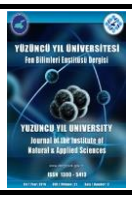
Yüzüncü Yıl Üniversitesi
Fen Bilimleri Enstitüsü Dergisi

https://dergipark.org.tr/tr/pub/yyufbed



Derleme Makalesi

Derin Sahte Ses Manipülasyonu Tespit Sistemleri Üzerine Bir Derleme

Gül TAHAOĞLU*, Muhammed KILIÇ, Beste ÜSTÜBİOĞLU, Güzin ULUTAŞ

Karadeniz Teknik Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 61080, Trabzon, Türkiye

Gül TAHAOĞLU, ORCID No: 0000-0002-8828-5674, Muhammed KILIÇ, ORCID No: 0000-0002-2402-8265, Beste ÜSTÜBİOĞLU, ORCID No: 0000-0001-7451-0634, Güzin ULUTAŞ, ORCID No: 0000-0001-5729-6613

*Sorumlu yazar e-posta: gultahaoglu@ktu.edu.tr

Makale Bilgileri

Geliş: 12.09.2023
Kabul: 15.02.2024
Online Nisan 2024

DOI:10.53433/yyufbed.1358880

Anahtar Kelimeler
Derin sahte ses manipülasyonu,
Derin sahte ses tespiti,
Ses doğrulama

Öz: Gerçek kişilerin konuşmalarını içeren dijital ses dosyalarının kullanılması ile gerçekleştirilen derin sahte ses manipülasyonu, sesi taklit edilecek kişinin sesini klonlayarak kişinin söylemediği bir şeyi söylemiş gibi içerikte ses dosyalarını oluşturan bir sahtecilik türüdür. Konuşmacının kimliğini doğrulamak için güvenlik adımı olarak kabul edilen Otomatik Konuşmacı Doğrulama Sistemlerinin derin sahte ses sahtecilikleri saldırılarına karşı savunmasızlığı söz konusudur. Ayrıca mahkemelerde karar merciini etkileyecek delil olarak sunulan ses dosyalarının orijinal olup olmadığı kontrolü önemli bir ihtiyaç haline gelmiştir. Bu tür sahteciliklerin uzman sistemler tarafından tespit edilebilmesi günümüz çağı için oldukça önem arz etmektedir. Bu sahtecilik türündeki saldırıların tespit edilebilmesi için literatürde çeşitli yöntemler önerilmiştir. Literatürdeki çalışmalarda performans değerlendirmesinde kullanılan ücretsiz erişimli veri setleri de mevcut olup sonuç kıyaslamasında kullanabilmesi mümkündür. Bu çalışmada literatürdeki yöntemler ve verisetleri incelenmiş, yöntemlerin bu verisetleri üzerindeki performans değerlendirmeleri, avantaj ve dezavantajları vurgulanmıştır.

A Review of Deepfake Audio Manipulation Detection Systems

Article Info

Received: 12.09.2023
Accepted: 15.02.2024
Online April 2024

DOI:10.53433/yyufbed.1358880

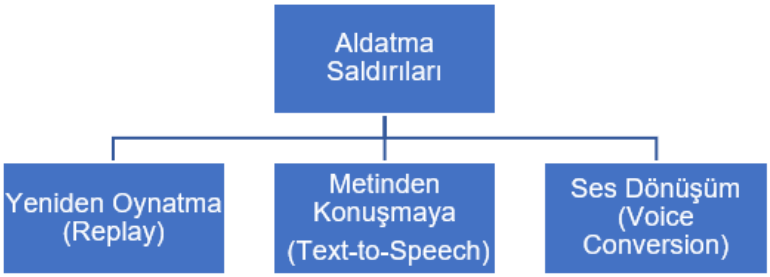
Keywords
Audio authentication,
Deep fake audio manipulation,
Deep fake audio detection

Abstract: Besides facilitating access to audio content on the Internet, developments in deep learning methods have made it possible to produce deep fake audio. Automatic Speaker Verification systems considered a security step to authenticate the speaker, are vulnerable to deep spoofing attacks. It is crucial for today's age that expert systems can detect such frauds. Deep fake audio spoofing is carried out to produce audio files in the content by cloning the speaker's voice that is planned to be changed as if he said something he did not say. Various methods are proposed in the literature to detect this type of forgery. There are free-access datasets used in performance evaluation in studies in the literature, and it is possible to use them in result comparison. The planned research aims to reduce or eliminate the noise that may exist in the audio file of the system by passing the preprocessing stage of the audio signal received as input. This paper examines the methods and datasets in the literature, and the advantages and disadvantages of the methods on these datasets are emphasized.

1. Giriş

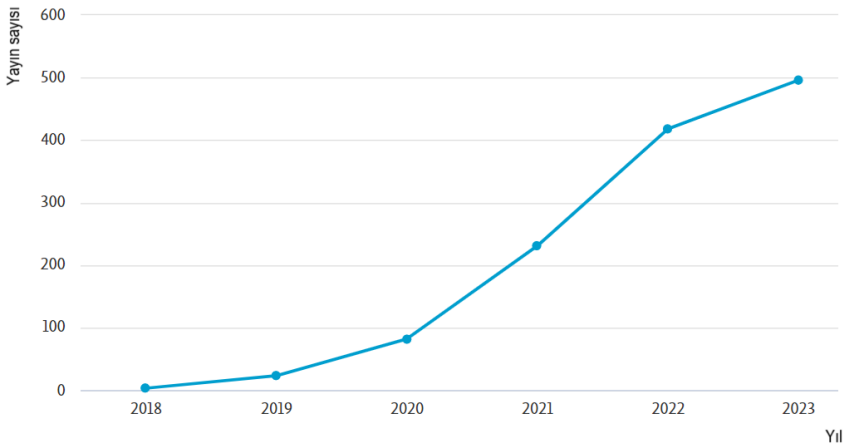
Derin öğrenme tekniklerindeki gelişmeler deepfake olarak isimlendirilen "deep learning" (derin öğrenme) ve "fake" (sahte) kelimelerinden türetilen gerçek bir çoklu ortam verisinin fark edilemeyecek şekilde başarılı bir şekilde sahtesinin üretilmesini mümkün kılmıştır (Patel & Patil, 2015). Bu sahtecilik türünden biri de derin sahte seslerin üretilmesidir. Derin sahte ses sahteciliği, değiştirilmesi planlanan konuşmacının sesinin klonlanarak, söylemediği bir şeyi söylemiş gibi içerikteki ses dosyalarının üretilmesi amacı ile gerçekleştirilmektedir (Patel & Patil, 2015). Yapay zekâ yaklaşımlarını kullanılarak oluşturulan bu sahtecilik yöntemi özellikle konuşmacı doğrulama sistemleri için büyük bir tehdit oluşturmaktadır. Ayrıca askeri ve politik konumlara sahip yetkilerin konuşmalarının bu atak türüne maruz kalması da ulusal güvenlik zafiyeti oluşturabilmektedir. Bu tür durumların günümüz çağında meydana gelme olasılığının yüksekliği, derin sahte seslerin tespit edilmesinin önemini ortaya koymaktadır.

Derin sahte ses manipülasyonu saldırıları, sahte ses dosyasının oluşturulmasındaki tekniklere göre gruplandırılabilmektedir. Şekil 1'de verildiği gibi en popülerler saldırı türleri; Yeniden oynatma (replay), Metinden Konuşmaya (Text-to-Speech, TTS) ve Ses Dönüşüm (Voice Conversion, VC) olup Aldatma Saldırıları sınıfında değerlendirilmektedirler. Bunlardan Metinden Konuşmaya ve Ses Dönüşüm saldırıları daha sık karşılaşılan atak türüdür ve biyometrik doğrulama sistemleri için büyük bir tehdit oluşturmaktadır.



Şekil 1. Derin sahte ses manipülasyonu saldırıları.

Konunun önemi gereği literatürde ses dosyaları üzerinde gerçekleştirilen aldatma saldırılarının tespit edilebilmesi üzerine gerçekleştirilen çalışmalar yeniliğini korumasına karşın popülaritesinde artış gözlemlenmektedir (www.scopus.com). Şekil 2'de scopus veritabanında "deepfake audio" anahtar kelimesi ile yapılan yayın sorgusu durumunda yıllara göre yayın sayısının grafiği Şekil 2'de verilmiştir. Grafikte de görüldüğü gibi araştırmacılar tarafından da konuya artan ilgi olduğu söylenebilir.



Şekil 2. "deepfake audio" anahtar kelimesi ile yapılan sorguda yıllara göre yayın sayısı.

Derin sahte seslerin tespiti için ortaya konulan sistemlerin performans değerlendirilmesinin gerçekleştirileceği ASVSpoof meydan okuma serisi oluşturulmuş ve bu kapsamda hem orijinal hem de sahte ses dosyalarını içeren verisetleri oluşturulmuştur (Wu ve ark., 2015 ve 2017; Nautsch ve ark., 2021; Yamagishi ve ark., 2021). Oluşturulan verisetleri literatürdeki çalışmaların performanslarını değerlendirilmek üzere kullanılmaktadır. Ancak aldatma saldırılarına karşı sistemlerini daha güçlü hale getirebilmek için literatürde yeni Aldatma Saldırısı engelleme yöntemlerinin araştırılması ve geliştirilmesi ihtiyacı devam etmektedir. Literatürde yapılan çalışmaları ses manipülasyonunun tespiti amacı ile ses dosyasından elde edilen özelliklerin türü açısından değerlendirildiğinde iki temel sınıfın ortaya çıktığı gözlemlenmiştir. Bunlar geleneksel yöntemler ile elde edilen özelliklerin kullanılmasına dayalı yöntemler ve derin öğrenme yaklaşımlarına dayalı özellikler olarak isimlendirilebilir.

Bu çalışmada derin sahte seslerin tespiti için literatürde önerilen yöntemler 2. Bölümde sınıflandırılmış, öne çıkan bazı yöntemlerin avantaj ve dezavantajlarından bu bölümde bahsedilmiştir. Yöntemlerin geliştirilmesinde faydalanılan verisetleri ve metriklerden 3. Bölümde verilmiştir. 2. Bölümde verilen yöntemlere ek olarak bu alandaki diğer çalışmalar, genel akışlarına ilişkin şekiller ve kullanılan verisetleri üzerindeki performans değerlendirmeleri ile daha detaylı olarak Bölüm 4'te sunulmuştur. Çalışmaların öne çıkan zorluklarına ve genel değerlendirmelerine son bölümde yer verilmiştir.

