespitine ilişkin literatürde yapılan çalışmalar Geleneksel Yöntemler ile elde edilen özellikler ve Derin öğrenme yöntemleri kullanılarak yapılan çalışmalar olarak iki kategoride incelenmektedir. Literatürde çok fazla çalışma mevcut olmakla birlikte başlıcaları bu tez kapsamında incelenmiştir.

Zhang ve arkadaşlarının önerdiği çalışmada (Zhang vd., 2021), giriş sesinden 60 boyutlu LFCC özellikleri çıkarılmıştır. Bu özelliklerin sınıflandırılmasında, Küresel Ortalama Havuzlama Katmanı'nın yerine Dikkatli Zamansal Havuzlama ile değiştirilmiş ResNet-18 mimarisi kullanılmıştır. Sahte konuşma verilerinin gerçek verilerden belirli bir marjla uzak tutulduğu ve gerçek konuşma özelliklerinin kompakt sınıra sahip olduğu özellik

uzayını öğrenmek için Tek Sınıf Softmax (One Class-Softmax) adlı kayıp fonksiyonunun kullanılması önerilmiştir.

Gupta ve Patil tarafından önerilen çalışma sesten Doğrusal Frekans Artık Kepstral Katsayıları (Linear Frequency Residual Cepstral Coefficients, LFRCC) özelliğinin çıkarılma işlemine dayanmaktadır (Gupta & Patil, 2022). LFRCC özellik setinin çıkarılma aşaması Doğrusal Tahmin (Linear Prediction, LP) işleminin yapılmasıyla başlamaktadır. Ses kaydının Doğrusal Tahmin (Linear Prediction, LP) ile elde edilen artık bilgisi, yeniden oynatma saldırılarını tespit etmek için kepsral alanda kullanılır. LP'nin uygun sırasının seçilmesiyle sonuçlarda iyileştirmeler elde edilmiştir. Sınıflandırma Gaussian Karışım Modeli ve Evrişimli Sinir Ağı kullanılarak yapılmıştır.

Mewada ve arkadaşlarının (2023) önerdikleri çalışmada (Mewada vd., 2023), giriş sesinden Gauss Filtresi Tabanlı MFCC, Gamaton Filtresi bazlı GTCC, Spektral Entropi, Spektral Düzlük ve Pitch Bilgisi gibi akustik özellikler çıkarılmıştır. Bu akustik özelliklerin kombinasyonu, Bayes algoritması tarafından optimize edilmiş BiLSTM kullanılarak derin özellikler elde edilmiş ve sınıflandırılmıştır.

Kwak ve arkadaşlarının önerdiği çalışmada (Kwak vd., 2023), giriş olarak alınan ses sinyallerinin süreleri dokuz saniyeye sabitlenmiştir. Sesin süresi dokuz saniyeden uzun olduğunda, fazla kalan kısmın kesilmesi; kısa olduğunda ise aynı ses dosyasının başına eklenerek genişletme işlemi gerçekleştirilmiştir. Ses sinyalleri, dönüşüm tabanlı yaklaşım olan CQT yöntemi kullanılarak elde edilen özellik vektörleriyle iki boyutta haritalanmıştır. Bu iki boyutta temsil edilen özellik vektörleri, Maksimum özellik haritası konseptini LCNN'den ve bağlantı atlama konseptini ResNet'den birleştiren Maksimum Özellik Haritasına sahip Artık Ağ (ResMax) blokları ile oluşturulan derin ağ mimarisiyle sınıflandırılmıştır. Yazarlar, ResMax mimarisini daha hafif hale getirmek için, evrişim katmanlarını derinlemesine ayrılabilir evrişim katmanlarına (Depthwise Separable Convolution) dönüştürmüşlerdir.

Ulutaş ve arkadaşları tarafından önerilen çalışmada giriş sesinden Sabit-Q Transform (Constant-Q Transform, CQT) akustik özelliği çıkarılmaktadır (Ulutaş vd., 2023). Bu özellik 3 kanallı olarak 2 boyutlu uzaya haritalanmıştır. Sınıflandırma aşamasında Görüntü Transformatörü (Vision Transformer, ViT) tercih edilmiştir.

Tan ve arkadaşlarının önerdikleri çalışmada (Tan vd., 2023), giriş sesinden 60 boyutlu LFCC akustik özelliği çıkarılmıştır. Derin özellik çıkarımı için, Küresel Ortalama Havuzlama Katmanı'nın yerine Dikkatli Zamansal Havuzlama uygulanmış ResNet-18

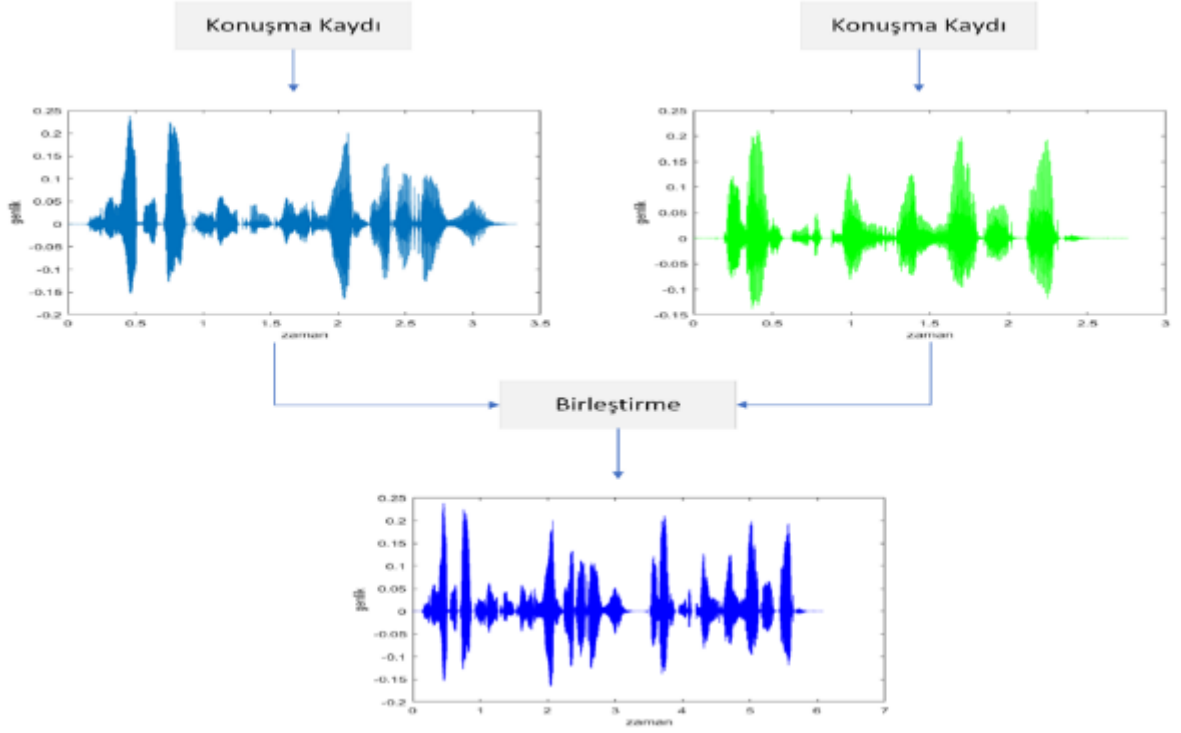
mimarisi önerilmiştir. Sınıflandırma işlemi SVM kullanılarak gerçekleştirilmiştir. SVM'nin Radyal Tabanlı Fonksiyon, Doğrusal, Polinom ve Sigmoid olmak üzere 4 farklı çekirdek kullanılarak etkinliği araştırılmıştır.

Abdzadeh ve arkadaşlarının önerdikleri çalışmada (Abdzadeh vd., 2023), giriş sesinin 4 saniyelik eşit uzunlukta olması için kırpma veya doldurma işlemleri gerçekleştirilmiştir. Akustik özellikler olarak STFT, LFCC, MFCC ve CQT Power Spectrogram çıkarılmıştır. Her akustik özellik, Kendine-Dikkat (Self-Attention) katmanına sahip ResNet-18 mimarisine verilerek gömme vektörü çıkarılmıştır. Elde edilen derin özellikler, 3 katmanlı Çok Katmanlı Algılayıcı MLP kullanılarak ve Tek Sınıf-Softmax (One Class-Softmax, OC-Softmax) ile sınıflandırılmıştır.

Lee tarafından önerilen çalışmada Fourier analizinde spektrumdaki belirli frekans bileşeninin analizi incelenerek yeni bir sinyal analiz kavramı türetilmiştir (Lee, 2024). Üretilen kavram Keyfi Ayrık Fourier Analizi (Arbitrary Discrete Fourier Analysis, ADFA), Mel-Ölçekli Ayrık Fourier Analizi (Mel-Scale Discrete Fourier Analysis, MDFA), ve Sabit Q Analizi (Constant Q Analysis, CQA) olarak adlandırılan üç sinyal analizi yöntemi geliştirmek için kullanılmıştır. Derin sahte ses tespiti için yaptıkları çalışmada giriş sesinden Keyfi Ayrık Fourier Analizi (Arbitrary Discrete Fourier Analysis, ADFA), Mel-Ölçekli Ayrık Fourier Analizi (Mel-Scale Discrete Fourier Analysis, MDFA), ve Sabit Q Analizi (Constant Q Analysis, CQA) akustik özellikleri çıkarılmıştır. Sınıflandırma işlemi Hafif Evrişimli Sinir Ağları (Light Convolutional Neural Network, LCNN) kullanılarak yapılmıştır.

1.2.2. Ses Montaj Sahteciliği

Ses montajı, bir ses dosyasına eklenen, silinen veya birleştirilen ses sinyallerinin işlenmesi işlemidir (Zhao, Chen, Wang, & Malik, 2017). Bu işlem genellikle kişinin konuşmasını manipüle etmek için kullanılır. Örneğin, ses kaydına farklı konuşma parçaları eklenerek veya çıkarılarak orijinal kaydın içeriği değiştirilebilir. Birleştirme yöntemi ile gerçekleştirilen ses montaj örneği Şekil 1'de gösterilmiştir. Aynı kişinin iki farklı zamanda alınmış ses kayıtlarının birleştirilmesiyle oluşturulmuştur. Ses montajı, ses kaydının bütünlüğünü ve doğruluğunu etkileyebilecek bir tür ses manipülasyonudur. Bu tür manipülasyonlar, sahte ses kayıtları oluşturmak veya mevcut ses kayıtlarını yanıltıcı şekillerde değiştirmek amacıyla kullanılabilir.



Şekil 1. Ses Montaj Sahteciliği Oluşturulma Adımları

1.2.2.1. Literatür Araştırması

Ses montaj sahteciliği konusunda literatürde birçok çalışma mevcuttur. Bu yöntemler Geleneksel Yöntemler kullanan çalışmalar ve Derin Öğrenme yöntemleri ile sahtecilik tespiti olarak iki sınıfta incelenmektedir.

1.2.2.1.1. Geleneksel Yöntemler

Capoferri ve arkadaşları tarafından konuşma kaydında yankı süresindeki tutarsızlıklara bağlı bir yöntem önerilmiştir (Capoferri vd., 2020). Giriş sesinden Kısa Vadeli Fourier Transform çıkarılmıştır. Ardından kayan pencereler üzerinde yankı süreleri tespit edilmiş ve olası tutarsızlıklar aranmıştır. Eğer yankı süresi aniden değişirse, ses parçası sahte olarak etiketlenmiştir.

Yan ve arkadaşları tarafından önerilen çalışmada (Yan vd., 2021), ilk olarak ses sinyalindeki gürültünün tahmini için Sorensen algoritması kullanılmıştır. Tahmini gürültü çerçevelenmiş ve Mel-Frekans Kepstral Katsayıları çıkarılmıştır. Sahteciliğin tespiti ve

lokalisasyonu için katsayıların varyansı hesaplanmıştır. Son olarak Lavielle'nin algoritması sahteciliğin tespiti ve yerinin belirlenmesi amacıyla bu katsayılara uygulanmıştır.

1.2.2.1.2. Derin Öğrenme Tabanlı Yöntemler

Jadhav ve arkadaşları tarafından önerilen çalışmada giriş sesinden Kısa Zamanlı Fourier Transform akustik özelliği çıkarılmıştır (Jadhav vd., 2019) . Derin özelliklerin çıkarılması işlemi Evrişimli Sinir Ağları kullanılarak gerçekleştirilmiştir. Sınıflandırma modelin son katmanında kullanılan Softmax aktivasyon fonksiyonu kullanılarak yapılmıştır.

Zeng ve Wu tarafından önerilen çalışmada giriş sesi birbiri ile örtüşen pencerelere bölünmüştür (Zeng & Wu, 2022). Mel Spektrogram özelliği bu pencerelerden çıkarılarak 18 katmanlı ResNet mimarisine giriş olarak verilmiştir. Yöntem ResNet mimarisinden alınan ikili karar sonuçları ile sahteciliğin lokalisasyonu hakkında fikir vermiştir.

Moussa ve arkadaşları tarafından yapılan çalışmada sesin 3 farklı akustik özelliği birleştirilerek Transformer yapısı olan seq2seq mimarisine verilmiştir (Moussa vd., 2022). Bunlar Mel Spektrogram, Mel Frekans Kepstral Katsayıları ve sesin parlaklığını tanımlayan Spektral Merkez akustik özellikleridir. Transformer kodlayıcı ve kod çözücü bölümlerinden oluşmaktadır. Derin özellikler çıkarıldıktan sonra tamamen bağlı katmanların sınıflandırmada kullanılması sonucunda sahtecilik ve lokalisasyonunun tespiti gerçekleştirilmiştir.

Zhang ve arkadaşları tarafından önerilen çalışmada kodlayıcı-kod çözücü mimarisine (ASLNet) dayalı ses sahteciliği tespit yöntemi önerilmiştir (Zhang vd., 2022). Öncelikle ses birkaç küçük ses segmentine bölünmüş ve her ses segmentinden Doğrusal Frekans Kepstral Katsayıları (Linear frequency cepstral coefficient, LFCC) ve Mel Frekans Kepstral Katsayıları (Mel-frequency cepstral coefficient, MFCC) akustik özellikleri ve karşılık gelen ikili gerçek veri maskesi (binary ground truth mask) oluşturulmuştur. Daha sonra, sesin tüm segmentlerinden akustik özellik matrisleri birleştirilmiş ve kodlayıcı-kod çözücü mimarisine gönderilmiştir. Mimari her ses için ikili çıkış maskesi döndürmüştür. Sonraki adımda, ikili çıkış maske küçük segmentlere bölünmüş ve küçük segmentte 1 olan öğelerin sayısının tüm öğelerin sayısına oranı hesaplanmıştır. Bu oran önceden belirlenen eşik ile karşılaştırılarak ses sahte veya gerçek olarak etiketlenmiştir.

Zhang ve arkadaşları tarafından önerilen çalışmada (Zhang vd., 2022), giriş sesinden Yüksek Frekans bileşenli Tekillik Algılama Özelliği(Singularity Detection Feature, SDF) ve

Doğrusal Frekans Cepstral Katsayıları (Linear-Frequency Cepstral Coefficients,LFCC) akustik özellikleri çıkarılmıştır. Sıralı bilginin öğrenilebilmesi için SDF özelliği Uzun Kısa Vadeli Bellek (Long-Short Term Memory,LSTM) ile kullanılmıştır. LSTM çıktısı özellik eksenini boyunca LFCC ile birleştirilmiş ve Hafif Evrişimli Sinir Ağı (Light Convolution Neural Network,LCNN) mimarisine gönderilmiştir.

Wang ve arkadaşları tarafından önerilen çalışmada (Wang vd., 2022), giriş kayan pencereler kullanılarak parçalara bölünmüş ve her parçadan akustik özellik çıkarılmıştır. Çalışmada Sabit-Q Transform (Constant-Q Transform, CQT) akustik özelliği kullanılmıştır. Derin öğrenme mimarisi olarak SE-Res2Net34 önerilmiştir. Ses parçalarından elde edilen akustik özellik mimariye verilerek derin özelliklerin çıkarılması sağlanmıştır. Modelin sahte ses parçacıklarında ekleme yapılan yerleri tespit etmesi beklenmiştir.

Leonzio ve arkadaşları tarafından önerilen çalışmada (Leonzio vd., 2023), giriş sesinden Kısa Vadeli Fourier Transform alınmıştır. Frekans temsiliinin kayıt cihazı sınıflandırma amacıyla geliştirilmiş CNN yapısı kullanılarak derin özelliği çıkarılmıştır. Ekleme ses sahteciliği tespiti k-Means algoritması kullanılarak yapılmıştır. Giriş sesinden sahtecilik tespit edildiğinde lokalizasyon aşamasına geçilmiştir. Ses kaydında aynı kayıt cihazı ile kaydedilmiş komşu pencerelerin özellikleri birbiri ile karşılaştırılır. Kayıt cihazının değişiklik göstermesi durumunda uzaklık yüksektir. Buna karşın aynı kayıt cihazıyla alınan pencerelerin uzaklığı düşüktür.

Moussa ve arkadaşları tarafından önerilen yöntemde kodlayıcı ve kod çözücü yapısına sahip mimari kullanılmıştır (Moussa vd., 2023). Mimariye giriş olarak ham ses verileri verilmiştir. Kodlayıcı bölümü Transformer yapısı ile tamamen uyumludur. Ancak kod çözme aşaması çoklu kafa dikkat katmanlarından (Multi-Head Attention Layers) oluşur. Önerdikleri SigPointer yapısı şu şekilde çalışmaktadır; Çıkış dizisini tahmin etmek için tüm giriş konumları üzerindeki koşullu olasılık dağılımı kullanılmaktadır. Bu nedenle, tüm dikkat katmanının (Attention Head) değerleri ortalamaya normalize edilmiştir.

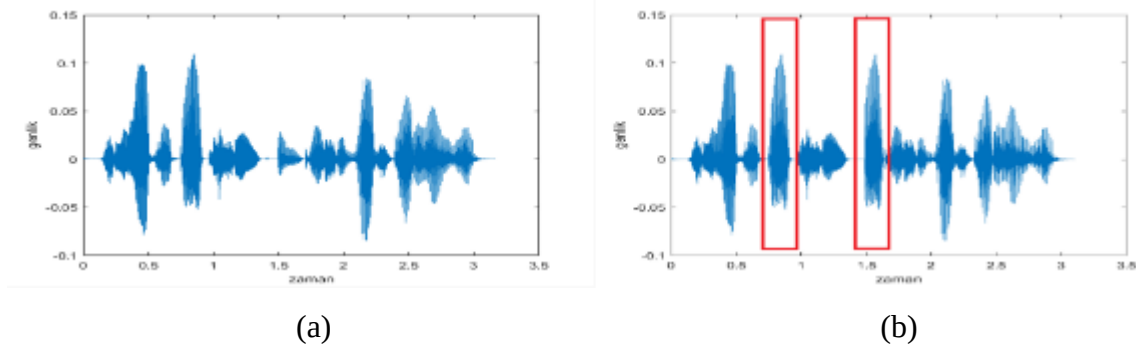
Zeng ve arkadaşları tarafından ENF Salınlı Süper-Vektör (ENF-FSV) ve derin özellik öğrenme temsili esas alınarak ses sahteciliği tespit yöntemi önerilmiştir (Zeng vd., 2023). Eğitim için yeterli sayıda sahte konuşma örneği bulunamadığında, ENF Evrensel Arka Plan Modeli (ENF-UBM) oluşturularak ve ortalama vektör güncellenerek ENF-FSV çıkarılır. ENF özelliklerinin önemli özellikleri vurgulamaması sorununu göz önünde bulundurarak, dikkat (attention) mekanizmasıyla ani dalgalanma bilgisine olan dikkati güçlendirmek için uçtan uca bir derin sinir ağı inşa edilir. Son olarak, Residual Network

(ResNet) modülüyle çıkarılan derin ENF-FSV özellikleri tasarlanan sınıflandırma ağına beslenir.

1.2.3. Kopyala Yapıştır Ses Sahteciliği

Kopyala yapıştır ses sahteciliği konuşma kaydındaki kelime veya kelime gruplarının kopyalanıp yine aynı konuşma kaydındaki farklı lokalizasyona yapıştırılmasıyla oluşturulmaktadır (Xiao, Jia, Fu, Huang, Li, & Shi, 2014). Bu işlem mevcut ses düzenleme teknolojileriyle kolaylıkla gerçekleştirilmektedir.

Örneğin “Ben olay anında odada o ve maktulün bulunduğunu söylüyorum” şeklindeki konuşma kaydını ele geçiren saldırgan cümleyi “**Ben** olay anında odada **ben** ve maktulün bulunduğunu söylüyorum” şeklinde değiştirebilir. Bu örnek kopyala yapıştır sahteciliğinin tehlikesini gözler önüne sermektedir. Birkaç basit adımla gerçekleştirilebilen bu sahtecilik büyük problemlere sebep olmaktadır. Dolayısıyla, teknolojinin getirdiği kolaylıkların yanı sıra, bu tür kötü niyetli uygulamalara karşı da dikkatli olunması gerekmektedir. Sahte ses oluşturma örneği Şekil 2b.’de gösterilmektedir. TIMIT veri tabanından alınan Şekil 2a’da ki orijinal konuşma kaydında ("sa1.wav") ikinci sözcük kopyalanarak aynı konuşma kaydının dördüncü sözcüğünün yerine yapıştırılmıştır. Bu işlem sonucunda, Şekil 2b’de gösterilen sahte ses elde edilmiştir.



Şekil 2. (a) Orijinal Ses (b) Kopyala Yapıştır Sahte Ses

Şekil 2b.’de kırmızı dikdörtgen kopyala yapıştır bölgelerini göstermektedir. Soldaki kırmızı kutucukla işaretlenmiş kelime kopyalanıp sağdaki kutucukla işaretlenen bölgeye yapıştırılarak sahte ses oluşturulmaktadır.

Kopyala-yapıştır sahteciliği, diğer ses sahteciliği türlerinden belirgin şekilde farklılık gösterir. Bu tür sahtecilikte, kopyalanan ses kesiti farklı ses kaydına değil, aynı kaydın

içindeki başka bölüme yapıştırılır. Dolayısıyla, kopyala-yapıştır yapılmış kelimeler için sesin genlik, frekans, uzunluk, gürültü ve ton gibi temel özellikleri değişmez. Bu durum, bir uzman tarafından dilbilimsel veya ses bilimsel analiz yapılmadığı sürece sesin otomatik olarak orijinal veya sahte olarak değerlendirilmesini zorlaştırır. Kopyala-yapıştır sahteciliğinin tespitini daha da zorlaştıran durum ise saldırganların sahtecilik izlerini gizlemek için uyguladığı son işlem operasyonlarıdır. Bu operasyonlar, sıkıştırma, gürültü ekleme, medyan filtreleme gibi işlemleri içerebilir ve ses dalgasının şeklini değiştirerek sahtecilik tespitini daha karmaşık hale getirir.

1.2.3.1. Literatür Araştırması

Bu bölümde kopyala yapıştır ses sahteciliği tespitine ilişkin literatürde yapılan çalışmalar incelenecektir.

Literatürde, ses kopyala-yapıştır sahteciliğini tespit etmek için önerilen yöntemler üç kategori altında incelenmektedir: Pencere tabanlı, Konuşma Etkinliği Algılama (Voice Activity Detection, VAD) tabanlı, Spektrogram tabanlı.

Pencere tabanlı yöntemler de konuşma kaydı önce örtüşmeyen veya eşit uzunluktaki örtüşen pencerelere bölünerek işlenir. Ardından, bu pencerelerden son işlem operasyonlarına dayanaklı özellikler elde edilir.

VAD Tabanlı yöntemler de Konuşma Etkinliği Algılama (Voice Activity Detection, VAD) veya Perde İzleme gibi yöntemler kullanılarak sesli bölgeler belirlenir ve bu bölgelerden özellikler çıkarılır. Sonrasında, elde edilen özellikler arasındaki benzerlik değerlendirilerek sahte bölümler tespit edilir ve işaretlenir.

Spektrogram Tabanlı yöntemlerde konuşma kaydı üzerinde herhangi ses etkinliği algılama modülü uygulanmaz veya pencerelere bölünmez. Konuşma kaydının tamamından özellik çıkarılır.

1.2.3.1.1. Pencere Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti

Kopyala yapıştır ses sahteciliği tespitine ilişkin ilk çalışma Xiao ve arkadaşları tarafından gerçekleştirilmiştir (Xiao vd., 2014). Ses dosyalarını analiz etmek için pencere tabanlı yöntem önerilmiştir. Bu yöntemde, öncelikle ses dosyası belirli periyoda sahip pencerelere bölünmüştür. Ardından, pencereler arasındaki benzerlik ölçülmüş ve bu işlem

için hızlı evrimsel algoritma kullanılmıştır. Belirli eşik üzerinde benzerlik gösteren pencereler, sahte olarak tanımlanmıştır.

Su ve arkadaşlarının yaptıkları çalışmada (Su vd., 2023), konuşma kaydı içindeki sesli bölgelerde ve bu bölgeler arasında kopyala-yapıştır sahteciliğini tespit etmek için iki kayan pencere (SW) stratejisi önerilmiştir. Konuşma kaydında sesli bölgelerin belirlenmesi için VAD yöntemi kullanılmıştır. Çalışmada, Sabit-Q Cepstral Katsayıları (Constant Q Cepstral Coefficients, CQCC) özelliği kullanılarak pencerelerden özellikler çıkarılmıştır. Farklı kısa pencereler arasındaki benzerlik, Pearson korelasyon katsayıları kullanılarak hesaplanmıştır.

1.2.3.1.2. VAD Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti

Yan ve arkadaşları tarafından önerilen çalışmada (Yan vd., 2015), başlangıç olarak ses dosyasından Perde dizisi (Pitch) çıkartılmıştır. Elde edilen perde dizilerinin benzerliğini belirlemek için ise Ortalama Fark (AD) ve Pearson Korelasyon Katsayısı (Pearson Correlation Coefficient-PCC) gibi yöntemlerden faydalanılmıştır. Bu yöntemler, perde dizileri arasındaki ilişkiyi değerlendirme ve benzerlik düzeyini nicel olarak ölçme amacıyla kullanılmıştır. Ses dosyalarının işlenmesinde bu yöntemlerin kullanımı, sahte seslerin tespiti ve gerçek seslerin ayırt edilmesi gibi kritik adımlarda etkili olmuştur.

Wang ve arkadaşları tarafından yapılan çalışmada (Wang vd., 2017), Tekil Değer Ayırıştırma (Singular Value Decomposition, SVD) dönüşümü ile Ayrık Kosinüs Dönüşümü (Discrete Cosinus Transform, DCT) kullanılarak ses kopyala-yapıştır sahteciliğini tespit etmek için yöntem önerilmiştir. Önerilen yöntemde, konuşma kaydını öncelikle sesli bölümlere ayırmak için VAD yöntemi kullanılmıştır. Daha sonra, her ses parçasından DCT katsayıları elde edilmiştir. Ardından, bu katsayıların kare matrisine SVD uygulanarak özvektörler elde edilmiştir. Benzerlik hesaplamasında ise Öklid Uzaklığından (Euclidean Distance, ED) faydalanılmıştır.

Liu ve Lu önerdiği çalışmada (Liu & Lu, 2017), konuşma kaydı öncelikle sesli kısımlara ayrılmıştır. Ardından, sesli kısımlardan Discrete Fourier Transform (DFT) özelliği elde edilmiştir. Özelliklerin hesaplanma sürecini hızlandırmak için sıralama yapılmış ve sadece yakın segmentlerle karşılaştırma işlemi gerçekleştirilmiştir. Bu karşılaştırma işlemi, Pearson Korelasyon Katsayısı kullanılarak gerçekleştirilmiştir.

İmran ve arkadaşları yaptığı çalışmada (İmran vd., 2017), ses kopyala-yapıştır sahteciliğini tespit etmek için Yerel İkili Örüntü (Local Binary Pattern-LBP) yöntemine

dayalı yöntem önermiştir. Bu yöntemde, diğer VAD tabanlı yöntemler gibi öncelikle sesli bölümler, önerdikleri VAD yöntemiyle elde edilmiştir. Daha sonra, her sesli bölümden özellik olarak LBP histogramları üretilmiştir. Histogramların benzerlik hesaplaması, Ortalama Kare Hatası (Mean Squared Error, MSE) ve Hata oranı (Error Rate, ER) metrikleri kullanılarak gerçekleştirilmiştir.

Xie ve arkadaşlarının yaptıkları çalışmada (Xie vd., 2018), sesli bölümlerden elde edilen ses perdesi, ayrık Fourier dönüşüm katsayıları (Discrete Fourier Transform-DFT), Mel frekans cepstral katsayıları (Mel-frequency cepstral coefficient-MFCC) ve Gamatonlar gibi dört farklı özellik çıkarılmıştır. Bu özelliklerin benzerliklerini değerlendirmek için PCC (Pearson Correlation Coefficient) ve AD (Average Difference) yöntemleri kullanılmıştır. Dört farklı özellikten elde edilen tespit sonuçları, nihai kararı elde etmek için C4.5 karar ağacıyla birleştirilmiştir.

Yan ve arkadaşları yaptıkları çalışmada (Yan vd., 2019), ses dosyasının sesli bölümleri Normalleştirilmiş Düşük Frekanslı Enerji Oranı kullanılarak belirlenmiştir. Ardından, sesli bölümlerden Perde ve Formant dizileri çıkarılmıştır. Daha sonra, DTW (Dynamic Time Warping) yöntemi kullanılarak bu diziler arasında benzerlik hesaplanmıştır.

Mannepalli ve arkadaşları yaptıkları çalışmada (Mannepalli vd., 2019), konuşma kaydının sesli kısımlarından Mel-Frekans Cepstral Katsayıları (Mel-Frequency Cepstral Coefficients, MFCC) özellikleri elde edilmiştir. Ayrıca, bu özellik vektörlerinin benzerliği Dinamik Zaman Atlama (Dynamic Time Warping, DTW) mesafesi kullanılarak karşılaştırılmıştır.

Li ve arkadaşlarının yaptıkları çalışmada (Li vd., 2019), konuşma kaydının ön işlem adımında karesi alınır. Ardından, Perde Dizileri çıkarılarak sesli ve sessiz kısımlar belirlenir. Sesli bölgelerin perde dizileri, işlemlerin daha hızlı gerçekleştirilebilmesi için Sonlu Dürtü Tepkisi (Finite Impulse Response, FIR) kullanılarak karşılaştırılır.

Anh ve arkadaşları tarafından önerilen çalışmada ses dosyasından elde edilen sesli bölümlerden fonetik diziler çıkarılmıştır (Anh vd., 2019). Çeşitli fonetik diziler arasındaki benzerlik en küçük sapmalarla hesaplanmıştır.

Huang ve arkadaşları tarafından önerilen yöntemde Huang vd. (2020), öncelikle belirledikleri VAD algoritmasıyla konuşma kaydının sesli kısımlarını tespit etmişlerdir. Ardından, bu sesli kısımlardan özellik çıkarmak için Ayrık Fourier dönüşümü (DFT) ve Mel-Frekans Cepstral Katsayıları (Mel-Frekans Cepstral Katsayıları, MFCC) yöntemlerinden

faaydalanmıřlardır. Elde edilen benzer  zellikler aracılıęıyla kopyala-yapıřtır yapılmıř kelimeleri tespit etmiřlerdir.

K         ve arkadařları  nerdikleri y ntemde (K         vd., 2020), bařlangı ta konuřma kaydı Perde  zleme y ntemi kullanılarak hecelere b l nm řt r. Her bir hece, 1 boyutlu yerel ikili model operat r yle analiz edilmiřtir. Bu operat r, her hecenin histogramını hesaplamaktadır. Histogramların benzerliklerini deęerlendirmek i in ortalama kare hatası kullanılmaktadır. Benzer histogramlara sahip olan heceler, sahte ses b l mleri olarak etiketlenmiřtir.

Akdeniz ve Becerikli yaptıkları  alıřmada (Akdeniz & Becerikli, 2021), ses sinyallerinde kopyala-yapıřtır sahtecilięini tespit etmek i in Mel Frekansı ve Delta Mel Frekansı tabanlı y ntemler  nerilmiřtir. Konuřma kaydındaki sesli b lgeleri belirlemek i in YAAPT (Yet Another Algorithm for Pitch Tracking) y ntemi kullanılmıřtır.  znitelik  ıkarmak i in Mel Frekansı Kepstrum Katsayıları (MFCC) ve Delta-Mel Frekansı Kepstrum Katsayıları (Delta-MFCC) kullanılmıřtır. Her b lge i in 13'er tane MFCC ve Delta-MFCC katsayıları elde edilmiřtir, bu katsayıların ortalaması alınarak her b lge i in temsil edici vekt r elde edilmiřtir. Benzerlięi belirlemek i in Pearson Korelasyon katsayısı hesaplanmıřtır. Son ařamada ise, elde edilen korelasyon matrisine dayanarak bir eřik deęeri belirlenerek ses sinyallerinde kopyala-yapıřtır sahtecilięinin varlıęı tespit edilmiřtir.

Akdeniz ve Becerikli  nerdikleri  alıřmada (Akdeniz & Becerikli, 2022), Perde  zleme y ntemi kullanılarak konuřma kaydındaki sesli b lgeler tespit edilmiřtir. Bu segmentlerden 5, 10 ve 15'lik Doęrusal Tahmin Katsayıları (LPC) elde edilmiřtir. Elde edilen katsayılar arasındaki benzerlik Pearson Korelasyon Katsayısı (PCC) kullanılarak hesaplanmıřtır. Deneysel b l mde, farklı katsayıların  alıřmanın bařarısına etkisi analiz edilmiř ve elde edilen sonu lar tablolar halinde sunulmuřtur.

 st bioęlu ve arkadařları  nerdikleri y ntemde ( st bioęlu vd., 2022), YAAPT y ntemi kullanılarak konuřma kaydından sesli b lgeler tespit edilmiřtir. Ardından, bu sesli b lgelerden Modifiye Ayrık Kosin s D n ř m  (MDCT) katsayıları elde edilmiř ve  zellik olarak katsayı matrisinin transpozisinin ortalaması alınmıřtır. Sesli b l mlerin  zellikleri arasındaki benzerlikleri  l mek i in  klid Uzaklıęı (Euclidean Distance, ED) kullanılmıřtır.

Xianfeng ve arkadařları tarafından  nerilen y ntem Xianfeng vd. (2023), ses kopyala-yapıřtır sahtecilięini tespit etme ve konumlandırmaya odaklanmaktadır. İlk adımda, konuřma kaydından Perde  zleme teknięiyle sesli ve sessiz segmentler ayrıştırılmıř ve her sesli segment  zerinde Hilbert-Huang d n ř m  uygulanarak spektrogram  zellikleri elde

edilmiştir. Daha sonra, her spektrogram özelliğinin benzerliğini hesaplamak için Dinamik Zaman Atlama (DTW) yöntemi kullanılmıştır. Belirli bir eşik değeri aşıldığında, Hilbert spektrogram özelliklerinin DTW değerlerini karşılaştırarak kopyala-yapıştır sahteciliğinin varlığı tespit edilmiş ve sesli segmente karşılık gelen DTW değeri indeksine göre kopyala-yapıştır sahteciliğinin konumu belirlenmiştir.

Üstübioğlu ve Üstübioğlu önerdiği yöntemeye göre (Üstübioğlu & Üstübioğlu, 2024), konuşma kaydı öncelikle perde bazlı ses aktivite algılama (Voice Activity Detection-VAD) yöntemiyle kelimelere ayrılmıştır. Daha sonra, her kelime Mel Spektrogram görüntüsüne dönüştürülmüştür. Spektrogram görüntüleri arasındaki benzerliği hesaplamak için yapısal farklılık (Structural Similarity-DSSIM) kullanılmıştır. Kelime görüntüleri arasındaki DSSIM değerlerine göre sahte kelimeler işaretlenmiştir.

Akdeniz ve Becerikli tarafından konuşma kaydındaki sesli ve sessiz bölgeler Perde izleme yöntemi ile belirlendikten sonra (Akdeniz & Becerikli, 2024), özellikler Mel frekans Kepstrum Katsayıları, Delta Mel frekans Kepstrum Katsayıları ve Delta Delta Mel Frekans Kepstrum Katsayıları ile birlikte Doğrusal Öngörme Katsayıları (LPC) kullanılarak sinyallerden elde edilmiştir. Sırasıyla, bu özellikler yapay sinir ağları kullanılarak sınıflandırılmıştır.

1.2.3.1.3. Spektrogram Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti

Ulutaş ve arkadaşları tarafından önerilen yöntemde (Ulutaş vd., 2022), konuşma kaydının frekans temsili çıkarılmıştır. Mel spektrogram görüntüsü örtüşen alt bloklara bölünmüş ve her RGB renk kanalı için bu bloklardan Gri Seviye Eş Oluşum Matrisi aracılığıyla özellikler çıkarılmıştır. Elde edilen özellik vektörleri, benzer vektörleri daha yakın hale getirmek için sözlüksel olarak sıralanmıştır. Bloklar arasındaki benzerlik, sahtecilik konusunda önemli ipucu sağlamıştır.

Yazıcı ve arkadaşları tarafından önerilen yöntemde (Yazıcı vd., 2022), ses dosyasından elde edilen Mel spektrogram görüntüleri kullanılmıştır. Bu görüntüler öncelikle gri seviye görüntülere dönüştürülmüş ve ardından örtüşen alt bloklara bölünmüştür. Her blok için özellikler, İkili Gradyan Modeli aracılığıyla çıkarılmıştır. Elde edilen özellik vektörleri, benzerliklerini belirlemek için leksikografik olarak sıralanmıştır.

Ulutaş ve arkadaşları tarafından önerilen yöntemde (Ulutaş vd., 2022), şüpheli ses dosyasından Mel spektrogram görüntüsü oluşturulmuştur. Bu görüntü üzerinde her RGB

renk kanalından SURF (Speeded Up Robust Features) anahtar noktaları çıkarılmıştır. Her kanaldan elde edilen anahtar noktalar, özellik vektörleri aracılığıyla eşleştirilerek ses dosyasının sahte mi yoksa orijinal mi olduğu belirlenmiştir. Son olarak, olası yanlış eşleşmeleri ortadan kaldırmak için önerilen son işleme adımı uygulanmıştır. Yöntem, deneysel çalışmalarla belirlenen eşleşme sayısının eşik değerine göre yeterli nihai eşleşen anahtar noktaların sağlanmasıyla, ses dosyasının sahte olduğunu ortaya koymaktadır.

Üstübioğlu ve arkadaşları önerdikleri metodolojide Mel Spektrogram görüntüsündeki anahtar noktaları tespit etmek için SIFT (Scale-Invariant Feature Transform) yöntemini kullanmışlardır (Üstübioğlu vd., 2023). Görüntü kanallarından elde edilen anahtar noktalar, özellik vektörleri aracılığıyla eşleştirilmiş ve görüntü alt blokları, anahtar noktaların merkezi olarak belirlenmiş ve sahte bloklar olarak işaretlenmiştir.

Üstübioğlu ve arkadaşları çalışmasında (Üstübioğlu vd., 2023), şüpheli sesi görselleştirmek için Süper Çözünürlüklü Spektrogram görüntülerinden yararlanılmıştır. Bu görüntülerden BRIEF (Binary Robust Independent Elementary Features) yöntemiyle anahtar noktalar ve tanımlayıcılar çıkarılmıştır. Ardından, ilgili tanımlayıcıları eşleştirmek için OPTICS (Ordering Points To Identify the Clustering Structure) yöntemi kullanılmıştır. Yöntem, spektrogram görüntüsündeki kümelerde yer alan anahtar noktaların konumuna bağlı olarak kopyalanmış bölgeler işaretlenmiştir. Aynı çalışma kapsamında, AKAZE, SIFT ve SURF anahtar nokta ve tanımlayıcı çıkarma yöntemleri denenmiştir. Farklı kümeleme yöntemleri olarak g2NN ve k-means metodlarının etkinliği de araştırılmıştır.

Üstübioğlu ve Tahaoğlu önerdiği yöntemde (Üstübioğlu & Tahaoğlu, 2023), başlangıçta ses dosyasından yüksek çözünürlüklü spektrogram görüntüsü elde edilmiştir. Daha sonra, Akaze yöntemi kullanılarak spektrogram görüntüsünden anahtar noktalar ve özellik tanımlayıcıları çıkarılmıştır. Elde edilen özellikler, g2NN (nearest neighbor graph) algoritmasıyla eşleştirilmiştir. Spektrogramdaki noktaların ses dosyasına iz düşürülmesi ile, ses kopyala-yapıştır sahteciliği tespit edilmiştir.

Üstübioğlu ve arkadaşları tarafından yapılan çalışmada (Üstübioğlu vd., 2023), kopyala yapıştır sahte seslerin tespiti için literatürde ilk kez Mel spektrogram ile derin öğrenme yöntemi kullanılmıştır. Bu çalışmada önerilen CNN mimarisi, şüpheli Mel spektrogram görüntülerini orijinal ve sahte olarak etiketlemektedir. Ayrıca, mimarinin öğrenme kapasitesini artırmak için spektrogram görüntülerine veri artırma yöntemleri uygulanmıştır.

Su ve arkadaşlarının önerdiği yöntemde (Su vd., 2023), Sabit Q Spektral Çizimlerine (CQSS) dayalı olarak özelleştirilmiş Genetik Algoritma (GA) ve Destek Vektör Makinesi (SVM) entegrasyonuna dayalı yaklaşım sunulmuştur. Bu yöntemde, CQSS özellikleri öncelikle kare-büyüklik sabiti Q dönüşümünün logaritmasının ortalaması alınarak çıkarılmıştır. Daha sonra, CQSS özellik seti, SVM ile birleştirilmiş özelleştirilmiş GA tarafından otomatik olarak optimize edilmiş ve aynı anda en iyi özellik alt kümesi ve sınıflandırma modeli elde edilmiştir.

1.3. Kopyala Yapıştır Sahte Ses Tespitinde Kullanılan Yöntemler

Bu bölümde yüksek lisans tez kapsamında kopyala yapıştır ses sahteciliği alanında kullanılan metodlar detaylı bir şekilde açıklanacaktır.

1.3.1. Yüksek Çözünürlüklü Spektrogram

STFT tabanlı spektrogramlar (Smith, 2010), ses sinyallerinin görsel temsilinde sıkça kullanılan yöntemdir. Bu yaklaşımlar, zaman ve frekans çözünürlüklerinin ilişkisini vurgular, böylece sesin zaman içindeki değişimlerini ve frekans bileşenlerini anlamamıza yardımcı olur. Ancak, STFT'nin bir dezavantajı vardır: aynı anda hem zaman hem de frekansda yüksek çözünürlük elde etmek mümkün değildir. STFT'nin hesaplanması için, sinyalin bir dizi segmenti üzerinde kayan pencere kullanılır. Bu pencere kısa olduğunda, sinyalin zamansal değişiklikleri net bir şekilde gözlemlenebilir ancak frekans eksenindeki harmonik pikler bulanıklaşır. Öte yandan, pencere uzun olduğunda, harmonik pikler daha belirgin hale gelir ancak zaman eksenindeki zamansal değişikliklerin netliği düşer. Tez kapsamında, zaman ve frekans açısından yüksek çözünürlük sağlayan spektrogram çıkarma tekniği benimsenmiştir. Bu tekniğin temelinde, Olasılıksal Gizli Bileşen Analizi (Probabilistic Hidden Component Analysis, PLCA) kullanılarak elde edilen iki spektrogramın birleştirilmesi yatar (Nam vd., 2010). Bu birleştirme adımı, yüksek zaman çözünürlüğüne sahip olan spektrogram ile yüksek frekans çözünürlüğüne sahip olan spektrogramın kombinasyonunu sağlar. Sonuç olarak, zaman ve frekans açısından yüksek çözünürlüklü ses spektrogramı elde edilir. Bu yöntem, ses sinyallerinin detaylı analizinde ve işlenmesinde önemli rol oynamaktadır.

PLCA , spektrogramı pozitif olmayan spektral ve zamansal bileşenlerin dış çarpımlarıyla toplamak için kullanılır ve Eşitlik (1)' de ifade edilmektedir. Örneğin Eşitlik (1)' de verilen V_{ft} f ve t ile temsil edilmiş spektrogram olsun.

$$V_{ft} = Y \sum_z P(z)P(f|z)P(t|z) \quad (1)$$

$P(f|z)$ spektral bileşenleri belirtirken, $P(t|z)$ ilgili spektral bileşenlerinin zaman içindeki varlığını, $P(z)$ ağırlıkların dağılımını, Y ölçekleme faktörünü ifade eder. Spektrogram, genellikle çoklu ağırlık dağılımlarının parametrelerinin tahmin edilmesiyle ayrıştırılabilir. Bu tahminler, Beklenen Maksimum (Expectation Maximization, EM) algoritması kullanılarak gerçekleştirilir. EM algoritması E ve M olmak üzere 2 adımda gerçekleştirilir. E adımı,

$$P(z|f, t) = \frac{P(z)P(f|z)P(t|z)}{\sum_z P(z)P(f|z)P(t|z)} \quad (2)$$

biçiminde hesaplanır. M adımı,

$$P(f|z) = \frac{\sum_t V_{ft}P(z|f, t)}{\sum_f \sum_t V_{ft}P(z|f, t)} \quad (3)$$

$$P(t|z) = \frac{\sum_f V_{ft}P(z|f, t)}{\sum_t \sum_f V_{ft}P(z|f, t)} \quad (4)$$

$$P(z) = \frac{\sum_f \sum_t V_{ft}P(z|f, t)}{\sum_z \sum_f \sum_t V_{ft}P(z|f, t)} \quad (5)$$

biçiminde hesaplanır.

PLCA, sadece zaman veya frekans alanı için kullanılabilir, ancak her ikisini birden kapsamaz. Yüksek frekans çözünürlüğüne sahip spektrogramu hesaplamak için, Eşlenmiş PLCA (Coupled PLCA) yöntemi kullanılır. Bu yöntemde, uzun pencereyle elde edilen

yüksek frekans çözünürlüğüne sahip V_F ve kısa pencereyle elde edilen yüksek zaman çözünürlüğüne sahip V_T olmak üzere iki spektrogram bulunur.

Yüksek Frekans çözünürlük dağılımı $P_F(f|z)$, $P_F(t|z)$ ve $P_F(z)$ iken Yüksek Zaman çözünürlük dağılımı $P_T(f|z)$, $P_T(t|z)$ ve $P_T(z)$ olarak belirlenmektedir.

Bu spektrogramlar, bulanık (blurred) sürüm elde edilmesinin ardından yorumlanır. Bulanıklık fonksiyonları, bu spektrogramların çıkarılmasında kullanılan uzun ve kısa pencereler arasındaki ilişkiyle belirlenir. Bulanıklık fonksiyonu, zaman alanında kısa pencereyi uzun pencereye dönüştüren lineer filtre olarak düşünülebilir. Spektrogram, STFT'nin büyüklüğü olarak verildiği için, zaman ve spektral bulanıklık işlemlerinin yalnızca bir Fourier dönüşüm çiftini yaklaşık olarak temsil etmesi beklenir. Zaman bulanık fonksiyonu B_T ve frekans bulanık fonksiyonu B_F , sırasıyla Eşitlik (6) ve Eşitlik (7)' de ifade edilmektedir.

$$P_T(f|z) = B_F(P_F(f|z)) \quad (6)$$

$$P_F(t|z) = B_T(P_T(t|z)) \quad (7)$$

Bu ilişkiden yararlanarak, yöntem $P_T(f|z)$ ve $P_F(t|z)$ bileşenlerini EM algoritmasıyla elimine eder. E adımı

$$P_F(z|f, t) = \frac{P_F(z)P_F(f|z)B_T(P_T(t|z))}{\sum_z P_F(z)P_F(f|z)B_T(P_T(t|z))} \quad (8)$$

$$P_T(z|f, t) = \frac{P_T(z)B_F(P_F(f|z))P_T(t|z)}{\sum_z P_T(z)B_F(P_F(f|z))P_T(t|z)} \quad (9)$$

biçiminde hesaplanır. M adımı,

$$P_F(f|z) = \frac{\sum_t V_{ft}^{(F)} P_F(z|f, t)}{\sum_f \sum_t V_{ft}^{(F)} P_F(z|f, t)} \quad (10)$$

$$P_T(t|z) = \frac{\sum_t v_{ft}^{(T)} P_T(z|f,t)}{\sum_t \sum_f v_{ft}^{(T)} P_T(z|f,t)} \quad (11)$$

$$P_F(z) = \frac{\sum_f \sum_t v_{ft}^{(F)} P_F(z|f,t)}{\sum_z \sum_f \sum_t v_{ft}^{(F)} P_F(z|f,t)} \quad (12)$$

$$P_T(z) = \frac{\sum_f \sum_t v_{ft}^{(T)} P_T(z|f,t)}{\sum_z \sum_f \sum_t v_{ft}^{(T)} P_T(z|f,t)} \quad (13)$$

biçiminde hesaplanır. Bulanıklaştırma işlevleri, iki spektrogram oluşturmak için kullanılan uzun ve kısa pencereler arasındaki ilişkiden türetilir. Spektrogram, STFT'nin büyüklüğü olarak verildiğinden, zamansal ve spektral bulanıklık işlevlerinin yalnızca bir Fourier dönüşüm çiftini yaklaşık olarak temsil etmesi beklenir. Bulanıklık fonksiyonu, zamansal alanda kısa pencereyi uzun pencereye veya frekans alanında keskin büyüklük tepkisini geniş tepkiye dönüştüren bir lineer filtre olarak düşünülebilir.

$$w_F(n) = B_F(\omega_T(n)) = \sum_i b_T(i) \omega_T(n - ih) + e_T(n) \quad (14)$$

$$W_T(k) = B_T(W_F(k)) = \sum_j b_F(j) W_F(k - j) + e_F(k) \quad (15)$$

h STFT atlama boyutu (hop size), $\omega_T(n)$ ve $\omega_F(n)$ kısa ve uzun pencereler, $W_T(k)$ ve $W_F(k)$ büyüklük (magnitude) tepkileri, $b_T(n)$ ve $b_F(k)$ zaman ve frekans bulanıklık fonksiyonlarının katsayılarıdır. Zaman dizini i , uzun pencerenin genişliğini kapsarken, frekans dizini j , yanlob seviyesinin yeterince düşük olduğu varsayılan geniş büyüklük tepkisinin ana lobuna sınırlıdır. Yukarıda türettiğimiz bulanıklık fonksiyonları, süper çözünürlük spektrogramundan yüksek frekans çözünürlüğüne veya yüksek zaman çözünürlüğüne sahip spektrograma dönüşüm sırasında oluşan zaman veya frekans bulanıklığını yaklaşık olarak gösterir. Dolayısıyla, bulanıklaştırma fonksiyonu ve eşleştirilmiş PLCA kullanılarak yüksek frekans çözünürlüğüne sahip $P_F(f|z)$ ve yüksek

zaman çözünürlüğüne sahip $P_T(t|z)$ hesaplanır. Bu bileşenlerin kullanımıyla, süper çözünürlüklü spektrogram şu şekilde oluşturulur:

$$V_{ft}^{(s)} = \gamma \sum_z P(z)P(f|z)P(t|z) \quad (16)$$

1.3.2. Anahtar Nokta ve Tanımlayıcıların Çıkarılması

Görüntüde anahtar noktalar ve bunların tanımlayıcılarının çıkarılması, görüntü işleme ve bilgisayarlı görme alanlarında yaygın olarak kullanılan tekniktir. Anahtar noktalar, görüntünün belirgin veya dikkate değer bölgelerini temsil eder ve genellikle görüntü tanıma, nesne algılama ve eşleştirme gibi uygulamalarda kullanılırlar. Anahtar noktaların ve tanımlayıcılarının çıkarılması birkaç adımdan oluşmaktadır. Görüntü genellikle gürültüyü azaltmak, kontrastı artırmak veya özelliklerin belirginleştirilmesi için ön işleme adımlarından geçirilir. Görüntüdeki belirgin noktaların tespit edilmesi için tespit yöntemi kullanılır. Bu yöntem, genellikle kenarlar, köşeler veya tepe noktalar gibi karakteristik özellikleri bulur. Belirlenen anahtar noktaların konumları daha hassas şekilde belirlenir. Bu, anahtar noktaların yerel pikseller arasındaki maksimumlardan tespit edilmesini içerebilir. Her anahtar nokta için yönlendirme veya oryantasyon atanır. Bu, anahtar noktanın çevresindeki yoğunluk dağılımının analiz edilmesi ile gerçekleştirilir. Her anahtar nokta etrafındaki bölgede bir tanımlayıcı vektör çıkarılır. Bu vektör, anahtar noktanın etrafındaki piksel yoğunluğu, gradient bilgileri veya yoğunluk dağılımının kombinasyonunu içerebilir.

1.3.2.1. Ölçekten Bağımsız Öznitelik Dönüşümü (SIFT)

1999 yılında David G. Lowe tarafından önerilen Ölçekten Bağımsız Öznitelik Dönüşümü (SIFT) algoritması (Lowe, 1999), görüntüde belirgin yerel özellikleri tanımlamak için kullanılır. Bu algoritma, görüntüdeki ölçekleme, öteleme ve dönme gibi farklı değişikliklere dayanıklı öznitelikler elde etmek amacıyla tasarlanmıştır. SIFT algoritmasının ana adımları arasında, ölçek uzayındaki ekstrem noktaların tespiti, anahtar noktaların seçilmesi, yönlendirme işlemi ve anahtar nokta tanımlayıcıların belirlenmesi yer alır.

Aday konumları belirlemek için etkili algoritmalar kullanan bir kademeli filtreleme yaklaşımı kullanılarak anahtar noktalar tespit edilmektedir. Anahtar nokta tespitinin ilk

aşaması, aynı nesnenin farklı görünümünde tekrarlanabilir şekilde atanan konumları ve ölçekleri belirlemektir. Görüntünün ölçek değişimine karşı değişmez konumları tespit etmek, ölçek uzayı olarak bilinen sürekli bir ölçek işlevi kullanılarak tüm olası ölçeklerde kararlı özellikler arayarak gerçekleştirilmektedir.

Ölçek uzayı çekirdeği Gauss fonksiyonu olarak belirlenmiştir. Bu nedenle, görüntünün ölçek uzayı, değişken ölçekli Gauss çekirdeği ile giriş görüntüsünün konvolüsyonundan elde edilen fonksiyon olarak tanımlanır. Bu konvolüsyon işlemi Eşitlik (17)'te verilmektedir.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (17)$$

$I(x, y)$ giriş görüntüsü, $G(x, y, \sigma)$ değişken ölçekli Gauss çekirdeğidir ve Eşitlik (17)'te tanımlanmıştır.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (18)$$

Lowe, Eşitlik (18) tarafından verilen Gauss filtresinin hesaplanmasında, birinci ve ikinci ölçek arasındaki oranın k olduğunu kabul etmiştir (Lowe, 1999). Bu durumda, Gauss çekirdeği ölçek uzayı standart sapma değerini Eşitlik (19)'da olduğu gibi hesaplar. Bu yaklaşım, Lowe'un k değerini $\sqrt{2}$ olarak seçtiği durumu açıklamaktadır.

$$\sigma_n = k^{n-1} \sigma = \sqrt{2}^{n-1} \sigma \quad (19)$$

Ölçek uzayı, üç boyutlu karşılaştırma için (x, y, σ) parametrelerine bağlı olarak $k = 2^{1/s}$ çarpım faktörü ile ayrı serilere, yani oktavlara bölünür. Her bir oktav, $s+3$ adet yumuşatılmış görüntü içerir. İkinci oktav, ilk oktavın σ 'nin yarı değeri kadar alt örnekleme yapılarak ve boyutu yarı oranına indirilmiş görüntü ile başlar. Daha sonra, bu işlem önceki oktavdan elde edilen bilgiler kullanılarak oktav sayısı kadar tekrarlanır.

Görüntünün ölçek uzayının belirlenmesi adımını takiben, her oktav için Gauss uzayı farkı (Difference of Gaussian, DoG) hesaplanır. DoG uzayı, farklı standart sapmalara sahip Gauss filtreleri ile görüntülerin konvolüsyonu yapıp elde edilen sonuçların farklarının alınmasıyla oluşturulur.

Ölçek uzayındaki istikrarlı anahtar nokta konumlarını etkin şekilde tespit etmek için, ekstremumlar, görüntüyle konvolüsyonu yapılan farklı Gauss fonksiyonundan elde edilen $D(x, y, \sigma)$ olarak ifade edilen Eşitlik (20)'daki fonksiyon kullanılarak hesaplanabilir. Bu fonksiyon, sabit çarpan faktörü k ile ayrılmış iki yakın ölçek arasındaki farktan oluşmaktadır.

$$\begin{aligned} D(x, y, \sigma) &= \left(G(x, y, k\sigma) - G(x, y, \sigma) \right) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (20)$$

Ekstremum noktalarının tespit edilmesi aşamasında, her bir DoG görüntüsündeki her piksel, alt, üst ve aynı ölçekteki dokuzar pikselin yanı sıra komşu ölçeklerdeki dokuzar piksel ile karşılaştırılır. Bir piksel, bulunduğu bölgede yerel maksimum veya yerel minimumsa, aday anahtar nokta olarak belirlenir.

DoG uzayıyla elde edilen aday özellik noktalarının elemesi sonraki adımda gerçekleştirilir. DoG operatörünün yoğun kenarlardan etkilenme eğilimi ve gürültüye hassasiyeti nedeniyle, düşük çözünürlükteki ve kenarlarda bulunan anahtar noktaların sayısı azaltılır.