2. Derin Sahte Ses Tespiti

Özellikle ses tabanlı kimlik doğrulama sistemleri, ses üzerinden iletişimde bulunan sistemler veya askeri ve politik durumlarda girdi olarak alınan seslerin derin sahte ses sahteciliği ile oluşturulup oluşturulmadığının ortaya koyan doğrulama adımı ihtiyacı giderek artmaktadır. Derin sahte seslerin üretiminde kullanılan teknolojilerdeki hızlı gelişim sayesinde, sahte bir sesin gerçek insan sesinden ayırt edilebilmesi gittikçe zorlaşmaktadır. Derin sahte seslerin tespiti için en sık başvurulan ve en güvenilir sistemler makine öğrenmesi yaklaşımlarının kullanıldığı otomatik sistemlerdir. Bu sistemler büyük veri kümesi içinde sahte ve gerçek ses örnekleriyle eğitilerek sahte sesleri tespit etmek için kullanılmaktadır. Bu amaçla literatürde önerilen yöntemler ses dosyasına geleneksel yöntemlerle veya derin öğrenme yaklaşımları ile özellikler elde edilerek bu özelliklerin sınıflandırılmasına dayanmaktadır. Buna göre bu alandaki çalışmaları iki sınıfta değerlendirmek mümkündür;

1. Geleneksel yöntemler ile özellik elde eden çalışmalar

2. Derin öğrenmeye dayalı özellikler kullanan çalışmalar

Literatürde ses ait özelliklerin elde edilmesinde ve bunların sınıflandırılmasında birçok yaklaşımı kapsayan çalışma bulunmaktadır. Bu çalışmalar gözlemlendiğinde, sistemin eğitimi sırasında kullanılmayan bir aldatma saldırısının sisteme girdi olarak verildiği zaman yöntemlerin başarısız kaldığı bilgisi edinilmiştir. Bu durum literatürde genelleştirme problemi olarak isimlendirilmektedir. Ayrıca yapılan çalışmaların birçoğunda ele alınan problem ikili sınıflama problemi olarak değerlendirilmekte ve girişten alınan ses dosyası gerçek veya sahte olarak etiketlendirilmektedir. Bunu gerçekleştirebilmek için sistemin eğitimi için kullanılan eğitim ve test verisetlerinde de benzer bir dağılımın varlığı kabul edilmektedir. Literatürdeki yöntemlerde sunulan modellerin geliştirilmesinde verisetleri oldukça önem arz etmektedir. Bir sonraki bölümde bu modellerinin eğitimi ve testinde kullanılan verisetleri ve performans değerlendirme metrikleri verilecek olup dördüncü bölümde yöntemlere ilişkin daha detaylı analizler yapılacaktır. Anlatımda kullanılan bazı terimler ve kısaltmaları Çizelge 1'de verilmiştir.

Çizelge 1. Makale içindeki kısaltmalar ve açıklamaları

Kısaltma	Açıklama	Kısaltma	Açıklama
TTS	Text-to-Speech	STLT	Short Term Long Term
VC	Voice Conversion	LCNN	Light CNN
SS	Speech Synthesis	RNN	Recurrent Neural Network
C1	First Mel-Cepstral Coeffcient	LSTM–RNN	Long Short Term Memory-Recurrent Neural Network

Çizelge 1. Makale içindeki kısaltmalar ve açıklamaları (devam)

Kısaltma	**Açıklama**	**Kısaltma**	**Açıklama**
C1	First Mel-Cepstral Coeffcient	LSTM–RNN	Long Short Term Memory-Recurrent Neural Network
HTS3	Hidden Markov Model Based Speech Synthesis System	LDA	Linear Discriminant Analysis
GMM	Gaussian Mixture Models	GDF	Gaussian Density Function
LSP	Line Spectrum Pair	HFCC	High-Frequency Cepstral Coefficients
KPLS	Kernel-Based Partial Least Square	SSAD	Self-Supervised Spoofing Audio Detection
MaryTTS	MARY Text-To-Speech	TCN	Temporal Convolutional Network
RC	Replay Configuration	GRU	Gated Recurrent Unit
LA	Logical Access	MSE	Mean Squared Error
PA	Physical access	LCNN-big	Light Convolutional Neural Network-big
VC	Voice Coder	LCNN-small	Light Convolutional Neural Network-small
WC	Waveform Concatenation	A-Softmax	Angular Softmax
VAE	Variational Auto-Encoder	TSSDNet	Time-domain Synthetic Speech Detection Net
MFCCs	Mel-Frequency Cepstral Coefficients	GNA	Gaussian Noise Addition
PSTN	Public Switched Telephone Network	SnrNA	Signal-to-Noise Ratio Noise Addition,
DF	Deep Fake	LPS	Log Power Spectrum
EER	Equal Error Rate	TC	Temporal Convolution
HTER	Half Total Error Rate	SC	Spatial Convolution
FAR	False Acceptance Rate	LSTM	Long Short-Term Memory
FRR	False Rejection Rate	AEXANet	Audio Examiner RawNet
t-DCF	Tandem Detection Cost Function	MFM	Maximum Feature Mapping
CM	countermeasure	GTCC	Gammatone Filter based Cepstrum Coefficient
fa	false alarm	BiLSTM	Bidirectional Long Short-Term Memory Network
MFCCs	Mel-Frequency Cepstral Coefficients	SVM	Support Vector Machine
MGD	Modified Group Delay Cepstral Coefficients	RBF	Radial Basis Function
LMS	Log Magnitude Spectrum	OC-Softmax	One Class-Softmax
GD	Group Delay	GAP	Global Average Pooling
IF	Instantaneous frequency derivative	SFFCC	Single Frequency Filter Cepstral Coefficients

Çizelge 1. Makale içindeki kısaltmalar ve açıklamaları (devam)

Kısaltma	**Açıklama**	**Kısaltma**	**Açıklama**
BPD	Baseband Phase Difference	ZTWCC	Zero Time Windowing Cepstral Coefficients
PSP	Pitch Synchronous Phase	IFCC	Instantaneous Frequency Cepstral Coefficients
MLP	Multi Layer Perceptron	Logspec	Log Power Magnitude Spectra
CFCC	Cochlear Filter Cepstral Coefficients	CRNN	Convolutional Recurrent Neural Network
CFCCIF	Cochlear Filter Cepstral Coefficients Instantaneous Frequency	SDA	Static, Delta and Acceleration,
FBNN	Filter Bank Neural Network	LTAS	Long-Term-Average-Spectrum
DNN-FBCC	Deep Neural Network Filter Bank Cepstral Coefficients	VoIP	Voice Over Internet Protocol
IMFCC	Inverted Mel-Frequency Cepstral Coefficients	LFB	Linear Filter Banks
LPCCres	Linear Prediction Cepstral Coefficients Residual	LMCL	Large Margin Cosine Loss
SC	Scattering Coefficients	LT	Long-Term
DCT	Discrete Cosine Transform	ST	Short-Term
SCC	Scattering Cepstral Coefficients	RLMS	Residual log magnitude spectrum
GMM-UBM	Gaussian Mixture Model - Universal Background Model	MSE-CC	MSE Cepstral Coefficient
RFCCs	Rectangular Filter Cepstral Coefficients	STCC	Short Term Cepstral Coefficients
LPCCs	Linear Prediction Cepstral Coefficients	GMM	Gaussian Mixture Model
SSFCs	Subband Spectral Centroid Frequency Coefficients	FMD	Frequency Modulation Deviation
SCMCs	Subband Spectral Centroid Magnitude Coefficients	SCD	Spectral Centroid Deviation,
CMN	Cepstral Mean Normalization	SCF	Spectral Centroid Frequency
STSSI	Short-Term Spectral Statistics Information	MelFbanks	Mel scale filter banks
OPI	Octave-Band Principal Information	MGD	Modified Group Delay Function
FPI	Full-band Principal Information	CQTMGD	Modified Group Delay Function based on Constant-Q-Transform
CQSPIC	Constant-Q-Statistics-Plus-Principal Information Coefficient	LPCCs	Linear Predictive Cepstral Coefficients
MCF	Modulation Centroid Frequency	CCs	Complex Kepstral Coeffcients
MCF-CC	MCF cosine coefficients	STFT	Short-Time Fourier Transform
MSE	Modulation Static Energy	GD gram	Group Delay gram

3. Verisetleri ve Metrikler

Derin sahte ses tespiti modellerinin başarılı olabilmesi için yüksek oranda sahte ve orijinal veri örneği içeren verisetlerinde eğitilmesi gerekmektedir. Literatürde bu ihtiyacı karşılayacak şekilde oluşturulmuş verisetleri bulunmaktadır. Bu veri setleri ASVSpoof, BTAS ve FoR verisetleridir. Önerilen modellerin başarısı performans metrikleri kullanılarak rapor edilmektedir. Bu bölümde ilk olarak literatürde açık erişimli olarak yayınlanan derin sahte ses tespiti modellerinin eğitimi ve testi için kullanılan verisetlerinden ve sonrasında performans analizinde kullanılan performans metriklerden bahsedilecektir.

3.1. ASVSpoof verisetleri

ASVSpoof verisetleri, 2015 yılında başlayan ASVSpoof meydan okuma serileri sırasında oluşturulmuştur ve verisetlerinin detayları aşağıdaki gibidir.

• **ASVSpoof 2015 veriseti:** ASVSpoof 2015 veriseti orijinal ve sahte ses kayıtlarından oluşmaktadır. Orijinal ses kayıtları, 106 konuşmacının (45 erkek ve 61 kadın) herhangi bir değişiklik yapılmadan ve v arka plan gürültü efektleri olmadan kaydedilir (Wu ve ark., 2015). Sahte ses kayıtları Konuşma Sentezi (Speech Synthesis, SS) ve Ses Dönüştürme (Voice Conversion, VC) algoritmaları aracılığı ile orijinal ses kayıtlarından oluşturulmuştur. Veriseti eğitim, geliştirme ve test setlerine bölünmüş olmakla birlikte hiçbirinde ortak konuşmacı bulunmamaktadır.

Çizelge 2'de bu verisetinde yer alan sahte ve orijinal ses kayıtları ve orijinal ses kayıtlarında kullanılan konuşmacı sayıları yer almaktadır. Çizelgede de görüldüğü gibi eğitim seti, 25 konuşmacıdan (10 erkek, 15 kadın) toplanan 3750 orijinal ve 12625 sahte sesleri içermektedir. Eğitim setinde kullanılan sahte sesler S1-S5 algoritmalarından biri kullanılarak oluşturulmuştur. Geliştirme veriseti, 15 erkek 20 kadın olmak üzere 35 konuşmacıdan alınan orijinal ve bu orijinal seslerden oluşturulmuş sahte sesleri içermektedir. Geliştirme veriseti, eğitim setindeki sahte sesleri oluşturmak için kullanılan sahtecilik algoritmalarıyla oluşturulmuştur. Değerlendirme seti, 20 erkek ve 26 kadın olmak üzere toplam 46 konuşmacıdan toplanan 9404 orijinal, 184000 sahte seslerden oluşmaktadır. Sahte sesler eğitim ve geliştirme setlerinde kullanılan 5 algoritmanın yanında, 5 farklı sahtecilik algoritması(S6-S10) kullanılarak elde edilmiştir.

Çizelge 2. ASVSpoof2015 veri setinin oluşturulmasında faydalanılan konuşmacı ve alt veri gruplarında yer alan ses

		Eğitim	**Geliştirme**	**Test**
Sesler	Orijinal	3750	3497	9404
	Sentetik	12625	49875	184000
Konuşmacı		25	35	46

Sahte seslerin oluşturulmasında kullanılan S1-S10 algoritmaları şu şekildedir:

- S1; basitleştirilmiş çerçeve seçimi tabanlı ses dönüştürme (Voice Conversion, VC) algoritmasıdır.
- S2; kaynak spektrumunun eğimini hedefe kaydırmak için yalnızca birinci Mel-Kepstral Katsayısını (First Mel-Cepstral Coeffcient, C1) kullanan en basit ses dönüştürme algoritmasıdır.
- S3 ve S4; gizli Markov modeli tabanlı konuşma sentezi sistemi (Hidden Markov Model Based Speech Synthesis System, HTS3) uygulayan konuşma sentezi (Speech Synthesis, SS) algoritmasıdır.
- S5; ses dönüştürme araç seti ve Festvox sistemi uygulanan bir ses dönüştürme algoritmasıdır.
- S6; Gauss Karışım Modellerine (Gaussian Mixture Models, GMM) ve küresel varyansı göz önünde bulundurarak maksimum olasılık (maximum likelihood) parametresi üretimine dayanan ses dönüştürme algoritmasıdır.
- S7; Çizgi Spektrum Çifti (Line Spectrum Pair, LSP) kullanan ses dönüştürme algoritmasıdır.
- S8; Japonca veriseti kullanılarak yapılan tensör tabanlı ses dönüştürme (Voice Conversion, VC) algoritmasıdır.

- S9; doğrusal olmayan bir dönüşüm fonksiyonunu uygulamak için çekirdek tabanlı kısmi en küçük kareleri (Kernel-Based Partial Least Square, KPLS) kullanan ses dönüştürme (Voice Conversion, VC) algoritmasıdır.
- S10; açık kaynaklı MARY Metinden Konuşmaya (MARY Text-To-Speech, MaryTTS) uygulanan Konuşma Sentezi (Speech Synthesis, SS) algoritmasıdır. S1, S2, S3 ve S4ün tümü sentez için aynı STRAIGHT ses kodlayıcı kullanırken S5, bir MLSA ses kodlayıcıyı kullanır.

• **ASVSpoof 2017 veriseti:** ASVSpoof 2017 veriseti, Metne-Bağımlı RedDots veriseti (Text-Dependent RedDots kullanılarak oluşturulmuştur. Metne-Bağımlı RedDots setinden alınan orijinal sesler ile yeniden oynatılmış sahte ses kayıtları oluşturulmuştur (Wu ve ark., 2017). Sahte sesler çeşitli cihazlar, hoparlörler ve kayıt cihazlarından oluşan çeşitli yeniden oynatma yapılandırmaları ve çeşitli ayarlar altında yeniden oynatılmasıyla oluşturulmuştur. Çizelge 3'de bu veri setinin oluşturulması için ses kayıtlarının alındığı konuşmacı sayıları ve her bir alt veri setlerinde yer alan orijinal/sahte ses sayıları verilmiştir. Çizelgede görüldüğü gibi eğitim setinde 10 erkek konuşmacının gerçek ve sahte konuşmaları yer alır. Sahte konuşma, 6 farklı oturumda 3 farklı tekrar yapılandırmasıyla üretilir. Geliştirme setinde toplam 8 konuşmacının orijinal ve sahte konuşma kayıtları mevcuttur. Sahte ses örnekleri farklı oynatma ve kayıt cihazlarıyla 10 farklı tekrar oynatma oturumundan oluşturulmuştur.

Çizelge 3. ASVSpoof2017 verisetinin oluşturulmasında faydalanılan konuşmacı ve alt veri gruplarında yer alan ses sayıları

		Eğitim	**Geliştirme**	**Test**
Sesler	Orijinal	3750	3497	9404
	Sentetik	12625	49875	184000
Konuşmacı		25	35	46

Orijinal seslerin yeniden oynatılması ile sahte seslerin oluşturulması sürecinde akustik etkiler, bir oynatma cihazının ve kayıt cihazının ve sesin yayıldığı bir akustik ortamın etkilerini kapsar. Bunların bir kombinasyonuna Tekrar Oynatma Konfigürasyonu (Replay Configuration, RC) adı verilir. Akustik ortam, orijinal konuşma verilerinin yeniden oynatıldığı ve yeniden kaydedildiği fiziksel alandır. Bu verisetinde sahte sesler 26 farklı ortamda oluşturulmuştur. Sekiz farklı ev koşulu, orta ortam gürültüsü seviyeleri ile karakterize edilen oturma odaları ve yatak odalarında yapılan kayıtları içermektedir. 10 ofis koşulu klima sistemleri tarafından üretilen orta düzeyde ortam gürültüsünü içerirken aynı zamanda yankılanma da sergiler. İki balkon ve bir kantin koşulları, ortam gürültüsünün yüksek olduğu ortamlardır. Dört düşük gürültü koşulu vardır. Yankısız oda kayıtları çok düşük toplam gürültü ve yankılanma sergilerken stüdyo kayıtları da benzer şekilde düşük düzeyde ortam gürültüsü ve aynı zamanda bir dereceye kadar yankılanma içerir. Analog kablo koşulları, fiziksel ses yayılımı olmadan, ancak bir oynatma cihazından doğrudan bir ASV sistemine elektriksel yayılımla yapılan kayıtları simüle eder. 26 oynatma cihazı ve 25 adet kayıt cihazı bulunmaktadır. 26 oynatma cihazı, 25 kayıt cihazı ve 26 ortamın kullanıldığı konfigürasyonlar toplamda 16900 adettir.

•**ASVSpoof 2019 veriseti:** Bu veriseti Mantıksal Erişim (Logical Access, LA) ve Fiziksel Erişim (Physical access, PA) olmak üzere iki farklı senaryoya ayrılmıştır (Nautsch ve ark., 2019). Mantıksal Erişim, anadili İngilizce olan 46 erkek, 61 kadın toplamda 107 kişiden elde edilen orijinal seslerin yanında, 17 farklı yöntemle oluşturulan sentetik seslerden oluşmaktadır. Örnekleme frekansı 16000 Hz. olmakla birlikte veri seti kayıpsız bir ses kodlama formatında oluşturulmuştur. Eğitim seti 8 erkek ve 12 kadın toplam 20 kişinin orijinal konuşma kayıtlarını ve 6 atakla oluşturulmuş sahte sesleri içerir. Geliştirme seti ise 4 erkek ve 6 kadın toplam 10 kişinin orijinal konuşma kayıtlarını ve 6 atakla oluşturulmuş sahte seslerden oluşmaktadır. Eğitim ve geliştirme setlerinde sentetik sesler A01- A06 arasında 6 farklı atak türüyle oluşturulmuştur. Buna karşın test seti ise 21 erkek ve 27 kadın olmak üzere toplam 48 kişinin orijinal konuşma kayıtlarını ve 13 yöntemle (A07- A19 ataklarıyla) elde edilen sentetik konuşmaları içerir. A16 ve A19 aslında sırasıyla A04 ve A06 ile çakışmaktadır. Dolayısıyla eğitim ve test setinde sadece 2 atak türü ortak iken 11 atak türü tamamen farklıdır.

Çizelge 4. ASVSpoof2019 LA veriseti oluşturulmasında faydalanılan konuşmacı ve alt veri gruplarında yer alan ses sayıları

		Eğitim	Geliştirme	Test
Sesler	Orijinal	2580	2548	7355
	Sentetik	22800	22296	63882
Konuşmacı		20	10	48

Atakların oluşturulması birkaç farklı yönteme dayanmaktadır. Eğitim ve geliştirme setinde yer alan A01 ile test setinde yer alan A07, A08, A10, A11, A12 ve A13 atakları Sinir Ağı kullanılarak oluşturulmuştur. Ses Kodlayıcı (Voice Coder, VC) kullanılarak oluşturulan ataklar eğitim ve geliştirme setinde yer alan A02, A03, A05 ve A06 ataklarıyla birlikte test setindeki A09, A14, A15, A17, A18 ve A19'dur. Eğitim ve geliştirme setindeki A04 ve test setindeki A16 atağı Dalga Biçimi Birleştirme (Waveform Concatenation, WC) yöntemi ile oluşturulmuştur.