Düşük çözünürlüğe sahip özellik noktalarının azaltılması için DoG uzayında $D(x, y, \sigma)$ fonksiyonunun ikinci dereceden Taylor serisi kullanılır. Bu yaklaşım, eşleşme ve kararlılık açısından önemli gelişmeler sağlar. Eşitlik (21) kullanılarak özellik noktalarının yeni konumları belirlenir ve $D(\bar{x})$ hesaplanır.

$$\bar{x} = \frac{d^2 D^{-2}}{dx^2} \frac{dD}{dx}, \quad D(\bar{x}) = D + \frac{1}{2} \frac{dD^T}{dx} \bar{x} \quad (21)$$

Her aday özellik noktası için bu fonksiyon hesaplanır ve $|D(\bar{x})|$ değeri belirli eşik değerinden küçükse, ilgili aday nokta elemeye tabi tutulur. Kenar bölgesinde tespit edilen aday anahtar noktalarını elemek için Heissan matrisi, Eşitlik (22)'de tanımlandığı şekilde kullanılır.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (22)$$

Heissan matrisi hesaplandıktan sonra, öz değerler yerine, değerlerin oransal olarak ele alınması gerektiğinden sadece oransal bilgilere odaklanmak yeterlidir. Eşitlik (23)'deki determinantın negatif olması durumunda nokta elemeye tabi tutulur.

$$Tr(H) = Dxx + Dyy = a + \beta, Det(H) = DxxDyy - (D_{xy}^2) = a\beta \quad (23)$$

Eşitlik (23) kullanılarak hesaplanan oran, belirli eşik değerinden küçükse, aday özellik noktası elemeye alınır ve sonunda kesin anahtar noktalar belirlenir.

$$\frac{Tr(H)^2}{Det(H)} = \frac{(a+\beta)^2}{a\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r}, \frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \quad (24)$$

Dönmeden bağımsızlık kazandırmak için, bir sonraki adımda belirlenen anahtar noktalara yönelik atama yapılır. Her özellik noktası etrafında gradyan yönleri ve büyüklükleri hesaplanır ve bu bölgedeki en belirgin yöne karar verilir. Bulunan en belirgin yönelim, özellik noktasının yönelimi olarak atanır.

Her $L(x, y)$ görüntü örneği için, özellik noktasının etrafındaki her piksel için, pikseller arasındaki farklılıklardan yola çıkılarak Eşitlik (25) formülü kullanılarak ilgili nokta için gradyan büyüklüğü $m(x, y)$ ve yönelimi $\theta(x, y)$ hesaplanır.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (25)$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (26)$$

360 derecelik yön aralığını kapsayan ve her biri diğerinden 10 derece açı farkına sahip olan 36 adet bine bölünmüş yön histogramı oluşturulur. Bu histogram, anahtar noktanın ölçeği olan σ 'nın 1,5 katı genişlikteki Gauss filtreli dairesel penceresindeki özellik noktalarının gradyan büyüklük değeri $m(x, y)$ ile birleştirilerek elde edilir.

Önceki adımlarda konum, ölçek ve yönler atanmış anahtar noktaların birbirinden ayırt edilebilmesi için her özellik noktası için tanımlayıcılar oluşturulur. İlk olarak, her özellik noktası etrafında 16x16 boyutunda bloklar alınır ve 45 derece açı farkı olan sekiz yönelime sahip yönelim histogramı içeren 4x4'lük 16 bloğa bölünür, böylece örnekleme

alanı oluşturulur. Her bloğun yönelimi histogramın yönelim bilgisini temsil ederken, büyüklüğü hesaplanan gradyan büyüklüğünü ifade eder.

Sekiz yönelimli histogramdaki etkinin diğer histogramlara ya da yönelimlere yumuşak şekilde kayması, şöyle ki değişikliklerin sınır etkisini azaltmak için üçlü-doğrusal (tri-linear) interpolasyon kullanılarak her gradyan örneği komşu histogram binlerine dağıtılır. Bir başka ifadeyle, her bir bindeki giriş, her boyut için $(1 - d)$ ağırlığıyla çarpılır. Burada d , binin merkez değerinden uzaklığını ifade eder.

1.3.2.2. SURF Yöntemi

SURF (Speeded Up Robust Features), 2006 yılında Herbert Bay tarafından geliştirilen özellik çıkarma algoritmasıdır (Bay vd., 2008). Bu algoritma, görüntüdeki dönme, ölçekleme ve ötelemeden bağımsız olarak yerel özellik noktalarını belirlemek için kullanılır. SIFT (Scale-Invariant Feature Transform) yöntemiyle benzer adımlara sahip olan SURF algoritması, önemli iyileştirmelere sahiptir. SIFT'e kıyasla, SURF algoritması anahtar noktalarını daha hızlı şekilde tespit edebilir ve bunu yaparken performans kaybı yaşatmaz.

Anahtar noktası tespiti için kullanılan yaklaşım, temel Hessian matrisi yaklaşımına dayanmaktadır. Bu, hesaplama süresini önemli ölçüde azaltan Viola ve Jones tarafından popüler hale getirilen integral görüntülerin kullanımına olanak tanır (Viola & Jones, 2001).

İntegral görüntüler kutu tipi konvolüsyon filtrelerinin hızlı hesaplanmasına olanak tanır ve Eşitlik (24)'de tanımlanmıştır. Bir integral görüntüsünün $I_{\Sigma}(x)$ girişi, $x = (x, y)^T$ konumunda, giriş görüntüsü I içinde ve x tarafından oluşturulan dikdörtgen bölgedeki tüm piksellerin toplamını temsil eder. İntegral görüntü hesaplandıktan sonra, herhangi dikey dikdörtgen alanın yoğunluklarının toplamını hesaplamak için üç toplama işlemi yapmak yeterlidir.

$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (27)$$

Hassaslığı nedeniyle, dedektörde Hessian matrisi kullanılmaktadır. Determinantın maksimum olduğu yerlerde blog benzeri yapılar tespit edilmektedir.

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (28)$$

$Lxx(x, \sigma)$ Gaussian ikinci derece türevinin $\frac{\partial^2}{\partial x^2} g(\sigma)$ x noktasında I görüntüsünün konvolüsyonudur.

Gauss fonksiyonları, ölçek-uzay analizi için optimaldir, ancak pratikte diskretize edilmeli ve kırılmalıdır. Bu, $\frac{\pi}{4}$ nin tek sayı katları etrafında görüntü dönüşümlerinde tekrarlama kaybına yol açar. Bu zayıflık genel olarak Hessian tabanlı dedektörler için geçerlidir. Tekrarlama, $\frac{\pi}{2}$ nin katları etrafında maksimuma ulaşır. Bu, filtrelerin kare şekline kaynaklanır. Hessian matrisi yaklaşımını kutu filtrelerle kullanılmaktadır. Bu, yaklaşık ikinci derece Gaussian türevleri ve integral görüntüler kullanılarak çok düşük hesaplama maliyetiyle sonuçlanır. Hessian determinantının Eşitlik (29)'da verilmektedir. Hesaplama süresi filtre boyutundan bağımsızdır. Filtre tepkilerinin göreceli ağırlığı w , Hessian determinantının ifadesini dengelemek için kullanılır. Bu, Gauss çekirdekleri ile yaklaşıklıkla elde edilen Gauss çekirdekleri arasındaki enerji korunumunu sağlamak için gereklidir.

$$\det(H) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (29)$$

$$w = \frac{|L_{xy}(1.2)|_F |D_{yy}(9)|_F}{|L_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.9 \quad (30)$$

$|x|_F$ Frobenius normunu ifade eder. Ağırlık ölçeğe bağlı olarak değişir.

Ölçek uzayları genellikle bir görüntü piramidi olarak uygulanır. Görüntüler iteratif olarak Gauss ile yumuşatılır ve daha yüksek piramit seviyesi elde etmek için örnekleme yapılır. Lowe, bu piramit katmanlarını çıkararak kenarların ve blogların bulunabildiği Gauss Farkı (DoG - Difference of Gaussians) görüntülerini elde eder.

Kutu filtrelerinin ve integral görüntülerin kullanılmasından dolayı, önceki filtrelenmiş katmanın çıktısına aynı filtre iteratif olarak uygulanmaz, ancak isteğe bağlı olarak herhangi boyuttaki kutu filtrelerini tam olarak aynı hızda doğrudan orijinal görüntüye uygulanır. Bu nedenle, Ölçek uzayı, görüntü boyutunu iteratif olarak azaltmak yerine filtre boyutunu yükselterek analiz edilir. Solda görüntü boyutunun küçültülerek filtre uygulanması gösterilirken sağda filtre boyutunun büyütülerek görüntüye uygulanması verilmektedir.

Ölçek uzayı, oktavlara bölünmüştür. Bir oktav, aynı giriş görüntüsünün giderek artan boyutta filtre ile konvolüsyonu ile elde edilen bir dizi filtre tepki haritasını temsil eder.

Toplamda, bir oktav, bir ölçek faktörünü kapsar. Her oktav, sabit sayıda ölçek seviyesine bölünmüştür.

Her yeni oktav için, filtre boyutu artışı iki katına çıkarılır (6–12'den 24–48'e). Aynı zamanda, anahtar noktaların çıkarılması için örnekleme aralıkları her yeni oktav için iki katına çıkarılabilir. Bu, hesaplama süresini azaltır. İkinci oktav için filtre boyutları 15, 27, 39, 51'dir. Üçüncü oktav, filtre boyutları 27, 51, 75, 99 ile hesaplanır ve eğer orijinal görüntü boyutu hala karşılık gelen filtre boyutlarından daha büyükse, ölçek uzayı analizi dördüncü oktav için gerçekleştirilir, bu da filtre boyutları 51, 99, 147 ve 195 kullanılarak yapılır.

Ölçek-uzay yapısında, Hessian determinantlarının değerlerine dayanarak özellik noktaları belirlenir. Ardışık üç ölçekten 3x3'lük alanlar seçilir ve toplamda 27 piksel arasından en yüksek gradyan değerine sahip olanı özellik noktası olarak belirlenir.

Özellik noktalarından tanımlayıcıyı çıkarmak için ilk adım, özellik noktasının etrafında merkezlenmiş kare bölge oluşturmaktır. Bu pencerenin boyutu 20'dir. Bölge daha sonra küçük 4x4 kare alt bölgelere bölünür. Bu, önemli mekansal bilgilerin korunmasını sağlar. Her alt bölge için, 5x5 düzenli aralıklı örnek noktasında Haar dalgaletini yanıtlarını hesaplanır. Yatay yönde Haar dalgaletini yanıtını dx ve dikey yönde Haar dalgaletini yanıtını dy olarak adlandırılmaktadır. Geometrik deformasyonlara ve lokalizasyon hatalarına karşı sağlamlığı artırmak için, dx ve dy önce Gaussian ($\sigma = 3.3s$) ile ağırlıklı hale getirilir.

Ardından, dalgalet yanıtları dx ve dy her alt bölge üzerinde toplanır ve özellik vektöründeki ilk girişlerin setini oluşturur. Yoğunluk değişikliklerinin polaritesi hakkında bilgi getirmek için, yanıtların mutlak değerlerinin toplamı, $|dx|$ ve $|dy|$ de hesaplanmaktadır. Bu nedenle, her alt bölge için 4D tanımlayıcı vektörü $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$ vardır ve bu, temel yoğunluk yapısı için vektördür. Tüm 4x4 alt bölgeler için bunları birleştirerek, 64 uzunluğunda tanımlayıcı vektör elde edilir.

1.3.2.3. BRIEF Anahtar Nokta ve Tanımlayıcıları

Önerilen bu metod spektrogram görüntülerinden tespit edilebilen belirgin noktaları ve bunların özellik tanımlayıcılarını çıkarmak için kullanılmaktadır. Genel olarak, belirgin noktalar, görüntünün mekansal konumları veya dikkat çekici alanlarıdır. Nokta tespiti aslında blob tespitinin bir alt kümesidir ve görüntüde dikkate değer bölgeleri veya mekansal alanları belirlemeyi hedefler. Bu noktalar, ek saldırılara maruz kalsa bile değişmezler. Anahtar noktaları Merkez Çevre Ekstrema (Center Surround Extrema, CenSurE) yöntemiyle

ve özellik tanımlayıcılarını ise İkili Sağlam Bağımsız Temel Özellikler (Binary Robust Independent Elementary Features, BRIEF) ile elde edilmektedir.

CenSurE'de, merkez-çevre filtresi, görüntüdeki tüm konumlarda ve ölçeklerde uygulanır ve zayıf köşe noktalarını belirlemek için Harris fonksiyonu kullanılır (Agrawal vd., 2008). Bu yöntemde, çekirdek içindeki merkezi piksel ile çekirdek bölgesindeki diğer pikseller arasındaki piksel karşılaştırmaları kaydedilir. Temelde, tüm konumlarda ve ölçeklerde basitleştirilmiş merkez-çevre filtresi hesaplanır ve ardından yerel komşulukta ekstremumlar belirlenir. Son adımda, elde edilen ekstremumlar, Harris ölçüsü hesaplanarak filtrelenir ve zayıf köşe tepkisine sahip olanlar elimine edilir.

Her pikselin çevresindeki bölgenin ve merkezindeki pikselin farkı hesaplanır. Bu fark, merkezdeki pikselin ve çevresindeki piksellerin yoğunlukları arasındaki değişiklikleri yansıtır. Ardından, bu farklar belirli özelliklerle değerlendirilir ve belirlenen kriterlere göre anahtar noktalar seçilir. Genellikle, piksel değerlerindeki büyük değişikliklere sahip bölgeler, potansiyel anahtar noktalar olarak kabul edilir. Son olarak, bu aday anahtar noktaları filtrelenir ve düzenlenir, gereksiz olanlar kaldırılır ve önemli olanlar belirlenir. Bu şekilde, CenSurE görüntülerdeki önemli yapısal özellikleri temsil eden anahtar noktaları belirler. Bu yöntem, özellikle hızlı hesaplama yetenekleri ve gerçek zamanlı uygulama olasılıkları nedeniyle yaygın olarak kullanılır.

Harris köşe algılayıcısı, görüntülerdeki köşe benzeri yapıları tespit etmek için kullanılan yöntemdir. Bu yöntem, piksel grubunun yoğunluk değişimlerinin belirli yönde en fazla olduğu noktaları tespit ederek köşe noktalarını bulur. İlk adımda, her pikselin x ve y yönlerindeki yoğunluk gradyanları hesaplanır. Bu, görüntünün her bir pikselinde yoğunluk gradyanı vektörlerinin elde edilmesini sağlar. Yoğunluk gradyanlarından elde edilen vektörler kullanılarak, her pikselin etrafındaki piksellerin yoğunluk gradyanlarının kovaryans matrisi hesaplanır. Bu matris, pikselin etrafındaki yoğunluk gradyanlarının dağılımını gösterir. Kovaryans matrisi kullanılarak, her piksel için bir köşe ölçüsü hesaplanır. Bu ölçü, pikselin etrafındaki yoğunluk gradyanlarının dağılımının köşe benzeri yapı oluşturup oluşturmadığını değerlendirir. Hesaplanan köşe ölçüleri, eşik değeriyle karşılaştırılarak köşe olarak kabul edilip edilmeyeceğine karar verilir. Eşik değerini geçen pikseller, köşe olarak işaretlenir ve anahtar noktaları olarak belirlenir.

Anahtar noktaları çıkardıktan sonra, onları temsil etmek için tanımlayıcılar elde edilmektedir. Literatürde birçok özellik tanımlayıcı yöntemleri mevcuttur, ancak bunların birçoğunun hesaplama maliyeti yüksektir. Bazı çalışmalarda, özellik vektörlerini elde

etikten sonra eşleme aşamasını hızlandırmak için ek boyut indirgeme veya hash (hashing) yaklaşımları kullanılır. Ancak, bu ek işlemlerin de hesaplama maliyeti olduğu açıktır. Doğrudan görüntü yamalarından ikili dizileri hesaplayan İkili Sağlam Bağımsız Temel Özellikleri (BRIEF) kullanmaktadır (Calonder vd., 2010). İkili özellik tanımlayıcıları, özellik tanımlayıcıları eşleştirilirken hesaplama maliyeti açısından büyük avantaj sunar. BRIEF yöntemi, temel olarak, ilgi noktası etrafındaki kare görüntü yamalarını ve bu yamaların Gauss çekirdeğine sahip düzleştirilmiş versiyonunu karşılaştırır. Amaç, yamaların düzleştirilmiş formu ile karşılaştırarak gürültüye karşı sağlamlık elde etmektir.

BRIEF tanımlayıcıları, görüntüdeki özellikleri temsil etmek için kullanılan ikili (binary) vektörlerdir. Bu tanımlayıcılar, anahtar noktaların etrafındaki piksel çiftlerinin karşılaştırılmasıyla elde edilir ve özellikle hızlı hesaplama ve yüksek işlemci performansı gerektiren uygulamalarda tercih edilir. İlk adımda, anahtar noktaların etrafında belirli aralıklarla piksel çiftleri seçilir. Bu piksel çiftleri, anahtar noktanın etrafında rastgele şekilde belirlenebilir. Her piksel çifti için, biri diğerine göre ne kadar farklı olduğunu belirlemek için karşılaştırma işlemi gerçekleştirilir. Bu karşılaştırma, piksel çiftlerinin gri tonlamalı değerleri arasındaki farkı kullanır. Karşılaştırma sonuçları, eşik değeri ile karşılaştırılır ve ikili değere dönüştürülür. Örneğin, fark negatifse, ikili değer 0 olarak atanır; aksi halde, 1 olarak atanır. Bu süreç Eşitlik (31)'de verilmektedir. Tüm piksel çiftleri için yapılan karşılaştırmalar sonucunda elde edilen ikili değerler bir araya getirilerek özellik vektörü oluşturulur. Bu vektör, anahtar noktanın özelliklerini temsil eder.

Görüntü yaması etrafındaki bir nokta x 'in piksel yoğunluğunu $\rho_i(x)$ olarak belirtelim. Benzer şekilde, anahtar nokta y ve $p_i(y)$ de piksel yoğunluğunu ifade eder. Yöntem, bu yoğunluklar için ikili testler gerçekleştirir ve

$$\tau(\rho_i; x, y) = \begin{cases} 1 & : \rho_i(x) < p_i(y) \\ 0 & : \rho_i(x) \geq p_i(y) \end{cases} \quad (31)$$

şeklinde hesaplanır. Sonra, Eşitlik (31)'de verilen τ cevabından ikili özellik vektörü $f_d(P_i)$ şu şekilde oluşturulur.

$$f_d(P_i) = \sum_{1 \leq j \leq d} 2^{j-1} \tau(P_i; x_j, y_j) \quad (32)$$

x ve y , görüntü yamasındaki rastgele seçilmiş görüntü noktalarıdır ve bu rastgele çiftlerin konumlarını belirlemek için çeşitli dağılım yaklaşımları önerilmiştir, bunlar arasında uniform, Gauss ve kaba polar ızgara bulunmaktadır.

Oluşturulan özellik vektörü, anahtar noktanın tanımlayıcı özelliklerini sağlar. Bu tanımlayıcılar, genellikle diğer özellik vektörleriyle karşılaştırılarak nesne tanıma, eşleştirme veya sınıflandırma gibi görevlerde kullanılır. BRIEF tanımlayıcılar, özellikle hızlı hesaplama yetenekleri ve yüksek performanslı bilgisayarlar için uygun olmaları nedeniyle geniş bir uygulama alanına sahiptir. Bu tanımlayıcılar, anahtar noktaların özelliklerini basit ve etkili bir şekilde temsil etmek için yaygın olarak kullanılır ve bilgisayar görüşü ve benzeri alanlarda sıkça tercih edilirler.

1.3.2.4. AKAZE

Literatürde en çok kullanılan anahtar nokta ve özellik çıkarım yöntemleri SIFT ve SURF'tur, SURF, SIFT'ten daha az hesaplama gerektirir. Her iki yaklaşım da Gauss ölçek uzayından yararlanır, ya SIFT gibi piramidal bir çerçevede Gauss ölçek uzayı oluşturarak ya da SURF gibi Gauss türevlerini kutu filtreleri aracılığıyla kullanarak. Bu durum, Gauss bulanıklığının nesne sınırlarını korumadığı ve tüm ölçeklerde ayrıntıları ve gürültüyü aynı ölçüde yumuşattığı ve belirginliği bozduğu dezavantajlarına sahiptir. Bu tür dezavantajları ortadan kaldırmak için AKAZE yöntemi önerilmiştir (Alcantarilla vd., 2013).

AKAZE, Fast Explicit Diffusion (FED) adı verilen hızlı ve doğrusal olmayan yayılma filtreleme yöntemlerini kullanarak özellikleri tespit etme ve tanımlama işlemlerini hızlandırır. AKAZE, nesne sınırlarını koruyarak ayrıntıları belirsizleştiren ve nesneler arasındaki benzersiz özellikleri vurgulayan doğrusal olmayan ölçek uzaylarını kullanır. Özellikleri tanımlamak için kullanılan yöntemler sayesinde dönme ve ölçek değişikliklerine karşı dayanıklıdır. düşük depolama gereksinimlerine sahip olduğu için bellek kullanımını optimize eder.

Doğrusal olmayan dağılım filtreleme, görüntünün parlaklığının artan ölçek seviyelerinde belirli bir akış fonksiyonunun çeşitli diverjansı olarak tanımlanır ve bu fonksiyon difüzyon sürecini kontrol eder. Bu yaklaşımlar, görüntünün parlaklığını doğrusal olmayan ölçek uzayı boyunca difüzyonunu sağlayan diferansiyel denklemlere dayanır. Dağılım denklemi Eşitlik (33)'de verilmektedir.

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \quad (33)$$

Burada div ve ∇ sırasıyla diverjans ve gradyan operatörleridir ve L görüntünün parlaklığını temsil eder. Difüzyon denkleminde iletkenlik fonksiyonu c tanıtılması sayesinde, difüzyonun yerel görüntü yapısına adapte edilmesi mümkün olur. Fonksiyon c , yerel görüntü farkı yapısına bağlıdır ve bu fonksiyon skaler veya tensör olabilir. Zaman t ölçek parametresidir ve daha büyük değerler basitleştirilmiş görüntü temsillerine yol açar. Anizotropik difüzyonda görüntü gradyan büyüklüğü, her ölçek seviyesinde difüzyonu kontrol eder. Bu şekilde, iletkenlik fonksiyonu c aşağıdaki gibi tanımlanır:

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|) \quad (34)$$

Burada, ∇L_σ fonksiyonu, orijinal görüntünün Gauss filtresiyle yumuşatılmış versiyonunun gradyanını ifade eder. Perona ve Malik'in çalışmasında tanıtılan iki iletkenlik fonksiyonundan birini kullanılmaktadır (Perona & Malik, 1990). g iletkenlik fonksiyonu, daha geniş bölgeleri küçük olanlara tercih eder:

$$g = \frac{1}{\frac{1+|\nabla L_\sigma|^2}{\lambda^2}} \quad (35)$$

λ kontrast faktörüdür ve difüzyon seviyesini kontrol eder. Hangi kenarların artırılacağını veya korunacağını ve hangilerinin iptal edileceğini belirler. Doğrusal olmayan difüzyon denklemini çözmek için analitik çözümler olmadığından, örnekleme yöntemiyle ayırtırmak gerekmektedir. Bunun için yazarlar Fast Explicit Diffusion (FED) şemasını kullanmaktadırlar (Grewenig vd., 2010).

Bu şemada temel fikir, kutu filtresinin faktörizasyonundan kaynaklanan değişen adımlı τ_j adımlarıyla n adet açık difüzyon adımının M döngüsünü gerçekleştirmektir.

$$\tau_j = \frac{\tau_{max}}{2 \cos^2\left(\pi \frac{2j+1}{4n+2}\right)} \quad (36)$$

burada τ_{max} , kararlılık koşulunu ihlal etmeyen maksimum adım büyüklüğüdür. Bir FED döngüsünün karşılık gelen durma zamanı θ_n ,

$$\theta_n = \sum_{j=0}^{n-1} \tau_j = \tau_{max} \frac{n^2+n}{3} \quad (37)$$

şeklinde ifade edilir. Eşitlik (37)'nin ayrıklaştırılması, vektör-matris gösterimi olarak ifade edilebilir:

$$\frac{L^{i+1}-L^i}{\tau} = A(L^i)L^i \quad (38)$$

Burada $A(L^i)$ görüntünün iletkenliğini kodlayan matristir ve τ ifadesi de $\tau < \tau_{max}$ eşitliğini sağlayan sabittir.

$$L^{i+1} = (I + \tau A(L^i)) L^i \quad (39)$$

I , kimlik matrisidir. Önceki varsayımlardan yola çıkarak, $L^{i+1,0} = L^i$ olacak şekilde, n adet τ_j adım boyutlu FED döngüsü şu şekilde elde edilir:

$$L^{i+1,j+1} = (I + \tau_j A(L^i)) L^{i+1,j} \quad (40)$$

FED döngüsü sırasında, $A(L^i)$ 'nin değeri sabittir ve döngü tamamlandıktan sonra, $A(L^i)$ matrisinin yeni değerleri hesaplanır.

Doğrusal olmayan ölçek uzayını oluşturmak için belirli evrim süresi kümesi tanımlanması gerekmektedir. Bu ölçek uzayı, O adet oktav ve S adet alt-seviyede dizi olarak düzenlenir. Her oktav ve alt-seviye, tam sayı oktav indisine (o) ve alt-seviye indisine (s) sahiptir. Oktav ve alt-seviye indisleri, karşılık gelen ölçek değerine (s piksel cinsinden) karşılık gelir.

$$\sigma_i(o, s) = 2^{0+s/S}, o \in [0 \dots O-1], s \in [0 \dots S-1], i \in [0 \dots M] \quad (41)$$

M toplam filtrelenmiş görüntü sayısıdır. Doğrusal olmayan difüzyon filtreleme zaman birimlerinde çalıştığı için piksel birimlerindeki σ_i ölçek seviyeleri kümesini zaman

birimlerine dönüştürmek gerekmektedir. Piksel birimlerinden zaman birimlerine dönüştürmek için Eşitlik (42)'de tanımlanan denklem kullanılmaktadır.

$$t_i = \frac{1}{2} \sigma_i^2, \quad i = \{0 \dots M\} \quad (42)$$

Ek olarak, giriş görüntüsüne, gürültüyü azaltmak için standart sapması σ_0 olan Gauss filtresi uygulanır. Düzleştirilmiş giriş görüntüsünden, kontrast faktörü λ , gradyan histogramının %70'lik dilimine karşılık gelen değerden otomatik olarak hesaplanır.

Ölçek uzayının oluşturulması, verilen giriş görüntüsü ve kontrast faktörü ile $M - 1$ kez çağırılarak FED döngüsünün yürütülmesi ile gerçekleştirilir. Her FED döngüsü için minimum adım sayısı olan n hesaplanır. Giriş görüntüsünün iki boyutlu olduğu durumda, maksimum adım büyüklüğü $\tau_{max} = 0.25$ olarak kabul edilir. Her FED döngüsünün zamanı $T = t_{i+1} - t_i$ olarak hesaplanır.

Her oktavda son alt-seviyeye ulaşıldığında, görüntüyü $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$ şeklindeki düzleştirme maskesiyle 2 katına kadar örnekleme yapılır ve bu örneklenmiş görüntüyü bir sonraki oktavdaki sonraki FED döngüsü için başlangıç görüntüsü olarak kullanılır. Görüntüyü örnekledikten sonra, kontrast parametresi λ 'de değiştirilir. Düzleştirme maskesi ideal adım kenarın kontrastını %25 azaltır ve bu nedenle kontrast parametresi 0.75 ile çarpılmalıdır. Doğrusal olmayan difüzyon filtreleme için piramidal FED yaklaşımını Şekil M'de verilmektedir. Aynı şekilde, FED iç döngü iterasyonları, her görüntünün doğrusal olmayan ölçek uzayındaki oktavını dikkate alan normalleştirilmiş ölçek faktörü kullanılarak Algoritma 1'de açıklanmıştır.

Algoritma 1. Doğrusal olmayan ölçek uzayının oluşturulması algoritması

Girdi. Orijinal görüntü L^0 , kontrast parametresi λ , τ_{max} ve evrim zamanı dizisi t_i

Çıktı. Filtrelenmiş görüntüler dizisi L^i $i = 0..M$

Adım 1. Filtrelenmiş görüntüler dizisi L^i tanımla

Adım 2. İletkenlik matrisinin $A(L^i)$ Hesaplanması

Adım 3. FED döngüsünün süresi $T = t_{i+1} - t_i$

Adım 4. FED adım sayısı n ' in hesaplanması

Adım 5. Adım büyüklüğü τ_j 'nin hesaplanması

Adım 6. $L^{i+1} = \mathbf{FED}(L^{i+1,0}, A(L^i), \tau_j)$

Adım 7. if $\sigma_{i+1} > \sigma_i$

$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ maskesiyle birlikte L^{i+1} görüntüsünün aşağı örneklendirilmesi

$\lambda = \lambda \times 0.75$

Adım 8. Adım 6 ve 7'i $L^{i+1,0} = L^i$ koşulu sağlanana kadar devam ettir.

Adım 9. Adım 2-6'i M-1 sayısınca çalıştır.

Algoritma 2. FED döngüsü algoritması

Girdi. $L^{i+1,0}, A(L^i), \tau_j$

Çıktı. $L^{i+1,n}$

Adım 1. $L^{i+1,j+1} = (I + \tau_j A(L^i)) L^{i+1,j}$

Adım 2. Adım 1'i n – 1 sayısınca tekrarla

Doğrusal olmayan ölçek uzayındaki her filtrelenmiş görüntü L^i için Hessian determinantı Eşitlik (44)'daki gibi hesaplanır. Farklı ölçeklerdeki diferansiyel çok ölçekli operatörler, ölçekle ilgili olarak normalize edilir; bu, doğrusal olmayan ölçek uzayındaki her görüntünün oktavını dikkate alan normalize edilmiş ölçek faktörü kullanılarak yapılır. Normalizasyon işlemi Eşitlik (43)'de verilmektedir.

$$\sigma_{1,norm} = \sigma_i / 2^{0^i} \quad (43)$$

$$L_{Hessian}^i = \sigma_{i,norm}^2 (L_{xx}^i L_{yy}^i - L_{xy}^i L_{xy}^i) \quad (44)$$

Mekansal konumdaki detektör tepkisinin maksimumları aranır. Her evrim seviyesi i için, detektör tepkisinin önceden tanımlanmış eşikten yüksek olduğunu ve 3×3 piksellik pencerede maksimum olduğu kontrol edilir. Bu, maksimum olmayan tepkileri hızlı şekilde eler. Ardından, her potansiyel maksimum için, tepkinin, sırasıyla üstteki ve alttaki $i + 1$ ve $i - 1$ seviyelerinden diğer anahtar noktalara göre maksimum olup olmadığını kontrol edilir, bu da $\sigma_i \times \sigma_i$ piksellik pencerede yapılır.

Doğrusal olmayan ölçek uzayından gradyan ve yoğunluk bilgisini kullanan Modifiye-Yerel Fark Binomo (M-LDB) adını taşıyan tanımlayıcı kullanılmaktadır. LDB tanımlayıcısı

BRIEF ile aynı prensibi takip eder, ancak ekstra sağlamlık için tek pikseller yerine alanların ortalaması arasında ikili testler kullanır. Yoğunluk değerlerinin yanı sıra, karşılaştırılan alanlardaki yatay ve dikey türevlerin ortalaması da kullanılır, bu da karşılaştırma başına 3 bit sağlar. LDB, 2x2, 3x3, 4x4 vb. ızgaraların yama bölgesine bölünmesini önerir. Ancak bu alt bölmelerin ortalamalarını hesaplamak, tanımlayıcı yukarı doğru (dönme değişmez) olduğunda integral görüntüler kullanılarak çok hızlıdır. Bununla birlikte, anahtar noktaların rotasyonunu dikkate aldığımızda integral görüntüler kullanılamaz ve dönüştürülmüş bölümdeki tüm noktaları ziyaret etmek nispeten hesaplama süresinde maliyete neden olur. Dönme invariansı, anahtar noktanın ana yönelimini KAZE'de olduğu gibi tahmin ederek ve LDB'nin ızgarasını buna göre döndürerek elde edilir. Izgaranın her alt bölümündeki tüm piksellerin ortalamasını kullanmak yerine, ızgaraları özellik ölçeğinin fonksiyonu olan adımlarda alt örnekleme yapılır.

1.3.3. Özellik Tanımlayıcıların Kümelendirilmesi Yaklaşımları

Bu bölümde görüntü üzerinde Anahtar Nokta ve Tanımlayıcıların çıkarılması algoritmaları kullanılarak çıkarılan Anahtar Nokta ve onların Tanımlayıcılarının eşleştirilmesi amacıyla kullanılan kümelendirme yapıları detaylı şekilde açıklanmaktadır.

1.3.3.1. K-means kümeleme

K-means kümeleme, N adet veri nesnesinden oluşan veri kümesini K adet kümeye bölme yöntemidir (Khaled vd., 1998). Temel amaç, kümelerin içindeki veri noktalarının birbirlerine olan benzerliğini maksimize etmek ve kümeler arasındaki farklılıkları minimize etmektir. Bu yöntem, veri setini belirli sayıda kümeye ayırarak her kümeyi bir merkez noktasıyla temsil eder ve bu merkez noktalarını iteratif olarak güncelleyerek en iyi kümeleri bulmaya çalışır.

K-means, yaygın olarak kullanılan kümeleme algoritmasıdır çünkü basit ve verimli şekilde uygulanabilir. Büyük veri kümelerini hızlı şekilde işleyebilir. K sabiti, algoritmanın başlamadan önce belirlenmesi gereken küme sayısını temsil eder. Algoritma, her veri noktasını bir kümeye atayarak toplam uzaklığı minimize etmeye çalışır. K-means, kare hatayı minimize edecek en uygun K kümesini bulmaya odaklanır.

K-means ile küme içi benzerlik büyük, kümeler arası benzerlik ise küçük olduğu sürece kümelenmenin doğruluğundan söz edilebilir. K-means algoritması iteratif (tekrarlayıcı) yaklaşım ile genelde iyi bir çözüm verir.

K-means kümeleme, çok boyutlu gerçel vektörlerden oluşan veri kümesi $\{x_1, x_2, \dots, x_N\}$ ve kümeleme için belirlenen K sayısı verildiğinde, kare hatayı en aza indirmek için veriyi K kümeye bölme işlemidir, özetle $S = \{s_1, s_2, \dots, s_K\}$ kümelerini oluşturmayı hedefler. Küme içindeki noktaların ortalaması Eşitlik (45)'ta verilen ifadeyle hesaplanmaktadır.

$$\mu_i = \frac{1}{|s_j|} \sum_{x_i \in s_j} x_i \quad (45)$$

μ_i, s_j 'deki noktaların ortalamasıdır. K-means kümeleme yaklaşımında amaç tüm kümeler için Eşitlik (46)'ın sağlanmasıdır.

$$\underset{S}{argmin} \sum_{j=1}^K \sum_{x_i \in s_j} \|x_i - \mu_i\|^2 \quad (46)$$

K-means algoritması, veri noktalarını belirli sayıda küme merkeziyle gruplamak için kullanılan kümeleme algoritmasıdır. İşleyişi şu adımlarla gerçekleşir:

- İlk olarak, her biri bir küme merkezi temsil edecek şekilde rastgele k merkez seçilir.
- Her veri noktası, en yakın olan küme merkezine atanır. Bu atama, genellikle Öklid mesafesi veya benzer bir ölçü ile yapılır.
- Her küme için yeni merkez hesaplanır. Bu, o kümeye atanmış olan tüm veri noktalarının ortalamasının alınmasıyla gerçekleştirilir.
- Atama ve merkez hesaplama adımları arka arkaya tekrarlanır. Atama yapıldıktan sonra, merkezler güncellenir ve bu güncellenmiş merkezlerle tekrar atama yapılır. Bu işlem, merkezlerin konumlarının sabitlenene veya durma kriteri sağlanana kadar devam eder.
- Algoritma, merkezlerin hareket etmeyeceği veya atama değişmeyeceği noktaya ulaştığında sona erer.

K-means, genellikle basit ve etkili kümeleme algoritması olarak kullanılır, ancak başlangıç merkezlerinin rastgele seçilmesi ve algoritmanın global optimuma ulaşmayabileceği gibi bazı sınırlamalara sahiptir.

1.3.3.2. Genelleştirilmiş 2 En Yakın Komşu

f^A veri noktaları olarak verilen özelliklerin g2NN ile kümelendirilmesi aşağıdaki yöntemle gerçekleştirilmektedir (Lu vd., 2022).

$$f^A = \{ f_1, f_2, \dots, f_n \} \quad (47)$$

f^A n boyutludur. X_i f_i^A ve f_j^A diğer $n - 1$ özelliğin Öklid Uzaklığı olsun.

$$X_i = \{d_{i,1}, d_{i,2}, \dots, d_{i,n}\} \quad (48)$$

$d_{i,j}$ ($i, j = 1, 2, \dots, n; i \neq j$) f_i^A ve f_j^A arasında Öklid uzaklığıdır.

$$d_{i,j} = \|f_i^A - f_j^A\|^2 \quad (49)$$

Uygun eşik değerini bulmak için, tanımlayıcılar arasındaki benzerlik $d_{i,j}^2$ (Öklid uzaklık karesi) kullanarak ölçülür. Bu nedenle, tüm f^A özellikleri için, $n \times (n-1)$ boyutunda bir ξ matrisi oluşturulur.

$$\xi = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix} = \begin{bmatrix} d_{1,2}^2 & \dots & d_{1,n}^2 \\ \vdots & \ddots & \vdots \\ d_{n,1}^2 & \dots & d_{n,n-1}^2 \end{bmatrix} \quad (50)$$

ξ uzaklık matrisinin her satırı üzerinde 2 en yakın komşu (2NN) algoritması tekrarlanır. X_i 'yi örnek olarak kullanırsak, iterasyon, şu durumlarda durur:

$$d_{i,j}^2 / d_{i,j+1}^2 < T \quad (51)$$

$T \in (0, 1)$ aralığında. Eğer iterasyon $d_{i,k}^2$ 'de durursa, İncelenen anahtar nokta için $\{d_{i,1}^2, d_{i,2}^2, \dots, d_{i,k}^2\}$ ($k = 1, 2, \dots, n; k \neq i$) içindeki her uzaklık, eşleşme olarak kabul edilir.

Daha yüksek eşik değeri daha fazla yanlış eşleşmeye yol açarken, daha düşük eşik değeri bazı doğru eşleşmeleri kaçırabilir. Uygun eşik değerinin sadece mümkün olduğunca çok sayıda doğru eşleşmeyi elde etmekle kalmayıp, aynı zamanda yanlış eşleşme sayısını kabul edilebilir sınırlar içinde tutması gerektiğine inanılır.

Genelleştirilmiş 2 En Yakın Komşu (generalized 2 nearest neighbor), veri noktalarının birbirlerine olan benzerliğini değerlendirmek için kullanılan algoritmadır. Genellikle, veri noktasının en yakın iki komşusunu belirlemek için kullanılır. Bunun için, her veri noktası için diğer tüm veri noktalarıyla olan uzaklıklar hesaplanır ve bu uzaklıklar sıralanır. Daha sonra, her veri noktası için en yakın iki komşu seçilir. Genelleştirilmiş 2 en yakın komşu algoritması, veri noktasının sadece en yakın iki komşusunu değil, aynı zamanda belirli eşik değerinin altında olan tüm komşularını da dikkate alır. Bu şekilde, veri noktasının benzerliklerini değerlendirirken daha esnek yaklaşım sunar.

Algoritma genellikle şu adımları izler:

- Her veri noktası için diğer tüm veri noktalarıyla olan uzaklıklar hesaplanır.
- Uzaklıklar sıralanır ve en yakın iki komşu belirlenir.
- Belirli eşik değerinin altında olan tüm komşular dikkate alınarak eşleşme tamamlanır.

Gerçek veri setlerinin birçoğunun önemli özelliği, içsel kümelenme yapısının global yoğunluk parametreleri kullanılarak doğru şekilde tanımlanamamasıdır. Farklı bölgelerdeki kümelerin belirlenmesi için, farklı yerel yoğunluklara ihtiyaç duyulabilir.

1.3.3.3. OPTICS

Gerçek veri setlerinin birçoğunun önemli özelliği, içsel kümelenme yapısının global yoğunluk parametreleri kullanılarak doğru şekilde tanımlanamamasıdır. Farklı bölgelerdeki kümelerin belirlenmesi için, farklı yerel yoğunluklara ihtiyaç duyulmaktadır. Global

yoğunluk parametresine uygun olmayan örnekler gürültü olarak tanımlanmakta ve kümelenmemektedir.

Kümeleme yapılarını tespit etmek ve analiz etmek için birinci alternatif, örneğin tek bağlantı yöntemi gibi hiyerarşik kümeleme algoritması kullanmaktır. Ancak, bu seçeneğin iki dezavantajı vardır. Genel olarak, tek bağlantı etkisiyle oldukça muzdariptir; bir başka ifadeyle, küçük bir nesne arası mesafeye sahip birkaç nokta hattı ile birbirine bağlı olan kümeler ayrılamazlar. İkinci olarak, hiyerarşik algoritmalar tarafından üretilen sonuçlar olan dendrogramları anlamak veya analiz etmek, birkaç yüzden fazla nesne için zor olabilir.

İkinci alternatif, farklı parametre ayarlarıyla yoğunluk tabanlı bölütleme algoritması kullanmaktır. Ancak, bu yaklaşımın da bazı dezavantajları vardır. Sonsuz sayıda olası parametre değeri olduğundan, sonuçları analiz etmek zordur. Birçok farklı değeri kullanmak, her nokta için farklı küme üyeliklerini saklamak için çok miktarda ikincil bellek gerektirir ve ilginç kümelenme seviyelerini kaçırma riskini taşır.

Bu sorunları aşmak için temel fikir, veri setinin yoğunluk tabanlı kümeleme yapısına göre özel sıralama oluşturan ve her kümeleme seviyesi hakkında bilgi içeren algoritma kullanmaktır. Bu algoritma sonuçları çok kolay analiz edilebilir ve anlaşılabilir kılar.

Yoğunluk tabanlı kümeleme (Density-Based Clustering)'nin temel fikri, kümenin her nesnesi için verilen yarıçap (ϵ) içindeki komşulukta en az bir minimum nesne sayısının *MinPts* bulunması gerektiğidir; kısacası, komşuluk kümesinin büyüklüğünün eşik değerini aşması gerekmektedir.

DBSCAN algoritması (Ester vd., 1996), veritabanındaki kümeleri ve gürültüyü keşfeden algoritmadır. Bu algoritma, kümenin, kümedeki herhangi bir çekirdek nesnesinden yoğunluk bakımından erişilebilen tüm nesnelerin kümesine eşit olduğu gerçeğine dayanmaktadır. Yoğunluk bakımından erişilebilen nesnelerin alınması, doğrudan yoğunluk bakımından erişilebilen nesnelerin toplanmasıyla iteratif olarak gerçekleştirilir. DBSCAN, veritabanındaki her noktanın ϵ komşulukları (e-komşuluklarını) kontrol eder. Bir nokta p 'nin ϵ -komşuluk alanı $N_\epsilon(p)$ *MinPts* noktadan daha fazlaysa, $N_\epsilon(p)$ içindeki nesneleri içeren yeni küme C oluşturulur. Ardından, henüz işlenmemiş C içindeki tüm q noktalarının ϵ -komşulukları kontrol edilir. Eğer $N_\epsilon(q)$ *MinPts* noktadan daha fazla içeriyorsa, zaten C içinde bulunmayan q 'nun komşuları kümeye eklenir ve bir sonraki adımda onların ϵ -komşulukları kontrol edilir. Bu işlem, artık mevcut küme C 'ye yeni nokta eklenemeyene kadar tekrarlanır.

Yoğunluk tabanlı küme sıralaması kavramını tanıtmak için öncelikle şu gözlem tanıtılmalıdır: Sabit $MinPts$ değeri için, daha yüksek yoğunluğa (daha düşük ε değerine) göre yoğunluk tabanlı kümeler, daha düşük yoğunluğa (daha yüksek ε değerine) göre yoğunlukla bağlantılı kümelere tamamen içerir.

DBSCAN algoritması, aynı anda birkaç mesafe parametresini işleyecek şekilde genişletilmiştir, bir başka ifadeyle farklı yoğunluklara göre yoğunluk tabanlı kümeler aynı anda oluşturulur. Ancak, tutarlı sonuç üretmek için kümenin genişletilmesi sırasında nesnelerin işleme sırasına uyulması gerekir. Her zaman, daha düşük ε değerine göre yoğunlukla erişilebilir olan nesne seçilir, böylece daha yüksek yoğunluğa (daha küçük ε değerlerine) göre kümelerin önceden tamamlanması garanti edilir.

OPTICS, ilkesel olarak sonsuz sayıda ε_i , “kullanıcının belirlediği” ε ($0 \leq \varepsilon_i \leq \varepsilon$) olan mesafe parametreleri için genişletilmiş DBSCAN algoritması gibi çalışır (Ankerst vd., 1999). Tek fark, küme üyeliklerinin atanmıyor olmasıdır. Bunun yerine, nesnelerin işleme sırasını ve genişletilmiş DBSCAN algoritmasının küme üyeliklerini atamak için kullanacağı bilgi tutulur (eğer bu sonsuz sayıda parametre için mümkün olsaydı). Bu bilgi, her nesne için yalnızca iki değerden oluşur: çekirdek mesafe (core-distance) ve ulaşılabilirlik mesafesi (reachability-distance).

p veritabanı D 'den bir nesne olsun, ε mesafe değeri olsun, $N_\varepsilon(p)$ p 'nin ε -komşuluğu olsun, $MinPts$ doğal sayı olsun ve $MinPts - uzaklık(p)$ p 'nin $MinPts$ komşusuna olan mesafe olsun. Dolayısıyla, p 'nin çekirdek mesafesi aşağıdaki gibi tanımlanır:

$$çekirdek\ mesafe_{\varepsilon, MinPts}(p) = \begin{cases} TANIMSIZ, & SAYI(N_\varepsilon(p)) < MinPts \\ MinPts - uzaklık(p), & diğer\ durum \end{cases} \quad (52)$$

Bir nesnenin p çekirdek mesafesi, ε 'ye göre p 'nin ε -komşuluğundaki nesne ile p arasındaki en küçük mesafe olarak tanımlanır. Bu komşu, $N_\varepsilon(p)$ içinde bulunuyorsa, p 'nin ε 'ye göre çekirdek nesne olacağı mesafedir. Aksi takdirde, çekirdek mesafe TANIMSIZ'dır.

p ve o , D veritabanından nesneler olsun, $N_\varepsilon(o)$ o 'nun ε -komşuluğu olsun ve $MinPts$ doğal sayı olsun. Bu durumda, p 'nin o 'ya göre ulaşılabilirlik mesafesi aşağıdaki gibi tanımlanı.

$$ulaşılabilirlik\ mesafesi_{\varepsilon, MinPts}(p, o) =$$

$$\begin{cases} \text{TANIMSIZ, } |N_\varepsilon(o)| < \text{MinPts} \\ \max(\text{çekirdek uzaklığı}(o), \text{uzaklık}(o, p)), \text{ diğ er durum} \end{cases} \quad (53)$$

Kavramsal olarak, nesne p 'nin diğ er nesne o 'ya göre ulaşılabilirlik mesafesi, p 'nin o 'dan doğrudan yoğunluk-erişilebilir (density-reachable) olduğ u en küçük mesafedir, eğer o bir çekirdek nesne ise. Bu durumda, ulaşılabilirlik mesafesi, o 'nun çekirdek mesafesinden daha küçük olamaz çünkü daha küçük mesafelerde hiçbir nesne doğrudan yoğunluk-erişilebilir değildir. Aksi takdirde, eğer o bir çekirdek nesnesi değilse, hatta üretilen mesafe ε 'de bile, p 'nin o 'ya göre ulaşılabilirlik mesafesi TANIMSIZ'dır. Bir nesnenin ulaşılabilirlik mesafesi, hesaplandığı çekirdek nesneye bağlıdır.

OPTICS, veritabanının kümelendirilmesini oluştururken, her nesne için çekirdek mesafesini ve uygun ulaşılabilirlik mesafesini de ek olarak saklar. Bu bilgi, kullanıcı tarafından belirlenen ε 'den daha küçük herhangi bir mesafe ε 'ye göre tüm yoğunluk tabanlı kümeleri çıkarmak için yeterlidir.

Sınıflandırma için kullanılacak nesnenin ε -komşulukları öncelikle alınır, ulaşılabilirlik mesafesini TANIMSIZ olarak ayarlanır ve çekirdek mesafesi belirlenir. Ardından, nesne Sıralı Liste'ye yazılır. Nesnenin çekirdek nesne özelliği kontrol edilir ve ε 'de çekirdek nesne değilse, bir sonraki nokta üzerinde işlemler devam eder. Aksi halde, nokta ε mesafesinde çekirdek nesnesi ise, ε ve MinPts 'ye göre doğrudan yoğunlukla erişilebilen nesneler iteratif olarak toplanır. Mevcut çekirdek nesnesinden doğrudan yoğunlukla erişilebilen nesneler, ileride genişleme için Tohum Listesine eklenir. Tohum Listesi içindeki nesneler, doğrudan yoğunlukla erişildikleri en yakın çekirdek nesneye olan ulaşılabilirlik mesafelerine göre sıralanır. Tohum-listesinde en küçük ulaşılabilirlik mesafesine sahip nesne seçilir. Bu nesnenin ε -komşulukları ve çekirdek mesafesi belirlenir. Daha sonra, nesne, çekirdek mesafesi ve mevcut ulaşılabilirlik mesafesi ile Sıralı Listeye basitçe eklenir. Eğer nesne çekirdek nesnesi ise, genişleme için daha fazla aday Tohum-listesine eklenebilir.

Tohum Listesine ekleme ve ulaşılabilirlik mesafelerinin merkezi nesneye göre belirlenir. Tohum Listesi öncelik kuyruğunda henüz bulunmayan nesneler, sadece ulaşılabilirlik mesafeleriyle birlikte basitçe eklenir. Kuyrukta zaten bulunan nesnelerin yeni ulaşılabilirlik mesafesi, önceki ulaşılabilirlik mesafesinden daha küçükse, kuyruğun üstüne taşınır.

Bir nesnenin ε -komşuluğunu almak için, merkezi o ve yarıçapı ε olan bölge sorgulanır. ε ve MinPts 'ye göre veritabanının artırılmış küme sıralamasını oluşturduktan

sonra, bu sıradan $MinPts$ ve $\varepsilon' \leq \varepsilon$ ye göre herhangi yoğunluk tabanlı kümeleme çıkarılır, basitçe küme sıralaması taranır ve nesnelerin ulaşılabilirlik mesafesine ve çekirdek mesafesine bağlı olarak küme üyelikleri atanır.

Öncelikle, mevcut nesnenin ulaşılabilirlik mesafesinin kümeleşme mesafesi ε' den büyük olup olmadığını kontrol edilir. Bu durumda, nesne, küme sıralamasındaki mevcut nesneden önceki herhangi nesneden ε' ve $MinPts'$ ye göre ε' ye göre yoğunlukla erişilemez. Bu açıktır, çünkü ε' ve $MinPts'$ ye göre önceki nesneden nesne, en fazla ε' ye kadar ulaşılabilirlik mesafesine sahip olurdu. Bu nedenle, ulaşılabilirlik mesafesi ε' den büyükse, Nesne'nin çekirdek mesafesine bakarız ve Nesne çekirdek nesnesi ise, yeni küme başlatılır; aksi halde, Nesne GÜRÜLTÜ'ye atanır (kümeleme sıralamasındaki ilk nesnenin ulaşılabilirlik mesafesi her zaman TANIMSIZ'dır ve TANIMSIZ'ın herhangi tanımlı mesafeden büyük olduğu varsayılmıştır). Eğer mevcut nesnenin ulaşılabilirlik mesafesi ε' den küçükse, bu nesne basitçe mevcut kümelere atanabilir çünkü o zaman bu nesne, kümeleme sıralamasındaki önceki çekirdek nesneden ε' ve $MinPts'$ ye göre yoğunlukla erişilebilir.

1.3.4. Hatalı Eşleşmelerin Ortadan Kaldırılması

Kümeleme yapıları kullanılarak özellikler eşleştirildikten sonra aykırı eşleşmelerin ortadan kaldırılması gerekmektedir. Bu bölümde tez kapsamında bu tür eşleşmeleri ortadan kaldıracak yöntemler tanıtılmıştır.

1.3.4.1. RANSAC Metodu

Fischler ve Bolles tarafından önerilen RANdom SAMple Consensus (RANSAC) algoritması (Fischler & Bolles, 1981), girdi verilerindeki büyük oranda aykırı değerlerle başa çıkmak için tasarlanmış genel parametre tahmini yaklaşımıdır.

RANSAC, temel model parametrelerini tahmin etmek için gereken minimum gözlem (veri noktaları) sayısını kullanarak aday çözümler üreten yeniden örnekleme yöntemidir. Geleneksel örnekleme tekniklerinden farklı olarak, başlangıçta tüm veriyi kullanarak başlangıç çözüm elde etmek ve ardından aykırı değerleri elemek yerine, RANSAC en küçük mümkün veri kümesini kullanır ve ardından bu kümesi tutarlı veri noktalarıyla genişletir.

RANSAC Algoritması şu şekilde çalışmaktadır

- Model parametrelerini belirlemek için gereken minimum nokta sayısı rastgele seçilir.
- Modelin parametreleri çözülür.
- Tüm noktalar kümesinden önceden belirlenmiş bir toleransa uygun olan noktaların sayısı belirlenir.
- Eğer nokta sayısının toplam nokta sayısına oranı önceden belirlenmiş eşik τ değerini aşıyorsa, tüm belirlenen noktaları kullanarak model parametreleri yeniden tahmin edilir ve işlem sonlandırılır.
- Aksi takdirde, adımları 1 ila 4 arasında (en fazla N kez) tekrarlayın.

RANSAC algoritması için deneme sayısı Eşitlik (54)'a göre belirlenir. Bu eşitliğe göre deneme sayısı N , ρ ve u değerlerinin bilinmesiyle hesaplanabilir. Burada, ρ sabit değer (doğru modelin seçilme olasılığı), u ise seçilen noktaların kurulan modele uygun olma olasılığıdır.

$$1 - \rho = (1 - u^m)^N \quad (54)$$

N iterasyon sayısıdır, genellikle en az bir rastgele örneklemin aykırı değer içermediği olasılığını (genellikle 0.99 olarak ayarlanır) sağlamak için yeterince yüksek seçilir. Herhangi seçilen veri noktasının içerik olasılığını temsil etmek üzere u , ve $v = 1 - u$ olarak aykırı değerlerin gözlemlenme olasılığını temsil eder, m belirtilen minimum nokta sayısını temsil eder. Eşitlik (54)'da her iki tarafın logaritması alındığında Eşitlik (55) elde edilmektedir.

$$N = \frac{\log(1-\rho)}{\log(1-(1-v)^m)} \quad (55)$$

1.3.4.2. Parametre Normalizasyon Metodu

Hatalı eşleşmelerin ortadan kaldırılması için kullanılan parametre normalizasyon işlemi şu şekilde çalışmaktadır. Veriseti (S) kümesi olarak belirlenen örneklerin her biri N boyutlu olarak ele alınır. $N = \{1, 2, 3, \dots\}$ olabilir.

Bir kümede s nesnesi, S kümesi içinde eleman olsun. Bu elemanları $s = [s_1, s_2, s_3, \dots, s_N]$ şeklinde temsil edilmektedir. Kullanıcı belirlediği parametreye göre, her N elemanı bölünür ve ondalık kısmı atılır. Bu işlem, kümedeki tüm elemanlar için

gerçekleştirilir ve bölüm sonuçları diziye kaydedilir. Dizinin histogramı alınarak, en yüksek frekansa sahip olan bölge belirlenir ve diğer küme elemanları silinir. Bu süreç, yoğunluğun olduğu bölgeyi korurken seyrek elemanları elemiş olur.

1.3.5. Perde İzleme Yöntemi

Perde izleme yöntemi, konuşma kaydında sesli bölgeleri belirlemek için kullanılan etkili tekniktir. Bu yöntemde, sesli kısımlar, sıfırdan farklı frekans değerlerine sahip perde dizileri ile temsil edilirken, sessiz kısımlar sıfır ile ifade edilmektedir. Konuşma kaydındaki perde dizilerinin elde edilmesinde, Yet Another Algorithm for Pitch Tracking (YAAPT) metodu kullanılmaktadır (Kasi & Zahorian, 2002). YAAPT, ses sinyallerinin analizinde güvenilir ve doğru sonuçlar veren algoritma olarak bilinmektedir. Bu yöntem sayesinde, konuşma kayıtlarındaki sesli ve sessiz bölgeler etkin şekilde ayrıştırılarak analiz edilebilmektedir. Bu yöntem 4 adımdan oluşmaktadır.

1.3.5.1. Konuşma Sinyali Ön İşleme

Konuşma sinyalinin işlenmesinde iki tür doğrusal olmayan işlem kullanılmaktadır: mutlak değer ve kare değer. Öncelikle, sinyalin mutlak değeri ve kare değeri hesaplanır. Ardından, bu doğrusal olmayan işlemlerden geçmiş sinyaller, perde çıkarımı amacıyla kullanılır.