A01, WaveNet Sinir Ağı kullanılarak gerçekleştirilen Metinden Konuşmaya sahteciliği ile oluşturulmuştur. A02, WORLD ses kodlayıcının kullanıldığı A01'e benzeyen Metinden Konuşmaya sahteciliği kullanılmıştır. A03, Merlin adı verilen açık kaynaklı TTS araç setini kullanan, A02'ye benzer Metinden Konuşmaya (Text-To-Speech, TTS) sahteciliğidir. A04, MaryTTS platformunu temel alan bir dalga biçimi birleştirme (Waveform Concatenation, WC) uygulanmıştır. A05, dalga biçimi üretimi için değişken bir otomatik kodlayıcı (Variational Auto-Encoder, VAE) ve WORLD ses kodlayıcı kullanan Sinir Ağı tabanlı bir Ses Dönüştürme (Voice Conversion, VC) sahteciliğidir. A06, transfer fonksiyonu (transfer-function) tabanlı bir Ses Dönüştürme (Voice Conversion, VC) sahteciliği ile oluşturulmuştur. Bu yöntem, bir konuşmacının sesini başka bir konuşmacının sesine dönüştürmek için ses kodlayıcı kullanır. A07'de dalga biçimi WORLD ses kodlayıcısı kullanılarak sentezlendikten sonra konuşmayı daha doğal hale getirmek için zaman alanlı sinir filtresi olan WaveCycleGAN2 kullanılır. A07 bir Metinden Konuşmaya sahteciliği ile oluşturulmuştur. A08, A01'e benzer bir Metinden Konuşmaya sahteciliğine uygulanmış sesleri kapsar. WaveNet'ten daha hızlı olan bir sinir kaynağı filtresi dalga biçimi (neural-source-filter waveform model) modelini kullanır. A09, Vocaine ses kodlayıcısı kullanılarak oluşturulan bir Metinden Konuşmaya sahteciliği ile oluşturulmuştur. A10, WaveRNN sinirsel ses kodlayıcı aracılığıyla gerçekleştirilen Metinden Konuşmaya sahteciliğidir. A11, dalga formları oluşturmak için Griffin-Lim algoritmasını kullanması dışında A10 ile aynı olan Metinden Konuşmaya (Text-To-Speech, TTS) sahteciliğidir. A12, WaveNet tabanlı Metinden Konuşmaya (Text-To-Speech, TTS) sahteciliğidir. A13, giriş dalga biçimini doğrudan değiştiren, birleşik Sinir Ağları tabanlı Ses Dönüştürme ve Metinden Konuşmaya sahtecilik sistemidir. A14, STRAIGHT ses kodlayıcıyı kullanan birleşik Sinir Ağları tabanlı Ses Dönüştürme ve Metinden Konuşmaya sahtecilik sistemidir. A15, STRAIGHT ses kodlayıcı yerine hoparlöre bağlı WaveNet ses kodlayıcılar yoluyla dalga formları üretir. A16, A04 ile aynı algoritmayı kullanan bir dalga biçimi birleştirme Metinden Konuşmaya (Text-To-Speech, TTS) sahteciliğidir. Ancak A16, A04'ten farklı bir eğitim setinden oluşturulmuştur. A17, A05 ile aynı VAE tabanlı çerçeveyi kullanan Sinir Ağı tabanlı bir Ses Dönüştürme sistemidir. WORLD ses kodlayıcıyı kullanmak yerine, A17 genelleştirilmiş doğrudan dalga biçimi modifikasyon yöntemini kullanır. A18, Mel-Frekans Kepstral Katsayıları (Mel-Frequency Cepstral Coefficients, MFCCs)'dan konuşma oluşturmak için bir ses kodlayıcı kullanan bir Ses Dönüştürme (Voice Conversion, VC) sahteciliğidir. A19, A06 ile aynı algoritmayı kullanan transfer fonksiyonu tabanlı bir Ses Dönüştürme sistemidir. Ancak A19, A06'dan farklı bir eğitim setinden başlayarak oluşturulmuştur.

Çizelge 5. ASVSpoof2019 PA veriseti oluşturulmasında faydalanılan konuşmacı ve alt veri gruplarında yer alan ses sayıları

		Eğitim	Geliştirme	Test
Sesler	Orijinal	5400	5400	18090
	Sentetik	48600	24300	119367
Konuşmacı	Orijinal	20	10	48

Fiziksel erişim senaryosu orijinal seslerin gizlice kaydedildikten sonra tekrardan oynatılmasını yani yeniden oynatma saldırılarını içerir. Fiziksel erişim senaryosunda sahte seslerin oluşturulması akustik ortam ile kayıt ve sunum cihazlarındaki farklılıkların kombinasyonu ile oluşur. Oda büyüklükleri 3 farklı aralıkta sınıflandırılmaktadır. 2-5 m^2 odalar a, 5-10 m^2 odalar b, 10-20 m^2 odalar c sınıfıdır. Konuşmacı ile ASV arasındaki mesafe 10-50 cm'lik kısa mesafeler a, 50-100 cm'lik orta mesafeler b ve 100-150 cm'lik uzun mesafeler c sınıfıdır. Duvar, tavan ve zemin emme katsayıları ve odadaki konum gibi alanlar arasındaki farklara göre yankılanma değişkenliği sergilediği varsayılmaktadır. Yankılanma düzeyi R ile gösterilen T60 yankılanma süresi cinsinden belirtilir. T60 değerlerinin üç kategorisi vardır. Kısa 50-200 ms'lik T60 süresiyle a, 200-600 ms T60'a sahip orta b, 600-1000 ms'lik T60 süresiyle yüksek c kategorileri vardır. Kayıtlar 3 farklı bölgede yapılmaktadır. A bölgesinde mesafe 10-50 cm, B bölgesinde 50-100 cm ve C bölgesinde 100 cm'in üzerindedir. A bölgesi konuşmacıya en yakın bölge olmakla birlikte bu bölgede çekilen kayıtların konuşmacıdan daha uzakta B ve C bölgelerinde yapılan kayıtlardan daha yüksek kalitede olması beklenmektedir. Orijinal ve sahte sesler 27 farklı ortamda simüle edilmiştir. Oda boyutu, yankılanma düzeyi ve konuşmacı ile ASV arasındaki mesafenin birleşimine göre ortam tanımlanmıştır. Örneğin 'aaa' ortamı 500-200 ms T60 yankılanmasına ve 10-50 cm konuşmacı mesafesine sahip 2-5 m^2 lik küçük bir odaya karşılık gelir. Her biri üç aralığa kategorize edilen üç parametre, 27 farklı akustik ortamın tam setini verir. Her akustik ortamda, yeniden oynatma dokuz farklı saldırı tipine göre simüle edilir. Her biri üç farklı parametre AA, AB, AC, … CB, CC gibi dokuz farklı yeniden oynatma yapılandırması sağlar.

•**ASVSpoof 2021 Veriseti:** ASVSpoof 2021 veriseti 3 farklı senaryo altında değerlendirme veriseti olarak yayınlanmıştır. Eğitim ve geliştirme seti olarak ASVSpoof 2019'un kullanılması ifade edilmiştir (Yamagishi ve ark., 2021). Çizelge 5'te değerlendirme setinin farklı senaryolarla oluşturulan sahte ses sayıları verilmiştir.

Çizelge 6. ASVSpoof2021 verisetinde yer alan ses sayıları

ASVSpoof 2021 Senaryolar	Test Ses Sayısı
Mantıksal Erişim	181566
Fiziksel Erişim	943110
Konuşma Derin Sahte	611829

Mantıksal Erişim (Logical Access, LA) veriseti için ASVSpoof 2021'in odak noktası, kodek ve iletim kanalı değişkenliğine karşı dayanıklı sahtecilik karşı önlemlerinin geliştirilmesi amacıyla oluşturulmuştur. Metinden Konuşmaya, Ses Dönüştürme veya hibrit algoritmalar (sentetik konuşmayla beslenen ses dönüştürme sistemleri) ile oluşturulan orijinal ve sahte konuşma verileri, ya genel anahtarlamalı telefon ağı (Public Switched Telephone Network, PSTN) üzerinden iletilir ya da belirli bir kodek bileşeni kullanan bir İnternet Protokolü (VoIP) ağı üzerinden. Veriler, Alaw ve G.722 kodek bileşenlerini kullanan VoIP ağları üzerinden iletilen denemeleri içerirken, diğer codec bileşenleri de kullanılacaktır. Konuşma verileri bu nedenle kodlama ve iletim bozukluklarının yanı sıra bant genişliği ve örnekleme frekanslarındaki farklılıklara ek olarak sahtecilikle ilgili bozulmalar da gösterebilir, ancak ilave gürültü göstermez. Bant genişliği ne olursa olsun tüm veriler, ortak 16 kHz, örnek başına 16 bit FLAC formatında sağlanacaktır.

Fiziksel Erişim (Physical Access, PA) grubunun oluşturulmasındaki amaç, yeniden oynatma saldırılarının tespitidir. ASVSpoof 2019'daki PA verisetinden farklı olarak daha küçük bir oranda simüle edilmiş tekrarlanan konuşmanın yanı sıra, ağırlıklı olarak gerçek tekrarlanan konuşmayı içermektedir. Kayıtların gerçek fiziksel alanlarda (odalarda) toplanması sonucunda tüm veriler düşük düzeyde arka plan gürültüsü içerir. Ses dosyaları LA göreviyle aynı FLAC dosya formatında oluşturulmuştur.

Konuşma Derin Sahte (Speech deepfake, DF) veri grubunda yer alan sesler LA'ya benzer şekilde, DF görevi orijinal konuşmayı, sıkıştırılmış yapay olarak oluşturulmuş veya dönüştürülmüş konuşmadan ayırmayı içerecek şekilde oluşturulmuştur. Ancak LA senaryosunun aksine DF senaryosu telefon yerine genel ses sıkıştırmasını temsil eder ve ASV sistemi yoktur. Herhangi bir kullanım durumu belirtilmese de senaryo potansiyel olarak sosyal medya veya adli tıp uygulamalarıyla ve bir saldırganın amacının ASV sistemi yerine insan dinleyiciyi kandırmak olduğu durumlarla ilgilidir. Bilinen sıkıştırma algoritmaları mp3 ve m4a'yı içerirken ek teknikleri de içerebilir. Burada sıkıştırma, sıkıştırma-açma

anlamına gelir; Ses dosyaları sıkıştırma bilgisi olmadan paylaşılmıştır. Bunun anlamı veriye sıkıştırma uygulanıp uygulanmadığı veya uygulandıysa bunun ayrıntı bilgisinin verilmemesidir. Amaç çoklu ve bilinmeyen koşullarda konuşma derin sahtekarlıklarının tespitine yönelik yöntemlerin geliştirilmesini teşvik etmektir.

3.2. BTAS 2016

BTAS 2016 veri seti orijinal seslerle birlikte konuşma sentezi ve dizüstü bilgisayar gibi cihazlar yardımıyla yeniden oynatılan ses dönüştürme sentetik konuşma saldırılarını içerir. Veri seti eğitim, geliştirme ve test olarak üçe ayrılmıştır (Korshunov ve ark., 2016). Çizelge 7'de de verildiği üzere eğitim setinde 4973 orijinal seslerin yanında 8 farklı ataktan toplam 38580 sahte ses mevcuttur. Geliştirme setinde ise 4995 orijinal sesler ve eğitim setindeki ataklardan oluşan 38580 sahte ses vardır. Test setinde ise 5576 orijinal seslerin yanında eğitim ve geliştirme setinde bulunan ataklarla birlikte 2 farklı atak ile toplam 44920 sahte ses mevcuttur.

Çizelge 7. BTAS 2016 veriseti alt gruplarında yer alan ses sayıları

		Eğitim	**Geliştirme**	**Test**
Sesler	Orijinal	4973	4995	5576
	Sahte	38580	38580	44920

Verisetinde SS-LP-LP, SS-LP-HQ_LP, VC-LP-LP VC-LP-HQ_LP, RE-PH1-LP, RE-PH2-HQ_LP şeklinde isimlendirilen ataklara maruz bırakılmış sesler bulunmaktadır. Ayrıca sadece test setinde mevcut olan ataklar RE-PH2-PH3 ve RE-LPPH2-PH3 ataklarla oluşturulan sesler bulunmaktadır. Bu ataklar şu şekilde oluşturulmuştur;

•SS-LP-LP atağı; Konuşma Sentezi (Speech Synthesis, SS) ile oluşturulmuş sesin dizüstü bilgisayar (Laptop, LP) kullanılarak oynatıldığı ve hedef doğrulama sisteminin yine bir dizüstü bilgisayarda çalıştığı ataklardan oluşmuştur.

•SS-LP-HQ_LP atağı; Konuşma Sentezi ile oluşturulmuş sesin yüksek kaliteli hoparlörün bağlı olduğu dizüstü bilgisayar kullanılarak oynatıldığı ve hedef doğrulama sisteminin bir dizüstü bilgisayarda da çalıştığı ataklardan oluşmuştur.

•VC-LP-LP atağı; Ses Dönüştürme (Voice Conversion, VC) ile oluşturulmuş sesin dizüstü bilgisayar kullanılarak oynatıldığı ve hedef doğrulama sisteminin bir dizüstü bilgisayar da çalıştığı ataklardan oluşmuştur.

•VC-LP-HQ_LP atağı; Ses Dönüştürme ile oluşturulmuş sesin yüksek kaliteli hoparlörün bağlı olduğu dizüstü bilgisayar kullanılarak oynatıldığı ve hedef doğrulama sisteminin bir dizüstü bilgisayar da çalıştığı ataklardan oluşmuştur.

•RE-LP-LP atağı; Yeniden Oynatma (REplay) ile oluşturulmuş sesin dizüstü bilgisayar kullanılarak oynatıldığı ve hedef doğrulama sisteminin bir dizüstü bilgisayarda çalıştığı ataklardan oluşmuştur.

•RE-PH1-LP atağı; Yeniden Oynatma ile oluşturulmuş sesin Samsung Galaxy S4 telefon kullanılarak oynatıldığı ve hedef doğrulama sisteminin bir dizüstü bilgisayar (Laptop, LP) da çalıştığı ataklardan oluşmuştur.

•RE-PH2-HQ_LP atağı; Yeniden Oynatma ile oluşturulmuş sesin iPhone 3GS kullanılarak oynatıldığı ve hedef doğrulama sisteminin bir dizüstü bilgisayar da çalıştığı ataklardan oluşmuştur.

•RE-PH2-PH3 atağı; Yeniden Oynatma ile oluşturulmuş sesin iPhone 3GS telefon kullanılarak oynatıldığı ve hedef doğrulama sisteminin iPhone 6S de çalıştığı ataklardan oluşmuştur.

•RE-LPPH2-PH3 atağı; dizüstü bilgisayarla kaydedilen orijinal veriler iPhone 3G kullanılarak (yani saldırgan tarafından çalınmış) iPhone 6S'te yeniden oynatılmasıyla oluşturulmuştur.

3.3. FoR (Fake or Real) veriseti

Bu verisetinde yer alan sentetik seslerin oluşturulması birkaç farklı derin öğrenme yöntemlerine dayanmaktadır (Reimao & Tzerpo, 2019). Deep Voice 3, Amazon AWS Polly, Baidu TTS, Google Traditional TTS, Google Cloud TTS, Google Wavenet TTS ve Microsoft Azure TTS kullanılarak

oluşturulan sesler sırasıyla 2645, 21160, 7935, 2645, 5290, 5290 ve 42320 tane olmak üzere toplam 87285 adettir. Orijinal seslerin toplanması ise oldukça karmaşık bir süreçte gerçekleştirilmiştir. Örneğin seslerin çeşitli mikrofonlarla kaydedilmesi gerekir, aksi takdirde makine öğrenmesi algoritması, sentetik bir ifade ile gerçek bir ifade arasındaki farkları öğrenmek yerine, bir kayıt cihazına özgü özelliklere göre sınıflandırmayı öğrenebilir. Her cinsiyetten çok çeşitli seslerin yanı sıra iyi bir aksan çeşitliliğine sahip olunması da önemsenmiştir. Orijinal seslerin toplanması ise Youtube videolarında alınan konuşmalar gibi internet kayıtlarının yanı sıra, Arctic veriseti, LJSpeech veriseti ve VoxForge verisetine dayanmaktadır. Arctic veriseti çok çeşitli aksanları ve tüm cinsiyetleri içeren 7 farklı konuşmacının seslendirdiği 7924 sese sahiptir. LJSpeech veriseti bir kadın konuşmacının 13100 konuşma kaydını içerir. Bu veriseti aynı zamanda Deep Voice 3 modelini eğitmek için kullanılmıştır. VoxForge veriseti herhangi bir kişinin sözlerini kaydedip projeye gönderebileceği açık kaynaklı bir gerçek konuşma verisetidir. Bu çok çeşitli seslere, kayıt cihazlarına ve hatta ses kalitesine sahip bir veriseti oluşturur. Çok çeşitli kayıt cihazlarını kullanan 1200'den fazla kişiden gelen 86000'den fazla ses kaydını içerir. Sosyal medya platformlarından 140 konuşmacının toplam 3720 ses kaydı alınmıştır. Verisetinin 4 farklı versiyonu mevcuttur. Orijinal olarak adlandırılan versiyonda herhangi bir değişiklik veya sınıf/cinsiyet dengelemesi olmaksızın, konuşma kaynaklarından toplanan dosyaları içerir. Bu veriseti versiyonunda toplam 195541 ses mevcuttur.

İkinci versiyon olarak For-norm versiyonu mevcuttur. For-norm olarak adlandırılan normalleştirilmiş veriseti, orijinal verisetiyle aynı dosyaları içerir, ses WAV'a dönüştürülür, 0dBFS'ye normalleştirilir, 16kHz örnekleme hızına alt örneklenir, monoya dönüştürülür ve sesin başında ve sonundaki sessizlikler kaldırılır. Son olarak verisetinde cinsiyet dağılımı da dengelenerek toplamda 69400 se alınmıştır. Üçüncü versiyon olarak For-2seconds adında tüm seslerin 2 saniye ile sınırlandırıldığı yeni bir veriseti oluşturuldu. 2 saniyeden kısa sesler atılırken, 2 saniyeden daha uzun sesler 2 saniyeye kırpıldı. Bu verisetinde toplam 17870 ses mevcuttur. Dördüncü ve son versiyon for-rerecorded sürümüdür. Verisetindeki sesleri normal bir bilgisayar hoparlörü kullanarak 2 saniye boyunca oynatılmış ve bir mikrofon ile kaydetmişlerdir.

3.4. Performans değerlendirme metrikleri

Derin sahte ses tespiti için önerilen modellerin başarımının testi için en sık başvurulan metrikler Eşit Hata Oranı (Equal Error Rate, EER), Toplam Hata Oranının Yarısı (Half Total Error Rate, HTER) ve t-DCF metrikleridir. Bu bölümde bu metriklerden bahsedilecektir.

Eşit hata oranı (Equal error rate, EER):

P_{fa} ve P_{miss} $\emptyset$ eşiği uygulanmış skorların (countermeasure,cm) yanlış alarm ve kaçırma oranlarını gösterir. Yanlış alarm Eşitlik (1)'deki gibi hesaplanır ve yanlış kabul oranını ifade ederken, kaçırma oranı Eşitlik (2)'deki gibi hesaplanır ve yanlış reddetme oranını ifade eder.

$$P_{fa}(\emptyset) = \frac{\#\{sahte\ seslerin\ skorları > \emptyset\}}{\#\{toplam\ sahte\ sayısı\}} \quad (1)$$

$$P_{miss}(\emptyset) = \frac{\#\{orijinal\ seslerin\ skorları \leq \emptyset\}}{\#\{toplam\ orijinal\ sayısı\}} \quad (2)$$

P_{fa} ve P_{miss} sırasıyla monoton olarak azalan ve artan fonksiyonlardır. EER Eşitlik (1) ve Eşitlik (2)'nin eşit olduğu EER eşiğine karşılık gelir ve Eşitlik (3)'deki gibi ifade edilir.

$$EER = P_{fa}(\emptyset_{EER}) = P_{miss}(\emptyset_{EER}) \quad (3)$$

Toplam hata oranının yarısı (Half total error rate, HTER): Otomatik Konuşmacı Doğrulama sistemlerinin başarısı belirli bir eşiğe ($\emptyset$) bağlı olan ve Eşitlik (4) ve Eşitlik (5) de sırasıyla verilen Yanlış Kabul Oranı (False Acceptance Rate, FAR) ve Yanlış Ret Oranı (False Rejection Rate, FRR) temel alınarak yapılmıştır (Witkowski ve ark., 2018).

$$FAR(\emptyset) = \frac{|\{h_{atak} \mid h_{atak} \geq \emptyset\}|}{|\{h_{atak}\}|} \quad (4)$$

$$FRR(\emptyset) = \frac{|\{h_{gerçek} \mid h_{gerçek} < \emptyset\}|}{|\{h_{gerçek}\}|} \quad (5)$$

Burada $h_{gerçek}$ original verinin skorudur, h_{atak} sahte verinin skorudur. Değerlendirme setinin EER değerine dayalı eşik sapmasını belirlemek için geliştirme seti kullanılır. Eşik sapması Eşitlik (6)'da verilmiştir. Son aşamada değerlendirme seti performansı Toplam Hata Oranının Yarısı (Half Total Error Rate, HTER) Denklem (7)'deki gibi hesaplanır.

$$\emptyset_{dev} = arg_{\emptyset} \min \frac{FAR_{dev}(\emptyset) + FRR_{dev}(\emptyset)}{2} \quad (6)$$

$$HTER_{eval}(\emptyset_{dev}) = \frac{FAR_{eval}(\emptyset_{dev}) + FRR_{eval}(\emptyset_{dev})}{2} \quad (7)$$

Burada dev verisetindeki geliştirme setini ifade eder.