1.3.5.2. Frekans Alanında Perde Dizisi Tahmini

Bu adımda, spektral analiz temelinde perde adayları belirlenmektedir. İlk olarak, doğrusal olmayan konuşma sinyalinin spektrumundan perde adaylarını çıkarmak amacıyla Spektral Harmonik Korelasyon (SHC) yöntemini kullanılmaktadır. SHC

$$SHC(t, f) = \sum_{f'=-L/2}^{L/2} \prod_{r=1}^{N+1} S(t, rf + f') \quad (56)$$

şeklinde hesaplanmaktadır. $S(t, f)$ t anındaki çerçevenin f frekansındaki genlik spektrumunu ifade eder. L , frekanstaki spektral pencere uzunluğudur. N , harmonik sayısını ve f frekansı temsil eder. $SHC(t, f)$ Her çerçeve için 0 ile 1.0 arasında normalize edilmiştir.

$SHC(t, f)$, perdenin tam katlarında yüksek genlik gösterecek olup, bu da konuşma sinyalinin frekans alanında perde adaylarını bulmada yardımcı olmaktadır.

Normalleştirilmiş düşük frekans enerjisi oranı (NLFER), perde izleme yönteminde sesli/sessiz kararları vermede yardımcı olarak kullanılır. Her çerçeve için düşük frekans bölgelerindeki ortalama enerji hesaplanır ve bu, sesli çerçeveler ile sessiz çerçeveler arasında ayırım konusunda yardımcı olur. NLFER

$$NLFER(t) = \frac{\sum_{f=2 \times F_0min}^{F_0max} S(t, f)}{\frac{1}{N} \sum_{t=1}^N \sum_{f=2 \times F_0min}^{F_0max} S(t, f)} \quad (57)$$

şeklinde hesaplanmaktadır. N , çerçeve sayısını ifade eder. F_0min ve F_0max , perde aralığını tanımlar. $S(t, f)$, t anındaki çerçevenin ve f frekansındaki konuşma sinyalinin spektrumunu temsil eder. Sesli çerçeveler yüksek NLFER değerine sahipken, sessiz çerçeveler düşük NLFER değerine sahiptir. Dolayısıyla, NLFER, sesli ve sessiz çerçeveleri ayırt etmek için kullanılır.

1.3.5.3. Zaman Alanında Perde Dizisi Tahmini

Perde dizileri hem orijinal sinyallerden hem de doğrusal olmayan sinyallerden zaman alanında çıkarılmaktadır. Kısa süreli otokorelasyon fonksiyonu, zaman alanında geleneksel perde izleme yöntemidir. Normalize edilmiş çapraz korelasyon fonksiyonu (NCCF) olarak adlandırılan otokorelasyon fonksiyonunun değiştirilmiş hali kullanılmaktadır. NCCF şöyle tanımlanır:

$$NCCF(k) = \frac{1}{\sqrt{e_0 e_k}} \sum_{n=0}^{L-K_{max}} s^2(n) \quad (58)$$

$$e_0 = \sum_{n=0}^{L-K_{max}} s^2(n) \quad (59)$$

$$e_k = \sum_{n=k}^{L-K_{max}} s^2(n) \quad (60)$$

$$K_{min} \leq k \leq K_{max} \quad (61)$$

L , çerçeve uzunluğunu ifade eder. K_{min} ve K_{max} , perde arama aralığını kapsamak için kullanılan gecikme değerleridir. NCCF, kendiliğinden $[-1, 1]$ aralığında normalize edilir ve

NCCF'nin sonucu, perdenin tam katlarına eşit gecikme değerlerinde 1 olacaktır. NCCF geleneksel otokorelasyon fonksiyonundan daha iyidir.

1.3.5.4. Dinamik Programlama ile Perde Dizisi Belirleme

Frekans alanında ve zaman alanında, konuşma ifadesi boyunca perde adaylarının ve ilgili değer matrisinin bir matrisini elde edilir. En iyi perde adaylarını seçmek için geçiş maliyet fonksiyonu ve yerel maliyet fonksiyonu tasarlanır. Son olarak, bu iki tür maliyeti hesaplayarak ve en aza indirerek en iyi perde adayları belirlenir

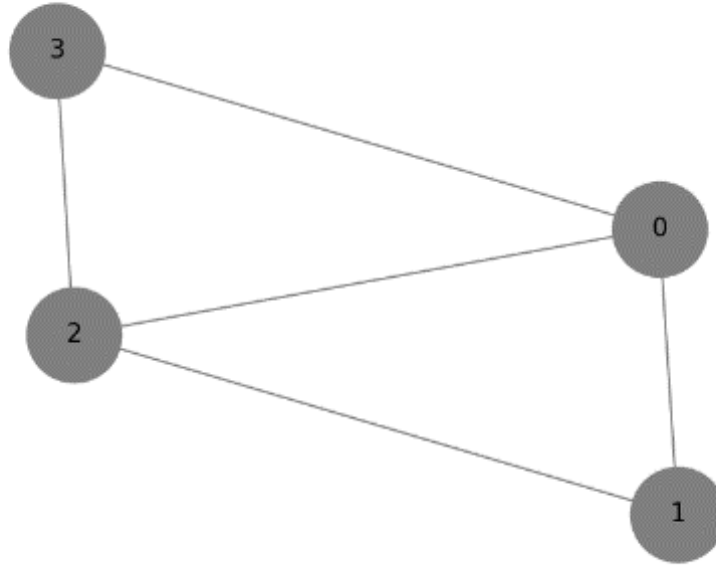
1.3.6. Graf Teorisi

Graf teorisi, matematiğin dalıdır ve grafları veya çizgeleri inceler. Graf, düğümler veya köşeler adı verilen noktaların ve bu noktaları birleştiren kenarların veya bağlantıların bir araya gelmesiyle oluşur. Bu yapılar, bir tür ağ yapısı olarak kabul edilir. Graf teorisi, 1736'da Leonhard Euler tarafından geliştirilen matematiksel teori olarak köken alır (Biggs vd., 1986). Graf teorisinin çalışmaları, Petri ağları gibi birçok yeni kavramın ortaya çıkmasına katkıda bulunmuştur.

1736 yılında, Leonhard Euler tarafından yazılan makale, günümüzde hala popülerliğini koruyan "Königsberg'in Yedi Köprüsü" problemiyle ilişkilendirilir ve bu, graf teorisinin kesin başlangıç tarihi olarak kabul edilir.

Graf teorisinde, G grafi genellikle iki küme ile temsil edilir: $G = (D, K)$. Burada, D düğümlerin kümesini ve K ise düğümler arasındaki ilişkileri temsil eden kenarların kümesini ifade eder.

Yönlü graf teorisinde, kenarlara giriş ve çıkış yönleri belirlenmişse, bu kenarlara "yönlü kenarlar" denir. Bir düğümden çıkan ve tekrar aynı düğüme giren kenar, genellikle "döngü" olarak adlandırılır. Eğer aynı yöne veya yönsüz olarak bir düğümden diğerine giden iki kenar varsa, bu kenarlara "paralel kenarlar" denir.



Şekil 3. Graf Örneği

Şekil 3’te verilen yönsüz graf için küme gösterimi aşağıdaki şemada verilmektedir.

$$D = \{0, 1, 2, 3\}$$

$$K = \{(0, 1), (0, 2), (0, 3), (1, 2), (2, 3)\}$$

$$G = \{D, K\}$$

Graf renklendirme, grafın düğümlerinin veya kenarlarının farklı renklere atanması işlemidir. Temel amaç, düğüm veya kenarın komşularıyla aynı rengi paylaşmamasını sağlayarak belirli kısıtlar altında grafi renklendirmektir. Graf renklendirmesi, genellikle belirli kısıtlar altında minimum renk sayısını bulma problemi olarak tanımlanır. Bu kısıtlar arasında komşu düğümlerin aynı renge boyanmaması, belirli düğüm kümesinin aynı renge boyanması gerekliliği veya diğer belirli koşullar yer alabilir. Graf renklendirme, bilgisayar bilimleri, iletişim ağları, zamanlama problemleri gibi birçok alanda uygulama bulur.

Welsh-Powell graf renklendirme algoritması (Olariu vd., 1989), grafi mümkün olan en az sayıda renkle boyamak için kullanılan algoritmadır. Bu algoritma, grafın düğümlerini renkle boyamak için en az renk sayısını hesaplamak üzere tasarlanmıştır.

Welsh-Powell algoritması adımları şu şekildedir:

- Grafın düğümleri, derecelerine göre azalan sırayla sıralanır.
- Düğümler sırayla ele alınır ve her düğüme bir renk atanır.
- Bir düğümün komşularıyla aynı rengi paylaşmaması sağlanır.

- Eğer düğüm daha önce boyanmış ise, komşularıyla aynı renge boyanmayacak şekilde farklı renk atanır.

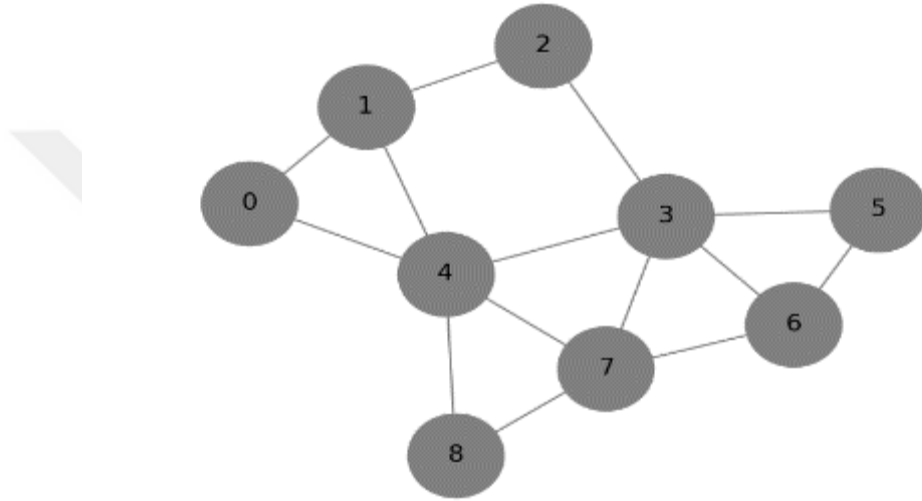
Şekil 4'te verilen grafın küme gösterimini çıkaralım:

$$D = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$$

$$K = \{(0, 1), (0, 4), (1, 2), (1, 4), (2, 3), (3, 4), (3, 5), (3, 6), (3, 7), (4, 7), (4, 8), (5, 6), (6, 7), (7, 8)\}$$

$$G = \{D, K\}$$

Grafın Welsh-Powell ile renklendirilmesi aşağıda adım adım incelenmektedir.



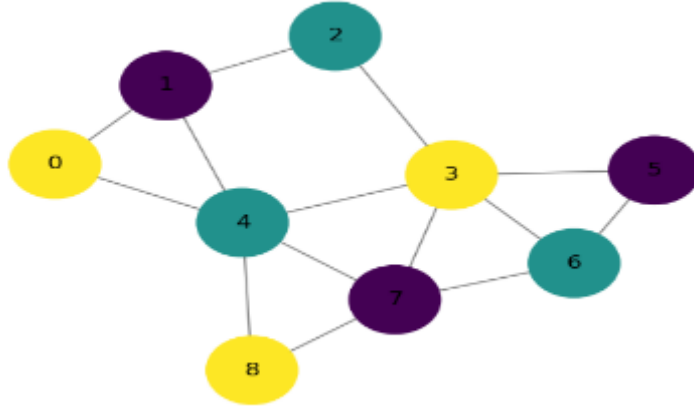
Şekil 4. Renklendirilmemiş Graf Örneği

- Düğümlerin dereceleri belirlenip, bunlar derece sıralamasına göre düzenlenir.

Düğüm	3	4	7	1	6	0	2	5	8
Derece	5	5	4	3	3	2	2	2	2

- En yüksek dereceli düğüme bir renk atanır ve ardından bu düğüme komşu olmayan tüm düğümlere aynı renk atanır. Örnekte 3 numaralı düğümden renklendirme işlemi başlatılır.
- Kalan boyanmamış düğümler arasında en yüksek dereceli olan seçilerek bu düğüme farklı renk atanır. Sonrasında bu düğüme komşu olmayan tüm düğümlere de aynı renk verilir.
- Tüm düğümler renklendirilene kadar bu işlem tekrarlanır.

Grafın renklendirilmiş hali Şekil 5'de verilmiştir.



Şekil 5. Renklendirilmiş Graf Örneği

Renklendirilmiş bir graf görüntüsü, düğümlerin komşuluk ilişkisi hakkında fikir vermektedir. Renklendirme algoritmasının çalışma prensibi gereği, komşu olan düğümler farklı renklere boyanmaktadır. Bu nedenle, yoğun düğümlü graflarda komşuluk ilişkisi görüntüye bakılarak yorumlanabilmektedir.

2. YAPILAN ÇALIŞMALAR

2.1. Giriş

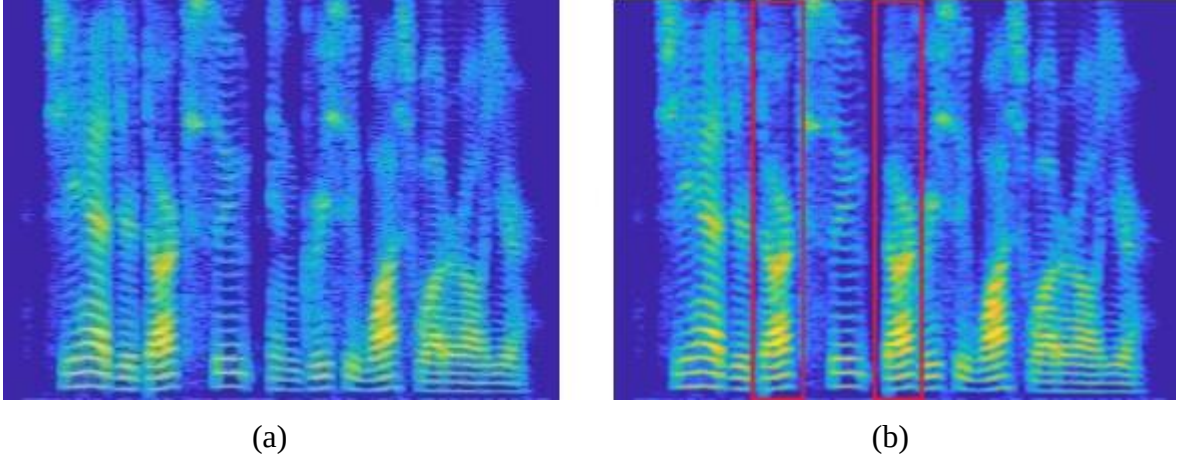
Bu tez çalışması iki farklı ses kopyalama ve yapıştırma sahteciliğini tespit etme yöntemi önermektedir. İlk yöntem, anahtar noktaların kümelendirilmesine dayanmaktadır. Konuşma kayıtlarından yüksek çözünürlüklü spektrogram görüntüleri elde edilmiştir. Bu görüntüler üzerinden BRIEF, AKAZE, SIFT ve SURF gibi yöntemler kullanılarak anahtar noktalar ve onların tanımlayıcıları çıkarılmıştır. Kümeleme işlemi ise OPTICS, g2NN ve k-Means algoritmaları kullanılarak gerçekleştirilmiştir.

Yapılan ikinci çalışma, graf görüntülerinin sınıflandırılmasına dayanmaktadır. Başlangıçta, giriş sesinden yüksek çözünürlüklü spektrogram görüntüleri elde edilir. Görüntü üzerinde SIFT anahtar noktası çıkarım yöntemi kullanılarak anahtar noktalar ve bunların tanımlayıcıları belirlenir. Kümeleme işlemi g2NN algoritması kullanılarak gerçekleştirilir. Ardından, görüntü üzerinde yoğun eşleşmenin olduğu frekans bölgesi tespit edilir ve bu bölgeye karşılık gelen frekans bilgileri kullanılarak sese band geçiren süzgeç

işlemi uygulanır. Ses üzerine Hızlı Fourier Dönüşümü işlemi uygulanarak frekans uzayına geçilir. Frekans bilgisinden graf görüntüleri elde edilir ve bu görüntüler kullanılarak sınıflandırma işlemi gerçekleştirilir.

2.2. Anahtar Noktalarının Kümelendirilmesi Yaklaşımıyla Kopyala Yapıştır Ses Sahteciliği Tespiti

Bu bölümde Yüksek Çözünürlüklü Spektrogram görüntülerinden elde edilen anahtar noktalarının kümelendirilmesi yaklaşımı kopyala yapıştır ses sahteciliği tespiti yöntemleri anlatılacaktır. Orijinal ve sahte konuşma kayıtlarının Yüksek Çözünürlüklü spektrogram görüntüleri Şekil 6’da verilmiştir. Kırmızı kutucuklar kopyala yapıştır segmentleri işaret etmektedir.



Şekil 6. Yüksek çözünürlüklü spektrogram görüntüsü (a) Orijinal ses (b) Sahte ses

Yüksek Çözünürlüklü Spektrogram görüntüleri zaman ve frekans açısından yüksek çözünürlük sağladığı için anahtar nokta çıkarım aşamasında oldukça başarılı sonuçlar sağlamaktadır.

2.2.1. OPTICS Kümeleme Yaklaşımı Kullanımı

BRIEF, AKAZE, SIFT ve SURF yapıları kullanılarak belirlenen anahtar noktalarının OPTICS algoritması ile kümelendirilmesi bu bölümde anlatılacaktır.

OPTICS (Ordering Points To Identify the Clustering Structure) kümeleme yaklaşımı, yoğunluk-temelli kümeleme yöntemlerinde önemli avantajlar sunar. OPTICS, veri setindeki farklı yoğunluktaki bölgeleri keşfetme yeteneği sayesinde, yoğunlukları farklı olan kümeleri

doğru şekilde tanımlayabilir ve sürekli yoğunluk düzeni oluşturur. Bu yöntem, kullanıcıların farklı yoğunluk seviyelerini keşfetmelerine olanak tanıyarak küme sayısını önceden belirleme gereksinimini ortadan kaldırır. Ayrıca, gürültü ve aykırı değerlere karşı dayanıklıdır; veri setindeki bu tür değerleri doğal olarak algılayıp kümelerden ayırabilir, böylece daha doğru ve güvenilir sonuçlar elde edilmesini sağlar. OPTICS, yalnızca küresel veya doğrusal kümeleri değil, aynı zamanda keyfi şekilli ve karmaşık kümeleri de başarıyla tanımlar. Her veri noktasının çekirdek ve erişim mesafeleri gibi değerleri hesaplayarak veri setinin yoğunluk yapısı hakkında detaylı bilgi sağlar, bu da veri setinin daha iyi anlaşılmasına yardımcı olur. Tüm bu özellikler, OPTICS'i özellikle yoğunluk-temelli ve karmaşık kümeleme problemlerinde güçlü ve esnek yöntem haline getirir.

2.2.1.1. BRIEF Anahtar Noktası Metodu

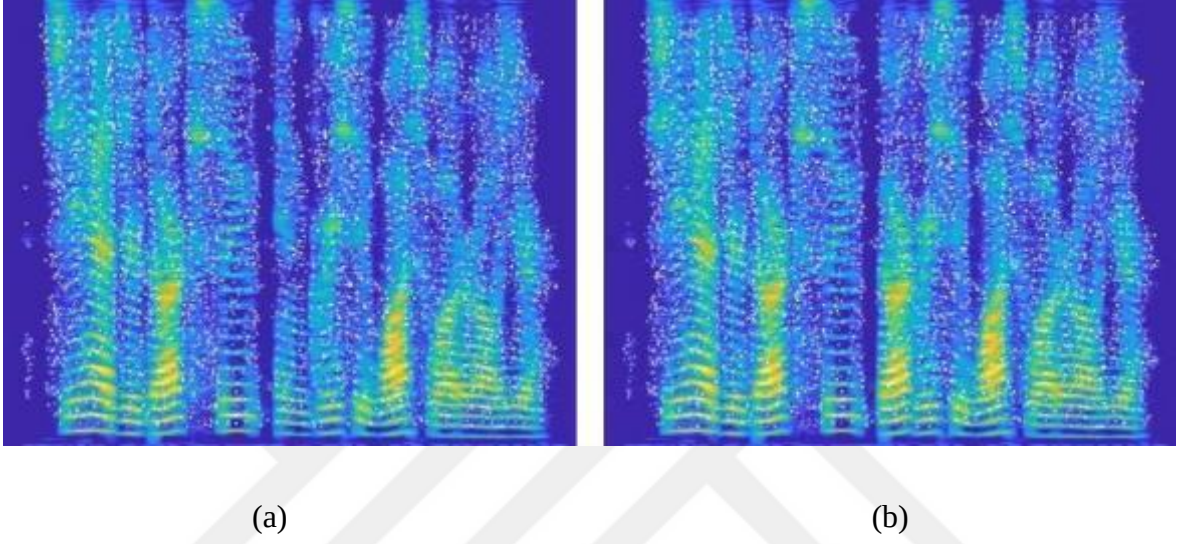
Kopyala yapıştır ses sahteciliği tespit etme yöntemleri arasında önerdiğimiz BRIEF anahtar noktalarının OPTICS ile kümelendirilmesi yaklaşımı giriş sesinden zaman ve frekans açısından Yüksek Çözünürlüklü Spektrogram görüntülerinin çıkarılması ile başlamaktadır.

Konuşma kayıtlarından Frekans bilgisi temsili için Yüksek Çözünürlüklü Spektrogram görüntüleri çıkarıldıktan sonra anahtar noktaları Merkez Çevre Ekstrema (Center Surround Extrema, CenSurE) yöntemiyle ve özellik tanımlayıcılarını ise İkili Sağlam Bağımsız Temel Özellikler (Binary Robust Independent Elementary Features, BRIEF) ile elde edilmektedir.

BRIEF (Binary Robust Independent Elementary Features) algoritması, özellikle hız ve verimlilik açısından diğer özellik çıkarma yöntemlerine göre belirgin avantajlar sunar. BRIEF, çok hızlı şekilde çalışır çünkü karmaşık hesaplamalar yerine basit ikili karşılaştırmalar kullanır, bu da gerçek zamanlı uygulamalar için ideal hale getirir. Hesaplama maliyetinin düşük olması, BRIEF'i düşük güçlü cihazlarda veya büyük veri setlerinde avantajlı kılar. Ayrıca, BRIEF'in ürettiği ikili vektörler, diğer yöntemlere göre daha az bellek kullanır ve Hamming mesafesi kullanılarak hızlı şekilde karşılaştırılabilir, bu da eşleştirme sürecini hızlandırır. Algoritmanın basitliği, uygulanmasını kolaylaştırır ve geliştirme sürecinde hata yapma olasılığını azaltır. BRIEF, farklı görüntü işleme problemlerinde etkili çözümler sunarak geniş uygulama yelpazesinde kullanılabilir. Ancak, dönüşüme ve ölçek değişikliklerine karşı dayanıklı olmaması, bazı durumlarda diğer

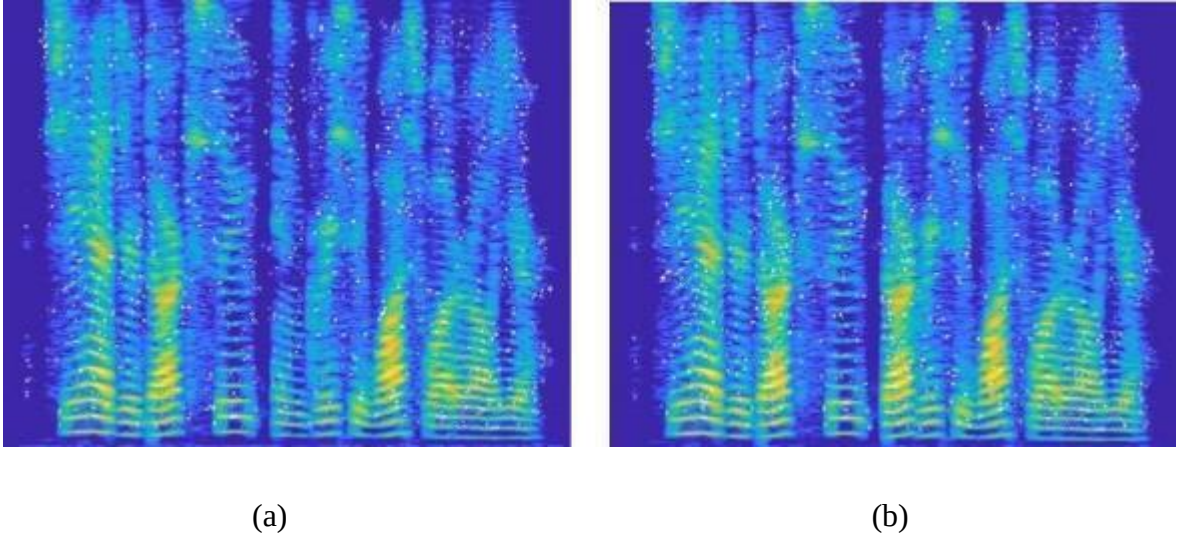
yöntemlerle birlikte kullanılmasını gerektirebilir. Yine de, BRIEF'in sağladığı hız ve verimlilik, onu birçok uygulamada tercih edilen yöntem yapmaktadır.

Şekil 7’de sırasıyla orijinal ve sahte ses spektrogramı üzerinde BRIEF tarafından belirlenen anahtar noktalar beyaz dairelerle işaretlenmektedir. Yöntemin başarısını etkileyen en önemli faktörler arasında görüntü üzerinde anahtar nokta sayısının yeterli miktarda olması gösterilmektedir.



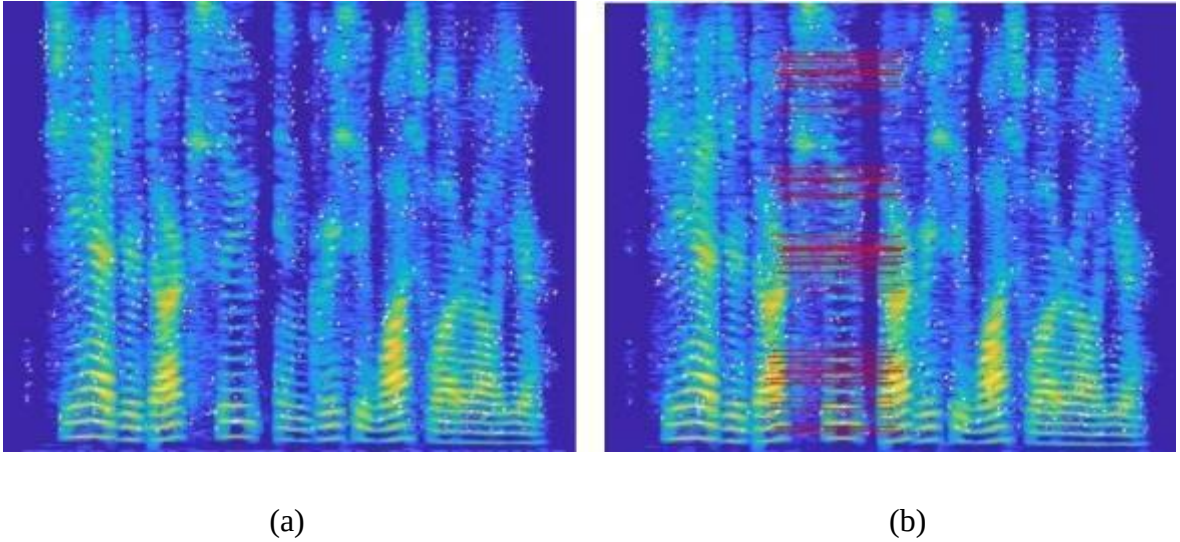
Şekil 7. Merkez Çevre Ekstrema tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

Belirlenen anahtar noktalar tanımlayıcı vektörler kullanılarak OPTICS algoritması tarafından kümelendirilmiştir. Herhangi sınıfa ait olmayan anahtar noktası ortadan kaldırılmıştır. Bu işlem sonucunda geriye kalan anahtar noktalar sırasıyla Şekil 8’de orijinal ve sahte ses spektrogramı üzerinde gösterilmektedir.



Şekil 8. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses

Kümelendirme işleminden sonra aynı kümeye sahip olanları kırmızı çizgilerle birleştirilmektedir. Bu işlem sahtecilik bölgesi hakkında ipucu vermektedir. Şekil 9 orijinal ve sahte spektrogramlar üzerindeki eşleşmeleri göstermektedir.

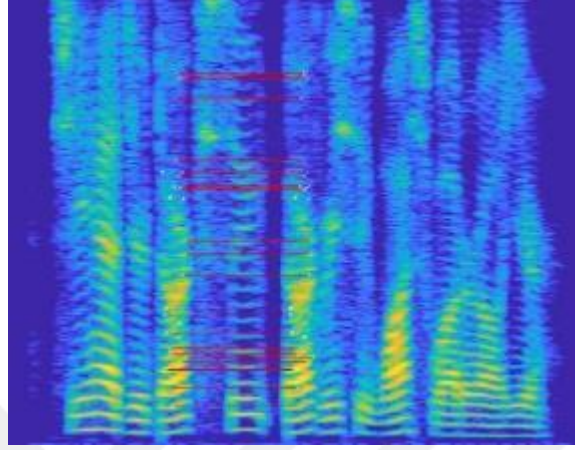


Şekil 9. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses

Orijinal sese ait Şekil 9a'da verilen spektrogramda herhangi eşleşme bulunamamıştır.

Şekil 9b'de, sahte sese ait yüksek çözünürlüklü spektrogram üzerindeki eşleşen anahtar noktalar kırmızı çizgilerle gösterilmektedir. Bu noktaların belirlenmesinin ardından, olası yanlış eşleşmeleri ortadan kaldırmak için RANSAC algoritması uygulanmaktadır. RANSAC, veri kümesi içinde aykırı verileri elemek amacıyla kullanılan yöntemdir. Sahte

sese ait yüksek çözünürlüklü spektrogramda, eşleşen noktalar RANSAC algoritması ile aykırı verilerden arındırıldıktan sonra, geriye kalan doğru eşleşmeler Şekil 10'da gösterilmektedir. Bu işlem sonucunda doğru eşleşmelerin sayısı azalsa da, sahtecilik bölgesi daha doğru biçimde temsil edilmektedir.



Şekil 10. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması

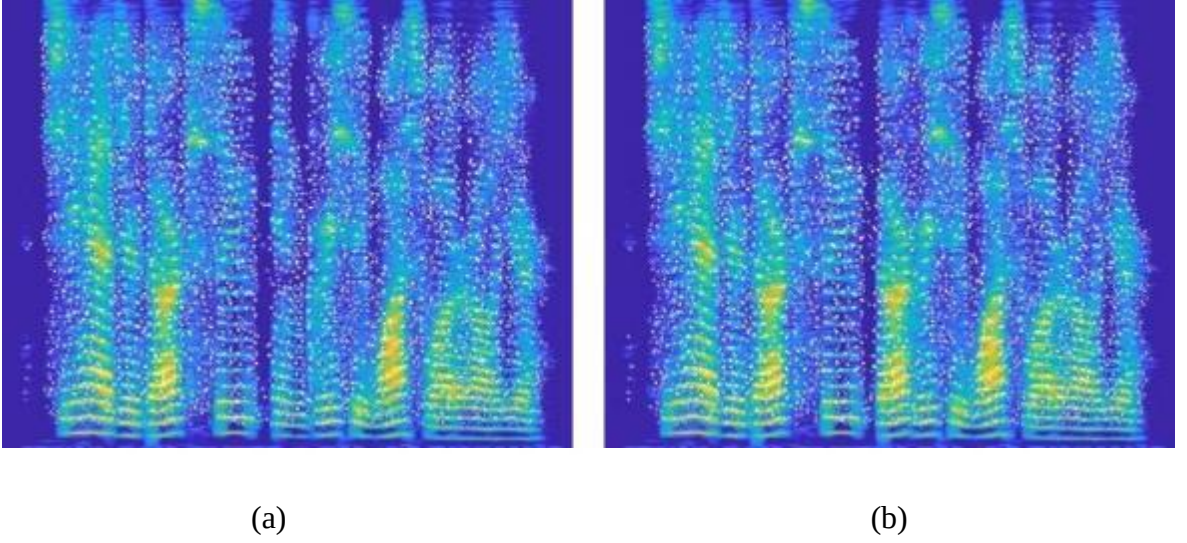
Spektrogram görüntülerinde yatay bileşenler zamanı temsil ederken, dikey bileşenler frekansı temsil etmektedir. Bu nedenle Şekil 10'da gösterilen kırmızı çizgilerin başlangıç ve bitiş noktalarının yatay eksene izdüşümü sahtecilik bölgesinin zaman bilgisini vermektedir.

2.2.1.2. AKAZE Anahtar Noktası Metodu

Kopyala-yapıştır sahteciliği tespit sistemi olarak önerdiğimiz farklı yöntem AKAZE ile anahtar noktaların çıkarılması ve OPTICS ile kümelendirilmesidir. Bu işlem için giriş görüntüsü Yüksek Çözünürlüklü Spektrogram tercih edilmiştir.

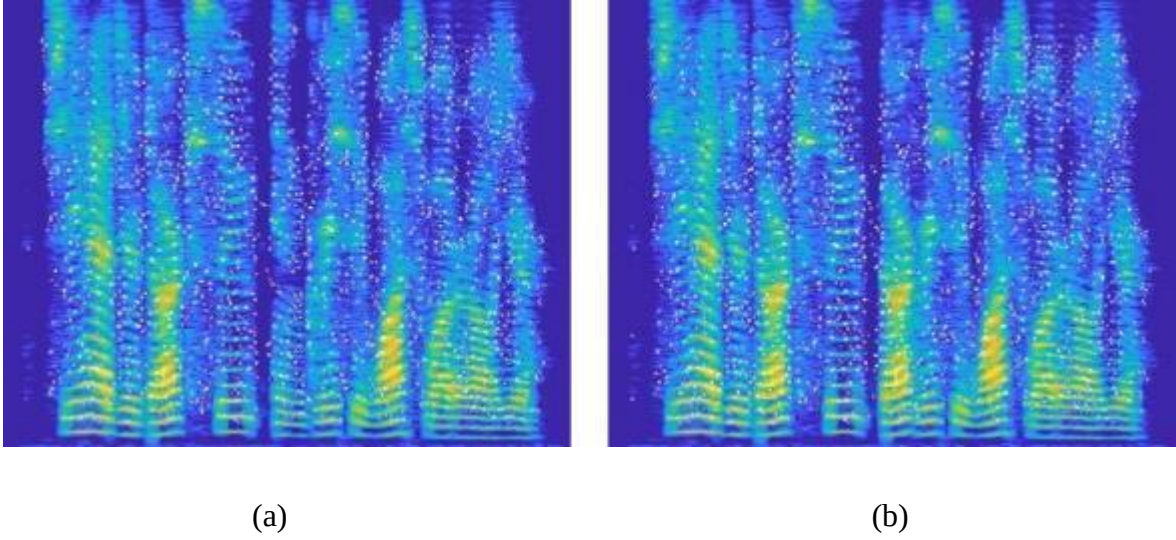
AKAZE (Accelerated-KAZE) anahtar nokta çıkarım yöntemi, görüntü analizi ve işleme alanında çeşitli avantajlar sunar. Öncelikle, bu yöntem hız ve verimlilik açısından oldukça üstündür; KAZE algoritmasının yavaşlığını gidererek daha hızlı performans sergiler. AKAZE, non-linear scale space (doğrusal olmayan ölçek uzayı) kullanarak, düşük kontrastlı ve bulanık alanlarda bile hassas anahtar noktalar tespit edebilir. Bu sayede, daha doğru ve güvenilir eşleştirmeler sağlar. Ayrıca, AKAZE'nin hesaplama maliyetleri düşüktür, bu da gerçek zamanlı uygulamalarda kullanımı mümkün kılar. Hem hız hem de doğruluk açısından sunduğu bu denge, AKAZE'yi çeşitli bilgisayarla görme ve görüntü işleme projelerinde tercih edilen yöntem haline getirmektedir.

Şekil 11.a ve 11b, sırasıyla orijinal ve sahte konuşma kayıtlarına ait yüksek çözünürlüklü spektrogramlar üzerinde AKAZE kullanılarak elde edilmiş anahtar noktaları göstermektedir. Bu anahtar noktalar ve onların tanımlayıcı vektörleri, spektrogramdan elde edilen özellikler olarak nitelendirildikleri için görüntü üzerinde çok sayıda bulunmalıdır. AKAZE algoritmasının etkinliği sayesinde, bu anahtar noktalar hem yüksek hassasiyetle tespit edilmekte hem de güvenilir tanımlayıcı vektörler oluşturularak, orijinal ve sahte kayıtlar arasındaki farkların belirlenmesine katkıda bulunmaktadır. Bu nedenle, görüntü üzerinde çok sayıda anahtar noktanın varlığı, spektrogramların ayrıntılı şekilde analiz edilmesini ve karşılaştırılmasını sağlamaktadır.



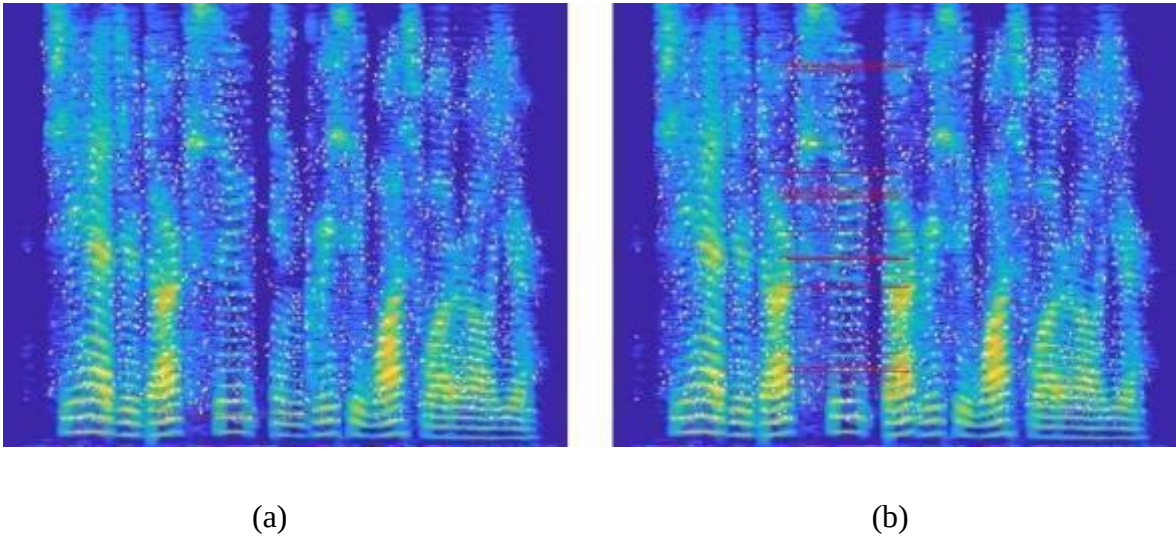
Şekil 11. AKAZE tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

AKAZE yapısı aracılığıyla çıkarılan anahtar noktaları OPTICS algoritması kullanılarak kümelendirilmektedir. Bu işlem ortak özellik barındıran anahtar noktaları aynı sınıfa dahil etmektedir. Şekil 12a ve 12b sırasıyla orijinal ve sahte spektrogramlardan AKAZE ile elde edilmiş anahtar noktaların OPTICS kullanarak kümelendirilmesi sonucu geriye kalan anahtar noktaları göstermektedir. Sınıflandırma işlemi anahtar noktaları elimine etmektedir.



Şekil 12. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses

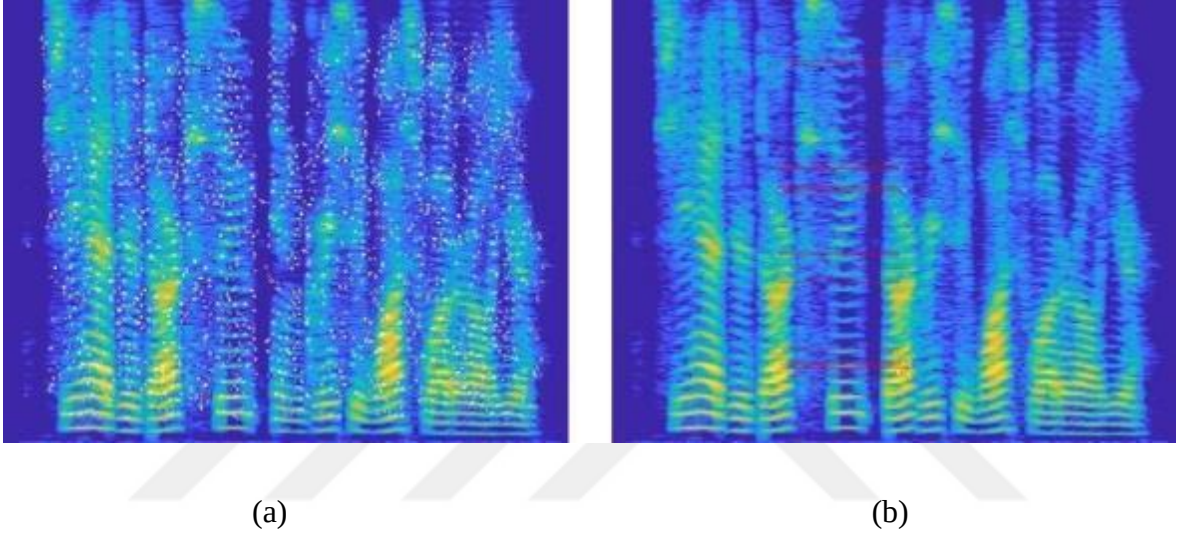
Aynı kümeye ait olan anahtar noktalar birleştirilerek birbirine bağlanmış olur. Bu bağlantının başladığı ve bittiği noktalar kopyala yapıştır bölgelerinin lokalizasyon bilgisini vermektedir. Şekil 13a ve 13b sırasıyla anahtar noktaların eşleşmelerini göstermektedir. Şekil 13a orijinal sese aittir, görüntüde görüldüğü üzere bu spektrogramda eşleşme bulunmaktadır. Şekil 13b sahte konuşma kaydına aittir ve eşleşmeler spektrogram üzerinde gösterilmektedir. AKAZE'nin anahtar noktalardan çıkardığı özellik tanımlayıcı vektörler kümelendirme aşamasında kullanılmıştır.



Şekil 13. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses

Eşleşme işlemleri sonucunda belirli bölgede işaretlemenin yoğun olması beklenmektedir. Ancak, bu durum gerçekleşmezse mevcut aykırı eşleşmeler göz ardı

edilerek ortadan kaldırılmaktadır. Bu süreç için RANSAC algoritması tercih edilmektedir. Şekil 14a'da, orijinal spektrograma ait olası hatalı eşleşmelerin ortadan kaldırılmış hali gösterilmektedir. Orijinal spektrogramda eşleşme olsa da yoğun şekilde olmadığı için mevcut kümelemeler göz ardı edilmekte ve ortadan kaldırılmaktadır. Şekil 14b'de, sahte spektrograma ait eşleşmeleri göstermektedir. Görüldüğü üzere, sınıflandırma sonucu oluşan doğrularda elimine işleminden sonra azalma görülmektedir. Bu yaklaşım, doğru eşleşmelerin korunmasını ve hatalı eşleşmelerin etkin şekilde elimine edilmesini sağlamaktadır.



Şekil 14. RANSAC tarafından hatalı eşleşmelerin kaldırılması (a) Orijinal ses (b) Sahte ses

Eşleşme bölgelerinin işaretlenmesi, sahtecilik lokalizasyonu hakkında bilgi verdiği için orijinal seslerde bu durumun olmaması beklenmektedir. Çünkü kelimenin kopyalanıp başka bölgeye taşınması, frekans bilgisi üzerinde değişiklik yapmamaktadır. Orijinal konuşma kaydına ait spektrogram üzerinde, zaman eksenini boyunca her bölge birbirinden farklı nitelikte piksellerle temsil edilmektedir. Bu durum, orijinal kayıtların sahtecilik içermediğini ve spektrogram üzerinde doğal dağılım gösterdiğini ortaya koymaktadır. Dolayısıyla, eşleşme bölgelerinin varlığı, sahtecilik tespiti için önemli gösterge olup, orijinal kayıtlarda bu tür eşleşmelerin bulunmaması beklenir.

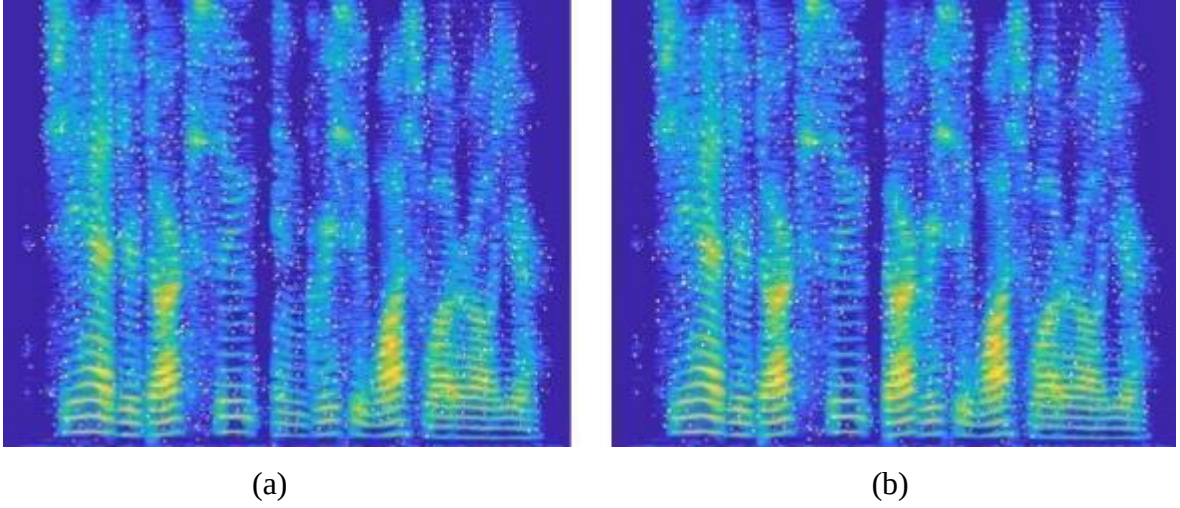
2.2.1.3. SIFT Anahtar Noktası Metodu

Bu yöntemde, SIFT anahtar noktaları ve tanımlayıcılarının OPTICS algoritması ile birlikte kullanılmasıyla yeni kopyala-yapıştır ses sahteciliği tespiti yöntemi önerilmiştir.

Öncelikle, sahtecilik şüphesi taşıyan konuşma kaydından yüksek çözünürlüklü spektrogram görüntüsü elde edilmektedir. Ardından, bu spektrogram görüntülerinden SIFT (Scale-Invariant Feature Transform) anahtar noktaları ve bu noktaların tanımlayıcı vektörleri çıkarılmaktadır. Elde edilen SIFT tanımlayıcıları, OPTICS (Ordering Points To Identify the Clustering Structure) algoritması ile analiz edilerek sahtecilik tespiti yapılmaktadır. Bu yaklaşım, ses dosyalarındaki kopyala-yapıştır sahtecilik girişimlerini yüksek doğrulukla belirlemeyi amaçlamaktadır.

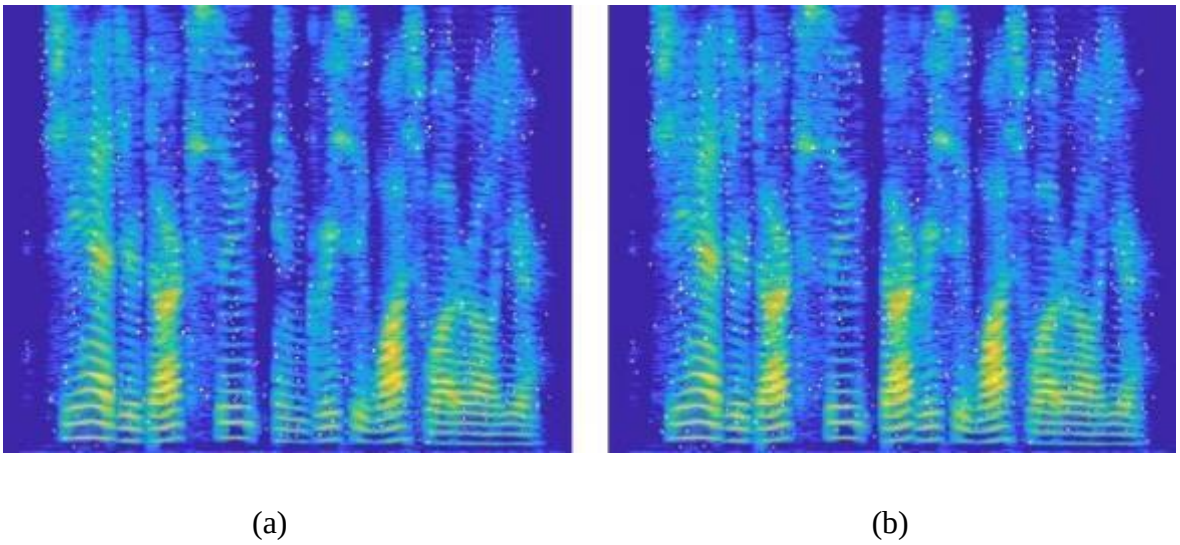
SIFT (Scale-Invariant Feature Transform) algoritmasının birçok avantajı bulunmaktadır. Öncelikle, ölçek ve döndürme değişmezliği sayesinde, farklı boyut ve açılarda çekilmiş görüntülerde dahi güvenilir anahtar noktalar tespit edilebilir. Bu özellik, SIFT'i çeşitli perspektif ve mesafelerden alınan görüntülerde etkin kılar. Ayrıca, SIFT, parlaklık ve kontrast gibi aydınlatma koşullarındaki değişimlere karşı dayanıklıdır, bu da farklı ışık koşullarında çalışmayı mümkün kılar. Algoritmanın tespit ettiği anahtar noktalar, yoğunlukları ve dağılımları bakımından yüksek ayrıştırma gücüne sahiptir, bu da nesnelerin hassas ve doğru şekilde tanımlanmasını sağlar. Son olarak, SIFT algoritması, geniş veri setlerinde dahi yüksek doğruluk ve hızla çalışabilen etkin yöntemdir, bu da onu bilgisayarla görme ve nesne tanıma uygulamalarında yaygın olarak kullanılabilir kılar.

Şekil 15a ve 15b sırasıyla orijinal ve sahte seslere ait spektrogram görüntülerinden çıkarılmış SIFT anahtar noktalarını göstermektedir. Görüntüde fazla anahtar nokta çıkarılması, görüntünün daha ayrıntılı ve kapsamlı şekilde temsil edilmesini sağlar. Bu durum, özellikle karmaşık veya ayrıntı açısından zengin bölgelerde, önemli özelliklerin kaçırılmaması açısından büyük önem taşımaktadır. Ayrıca, fazla sayıda anahtar nokta, farklı bakış açılarından veya perspektiflerden çekilmiş görüntülerin eşleştirilmesi ve hizalanmasında daha yüksek doğruluk sağlar. Bunun yanı sıra, fazla anahtar nokta çıkarılması, gürültüye ve bozulmalara karşı daha dayanıklı eşleştirme sağlar. Görüntüdeki bazı anahtar noktalar gürültü veya diğer bozulmalar nedeniyle kullanılamaz hale gelse bile, yeterli sayıda anahtar noktanın bulunması, doğru eşleştirme ve tanımanın devam etmesine olanak tanır.



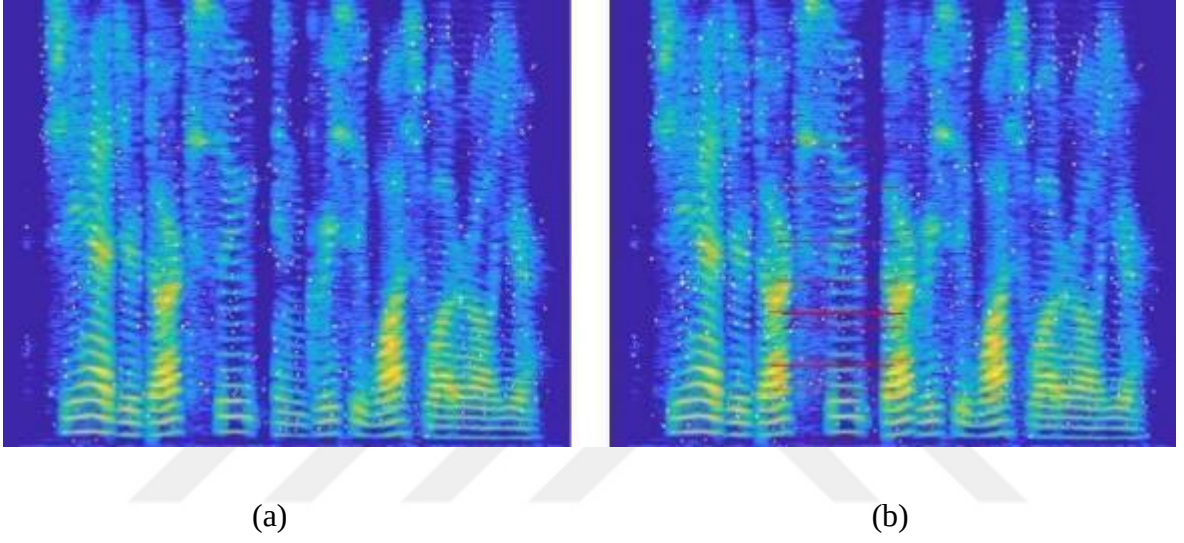
Şekil 15. SIFT tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

OPTICS yapısı tercih edilerek tanımlayıcı vektörler kullanılarak sınıflandırma işlemi gerçekleştirildikten sonra elde edilen anahtar noktaları Şekil 16a ve 16b'de sırasıyla orijinal ve sahte sesler için gösterilmektedir. Görüntüler üzerinde beyaz dairelerle belirtildiği üzere, oldukça az sayıda kümelenmiş anahtar noktası kalmaktadır. Kümelenmiş anahtar noktalarının sayıca azalması, sahte ses tespitine olumsuz olarak yansımaktadır. Bu durum, sahte seslere ait spektrogramda eşleşme sayısının fazla olması gerektiği beklentisi ile çelişmektedir. Eşleşme sayısının yetersiz olması, sahte seslerin doğru ve güvenilir şekilde tespit edilmesini zorlaştırmakta ve sahtecilik tespitinin etkinliğini azaltmaktadır. Dolayısıyla, kümelenmiş anahtar noktalarının yeterli sayıda olması, sahte ses tespitinde kritik rol oynamaktadır.



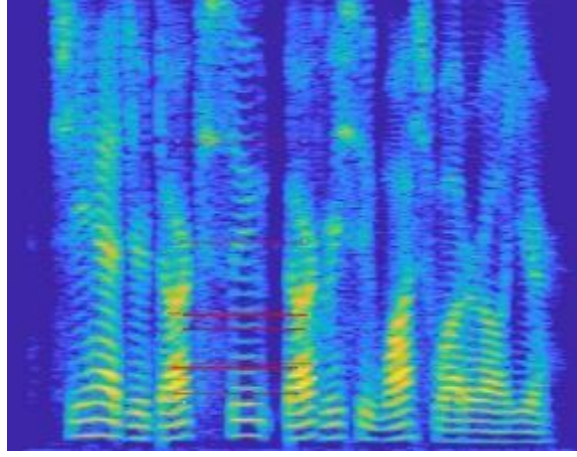
Şekil 16. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses

Elde edilen anahtar noktalar birleştirilerek kopyala-yapıştır sahte ses tespitinde orijinal ve sahte olarak etiketlenmesi ve sahte ses üzerinde lokalizasyonunun belirlenmesi gerçekleştirilmektedir. Bu işlem, sadece yatay doğruların çizilmesi ile yapılmaktadır. Bunun nedeni, dikey eksenin farklı frekansları temsil etmesidir. Şekil 17a ve 17b, sırasıyla orijinal ve sahte seslere ait eşleşmeleri göstermektedir. Yatay doğrular aracılığıyla yapılan bu eşleştirme, sahtecilik girişimlerinin tespit edilmesi ve lokalize edilmesine olanak tanımaktadır. Bu yöntem, farklı frekanslardaki ses bileşenlerinin doğru şekilde analiz edilmesini ve sahtecilik tespitinde yüksek doğruluk elde edilmesini sağlamaktadır.



Şekil 17. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses

Orijinal seste herhangi eşleşme bulunamamıştır. Buna karşın, sahte ses üzerinde birkaç eşleşme tespit edilmiştir. Şekil 17b incelendiğinde, aykırı eşleşmenin de tespit edildiği görülmektedir. Bu aykırı eşleşme, yalnızca bir tane olarak belirlenmiştir. Eğer sahtecilik, bu tek eşleşmenin olduğu bölgede gerçekleşmiş olsaydı, birden fazla eşleşmenin bu kısımda oluşması beklenirdi. Olası yanlış eşleşmelerin ortadan kaldırılması için RANSAC algoritması kullanılmıştır. Şekil 18, sahte sese ait yüksek çözünürlüklü spektrogramda muhtemel hatalı eşleşmelerin ortadan kaldırılmış halini göstermektedir. Bu işlem, sahtecilik tespitinin doğruluğunu artırmakta ve yalnızca güvenilir eşleşmelerin değerlendirilmesini sağlamaktadır.