Tandem algılama maliyet fonksiyonu (Tandem detection cost function, t-DCF):

Tandem Algılama Maliyet Fonksiyonu Eşitlik (8)'deki gibi hesaplanır (Kinnunen ve ark., 2020).

$$t - DCF(s) = C_1 P_{miss}^{cm}(s) + C_2 P_{fa}^{cm}(s) \quad (8)$$

Burada $P_{miss}^{cm}(s)$ s eşiğinde yanlış etiketleme oranını (miss rate), $P_{fa}^{cm}(s)$ yanlış alarmı (false alarm, fa) ifade eder ve Eşitlik (9) ve Eşitlik (10)'daki gibi hesaplanır. (CM: countermeasure, skor)

$$P_{fa}^{cm}(S) = \frac{\#\{sahte\ seslerin\ skorları > \emptyset\}}{\#\{toplam\ sahte\ sayısı\}} \quad (9)$$

$$P_{fa}^{cm}(S) = \frac{\#\{sahte\ seslerin\ skorları > \emptyset\}}{\#\{toplam\ sahte\ sayısı\}} \quad (10)$$

C_1 ve C_2 sabitleri Denklem (11)'de verilmiştir, t-DCF maliyetleri öncelikler ve ASV sistemi algılama hataları tarafından belirlenir:

$$C_1 = \pi_{tar}(C_{miss}^{cm} - C_{miss}^{asv} P_{miss}^{asv}) - \pi_{non} C_{fa}^{asv} P_{fa}^{asv}$$
$$C_2 = C_{fa}^{cm} \pi_{spoof}(1 - P_{miss,spoof}^{asv}) \quad (11)$$

Burada C_{miss}^{asv} ve C_{fa}^{asv} sırasıyla ASV sisteminin orijinal sesi sahte olarak etiketleme ve sahte sesi orijinal olarak etiketleme maliyetleridir. Benzer şekilde CM sistemleri için, orijinal denemesinin reddedilmesi ve bir sahte denemenin kabulü için sırasıyla C_{miss}^{cm} ve C_{fa}^{cm} olmak üzere 2 maliyet atanmıştır. Hedef (π_{tar}), hedef olmayan (π_{non}) ve sahte (π_{spoof}) priori olasılıklarını (priori probabilities) ileri sürülmüştür. Olasılıkların toplamı 1'e eşittir. Son olarak P_{miss}^{asv}, P_{fa}^{asv}, $P_{miss,spoof}^{asv}$ ASV sisteminin sabit tespit hatası oranlarıdır. Bunlardan ilk ikisi geleneksel yanlış etiketlenen orijinal sesler ve yanlış alarm oranı (doğru işaretlenen sahte sesler)' dır. Sonuncusu ise ASV tarafından reddedilen sahte seslerin oranıdır.

Ham t-DCF değerinin yorumlanması zor olabilir. Bu nedenle t-DCF için Eşitlik (12)'de verilen normalleştirme işlemi yapılır.

$$t - DCF_{norm}(s) = \frac{t - DCF(s)}{t - DCF_{default}} \quad (12)$$

Burada $t-DCF_{default}$, $t-DCF_{default} = \min\{C_1, C_2\}$ olarak tanımlanan varsayılan bir parametredir. C_1 ve C_2 Eşitlik (11)'den elde edilir. $P_{miss}^{cm}(s) = 1$ ve $P_{fa}^{cm}(s) = 0$ (CM eşiği s → ∞) ve $P_{miss}^{cm}(s) = 0$ ve $P_{fa}^{cm}(s) = 1$ (CM eşiği s → -∞) . İlk durumda normalleştirilmiş t-DCF Eşitlik (13)'deki gibi yazılır.

$$t-DCF_{norm}(s) = P_{miss}^{cm}(s) + \alpha P_{fa}^{cm}(s) \quad (13)$$

Burada $\alpha = C_2/C_1$, bu durumda,

$$t-DCF_{norm}(s) = \beta P_{miss}^{cm}(s) + P_{fa}^{cm}(s) \quad (14)$$

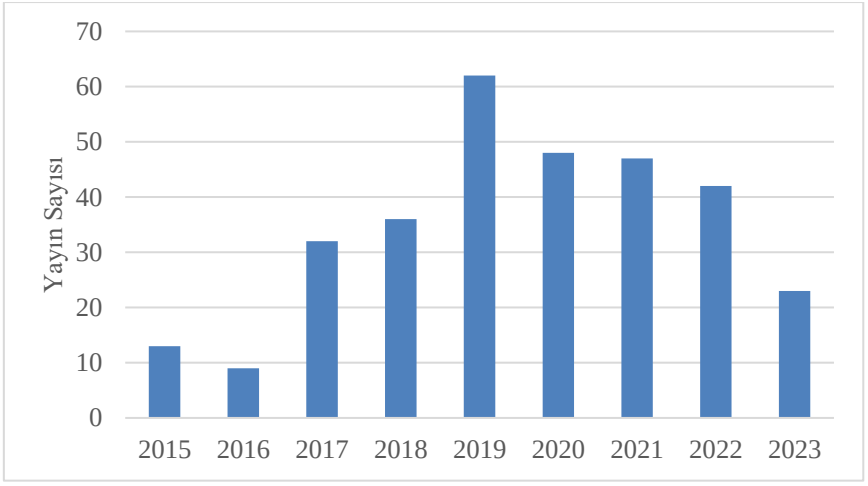
Burada $\beta = C_1/C_2$. ASVspoof 2019 için genellikle C1 > C2 olur, dolayısıyla Eşitlik (14)'de verilen ikinci durum geçerlidir. Normalleştirilmiş t-DCF, CM eşiğinin (s) bir fonksiyonudur. ASVSpoof 2019 eşik ayarına (kalibrasyona) odaklanmaz. Dolayısıyla minimum normalleştirilmiş t-DCF şu şekil Eşitlik (15)'deki gibi tanımlanır.

$$t-DCF_{norm}^{min} = t-DCF_{norm}(s_*) \quad (15)$$

Burada $s_* = \arg min_s\, t-DCF_{norm}(s)$ ground truth kullanılarak belirlenen optimal eşiktir.

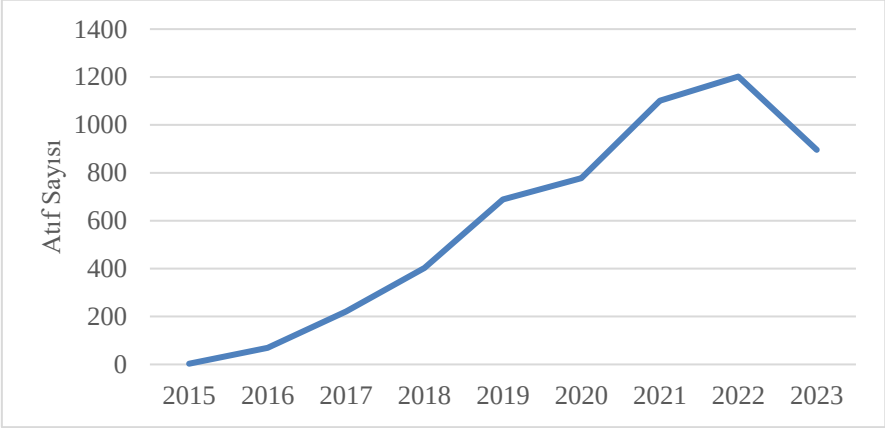
4. Derin Sahte Ses Tespitine Yönelik Literatürde Yapılan Çalışmalar

Bu bölümde derin sahte ses tespiti için literatürde önerilen yöntemlerin Bölüm 2'de yer alan verisetleri ve metrikler bazında detaylı analizleri verilecektir. Çalışmalar önceden de ifade edildiği gibi geleneksel yöntemler ile özellik elde eden çalışmalar ve derin öğrenmeye dayalı özelliklerin elde edildiği çalışmalar olmak üzere iki alt grupta değerlendirilecek şekilde sunulmuştur. Google Scholar platformunda derin öğrenme tabanlı yöntemler ile derin sahte ses manipülasyonu tespiti yapan yayınlar incelenmiştir. Bu amaçla arama işlemi "asvspoof, deep learning" anahtar kelimelerinin birleşimiyle yapılmıştır. Web of Science platformunda yapılan literatür araştırma yöntemi şu şekildedir: Konu kriteri altında "ASVspoof 2019", "ASVspoof 2017" ve "ASVspoof 2015" veriseti isimleri anahtar kelime olarak belirlenip ayrı ayrı taranmış ve yüksek atıf kriterine göre sıralanmıştır. Web of Science kullanılarak belirlenmiş yayınların yıl bazında sayısını gösteren grafik Şekil2'de verilmiştir. Makaleye eklenecek yayınlarda yayınlanma yılı kriterinden daha çok veriseti çeşitliliğine öncelik verilmiştir. Öte yandan derin sahte ses tespiti için kompleks yöntemler kullanan yayınlarda makalede tercih edilmemiştir. WoS indeksi sonucuna göre makalede en çok 2017 yılında ilgili kriterlerde çalışma bulunmaktadır.



Şekil 3. WoS tarafından indekslenen yıl bazında yayın sayısı.

Yıl bazında atıf sayısını gösteren grafik Şekil 2'de verilmiştir. Şekilde de görülen atıf sayısında yıllara göre artış konunun önemini vurgulamaktadır. En yüksek atıflı yayın 183 atıfla Lavrentyeva ve ark. (2017)'a aittir.



Şekil 4. WoS tarafından indekslenen yayınların yıl bazında atıf sayısı.

4.1. Geleneksel yöntemler ile özellik elde eden çalışmaların detaylı analizi

Derin sahte seslerin tespitinde girdi olarak alınan ses dosyasındaki akustik özelliklerden faydalanan ve geleneksel makine öğrenimi yaklaşımları ile sınıflama gerçekleştiren yöntemler bu bölüm kapsamında detaylandırılmıştır.