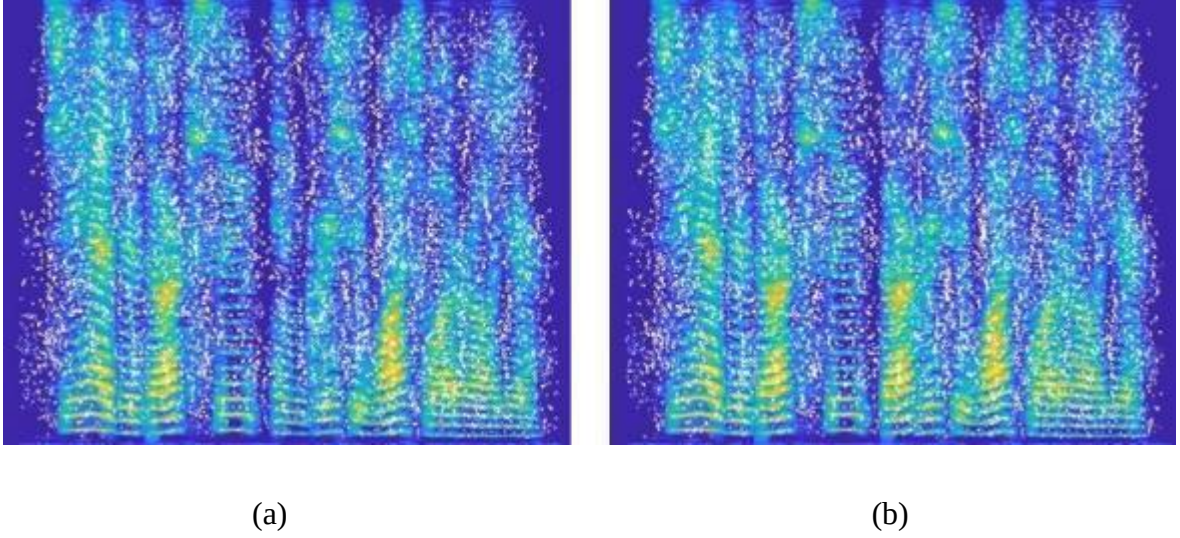
Şekil 18. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması

2.2.1.4. SURF Anahtar Noktası Metodu

Tez kapsamında kopyala yapıştır ses sahteciliği tespitinde SURF anahtar noktaları kullanılması üzerinde de çalışmalar yapılmıştır. Konuşma kaydından yüksek çözünürlüklü spektrogram görüntüleri çıkarılmış ve bu resimlerden SURF anahtar noktaları ve onların tanımlayıcı vektörleri çıkarılmıştır.

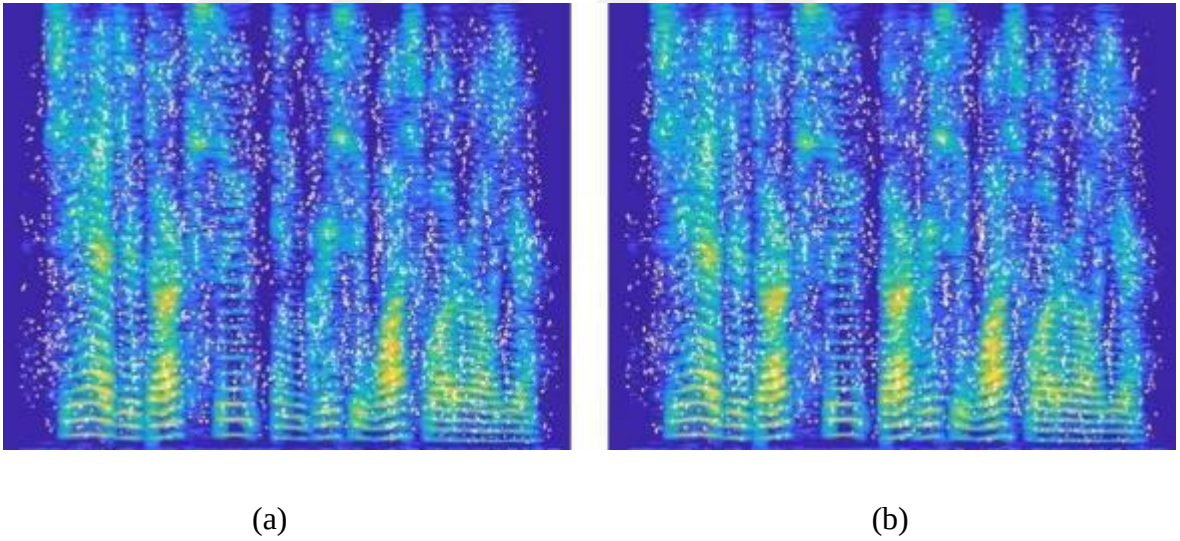
SURF (Speeded-Up Robust Features) anahtar nokta çıkarım yöntemi, diğer yöntemlere kıyasla birçok avantaja sahiptir. İlk olarak, SIFT gibi benzer yöntemlere göre çok daha hızlıdır, bu da gerçek zamanlı uygulamalarda ve büyük veri setlerinde önemli performans artışı sağlar. SURF, ölçek ve döndürme değişmezliği sayesinde farklı boyut ve açılardaki görüntülerde güvenilir anahtar noktaların tespit edilmesine olanak tanır. Ayrıca, Hessian matrisine dayalı tespit yöntemi ile yüksek hassasiyet ve doğruluk sunar, bu da anahtar noktaların görüntüdeki önemli özellikleri daha iyi temsil etmesini sağlar. SURF, aydınlatma ve kontrast değişikliklerine karşı dayanıklıdır, bu sayede farklı ışık koşullarında bile tutarlı sonuçlar elde edilebilir. Yüksek ayrıştırma gücüne sahip anahtar noktalar çıkarması, nesne tanıma ve izleme gibi uygulamalarda daha güvenilir ve doğru sonuçlar elde edilmesini sağlar. Bu özellikler, SURF'ü hız, hassasiyet ve dayanıklılık açısından güçlü araç haline getirerek, bilgisayarla görme ve görüntü işleme uygulamalarında geniş kullanım alanına sahip olmasını sağlar.

Şekil 19a ve 19b sırasıyla orijinal ve sahte seslere ait yüksek çözünürlüklü spektrogram görüntülerinde tespit edilmiş SURF anahtar noktaları temsil etmektedir. Spektrogram üzerinde oldukça fazla anahtar noktaların belirlenmesi yöntemin en büyük avantajı olarak gösterilebilir.



Şekil 19. SURF tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

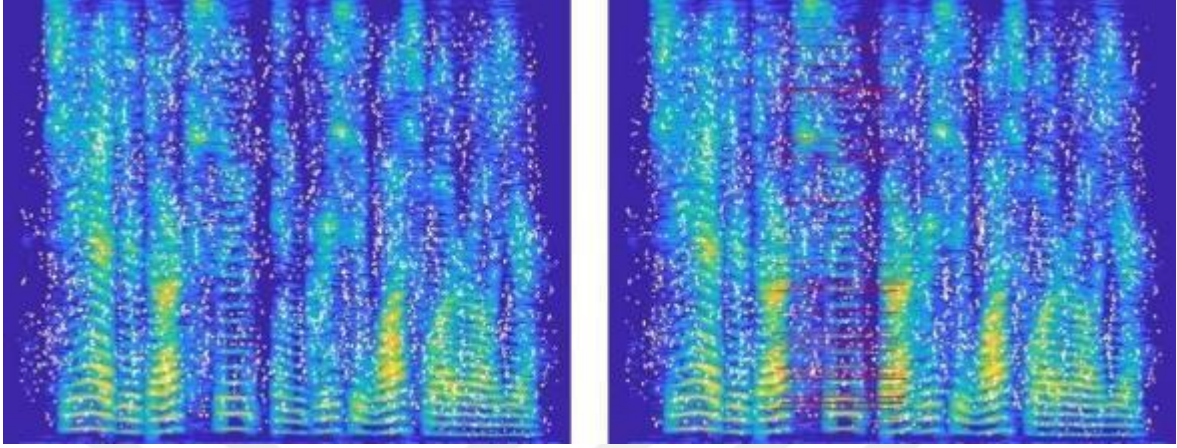
Görüntü üzerinde çıkarılan anahtar noktaları onların tanımlayıcı vektörleri kullanılarak OPTICS algoritması ile kümelendirilmiştir. Bu işlem sonucunda herhangi kümeye ait olan anahtar noktaları sırasıyla orijinal ve sahte seslere ait olan spektrogramlar üzerinde Şekil 20a ve Şekil 20b’de gösterilmiştir.



Şekil 20. OPTICS tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses

Makine öğrenmesi yapılarından OPTICS ile kümelendirilen SURF anahtar noktalarına ait olan özellik tanımlayıcıları görüntü üzerinde birleştirilmektedir. Bu süreç sonucunda, görüntü üzerinde çizgiler oluşturulmaktadır. Elde edilen bu çizgiler, kopyala-yapıştır sahte ses tespiti için işarettir. Spektrogram üzerinde yatay eksenin zamanı temsil etmesi nedeniyle, doğru başlangıç ve bitiş noktaları sahtecilik segmentini belirtir. Şekil 21a

ve 21b, sahte sese ait spektrogram görüntüsü üzerinde eşleşmeleri göstermektedir. Bu süreç, sahte sesin zaman açısından belirlenmesini ve sahtecilik segmentinin lokalizasyonunu sağlayarak, sahte ses tespitinde etkin yöntem olarak kullanılmaktadır.

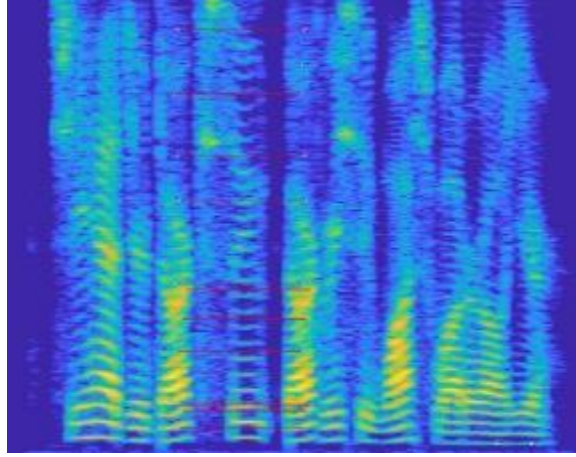


(a)

(b)

Şekil 21. OPTICS tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses

Orijinal spektrogramda herhangi eşleşme tespit edilmemiştir, ancak sahte spektrogramda oldukça fazla eşleşme belirlenmiştir. Bu eşleşmeler arasında, hatalı eşleşmeler de tespit edilmiştir. Olası yanlış eşleşmelerin ortadan kaldırılması için RANSAC algoritması kullanılmaktadır. RANSAC, aykırı verileri tanımlayarak, bu aykırıları dikkate almadan modelin oluşturulmasını sağlar ve böylece yanlış eşleşmelerin etkisini azaltır. Bu süreç, sahte spektrogramda meydana gelen yanlış eşleşmelerin doğru şekilde ele alınmasını ve sahte sesin doğru şekilde tespit edilmesini sağlar. Bu yöntem, sahte ses tespitindeki doğruluk ve güvenilirlik açısından önemli rol oynamaktadır. Şekil 22 sahte sese ait spektrogram üzerinde olası hatalı eşleşmelerin elimine edilmiş halidir.



Şekil 22. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması

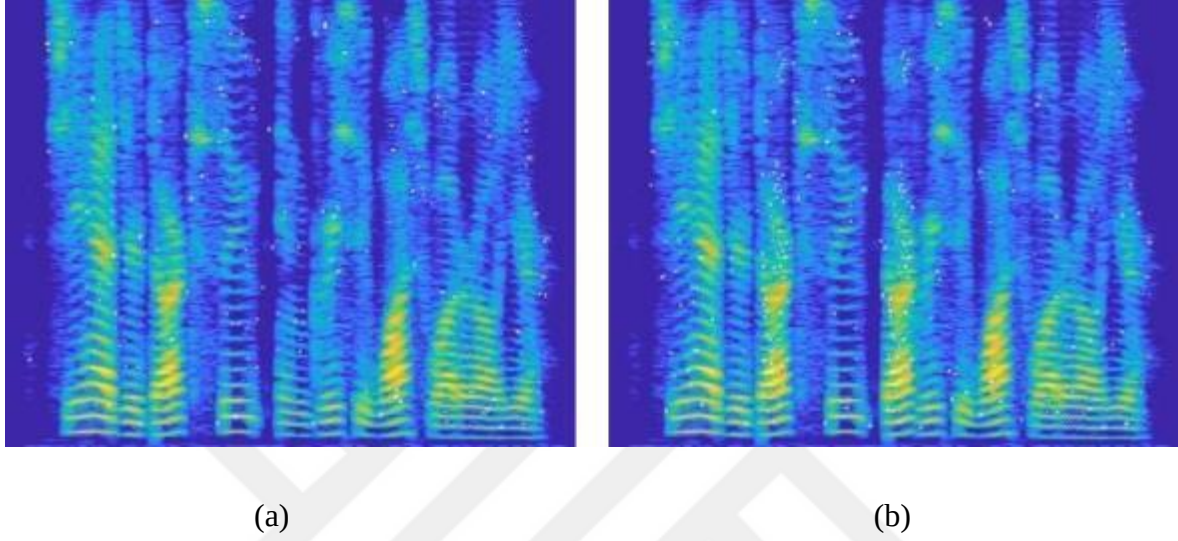
2.2.2. BRIEF Anahtar Noktalarının g2NN Kümeleme Yaklaşımıyla Kullanılması

Kopyala yapıştır ses sahteciliği tespitinde tez kapsamı içerisinde önerdiğimiz bu sistemde konuşma kaydından yüksek çözünürlüklü spektrogram görüntüsü alınmaktadır. Yüksek çözünürlüklü spektrogramlardan BRIEF anahtar noktaları çıkarıldıktan sonra kümeleme yapısı olarak g2NN kullanılarak anahtar noktalar tanımlayıcı vektörler aracılığıyla kümelendirilmiştir.

g2NN, graf tabanlı yaklaşım olduğundan, veri kümesindeki karmaşık ilişkileri ve yapıları daha etkin şekilde yakalayabilir. Bu, veri setlerindeki doğrusal olmayan ilişkilerin ve yapıların daha iyi modellenmesine olanak tanır. Ayrıca, g2NN, geleneksel kümeleme algoritmalarının aksine, veriye önceden belirlenmiş küme sayısı verme zorunluluğunu ortadan kaldırır. Bu, veri setindeki doğal küme sayısının bilinmediği veya değişken olduğu durumlarda daha esnek çözüm sunar. g2NN ayrıca, yoğun ve seyrek veri setlerinde iyi performans gösterir. Yoğun veri setlerinde, benzer özelliklere sahip veri noktaları arasındaki ilişkileri belirlemek için etkili şekilde çalışabilir. Seyrek veri setlerinde ise, veri noktaları arasındaki uzaklıkların tam olarak hesaplanmasına gerek olmadığından, hesaplama maliyetini azaltır. Bu, büyük ölçekli veri setlerinde daha verimli kümeleme süreci sağlar.

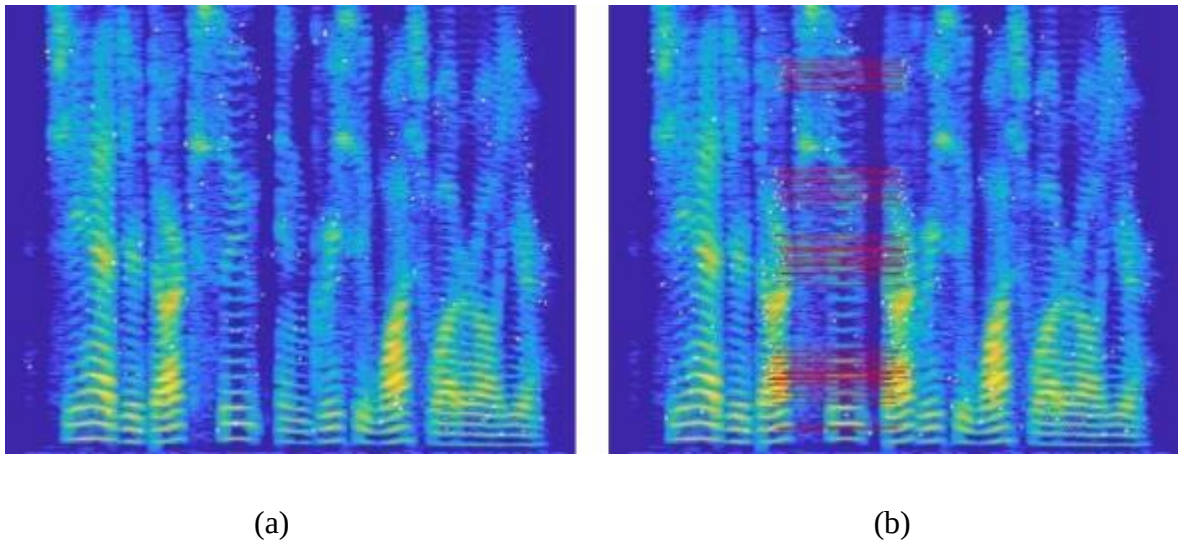
Son olarak, g2NN, aykırı verilere karşı dirençlidir. Graf tabanlı yaklaşımı sayesinde, aykırı verilerin kümelere etkisi azaltılabilir ve bu şekilde daha homojen ve temsilci kümeler elde edilebilir. Bu özellik, veri setlerindeki gürültüyü ve dengesizliği azaltarak daha tutarlı ve güvenilir sonuçlar elde edilmesini sağlar.

Konuşma kaydından yüksek çözünürlüklü spektrogram alındıktan sonra görüntü üzerinden BRIEF anahtar noktası ve onların tanımlayıcı vektörleri çıkarılmaktadır. Elde edilen bu noktalar tanımlayıcı özellik vektörleri kullanılarak g2NN algoritması ile kümelendirilmektedir. Bu işlem sonrasında kümelenmiş anahtar noktaları Şekil 23a ve Şekil 23b’de gösterilmektedir. Şekil 23a orijinal sese 23b sahte sese aittir.



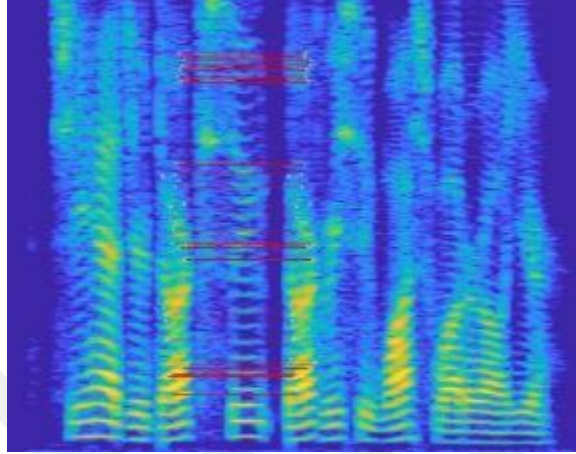
Şekil 23. Merkez Çevre Ekstrema tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

Aykırı verilerin g2NN ile etkisi azaltıldığı için kümelendirme sonucunda anahtar noktalarda oldukça azalma gözlemlenmektedir. Şekil 24a ve 24b sırasıyla orijinal ve sahte seslere ait spektrogramlar üzerinde BRIEF anahtar noktalarının g2NN ile kümelendirilmesi sonucu eşleşmeleri göstermektedir. Orijinal ses üzerinde herhangi eşleşme görülmemiştir. Buna karşın sahte ses üzerinde birçok eşleşme görülmektedir.



Şekil 24. g2NN tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses

Son olarak görüntü üzerinde olası hatalı eşleşmeleri ortadan kaldırmak için RANSAC işlemi belirlenen eşleşmelere uygulanmaktadır. Şekil 25 sahte sese ait son eşleşen doğruları temsil etmektedir.



Şekil 25. Sahte spektrogram üzerinde RANSAC tarafından hatalı eşleşmelerin kaldırılması

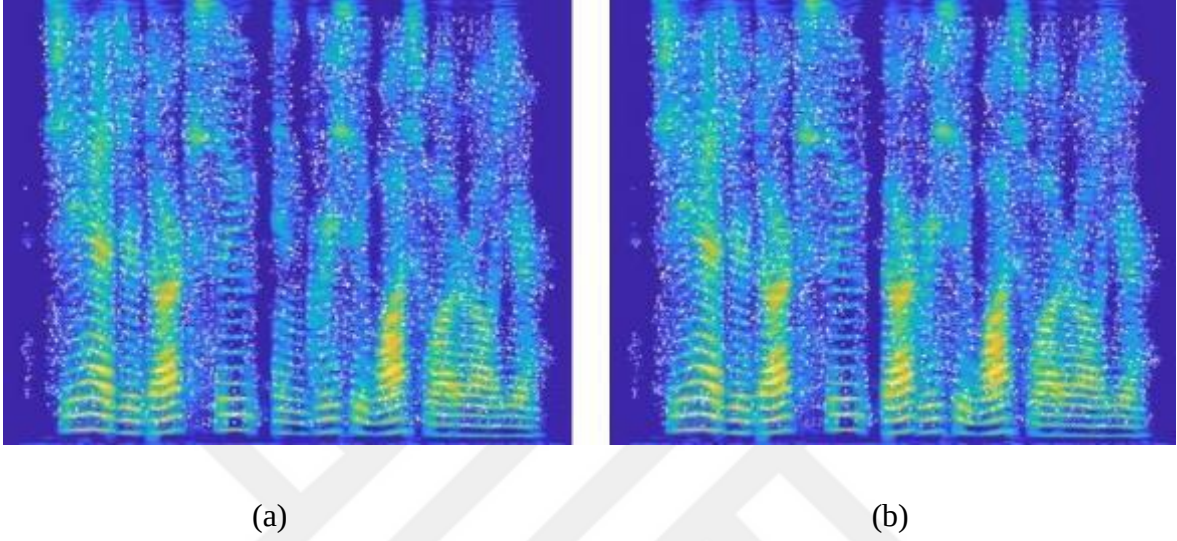
2.2.3. BRIEF Anahtar Noktalarının k-Means Kümeleme Yaklaşımıyla Kullanılması

Sahtecilik tespitinde önerilen diğer metotta, yüksek çözünürlüklü spektrogramdan BRIEF anahtar noktaları çıkarılmaktadır. Elde edilen bu noktalar, k-Means algoritması ile kümelendirilerek orijinal ve sahte seslerin etiketlenmesi yapılmaktadır. Eğer ses sahte ise, kopyala-yapıştır konumu hakkında bilgi edinmek için eşleşen doğruların başlangıç ve bitiş noktaları kullanılmaktadır. Bu yöntem, sahte sesin lokalizasyonunu ve sahtecilik segmentinin belirlenmesini sağlayarak, ses sahteciliğinin tespitinde etkili yaklaşım sunmaktadır.

k-Means algoritması basit ve anlaşılır yapıya sahiptir, bu da uygulanmasını ve yorumlanmasını kolaylaştırır. Hesaplama açısından verimli olan k-Means, büyük veri setlerinde hızlı şekilde çalışabilir ve bu nedenle yüksek performans gerektiren uygulamalarda tercih edilir. Ayrıca, k-Means, küme merkezlerini kullanarak verilerin doğrudan ayrıştırılmasını sağlar, bu da kümeler arasındaki mesafeleri minimize ederek homojen kümeler oluşturur. Bu yöntem, özellikle küme sayısının önceden bilindiği durumlarda etkili sonuçlar verir. Son olarak, k-Means algoritması, verilerin geometrik

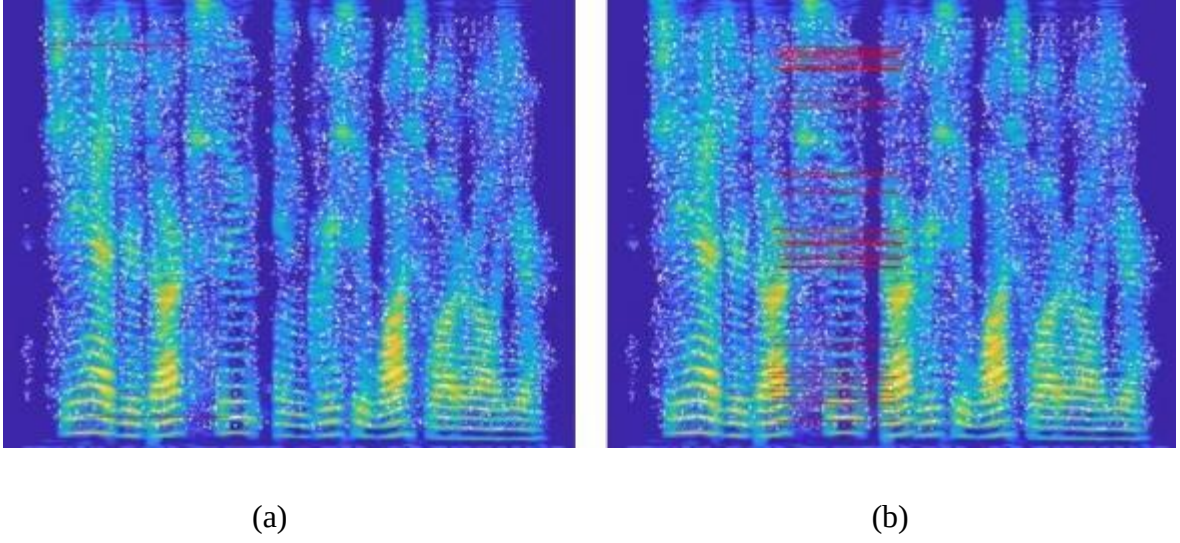
yapısını iyi yansıtarak kümelerin kompakt ve ayrık olmasını sağlar, bu da birçok uygulamada doğruluğu artırır ve kullanıcıların veri içindeki yapıları daha iyi anlamalarına yardımcı olur.

Şekil 26a ve 26b sırasıyla orijinal ve sahte seslere ait BRIEF ile elde edilmiş anahtar noktaların k-Means ile kümelendirilmesi sonucu elde edilen anahtar noktaları göstermektedir.



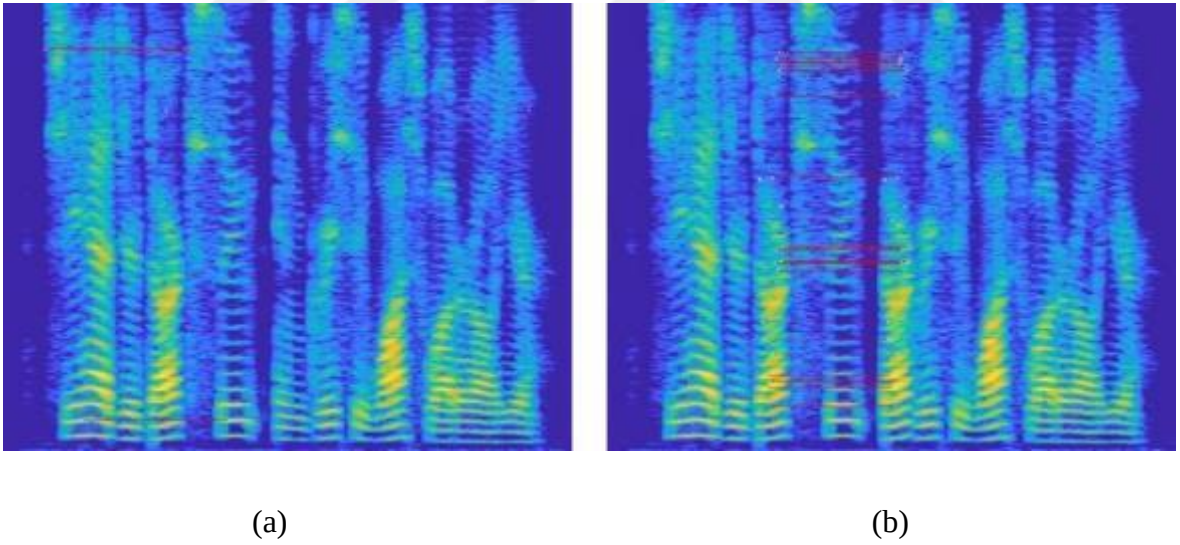
Şekil 26. Merkez Çevre Ekstrema tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

Kümelendirme işleminden sonra, yatay ekseninde aynı kümeye ait anahtar noktalar birleştirilerek eşleşmeler görüntüde ortaya çıkarılmaktadır. Orijinal seslerde, eşleşmenin çok az veya hiç olmaması beklenirken, sahte seslerde eşleşmenin oldukça fazla olması beklenmektedir. Bu durum, sahte seslerdeki kopyala-yapıştır işlemlerinin belirginliğini artırmakta ve tespit sürecini kolaylaştırmaktadır. Elde edilen eşleşmeler, sahtecilik segmentlerinin doğru şekilde lokalize edilmesine olanak tanır, bu da sahtecilik tespitinin doğruluğunu ve güvenilirliğini artırmaktadır. Şekil 27a ve 27b sırasıyla orijinal ve sahte seslere ait eşleşmeleri göstermektedir. Orijinal görüntü üzerinde birbirinden aykırı bölgelerde oluşmuş eşleşmeler gözlemlenmektedir. Bu eşleşmeler sahtecilik konusunda yanlış yönlendirmelere sebep olmaktadır.



Şekil 27. k-Means tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses

Son olarak seslere ait eşleşmelere olası hatalı eşleşmeleri ortadan kaldırmak için RANSAC algoritması uygulanmaktadır. Şekil 28a ve 28b sırasıyla orijinal ve sahte seslere ait eşleşmelerde hatalı olanlarının kaldırılmış hali verilmektedir.



Şekil 28. k-Means tarafından belirlenen eşleşmeler (a) Orijinal ses (b) Sahte ses

Önceki aşamalarda, çeşitli anahtar nokta çıkarma yöntemleri ve kümeleme yapıları kullanılarak eşleştirilmiş ve iki kümeye ayrılmıştır. Bu kümeler, ses spektrogramındaki yinelenen segmentlere karşılık gelmektedir. Spektrogramdaki segmentlerin seste hangi segmentlere karşılık geldiğini belirlemek amacıyla Algoritma (3) önerilmiştir. Önerilen algoritmada, ilk olarak kümelerdeki anahtar noktaların yatay eksenindeki ortalaması

alınmaktadır. Daha sonra, hesaplanan ortalama noktaların, algoritmada verilen dönüşüm formülü kullanılarak seste hangi noktalara karşılık geldiği belirlenmektedir. Son olarak, bu noktaların hangi segmentte yer aldığı incelenerek yinelenen segmentler tespit edilmektedir.

Algoritma 3. Segment Tespiti

Girdi. *Eşleşmelerin yatay eksen koordinatları (X eksen), Segmentler*

Çıktı. *Kopyalanmis_Segment, Yapistirilmis_Segment*

Adım 1. Eşleşmelerin Başlangıç noktalarının ortalaması, *baslangic_ortalama*

Adım 2. Eşleşmelerin Bitiş noktalarının ortalaması, *bitis_ortalama*

Adım 3. $Ses_x1 = (baslangic_ortalama / spektrogram_genisligi) * ses_uzunlugu$

Adım 4. $Ses_x2 = (bitis_ortalama / spektrogram_genisligi) * ses_uzunlugu$

Adım 5. if segment in *Ses_x1*

Kopyalanmis_Segment = segment

end if

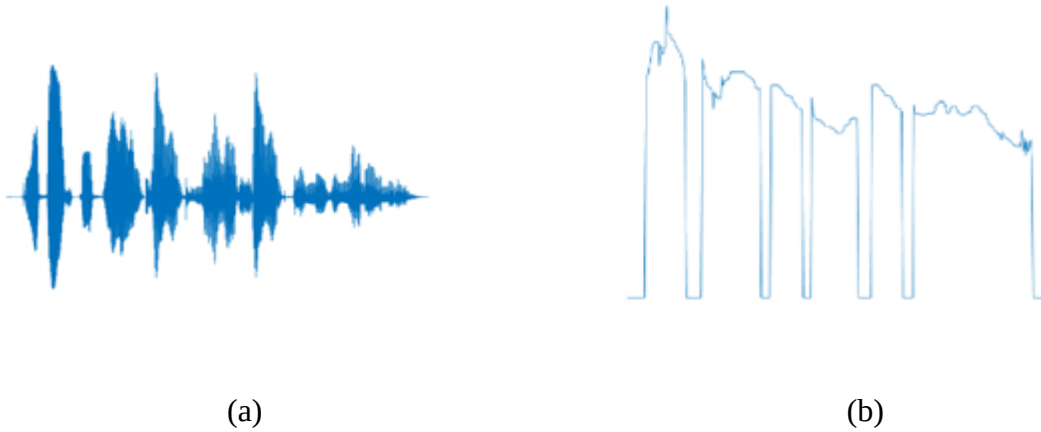
Adım 6. if segment in *Ses_x1*

Yapistirilmis_Segment = segment

end if

Adım 7. Adım 5-6 Perde dizisi üzerindeki tüm segmentler boyunca tekrarla

Algoritmanın girdi olarak aldığı segment dizisi Perde dizisi kullanılarak elde edilmektedir. Şekil 29a ve 29b sırasıyla sahte sese ve bu sese ait Perde dizisini göstermektedir. Kayıta konuşmacı kurduğu cümle 6 kelimeden ibarettir. Bu nedenle YAAPT metoduyla çıkarılan Perde dizileri 6 adettir.



Şekil 29. Sahte sesin dalga formu ve Perde dizisi (a) Dalga formu (b) Perde dizisi

Önerilen algoritma Spektrogram üzerinde yapılan işaretlerin başlangıç ve bitiş noktalarının ortalamasını Perde dizilerinden alınan segmentlere iz düşürerek kopyala yapıştır sahte bölgeleri belirlemektedir. İz düştüğü Perde dizisinin tamamı sahte segment olarak etikenmektedir. Son olarak Perde dizisinden sese geçilerek kopyalanmış ve yapıştırılmış ses parçaları belirlenmektedir.

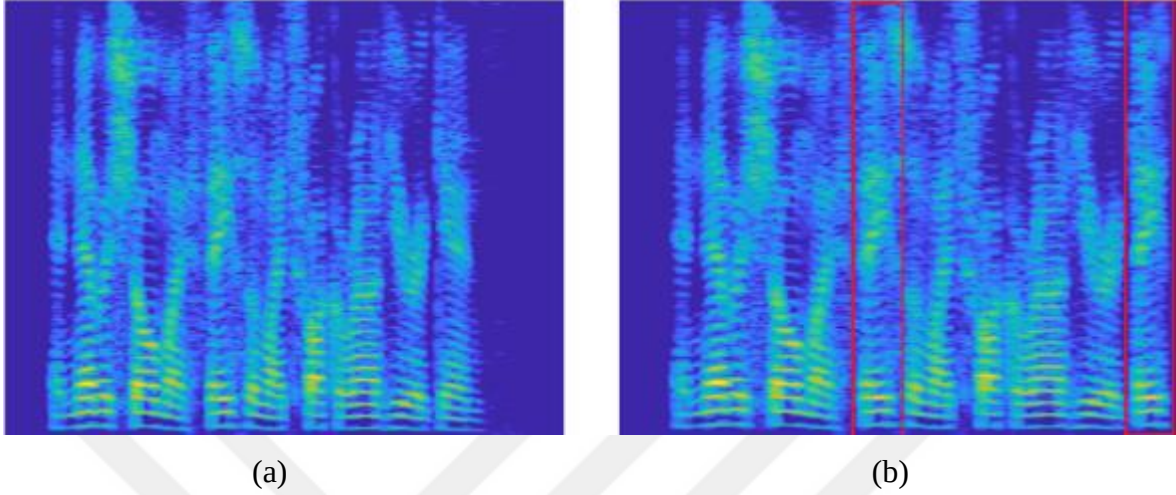
Eşleşmelerin başlangıç ve bitiş noktalarını sese izdüşürerek, seste yinelenen segmentler tespit edilebilir. Ancak, saldırganlar tarafından yapılan sahtekarlığı gizlemek amacıyla ses verilerine atak uygulanması, saldırısız spektrogram görüntülerine kıyasla spektrogramdaki anahtar noktaların sayısını azaltır.

2.3. Graf Tabanlı Kopyala Yapıştır Ses Sahteciliği Tespiti

Tez kapsamı içerisinde kopyala yapıştır ses sahteciliği tespitinde yapılan diğer çalışma konuşma kaydından elde edilen graf görüntülerinin sınıflandırılması olmuştur. Bu sınıflandırma işlemi ses sahteciliği tespitinde konuşma kaydını sahte veya orijinal olarak etiketler. Yöntemin aşamaları şu şekilde olmaktadır:

- Konuşma kaydından Yüksek Çözünürlüklü Spektrogram görüntüsü elde edilir.
- Literatürde sık kullanılan SIFT anahtar noktası ve tanımlayıcılarının çıkarılması yöntemi sesin frekans temsiline uygulanır.
- Çıkarılan anahtar noktaları onların tanımlayıcı vektörleri kullanılarak g2NN algoritması ile kümelendirilir.
- Sınıflandırılan anahtar noktaları görüntü üzerinde eşleştirilerek yatay ekseninde sahtecilik lokalizasyonu ve dikey ekseninde frekans bilgisi işaretlenmiş olur.
- Dikey ekseninde eşleşmelerin yoğun olduğu bölgenin frekans bilgisi kullanılarak konuşma kaydına band geçiren süzgeç işlemi yapılır.
- Filtreleme yapılarak işleme alınmış sese Hızlı Fourier Transform uygulanarak frekans uzayına geçilir.
- Süzgeç uygulandıktan sonra frekans uzayına dönüştürülmüş seste deneysel çalışmalarla belirlenen eşik kullanılarak veri üzerinde sıfırlama işlemi yapılır.
- Son aşamada graf görüntüsü çıkarılarak önerilen CNN mimarisi ve transfer öğrenme yapılarından VGG16 ve VGG19 kullanılarak sınıflandırma gerçekleştirilir ve nihai karar füzyon denklemi ile belirlenir.

Şekil 30a ve 30b sırasıyla orijinal ve sahte konuşma kayıtlarına ait yüksek çözünürlüklü spektrogram görselleridir. Sahte görüntü üzerindeki kırmızı kutucuklar kopyala yapıştır segmentleri işaret etmektedir.



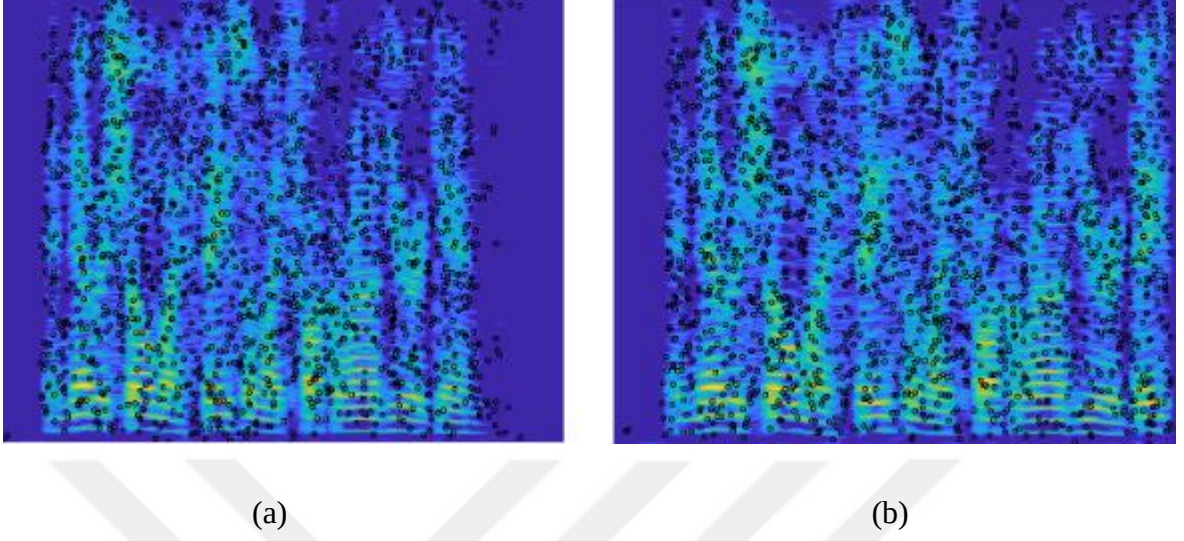
Şekil 30. Yüksek Çözünürlüklü Spektrogram (a) Orijinal ses (b) Sahte ses

Yüksek çözünürlüklü spektrogramlar zaman ve frekans eksenlerinde daha ince detaylar sağlayarak kısa süreli olayların ve yüksek frekanslı bileşenlerin net şekilde görülmesini mümkün kılar. Bu, özellikle müzik analizi, konuşma tanıma ve bioakustik araştırmalar gibi alanlarda, birbirine yakın frekans bileşenlerinin daha doğru şekilde ayırt edilmesini sağlar. Ayrıca, gürültü ve sinyal arasındaki farkların daha belirgin hale gelmesi, sinyal işleme ve filtreleme süreçlerini daha etkili kılarak, analizlerin doğruluğunu artırır.

2.3.1. Yüksek Çözünürlüklü Spektrogram Görüntülerinden SIFT Anahtar Noktası Çıkarılması

Mevcut ses düzenleme yazılımları kullanılarak gerçekleştirilen ses sahteciliği, kopyalanıp yapıştırılan sesin frekans bilgisi üzerinde değişiklik yapmayacağı için 2 boyutlu uzaya haritalanmış spektrogramlarda piksel bilgisinin de korunması beklenmektedir. Bu bağlamda zaman ekseninde aynı piksel dizilerine sahte seste karşılaşılması muhtemeldir. Bu benzerliği yakalayabilmek amacıyla görüntü üzerinde küçük çerçevelerle çalışan SIFT anahtar noktası çıkarım metodu kullanılarak spektrogramdan özellikler çıkarılmaktadır. Görüntü üzerinde pencereler kullanılarak işlem yapılmasının en uygun pencere boyutuna karar vermek gibi birçok sakıncaları vardır. Ancak anahtar noktası ile çalışmak faydalı özellikler ve pikseller ile işlemini gerçekleştirdiği için daha kullanışlıdır.

Şekil 31a ve 31b sırasıyla orijinal ve sahte görüntü üzerinde belirlenen anahtar noktaları ifade etmektedir.



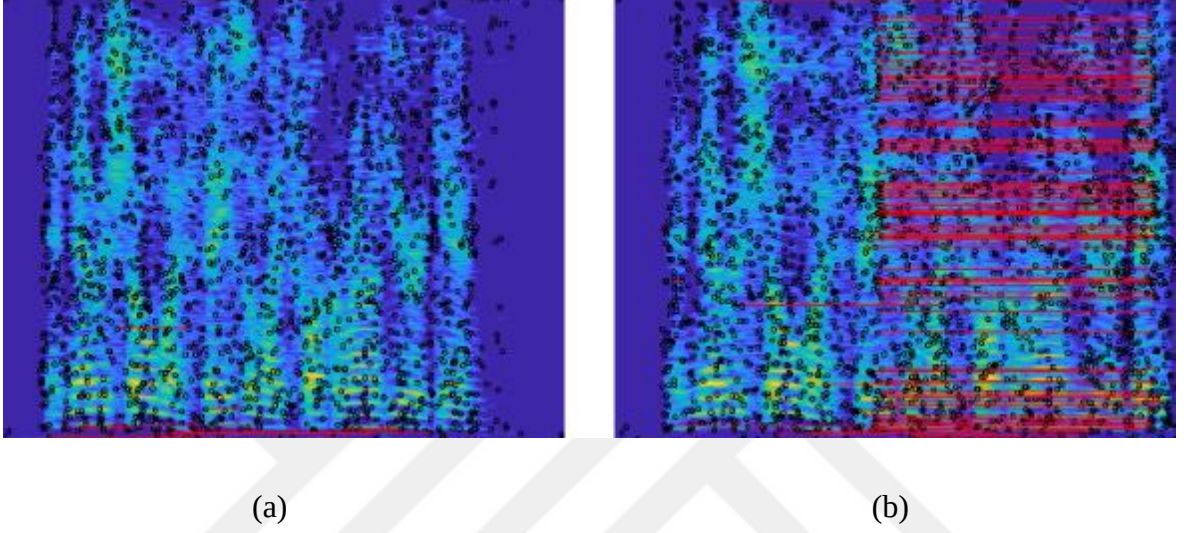
Şekil 31. SIFT tarafından belirlenen anahtar noktalar (a) Orijinal ses (b) Sahte ses

Ses kaydı üzerinde kopyalanıp taşınan parçanın spektrogram görüntüsünde frekans bilgisinin taşınmış olması, analiz açısından büyük önem taşır. Bu özellik, iki frekans segmenti karşılaştırıldığında benzer anahtar noktaların ve tanımlayıcıların kolayca belirlenmesini sağlar. Spektrogramda görülen frekans bilgileri, kopyalanan parçanın özgün yapısını koruyarak, analiz sırasında doğru ve güvenilir sonuçlar elde edilmesini mümkün kılar. Bu sayede, ses kayıtlarının incelenmesi ve karşılaştırılması sırasında, spektrogram görüntüleri üzerinde frekans bilgilerine dayanarak yapılan analizler, sesin belirli karakteristik özelliklerini ortaya çıkarmada etkili olur.

2.3.2. g2NN ile Anahtar Noktalarının Kümelendirilmesi

Kümeleme algoritmaları veri içerisindeki aynı özelliğe sahip elemanları ortak çatı altında toplamak amacıyla kullanılan yapılardır. Çalışma prensipleri tekrarlı şekilde çalışıp elemanları özelliklerine göre sınıfa atayıp etiketlemesine dayanmaktadır. Kopyalanıp yapıştırılan segmentler, spektrogram üzerinde birbirine çok yakın piksel değerlerine sahip olduğundan, bu bölümler anahtar noktalar ve tanımlayıcılar açısından birebir veya çok az farklılık gösterir. Bu bağlamda anahtar noktalarını g2NN kümeleme yaklaşımı kullanarak gruplandırıyoruz.

Kümelendirme işlemi anahtar noktasının sahip olduğu özellik tanımlayıcıları kullanılarak gerçekleştirilir. Söz konusu vektörler birbirleri arasında karşılaştırılarak g2NN tarafından belirlenen şartı sağlayan noktalara aynı sınıf etiketi atanmaktadır. Bu atama işlemi sonucunda aynı sınıfa sahip noktalar çevresine göre benzer niteliğe sahiptir. Şekil 32a ve 32b sırasıyla orijinal ve sahte sınıfa ait g2NN kümeleme işlemi sonucu eşleşmeleri göstermektedir.



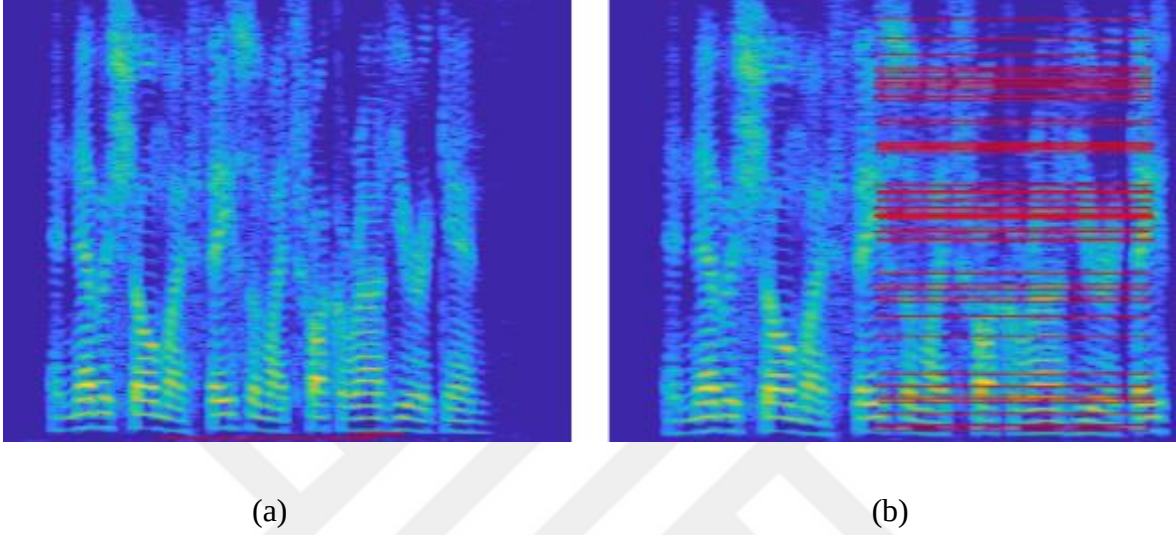
Şekil 32. g2NN tarafından kümelendirilmiş anahtar noktalar (a) Orijinal ses (b) Sahte ses

Eşleşmelerin birbirine çok aykırı olmaması beklenir; yatay ekseninde tüm doğruların başlangıç noktalarının birbirine yakın değerlere sahip olması gerekmektedir. Aynı durum bitiş noktaları için de geçerlidir. Başlangıç ve bitiş noktaları bu şarta uymayan eşleşmeler hatalı eşleşme olarak adlandırılır. Bu kriterlere göre, doğruların başlangıç ve bitiş noktalarındaki uyum, doğru eşleşmenin belirlenmesinde kritik öneme sahiptir.

2.3.3. Hatalı Eşleşmelerin Ortadan Kaldırılması

Spektrogram üzerinde kopyala yapıştır ses sahteciliği tespitinde, bir bölgede eşleşme tespit edilmesi, o bölgede benzer özelliklere sahip piksel bloklarının olduğunu gösterir. Ancak, sadece birkaç benzer bloğun bulunması, bu segmentlerin kopyala yapıştır segmentleri olduğunu garanti etmez. Kopyala yapıştır tespitinin güvenilir olabilmesi için, ilgili bölge üzerinde yoğun şekilde eşleşmelerin elde edilmesi beklenir. Bu şarta uymayan eşleşmeler, hatalı eşleşme olarak kabul edilir. Bu bağlamda, kümeleme algoritması kullanılarak elde edilen eşleşmeler, olası hatalı eşleşmelerin elimine edildiği algoritmaya

tabi tutulur. Bu süreç, kopyala yapıştır sahteciliğinin daha doğru ve güvenilir şekilde tespit edilmesini sağlar. Şekil 33a ve 33b sırasıyla orijinal ve sahte spektrogramlar üzerinde hatalı eşleşmelerin elimine edilmiş halini göstermektedir. Orijinal spektrogramda birkaç eşleşmenin kaldığı görülmektedir. Sahte spektrogramda eşleşmeler bu aşamada düzen içerisinde. Doğruların başlangıç ve bitiş değerleri birbirine yakın konumdadır.

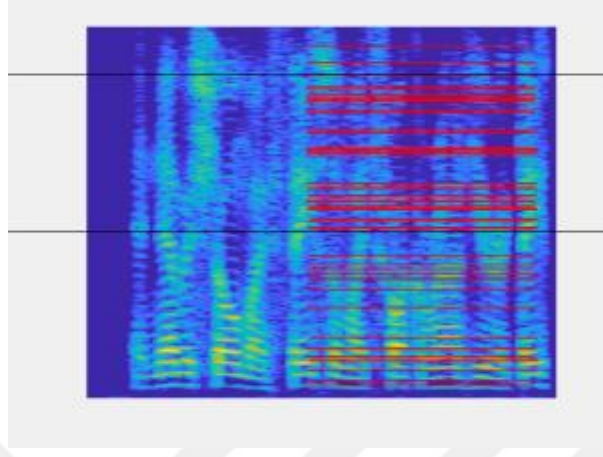


Şekil 33. Parametre Normalizasyonu tarafından hatalı eşleşmelerin kaldırılması (a) Orijinal ses (b) Sahte ses

2.3.3. Yoğun Kümelenme Bölgelerinin Tespit Edilmesi

Yöntemin amacı sahte ve orijinal konuşma kayıtlarını birbirinden ayırmaktır. Spektrogram üzerinde dikey eksen frekans bilgisini barındırmaktadır. Görüntü üzerindeki pikseller frekansın genliğini bir başka ifade ile şiddetini temsil etmektedir. Anahtar noktalarının kümelenmesi sonucu oluşan eksenlerin belli frekans bölgesinde yoğunlaşması beklenmektedir. Bu bölge sahte segmentlerin frekans bilgisini taşımaktadır. Öncelikli amaç eşleşme sonrası doğruların yoğunlaştığı kısmı tespit etmektir. Görüntü üzerinde deneysel çalışmalarla belirlenen sayıda boyutta çerçeve birbiri ile örtüşecek şekilde gezdirilerek eşleşmenin yoğun olduğu piksel aralıkları belirlenir. Belirlenen piksel aralıkları frekans uzayına iz düşürülerek dönüşüm gerçekleştirilir. Elde edilen başlangıç ve bitiş frekansları kopyala yapıştır ses segmentlerinin frekans aralıklarıdır. Orijinal spektrogram üzerinde bulunan eşleşme sayısı, deneysel çalışmalarla belirlenen sınırın altında olduğu için yoğun kümelenme bölgelerinin tespit edilmesi algoritması çalıştırılmamıştır. Buna karşın, sahte spektrogram üzerinde yoğun kümelenme bölgeleri tespit edilmiştir. Şekil 34, sahte spektrogram üzerinde belirlenen yoğun eşleşmeleri göstermektedir. Sınırları siyah çizgilerle

belirlenen alan, söz konusu bölgeyi işaret etmektedir. Bu bölgenin frekans bilgisi elde edilerek, sahte segmentlerin frekans bilgisi çıkarılmıştır.



Şekil 34. Sahte spektrogram üzerinde yoğun eşleşme bölgelerinin tespit edilmesi

2.3.4. Band Geçiren Süzgeç İşlemi

Band geçiren süzgeç işlemi sinyal üzerinde belirlenen frekans aralığını geçirip diğer frekansları bastırmaktadır. Bu işlem bir nevi ses sinyali üzerinde işaretleme yapmaktır. Konuşma kaydına uygulanan band geçiren süzgeç işlemi sinyalin frekansı üzerinde büyük etkiye sahiptir. Orijinal kayıtlarda anahtar noktası eşleşmesi öngörülmediğinden, band geçiren süzgeç işleminin uygulanması beklenmez. Band geçiren süzgeç uygulanması tüm frekans aralığının ses üzerinde saklanması anlamı taşımaktadır. Ancak sahte seslerde filtre uygulanacaktır. Bu uygulanan filtre frekans aralığı, sınırlı çerçeveye sahip piksel bilgisinden dönüştürüldüğü için, belirli aralıkta olmaktadır. Sonuç itibariyle bu aşamada konuşma kaydı frekans bilgisi açısından işaretlenmiştir. Şekil 35.a ve Şekil 35.b sırasıyla sahte ses kaydını ve bu kayda ait band geçiren süzgeç uygulanmış sinyali göstermektedir.



(a)

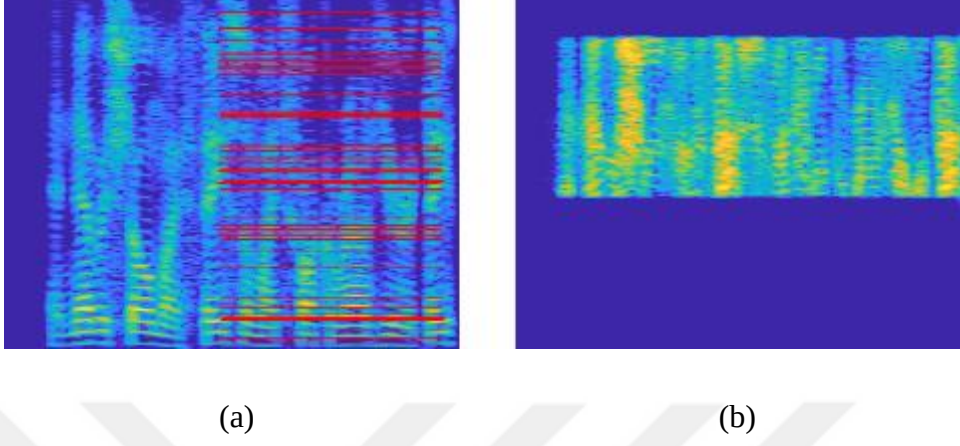


(b)

Şekil 35. (a) Sahte sese ait dalga formu
işlemi uygulanmış hali

(b) Sahte sese ait band geçiren süzgeç

Şekil 36.a ve Şekil 36.b sırasıyla yoğun eşleşme bölgelerinin belirlenmesi ve bu işlem sonucunda band geçiren süzgeç uygulanmış sesin yüksek çözünürlüklü spektrogramıdır.



Şekil 36. Sahte spektrogram üzerinde eşleşmeler (b). Band geçiren süzgeç işleminden sonra spektrogram görüntüsü

2.3.5. Hızlı Fourier Transform Uygulaması

Ses verisi üzerinde çalışırken, zaman domeninde çalışmak sesin uzunluğundan etkilenir. Örneğin, 1 saniyelik ses ile 10 saniyelik ses arasında, veri sayısı açısından önemli fark bulunmaktadır. Bu durum, farklı uzunluklardaki seslerin karşılaştırılmasını ve analiz edilmesini zorlaştırır. Ancak, frekans domenine geçildiğinde, sesin süresi ne olursa olsun veri ortak formata dönüştürülmüş olur. Bu nedenle, ses sinyallerinin analizinde ve işlenmesinde Hızlı Fourier Dönüşümü (FFT) kullanılarak frekans domenine geçmek yaygın yöntemdir. FFT uygulaması, zaman domenindeki sinyalleri frekans bileşenlerine ayırarak, süreye bağımlılığı ortadan kaldırır ve analizi daha tutarlı hale getirir. Şekil 37a ve Şekil 37b sırasıyla orijinal ve sahte seslerin frekans uzayını temsil etmektedir. Orijinal ses üzerinde band geçiren süzgeç işlemi uygulanmadığı için frekans bilgisinin her bölgeye dağıldığı gözlemlenmektedir. Ancak sahte ses verisi süzgeç işlemine alındığı için frekans uzayında belli kısımlar sıfıra çok yakın değere sahiptir.



Şekil 37. Sesin frekans uzayı (a) Band geçiren süzgeç işlemi uygulanmamış (b) Band geçiren süzgeç işlemi uygulanmış

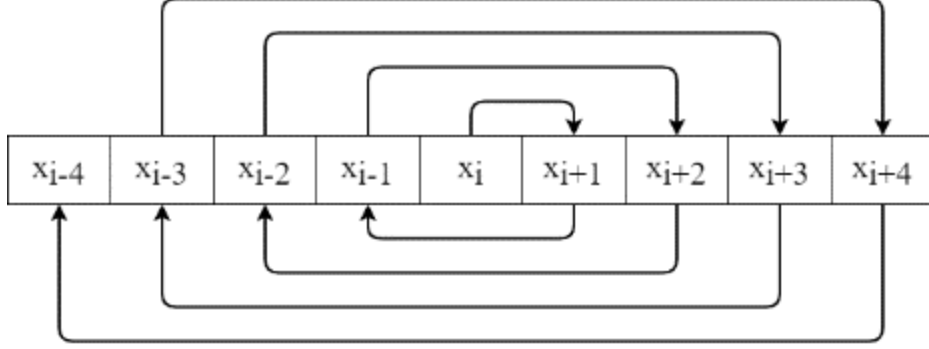
Sahte sesin frekans temsilinde sıfıra yakın değerler sıfıra eşitlenerek özellik çıkarma aşamasında kullanılan algoritmaya uygun hale getirilir. Bu aşama bir dizi olayların gerçekleştirilmesiyle mümkün olur. Sonuç itibariyle beklenen şartlar sağlanmaz ise değerlerin sıfıra eşitlenmesi gerçekleşmez.

- Öncelikle veri içerisindeki deneysel çalışmalarla belirlenen eşikten (T_1) küçük değerler sayılır. Bu sayım sonucunda değer elde edilir (S_1).
- Elde edilen toplam küçük sayılar, verinin eleman sayısına bölünerek oran (R) elde edilir.
- Alınan oranın (R) deneysel çalışmalarla belirlenen sayıdan (T_2) küçük olması durumunda sıfırlama işlemi gerçekleştirilir.
- Sıfırlama işlemi T_1 eşikinden küçük değerlerin sıfıra eşitlenmesi ile mümkün olmaktadır.

2.3.6. Frekans Uzayından Öznitelik Çıkarılması

Önişlemden geçirilmiş sesten graf görüntüsü çıkarmak için özellik çıkarma aşamasına geçilmektedir. Bu adımda frekans uzayına dönüştürülmüş ses dosyası boyutu 9 olacak şekilde birbiri ile örtüşmeyen çerçevelere bölünmektedir. Her çerçeveden bilgi üretimi spiral görüntü yöntemiyle uygulanmaktadır. Bu yöntemde Şekil 38’ de görüldüğü gibi merkez pikselin komşu pikselleri spiral mantığı ile gezilmekte ve göreceli büyüklük durumu değerlendirilmektedir. Büyüklük ilişkisi 1 ile temsil edilirken, eşitlik ve küçük olma durumları 0 ile kodlanmaktadır. Buna göre (X_i, X_{i+1}) , (X_{i+1}, X_{i-1}) , (X_{i-1}, X_{i+2}) , (X_{i+2}, X_{i-2}) ,

(X_{i-2}, X_{i+3}) , (X_{i+3}, X_{i-3}) , (X_{i-3}, X_{i+4}) ve (X_{i+4}, X_{i-4}) ikilileri arasındaki karşılaştırma durumlarına göre 8 bitlik değer üretilmektedir. Merkez örnek değeri üretilen 8 bitlik değer ile yer değiştirilerek özellik çıkarma aşaması gerçekleştirilmiş olmaktadır.



Şekil 38. Spiral örüntü yöntemiyle bilgi üretimi

Tüm veri için hesaplanan özellik V ile gösterilmektedir. $\{V\}=\{V_1, \dots, V_N\}$ Burada N , frekans uzayındaki çerçeve sayısıdır.

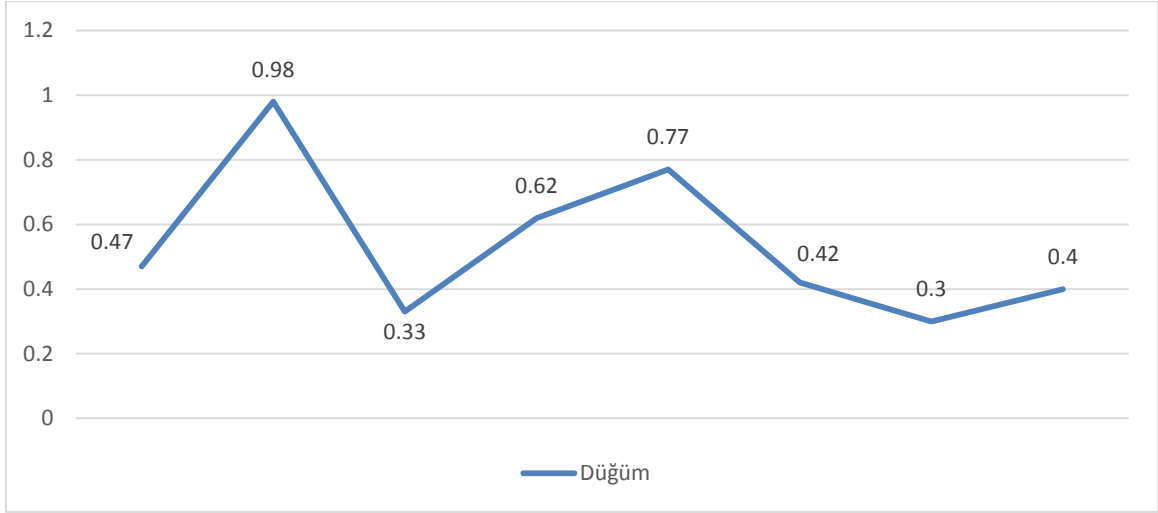
2.3.7. Graf Görüntüsünün Çıkarılması

Graf verisinde düğüm ve kenar olmak üzere 2 yapı vardır. Önerdiğimiz kopyala yapıştır ses sahteciliği yönteminde düğümlerimiz her çerçeveden elde edilen özneliliklerle temsil edilmektedir. Bu düğümleri birbirine bağlayan bağlantılar Eşitlik 62’de verilen algoritmayla atanmaktadır.

$$\frac{V_b - V_c}{b - c} > \frac{V_b - V_a}{b - a} \quad (62)$$

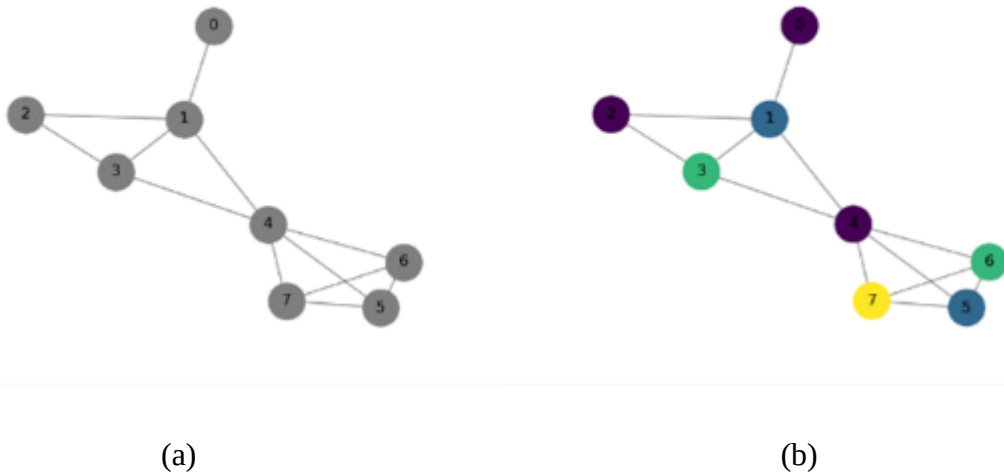
Graf içerisinde herhangi iki düğüm arasında kenar oluşturulabilmesi için, Eşitlik 62’de verilen denklemle belirtilen kriterin sağlanması gerekmektedir. Eğer V_a ile V_b arasındaki her (c, V_c) noktası bu kriteri karşılıyorsa, o zaman (a, V_a) ve (b, V_b) düğümleri arasında bağlantı kurulur. Kriterde (b, V_b) ve (c, V_c) noktalarından geçen doğrunun eğimi e_{bc} ile (a, V_a) ve (b, V_b) noktalarından geçen doğrunun eğimi e_{ba} arasında yapılan karşılaştırma, kenar bilgisinin oluşması hususunda katkı sağlayacaktır. İlgili koşul sağlandığında ($e_{bc} > e_{ba}$), V_a ile V_b arasında kenar oluşturularak graf güncellenmektedir. Aksi durumda her iki düğüm arasında

herhangi bağlantı oluşturulmayacaktır. Şekil 39’da verilen grafiğin graf görüntüsü Şekil 40’ta verilmektedir.



Şekil 39. Örnek bir öznitelik

Eğim denklemi gereğince ilk düğüm (0 numaralı düğüm) sadece ikinci düğümü görebilmektedir. Ancak ikinci düğüm (1 numaralı düğüm) üçüncü, dördüncü ve beşinci düğümü görebildiği için aralarında bağlantı kurulabilmiştir. Bu olay birbiri ardınca sıralanmış dağlara benzetilebilir. Bir dağın tepesinden baktığınızda önünüzde küçük dağlar var ise onları görebilirsiniz, lakin büyük dağ çıktığında sadece onu görür arkasındakileri göremezsiniz. Çıkarılan graf görüntüsü Welsh-Powell algoritması kullanılarak renklendirilmektedir. Verilen örneğe ait renklendirilmiş graf görüntüsü Şekil 40’de verilmektedir.



Şekil 40. Örnek öznitelğe ait graf görüntüsü (a) Renksiz (b) Renklendirilmiş

Şekil 41’de birbirine komşu olan düğümlerin farklı renklendirildiği graf görüntüsü yer almaktadır. Bu graf görüntüsü, band geçiren süzgeç işlemi uygulanmamış orijinal ses ile filtrenin uygulandığı sahte sese ait graf görüntülerini içermektedir. Orijinal konuşma kaydına ait graf resminde, tüm düğümler çevreye yayılmakta ve geniş dağılım göstermektedir. Buna karşın, sahte graf görüntüsünde birçok düğümün boğum oluşturduğu gözlemlenmektedir. Band geçiren filtre işlemi, frekans uzayında verileri sıfıra yakınlaştırmakta ve bu durum graf görüntüsünde düğümlerin boğumlanması şeklinde yansımaktadır. Bu olay, eğim denklemi sayesinde sıfır değerine sahip düğümler ile diğerleri arasında bağlantı kurulabilmesi sonucu ortaya çıkmaktadır. Sonuç olarak, birçok bağlantıya sahip düğüm, görüntü üzerinde boğum görünümü yaratmaktadır.



Şekil 41. Graf görüntüsü (a) Orijinal ses (b) Sahte ses

2.3.8. Graf Görüntüsünün Sınıflandırılarak Kopyala Yapıştır Sahte Ses Tespiti Yapılması

Yapay zeka ile görüntü sınıflandırma, günümüzde birçok alanda önemli uygulamalara sahip teknolojidir. Bu süreç, görüntülerin belirli kategorilere otomatik olarak ayrılmasını sağlar ve derin öğrenme yöntemleri ile büyük ölçüde geliştirilmiştir. Özellikle, konvolüsyonel sinir ağları kullanılarak, yüksek doğrulukla nesne tanıma, yüz tanıma ve tıbbi görüntü analizi gibi görevler gerçekleştirilmektedir. CNN'ler, görüntüdeki pikseller arasındaki ilişkileri öğrenerek, karmaşık desenleri ve özellikleri tanımlayabilme yeteneğine sahiptir. Bu sayede, sağlık, güvenlik, ulaşım ve eğlence gibi birçok sektörde yenilikçi çözümler sunulmakta, verimlilik ve doğruluk artırılmaktadır. Yapay zeka destekli görüntü

sınıflandırma, hem akademik araştırmalarda hem de endüstriyel uygulamalarda giderek daha fazla önem kazanmaktadır.

Yüksek çözünürlüklü spektrogramlardan çıkarılan anahtar noktaların kümelenilmesi ve yoğun eşleşme bölgelerinin bulunması ile süzgeç işleminin uygulanması, konuşma kaydından elde edilen graf görüntüleri üzerinde ayırt edici etkiye sahiptir. Bu etkinin yapay sinir ağları tarafından yakalanıp sınıflandırılması, önerilen modellerle gerçekleştirilmektedir.

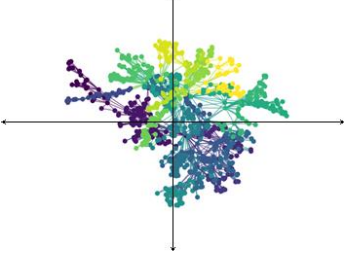
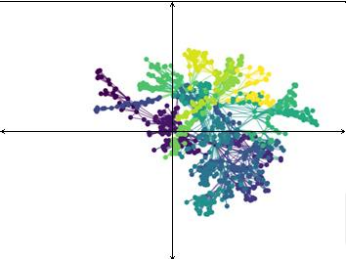
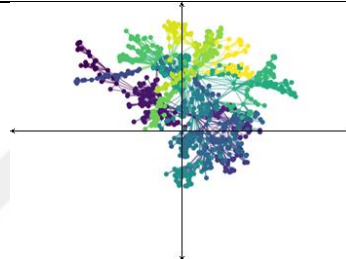
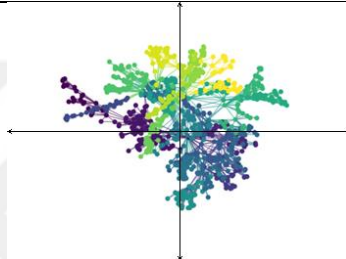
2.3.8.1. Veriseti Hazırlanması

Veri arttırma (Data Augmentation) makine öğrenimi ve özellikle derin öğrenme modellerinin performansını artırmak amacıyla kullanılan tekniktir. Bu yöntem, mevcut veri setini çeşitli dönüşümler uygulayarak yapay olarak genişletir. Özellikle görüntü işleme, ses tanıma ve doğal dil işleme gibi alanlarda yaygın olarak kullanılır.

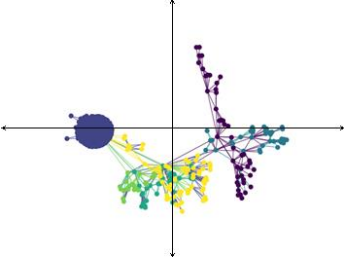
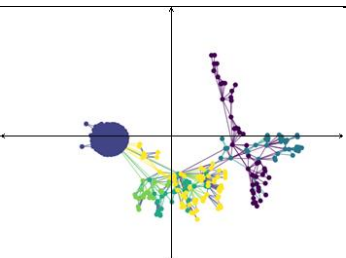
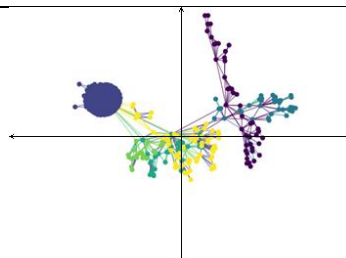
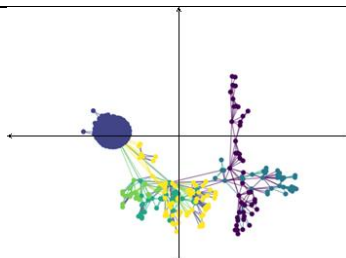
Bu işlemin başlıca amacı, modelin genelleme yeteneğini artırmak ve overfitting (aşırı uyum) sorununu azaltmaktır. Overfitting, modelin eğitim verilerine çok iyi uyum sağlarken, yeni ve görülmemiş verilere karşı performansının düşmesi durumudur. Bu işlem ile eğitim verisinin çeşitliliği artırılarak modelin farklı durumlara karşı daha dayanıklı ve esnek olması sağlanır.

Görüntü işleme alanında yaygın olarak kullanılan veri arttırma teknikleri arasında döndürme, ölçekleme, çevirme, parlaklık ve kontrast ayarlamaları, gürültü ekleme ve kırpma gibi işlemler bulunur. Sonuç olarak veri setini yapay olarak büyüterek modelin daha geniş veri yelpazesi üzerinde eğitilmesini sağlar ve böylece modelin performansını ve genelleme kabiliyetini artırır. Yatay öteleme, dikey öteleme ve döndürme veri artırma yöntemleri çeşitli parametrelerle uygulanarak yapay şekilde daha fazla miktarda eğitim verisi oluşturulmuştur. Tablo 1 ve Tablo 2 örnek graf görüntüleri üzerinde tez çalışması kapsamında uygulanan veri artırım yöntemlerini göstermektedir.

Tablo 1. Orijinal graf görüntülerine uygulanan veri artırım yöntemleri

	Orijinal graf görüntüsü	
		
Yatay Öteleme:0.30	Dikey Öteleme:0.20	Döndürme:30
		

Tablo 2. Sahte graf görüntülerine uygulanan veri artırım yöntemleri

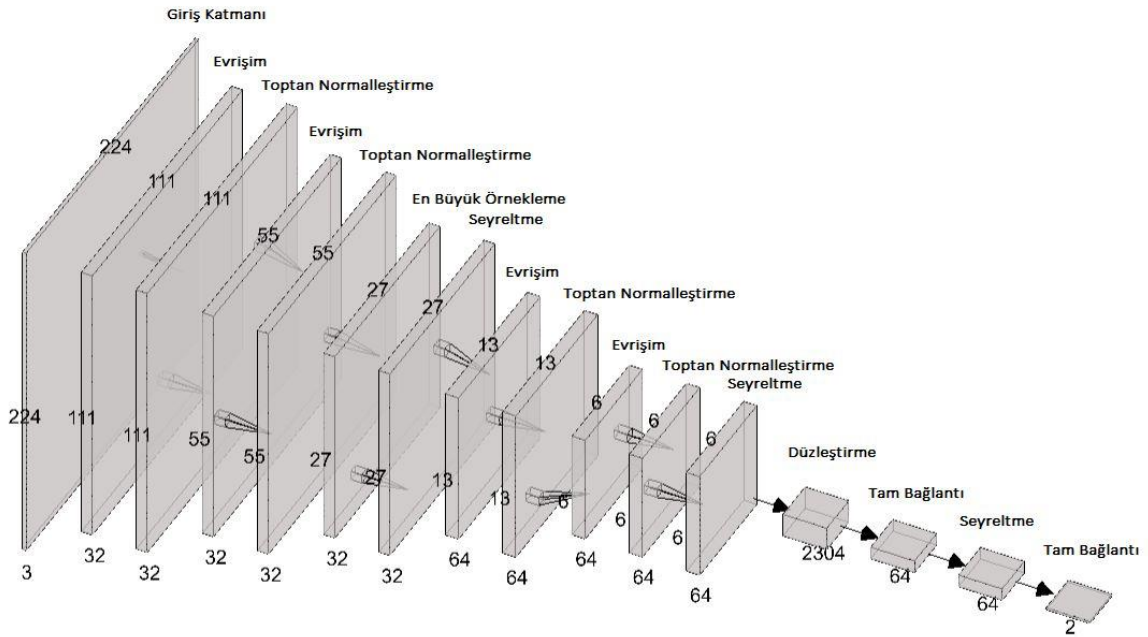
	Sahte graf görüntüsü	
		
Yatay Öteleme:0.30	Dikey Öteleme:0.20	Döndürme:30
		

Tez kapsamında gerçekleştirilen çalışmada, eğitim setine veri artırım yöntemi uygulanmıştır. Analiz amacıyla, veri artırımı yapılmış ve yapılmamış olmak üzere iki farklı eğitim seti kullanılmıştır.

2.3.8.2. Önerilen CNN Modeliyle Sınıflandırma

CNN, Evrişimsel Sinir Ağlarının (Convolutional Neural Network) kısaltmasıdır. Derin öğrenme alanında kullanılan yapay sinir ağı türüdür. Görüntü tanıma, ses tanıma, doğal dil işleme gibi alanlarda yaygın olarak kullanılır. CNN'ler, verinin özelliklerini otomatik olarak öğrenmek için girdi verisini işlemek için özel mimariye sahiptir. Özellikle bilgisayarlı görü sistemlerindeki başarılarıyla dikkat çekerler. Görsel veri gibi yapısal verileri işlerken, konvolüsyon (evrişim) işlemi ve havuzlama (pooling) işlemi gibi özellik çıkarma tekniklerini kullanırlar. Bu sayede, örüntülerin hiyerarşik yapılarını keşfedebilirler ve veriyi sınıflandırmak veya tanımak için kullanılabilirler.

Sahte ve orijinal graf görüntülerinin sınıflandırılması amacıyla yeni mimari önerilmiştir. Önerilen model Şekil 42’de verilmektedir. Mimariye giriş olarak graf görüntüsü verilmektedir.



Şekil 42. Önerilen mimari

Katman 1: Giriş katmanında, 224x224 boyutunda görüntüler için 3x3 boyutunda 32 çekirdek kullanılmıştır. Aktivasyon fonksiyonu olarak ReLU tercih edilmiş ve ardından Toplu Normalleştirme (Batch Normalization) uygulanmıştır.

Katman 2: Bu katmanda, giriş katmanına benzer şekilde 3x3 boyutunda 32 çekirdek ve ReLU aktivasyon fonksiyonu kullanılmıştır. Daha sonra Toplu Normalleştirme ve 2x2 adımlı maksimum havuzlama (max-pooling) gerçekleştirilmiş, böylece özellik haritasının boyutu azaltılmış ve %50 Seyreltme (Dropout) uygulanmıştır.

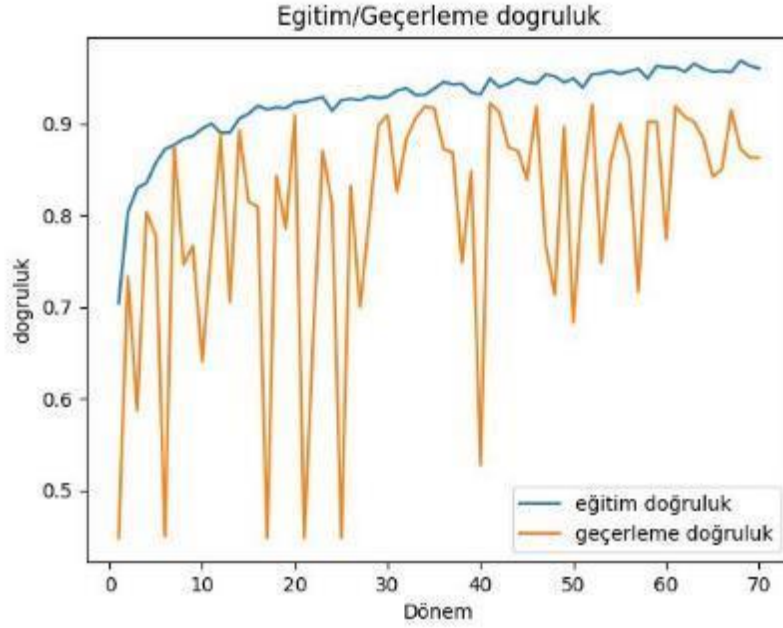
Katman 3: 3x3 boyutunda 64 çekirdek kullanılmış, aktivasyon fonksiyonu olarak yine ReLU tercih edilmiş ve Toplu Normalleştirme eklenmiştir.

Katman 4: 3x3 boyutunda 64 çekirdek kullanılmış, aktivasyon fonksiyonu olarak yine ReLU tercih edilmiş, Toplu Normalleştirme ve %50 Seyreltme eklenmiştir.

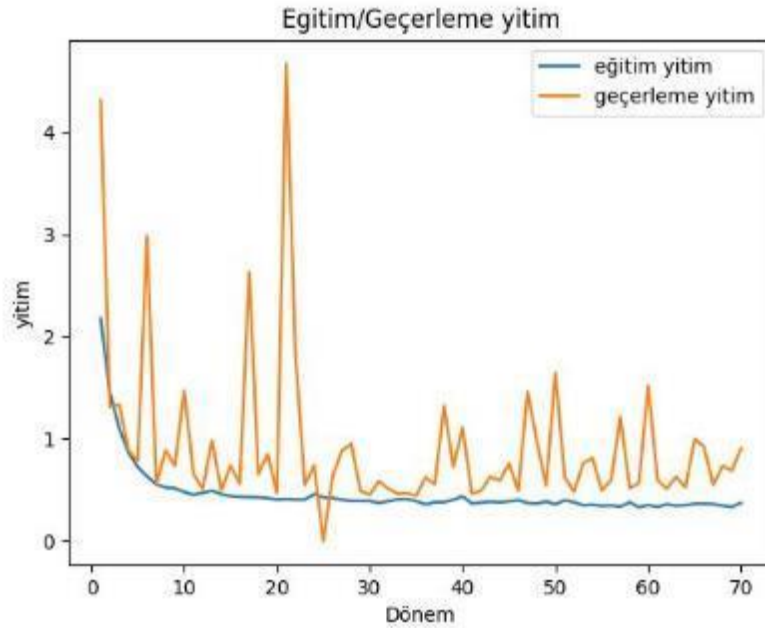
Katman 5: Bu katmanda, özellik haritası tek boyuta indirgenmiştir. 64 birimli Dense katmanı ve ReLU aktivasyon fonksiyonu kullanılarak Tam Bağlantılı (Fully Connected) katmanı oluşturulmuş ve %50 Seyreltme uygulanmıştır.

Katman 6: Önceki katmandan gelen veriler Softmax aktivasyon fonksiyonu kullanılarak 2 sınıfa ayrılmıştır.

Öğrenme oranı sabit olarak 0.0007 olarak belirlenmiş ve eğitim süresince değişmemiştir. Optimizasyon için Adam Optimizer tercih edilmiş, ikili çapraz entropi (binary cross entropy) kayıp (loss) fonksiyonu kullanılmıştır. Seyreltme (Dropout) oranı %50 olarak ayarlanmış ve eğitim verileri 8'lik yığınlar (batchler) halinde işlenmiştir. Şekil 43 ve 44'de artırısız eğitim seti ile yapılmış eğitimin eğrileri verilmektedir. Grafiklerde mavi çizgi eğitim eğrisini, sarı çizgi ise geçerleme grafiğini temsil etmektedir. Şekil 43'de eğitim doğruluğunun sürekli arttığı gözlemlenmekte iken geçerleme doğruluğu titreşimli yapıya sahiptir. Şekil 44'te üzerinde eğitim yitiminin sürekli azalma yöneliminde olduğu gözlemlenmekte iken, geçerleme yitimi titreşimli görünüme sahiptir. Yapılan eğitimlerde eğrilerin büyük genlikli titreşimlere sahip olmaması beklenmektedir.



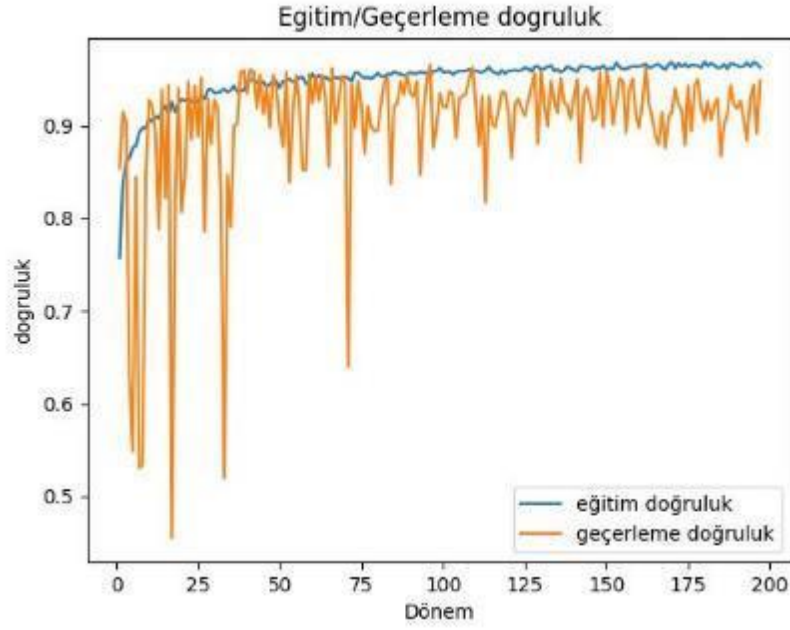
Şekil 43. Eğitim ve geçerleme doğruluk grafikleri



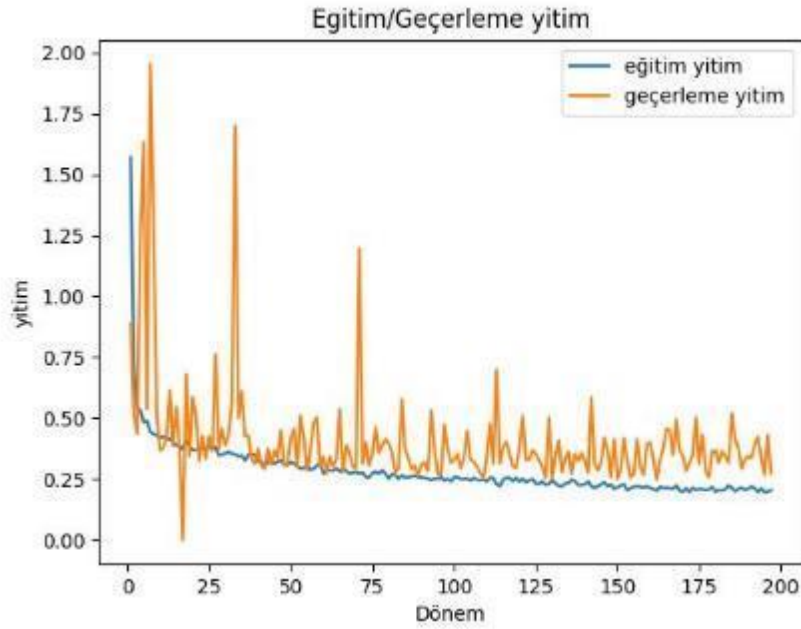
Şekil 44. Eğitim ve geçerleme yitim grafikleri

Artırımlı veri ile yapılan eğitimin eğrileri Şekil 45 ve 46 da gösterilmektedir. Her iki grafiğin incelenmesi sonucunda artırımlı veri ile yapılan eğitimde eğriler üzerinde titreşimin azaldığı gözlemlenmektedir. Titreşimsiz eğitim eğrisi, modelin eğitim sırasında dalgalanmalara veya aşırı uyuma (overfitting) yatkın olmadığını gösterir. Yani, modelin hem

eğitim veri setine hem de doğrulama veri setine iyi genelleştirdiği ve yeni veriler üzerinde de iyi performans gösterebileceği anlamına gelir.



Şekil 45. Eğitim ve geçerleme doğruluk grafikleri



Şekil 46. Eğitim ve geçerleme yitim grafikleri

2.3.8.3. Transfer Öğrenme Modelleri

Transfer öğrenme, makine öğrenimi yöntemi olarak, bir görevde elde edilen bilgi veya yeteneklerin, başka görevde veya bağlamda kullanılmasıdır. Bu yöntem, genellikle bir görevde yeterince veri veya kaynak olmadığında veya yeni görev için sıfırdan model eğitmek maliyetli veya zor olduğunda kullanılır. Temel fikir, bir görevde veya alanında öğrenilen bilgilerin, başka görevde veya alanında da işe yarayabileceğidir. Transfer öğrenme, öğrenme hızını artırabilir, modelin performansını iyileştirebilir ve genellikle daha verimli şekilde yeni görevlere adapte olmasını sağlayabilir. Bu nedenle, çeşitli makine öğrenimi ve derin öğrenme uygulamalarında yaygın olarak kullanılmaktadır. Tez çalışması kapsamında VGG16 ve VGG19 mimarisi uygulanmaktadır.

2.3.8.3.1. VGG16 Kullanımı

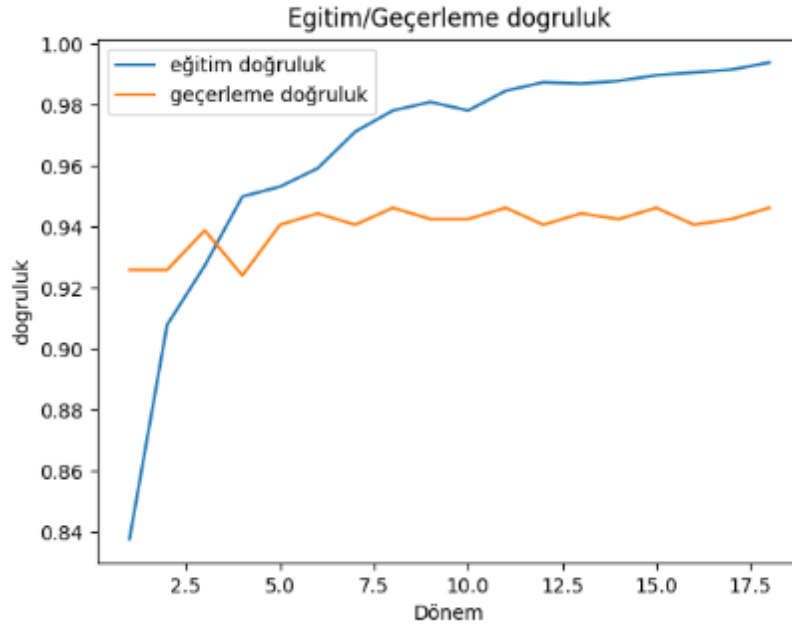
Bu çalışmada, görüntü sınıflandırma problemlerinde yaygın olarak kullanılan derin öğrenme modeli olan VGG16'nın kullanımı incelenmiştir. VGG16 modeli, 2014 yılında Visual Geometry Group (VGG) tarafından geliştirilmiş ve ImageNet veri kümesi üzerinde etkileyici sonuçlar elde etmiştir (Simonyan ve Zisserman, 2014). Bu model, derin mimariye sahiptir. Tez kapsamında, VGG16 transfer öğrenme modeli kullanılarak çeşitli görüntü sınıflandırma görevleri üzerinde deneyler yapılmış ve modelin performansı değerlendirilmiştir. VGG16 mimarisinin evrişim katmanları ImageNet kullanılarak elde edilen ağırlıklarla başlatılmıştır. Tam bağlantı katmanları ise 128, 64 ve 2 nöronlu olarak değiştirilmiştir. İlk tam bağlantı katmanından sonra 0.5 oranlı seyreltme katmanı eklenmiştir.



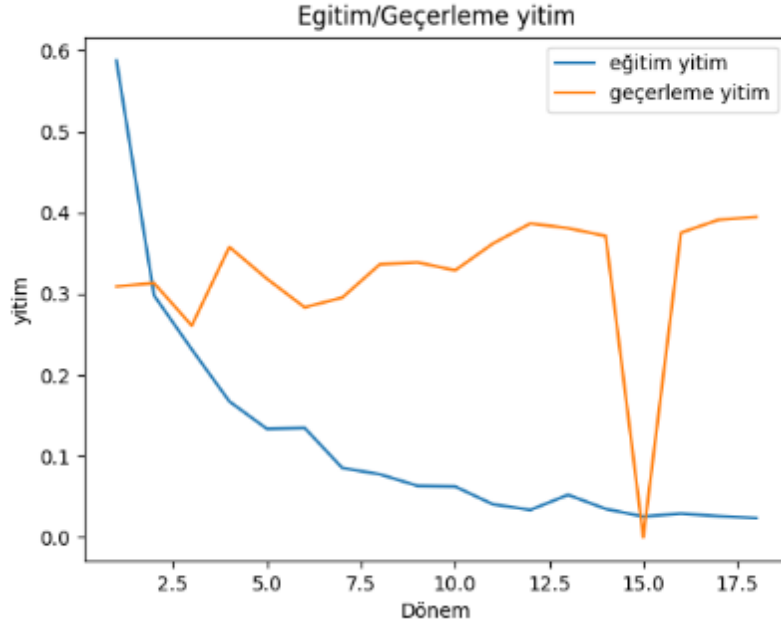
Şekil 47. Önerilen VGG16 Modeli

Öğrenme oranı sabit olarak 0.0007 olarak belirlenmiş ve eğitim süresince değişmemiştir. Optimizasyon için Adam Optimizer tercih edilmiş, ikili çapraz entropi (binary cross entropy) kayıp (loss) fonksiyonu kullanılmıştır. Bırakma (Dropout) oranı %50 olarak ayarlanmış ve eğitim verileri 8'lik yığınlar (batchler) halinde işlenmiştir. Toplamda 200 dönem (epoch) eğitim yapılması ayarlanmıştır. Ancak eğitim sırasında geçerleme kaybı incelenmiş, 15 dönem boyunca değişim olmadığında eğitim durdurulmuştur.

Şekil 48 ve 49'da artırısız eğitim seti ile yapılmış eğitimin eğrileri verilmektedir. Eğitim doğruluğunun sürekli artış içerisinde olduğu görülmektedir. Buna karşın geçerleme doğruluğu 0.9 üzerinde ve küçük genlikli dalgalanmalara sahiptir. Şekil 49 üzerinde eğitim yitiminin sürekli azalma yöneliminde olduğu gözlemlenmekte, geçerleme yitimi 0.3 civarında seyrederken ani sıçrama yaşamıştır.

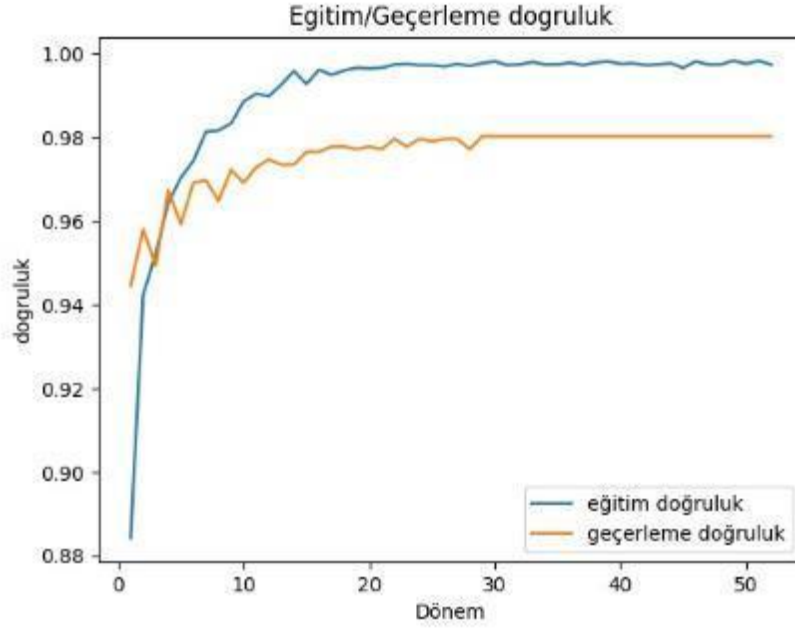


Şekil 48. Eğitim ve geçerleme doğruluk grafikleri

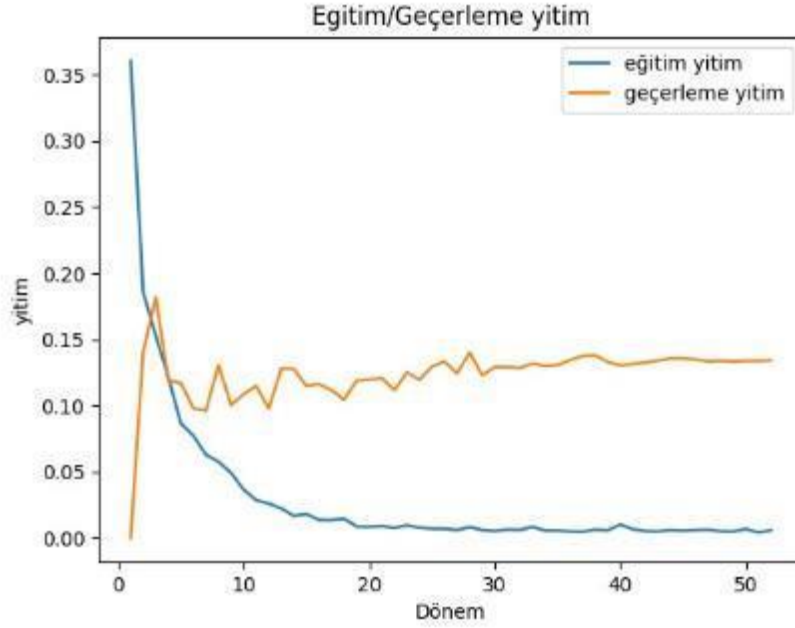


Şekil 49. Eğitim ve geçerleme yitim grafikleri

Artırımsız eğitime ek olarak, eğitim setine veri artırımı uygulandıktan sonra VGG16 modelinin eğitimi gerçekleştirilmiştir. Modelin eğitimi süresince elde edilen eğriler Şekil 50 ve Şekil 51’de gösterilmektedir. Bu eğriler, veri artırımı uygulanmış ve uygulanmamış durumlar arasındaki performans farkını net bir şekilde ortaya koymaktadır. Eğitim sürecinde kaydedilen bu grafikler, modelin doğruluğu ve kaybı üzerindeki etkileri karşılaştırmalı olarak değerlendirmemize olanak tanımaktadır.



Şekil 50. Eğitim ve geçerleme doğruluk grafikleri



Şekil 51. Eğitim ve geçerleme yitim grafikleri

Artırımlı veri ile yapılan eğitim sonucunda doğruluk eğrilerinde gözle görülür iyileşme gözlenmemiştir. Ancak, yitim eğrisindeki dalgalanma genliklerinin daha küçük aralıklarda olduğu tespit edilmiştir. Bu durum, veri artırmanın modelin kararlılığı üzerinde olumlu etki

yarattığını ve bu etkinin yitim eğrisindeki azalan dalgalanma genlikleri aracılığıyla gözlemlenebildiğini göstermektedir.

2.3.8.3.2. VGG19 Kullanımı

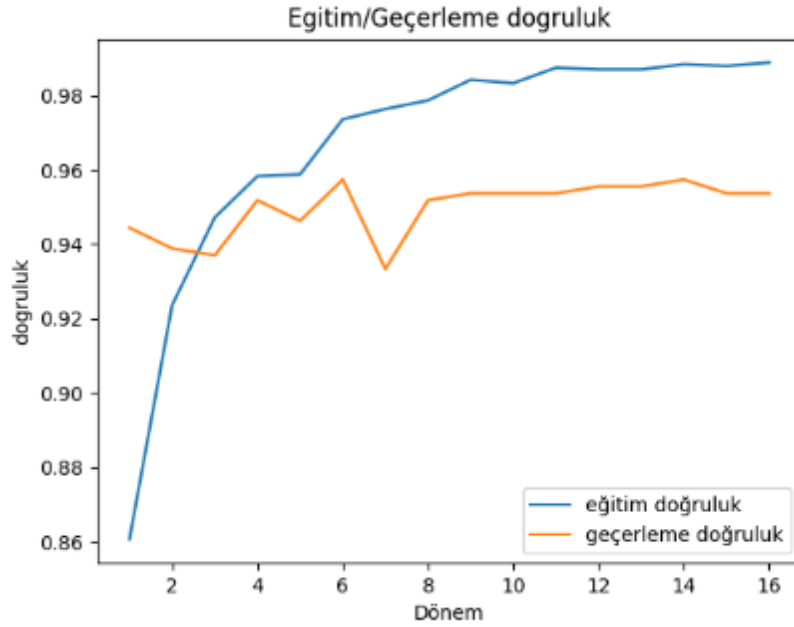
VGG19 modeli, derin öğrenme alanında yaygın olarak kullanılan konvolüsyonel sinir ağıdır (Simonyan ve Zisserman, 2014). Çok katmanlı derin yapısı ile özellikle görüntü sınıflandırma ve nesne tanıma gibi görevlerde yüksek performans sergilemektedir. Bu model, büyük ölçekli veri kümeleri üzerinde önceden eğitilmiş ağırlıklar kullanarak transfer öğrenme için idealdir ve bu sayede daha küçük veri setlerinde bile etkili sonuçlar elde edilebilmektedir. VGG19'un katman derinliği ve yapısal özellikleri, karmaşık veri temsillerini öğrenme kapasitesini artırmakta ve bu da onu birçok bilgisayarla görme uygulaması için tercih edilen model haline getirmektedir. Bu tez kapsamında, VGG19 modelinin kullanımıyla gerçekleştirilen deneyler ve elde edilen sonuçlar detaylı şekilde incelenmektedir. VGG19 mimarisinin evrişim katmanları, ImageNet kullanılarak elde edilen önceden eğitilmiş ağırlıklarla başlatılmıştır. Modelin tam bağlantı katmanları ise 128, 64 ve 2 nöronlu olarak yeniden yapılandırılmıştır. Ayrıca, ilk tam bağlantı katmanından sonra, modelin aşırı öğrenmesini engellemek amacıyla 0.5 oranında seyreltme (dropout) katmanı eklenmiştir. Bu yapılandırma, modelin genel performansını artırmayı ve daha iyi genelleştirme yapmasını sağlamayı hedeflemektedir. Kullanılan VGG19 yapısı Şekil 52'de verilmektedir.



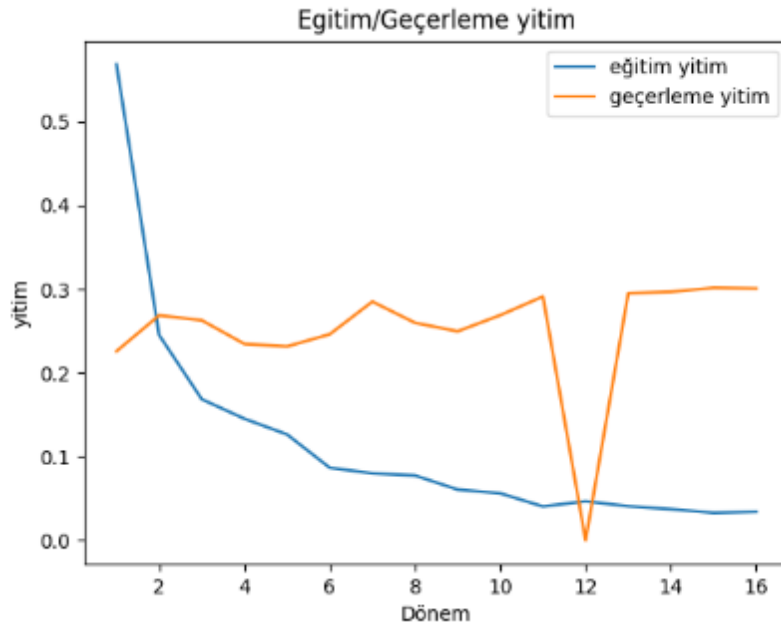
Şekil 52. Önerilen VGG19 modeli

VGG19 eğitim parametreleri VGG16 ile aynı şekilde yapılmaktadır. Bu durum iki mimari arasındaki farkın gözlemlenmesi için gereklidir. VGG19 mimarisi VGG16 yapısına göre daha derin evrişim katmanlarına sahip olması nedeniyle özellik çıkarma aşamasında

daha başarılıdır. Şekil 53 ve Şekil 54’de artırısız veri ile yapılan eğitim sonucunda elde edilen eğrileri göstermektedir. Eğitim doğruluğu sürekli artan formda olmasına karşın geçerleme doğruluğu 0.93 ile 0.95 arasında seyretmekte ve ani dalgalanma eğilimindedir.



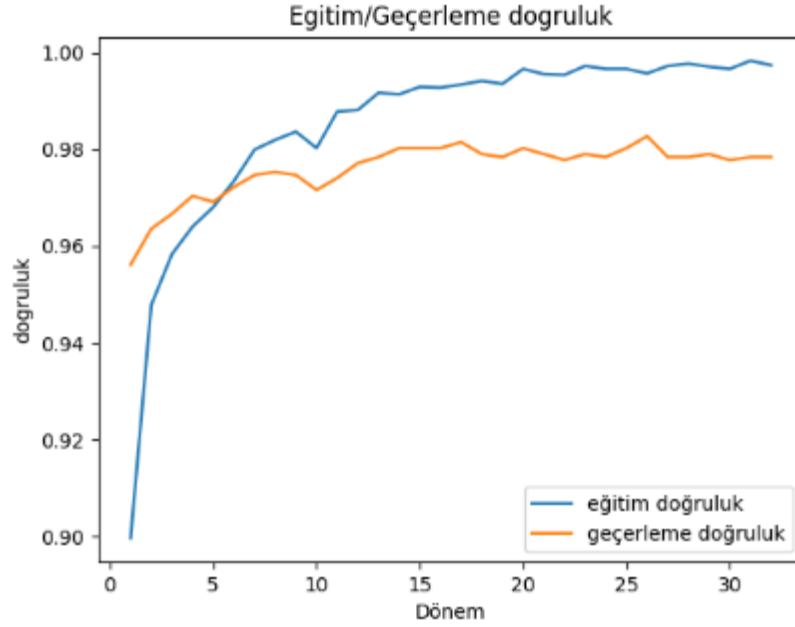
Şekil 53. Eğitim ve geçerleme doğruluk grafikleri



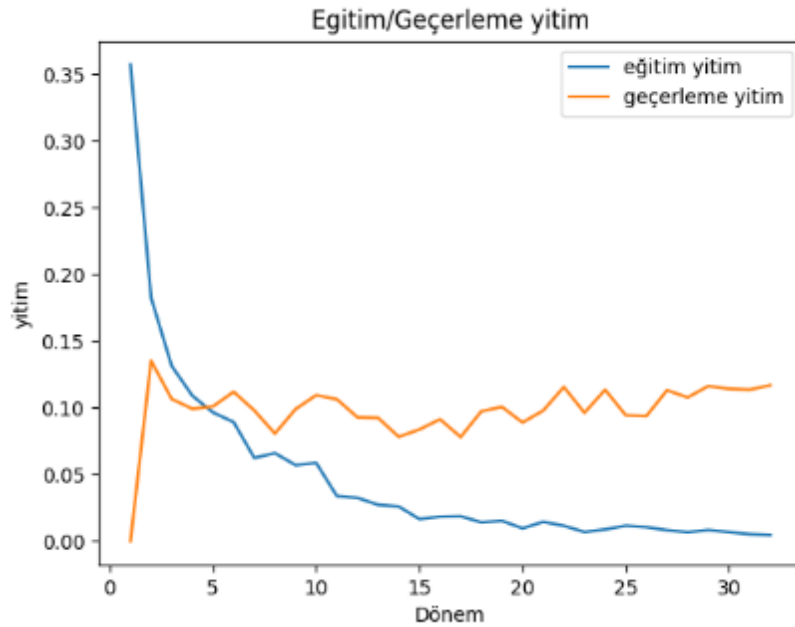
Şekil 54. Eğitim ve geçerleme yitim grafikleri

Artırısız veri ile yapılan VGG19 eğitim ve geçerleme eğrileri, VGG16 ile elde edilen çizelgelerle benzerlik göstermektedir. Bu durum, evrişim katmanlarının sayısının artmasının

eğriler üzerinde belirgin etkisi olmadığını ortaya koymaktadır. Son olarak, veri artırımı uygulanmış veri seti kullanılarak VGG19 mimarisi eğitilmiştir. Bu eğitimin sonuçlarına ilişkin grafikler Şekil 55 ve Şekil 56'da sunulmaktadır.



Şekil 55. Eğitim ve geçerleme doğruluk grafikleri



Şekil 56. Eğitim ve geçerleme yitim grafikleri

Artırılmış veri ile yapılan VGG19 mimarisine ait doğruluk ve yitim çizelgeleri, daha durağan ve titreşimsiz görünüm sergilemektedir. Bu durum, veri artırma yöntemlerinin ve

özellik çıkarma aşamasında kullanılan evrişim katmanlarının fazlalığının eğitime pozitif etkisi olduğunu göstermektedir. Eğitim sürecinde elde edilen bu daha stabil grafikler, modelin genel performansını ve kararlılığını artırmada önemli rol oynamaktadır.

2.3.8.4. Füzyon Uygulanması

Tez kapsamında geliştirilen CNN ve transfer öğrenme kullanılarak, VGG16 ve VGG19 mimarileriyle graf görüntüleri iki sınıfa ayrılarak sınıflandırılmaktadır. Kullanılan üç farklı mimariden elde edilen sonuçlar, önerilen füzyon denklemleri ile birleştirilerek, ses sahte veya orijinal olarak etiketlenmektedir. Bu yaklaşım, farklı mimarilerin güçlü yönlerini bir araya getirerek, sınıflandırma performansını artırmayı hedeflemektedir.

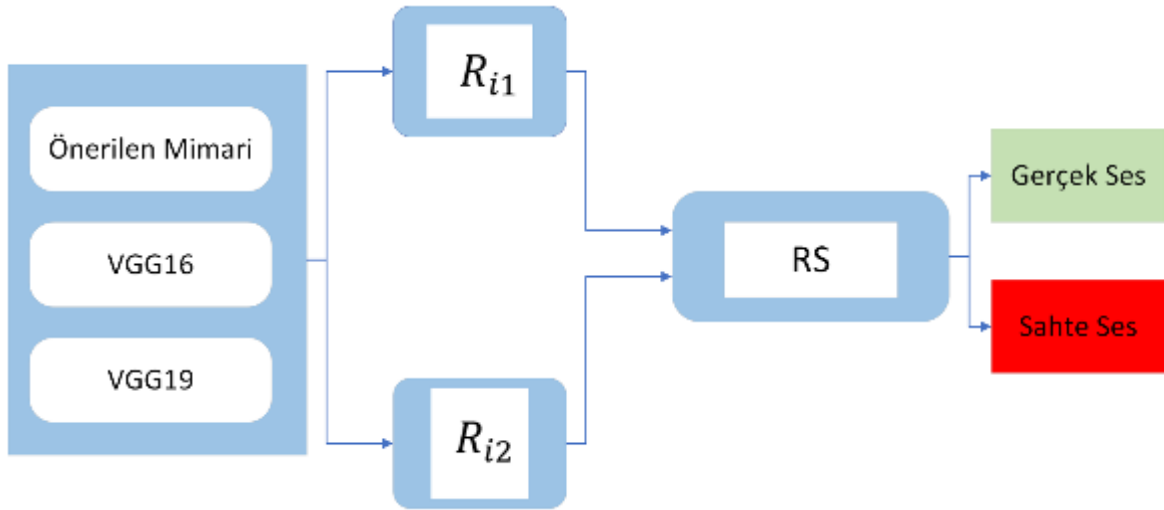
Füzyon denklemi 2 bölüme ayrılmaktadır. Birinci aşamada mimarilerin çıkış katmanı sayısınca rank denklemi mevcuttur. Eşitlik 63'te verilen R_{i1} ; her mimaride birinci nöronun skoru ile hesaplanmaktadır. Eşitlik 64'te gösterilen R_{i2} ise her mimaride ikinci nöronun skoru ile işlem görmektedir. Eşitlik 65'te ifade edilen son denklemde 3 farklı mimari içerisinde 1. ve 2. olasılık skorları çarpılıp toplanır ve füzyon skoru elde edilir.

$$R_{i1} = 1 - e^{-\left(\frac{P_i-1}{2}\right)^2} \quad (63)$$

$$R_{i2} = 1 - \tanh\left(\frac{P_i-1}{2}\right) \quad (64)$$

$$RS = \sum_i R_{i1} * R_{i2} \quad (65)$$

Burada i mimariyi temsil etmektedir ($i = 1,2,3$), P_i i. mimarinin skorunu vermekte, R ise rank skoru ifade etmekte, RS füzyon skorunu göstermektedir. Füzyon işlemi Şekil 57'de verilmektedir.



Şekil 57. Füzyon grafiği

Tez kapsamında füzyon işlemi artırılmış veri ile yapılan eğitimden elde edilen sonuçlara ve artırılmış veri ile alınan sonuçlar olmak üzere 2 farklı durumda uygulanmıştır.

3. BULGULAR VE TARTIŞMA

Bu bölümde, tez kapsamında yapılan çalışmaların veri setleri üzerindeki sonuçları incelenecektir. İlk olarak, anahtar noktalarının kümelenendirilmesi yöntemi ile kopyala yapıştır ses sahteciliği tespitine yönelik analizler ele alınacaktır. Son olarak, graf tabanlı yöntemin sonuçları değerlendirilecektir. Bu analizler, tezde önerilen yöntemlerin etkinliğini ve doğruluğunu ortaya koymayı amaçlamaktadır.

3.1. Değerlendirme Metrikleri

Tez çalışması kapsamında önerilen yöntemin başarısını değerlendirebilmek amacıyla literatürde sıkça kullanılan değerlendirme metriklerinden faydalanılmıştır. Kesinlik, duyarlılık, doğruluk ve F-skor metrikleri sırasıyla Eşitlik 66, 67, 68 ve 69'da belirtilmiştir.

$$Kesinlik = \frac{DP}{DP+YP} \quad (66)$$

$$Duyarlılık = \frac{DP}{DP+YN} \quad (67)$$

$$Doğruluk = \frac{DP+DN}{DP+DN+YN+YP} \quad (68)$$

$$F - Skor = 2 * \frac{Kesinlik * Duyarluluk}{Kesinlik + Duyarluluk} \quad (69)$$

DP sahte olarak doğru şekilde tespit edilen seslerin sayısı, DN orijinal olarak doğru tespit edilen seslerin sayısı, YP sahte olarak belirlenen orijinal seslerin sayısı, YN orijinal olarak tespit edilen sahte seslerin sayısıdır.