Ele alınan çalışmalardan ilki Wang ve ark. (2015) tarafından önerilmiştir. Şekil 5'te blok diyagramı verilen çalışmada giriş sesinden Mel-Frekans Kepstral, Değiştirilmiş Grup Gecikmesi Kepstral Katsayıları (Modified Group Delay Cepstral Coefficients, MGD) ve Fourier spektrumundan alınan Bağıl Faz (Relative Phase) Bilgisi özellik vektörü olarak çıkarılmıştır. Elde edilen özelliklerin sınıflandırılmasında Gauss Karışım Modeli kullanılmıştır. Önerilen model ASVSpoof 2015 veriseti ile eğitilmiş ve geliştirme setinde üç özellik vektörünün füzyonu ile en yüksek EER değerinin elde edildiği gösterilmiştir. Değerlendirme setinde yapılan deneylerde ise; Değiştirilmiş Grup Gecikmesi Kepstral Katsayıları ve Bağıl Faz füzyonu sonucunda %3.726 EER sonucunu elde ettiği gözlemlenmektedir. Bilinmeyen saldırılar karşısında yöntemin s10 haricinde başarısının oldukça yüksek olduğu deneysel sonuçlarla gösterilmiştir.

Şekil 5. Üç farklı akustik özelliğin Gauss Karışım Modeli ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Wang ve ark., 2015).

Xiao ve ark. (2015) tarafından önerilen bir diğer çalışmada giriş sesinden iki tip magnitüd-tabanlı ve beş tip faz-tabanlı (Phase-Based) olmak üzere toplam yedi farklı akustik özellik çıkarılmıştır. Magnitüd tabanlı özellikler Log Magnitüd Spektrumu (Log Magnitude Spectrum, LMS) ve Artık Log Magnitüd Spektrumu (Residual log magnitude spectrum, RLMS)'dur. Faz-Tabanlı özellikler ise Grup Gecikmesi (Group Delay,GD), Değiştirilmiş Grup Gecikmesi (Modified Group Delay, MGD), Anlık Frekans Türevi(Instantaneous frequency derivative, IF), Temel Bant Faz Farkı (Baseband Phase

Difference, BPD) ve Pitch Senkron Fazı (Pitch Synchronous Phase, PSP)'dır. Şekil 6'da temel adımları verilen yaklaşım sınıflandırma aşamasında Çok Katmanlı Algılayıcılardan (Multi Layer Perceptron, MLP) faydalanmaktadır. Modelin eğitim ve testi için ASVSpoof 2015 verisetini kullanan çalışmada, tekli özelliklerle elde ettiği sonuçların yanında füzyon sonuçlarını da rapor edilmiştir. Geliştirme setinde Grup Gecikmesi özelliği %0.114 EER ile en yüksek başarıya sahip iken füzyon işlemi sonrasında %0.001 EER değeri elde edilmiştir. Değerlendirme setinde füzyon sonucu %2.62 EER'dir. Önerilen yaklaşımın bilinen ve bilinmeyen ataklardaki başarısı sırası ile %0.29 EER ve %5.23 EER olarak raporlanmıştır.

Şekil 6. Yedi farklı akustik özelliğin çok katmanlı algılayıcı ile sınıflandırılması ile derin sahte ses tespiti yöntemi (Xiao ve ark., 2015).

Patel & Patil (2015) tarafından önerilen çalışmada giriş sesinden CFCC ve IF değişiminin kombinasyonuna dayalı CFCCIF ile Mel-Frekans Kepstral Katsayıları akustik özellik olarak çıkarılmıştır. Sınıflandırma Gauss Karışım Modeli-Log Olasılık kullanılarak yapılmıştır. Şekil 7 yöntemin genel akışını vermektedir. Yöntemin eğitim ve testinde ASVSpoof 2015 veriseti kullanılmıştır. Geliştirme setinde MFCC ve CFCCIF füzyonu ile %0.83 EER başarı elde edilirken değerlendirme setinde MFCC ve CFCCIF füzyonu ile ortalama %1.211 EER değerine ulaşılmıştır. Değerlendirme setinde bilinen ataklarda ortalama %0.407899 EER başarısı, bilinmeyen ataklarda ortalama %2.013 EER başarısı rapor edilmiştir.

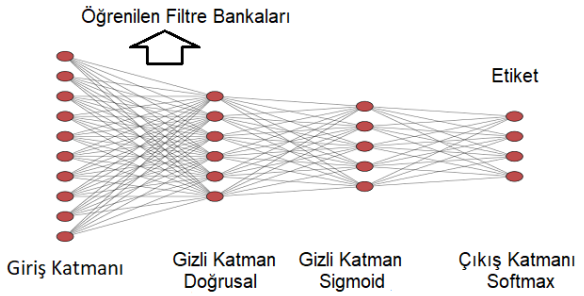
Şekil 7. Üç farklı akustik özelliğin Gauss Karışım Modellemesi ile sınıflandırılması ile derin sahte ses tespiti yöntemi (Patel & Patil, 2015).

Paul ve ark. (2017) önerilen çalışmada girişten alınan sesten Mel-Frekans Kepstral Katsayıları ve Sabit-Q-Kepstral Katsayılarını özellik olarak elde etmiş ve çıkarılan özelliklerin sınıflandırılmasında Gauss Karışım Modeli-Maksimum Olabilirlik yaklaşımından faydalanmıştır. Şekil 8'de yönteme ilişkin akış sunulmaktadır. Çalışmada sonuçların verilmesi sürecinde ASVSpoof2015 ve BTAS 2016 verisetleri kullanılmış ve BTAS 2016 verisetinde en iyi başarı geliştirme setinde 0 EER ve değerlendirme setinde 0.76 EER ile Sabit-Q-Kepstral Katsayıları kullanılarak elde edilmiştir. Mel-Frekans Kepstral Katsayılarının özellik vektörü olarak seçilmesi durumunda ise geliştirme setinde 0.19 ve değerlendirme setinde 3.67 EER alınmıştır. ASVSpoof 2015 değerlendirme verisetinde MFCC ve CQCC özellikleri tek başlarına düşük performans verse de CQCC özelliğinin 0.44 EER ile tüm genelleme senaryolarında daha iyi performans sağladığı görülmektedir. Yazarlar yöntemin genelleme

problemine sahip olduğunu ve ekstra ataklı seslerin eğitim setinde yer almaması durumunda bu seslerle yapılan testlerdeki başarının düştüğünü belirtmişlerdir.

Şekil 8. Mel-Frekans Kepstral Katsayıları ve Sabit-Q-Kepstral Katsayıları özelliklerinin Gauss Karışım Modeli ve Maksimum Olabilirlik yaklaşımları ile sınıflandırıldığı derin sahte ses tespiti yöntemi (Paul ve ark., 2017).

Yu ve ark. (2017) tarafından önerilen çalışmada giriş sesi 20ms uzunluğunda ve 10ms adım boyutlu çerçevelere bölünmüştür. Çerçevelerden güç spektrumu hesaplanmış ve ağın girişi olarak kullanılmıştır. Çalışmada Derin Sinir Ağı filtre bankası kepstral katsayıları ile genişletilerek yeni bir filtre bankası tabanlı kepstral özellik yaklaşımı Şekil 9'daki gibi önerilmiştir. DNN filtre bankası, orjinal ve sentetik ses verilerin Filtre Bankası Sinir Ağının eğitilmesi ile oluşturulmuştur. (Filter Bank Neural Network,FBNN). DNN giriş katmanı ile ilk gizli katman arasındaki öğrenilmiş ağırlık matrisi, 'özel öğrenilen filtre bankası' olarak kabul edilmiştir. Bu gizli katman düğümlerinin sayısı, filtre bankası kanallarının sayısına karşılık gelir ve ağırlık matrisinin her sütunu, her filtrenin frekans yanıtı olarak ele alınmıştır. Geleneksel elle tasarlanmış filtre bankalarının aksine, öğrenilen filtre bankasının filtreleri farklı kanallarda farklı şekillere sahiptir. Bu da gerçek ve sentetik konuşma arasındaki ayırt edici özelliğin daha etkin bir şekilde sınıflandırılmasını sağlamıştır. İlk gizli katmandan üretilen DNN özelliği, bir tür filtre bankası özelliği olarak ele alınmıştır. Öğrenilen filtre bankası DNN-FBCC olarak isimlendirilmiştir. Üretilen Katsayıların sınıflandırılması için Gauss Karışım Modeli-Maksimum Olasılık kullanılmıştır. Modelin eğitimi ve performans analizinde ASVSpoof 2015 verisetinden faydalanılmış ve Geliştirme setinde 0.09 EER, değerlendirme setinde ise 0.56 EER değerlerine ulaşıldığı rapor edilmiştir.

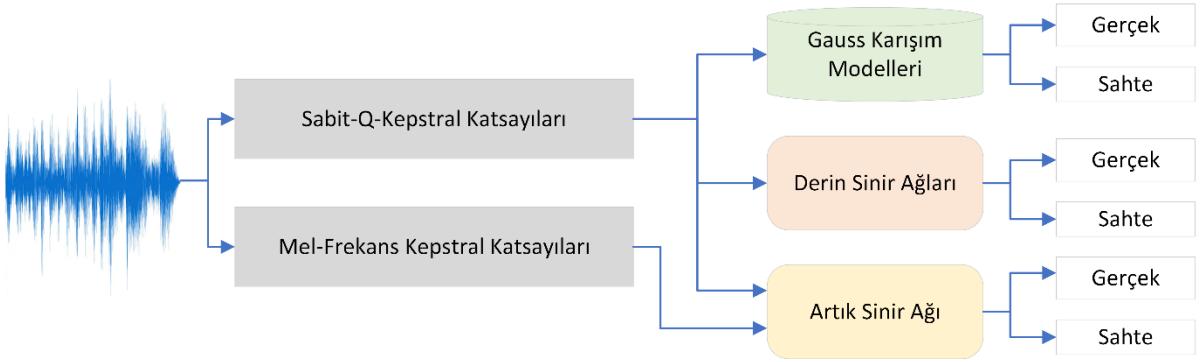


Şekil 9. Öğrenilen filtre bankaları kullanımına dayalı derin sahte ses tespiti yöntemi (Yu ve ark., 2017).