3.2. Anahtar Noktalarının Kümelenirilmesi Yaklaşımı Deneyisel Sonuçları

Bu bölümde Yüksek Çözünürlüklü Spektrogram görüntülerinden elde edilen anahtar noktalarının kümelenirilerek kopyala yapıştır ses sahteciliği yapan çalışmanın verisetleri üzerinde başarısı incelenmektedir. Öncelikle kullanılan verisetleri detaylı biçimde açıklanmaktadır.

3.2.1. Kullanılan Verisetleri

Bu yöntemde kullanılan verisetleri TIMIT veriseti ve Arabic Konuşma Külliyesi kullanılarak oluşturulmuş sahte ve orijinal seslerden oluşmaktadır.

3.2.1.1. TIMIT Kopyala-Yapıştır Veriseti

TIMIT veri seti, konuşma tanıma sistemlerinin geliştirilmesi ve değerlendirilmesi için yaygın olarak kullanılan veri setidir (Garofolo, Lamel, Fisher, Fiscus, & Pallett, 1993). Bu veri seti, farklı bölgesel aksanlara sahip konuşmacılar tarafından söylenen cümlelerin kayıtlarını içermektedir. Tez çalışması kapsamında Üstübioğlu ve arkadaşlarının TIMIT verisetinden oluşturduğu kopyala yapıştır sahte ses veriseti kullanılmaktadır (Üstübioğlu vd., 2022). Bu verisetinde TIMIT'ten 126 orijinal konuşma kaydı seçilmiş ve bu kayıtlar kullanılarak 368 adet kopyala yapıştır sahte ses oluşturulmuştur. Elde edilen sahte seslere 20dB ve 30dB gürültü ekleme, 32kbps ve 64kbps sıkıştırma ve medyan filtre atağı uygulanmıştır. TIMIT kopyala yapıştır sahte ses veriseti Tablo 3'de ayrıntılı olarak sunulmaktadır.

Tablo 3. Anahtar Noktaların Kümelenendirilmesi Yaklaşımı için oluşturulmuş TIMIT kopyala yapıştır veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	126	126
Sahte	Ataksız	368	2208
	Medyan Filtre	368	
	20dB gürültü	368	
	30dB gürültü	368	
	32kbps sıkıştırma	368	
	64kbps sıkıştırma	368	

Toplamda 2334 konuşma kaydının 2208 tanesi sahte ses, 126 tanesi ise orijinal sestir. Veri setinde, ataksız sahte seslerin yanı sıra, 5 farklı atak uygulanmış sahte sesler bulunmaktadır. Gerçek olmayan 368 ataksız kayda, 5 farklı atak türü uygulanarak sahte sesler elde edilmiştir. Bu çeşitlendirilmiş sahte ses kayıtları, ses sahteciliği tespiti çalışmalarında kapsamlı değerlendirme yapılabilmesi için hazırlanmıştır.

3.2.1.2. Arabic Kopyala Yapıştır Ses Veriseti

Arabic Speech Corpus veri seti, konuşma tanıma sistemlerinin geliştirilmesi ve değerlendirilmesi için kullanılan kapsamlı veri setidir (Alsulaiman, Muhammad, Abdelkader, Mahmood, & Ali, 2014). Bu veri seti, farklı Arap ülkelerinden ve çeşitli lehçelere sahip konuşmacılar tarafından seslendirilen cümlelerin yüksek kaliteli kayıtlarını içermektedir. Arabic Speech Corpus, konuşma tanıma, dil modelleme ve akustik analiz gibi alanlarda araştırma ve geliştirme faaliyetlerine önemli kaynak sağlamaktadır. Tez çalışması kapsamında Üstübioğlu ve arkadaşlarının Arabic Speech Corpus verisetinden oluşturduğu kopyala yapıştır sahte ses veriseti kullanılmaktadır (Üstübioğlu vd., 2023). Arabic Speech Corpus veri setinden 36 orijinal ses kaydı alınmış ve bu kayıtlarla 30 adet kopyala-yapıştır sahte ses oluşturulmuştur. Elde edilen sahte seslere dört farklı atak türü uygulanmıştır: medyan filtre, 20dB gürültü ekleme, 30dB gürültü ekleme ve 64kbps sıkıştırma. Tüm sahte seslere bu ataklar uygulanarak toplamda 186 ses kaydı elde edilmiştir. Arabic Speech Corpus veri setinden oluşturulan bu sahte seslerin detaylı açıklaması Tablo 4'te sunulmaktadır.

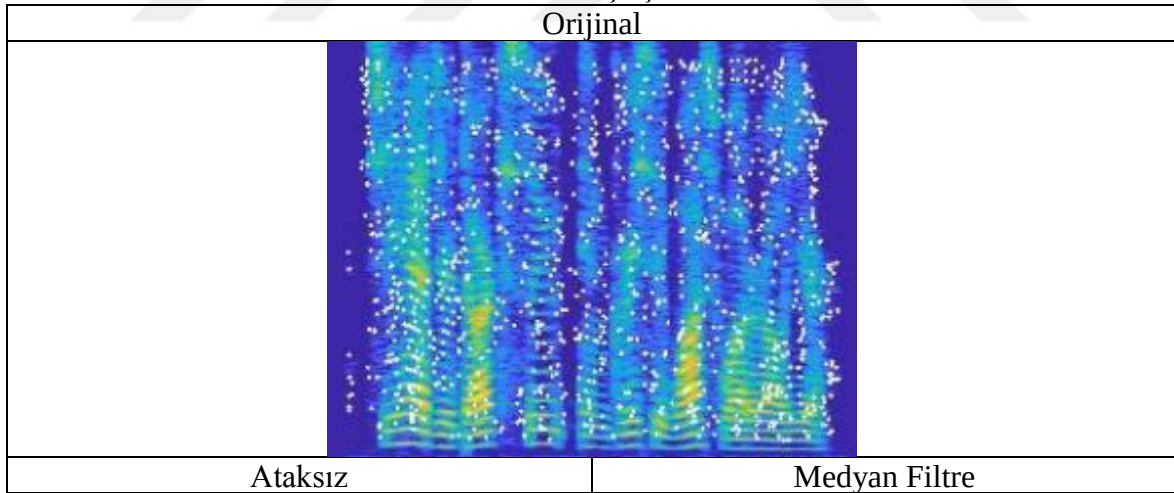
Tablo 4. Anahtar Noktaların Kümelenendirilmesi Yaklaşımı için oluşturulmuş Arabic kopyala yapıştır veriseti özeti

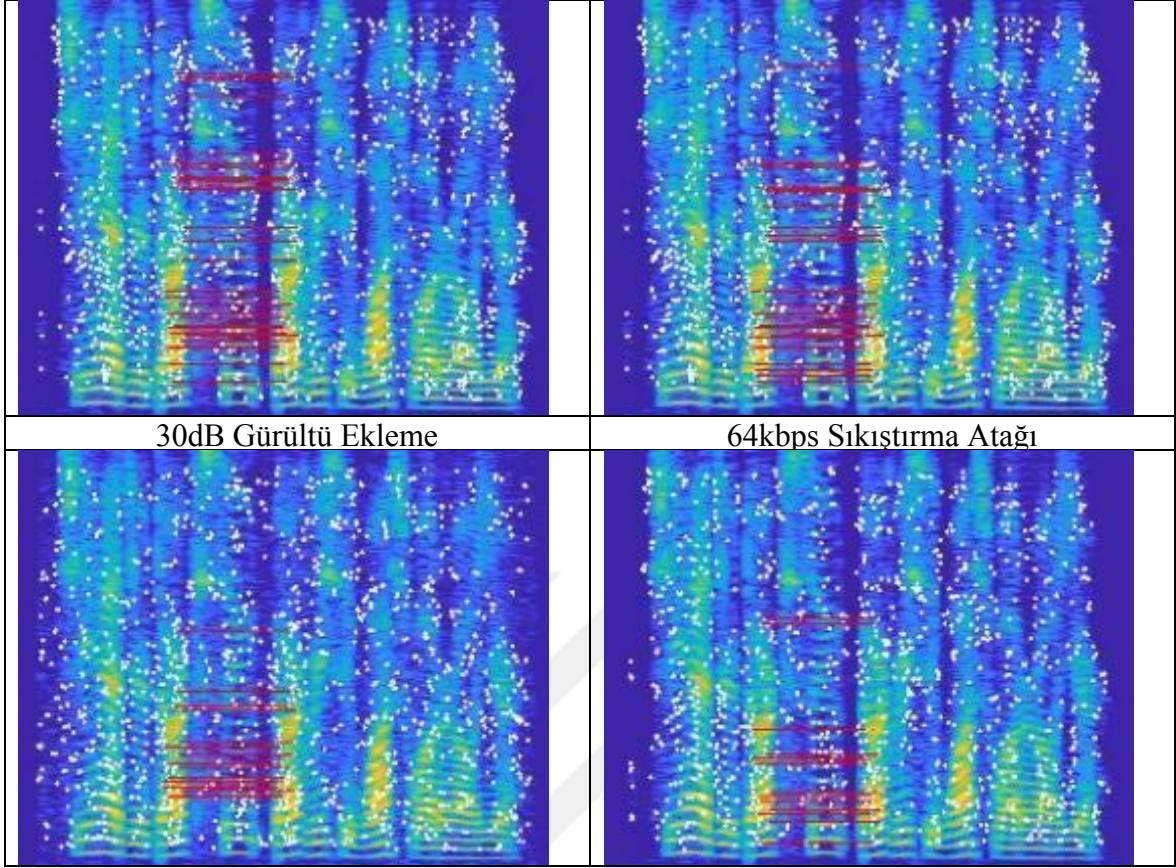
Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	36	36
Sahte	Ataksız	30	150
	Medyan Filtre	30	
	20dB gürültü	30	
	30dB gürültü	30	
	64kbps sıkıştırma	30	

3.2.2. BRIEF Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelenendirilmesi

BRIEF ve OPTICS'in beraber kullanıldığı yöntemde anahtar noktaları Merkez Çevre Ekstreması (CenSurE) yöntemi ile elde edilmiş ve özellik tanımlayıcıları İkili Sağlam Bağımsız Temel Özellikler (BRIEF) ile belirlenmiştir. Tablo 5'te orijinal, ataksız sahte ses ve medyan filtre, 30dB gürültü ve 64kbps sıkıştırma atağı uygulanmış seslere ait spektrogram görüntülerinden elde edilmiş anahtar noktaları ve eşleşmeleri göstermektedir.

Tablo 5. BRIEF ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler





Tablo 5'te verilen görsellere ait anahtar nokta ve eşleşme doğruluğu sayıları Tablo 6'de sunulmaktadır. Orijinal spektrogramlarda kopyala yapıştır sahte segmentleri bulunmadığından eşleşme tespit edilememiştir. En yüksek eşleşme sayısı, ataksız sahte seslerde elde edilmiştir. Buna karşın, sahte sesler arasında en düşük eşleşme sayısı 64kbps sıkıştırma atağında 13 eşleşme ile kaydedilmiştir. Orijinal seslerde 1213 anahtar noktası mevcutken, ataksız sahte seslerde 1333 anahtar noktası tespit edilmiştir. En düşük anahtar noktası sayısı ise 1208 ile 64kbps sıkıştırma atağında gözlemlenmiştir.

Tablo 6. BRIEF ve OPTICS'in beraber kullanıldığı yöntemle ait anahtar nokta ve eşleşme sayıları

Tür	Anahtar Nokta Sayısı	Eşleşme Sayısı
Orijinal	1213	0
Ataksız Sahte Ses	1333	31
Medyan Filtre	1317	26
30dB Gürültü	1362	16
64kbps Sıkıştırma	1208	13

Yöntemin TIMIT veriseti üzerinde alınan sonuçları Tablo 7'te verilmektedir. En yüksek F-skor metriği, ataksız sahte seslerde elde edilmiştir; en düşük metriğe ise 20 dB

gürültü uygulanmış seslerde ulaşılmıştır. Medyan filtre, 30 dB gürültü ve 64 kbps sıkıştırma ataklarında F-skoru 0.96 olarak tespit edilmiştir. 32 kbps sıkıştırma atağı ise 0.94 F-skoruna sahiptir. Tüm atak türlerinde ve ataksız sahte seslerde kesinlik değeri 0.99 olarak tespit edilmiştir. En yüksek duyarlılık skoru ataksız sahte seslerde elde edilirken, en düşük duyarlılık skoru 20 dB gürültü eklenmiş seslerde görülmüştür. 30 dB gürültü ve 64 kbps sıkıştırma ataklarında duyarlılık metriği 0.93 olarak belirlenmiştir. 32 kbps sıkıştırma ve medyan filtre atakları ise sırasıyla 0.89 ve 0.92 duyarlılık skorlarına sahiptir.

Tablo 7. TIMIT kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	0.99	0.94	0.97
Medyan Filtre	0.99	0.92	0.96
20 dB Gürültü Ekleme	0.99	0.82	0.9
30 dB Gürültü Ekleme	0.99	0.93	0.96
32 kbps Sıkıştırma	0.99	0.89	0.94
64 kbps Sıkıştırma	0.99	0.93	0.96

Yöntemin Arabic veri seti üzerinde elde edilen sonuçları Tablo 8'te sunulmuştur. Yapılan analizler sonucunda, 20 dB gürültü atağında 0.96 kesinlik metriği elde edilirken, diğer durumlarda bu skor 1 olarak belirlenmiştir. Ataksız sahte sesler ve 0.85 duyarlılık metriğine sahip 30 dB gürültü ekleme atağı uygulanmış sesler haricinde, tüm ataklı durumlarda duyarlılık metriği 0.86 olarak ölçülmüştür. Ayrıca, 30 dB gürültü atağı altında F-skor metriği 0.90 olarak tespit edilirken, diğer durumlarda bu değer 0.92'dir. Bu sonuçlar, yöntemimizin çeşitli gürültü seviyelerinde ve atak türlerinde yüksek performans gösterdiğini ortaya koymaktadır.

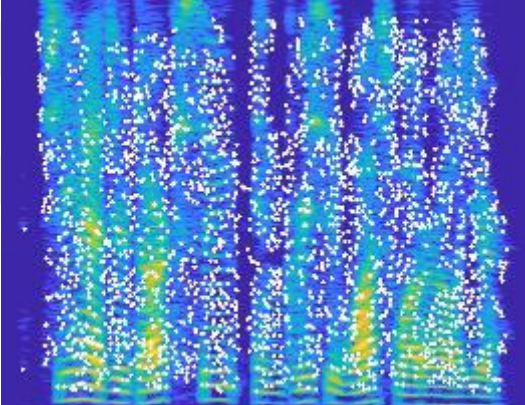
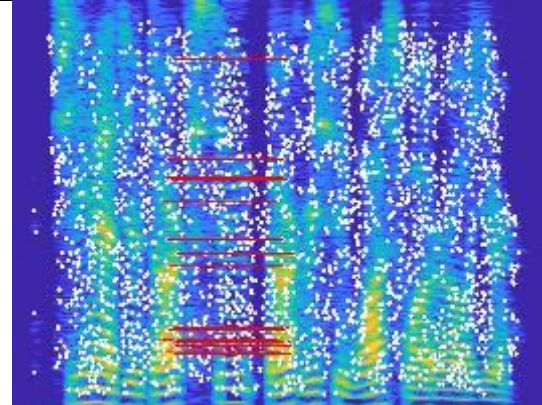
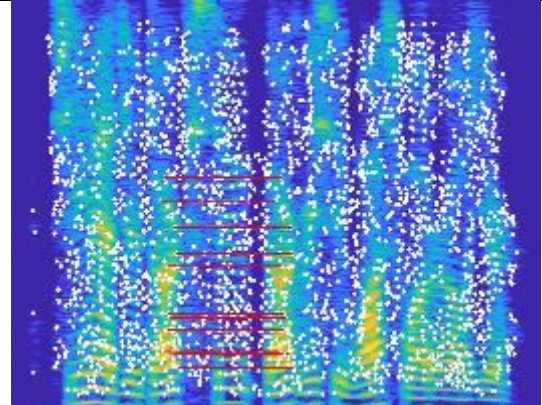
Tablo 8. Arabic kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları

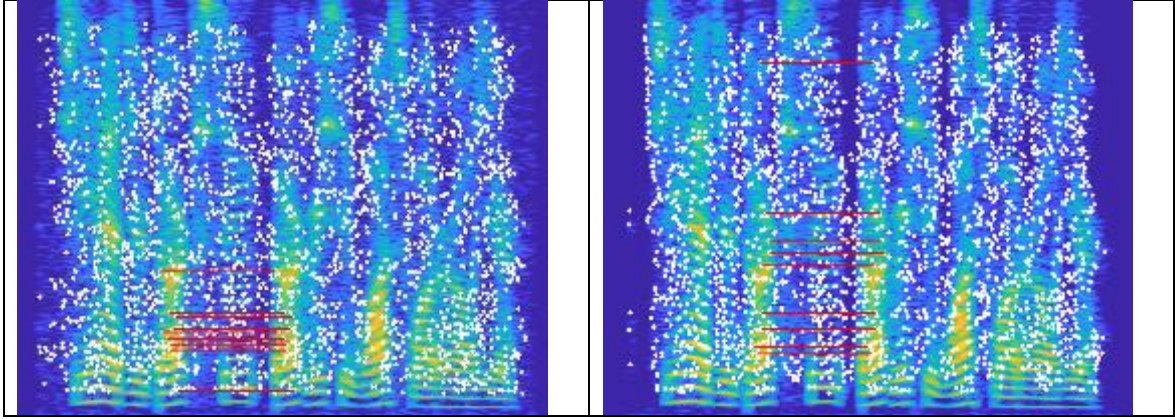
	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	1	0.86	0.92
Medyan Filtre	1	0.86	0.92
20 dB Gürültü Ekleme	0.96	0.86	0.92
30 dB Gürültü Ekleme	1	0.85	0.90
64 kbps Sıkıştırma	1	0.86	0.92

3.2.3. AKAZE Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi

AKAZE ve OPTICS kombinasyonu ile oluşturulan yöntemde yüksek çözünürlüklü spektrogram görüntülerinden AKAZE kullanılarak anahtar noktaları çıkarılmıştır. Kümelendirme işlemi OPTICS aracılığıyla gerçekleştirilmiştir. Tablo 9’da yönteme ait anahtar noktaları ve eşleşme doğruları örnek orijinal ses ve bununla gerçekleştirilmiş sahte ses ve atak uygulanmış hallerinin spektrogramı üzerinde gösterilmiştir. Orijinal spektrogram üzerinde eşleşme belirlenmediği görsel üzerinde görünmektedir. Buna sahte ses ve ataklı hallerinde eşleşme mevcuttur. Tablo 9’da verilen görsellere ait anahtar nokta ve eşleşme sayısı Tablo 10’da verilmektedir. Görsellerde de görüldüğü üzere orijinal spektrogram üzerinde herhangi eşleşme saptanmamıştır.

Tablo 9. AKAZE ve OPTICS’in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler

Orijinal	
	
Ataksız	Medyan Filtre
	
30dB Gürültü Ekleme	64kbps Sıkıştırma Atakı



Tablo 10. AKAZE ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları

Tür	Anahtar Nokta Sayısı	Eşleşme Sayısı
Orijinal	2906	0
Ataksız Sahte Ses	2986	16
Medyan Filtre	2950	12
30dB Gürültü	2870	10
64kbps Sıkıştırma	2933	9

Orijinal spektrogramda 2906 anahtar nokta sayısı tespit edilmiştir. Bu sayı, 30 dB gürültü uygulanmış seste elde edilen spektrogram üzerindeki anahtar nokta sayısından fazladır. Gürültü atağı, görüntü üzerindeki anahtar nokta sayısını olumsuz etkilemektedir. Sahte spektrogramlar üzerinde en yüksek anahtar nokta sayısı ataksız konuşma kaydına aittir. Bunu, 2950 sayısı ile medyan filtre ve 2933 sayısı ile 64 kbps sıkıştırma atağı takip etmektedir.

Eşleşme doğru sayılarında en yüksek skor, 16 ile ataksız sahte seslere ait iken, en düşük sayı 9 ile 64 kbps atağına aittir. 30 dB gürültü ekleme ve medyan filtre ataklarında sırasıyla 10 ve 12 eşleşme sayısı görülmüştür. Konuşma kayıtlarına uygulanan ataklar, elde edilen eşleşme sayısında düşüşe sebep olmaktadır.

Metodun TIMIT veriseti üzerinde elde edilen sonuçları Tablo 11'de sunulmaktadır. En yüksek F-skoru, ataksız sahte seslerde kaydedilmiş olup, en düşük F-skoru ise 20 dB gürültü eklenmiş seslerde görülmüştür. Medyan filtre, 30 dB gürültü ve 64 kbps sıkıştırma ataklarında F-skoru 0.94 olarak ölçülmüştür. Buna karşılık, 32 kbps sıkıştırma atağı 0.90 F-skoruna sahiptir. Tüm atak türlerinde ve ataksız sahte seslerde kesinlik değeri 1 olarak belirlenmiştir. En yüksek duyarlılık, ataksız sahte seslerde tespit edilirken, en düşük

duyarlılık 20 dB gürültü eklenmiş seslerde görülmüştür. 30 dB gürültü ve 64 kbps sıkıştırma ataklarında duyarlılık metriği 0.89 olarak kaydedilmiştir. 32 kbps sıkıştırma ve medyan filtre atakları sırasıyla 0.82 ve 0.90 duyarlılık skorlarına sahiptir.

Tablo 11. TIMIT veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	1	0.93	0.96
Medyan Filtre	1	0.90	0.94
20 dB Gürültü Ekleme	1	0.79	0.88
30 dB Gürültü Ekleme	1	0.89	0.94
32 kbps Sıkıştırma	1	0.82	0.90
64 kbps Sıkıştırma	1	0.89	0.94

Yöntemin Arabic veri seti üzerinde elde edilen sonuçları Tablo 12'te sunulmuştur. Yapılan analizler sonucunda, 20 dB gürültü atağı ve 64 kbps sıkıştırma atağında sırasıyla 0.96 ve 0.93 kesinlik metriği elde edilmiştir. Diğer durumlarda bu skor 1 olarak belirlenmiştir. Ataksız sahte sesler, medyan filtre ve 30 dB gürültü ekleme atağı altında duyarlılık metriği 0.78 olarak ölçülmüştür. Buna karşın, 20 dB gürültü ekleme atağı ve 64 kbps sıkıştırma atağı ile duyarlılık skoru 0.77'ye ulaşmıştır. Ayrıca, 64 kbps sıkıştırma ve 20 dB gürültü atağı altında F-skor metriği sırasıyla 0.84 ve 0.85 olarak tespit edilmişken, diğer durumlarda bu değer 0.88'dir. Bu sonuçlar, yöntemimizin çeşitli gürültü seviyelerinde ve atak türlerinde yüksek performans gösterdiğini ortaya koymaktadır.

Tablo 12. Arabic veriseti üzerinde yöntemin deneysel sonuçları

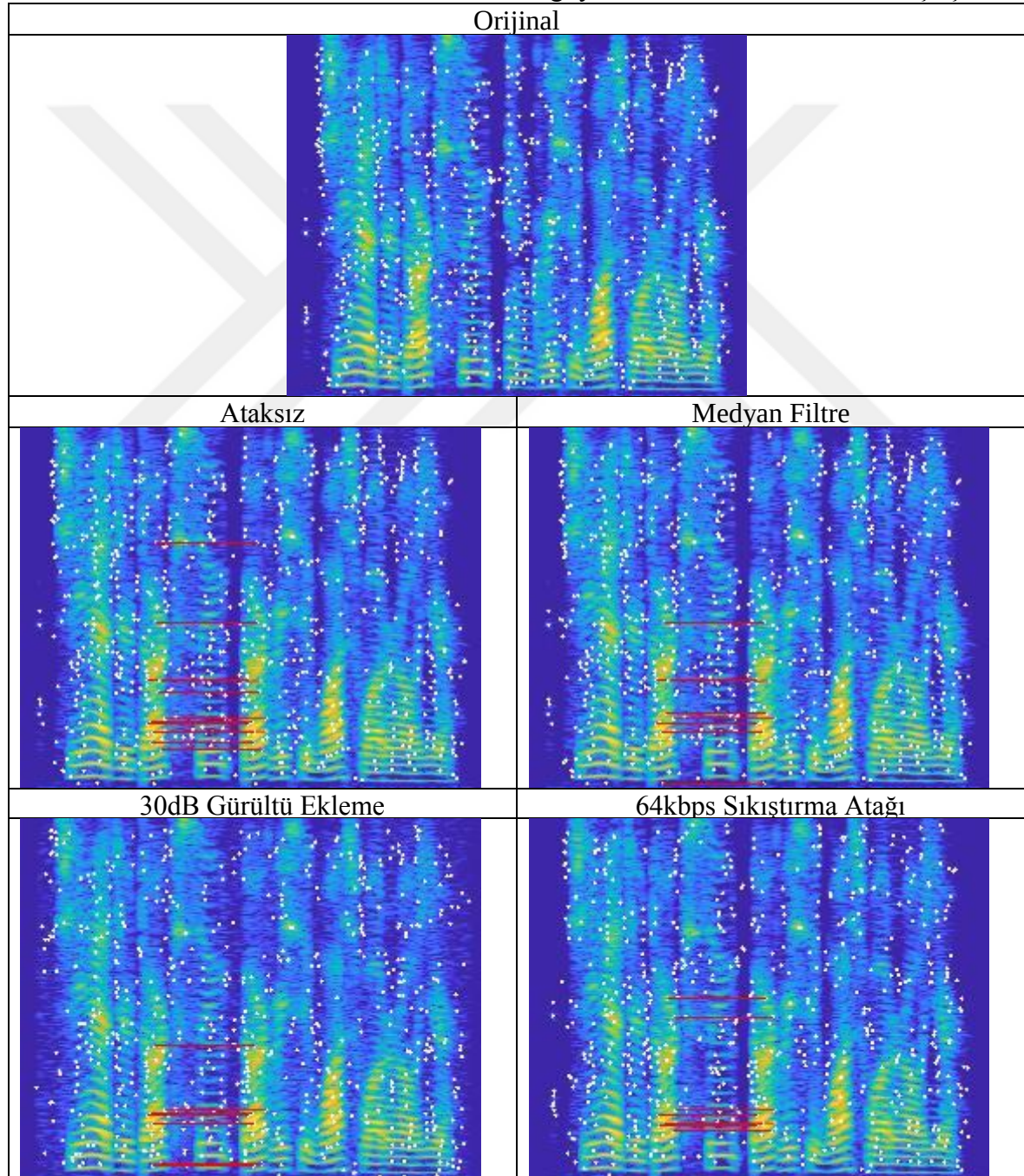
	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	1	0.78	0.88
Medyan Filtre	1	0.78	0.88
20 dB Gürültü Ekleme	0.96	0.77	0.85
30 dB Gürültü Ekleme	1	0.78	0.88
64 kbps Sıkıştırma	0.93	0.77	0.84

3.2.4. SIFT Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi

Anahtar noktası çıkarma yöntemlerinden SIFT seçilerek yüksek çözünürlüklü spektrogram görüntülerine uygulanmıştır. Elde edilen noktalar özellik tanımlayıcı vektörler kullanılarak OPTICS ile kümelendirilmiştir. Tablo 13'te, yöntemin anahtar noktaları ve eşleşme doğruları, orijinal ses, sahte ses ve çeşitli atakların uygulanmış hallerinin

spektrogramları üzerinde gösterilmiştir. Görsellerde, orijinal spektrogram üzerinde herhangi eşleşmenin belirlenmediği açıkça görülmektedir. Ancak, sahte ses ve atak uygulanmış hallerde eşleşmeler mevcuttur. Tablo 13'teki görsellerin anahtar noktaları ve eşleşme sayıları Tablo 14'te detaylandırılmıştır. Görsellerde görüldüğü üzere, orijinal spektrogramda hiçbir eşleşme tespit edilmemiştir. Ataksız sahte spektrogramda ve tüm ataklı spektrogram görsellerinde anahtar nokta sayısının diğer yöntemlere göre oldukça azaldığı gözlemlenmektedir.

Tablo 13. SIFT ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler



Yapılan analizler sonucunda, en yüksek anahtar nokta sayısının 877 ile ataksız sahte ses spektrogramında elde edildiği görülmüştür. Ancak, 64 kbps sıkıştırma işlemi uygulandığında bu sayı 771'e düşmüştür. Ayrıca, medyan filtre ve 30 dB gürültü atağı altında anahtar nokta sayıları sırasıyla 852 ve 781 olarak kaydedilmiştir. Orijinal seslerde ise anahtar nokta sayısı 787 olarak belirlenmiştir.

Analizler neticesinde, en düşük eşleşme sayısının 7 ile 30 dB gürültü atağı uygulanmış spektrograma ait olduğunu göstermektedir. Buna karşın, en yüksek eşleşme sayısı 12 ile ataksız sahte ses spektrogramında gözlemlenmiştir. Orijinal ses spektrogramında ise hiç eşleşme bulunamamıştır. Medyan filtre ve 64 kbps sıkıştırma ataklarında eşleşme sayısı 8 olarak tespit edilmiştir.

Tablo 14. SIFT ve OPTICS'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşme sayıları

Tür	Anahtar Nokta Sayısı	Eşleşme Sayısı
Orijinal	787	0
Ataksız Sahte Ses	877	12
Medyan Filtre	852	8
30dB Gürültü	781	7
64kbps Sıkıştırma	771	8

Metodun TIMIT veri seti üzerinde elde edilen sonuçları Tablo 15'te sunulmuştur. En yüksek F-skoru, ataksız sahte seslerde kaydedilmiş olup, en düşük F-skoru ise 20 dB gürültü eklenmiş seslerde gözlemlenmiştir. Medyan filtre ve 30 dB gürültü ataklarında F-skoru 0.91 olarak ölçülmüştür. Buna karşılık, 32 kbps ve 64 kbps sıkıştırma ataklarında F-skorları sırasıyla 0.79 ve 0.88 olarak belirlenmiştir. Tüm atak türlerinde ve ataksız sahte seslerde kesinlik değeri 1 olarak tespit edilmiştir. En yüksek duyarlılık, ataksız sahte seslerde kaydedilirken, en düşük duyarlılık 20 dB gürültü eklenmiş seslerde görülmüştür. Bununla birlikte, 30 dB gürültü eklenmiş seslerin spektrogramlarında 0.85 duyarlılık metriği elde edilmiştir. Medyan filtre, 32 kbps ve 64 kbps sıkıştırma ataklarında duyarlılık değerleri sırasıyla 0.84, 0.66 ve 0.79 olarak gözlemlenmiştir. SIFT anahtar noktasının kullanımında, duyarlılık metriği 20 dB gürültü ve 32 kbps sıkıştırma durumlarında 0.7'nin altında kalmıştır.

Tablo 15. TIMIT veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	1	0.91	0.95
Medyan Filtre	1	0.84	0.91
20 dB Gürültü Ekleme	1	0.64	0.78
30 dB Gürültü Ekleme	1	0.85	0.91
32 kbps Sıkıştırma	1	0.66	0.79
64 kbps Sıkıştırma	1	0.79	0.88

Arabic veri seti üzerinde yapılan çalışmanın sonuçları Tablo 16'da sunulmuştur. Ataksız sahte ses ve 30 dB gürültü atağı uygulanan spektrogramlar için 0.90 F-skor değeri elde edilmiştir. En yüksek F-skor değeri ise 0.93 ile medyan filtre uygulanmış sahte ses spektrogramlarıyla kaydedilmiştir. Ancak, 20 dB gürültü ve 64 kbps sıkıştırma ataklarında 0.87 ile en düşük F-skor metriği gözlemlenmiştir. Duyarlılık metriği ise 20 dB gürültü ve 64 kbps sıkıştırma ataklarında en düşük seviyede 0.86, medyan filtre atağında ise en yüksek seviyede 0.96 tespit edilmiştir. Ataksız sahte spektrogram ve 30 dB gürültü atağı altında duyarlılık seviyesi ise 0.90 olarak gözlemlenmiştir. Kesinlik metriği, ataksız sahte spektrogram, medyan filtre ve 30 dB gürültü atakları altında 0.90 seviyesindeyken, 20 dB gürültü ve 64 kbps sıkıştırma atakları altında 0.89 olarak ölçülmüştür. Yöntemin Arabic veri setinde, tüm metriklerde medyan filtre atağı altında daha iyi başarı elde ettiği gözlemlenmiştir.

Tablo 16. Arabic veriseti üzerinde yöntemin deneysel sonuçları

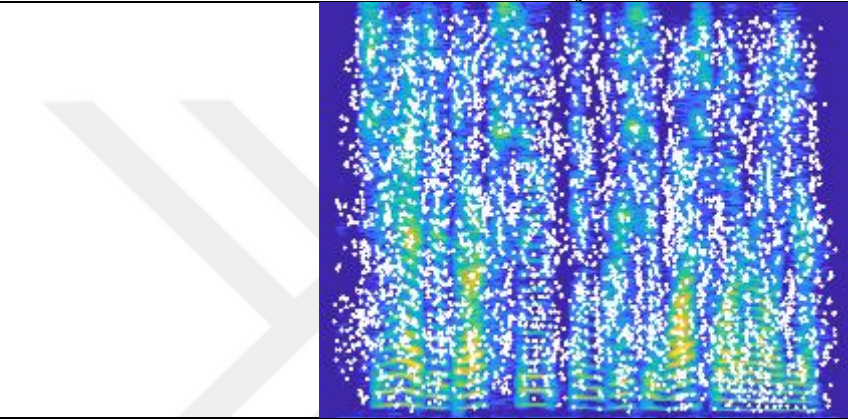
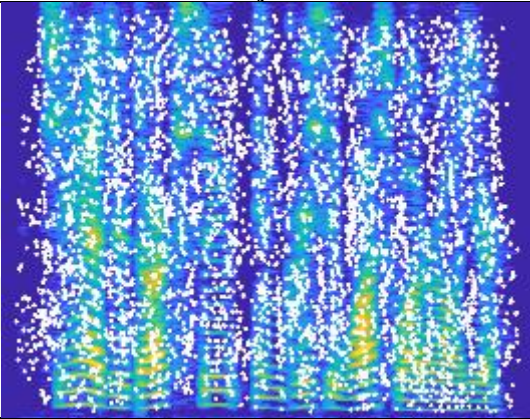
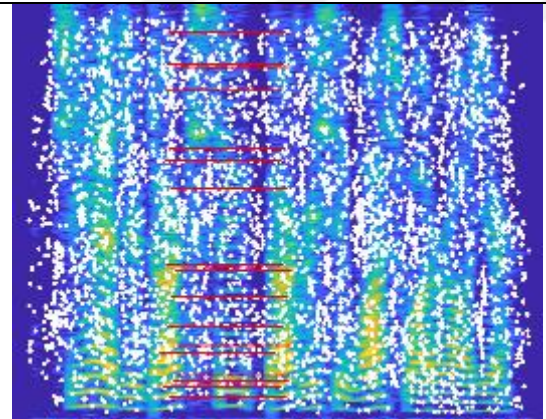
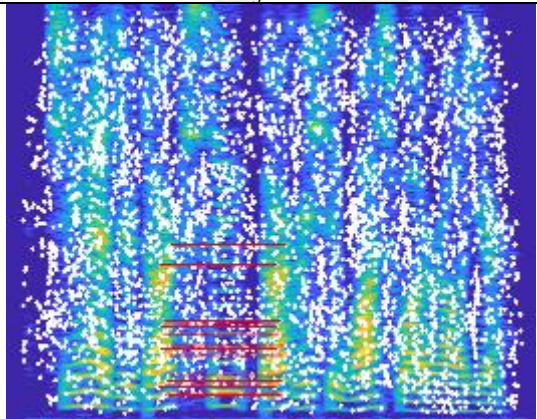
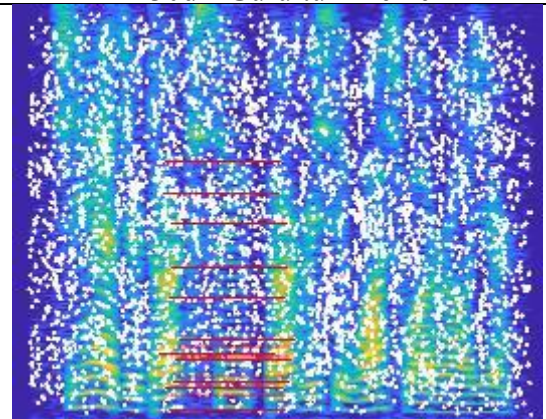
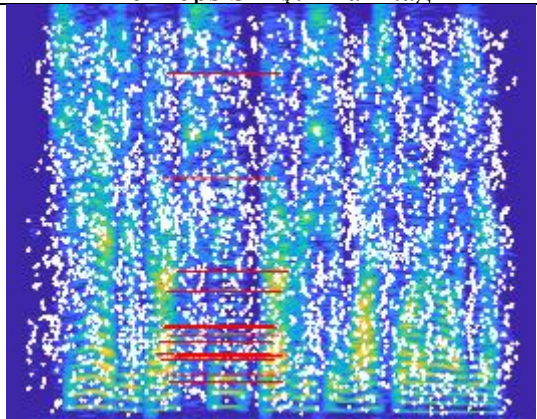
	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	0.90	0.90	0.90
Medyan Filtre	0.90	0.96	0.93
20 dB Gürültü Ekleme	0.89	0.86	0.87
30 dB Gürültü Ekleme	0.90	0.90	0.90
64 kbps Sıkıştırma	0.89	0.86	0.87

3.2.5. SURF Anahtar Noktalarının OPTICS Yaklaşımı ile Kümelendirilmesi

SURF yöntemi kullanılarak yüksek çözünürlüklü spektrogram görüntülerine anahtar nokta çıkarma işlemi gerçekleştirilmiştir. Elde edilen noktalar, özellik tanımlayıcı vektörler kullanılarak OPTICS algoritmasıyla kümelendirilmiştir. Tablo 17'de, yöntemin anahtar noktaları ve eşleşme doğruları orijinal ses, sahte ses ve çeşitli ataklar uygulanmış spektrogramlar üzerinde görselleştirilmiştir. Görsellerden açıkça anlaşıldığı üzere, orijinal

spektrogramda herhangi eşleşme belirlenmemiştir. Ancak, sahte ses ve atak uygulanmış spektrogramlarda eşleşmeler görülmektedir. Tablo 17'deki görsellerdeki anahtar nokta sayıları ve eşleşme detayları Tablo 18'te ayrıntılı olarak sunulmuştur. Görsellerde gözlendiği gibi, orijinal spektrogramda hiçbir eşleşme tespit edilmemiştir. Ataksız sahte spektrogramda ve tüm ataklı spektrogramlarda anahtar nokta sayısının diğer yöntemlere göre önemli ölçüde azaldığı gözlemlenmektedir.

Tablo 17. SURF ve OPTICS'in beraber kullanıldığı yönetime ait anahtar nokta ve eşleşmeler

Orijinal	
	
Ataksız	Medyan Filtre
	
30dB Gürültü Ekleme	64kbps Sıkıştırma Atak
	

Tablo 18'in analizine göre, en yüksek anahtar nokta sayısına 30 dB gürültü atağı uygulanmış sahte sesin spektrogramıyla ulaşılmıştır. Ancak, en düşük sayı 64 kbps sıkıştırma atağına maruz kalmış sahte sesin spektrogramıyla kaydedilmiştir. Orijinal spektrogramda bu sayının 6997 olduğu gözlemlenmiştir. Ataksız sahte ve medyan filtre ataklarına ait spektrogramlarda ise sırasıyla 7031 ve 7010 adet anahtar nokta tespit edilmiştir.

Orijinal spektrogramda herhangi eşleşme bulunmamaktadır. Bu durum Tablo 18'te gösterilmiştir. En yüksek eşleşme sayısı, ataksız sahte sesin spektrogramında elde edilmiştir. Ataklı seslerde, en düşük eşleşme sayısı medyan filtre uygulandığında gözlemlenmiştir. 64 kbps sıkıştırma ve 30 dB gürültü atağında eşleşme sayıları sırasıyla 14 ve 15 olarak belirlenmiştir. Sonuçlar, sese atak uygulanmasının eşleşme sayısında düşüşe yol açtığını göstermektedir.

Tablo 18. SURF ve OPTICS'in beraber kullanıldığı yöntemle ait anahtar nokta ve eşleşme sayıları

Tür	Anahtar Nokta Sayısı	Eşleşme Sayısı
Orijinal	6997	0
Ataksız Sahte Ses	7031	17
Medyan Filtre	7010	13
30dB Gürültü	7289	15
64kbps Sıkıştırma	6790	14

Metodun TIMIT veri seti üzerinde elde edilen sonuçları Tablo 19'te gösterilmiştir. En yüksek F-skoru, 0.96 ile ataksız sahte seslerde elde edilirken, en düşük F-skoru 0.79 ile 20 dB gürültü eklenmiş seslerde tespit edilmiştir. Medyan filtre ve 30 dB gürültü ataklarında F-skorları sırasıyla 0.94 ve 0.92 olarak belirlenmiştir. Diğer yandan, 32 kbps ve 64 kbps sıkıştırma ataklarında F-skorları sırasıyla 0.82 ve 0.91 olarak kaydedilmiştir. Tüm atak türlerinde ve ataksız sahte seslerde kesinlik değeri 1 olarak saptanmıştır. En yüksek duyarlılık ataksız sahte seslerde görülürken, en düşük duyarlılık 20 dB gürültü eklenmiş seslerde gözlemlenmiştir. Ayrıca, 30 dB gürültü eklenmiş seslerin spektrogramlarında duyarlılık metriği 0.85 olarak bulunmuştur. Medyan filtre, 32 kbps ve 64 kbps sıkıştırma ataklarında duyarlılık değerleri sırasıyla 0.89, 0.70 ve 0.84 olarak kaydedilmiştir. SURF anahtar noktalarının kullanıldığı durumda, duyarlılık metriği 20 dB gürültü atağı altında 0.7'nin altında kalmıştır.

Tablo 19. TIMIT veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skör
Ataksız Sahte Ses	1	0.92	0.96
Medyan Filtre	1	0.89	0.94
20 dB Gürültü Ekleme	1	0.65	0.79
30 dB Gürültü Ekleme	1	0.85	0.92
32 kbps Sıkıştırma	1	0.70	0.82
64 kbps Sıkıştırma	1	0.84	0.91

Arabic veri seti üzerinde yapılan çalışmanın sonuçları Tablo 20'te sunulmuştur. 20 dB gürültü eklenmiş sahte sesler ve 64 kbps sıkıştırma atağı uygulanmış spektrogramlar için 0.93 F-skör değeri elde edilmiştir. En yüksek F-skör değeri, 0.95 ile medyan filtre uygulanmış sahte ses spektrogramlarında kaydedilmiştir. Ataksız sahte sesler ve 30 dB gürültü atağı altında ise 0.94 F-skör metriği gözlemlenmiştir. Duyarlılık metriği, 20 dB gürültü ve 64 kbps sıkıştırma ataklarında en düşük seviyede 0.93, medyan filtre atağında ise en yüksek seviyede 0.97 olarak tespit edilmiştir. Ataksız sahte spektrogram ve 30 dB gürültü atağı altında duyarlılık seviyesi 0.96 olarak belirlenmiştir. Kesinlik metriği ise ataksız sahte spektrogram ve tüm ataklar altında 0.93 seviyesinde kalmıştır. Yöntemin Arabic veri setinde, tüm metriklerde medyan filtre atağı altında daha iyi başarı elde ettiği gözlemlenmiştir.

Tablo 20. Arabic veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skör
Ataksız Sahte Ses	0.93	0.96	0.94
Medyan Filtre	0.93	0.97	0.95
20 dB Gürültü Ekleme	0.93	0.93	0.93
30 dB Gürültü Ekleme	0.93	0.96	0.94
64 kbps Sıkıştırma	0.93	0.93	0.93

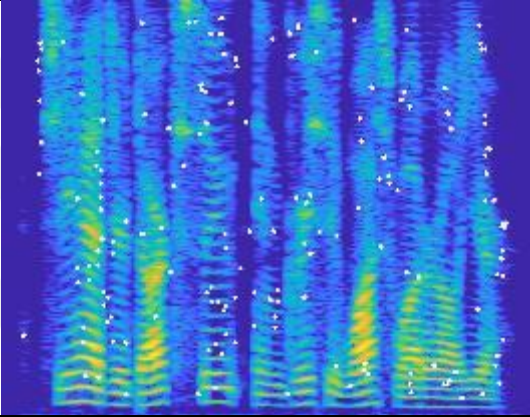
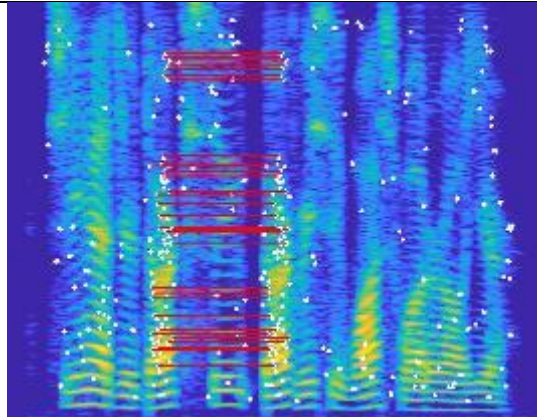
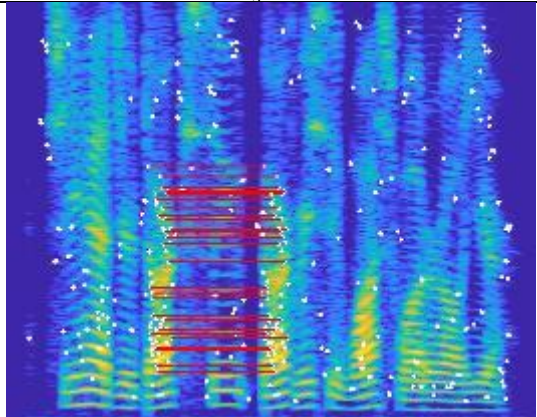
3.2.6. BRIEF Anahtar Noktalarının g2NN Yaklaşımı ile Kümelendirilmesi

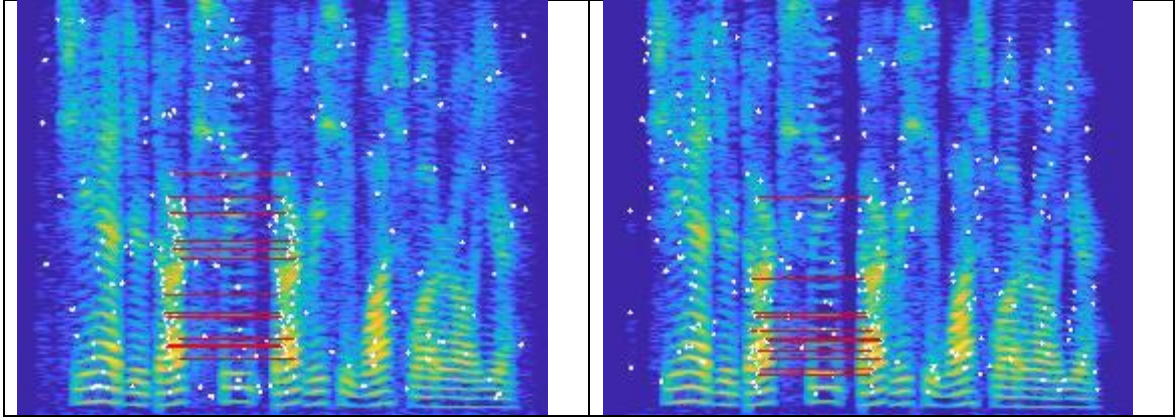
Anahtar noktalarının kümelendirilerek kopyala yapıştır ses sahteciliği tespit yapan diğer yöntemimiz BRIEF ve g2NN kombinasyonuna dayanmaktadır. Bu yaklaşımda yüksek çözünürlüklü spektrogram görüntülerinden anahtar noktaları Merkez Çevre Ekstreması (CenSurE) yöntemi ile elde edilmiş ve özellik tanımlayıcıları İkili Sağlam Bağımsız Temel Özellikler (BRIEF) ile belirlenmiştir. Kümelendirme aşamasında g2NN tercih edilmiştir. Tablo 21'te, yönteme ait anahtar noktalar ve eşleşme doğruları örnek orijinal ses, bununla oluşturulmuş sahte ses ve atak uygulanmış halleri üzerinden spektrogram üzerinde

gösterilmiştir. Orijinal spektrogramda eşleşme bulunmadığı görselde açıkça görülmektedir. Ancak, sahte ses ve atak uygulanmış hallerinde eşleşmeler mevcuttur. Tablo 21'te gösterilen görsellere ait anahtar nokta ve eşleşme sayıları Tablo 22'te belirtilmiştir. Görsellerde de görüldüğü gibi, orijinal spektrogramda herhangi eşleşme tespit edilmemiştir.

Yapılan analizler sonucunda, en yüksek anahtar nokta sayısının 410 ile ataksız sahte ses spektrogramında elde edildiği tespit edilmiştir. Ancak, 64 kbps sıkıştırma işlemi uygulandığında bu sayı 320'ye düşmüştür. Medyan filtre atağında anahtar nokta sayısı 390, 30 dB gürültü atağında ise 334 olarak belirlenmiştir. Orijinal seslerde ise anahtar nokta sayısı 244 olarak kaydedilmiştir.

Tablo 21. BRIEF ve g2NN'in beraber kullanıldığı yöntemle ait anahtar nokta ve eşleşmeler

Orijinal	
	
Ataksız	Medyan Filtre
	
30dB Gürültü Ekleme	64kbps Sıkıştırma Atağı



Yapılan incelemeler sonucunda, en düşük eşleşme sayısının 12 ile 64 kbps sıkıştırma atağı uygulanmış spektrogramda olduğu tespit edilmiştir. Buna karşılık, en yüksek eşleşme sayısı 32 ile ataksız sahte ses spektrogramında görülmüştür. Orijinal ses spektrogramında ise hiçbir eşleşme bulunamamıştır. Medyan filtre ve 30 dB gürültü ataklarında eşleşme sayıları sırasıyla 28 ve 14 olarak belirlenmiştir.

Tablo 22. BRIEF ve g2NN'in beraber kullanıldığı yöntemle ait anahtar nokta ve eşleşme sayıları

Tür	Anahtar Nokta Sayısı	Eşleşme Sayısı
Orijinal	244	0
Ataksız Sahte Ses	410	32
Medyan Filtre	390	28
30dB Gürültü	334	14
64kbps Sıkıştırma	320	12

Metodun TIMIT veri seti üzerinde elde edilen sonuçlar Tablo 23'te sunulmuştur. En yüksek F-skor, 64 kbps sıkıştırma atağı uygulanmış sahte seslerde 0.96 olarak belirlenirken, en düşük F-skor 20 dB gürültü eklenmiş seslerde 0.88 olarak gözlemlenmiştir. Medyan filtre, 30 dB gürültü ve 32 kbps sıkıştırma ataklarında F-skor metrikleri sırasıyla 0.94, 0.95 ve 0.92 olarak tespit edilmiştir. Tüm atak türlerinde ve ataksız sahte seslerde kesinlik değeri 1 olarak kaydedilmiştir. En yüksek duyarlılık, 64 kbps sıkıştırılmış sahte seslerde 0.93 olarak belirlenirken, en düşük duyarlılık 20 dB gürültü eklenmiş seslerde 0.80 olarak gözlemlenmiştir. Ayrıca, ataksız ve 30 dB gürültü atağı uygulanmış sahte seslere ait spektrogramlarda duyarlılık metriği 0.92 olarak kaydedilmiştir. Medyan filtre ve 32 kbps sıkıştırma ataklarında duyarlılık değerleri sırasıyla 0.90 ve 0.86 olarak tespit edilmiştir.

Tablo 23. TIMIT veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	1	0.92	0.95
Medyan Filtre	1	0.90	0.94
20 dB Gürültü Ekleme	1	0.80	0.88
30 dB Gürültü Ekleme	1	0.92	0.95
32 kbps Sıkıştırma	1	0.86	0.92
64 kbps Sıkıştırma	1	0.93	0.96

Arabic veri seti üzerinde yapılan çalışmanın sonuçları Tablo 24'te sunulmuştur. 20 dB gürültü eklenmiş sahte sesler ve 64 kbps sıkıştırma atağı uygulanmış spektrogramlar için 0.93 F-skor değeri elde edilmiştir. En yüksek F-skor değeri, 0.95 ile ataksız sahte seslerde, medyan filtre ve 30 dB gürültü atakları uygulanmış spektrogramlarda kaydedilmiştir. 64 kbps sıkıştırma ve 20 dB gürültü atakları altında ise 0.93 F-skor metriği gözlemlenmiştir. Duyarlılık metriği, ataksız sahte spektrogram ve tüm ataklar altında 0.90 seviyesinde kalmıştır. Kesinlik metriği ise ataksız sahte spektrogram, medyan filtre ve 30 dB gürültü ataklarında 1 olarak gözlemlenmiştir. Ancak, 20 dB gürültü ve 64 sıkıştırma ataklarında 0.96 kesinlik metriğine ulaşılmıştır.

Tablo 24. Arabic veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	1	0.90	0.95
Medyan Filtre	1	0.90	0.95
20 dB Gürültü Ekleme	0.96	0.90	0.93
30 dB Gürültü Ekleme	1	0.90	0.95
64 kbps Sıkıştırma	0.96	0.90	0.93

3.2.7. BRIEF Anahtar Noktalarının k-Means Yaklaşımı ile Kümelendirilmesi

Kümelendirme yöntemi olarak k-Means tercihinin tez çalışması kapsamında kullanımı sonuçları bu bölümde detaylı şekilde açıklanacaktır. Spektrogram görüntülerinden anahtar noktaları Merkez Çevre Ekstreması (CenSurE) yöntemi ile çıkarılmış ve özellik tanımlayıcıları İkili Sağlam Bağımsız Temel Özellikler (BRIEF) ile belirlenmiştir. Tablo 26'te, yöntemin anahtar noktaları ve eşleşme doğruları orijinal ses, sahte ses ve çeşitli ataklar uygulanmış spektrogramlar üzerinde görselleştirilmiştir. Görsellerden açıkça anlaşıldığı üzere, orijinal spektrogramda diğer yöntemlerden farklı olarak eşleşme belirlenmiştir.

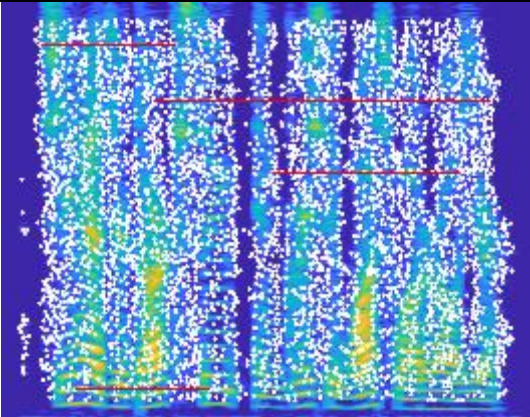
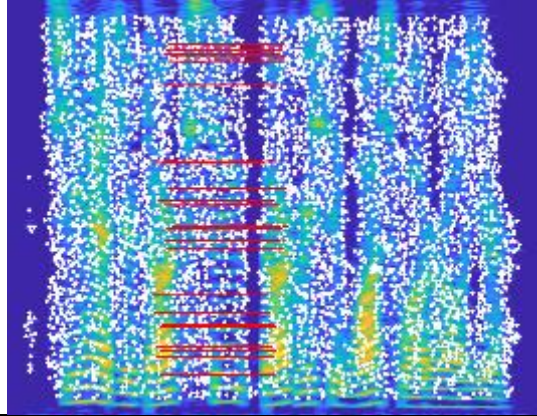
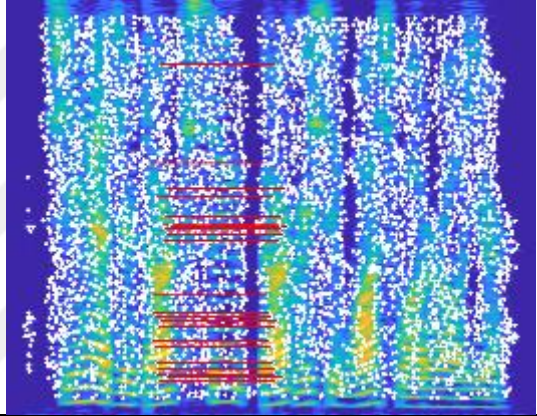
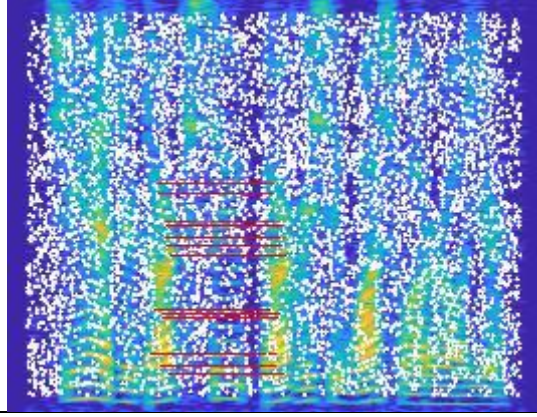
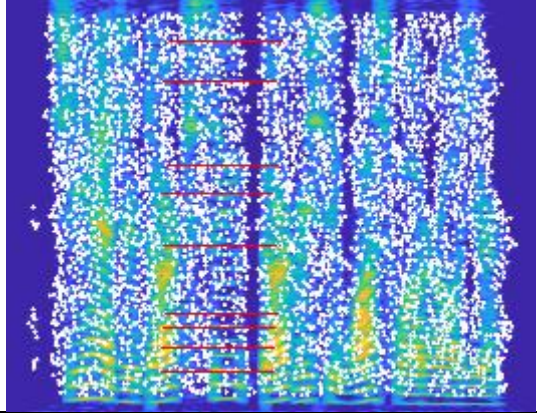
Ataksız sahte ses ve atak uygulanmış spektrogramlarda eşleşmeler göze çarpmaktadır. Tablo 26'daki görsellerdeki anahtar nokta sayıları ve eşleşme detayları Tablo 25'da ayrıntılı olarak sunulmuştur. Orijinal spektrogram üzerinde eşleşmelerin tespit edilmesi, yöntemin başarısını oldukça etkilemektedir. Bu durum, orijinal seslerin sahte olarak tanımlanmasına neden olabilir. En yüksek eşleşme sayısı, ataksız sahte seslere ait spektrogramlarda elde edilmiştir. Öte yandan, 64 kbps sıkıştırma uygulanmış konuşma kayıtlarına ait yüksek çözünürlüklü spektrogramlarda bu sayı en düşük seviyede 10 olarak ölçülmüştür. Medyan filtre ve 30 dB gürültü ataklarında ise sırasıyla 22 ve 17 eşleşme sayısı kaydedilmiştir. Orijinal spektrogram üzerinde 4 adet eşleşen bölge tespit edilmiştir.

Tablo 25'in analizine göre, en yüksek anahtar nokta sayısına 30 dB gürültü atağı uygulanmış sahte sesin spektrogramıyla ulaşılmıştır. Ancak, en düşük sayı 64 kbps sıkıştırma atağına maruz kalmış sahte sesin spektrogramıyla kaydedilmiştir. Orijinal spektrogramda bu sayının 4662 olduğu gözlemlenmiştir. Ataksız sahte ve medyan filtre ataklarına ait spektrogramlarda ise sırasıyla 4757 ve 4745 adet anahtar nokta tespit edilmiştir.

Tablo 25. BRIEF ve k-Means'in beraber kullanıldığı yöntemle ait anahtar nokta ve eşleşme sayıları

Tür	Anahtar Nokta Sayısı	Eşleşme Sayısı
Orijinal	4662	4
Ataksız Sahte Ses	4757	24
Medyan Filtre	4745	22
30dB Gürültü	5402	17
64kbps Sıkıştırma	4717	10

Tablo 26. BRIEF ve k-Means'in beraber kullanıldığı yönteme ait anahtar nokta ve eşleşmeler

Orijinal	
	
Ataksız	Medyan Filtre
	
30dB Gürültü Ekleme	64kbps Sıkıştırma Atağı
	

Metodun TIMIT veri seti üzerinde elde edilen sonuçlar Tablo 27'te sunulmuştur. En yüksek F-skor, ataksız sahte seslerde 0.81 olarak belirlenirken, en düşük F-skor 20 dB gürültü eklenmiş seslerde 0.67 olarak gözlemlenmiştir. Medyan filtre, 30 dB gürültü ve 64 kbps sıkıştırma ataklarında F-skor metriği 0.79 olarak tespit edilmiştir. Öte yandan 32 kbps sıkıştırma atağında 0.74 skoru kaydedilmiştir. En yüksek kesinlik metriği 0.73 ile ataksız sahte seslerde alınmıştır. Buna karşın en düşük 20 dB gürültü atağında 0.66 ile ulaşılmıştır.

Medyan filtre, 30 dB gürültü ve 64 kbps sıkıştırma ataklarında 0.72 kesinlik metriğine sahiptir. Diğer atak durumu 32 kbps sıkıştırma 0.70 kesinlik metrik değerine sahiptir. Ataksız sahte seslerde en yüksek duyarlılık seviyesi 0.92 olarak saptanırken, 20 dB gürültü eklenmiş seslerdeki en düşük duyarlılık 0.69 olarak kaydedilmiştir. Medyan filtre ve 64 kbps sıkıştırma atağı uygulanmış sahte seslere ait spektrogramlarda duyarlılık metriği ise 0.88 olarak ölçülmüştür. 32 kbps sıkıştırma ataklarında ise duyarlılık değeri 0.79 olarak belirlenmiştir.

Tablo 27. TIMIT veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	0.73	0.92	0.81
Medyan Filtre	0.72	0.88	0.79
20 dB Gürültü Ekleme	0.66	0.69	0.67
30 dB Gürültü Ekleme	0.72	0.89	0.79
32 kbps Sıkıştırma	0.70	0.79	0.74
64 kbps Sıkıştırma	0.72	0.88	0.79

Arabic veri seti üzerinde yapılan çalışmanın sonuçları Tablo 28'te sunulmuştur. 30 dB gürültü eklenmiş ve medyan filtre atağı uygulanmış spektrogramlar için 0.79 F-skor değeri elde edilmiştir. En yüksek F-skor değeri, 0.81 ile ataksız sahte seslerde ulaşılmıştır. 64 kbps sıkıştırma ve 20 dB gürültü atakları uygulanmış spektrogramlarda sırasıyla 0.74 ve 0.67 F-skor değeri belirlenmiştir.

Duyarlılık metriği en yüksek ataksız sahte spektrogram ile 0.92 olarak ölçülürken, en düşük seviyeyi 0.69 ile 20 dB gürültü atağı altında görmüştür. Bunların yanı sıra medyan filtre, 30 dB gürültü ve 64 kbps sıkıştırma ataklarında sırasıyla 0.88, 0.89 ve 0.79 duyarlılık metriği ölçülmüştür.

Kesinlik metriği ise ataksız sahte spektrogramda en yüksek seviyeyi 0.73 ile alırken, en düşük skora 0.66 ile 20 dB gürültülü spektrogramlarda görülmüştür. Öte yandan 64 kbps sıkıştırma atağı altında 0.70 seviyesindedir. Medyan filtre ve 30 dB gürültü ataklarında 0.72 olarak gözlemlenmiştir.

Tablo 28. Arabic veriseti kullanılarak oluşturulmuş veriseti üzerinde yöntemin deneysel sonuçları

	Kesinlik	Duyarlılık	F-skor
Ataksız Sahte Ses	0.73	0.92	0.81
Medyan Filtre	0.72	0.88	0.79
20 dB Gürültü Ekleme	0.66	0.69	0.67
30 dB Gürültü Ekleme	0.72	0.89	0.79
64 kbps Sıkıştırma	0.70	0.79	0.74

3.2.8. Önerilen Yöntemlerin Literatür ile Karşılaştırılması

Bu bölümde tez kapsamında gerçekleştirilen anahtar noktalarının kümelendirilerek kopyala yapıştır ses sahteciliği tespiti çalışmalarının birbirleri arasında ve literatürdeki benzer çalışmalarla karşılaştırılması detaylandırılacaktır.

Duyarlılık metriği kullanılarak 6 farklı yöntemin TIMIT veriseti üzerindeki sonuçları Tablo 29’da verilmektedir. BRIEF tanımlayıcı vektörlerinin OPTICS algoritmasıyla kümelendirilmesi, tez kapsamında yapılan çalışmalarda tüm durumlar arasında en yüksek metrik değerini yakalamıştır. Ataksız sahte seslerde en düşük metrik değeri 0.91 ile SIFT ve OPTICS kombinasyonu kullanılarak ulaşılmıştır. Öte yandan AKAZE anahtar noktalarının OPTICS ile kümelendirilmesi 0.93 puan ile BRIEF OPTICS kombinasyonuna metrik değeri bakımından çok yaklaşmıştır. Bunların yanı sıra SURF anahtar noktalarının OPTICS ile, BRIEF’in g2NN ve k-Means ile kümelendirilmesi sonucunda 0.92 metrik değeri elde edilmiştir.

Tablo 29. TIMIT veriseti üzerinde 6 farklı yöntemin deneysel sonuçları

Anahtar Nokta Çıkarım	BRIEF	AKAZE	SIFT	SURF	BRIEF	
Türler\Kümeleme	OPTICS				g2NN	k-Means
Ataksız Sahte Ses	0.94	0.93	0.91	0.92	0.92	0.92
Medyan Filtre	0.92	0.9	0.84	0.89	0.9	0.88
20 dB Gürültü Ekleme	0.82	0.79	0.64	0.65	0.8	0.69
30 dB Gürültü Ekleme	0.93	0.89	0.85	0.85	0.92	0.89
32 kbps Sıkıştırma	0.89	0.82	0.66	0.7	0.86	0.79
64 kbps Sıkıştırma	0.93	0.89	0.79	0.84	0.93	0.88

Medyan filtre atağı uygulanmış sahte seslere ait spektrogram görsellerinde en düşük metrik değeri 0.84 seviyesinde SIFT anahtar noktalarının OPTICS kümelendirilmesi yöntemi ile elde edilmiştir. AKAZE'nin OPTICS ile ve BRIEF'in g2NN ile kümelendirilmesi sonucu 0.90 metrik değeri gözlemlenmiştir. Bunların yanı sıra SIFT ve OPTICS, SURF ve OPTICS, BRIEF ve k-Means kombinasyonları sırasıyla 0.84, 0.89 ve 0.88 metrik değerlerine sahiptir.

20 dB gürültü atağı altında en düşük performans SIFT ve OPTICS'in birlikte kullanıldığı yöntemle 0.64 olarak alınmıştır. Bu skoru SURF ve OPTICS kombinasyonu 0.65 ve BRIEF ve k-Means kombinasyonu 0.69 ile yakından takip etmektedir. Öte yandan AKAZE ve OPTICS, BRIEF ve g2NN sırasıyla 0.79 ve 0.80 metrik değerlerine sahiptir. 30 dB gürültü uygulanmış konuşma kayıtlarına ait en düşük metrik değerini SIFT ve OPTICS, SURF ve OPTICS çalışmaları 0.85 ile aynı puanı paylaşmaktadır. Öte yandan AKAZE ve OPTICS ile BRIEF ve k-Means 0.89 metrik değerine sahiptir. BRIEF ve g2NN işbirliği 0.92 ile en yüksek skora sahip yöntemi yakından takip etmektedir.

32 kbps sıkıştırma atağı altında en düşük performans 0.66 metrik değeriyle SIFT ve OPTICS işbirliğinde alınırken, bunu 0.7 ile SURF ve OPTICS takip etmektedir. Öte yandan BRIEF ve k-Means, AKAZE ve OPTICS, BRIEF ve g2NN sırasıyla 0.79, 0.82 ve 0.86 metrik değerleri almıştır. 64 kbps sıkıştırma uygulanmış konuşma kayıtlarına ait spektrogramlar üzerinde en yüksek metrik değeri ünvanını BRIEF ve OPTICS, BRIEF ve g2NN 0.93 ile korumaktadır. En düşük skor ise SIFT ve OPTICS işbirliğinin kullanıldığı çalışmada gözlemlenmiştir. Bunlara ek olarak SURF ve OPTICS, AKAZE ve OPTICS, BRIEF ve k-Means sırasıyla 0.84, 0.89 ve 0.88 metrik değerleri göstermiştir.

BRIEF ve OPTICS kombinasyonunun birlikte kullanıldığı çalışma tez kapsamında gerçekleştirilen diğer yöntemlerden daha yüksek duyarlılık skoruna sahip olduğu için literatürdeki benzer çalışmalarla kıyaslanması için tercih edilmektedir. Yöntemin literatürdeki benzer çalışmalarla karşılaştırılması duyarlılık metriği kullanılarak Tablo 30'da verilmektedir. Ataksız sahte seslerde en yüksek skora 0.94 ile önerdiğimiz metodla ulaşılırken, en düşük skor 0.18 puanla İmran ve arkadaşlarının Yerel İkili Örüntü operatörüne dayanan yöntemiyle elde edilmiştir. Öte yandan, 0.82 duyarlılık metriği Üstübioğlu ve arkadaşlarının Mel Spektrogram görüntülerinden elde ettikleri SIFT anahtar noktalarının eşleştirilmesi yöntemiyle alınmıştır. Ataklı seslerde, en yüksek duyarlılık metriği 30 dB gürültü ekleme ve 64 kbps sıkıştırma altında 0.93 puandır. 30 dB gürültü atağı altında en düşük skor 0.16 ile Huang ve arkadaşlarının Ayırık Fourier dönüşümü ve Mel-

Frekans Kepstral Katsayıları özelliklerine dayanan yöntemiyle alınmıştır. 64 kbps atağında en düşük skor yine aynı yazarlara aittir. Medyan filtre atağı uygulanmış konuşma kayıtlarında, yöntemimiz 0.92 duyarlılık metriğine sahiptir. En düşük skor ise 0.1 ile İmran ve arkadaşlarının çalışmasına aittir. 20 dB gürültü atağı altında, yöntemimiz 0.82 skoruyla benzer çalışmalardan daha başarılı iken, en düşük skor Huang ve arkadaşlarının önerdikleri prosedürle alınmıştır. İmran ve arkadaşları ile Huang ve arkadaşları 0.2 skoruyla 32 kbps sıkıştırma ataklarında aynı başarıya sahip olmuşlardır. Önerilen yöntem ise bu atakta 0.89 metrik değerine sahiptir.

Tablo 30. TIMIT veriseti üzerinde Duyarlılık metriği kullanılarak BRIEF ve OPTICS'in beraber kullanıldığı yöntemin literatür sonuçları

	Önerilen	Üstübioğlu vd., 2023	İmran vd., 2017	Huang vd., 2020	Yan vd., 2019	Wang vd., 2017
Ataksız Sahte Ses	0.94	0.82	0.18	0.19	0.23	0.22
Medyan Filtre	0.92	0.8	0.1	0.21	0.28	0.22
20 dB Gürültü Ekleme	0.82	0.74	0.17	0.1	0.37	0.25
30 dB Gürültü Ekleme	0.93	0.78	0.19	0.16	0.31	0.29
32 kbps Sıkıştırma	0.89	0.73	0.2	0.2	0.31	0.29
64 kbps Sıkıştırma	0.93	0.75	0.2	0.19	0.3	0.3

Arabic veriseti üzerinde tez çalışması kapsamında gerçekleştirilen 6 farklı yöntemin duyarlılık metriği skorları Tablo 31'de verilmektedir. Tüm atak türleri altında SURF anahtar noktalarının OPTICS yaklaşımıyla kümelendirilmesi metodu en yüksek duyarlılık metriği skoruna sahip olmaktadır. Ataksız sahte seslere ait spektrogramlar üzerinde AKAZE ve OPTICS kombinasyonu 0.78 ile en düşük skora sahip olmuştur. Bunu 0.86 ile BRIEF OPTICS kombinasyonu takip etmektedir. En yüksek metrik değerine 0.96 ile SURF ve OPTICS işbirliği ile ulaşılırken, BRIEF ve k-Means çalışması 0.92 metrik değeri seviyelerindedir. Öte yandan SIFT ve OPTICS, BRIEF ve g2NN yöntemleri 0.90 ile aynı duyarlılık metriği puanını paylaşmaktadır.

Medyan filtre uygulanmış sahte seslerde en yüksek skor 0.97 ile SURF ve OPTICS kombinasyonu ile elde edilmiştir. Bunu SIFT ve OPTICS yöntemi 0.96 ile yakından takip etmektedir. BRIEF ve k-Means prosedürü 0.88 metrik puanına sahip olmakta, buna en yakın

seviyede BRIEF ve OPTICS çalışması bulunmaktadır. En düşük skora 0.78 ile AKAZE ve OPTICS birleşimi ulaşmakta, BRIEF ve g2NN'in birlikte kullanımı 0.90 metrik değerine sahip olmaktadır.

20 dB gürültü uygulanmış sahte seslere ait spektrogramlar üzerinde en düşük duyarlılık metriği değeri, BRIEF ve k-Means kombinasyonu ile elde edilmiştir. Bunun yanı sıra, BRIEF ve OPTICS ile SIFT ve OPTICS kombinasyonları 0.86 ile aynı seviyededir. Öte yandan, AKAZE ve OPTICS ile BRIEF ve g2NN kombinasyonları sırasıyla 0.77 ve 0.90 metrik değerlerine sahiptir. En yüksek seviyeye ise SURF ve OPTICS kombinasyonu ile ulaşılmıştır.

Tablo 31. Arabic veriseti üzerinde 6 farklı yöntemin deneysel sonuçları

Anahtar Nokta Çıkarım	BRIEF	AKAZE	SIFT	SURF	BRIEF	
Türler\Kümeleme	OPTICS				g2NN	k-Means
Ataksız Sahte Ses	0.86	0.78	0.9	0.96	0.9	0.92
Medyan Filtre	0.86	0.78	0.96	0.97	0.9	0.88
20 dB Gürültü Ekleme	0.86	0.77	0.86	0.93	0.9	0.69
30 dB Gürültü Ekleme	0.85	0.78	0.9	0.96	0.9	0.89
64 kbps Sıkıştırma	0.86	0.77	0.86	0.93	0.9	0.79

30 dB gürültü atağına ait spektrogramlar üzerinde en düşük duyarlılık metriği değeri, AKAZE ve OPTICS kombinasyonu ile elde edilmiştir. Bunun yanı sıra, BRIEF ve g2NN ile SIFT ve OPTICS kombinasyonları 0.90 ile aynı seviyededir. Öte yandan, BRIEF ve OPTICS ile BRIEF ve k-Means kombinasyonları sırasıyla 0.85 ve 0.89 metrik değerlerine sahiptir. En yüksek seviyeye ise 0.96 ile SURF ve OPTICS kombinasyonu ile ulaşılmıştır.

64 kbps sıkıştırılmış konuşma kayıtlarından oluşturulmuş spektrogramlar üzerinde BRIEF ve OPTICS, SIFT ve OPTICS 0.86 ile aynı metrik değerine sahiptir. En düşük skora AKAZE ve OPTICS kombinasyonu ile ulaşmakta, bu puana en yakın 0.79 ile BRIEF ve k-Means birleşimi bulunmaktadır. Bu istatistiklere ek olarak BRIEF ve g2NN çalışması 0.90 metrik değerine sahiptir. En yüksek seviyeye SURF ve OPTICS kombinasyonu ile ulaşılmıştır.

3.3. Graf Tabanlı Yöntemin DeneySEL Sonuçları

Bu bölümde, tez çalışması kapsamında gerçekleştirilen konuşma kayıtlarından elde edilen graf görüntülerinin yapay sinir ağı yöntemleriyle sınıflandırılarak sahte ve orijinal olarak etiketlendiği çalışmanın performans sonuçları incelenecektir.

3.3.1. Kullanılan Verisetleri

Graf tabanlı çalışmada kullanılan verisetleri eğitim ve test olarak 2'ye ayrılmaktadır. Eğitim seti Arabic veriseti kullanılarak oluşturulmuştur. Test seti ise Arabic, TIMIT, Librispeech, Chinspeech ve tez çalışması kapsamında geliştirdiğimiz Türkçe verisetinden oluşmaktadır.

3.3.1.1. Arabic Artırımsız Eğitim Seti

Graf tabanlı yöntemimiz anahtar noktalarının kümelendirilerek kopyala yapıştır ses sahteciliği yapan çalışmanın aksine modellerin eğitilmesi amacıyla eğitim setine de ihtiyaç duymaktadır. Arabic veriseti kullanılarak oluşturulmuş söz konusu veri seti Tablo 32'de ifade edilmiştir.

Tablo 32. Graf tabanlı yöntem için oluşturulmuş Arabic kopyala yapıştır eğitim veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	1500	1500
Sahte	Ataksız	1000	1200
	Medyan Filtre	40	
	20dB gürültü	40	
	30dB gürültü	40	
	32kbps sıkıştırma	40	
	64kbps sıkıştırma	40	

Tablo 32'de görüldüğü üzere, orijinal ve sahte ses sayısı sırasıyla 1500 ve 1200 adettir. Sahte konuşma kayıtları, ataksız ve medyan filtre, 20 dB ve 30 dB gürültü, 32 kbps ve 64 kbps sıkıştırma atakları uygulanmış verilerden oluşmaktadır. Kopyala yapıştır sahteciliği ile oluşturulmuş 40 sese beş farklı atak uygulanarak 200 ataklı sahte ses elde edilmiştir. Bu oluşturulan eğitim seti, artırımsız eğitim seti olarak adlandırılmaktadır.

3.3.1.2. Arabic Artırmalı Eğitim Seti

Artırım yöntemleri oluşturulan eğitim setine uygulanarak veriseti 3 kat artırılmıştır. Oluşturulan yeni eğitim seti Tablo 33'te detaylandırılmıştır.

Tablo 33. Graf tabanlı yöntem için oluşturulmuş Artırmalı Arabic kopyala yapıştır eğitim veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	4500	4500
Sahte	Ataksız	3000	3600
	Medyan Filtre	120	
	20dB gürültü	120	
	30dB gürültü	120	
	32kbps sıkıştırma	120	
	64kbps sıkıştırma	120	

Veri artırım yöntemi görüntü bazındadır ve graf görüntülerine uygulanmaktadır. Tablo 32'te verilen sayılar artırılmış veri içindir. Sahte ve orijinal graf görüntü sayısı arasındaki oran korunmuş, sadece 2 sınıf arasında ayrımın daha net yapılabilmesi için verisetindeki örnekler 3 katına çıkarılmıştır.

3.3.1.3. Arabic Test Seti

Veriseti 200 orijinal konuşma kaydı kullanılarak kopyala yapıştır sahteciliği oluşturulması sonucu oluşmaktadır. Tablo 34'te verilen medyan filtre, 20 dB ve 30 dB gürültü, 32 kbps ve 64 kbps sıkıştırma atakları bu seslere uygulanarak sahte veri sayısı 1200'e çıkarılmıştır.

Tablo 34. Graf tabanlı yöntem için oluşturulmuş Arabic kopyala yapıştır test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	200	200
Sahte	Ataksız	200	1200
	Medyan Filtre	200	
	20dB gürültü	200	
	30dB gürültü	200	
	32kbps sıkıştırma	200	
	64kbps sıkıştırma	200	

3.3.1.4. TIMIT Test Seti

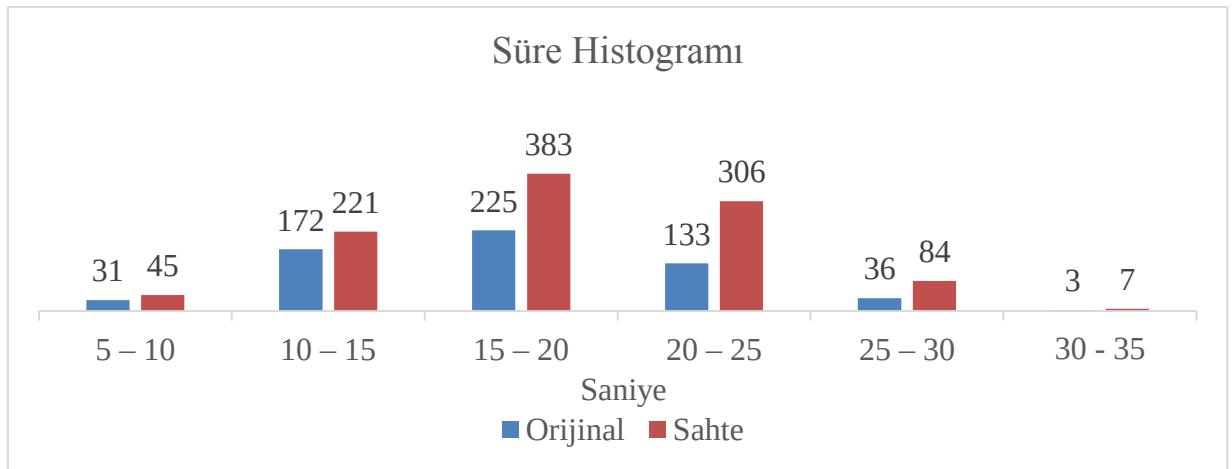
TIMIT veriseti kullanılarak gerçekleştirilen bu test seti 160 orijinal ve 960 sahte ses içermektedir. Veriseti Tablo 35’te detaylandırılmıştır. Ataksız sahte ses sayısı 160 tanedir ve bunların hepsine Tablo 35’te verilen 5 farklı saldırı yönteminin her biri uygulanmıştır.

Tablo 35. Graf tabanlı yöntem için oluşturulmuş TIMIT kopyala yapıştır test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	160	160
Sahte	Ataksız	160	960
	Medyan Filtre	160	
	20dB gürültü	160	
	30dB gürültü	160	
	32kbps sıkıştırma	160	
	64kbps sıkıştırma	160	

3.3.1.5. Türkçe Test Seti

Tübitak 122E013 projesi kapsamında literatüre Türkçe kopyala yapıştır sahteciliği veriseti kazandırılmıştır. Söz konusu veriseti günlük konuşma cümlelerini içeren orijinal konuşma kayıtları kullanılarak gerçekleştirilmiştir. Kopyala yapıştır sahteciliği sonrasında cümlelerin anlam bütünlüğü bozulmamıştır. Veri setinin saniye cinsinden histogramı Şekil 58’de verilmektedir. Şekil 58’de açıkça görülmektedir ki veri setindeki seslerin büyük kısmı 15-20 saniye aralığındaki uzunlukta seslerden oluşmaktadır.



Şekil 58. Türkçe verisetine ilişkin orijinal seslerin süre dağılımı

Tablo 36’da detaylı biçimde açıklanan Türkçe kopyala yapıştır veri seti, orijinal ve sahte ses kayıtlarından oluşmaktadır. Orijinal 600 konuşma kaydı, eşit sayıda kadın ve erkek konuşmacılardan alınmış olup, bu kayıtlar üç farklı ortamda (ofis, kantin ve sessiz oda) eşit sayıda gerçekleştirilmiştir. Oluşturulan sahte seslerin 348 tanesi ofis ortamından, 349’ar tanesi ise kantin ve sessiz oda ortamlarından elde edilmiştir. Graf tabanlı yöntemimizin performans sonuçlarının değerlendirilmesi amacıyla bu veri seti kullanılmıştır.

Tablo 36. Graf tabanlı yöntem için oluşturulmuş Türkçe kopyala yapıştır test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	600	600
Sahte	Ataksız	1046	6276
	Medyan Filtre	1046	
	20dB gürültü	1046	
	30dB gürültü	1046	
	32kbps sıkıştırma	1046	
	64kbps sıkıştırma	1046	

3.3.1.6. Librispeech Test Seti

İngilizce konuşma kayıtları içeren Librispeech kullanılarak oluşturulan 3 farklı veriseti tez çalışması kapsamında kullanılmaktadır. İlk veriseri Su ve arkadaşlarının gerçekleştirdikleri çalışmada oluşturulmuştur (Su vd., 2023). Bu veriseti 500 orijinal ve 500 sahte konuşma kayıtlarının yanında saldırı uygulanmış sahteciliklere de sahiptir. Söz konusu veriseti Tablo 37’de detaylandırılmıştır. Kopyala yapıştır işlemi ile oluşturulan sahte seslerin her birine Tablo 36’te verilen 128 kbps sıkıştırma, 20 dB ve 30 dB gürültü ekleme ve yeniden örnekleme atakları uygulanarak toplam sahte ses sayısı 2500’ e çıkarılmıştır. Yeniden örnekleme saldırısı sesin önce 8 kHz sonra 16 kHz ile örneklenmesi ile gerçekleştirilmiştir.

Tablo 37. Graf tabanlı yöntemin test edilmesinde kullanılan Librispeech V1 test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	500	500
Sahte	Ataksız	500	2500
	8kHz ve 16kHz Örnekleme	500	
	20dB gürültü	500	
	30dB gürültü	500	
	128kbps sıkıştırma	500	

İkinci Librispeech veriseti Su ve arkadaşlarının önerdiği yöntemde gerçekleştirilmiştir. (Su vd., 2023). Tablo 38’de detaylandırılan bu veri seti, 500 orijinal ve 500 sahte konuşma

kaydının yanı sıra, çeşitli saldırılara maruz kalmış sahtecilikleri de içermektedir. Kopyala yapıştır yöntemiyle oluşturulan sahte seslerin her birine 128 kbps sıkıştırma, 10 dB ve 20 dB gürültü ekleme ve yeniden örnekleme saldırıları uygulanmıştır. Bu işlemler sonucunda toplam sahte ses sayısı 2500'e ulaşmıştır. Yeniden örnekleme saldırısı, sesin önce 8 kHz'de, ardından 16 kHz'de örneklenmesi yoluyla gerçekleştirilmiştir.

Tablo 38. Graf tabanlı yöntemin test edilmesinde kullanılan Librispeech V2 test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	500	500
Sahte	Ataksız	500	2500
	8kHz ve 16kHz Örnekleme	500	
	10dB gürültü	500	
	20dB gürültü	500	
	128kbps sıkıştırma	500	

Üçüncü Librispeech veriseti Su ve arkadaşlarının yine aynı çalışmasında oluşturulmuştur. (Su vd., 2023). Bu veriseti sadece sahte seslerden oluşmaktadır. Diğerlerinden en büyük farkı sahtecilik segmentlerinin 0.1, 0.2, 0.3 ve 0.4 saniye ile sınırlandırılıp buna göre ayrılmasıdır. Tablo 39'da detayları verilen verisetinde toplam sahte ses sayısı 800 adettir.

Tablo 39. Graf tabanlı yöntemin test edilmesinde kullanılan Librispeech V3 test veriseti özeti

Sınıf	Sahte segment uzunluğu	Ses Sayısı	Toplam Ses Sayısı
Sahte	0.1	200	800
	0.2	200	
	0.3	200	
	0.4	200	

3.3.1.7. Chinspeech Test Seti

Chinspeech, internetten toplanan sesli kitaplardan türetilmiş Çin külliyatından oluşturulan verisetidir. Bu tez kapsamında 3 farklı versiyonu kullanılmaktadır. İlk veri seti, Su ve arkadaşlarının (2023) gerçekleştirdiği çalışma kapsamında oluşturulmuştur. Bu veri seti, 500 orijinal ve 500 sahte konuşma kaydının yanı sıra, saldırıya uğramış sahtecilikleri de içermektedir. Detaylar Tablo 40'da sunulmuştur. Kopyala yapıştır yöntemiyle oluşturulan sahte seslerin her birine, Tablo 40'da belirtilen 128 kbps sıkıştırma, 20 dB ve 30 dB gürültü ekleme ve yeniden örnekleme saldırıları uygulanarak toplamda 2500 sahte ses elde

edilmiştir. Yeniden örnekleme saldırısı, sesin önce 8 kHz, ardından 16 kHz ile örneklenmesi suretiyle gerçekleştirilmiştir.

Tablo 40. Graf tabanlı yöntemin test edilmesinde kullanılan Chinspeech V1 test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	500	500
Sahte	Ataksız	500	2500
	8kHz ve 16kHz Örnekleme	500	
	20dB gürültü	500	
	30dB gürültü	500	
	128kbps sıkıştırma	500	

İkinci Chinspeech veri seti, Su ve arkadaşlarının (2023) önerdiği yöntem kullanılarak oluşturulmuştur. Tablo 41’de ayrıntılı şekilde açıklanan bu veri seti, 500 orijinal ve 500 sahte konuşma kaydının yanı sıra, çeşitli saldırılara maruz kalmış sahtecilikleri de içermektedir. Kopyala yapıştır yöntemiyle oluşturulan sahte seslerin her birine 128 kbps sıkıştırma, 10 dB ve 20 dB gürültü ekleme ve yeniden örnekleme saldırıları uygulanmıştır. Bu işlemler sonucunda toplamda 2500 sahte ses elde edilmiştir. Yeniden örnekleme saldırısı, sesin önce 8 kHz, ardından 16 kHz ile örneklenmesi suretiyle gerçekleştirilmiştir.

Tablo 41. Graf tabanlı yöntemin test edilmesinde kullanılan Chinspeech V2 test veriseti özeti

Sınıf	Atak	Ses Sayısı	Toplam Ses Sayısı
Orijinal	-	500	500
Sahte	Ataksız	500	2500
	8kHz ve 16kHz Örnekleme	500	
	10dB gürültü	500	
	20dB gürültü	500	
	128kbps sıkıştırma	500	

Üçüncü Chinspeech veri seti, Su ve arkadaşlarının aynı çalışmasında oluşturulmuştur (Su vd., 2023). Bu veri seti yalnızca sahte ses kayıtlarından oluşmaktadır. Diğer veri setlerinden en büyük farkı, sahtecilik segmentlerinin 0.1, 0.2, 0.3 ve 0.4 saniye ile sınırlandırılarak bu kriterlere göre ayrılmış olmasıdır. Tablo 42’de ayrıntıları verilen bu veri setinde toplam 800 sahte ses bulunmaktadır.

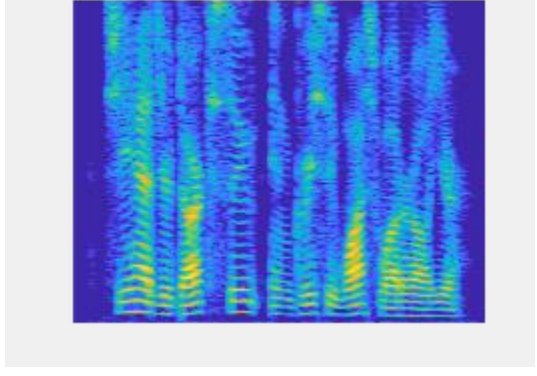
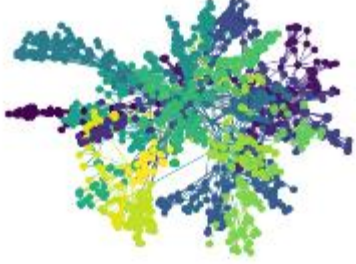
Tablo 42. Graf tabanlı yöntemin test edilmesinde kullanılan Chinspeech V3 test veriseti özeti

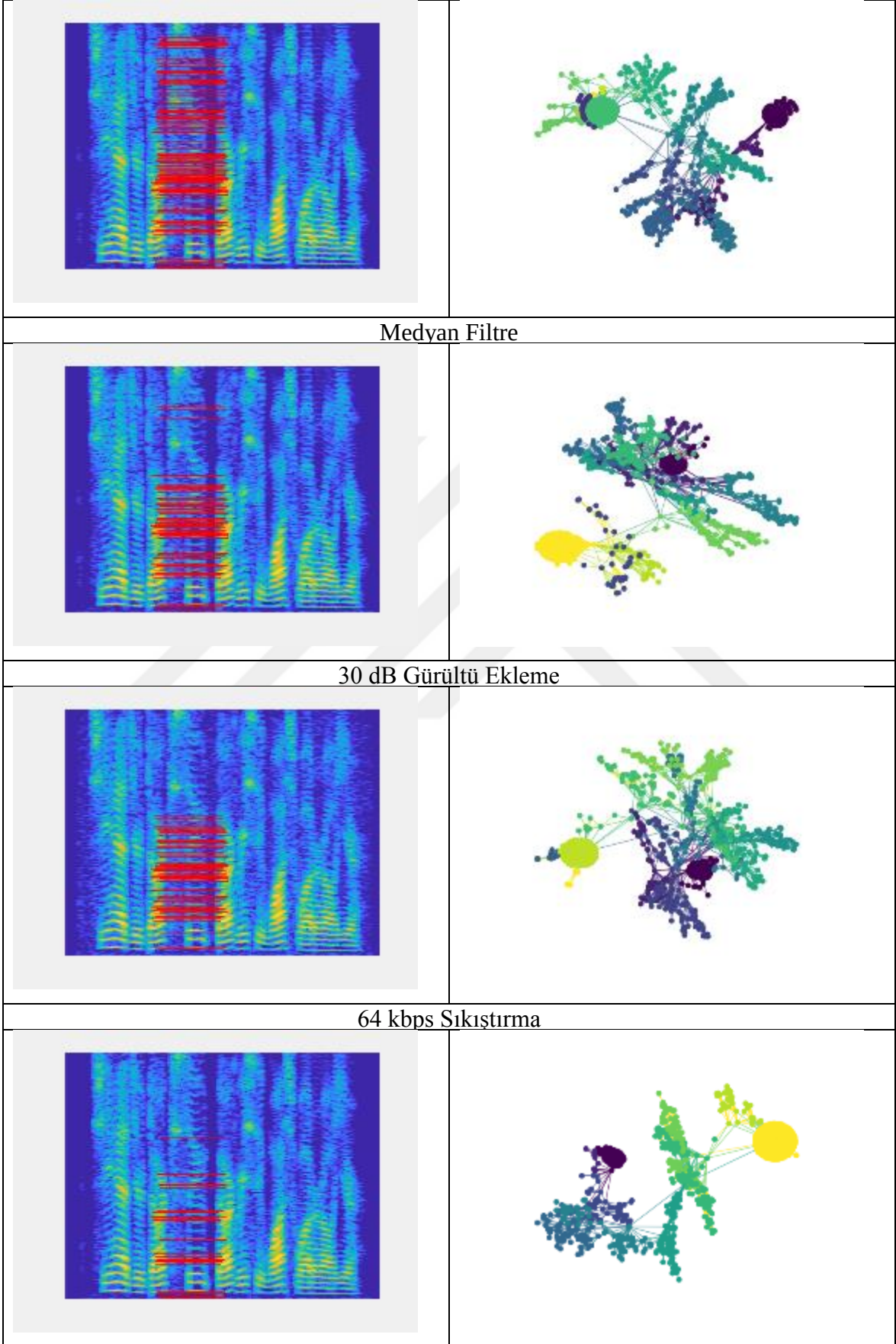
Sınıf	Sahte segment uzunluğu	Ses Sayısı	Toplam Ses Sayısı
Sahte	0.1	200	800
	0.2	200	
	0.3	200	
	0.4	200	

3.3.2. Graf Tabanlı Yöntemin Performans Sonuçları

Konuşma kayıtlarından SIFT anahtar noktalarının çıkarılarak g2NN ile kümelendirilmesi ve bu işlem sonrasında yoğun eşleşme bölgelerinin belirlenerek frekans eksenine izdüşürülmesi sonucu band geçiren süzgeç uygulanması ve graf görüntülerinin çıkarılması aşamalarına dayanan yöntemimizin farklı verisetleri altında deneysel sonuçları bu bölümde açıklanmaktadır. Tablo 43’de, örnek sesin, bu sesle oluşturulmuş sahte sesin ve çeşitli saldırılara maruz kalmış hallerinin yüksek çözünürlüklü spektrogram görüntüleri üzerinde eşleşme doğruları ve grafları gösterilmektedir. Görüntünün sol tarafında yüksek çözünürlüklü spektrogramlar yer alırken, sağ tarafında ise bu spektrogramlara ait graf görselleri bulunmaktadır. Örnek seslerden elde edilen yüksek çözünürlüklü spektrogram görsellerinin analizi sonucunda orijinal seste eşleşme bulunamadığı görülmekte, sahte verilerde ise oldukça fazla eşleşme tespit edildiği saptanmaktadır. Tablo 43’de görüleceği üzere orijinal ve sahte seslere ait graf resimleri birbirinden farklı özelliklere sahiptir. Tez kapsamında kullanılan mimarilerden bu farklılığı yakalaması beklenmektedir.

Tablo 43. Yüksek çözünürlüklü spektrogram görüntüleri üzerinde eşleşme doğruları ve graf görselleri

Orijinal	
	
Ataksız	



3.3.2.1. Artırımsız Veriseti Kullanılarak Eğitilen Mimarilerin Deneysel Sonuçları

Arabic veriseti kullanılarak oluşturulan eğitim seti ile kullanılan mimarilerin eğitimi gerçekleştirilmiştir. Tez çalışması kapsamında kullanılan farklı dilde ve özelliklerde test setleri ile mimarilerin performans sonuçları bu bölümde incelenmektedir. Arabic test seti üzerinde Füzyon yapısının performans sonuçları Tablo 44'te verilmektedir.

En yüksek duyarlılık metriği 0.96 ile ataksız ve medyan filtre atağı altında alınmakta, en düşük skora 0.815 ile 32 kbps sıkıştırılmış sahte seslere ait graf görüntüleriyle ulaşılmaktadır. Kesinlik metriği ataksız sahte sesler ve tüm saldırı durumlarında 0.96 civarında seyretmektedir. Doğruluk metriğinde en yüksek seviyeyi 0.965 puanla ataksız sahte sesler ve medyan filtre saldırısı beraber paylaşmaktadır. Öte yandan en düşük skor 0.8925 ile 32 kbps sıkıştırma saldırısına maruz kalmış verilerle alınmıştır. Diğer metriklerdeki gibi en yüksek F-skor değeri ataksız sahte ve medyan filtre saldırısı altında görülmektedir. Gürültü saldırısında farklı parametrelerin kullanımı Arabic test seti performans sonuçları üzerinde çok fazla etkisinin olmadığı Tablo 44'de gözükmektedir. Buna karşın sıkıştırma işlemlerinde aynı durum söz konusu değildir. 32 kbps sıkıştırma işlemi daha düşük duyarlılık, doğruluk ve F-skor metriklerine sebep olmaktadır.

Tablo 44. Arabic test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.96	0.9697	0.965	0.9648
Medyan Filtre	0.96	0.9697	0.965	0.9648
20 dB Gürültü Ekleme	0.935	0.9689	0.9525	0.9517
30 dB Gürültü Ekleme	0.93	0.9688	0.95	0.949
32 kbps Sıkıştırma	0.815	0.9645	0.8925	0.8835
64 kbps Sıkıştırma	0.905	0.9679	0.9375	0.9354

TIMIT test seti üzerinde Füzyon yapısının performans sonuçları Tablo 45'te verilmektedir. En yüksek duyarlılık metriği, 0.99 ile saldırısız sahte sesler altında elde edilmiştir. En düşük skor ise, 0.8063 ile 32 kbps sıkıştırılmış sahte seslerin graf görüntülerinde gözlenmiştir. Kesinlik metriği de benzer eğilim göstermiş, en yüksek değer 0.8641 olarak kaydedilmiştir. En düşük kesinlik değeri ise 0.8377 ile 32 kbps sıkıştırma saldırısına maruz kalmış seslerde görülmüştür. Doğruluk metriğinde en yüksek puan, 0.9188 ile saldırısız sahte seslerde elde edilmiştir. Buna karşılık, en düşük doğruluk skoru, 0.825 ile 32 kbps sıkıştırma saldırısına uğramış verilerden alınmıştır. Diğer metriklerde olduğu gibi, en yüksek F-skoru da saldırısız sahte konuşma kayıtları altında kaydedilmiştir. TIMIT test

seti üzerinde yapılan analizlerde, farklı gürültü parametrelerinin duyarlılık, doğruluk ve F-skor üzerinde büyük farklılıklar oluşturduğu görülmüştür. Aynı durum, sıkıştırma işlemlerinde kullanılan farklı parametreler için de geçerlidir.

Tablo 45. TIMIT test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.9938	0.8641	0.9188	0.9244
Medyan Filtre	0.9625	0.8603	0.9031	0.9086
20 dB Gürültü Ekleme	0.8125	0.8387	0.8281	0.8254
30 dB Gürültü Ekleme	0.9438	0.858	0.8938	0.8988
32 kbps Sıkıştırma	0.8063	0.8377	0.825	0.8217
64 kbps Sıkıştırma	0.95	0.8588	0.8969	0.9021

Türkçe test seti üzerinde performans sonuçları Tablo 46’da verilmektedir. Duyarlılık metriğinin en yüksek ve en düşük olduğu durum sırasıyla 0.97992 ile medyan filtre saldırısı ve 0.87189 ile 20 dB gürültü ekleme atağıdır. Bunlara ek olarak ataksız, 30 dB gürültü ve 64 kbps sıkıştırma saldırıları altında 0.9761, 0.97228 ve 0.94646 duyarlılık metrikleri not edilmiştir. Kesinlik metriği Türkçe veriseti altında 0.89 civarında seyretmektedir. Doğruluk metriğinde duyarlılık skorları gibi en yüksek seviyeyi 0.91555 ile medyan filtre saldırısı altında görmekte, en düşük puanı 0.8469 ile 20 dB gürültü altında almaktadır. Bunu 0.84419 ile 32 kbps sıkıştırılmış sesler takip etmektedir. Öte yandan medyan filtre, 30 dB gürültü ve seslere herhangi atak uygulanmaması yaklaşık olarak 0.93 puanlık F-skor değerine sebep olmaktadır.

Tablo 46. Türkçe test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.9761	0.8964	0.91312	0.93455
Medyan Filtre	0.97992	0.89676	0.91555	0.9365
20 dB Gürültü Ekleme	0.87189	0.88544	0.8469	0.87861
30 dB Gürültü Ekleme	0.97228	0.89604	0.91069	0.9326
32 kbps Sıkıştırma	0.88337	0.88676	0.85419	0.88506
64 kbps Sıkıştırma	0.94646	0.8935	0.89429	0.91922

Librispeech kopyala yapıştır veriseti versiyon 1 üzerinde yöntemimizin performans sonuçları duyarlılık, kesinlik, doğruluk ve F-skor metrikleri kullanılarak Tablo 47’de verilmektedir. 20 dB gürültü atağı altında duyarlılık, doğruluk ve F-skor metrikleri 0.98

civarındadır. Buna karşın ataksız sahte seslerde duyarlılık metrik değeri 1’dir. Diğer durumlarda tüm metrik puanları 0.99 seviyesindedir. Tablodaki verilerin analizi sonucunda veriseti içerisindeki konuşma kayıtlarına saldırı uygulamak sistemin performansı üzerinde olumsuz etkiye sebep olmamaktadır.

Tablo 47. Librispeech V1 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skör
Ataksız Sahte Ses	1	0.9921	0.996	0.996
Yeniden Örnekleme	0.998	0.992	0.995	0.995
20 dB Gürültü Ekleme	0.98	0.9919	0.986	0.9859
30 dB Gürültü Ekleme	0.994	0.992	0.993	0.993
128 kbps Sıkıştırma	0.994	0.992	0.993	0.993

Librispeech kopyala yapıştır veri setinin ikinci versiyonu üzerinde yapılan yöntemimizin performans sonuçları Tablo 48’de duyarlılık, kesinlik, doğruluk ve F-skör metrikleri kullanılarak sunulmuştur. Ataksız sahte sesler ve 128 kbps sıkıştırma atağı durumlarında duyarlılık metriği 0.9920 ile en yüksek değere ulaşmıştır. Buna karşılık, 10 dB gürültü atağı uygulanmış sahte seslerin sınıflandırılması 0.5860 duyarlılık metriği ile en düşük değeri göstermiştir. Kesinlik metriği ise tüm durumlarda 0.98’in üzerinde seyretmektedir. Doğruluk metriği açısından, ataksız sahte sesler ve 128 kbps sıkıştırılmış sesler 0.9930 ile en yüksek puana sahiptir. Bununla birlikte, 10 dB gürültü altında doğruluk metriği 0.7892 ile en düşük seviyede kalmıştır. F-skör metriğinde de 10 dB gürültü altında alınan 0.7362 puanı en düşük değer olup, en yüksek F-skör ataksız ve 128 kbps sıkıştırma saldırısı altında elde edilmiştir.

Tablo 48. Librispeech V2 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skör
Ataksız Sahte Ses	0.9920	0.9940	0.9930	0.9930
Yeniden Örnekleme	0.9840	0.9939	0.9890	0.9889
10 dB Gürültü Ekleme	0.5860	0.9899	0.7892	0.7362
20 dB Gürültü Ekleme	0.8500	0.9930	0.9220	0.9159
128 kbps Sıkıştırma	0.9920	0.9940	0.9930	0.9930

Sadece ataksız sahte seslerden oluşan Librispeech versiyon üç verisetine ait artırimsız eğitim verisi ile eğitilmiş mimarilerin füzyonuna ait performans sonuçları Tablo 49’da verilmektedir. Kopyala yapıştır segmentlerinin 0.2 saniye ile sınırlandırıldığı 200 adet ses

için duyarlılık metriği 0.99 olarak ölçülmüştür. Buna karşın 0.3 ve 0.4 saniyelik segmentli konuşma kayıtları için sırasıyla 0.9850 ve 0.9750 duyarlılık metriği gözlemlenmiştir.

Tablo 49. Librispeech V3 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık
0.2	0.9900
0.3	0.9850
0.4	0.9750

Chinspeech kopyala yapıştır versiyon 1 veriseti üzerinde yöntemimizin performans sonuçları Tablo 50’de verilmektedir. En yüksek metrik değerlerinin alındığı 30 dB gürültü ve 128 kbps sıkıştırma atakları altında duyarlılık, kesinlik, doğruluk ve F-skor sırasıyla 1.0, 0.9728, 0.9860 ve 0.9862 seviyesindedir. Buna karşın en düşük metrik değerlerine 20 dB gürültü uygulanmış sahte seslerden elde edilmiş spektrogramlarla duyarlılık, kesinlik, doğruluk ve F-skor metrikleri aynı puanla 0.9720 değerine sahiptir. Öte yandan ataksız sahte sesler ve yeniden örnekleme atağı ile duyarlılık, kesinlik, doğruluk ve F-skor metrikleri sırasıyla 0.9980, 0.9727, 0.9850 ve 0.9852 civarındadır.

Tablo 50. Chinspeech V1 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.9980	0.9727	0.9850	0.9852
Yeniden Örnekleme	0.9980	0.9727	0.9850	0.9852
20 dB Gürültü Ekleme	0.9720	0.9720	0.9720	0.9720
30 dB Gürültü Ekleme	1.0000	0.9728	0.9860	0.9862
128 kbps Sıkıştırma	1.0000	0.9728	0.9860	0.9862

Tablo 51’de Chinspeech versiyon 2 veri setine ait artırısız veriseti ile eğitilmiş mimarilerin füzyon sonuçları 4 farklı metrik kullanılarak verilmektedir. Tüm metrikler bazında en yüksek skorlara yeniden örnekleme ve 128 kbps sıkıştırma atakları altında duyarlılık, kesinlik ve doğruluk sırasıyla 0.9920, 0.99 ve 0.9910 değerleri almakta ve F-skor puanında doğruluk ile aynı seviyede olmaktadır. Seslere 10 dB gürültü atağı uygulanması en küçük metrik değerlerine sahip olmaktadır. 20 dB gürültü atağı ve ataksız sahte seslerle duyarlılık, kesinlik, doğruluk ve F-skor 0.9780, 0.9899, 0.9840, 0.9839 metrik değerleri göstermiştir. Gürültü atakları yüksek çözünürlüklü spektrogram görüntülerinden elde edilen

anahtar noktalarını negatif olarak etkiledikleri için graf tabanlı yöntemin performansını olumsuz olarak etkilemektedir.

Tablo 51. Chinspeech V2 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skör
Ataksız Sahte Ses	0.9780	0.9899	0.9840	0.9839
Yeniden Örnekleme	0.9920	0.9900	0.9910	0.9910
10 dB Gürültü Ekleme	0.8340	0.9882	0.9120	0.9046
20 dB Gürültü Ekleme	0.9780	0.9899	0.9840	0.9839
128 kbps Sıkıştırma	0.9920	0.9900	0.9910	0.9910

Chinspeech versiyon 3 veriseti üzerinde performans sonuçları Tablo 52’de verilmektedir. Farklı uzunlukta kopyala yapıştır segment uzunluğuna sahip olması, bu veriseti çalışmayı farklı yönüyle değerlendirme imkanı vermektedir. En yüksek duyarlılık metriği 0.9950 ile 0.3 saniyelik segment uzunluğuna sahip seslerde alınmaktadır. Buna karşın en düşük seviyeyi çalışma 0.2 uzunluktaki segmentlerde görmüştür. Öte yandan 0.9700 ile 0.2 saniyelik segmentlerde almıştır.

Tablo 52. Chinspeech V3 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık
0.2	0.9700
0.3	0.9950
0.4	0.9900

3.3.2.2. Artırılmış Veriseti Kullanılarak Eğitilen Mimarilerin Deneysel Sonuçları

Modellerin öğrenme kapasitesini artırmak amacıyla eğitim verisetine veri artırma teknikleri uygulanmıştır. Bu işlem sonucunda, verisetindeki grafik sayısı 3 katına çıkmıştır. Oluşturulan bu genişletilmiş veriseti, tez çalışması kapsamında geliştirilen mimari, VGG16 ve VGG19 yapılarının eğitilmesinde kullanılmıştır. Bu üç mimariden elde edilen skorlar, füzyon işlemine tabi tutulmuştur. Bu bölümde, farklı verisetleri üzerinde gerçekleştirilen füzyon işleminin performans sonuçları değerlendirilmektedir.

Arabic test seti üzerinde yöntemin performans sonuçları Tablo 53’de sunulmuştur. Medyan filtre ve 30 dB gürültü saldırıları, tüm metrikler açısından en yüksek skorlara ulaşmıştır. Bu durumda duyarlılık, kesinlik, doğruluk ve F-skör sırasıyla 0.9750, 0.9653,

0.97 ve 0.9701 olarak kaydedilmiştir. En düşük puanlar ise 32 kbps sıkıştırılmış sahte sesler kullanılarak elde edilen graf görüntülerinde, kesinlik metriği hariç, gözlemlenmiştir. Bu durum için duyarlılık 0.82, doğruluk 0.8925 ve F-skor 0.8841 olarak raporlanmıştır. Ancak kesinlik metriği 0.9591 ile diğer durumlara oldukça yakın kalmıştır. Duyarlılık metriği, saldırı olmadan, 20 dB gürültü ve 64 kbps sıkıştırma saldırıları altında sırasıyla 0.97, 0.9450 ve 0.91 değerlerine ulaşmıştır. Kesinlik metriği ise 64 kbps sıkıştırma, 20 dB gürültü ve saldırısız sahte sesler altında sırasıyla 0.9630, 0.9643 ve 0.9652 olarak belirlenmiştir. Doğruluk metriği 64 kbps sıkıştırma, 20 dB gürültü ve saldırısız sahte sesler altında sırasıyla 0.9375, 0.9550 ve 0.9675 olarak raporlanmıştır. 20 dB gürültü 0.9545, saldırısız sahte seslerden alınan graf görüntüleri 0.9676 ve 64 kbps sıkıştırma durumları 0.9357 metrik değerlerine ulaşmıştır.

Tablo 53. Arabic test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.97	0.9652	0.9675	0.9676
Medyan Filtre	0.9750	0.9653	0.9700	0.9701
20 dB Gürültü Ekleme	0.9450	0.9643	0.9550	0.9545
30 dB Gürültü Ekleme	0.9750	0.9653	0.9700	0.9701
32 kbps Sıkıştırma	0.8200	0.9591	0.8925	0.8841
64 kbps Sıkıştırma	0.9100	0.9630	0.9375	0.9357

TIMIT test seti üzerindeki çalışmanın performans sonuçları Tablo 54'te sunulmaktadır. En düşük ve en yüksek metrik değerleri sırasıyla 32 kbps sıkıştırılmış sahte seslerde ve saldırısız sahte konuşma kayıtlarında elde edilmiştir. En yüksek metrik değerler duyarlılıkta 0.9938, kesinlikte 0.8413, doğrulukta 0.9031 ve F-skorunda 0.9112 olarak ölçülmüştür. Buna karşın, en düşük metrik değerler duyarlılıkta 0.8188, kesinlikte 0.8137, doğrulukta 0.8156 ve F-skorunda 0.8162 olarak görülmüştür. 20 dB gürültü saldırısı, tüm metriklerde 0.85 seviyesinin altında kalmıştır. Medyan filtre uygulanmış sahte konuşma kayıtlarında duyarlılık, kesinlik, F-skor ve doğruluk metrikleri sırasıyla 0.9563, 0.8361, 0.8921 ve 0.8844 olarak kaydedilmiştir. Sıkıştırma parametresi 64 kbps olarak tercih edildiğinde ise kesinlik 0.8378, doğruluk 0.8906, F-skor 0.8986 ve duyarlılık 0.9688 olarak not edilmiştir.

Tablo 54. TIMIT test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.9938	0.8413	0.9031	0.9112
Medyan Filtre	0.9563	0.8361	0.8844	0.8921
20 dB Gürültü Ekleme	0.8250	0.8148	0.8188	0.8199
30 dB Gürültü Ekleme	0.9438	0.8343	0.8781	0.8856
32 kbps Sıkıştırma	0.8188	0.8137	0.8156	0.8162
64 kbps Sıkıştırma	0.9688	0.8378	0.8906	0.8986

Türkçe test seti üzerinde yöntemin performans sonuçları Tablo 55'te ifade edilmiştir. Medyan filtre atağı ve ataksız sahte seslerden elde edilen graf görüntüleri ile yöntemimiz duyarlılık, kesinlik, doğruluk ve F-skor metrik değerleri sırasıyla 0.9790, 0.9078, 0.9235 ve 0.9420 değerlerine ulaşmıştır. Buna karşın en düşük 20 dB gürültü atağı altında 0.8815 duyarlılık, 0.8986 kesinlik, 0.8615 doğruluk ve 0.8899 F-skor metrik puanı seviyesindedir. En düşük skoru 32 kbps sıkıştırma atağı duyarlılık, kesinlik, doğruluk ve F-skor metrikleri sırasıyla 0.8872, 0.8992, 0.8651 ve 0.89317 skorları takip etmektedir. 64 kbps sıkıştırma saldırısı tüm metriklerde 0.90 ile 0.95 arasında seyretmektedir.

Tablo 55. Türkçe test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.9789	0.9078	0.92345	0.94204
Medyan Filtre	0.9790	0.9078	0.9235	0.94204
20 dB Gürültü Ekleme	0.8815	0.8986	0.8615	0.88996
30 dB Gürültü Ekleme	0.9751	0.9075	0.9210	0.94009
32 kbps Sıkıştırma	0.8872	0.8992	0.8651	0.89317
64 kbps Sıkıştırma	0.9551	0.9057	0.9083	0.92973

Librispeech kopyala yapıştır veriseti versiyon 1 üzerindeki yöntemimizin performans sonuçları, duyarlılık, kesinlik, doğruluk ve F-skor metrikleri kullanılarak Tablo 56'da sunulmuştur. Saldırızsız sahte seslerde duyarlılık, kesinlik ve F-skor metrikleri sırasıyla 1.0, 0.9881 ve 0.9940 olarak belirlenmiş olup, doğruluk metrikleri F-skor ile aynı değeri almıştır. Yeniden örnekleme, 30 dB gürültü ve 128 kbps sıkıştırma saldırıları altında ise duyarlılık 0.998, kesinlik 0.9881, doğruluk 0.9930 ve F-skor doğrulukla aynı puana ulaşmıştır. Ayrıca, 20 dB gürültü saldırısı durumunda duyarlılık 0.9820, kesinlik 0.9879 ve doğruluk 0.9850 olarak kaydedilmiş ve doğruluk metrikleri F-skorla aynı değere sahip olmuştur.

Tablo 56. Librispeech V1 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skör
Ataksız Sahte Ses	1.0000	0.9881	0.9940	0.994
Yeniden Örnekleme	0.998	0.9881	0.993	0.993
20 dB Gürültü Ekleme	0.9820	0.9879	0.9850	0.985
30 dB Gürültü Ekleme	0.9980	0.9881	0.9930	0.993
128 kbps Sıkıştırma	0.9980	0.9881	0.9930	0.993

Librispeech veri setinin kopyala-yapıştır verisi versiyon 2 üzerindeki yöntemimizin performans sonuçları, duyarlılık, kesinlik, doğruluk ve F-skör metrikleri kullanılarak Tablo 57’de sunulmuştur. Saldırızsız sahte ve 128 kbps sıkıştırma saldırısı durumlarında duyarlılık, kesinlik ve F-skör metrikleri sırasıyla 0.9940, 0.99 ve 0.9920 olarak belirlenmiştir. Doğruluk metrikleri ise F-skör ile aynı değeri almıştır. Yeniden örnekleme atağı tüm metrik değerlerinde 0.99 seviyesindedir. 10 dB gürültü saldırısı durumunda duyarlılık 0.5880, kesinlik 0.9833, doğruluk 0.7890 ve F-skör 0.7359 başarı kaydetmiştir. Ayrıca, 20 dB gürültü saldırısı durumunda duyarlılık 0.9940, kesinlik 0.99 ve doğruluk 0.9920 olarak kaydedilmiş ve doğruluk metrikleri F-skör ile aynı değere sahip olmuştur.

Tablo 57. Librispeech V2 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skör
Ataksız Sahte Ses	0.9940	0.9900	0.9920	0.992
Yeniden Örnekleme	0.99	0.99	0.99	0.99
10 dB Gürültü Ekleme	0.5880	0.9833	0.7890	0.7359
20 dB Gürültü Ekleme	0.8500	0.9884	0.9200	0.914
128 kbps Sıkıştırma	0.9940	0.9900	0.9920	0.992

Sadece ataksız sahte sesleri içeren Librispeech versiyon 3 verisetine ait artırısız eğitim verisiyle eğitilmiş mimarilerin füzyon performans sonuçları Tablo 58’de gösterilmektedir. 200 adet ses için kopyala yapıştır segmentleri 0.2 saniye ile sınırlandırıldığında, duyarlılık metriğı 0.995 olarak ölçölmüştür. Buna karşılık, 0.3 saniyelik segmentler için duyarlılık 0.99, 0.4 saniyelik segmentler için ise 0.98 olarak belirlenmiştir.

Tablo 58. Librispeech V3 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık
0.2	0.9950
0.3	0.9900
0.4	0.9800

Chinspeech kopyala yapıştır veriseti versiyon 1 üzerinde yöntemimizin performans sonuçları duyarlılık, kesinlik, doğruluk ve F-skör metrikleri kullanılarak Tablo 59’da verilmektedir. En yüksek başarıya yeniden örnekleme, 30 dB gürültü ve 128 kbps sıkıştırma atağı uygulanmış sahte seslere ait graf görselleriyle duyarlılık, kesinlik ve aynı skora sahip F-skör ile doğruluk metriklerinde sırasıyla 1, 0.9881 ve 0.9940 değerleri gözlemlenmiştir. Ataksız sahte seslerde bu metrikler aynı sıra ile 0.9980, 0.9881, 0.9930 skorlarına sahiptir. Aynı zamanda 20 dB gürültü atağı altında duyarlılık 0.9740, kesinlik 0.9878, doğruluk 0.9810 ve F-skör 0.9809 olarak not edilmiştir.

Tablo 59. Chinspeech V1 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skör
Ataksız Sahte Ses	0.9980	0.9881	0.9930	0.993
Yeniden Örnekleme	1	0.9881	0.994	0.994
20 dB Gürültü Ekleme	0.9740	0.9878	0.9810	0.9809
30 dB Gürültü Ekleme	1	0.9881	0.9940	0.994
128 kbps Sıkıştırma	1	0.9881	0.9940	0.994

Chinspeech kopyala yapıştır veriseti versiyon 2 üzerinde yöntemimizin performans sonuçları duyarlılık, kesinlik, doğruluk ve F-skör metrikleri kullanılarak Tablo 60’ta verilmektedir. En yüksek başarıya 128 kbps sıkıştırma atağı uygulanmış sahte seslere ait graf görselleriyle duyarlılık, kesinlik ve aynı skora sahip F-skör ile doğruluk metriklerinde sırasıyla 0.9960, 0.9920 ve 0.9940 değerleri gözlemlenmiştir. Bu skorları yeniden örnekleme saldırısı altında alınmış metrik değerleri 0.994 duyarlılık, 0.992 kesinlik, 0.993 doğruluk ve F-skör takip etmektedir. Ataksız sahte seslerle 0.99 duyarlılık, 0.992 kesinlik, 0.991 doğruluk ve F-skör not edilmiştir. 20 dB gürültü uygulanması durumunda duyarlılık, kesinlik, doğruluk ve F-skör metrikleri sırasıyla 0.98, 0.9919, 0.986 ve 0.9859 olarak gözlemlenmiştir. Bunların yanı sıra en düşük başarıya 10 dB gürültü uygulanmış sahte seslerle 0.8380 duyarlılık, 0.9905 kesinlik, 0.9150 doğruluk ve 0.9079 F-skör değeri alınmıştır.

Tablo 60. Chinspeech V2 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık	Kesinlik	Doğruluk	F-skor
Ataksız Sahte Ses	0.9900	0.9920	0.9910	0.991
Yeniden Örnekleme	0.994	0.992	0.993	0.993
10 dB Gürültü Ekleme	0.8380	0.9905	0.9150	0.9079
20 dB Gürültü Ekleme	0.98	0.9919	0.986	0.9859
128 kbps Sıkıştırma	0.9960	0.9920	0.9940	0.994

Sadece ataksız sahte sesleri içeren Chinspeech versiyon 3 verisetine dayalı artırımı eğitim verisiyle eğitilen mimarilerin füzyon performans sonuçları Tablo 61’de sunulmuştur. 200 adet ses için kopyala yapıştır segmentleri 0.3 ve 0.4 saniye ile sınırlandırıldığında duyarlılık metriği 0.9950 olarak kaydedilmiştir. Buna karşın, 0.3 saniyelik segmentler için duyarlılık 0.98 olarak ölçülmüştür.

Tablo 61. Chinspeech V3 test seti üzerinde Füzyon işleminin performans sonuçları

	Duyarlılık
0.2	0.9800
0.3	0.9950
0.4	0.9950

3.3.2.3. Yöntemlerin Birbirleriyle ve Literatür Karşılaştırılması

Bu bölümde, tez kapsamında gerçekleştirilen çalışmada, konuşma kayıtlarından graf görüntülerinin çıkarılması ve bu görüntülerin 3 farklı mimari ile sınıflandırılması sonucu yapılan füzyon işlemi ile kopyala yapıştır ses sahteciliği tespitinin sonuçları detaylandırılacaktır. Ayrıca, elde edilen bu sonuçlar, hem birbirleri arasında hem de literatürdeki benzer çalışmalarla karşılaştırılacaktır.

F-skor metriği kullanılarak eğitilmiş iki farklı verisiyle mimarilerin füzyon sonuçları, Arabic test veriseti üzerindeki performansları Tablo 62’de sunulmuştur. İlk sütun, verisetindeki konuşma kayıtlarındaki atak durumlarını temsil ederken, ikinci sütun artırımı eğitim verisiyle eğitilen mimarilerin füzyon sonuçlarına ilişkin F-skor performans sonuçlarını göstermektedir. Son sütunda ise artırımı veri ile eğitilmiş mimarilerin skorlarına yer verilmiştir.

Tablo 62'in analizi sonucunda verisetine artırım yöntemleri uygulamak performansın artmasına sebep olmaktadır. Ataksız sahte seslerde ve tüm saldırı şartları altında artırımı veriseti ile eğitilen mimariler daha yüksek F-skor değerleri göstermiştir.

Artırımı eğitim veriseti kullanıldığında en yüksek F-skor seviyesine medyan filtre ve 30 dB gürültü atağı altında 0.9701 metrik değeri kaydedilmiştir. Buna karşın en düşük skora 32 kbps sıkıştırma saldırısı koşulunda 0.8841 puanla ulaşılmıştır.

Tablo 62. Artırımı ve artırısız eğitim seti ile eğitilmiş mimarilerin Arabic test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Sei	Arabic Test Seti	
	Artırısız	Artırımı
Eğitim Veri Seti		
Ataksız Sahte Ses	0.9648	0.9676
Medyan Filtre	0.9648	0.9701
20 dB Gürültü Ekleme	0.9517	0.9545
30 dB Gürültü Ekleme	0.949	0.9701
32 kbps Sıkıştırma	0.8835	0.8841
64 kbps Sıkıştırma	0.9354	0.9357

En düşük metrik değerini Wang ve arkadaşlarının önerdiği Tekil Değer Ayırıştırma ve Ayrık Kosinüs Dönüşümüne dayanan ses kopyala-yapıştır sahteciliği ile elde edilmiştir. Öte yandan İmran ve arkadaşlarının Yerel İkili Örüntü operatörüne dayanan çalışması ataksız sahte, medyan filtre, 20 dB ve 30 dB gürültü, 32 kbps ve 64 kbps sıkıştırma saldırıları altında 0.3015, 0.2622, 0.5336, 0.4531, 0.3267 ve 0.2779 metrik değerlerini göstermiştir. Huang ve arkadaşlarının Ayrık Fourier dönüşümü ve Mel-Frekans Kepstral Katsayıları özelliklerine dayanan prosedürü ataksız sahte seslerde 0.4028, medyan filtre altında 0.4104, 20 dB gürültü atağında 0.4434, 30 dB gürültü saldırısında 0.4506, 32 kbps ve 64 kbps sıkıştırılmış seslerde sırasıyla 0.4215 ve 0.4178 metrik seviyesini kaydetmiştir. Perde ve Formant dizilerinin kullanıldığı Yan ve arkadaşlarının çalışmasında atak uygulanmamış sahte seslerde 0.2289, medyan filtre saldırısı 0.2244, 20 dB ve 30 dB gürültü ataklarında sırasıyla 0.5011 ve 0.4072, 32 kbps ve 64 sıkıştırma durumunda 0.2634 ve 0.2420 metrik değerlerini not edilmiştir.

Tablo 63. Artırılmış eğitim seti ile eğitilmiş mimarının Arabic test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri

	Önerilen	İmran vd., 2017	Huang vd., 2020	Yan vd., 2019	Wang vd., 2017
Ataksız Sahte Ses	0.9676	0.3015	0.4028	0.2289	0.1141
Medyan Filtre	0.9701	0.2622	0.4104	0.2244	0.1141
20 dB Gürültü Ekleme	0.9545	0.5336	0.4434	0.5011	0.1141
30 dB Gürültü Ekleme	0.9701	0.4531	0.4506	0.4072	0.1141
32 kbps Sıkıştırma	0.8841	0.3267	0.4215	0.2634	0.1212
64 kbps Sıkıştırma	0.9357	0.2779	0.4178	0.2420	0.1212

Artırılmış ve artırılmış eğitim setleri kullanılarak eğitilen mimarilerin TIMIT test seti üzerinde füzyon sonuçları Tablo 64’te gösterilmektedir. Söz konusu tablonun analizi sonucunda F-skor metrik değerlerinin birbirine çok yakın seviyede olduğu saptanmıştır. Ataksız sahte seslerde artırılmış ve artırılmış eğitim veriseti ile sırasıyla 0.9244 ve 0.9112 metrik değerleri alınmıştır. Medyan filtre atağı altında bu değerler artırılmış eğitim verisetiyle 0.9086, artırılmış eğitim verisetiyle 0.8921 olarak ölçülmüştür. Gürültü seviyesinin 20 dB olarak uygulandığı sahte konuşma kayıtlarında artırılmış eğitim veriseti ile 0.8254, artırılmış eğitim verisetiyle 0.8199 metrik değeri saptanmıştır. Bunlara ek olarak 30 dB gürültü uygulanmış sahte seslerde 0.8988 metrik değeri alınmış, graf sayısının fazla olduğu eğitim setinde ise 0.8856 metrik değeri görülmüştür. 32 kbps sıkıştırma saldırısı artırılmış eğitim verisetiyle 0.8217 metrik değeri kaydederken, artırılmış eğitim verisetiyle 0.8162 seviyesinde seyretmektedir. Sıkıştırmanın 64 kbps olarak uygulandığı seslerde bu değerler sırasıyla 0.9021 ve 0.8986 olmaktadır.

Tablo 64. Artırılmış ve artırılmış eğitim seti ile eğitilmiş mimarilerin TIMIT test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Seti	TIMIT Test Seti	
	Artırılmış	Artırılmış
Eğitim Veri Seti		
Ataksız Sahte Ses	0.9244	0.9112
Medyan Filtre	0.9086	0.8921
20 dB Gürültü Ekleme	0.8254	0.8199
30 dB Gürültü Ekleme	0.8988	0.8856
32 kbps Sıkıştırma	0.8217	0.8162
64 kbps Sıkıştırma	0.9021	0.8986

Artırılmış veriseti kullanılarak eğitilen mimarilerin füzyon sonucu TIMIT test seti üzerinde Tablo 65'te ifade edilen literatürdeki benzer çalışmalardan daha yüksek F-skor metriğine sahiptir.

Ataksız sahte seslerde, medyan filtre, 32 kbps ve 64 sıkıştırma saldırılarında sırasıyla 0.4877, 0.3667, 0.4039 ve 0.4639 seviyeleriyle en düşük metrik değerleri İmran ve arkadaşlarının Yerel İkili Örüntü operatörüne dayanan ses kopyala-yapıştır sahteciliği ile elde edilmiştir. Huang ve arkadaşlarının Ayırık Fourier dönüşümü ve Mel-Frekans Kepstral Katsayıları özelliklerine dayanan çalışması 20 dB gürültü atağı altında 0.5443, 30 dB gürültü atağı altında 0.4564 metrik değerleri kaydederek, bu saldırılar altında en düşük skora sahip olmuştur. Perde ve Formant dizilerinin kullanıldığı Yan ve arkadaşlarının önerdiği yöntem ile Wang ve arkadaşlarının önerdiği Tekil Değer Ayırıştırma ve Ayırık Kosinüs Dönüşümüne dayanan metodları tüm saldırı durumlarında 0.45 ile 0.65 arasında metrik değerlerine sahip olmuştur.

Tablo 65. Artırılmış eğitim seti ile eğitilmiş mimarinin TIMIT test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri

	Önerilen	İmran vd., 2017	Huang vd., 2020	Yan vd., 2019	Wang vd., 2017
Ataksız Sahte Ses	0.9112	0.4877	0.5535	0.5697	0.5655
Medyan Filtre	0.8921	0.3667	0.5489	0.5741	0.5569
20 dB Gürültü Ekleme	0.8199	0.6040	0.5759	0.6491	0.5988
30 dB Gürültü Ekleme	0.8856	0.5460	0.5443	0.5872	0.5740
32 kbps Sıkıştırma	0.8162	0.4039	0.4564	0.4702	0.5305
64 kbps Sıkıştırma	0.8986	0.4639	0.4868	0.5097	0.5169

Tablo 66, artırılmış ve artırılmıy eğitim setleri kullanılarak eğitilen mimarilerin Türkçe test seti üzerindeki füzyon sonuçlarını göstermektedir. Bu tablo incelendiğinde, F-skor metrik değerlerinin birbirine oldukça yakın olduğu tespit edilmiştir. Ataksız sahte seslerde, artırılmıy eğitim veriseti ile 0.93455, artırılmış eğitim veriseti ile ise 0.94204 metrik değerleri elde edilmiştir. Medyan filtre atağı altında, bu skorlar artırılmıy eğitim veriseti için 0.9365, artırılmış eğitim veriseti için ise 0.94204 olarak ölçülmüştür. 20 dB gürültü seviyesi uygulanmış sahte konuşma kayıtlarında, artırılmıy eğitim veriseti ile 0.87861, artırılmış eğitim veriseti ile 0.88996 metrik değerleri bulunmuştur. Ayrıca, 30 dB gürültü uygulanmış sahte seslerde 0.9326 metrik değeri elde edilirken, daha fazla graf içeren eğitim setinde

0.94009 metrik değerine ulaşılmıştır. 32 kbps sıkıştırma saldırısı altında, artırısız eğitim veriseti 0.88506 metrik değeri kaydederken, artırımı eğitim veriseti 0.89317 seviyesinde skor göstermiştir. Sıkıştırmanın 64 kbps olarak uygulandığı seslerde, bu değerler sırasıyla 0.91922 ve 0.92973 olarak kaydedilmiştir.

Tablo 66. Artırımı ve artırısız eğitim seti ile eğitilmiş mimarilerin Türkçe test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Sei	Türkçe Test Seti	
Eğitim Veri Seti	Artırısız	Artırımı
Ataksız Sahte Ses	0.93455	0.94204
Medyan Filtre	0.9365	0.94204
20 dB Gürültü Ekleme	0.87861	0.88996
30 dB Gürültü Ekleme	0.9326	0.94009
32 kbps Sıkıştırma	0.88506	0.89317
64 kbps Sıkıştırma	0.91922	0.92973

Artırımı veri seti kullanılarak eğitilen mimarilerin füzyon sonuçları, TIMIT test seti üzerinde literatürdeki benzer çalışmalara kıyasla daha yüksek F-skor metriklerine ulaşmıştır. Ataksız sahte sesler, medyan filtre, 20 dB ve 30 dB gürültü, 32 kbps ve 64 sıkıştırma saldırılarında sırasıyla 0.2407, 0.2332, 0.1833, 0.2224, 0.1097 ve 0.1377 metrik değerleriyle en düşük başarıya Yan ve arkadaşlarının Perde ve Formant dizilerinin kullanıldığı çalışmaya dayanan ses kopyala-yapıştır sahteciliği ile elde edilmiştir.

Tablo 67. Artırımı eğitim seti ile eğitilmiş mimarinin Türkçe test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri

	Önerilen	İmran vd., 2017	Huang vd., 2020	Yan vd., 2019	Wang vd., 2017
Ataksız Sahte Ses	0.94204	0.5647	0.5234	0.2407	0.5934
Medyan Filtre	0.94204	0.5639	0.5279	0.2332	0.5992
20 dB Gürültü Ekleme	0.88996	0.3794	0.5152	0.1833	0.5669
30 dB Gürültü Ekleme	0.94009	0.3794	0.5270	0.2224	0.5617
32 kbps Sıkıştırma	0.89317	0.4435	0.4589	0.1097	0.5245
64 kbps Sıkıştırma	0.92973	0.3635	0.4707	0.1377	0.5221

Artırısız ve artırımı eğitim setleri kullanılarak eğitilen mimarilerin Librispeech V1 test seti üzerinde füzyon sonuçları Tablo 68’te gösterilmektedir. Söz konusu tablonun analizi sonucunda F-skor metrik değerlerinin birbirine çok yakın seviyede olduğu saptanmıştır. Ataksız sahte seslerde artırısız ve artırımı eğitim veriseti ile sırasıyla 0.996 ve 0.994

metrik değerleri alınmıştır. Yeniden örnekleme atağı altında bu değerler artırısız eğitim verisiyle 0.995, artırılı eğitim verisiyle 0.993 olarak ölçülmüştür. Gürültü seviyesinin 20 dB ve 30 dB, sıkıştırma parametresinin 128 kbps olarak belirlendiği saldırı durumlarında sırasıyla 0.985, 0.993 ve 0.993 metrik değerleri kaydedilmiş olup her iki eğitim verisi durumunda aynı değere ulaşılmıştır.

Tablo 68. Artırılı ve artırısız eğitim seti ile eğitilmiş mimarilerin Librispeech V1 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Seti	Librispeech V1 Test Seti	
Eğitim Veri Seti	Artırısız	Artırılı
Ataksız Sahte Ses	0.996	0.994
Yeniden Örnekleme	0.995	0.993
20 dB Gürültü Ekleme	0.9859	0.985
30 dB Gürültü Ekleme	0.993	0.993
128 kbps Sıkıştırma	0.993	0.993

Üstübioğlu ve arkadaşlarının Modifiye Edilmiş Ayrık Kosinüs Dönüşümü kullandığı ses kopyala-yapıştır sahteciliği tespiti yöntemi, ataksız sahte seslerde, yeniden örnekleme, 20 dB gürültü ve 128 kbps sıkıştırma saldırılarında sırasıyla 0.5110, 0.4988, 0.4725 ve 0.5168 ile en düşük metrik değerleri elde etmiştir. 30 dB gürültü atağı uygulanmış konuşma kayıtlarında ise en düşük skor, 0.5193 ile Imran ve arkadaşlarının Yerel İkili Örüntü operatörüne dayanan çalışmasında kaydedilmiştir. Öte yandan, önerilen yöntemimiz tüm durumlarda en yüksek metrik değerlerine ulaşmıştır.

Tablo 69. Artırımlı eğitim seti ile eğitilmiş mimarinin Librispeech V1 test seti üzerinde literatürdeki benzer çalışmaların F-skor değerleri

	Önerilen	Su vd., 2023	Üstübioğlu vd., 2022	İmran vd., 2017	Huang vd., 2020	Yan vd., 2019
Ataksız Sahte Ses	0.9940	0.9860	0.511	0.6615	0.7839	0.856
Yeniden Örnekleme	0.993	0.934	0.4988	0.6533	0.7818	0.80
20 dB Gürültü Ekleme	0.9850	0.925	0.4725	0.5015	-	0.772
30 dB Gürültü Ekleme	0.993	0.956	0.5339	0.5193	0.6312	0.81
128 kbps Sıkıştırma	0.9930	0.934	0.5168	0.6906	0.7880	0.84

Artırımsız ve artırımlı eğitim setleri kullanılarak eğitilen mimarilerin Librispeech V2 test seti üzerindeki füzyon sonuçları Tablo 70'de sunulmuştur. Tablo incelendiğinde, F-skor metrik değerlerinin birbirine oldukça yakın olduğu gözlenmiştir. Ataksız sahte seslerde, artırımsız ve artırımlı eğitim veri setleri ile sırasıyla 0.993 ve 0.992 metrik değerleri elde edilmiştir. Yeniden örnekleme atağı altında bu skorlar, artırımsız eğitim veri seti için 0.9889, artırımlı eğitim veri seti için ise 0.99 olarak ölçülmüştür. 10 dB gürültü seviyesine sahip seslerde, artırımsız eğitim veri seti ile 0.7362, artırımlı eğitim veri seti ile ise 0.7359 metrik değerleri elde edilmiştir. 20 dB gürültü ve 128 kbps sıkıştırma parametreleri ile yapılan saldırılarda, her iki eğitim veri seti ile sırasıyla 0.91 ve 0.99 metrik değerleri kaydedilmiş ve bu durumlarda metrik değerlerin yaklaşık olarak aynı olduğu görülmüştür.

Tablo 70. Artırımlı ve artırımsız eğitim seti ile eğitilmiş mimarilerin Librispeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Seti	Librispeech V2 Test Seti	
Eğitim Veri Seti	Artırımsız	Artırımlı
Ataksız Sahte Ses	0.9930	0.992
Yeniden Örnekleme	0.9889	0.99
10 dB Gürültü Ekleme	0.7362	0.7359
20 dB Gürültü Ekleme	0.9159	0.914
128 kbps Sıkıştırma	0.9930	0.992

Artırılmış veriseti ile eğitilmiş mimarilerin füzyonu Librispeech V2 ile test edilmesi sonucu oluşan sonuçların literatürdeki benzer çalışmalarla kıyaslanması Tablo 70’te verilmektedir. Gürültü saldırıları haricinde yöntemimiz Tablo 71’te verilen literatürdeki benzer çalışmalardan daha yüksek metrik değerine sahiptir. Ataksız sahte seslerde, yeniden örnekleme, 20 dB gürültü ve 128 kbps sıkıştırma saldırıları altında Liu ve Lu’nun Ayrık Fourier Transform ve Pearson Korelasyon Katsayısının beraber kullanıldığı çalışma sırasıyla 0.81, 0.72, 0.78 ve 0.76 metrik değerlerine sahiptir. 10 dB gürültü uygulanmış sahte seslerde en yüksek metrik değeri 0.95 ile Su ve arkadaşlarının Sabit Q Spektral Çizimlerine dayalı olarak özelleştirilmiş Genetik Algoritma ve Destek Vektör Makinesi entegrasyonuna dayalı yöntemiyle alınmıştır. Aynı atak için 0.73 ile en düşük metrik puanı Yan ve arkadaşlarının Perde dizileri arasında benzerlik baktığı çalışmayla alınmıştır.

Tablo 71. Artırılmış eğitim seti ile eğitilmiş mimarilerin Librispeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

	Önerilen	Su vd., 2023	Yan vd., 2015	Liu & Lu 2017	İmran vd., 2017	Xie vd., 2018
Ataksız Sahte Ses	0.992	0.99	0.90	0.81	0.88	0.93
Yeniden Örnekleme	0.99	0.98	0.87	0.72	0.85	0.85
10 dB Gürültü Ekleme	0.7359	0.95	0.73	0.77	0.70	0.86
20 dB Gürültü Ekleme	0.914	0.97	0.86	0.78	0.85	0.88
128 kbps Sıkıştırma	0.992	0.99	0.88	0.76	0.86	0.85

Artırılmış ve artırılmış eğitim setleri kullanılarak eğitilen mimarilerin Librispeech V3 test seti üzerindeki füzyon sonuçları Tablo 72’de yer almaktadır. Tablo incelendiğinde, duyarlılık metrik değerlerinin birbirine 0.05 puan farkla çok yakın olduğu fark edilmektedir.

Tablo 72. Artırılmış ve artırılmı eğitim seti ile eğitilmiş mimarilerin Librispeech V3 test seti üzerinde Duyarlılık metrięi ile füzyon performans sonuçları

Test Seti	Librispeech V3 Test Seti	
Eğitim Veri Seti	Artırılmı	Artırılmış
0.2	0.9900	0.9950
0.3	0.9850	0.9900
0.4	0.9750	0.9800

Artırılmış veri seti ile eğitilmiş mimarilerin Librispeech V3 test sonuçlarının, literatürdeki benzer çalışmalarla karşılaştırılması Tablo 73’de sunulmaktadır. Kopyala yapıştır segment uzunluğunun 0.2 ve 0.3 ile sınırlandırıldığı sahte konuşma kayıtlarında yöntemimiz sırasıyla 0.9950 ve 0.99 ile literatürdeki benzer çalışmalardan daha yüksek metrik değerlerine sahiptir. Bu durumlarda en düşük seviyeyi Yan ve arkadaşlarının Perde dizilerini kullandığı metot ile sırasıyla 0.81 ve 0.84 metrik değerleriyle alınmıştır. Öte yandan 0.4 saniyelik sahte segmentlerden oluşan konuşma kayıtlarında en düşük başarı aynı çalışmaya ait olmakta, en yüksek metrik değerini Su ve arkadaşlarının Sabit Q Spektral Çizimlerine dayalı olarak özelleştirilmiş Genetik Algoritma ve Destek Vektör Makinesi entegrasyonuna dayalı yöntemiyle elde edilmiştir.

Tablo 73. Artırılmış eğitim seti ile eğitilmiş mimarilerin Librispeech V3 test seti üzerinde Duyarlılık metrięi ile füzyon performans sonuçları

	Önerilen	Su vd., 2023	Yan vd., 2015	Liu & Lu 2017	İmran vd., 2017	Xie vd., 2018
0.2	0.9950	0.98	0.81	0.83	0.90	0.86
0.3	0.9900	0.985	0.84	0.91	0.91	0.91
0.4	0.9800	0.99	0.88	0.93	0.93	0.93

Artırılmı ve artırılmış eğitim setleri kullanılarak eğitilen mimarilerin Chinspeech V1 test seti üzerinde füzyon sonuçları Tablo 74’te gösterilmektedir. Söz konusu tablonun analizi sonucunda artırılmış veriseti kullanılarak eğitilen mimarilerin füzyon sonucu daha yüksek F-skor metrik değerlerine sahip olduğu saptanmıştır.

Ataksız sahte seslerde artırılmı ve artırılmış eğitim veriseti ile sırasıyla 0.9852 ve 0.993 metrik değerleri alınmıştır. Yeniden örnekleme atağı altında bu değerler artırılmı eğitim verisetiyle 0.9852, artırılmış eğitim verisetiyle 0.994 olarak ölçülmüştür. 20 dB gürültü atağı uygulanmış sahte konuşma kayıtlarında artırılmı eğitim verisetiyle 0.9720, artırılmış ile 0.9809 F-skor değeri ölçülmüştür. Aynı saldırı türünün 30 dB içeren parametresi ve 128

kbps sıkıştırma atağı şartları altında bu değerler sırasıyla artırısız verisetiyle 0.9862, artırımı verisetiyle 0.994 seviyesini görmüştür.

Tablo 74. Artırımı ve artırısız eğitim seti ile eğitilmiş mimarilerin Chinspeech V1 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Seti	Chinspeech V1 Test Seti	
Eğitim Veri Seti	Artırısız	Artırımı
Ataksız Sahte Ses	0.9852	0.993
Yeniden Örnekleme	0.9852	0.994
20 dB Gürültü Ekleme	0.9720	0.9809
30 dB Gürültü Ekleme	0.9862	0.994
128 kbps Sıkıştırma	0.9862	0.994

Yöntemimizin literatürdeki benzer çalışmalarla kıyaslanması Tablo 74’te verilmektedir. Çalışmamız Tablo 75’te ifade edilen tüm benzer çalışmalardan daha yüksek F-skor değerine sahiptir. Ataksız sahte, yeniden örnekleme ve 128 kbps sıkıştırma saldırıları durumunda Üstübioğlu ve arkadaşlarının Modifiye Ayrık Kosinüs Dönüşümü ve Öklid Uzaklığının birlikte kullanıldığı çalışmada sırasıyla 0.5469, 0.5528 ve 0.5292 metrik değerleri alarak en düşük başarıyı elinde bulundurmaktadır. Ancak gürültü atakları altında en düşük metrik değerleri 20 dB ve 30 dB durumlarında sırasıyla 0.5061 ve 0.5362 ile İmran ve arkadaşlarının Yerel İkili Örüntü operatörünü kullandığı ses kopyala-yapıştır sahteciliği tespiti yöntemi ile elde edilmiştir.

Tablo 75. Artırımı eğitim seti ile eğitilmiş mimarilerin Chinspeech V1 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

	Önerilen	Su vd., 2023	Üstübioğlu vd., 2022	İmran vd., 2017	Huang vd., 2020	Yan vd., 2019
Ataksız Sahte Ses	0.993	0.976	0.5469	0.7378	0.7767	0.861
Yeniden Örnekleme	0.994	0.943	0.5528	0.7411	0.7769	0.809
20 dB Gürültü Ekleme	0.9809	0.92	0.5481	0.5061	-	0.774
30 dB Gürültü Ekleme	0.994	0.954	0.5376	0.5362	-	0.822
128 kbps Sıkıştırma	0.994	0.966	0.5292	0.75595	0.7792	0.847

Tablo 76, Chinspeech V2 test seti üzerinde artırimsız ve artırılmış eğitim setleri kullanılarak eğitilen mimarilerin füzyon sonuçlarını göstermektedir. Bu tablo incelendiğinde, artırılmış veri seti kullanılarak eğitilen mimarilerin füzyon sonucunun daha yüksek F-skor metrik değerlerine ulaştığı tespit edilmiştir.

Ataksız sahte sesler üzerinde artırimsız eğitim veri setiyle 0.9839, artırılmış eğitim veri setiyle ise 0.991 metrik değerleri elde edilmiştir. Yeniden örnekleme atağı altında, bu skorlar artırimsız eğitim veri setiyle 0.9910, artırılmış eğitim veri setiyle 0.993 olarak ölçülmüştür. 10 dB gürültü atağı uygulanan sahte konuşma kayıtlarında, artırimsız eğitim veri setiyle 0.9046, artırılmış eğitim veri setiyle 0.9079 F-skor değerleri tespit edilmiştir. Aynı saldırı türünün 20 dB içeren parametresi kullanıldığında, artırimsız eğitim setiyle 0.9839, artırılmış eğitim setiyle 0.9859 metrik değeri kaydedilmiştir. 128 kbps sıkıştırma atağı altında ise bu skorlar artırimsız veri setiyle 0.9910, artırılmış veri setiyle 0.994 seviyesinde bulunmuştur.

Tablo 76. Artırılmış ve artırimsız eğitim seti ile eğitilmiş mimarilerin Chinspeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

Test Seti	Chinspeech V2 Test Seti	
Eğitim Veri Seti	Artırimsız	Artırılmış
Ataksız Sahte Ses	0.9839	0.991
Yeniden Örnekleme	0.9910	0.993
10 dB Gürültü Ekleme	0.9046	0.9079
20 dB Gürültü Ekleme	0.9839	0.9859
128 kbps Sıkıştırma	0.9910	0.994

Chinspeech V2 Test Seti üzerinde yöntemimiz ve literatürdeki benzer çalışmaların F-skor değerleri Tablo 77’de verilmektedir. Ataksız sahte konuşma kayıtları ve tüm saldırı durumlarında, yöntemimiz Yan ve arkadaşları tarafından önerilen Perde Dizisi, Ortalama Fark ve Pearson Korelasyon Katsayılarının kullanıldığı çalışmadan ve Imran ve arkadaşlarının Yerel İkili Örüntü operatörüne dayanan yönteminden daha yüksek metrik değerlerine sahiptir. Ayrıca, ataksız sahte seslerde çalışmamız, Xie ve arkadaşları tarafından gerçekleştirilen metodla 0.99 ile aynı metrik değerine sahiptir. Bununla birlikte, Su ve arkadaşlarının Sabit Q Spektral Çizimlerine dayalı olarak özelleştirilmiş Genetik Algoritma ve Destek Vektör Makinesi entegrasyonuna dayalı yöntemi, ataksız sahte seslerde 1.0 metrik puanıyla en yüksek değere sahiptir.

Tablo 77. Artırılmış eğitim seti ile eğitilmiş mimarilerin Chinspeech V2 test seti üzerinde F-skor metriği ile füzyon performans sonuçları

	Önerilen	Su vd., 2023	Yan vd., 2015	Liu & Lu 2017	İmran vd., 2017	Xie vd., 2018
Ataksız Sahte Ses	0.991	1.0	0.90	0.96	0.95	0.99
Yeniden Örnekleme	0.993	0.98	0.90	0.93	0.92	0.95
10 dB Gürültü Ekleme	0.9079	0.96	0.90	0.92	0.90	0.92
20 dB Gürültü Ekleme	0.9859	0.98	0.90	0.93	0.93	0.95
128 kbps Sıkıştırma	0.994	0.99	0.90	0.955	0.94	0.96

Artırılmış ve artırılmı eğitim setleri kullanılarak eğitilen mimarilerin Chinspeech V3 test seti üzerindeki füzyon sonuçları Tablo 78'de yer almaktadır. Tablo incelendiğinde, duyarlılık metrik değerlerinin artırılmış veriseti ile eğitilmiş mimarilerin füzyonu sonucunda daha yüksek değerlerde olduğu görülmektedir.

Tablo 78. Artırılmış ve artırılmı eğitim seti ile eğitilmiş mimarilerin Chinspeech V3 test seti üzerinde Duyarlılık metriği ile füzyon performans sonuçları

Test Seti	Chinspeech V3 Test Seti	
Eğitim Veri Seti	Artırılmı	Artırılmış
0.2	0.9700	0.9800
0.3	0.9950	0.9950
0.4	0.9900	0.9950

Artırılmış veri seti ile eğitilmiş mimarilerin Chinspeech V3 test sonuçlarının, literatürdeki benzer çalışmalarla karşılaştırılması Tablo 79'da sunulmaktadır. Kopyala yapıştır segment uzunluğunun 0.2, 0.3 ve 0.4 ile sınırlandırıldığı bu veriseti üzerinde yöntemimiz tüm uzunluklarda 1.0 metrik değerine sahip Su ve arkadaşlarının Sabit Q Spektral Çizimlerine dayalı olarak özelleştirilmiş Genetik Algoritma ve Destek Vektör Makinesi entegrasyonuna dayalı yönteminden sonra en yüksek metrik değerlerine sahiptir. En düşük seviyeyi Yan ve arkadaşlarının Perde dizilerini kullandığı çalışma da 0.2 saniyelik segmentlerde 0.78, 0.3 saniyelik segmentlerde 0.87 ve 0.4 uzunluklu kopyala yapıştır bölümlerinde 0.88 olarak görmüştür.

Tablo 79. Artırılmış eğitim seti ile eğitilmiş mimarilerin Chinspeech V3 test seti üzerinde Duyarlılık metriği ile füzyon performans sonuçları

	Önerilen	Su vd., 2023	Yan vd., 2015	Liu & Lu 2017	İmran vd., 2017	Xie vd., 2018
0.2	0.9800	1.0	0.78	0.88	0.87	0.96
0.3	0.9950	1.0	0.87	0.95	0.95	0.98
0.4	0.9950	1.0	0.88	0.97	0.97.5	0.98

4. SONUÇLAR

Tez çalışması kapsamında, kopyala yapıştır ses sahteciliği tespitine yönelik olarak 2 farklı yöntem önerilmiştir. Birinci yöntemimiz, BRIEF, AKAZE, SIFT ve SURF gibi dört farklı anahtar nokta çıkarma yöntemini kullanarak özellik tanımlayıcıları çıkarmakta ve bu tanımlayıcıları OPTICS, g2NN ve k-Means gibi kümelendirme algoritmaları ile gruplandırmaktadır. Bu kombinasyonlar altında elde edilen yöntemler toplamda altı farklı çalışma olarak değerlendirilmektedir. Deneysel sonuçlar, sahte seslere uygulanan çeşitli saldırıların yöntemin başarısını olumsuz yönde etkilediğini açıkça göstermiştir. Özellikle, 20 dB gürültü eklenmiş sahte sesler ve 32 kbps sıkıştırma saldırısına maruz kalmış konuşma kayıtları, yöntemin performansını en çok etkileyen saldırılar arasında yer almaktadır. Bu tür saldırılar, Yüksek Çözünürlüklü Spektrogram üzerinde elde edilen anahtar noktaların kalitesini olumsuz yönde etkilemektedir.

Graf tabanlı yöntemimiz, artırılmış ve artırılmı  eğitim setleri ile eğitilmiş olup önerilen CNN modeli, VGG16 ve VGG19 mimarileri üzerinde çalışmaktadır. Eğitilen modeller, füzyon algoritması kullanılarak skor seviyesinde birleştirilmiştir. Çalışmanın performansını olumsuz etkileyen saldırılar arasında 10 dB ve 20 dB gürültü ile 32 kbps sıkıştırma en belirgin şekilde yer almaktadır. Diğer yandan, ataksız sahte sesler ve diğer saldırılar altında yöntemin başarısı oldukça umut vericidir. Graf tabanlı yöntemimiz öğrenmeye dayalı olduğundan, kullanılan eğitim setinin kalitesi önem arz etmektedir. Eğitim için kullanılan veri miktarı, yöntemin başarısını olumsuz yönde etkileyebilecek potansiyele sahiptir.

5. ÖNERİLER

Yüksek Lisans tezi olarak sunulan bu çalışmada kopyala yapıştır ses sahteciliği tespitine ilişkin 2 yöntem önerilmiştir. Bu alanda çalışma yapmak isteyen araştırmacılar için öneriler aşağıda belirtilmiştir.

- Kopyala yapıştır yöntemiyle oluşturulan sahte sesler üzerinde gerçekleştirilen saldırılar, bu tür sahteciliklerin tespit edilmesini zorlaştırmaktadır. Bu nedenle gürültü gibi atakların etkisini ortadan kaldırmak için öncelikle konuşma kaydındaki gürültü seviyesi belirlenebilir ve uygun parametre seçimiyle gürültü etkisi ortadan kaldırılabilir.
- Graf tabanlı yöntemin başarısı, yüksek çözünürlüklü spektrogram görüntülerine anahtar noktaların çıkarımından önce uygulanacak ön işlem adımları ile artırılabilir.

6. KAYNAKLAR

- Abelson, R. P. (1997). A retrospective on the significance test ban of 1999 (If there were no significance tests, they would be invented). L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Dü) içinde, *What if there were no significance tests?* (s. 117-141). New Jersey, Mahwah: Erlbaum.
- Altın Yavuz, A. (2021). İlerleyen Sansürlü Örneklemelere Dayalı Olarak Weibull Dağılımının Şekil Parametresinin Sağlam Tahmin Edicileri. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(2), 410-418.
- Anderson, T. (1962). On the distribution of the two-sample Cramer-von Mises criterion. *The Annals of Mathematical Statistics*, 33(3), 1148-1159.
- Anderson, T., & Darling, D. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2), 193-212.
- Anderson, T., & Darling, D. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49(268), 765-769.
- Angus, J. (1994). The probability integral transform and related results. *SIAM review*, 36(4), 652-654.
- Arizono, I., & Ohta, H. (1989). A test for normality based on Kullback—Leibler information. *The American Statistician*, 43(1), 20-22.
- Arshad, M., Rasool, M., & Ahmad, M. (2003). Anderson Darling and modified Anderson Darling tests for generalized Pareto distribution. *Pakistan Journal of Applied Sciences*, 3(2), 85-88.
- Bain, L., & Engelhardt, M. (1987). *Introduction to probability and mathematical statistics*. Brooks/Cole.
- Balakrishnan, N., & Nevzorov, V. (2004). *A primer on statistical distributions*. Hoboken, New Jersey: John Wiley & Sons Inc.
- Balakrishnan, N., Voinov, V., & Nikulin, M. (2013). *Chi-squared goodness of fit tests with applications*. United States of America: Elsevier.
- Baringhaus, L., & Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, 35(1), 339-348.
- Batsidis, A., Economou, P., & Bar-Lev, S. (2022). A Comparative Study of Goodness-of-Fit Tests for the Laplace Distribution. *Austrian Journal of Statistics*, 51(2), 91-123.
- Bayoud, H. (2021). Tests of normality: new test and comparative study. *Communications in Statistics-Simulation and Computation*, 78(1), 1-22.

- Bera, A., & Jarque, C. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters*, 7(4), 313-318.
- Bonett, D., & Seier, E. (2002). A test of normality with high uniform power. *Computational Statistics & Data Analysis*, 40(3), 435-445.
- Bontemps, C., & Meddahi, N. (2005). Testing normality: a GMM approach. *Journal of Econometrics*, 124(1), 149-186.
- Brys, G., Hubert, M., & Struyf, A. (2008). Goodness-of-fit tests based on a robust measure of skewness. *Computational statistics*, 23(3), 429-442.
- Brys, G., Hubert, M., & Struyf, A. (2008). Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, 23(3), 429-442.
- Cabaña, A., & Cabaña, E. (2003). Tests of normality based on transformed empirical processes. *Methodology and Computing in Applied Probability*, 5(3), 309-335.
- Chaichatschwal, R., & Budsaba, K. (2007). A power comparison of goodness-of-fit tests for normality based on the likelihood ratio and the non-likelihood ratio. *Thailand Statistician*, 5(2007), 57-68.
- Chen, L., & Shapiro, S. (1995). An alternative test for normality based on normalized spacings. *Journal of Statistical Computation and Simulation*, 53(3-4), 269-287.
- Cochran, W. (1952). The χ^2 test of goodness of fit. *The Annals of Mathematical Statistics*, 23(3), 315-345.
- Coin, D. (2008). A goodness-of-fit test for normality based on polynomial regression. *Computational Statistics & Data Analysis*, 52(4), 2185-2198.
- Cox, D. (1961). Tests of separate families of hypotheses. J. Neyman (Dü.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*. içinde 1, s. 105-123. University of California Press.
- Cox, D. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24(2), 406-424.
- Cramér, H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1928(1), 13-74. doi:10.1080/03461238.1928.10416862
- Crzcgorzewski, P., & Wirczorkowski, R. (1999). Entropy-based goodness-of-fit test for exponentiality. *Communications in Statistics-Theory and Methods*, 28(5), 1183-1202.
- D'Agostino, R., & Stephens, M. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker Inc., New York.
- D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2), 341-348.

- D'agostino, R., & Pearson, E. (1973). Tests for departure from normality. Empirical results for the distributions of b^2 and $\sqrt{b}$. *Biometrika*, 60(3), 613-622.
- Dekking, F., Kraaikamp, C., Lopuhaä, H., & Meester, L. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how* (Cilt 488). Delft, Netherlands: Springer.
- del Barrio, E., Cuesta-Albertos, J. A., Matrán, C., & Rodríguez-Rodríguez, J. M. (1999). Tests of goodness of fit based on the L2-Wasserstein distance. *Annals of Statistics*, 27(4), 1230-1239.
- Dodge, Y. (2006). *The Oxford dictionary of statistical terms*. New York: Oxford University Press .
- Domanski, C., & Szczepocki, P. (2020). Comparison of selected tests for univariate normality based on measures of moments. *Statistics in Transition New Series*, 21(5), 151-179.
- Dong, L., & Giles, D. (2007). An empirical likelihood ratio test for normality. *Communications in Statistics—Simulation and Computation*, 36(1), 197-215.
- Doornik, J., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(1), 927-939.
- Dudewicz, E. J., van der Meulen, E. C., SriRam, M. G., & Teoh, N. K. (1995). Entropy-based random number evaluation. *American Journal of Mathematical and Management Sciences*, 15(1-2), 115-153.
- Epps, T. (2005). Tests for location-scale families based on the empirical characteristic function. *Metrika*, 62(1), 99-114.
- Epps, T., & Pulley, L. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, 70(3), 723-726.
- Epps, T., Singleton, K., & Pulley, L. (1982). A test of separate families of distributions based on the empirical moment generating function. *Biometrika*, 69(2), 391-399.
- Ersan, E., & Mersin, S. (2023). Kaynakça Nasıl Yazılır. *Fen Bilimleri Tez Dergisi*, 12(5), s. 123-135.
- Esteban, M. D., Marhuenda, Y., Morales, D., & Sanchez, A. (2007). New goodness-of-fit tests based on sample quantiles. *Communications in Statistics—Simulation and Computation*, 36(3), 631-642.
- Esteban, M., Castellanos, M., Morales, D., & Vajda, I. (2001). Monte Carlo comparison of four normality tests using different entropy estimates. *Communications in Statistics-Simulation and Computation*, 30(4), 761-785.
- Evren, A., & Tuna, E. (2012). On some properties of goodness of fit measures based on statistical entropy. *International Journal of Research and Reviews in Applied Sciences*, 13(1), 192-205.

- Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, 91(434), 674-688.
- Fan, Y. (1997). Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. *Journal of Multivariate Analysis*, 62(1), 36-63.
- Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory*, 14(5), 604-621.
- Filliben, J. (1975). The probability plot correlation coefficient test for normality. *Technometrics*, 17(1), 111-117.
- Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical distributions* (Fourth Edition b.). Hoboken, New Jersey: John Wiley & Sons Inc.
- Fortiana, J., & Grané, A. (2003). Goodness-of-fit tests based on maximum correlations and their orthogonal decompositions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 115-126.
- Freeman, M., & Tukey, J. (1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, 21(4), 607-611.
- Frosini, B. V. (1987). On the distribution and power of a goodness-of-fit statistic with parametric and nonparametric applications. P. Revesz, K. Sarkadi, & P. K. Sen (Dü) içinde, *Goodness-of-fit* (s. 133-154). Amsterdam: North Holland Publishing Company.
- Gel, Y., & Gastwirth, J. (2008). A robust modification of the Jarque-Bera test of normality. *Economics Letters*, 99(1), 30-32.
- Glen, A., Leemis, L., & Barr, D. (2001). Order statistics in goodness-of-fit testing. *IEEE Transactions on Reliability*, 50(2), 209-213.
- Gokhale, D. (1983). On entropy-based goodness-of-fit tests. *Computational Statistics & Data Analysis*, 1, 157-165.
- Grané, A. (2012). Exact goodness-of-fit tests for censored data. *Annals of the Institute of Statistical Mathematics*, 64(6), 1187-1203.
- Hegazy, Y., & Green, J. (1975). Some New Goodness-Of-Fit Tests Using Order Statistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(3), 299-308.
- Heyde, C. (1975). A supplement to the strong law of large numbers. *Journal of Applied Probability*, 12(1), 173-175.
- Hosking, J. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 105-124.
- Hubbard, R., & Lindsay, R. (2008). Why P values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology*, 18(1), 69-88.

- Jarque, C., & Bera, A. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3), 255-259.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (2019). Genesis. N. Balakrishnan, & A. P. Basu (Dü) içinde, *Exponential Distribution: Theory, Methods and Applications* (s. 1-4). New York: Taylor & Francis Group.
- Johnson, N., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2* (Second Edition b., Cilt 289). Hoboken, New Jersey: John Wiley & Sons Inc.
- Khinchin, A. I. (1957). *Mathematical foundations of information theory*. (R. A. Silverman, & M. D. Friedman, Çev.) New York: Dover Publication Inc.
- Kobayashi, H., Mark, B., & Turin, W. (2012). *Probability, random processes, and statistical analysis*. New York, USA: Cambridge University Press.
- Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer.
- Kuiper, N. (1960). Tests concerning random points on a circle. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A*, 63, s. 38-47.
- Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- Laio, F. (2004). Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*, 40(9).
- Ledwina, T. (1994). Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association*, 89(427), 1000-1005.
- Lilliefors, H. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318), 399-402.
- Mardia, K. V. (1972). *Statistics of Directional Data*. London, England: Academic Press.
- Marques de Sá, J. (2007). *Applied statistics using SPSS, statistica, Matlab and R*. USA: Springer Company.
- Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253), 68-78.
- Matsui, M., & Takemura, A. (2005). Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Annals of the Institute of Statistical Mathematics*, 57(1), 183-199.
- Mendes, M., & Pala, A. (2003). Type I error rate and power of three normality tests. *Pakistan Journal of Information and Technology*, 2(2), 135-139.

- Nakas, C. (2007). Performance of the One-Sample Goodness-of-Fit P-P-Plot Length Test. *Communications in Statistics—Simulation and Computation*, 36(5), 1053-1059.
- Neyman, J. (1949). Contribution to the theory of χ^2 test. J. Neyman (Dü.), *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability* içinde (s. 239-273). University of California Press.
- Noughabi, H., & Arghami, N. (2011). Monte Carlo comparison of seven normality tests. *Journal of Statistical Computation and Simulation*, 81(8), 965-972.
- Onyper, S., Thacher, P., Gilbert, J., & Gradess, S. (2012). Class start times, sleep, and academic performance in college: A path analysis. *Chronobiology International*, 29(3), 318-335.
- Owen, A. (2001). *Empirical likelihood* (1 b.). New York: Chapman and Hall/CRC.
- Park, S. (1999). A goodness-of-fit test for normality based on the sample entropy of order statistics. *Statistics & probability letters*, 44(4), 359-363.
- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157-175.
- Pettitt, A. (1977). Testing the normality of several independent samples using the Anderson-Darling statistic. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 26(2), 156-161.
- Rahman, M., & Govindarajulu, Z. (1997). A modification of the test of Shapiro and Wilk for normality. *Journal of Applied Statistics*, 24(2), 219-236.
- Ramachandran, K., & Tsokos, C. (2015). *Mathematical Statistics with Applications in R* (Second Edition b.). Academic Press.
- Ramachandran, K., & Tsokos, C. (2021). *Mathematical statistics with applications in R* (Third Edition b.). Academic Press.
- Razali, N., & Yap, B. W. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.
- Reschenhofer, E., & Bomze, I. (1991). Length tests for goodness of fit. *Biometrika*, 78(1), 207-216.
- Romao, X., Delgado, R., & Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation*, 80(5), 545-591.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences*, 42(1), 43-47.

- Rosin, P., & Rammler, E. (1933). Laws governing the fineness of a powdered coal. *Journal of Institute of Fuel*, 7, 29-36.
- Saksena, R. S. (1980). *A Handbook of Statistics* (Cilt 1). (P. R. Krishnaiah, Dü.) Motilal Banarsidass Publishe.
- Seier, E. (2002). Comparison of tests for univariate normality. *InterStat Statistical Journal*, 1(2002), 1-17.
- Shan, G., Vexler, A., Wilding, G., & Hutson, A. (2010). Simple and exact empirical likelihood ratio tests for normality based on moment relations. *Communications in Statistics—Simulation and Computation*, 40(1), 129-146.
- Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423.
- Shapiro, S. S., & Francia, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337), 215-216.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611.
- Shapiro, S., Wilk, M., & Chen, H. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63(324), 1343-1372.
- Siraj-Ud-Doula, M. (2019). A Comparison among Twenty-Seven Normality Tests. *Research & Reviews: Journal of Statistics*, 8(3), 41-59.
- Smirnov, N. (1939). Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, 2(2), 3-16.
- Song, K.-S. (2002). Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory*, 48(5), 1103-1117.
- Stephens, M. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347), 730-737.
- Student. (1908). The probable error of a mean. *Biometrika*, 1(6), 1-25. doi:<https://doi.org/10.2307/2331554>
- Sullivan, G., & Feinn, R. (2012). Using effect size-or why the P value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282.
- Tolikas, K., & Heravi, S. (2008). The Anderson-Darling goodness-of-fit test statistic for the three-parameter lognormal distribution. *Communications in Statistics—Theory and Methods*, 37(19), 3135-3143.
- Torabi, H., Montazeri, N., & Grané, A. (2016). A test for normality based on the empirical distribution function. *Statistics and Operations Research Transactions*, 40(1), 55-88.

- Towhidi, M., & Salmanpour, M. (2007). A new goodness-of-fit test based on the empirical characteristic function. *Communications in Statistics—Theory and Methods*, 36(15), 2777-2785.
- Uhm, T., & Yi, S. (2021). A comparison of normality testing methods by empirical power and distribution of P-values. *Communications in Statistics-Simulation and Computation*, 1-14.
- Uyanto, S. (2022). An Extensive Comparisons of 50 Univariate Goodness-of-fit Tests for Normality. *Austrian Journal of Statistics*, 51(3), 45-97.
- van Zyl, J. (2018). The performance of univariate goodness-of-fit tests for normality based on the empirical characteristic function in large samples. *Communications in Statistics-Simulation and Computation*, 47(4), 1146-1156.
- Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1), 54-59.
- von Plato, J. (2005). AN Kolmogorov, Grundbegriffe der wahrscheinlichkeitsrechnung (1933). I. G. Guinness (Dü.) içinde, *Landmark Writings in Western Mathematics 1640-1940* (s. 960-969). Elsevier.
- Wang, J. (2012). Sample distribution function based goodness-of-fit test for complex surveys. *Computational Statistics & Data Analysis*, 56(3), 664-679.
- Watson, G. (1961). Goodness-of-fit tests on a circle. *Biometrika*, 48(1/2), 109-114.
- Watson, G. (1962). Goodness-of-fit tests on a circle. II. *Biometrika*, 49(1/2), 57-63.
- Weibull, W. (1939). A statistical theory of the strength of material. *Ingeniors Vetenskapa Acadamiens Handligar*, 151, 5-45.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18, 192-297.
- Wijekularathna, D., Manage, A., & Scariano, S. (2019). Power analysis of several normality tests: A Monte Carlo simulation study. *Communications in Statistics-Simulation and Computation*, 51(3), 757-773.
- Wilks, S. S. (1935). The likelihood test of independence in contingency tables. *The Annals of Mathematical Statistics*, 6(4), 190-196.
- Yap, B., & Sim, C. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141-2155.
- Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, 77(2), 175-183.
- Yıldırım, N., & Gökpınar, F. (2012). Bazı Normallik Testlerinin 1. Tip Hataları ve Güçleri Bakımından Kıyaslanması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 16(1), 109-115.

- Zghoul, A. (2010). A goodness of fit test for normality based on the empirical moment generating function. *Communications in Statistics-Simulation and Computation*, 39(6), 1292-1304.
- Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 281-294.
- Zhang, J., & Wu, Y. (2005). Likelihood-ratio tests for normality. *Computational Statistics & Data Analysis*, 49(3), 709-721.
- Zhang, P. (1999). Omnibus test of normality using the Q statistic. *Journal of Applied Statistics*, 26(4), 519-528.



ÖZGEÇMİŞ

Muhammed KILIÇ, İlköğrenimini Gündoğdu 29 Ekim İlköğretim Okulu'nda bitirdikten sonra lise eğitimini 2015 yılında Rize Anadolu Lisesi'nde tamamladı. 2016-2017 öğretim yılında Karadeniz Teknik Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği bölümünde lisans eğitimine başladı. 2020 yılında bu bölümden mezun oldu. Aynı yıl Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim dalında tezli yüksek lisans programına başladı. 2023 yılından itibaren Karadeniz Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü'nde araştırma görevlisi olarak görev yapmaktadır. Orta derece İngilizce bilmektedir.

Ustubioglu, B., Tahaoglu, G., Ulutas, G., Ustubioglu, A., ve Kilic, M. (2024). Audio forgery detection and localization with super-resolution spectrogram and keypoint-based clustering approach. The Journal of Supercomputing, 80(1), 486-518.