Witkowski ve ark. (2017) tarafından önerilen ve detayları Şekil 10'da verilen çalışmada giriş sesinden Sabit-Q-Kepstral Katsayıları, Kepstrum, Mel-Frekans Kepstral Katsayıları, IMFCC, LPCCres olmak üzere beş farklı akustik özellik üretilmektedir. Çıkarılan özelliklerin sınıflandırılmasında ise Gauss Karışım Modellemesi yaklaşımı kullanılmış ve yapılan deneysel çalışmalarda akustik özelliklerin farklı frekans aralıklarının başarıma etkisi değerlendirilmiştir. ASVSpoof 2017 yeniden oynatma alt verisetinde yer alan geliştirme Setinde Mel-Frekans Kepstral Katsayıları ile en yüksek başarıya 4000-8000 Hz. frekans aralığında 3.16 EER ile ulaşıldığı görülmektedir. Değerlendirme setinde Sabit-Q-Kepstral Katsayıları kullanıldığında 17.31 EER sonucu rapor edilmiştir.

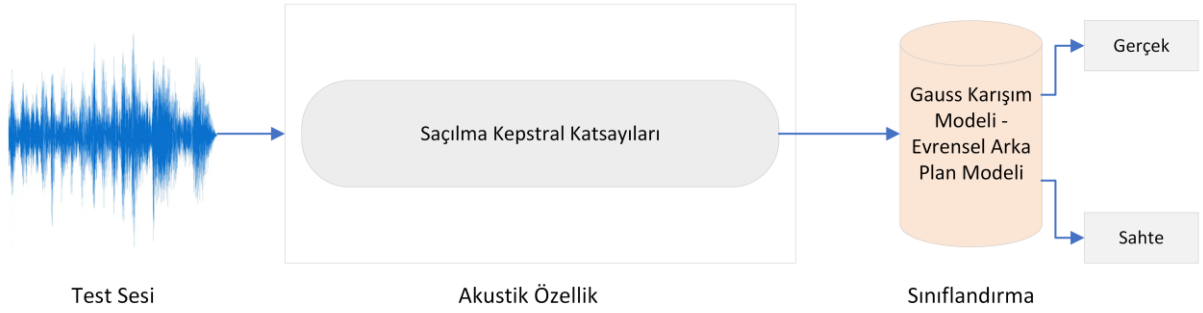
Şekil 10. Beş farklı akustik özelliklerin Gauss Karışım Modellemesi ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Witkowski ve ark., 2017).

Chen ve ark. (2017) tarafından önerilen bir diğer çalışmada giriş sesinden Sabit-Q-Kepstral Katsayıları ve Mel-Frekans Kepstral Katsayıları olmak üzere iki farklı akustik özellik çıkarılmıştır. Çıkarılan özelliklerin sınıflandırılmasında ise Gauss Karışım Modelleri, Derin Sinir Ağları ve Artık Sinir Ağı (ResNet) modellerinin kullanılması önerilmiştir. Şekil 10'da çalışmayan ilişkin genel gösterim sunulmaktadır. Modelin eğitim ve testinde ASVSpoof 2017 verisetinden faydalanılmıştır. Sabit-Q-Kepstral Katsayılarının Gauss Karışım Modelleri, Artık Sinir Ağı ile ve Mel-Frekans Kepstral Katsayılarının Artık Sinir Ağı ile sınıflandırılmasının üçlü füzyonu sonucunda geliştirme setinde 2.58 EER, değerlendirme setinde ise 13.30 EER elde edilmiştir.



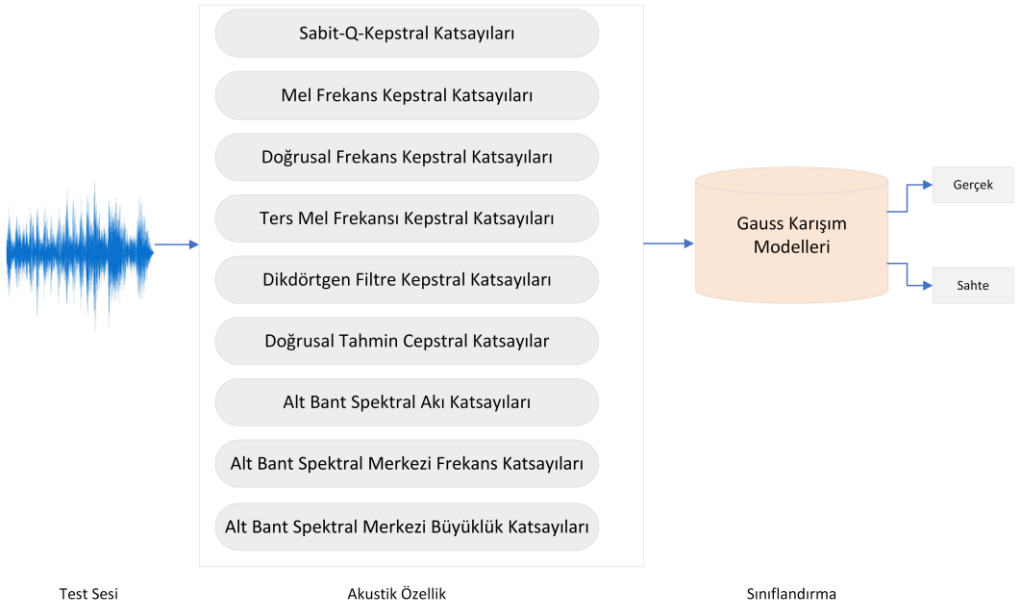
Şekil 11. Sabit-Q-Kepstral Katsayıları ve Mel-Frekans Kepstral Katsayılarının üç farklı yaklaşımla sınıflandırılarak değerlendirilmesine dayalı derin sahte ses tespiti yöntemi (Chen ve ark., 2017).

Sriskandaraja ve ark. (2017) tarafından önerilen çalışmada giriş sesinin SC katsayılarının logaritmalarından üretilen vektör üzerinde DCT uygulandıktan sonraki ilk altmış katsayı ile özellik vektörü elde edilmiştir. Saçılma Kepstral Katsayıları (SCC) olarak adlandırılan bu özelliklerin sınıflandırmasında GMM-UBM'den faydalanılmıştır. Çalışmada deneysel sonuçların elde edilmesi aşamasında SAS Corpus ve ASVSpoof 2015 verisetleri kullanılmıştır. SCC özelliği çıkarma aşamasında 256ms pencere boyutu (window size) kullanıldığında SAS Corpus geliştirme setinde 0.31 EER başarısı elde edilmiştir. 1. seviye özellik kullanıldığında ise bilinen ataklarda 0.55, bilinmeyen ataklarda 4.90 EER değeri raporlanmıştır. SCC ikinci seviye özelliğinde ise bilinen ataklarda 0.04, bilinmeyen ataklarda 3.96 ve genel EER durumunda 2.92 EER değerleri elde edilmiştir. ASVSpoof 2015 test setinde ise bilinen ataklarda 0.02, bilinmeyen ataklarda 0.33 EER değerleri alınmış ve genel EER başarısı 0.18 olarak raporlanmıştır.



Şekil 12. Saçılma Kepstral Katsayılarının Gauss Karışım Modeli-Evrensel Arka Plan Modeli yaklaşımıyla sınıflandırılarak değerlendirilmesine dayalı derin sahte ses tespiti yöntemi (Sriskandaraja ve ark., 2017).

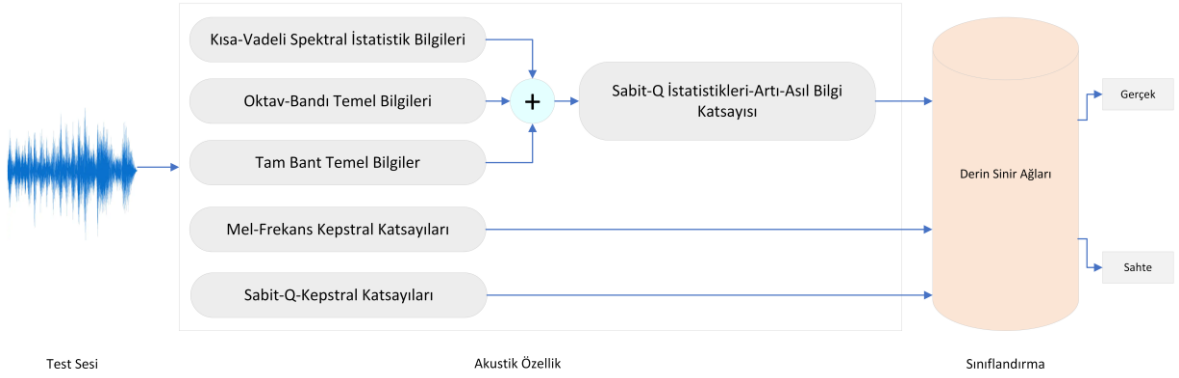
Font ve ark. (2017) tarafından önerilen çalışmada giriş sesinden Sabit-Q-Kepstral Katsayıları, Mel Frekans Kepstral Katsayıları, Doğrusal Frekans Kepstral Katsayıları, IMFCC, RFCC, LPCC, SSFC, SCFC ve SCMC akustik özellikleri çıkarılır. Sınıflandırma Gauss Karışım Modelleri kullanılarak yapılmıştır. Sabit-Q-Kepstral Katsayılar CMN ile normalize edilmiştir. Çalışmada ASVSpoof 2017 ve BTAS 2016 verisetlerinden faydalanılmıştır. İki farklı çalışma stratejisi izlenen çalışmada Birinci stratejide mümkün olduğunca düşük geliştirme seti hatası elde etmek amaçlanmışken, ikinci stratejide veritabanları arası deneyler gerçekleştirilerek, tutarlı performans gösteren bir konfigürasyon bulunmaya çalışılmıştır. IMFCC kullanıldığında ASVSpoof 2017 geliştirme setinde 3.85 EER, değerlendirme setinde %30.91 EER başarısı elde edilmiştir. SCMC kullanımında geliştirme ve değerlendirme setlerinde sırası ile 9.32 ve 11.49 EER değerleri üretilmiştir. Veritabanları arası (Cross-database) çalışmada yapılmış ve BTAS 2016 ile eğitim yapılıp (LPCCs kullanımında) BTAS 2016 geliştirme setinde test yapıldığında 1.09 EER değere ulaşılırken ASVSpoof 2017 geliştirme setinde 22.81 EER skoru üretilmiştir. Aynı özellikle ASVSpoof 2017'de eğitim yapıldığında ise BTAS 2016 geliştirme setinde 13.88 EER, ASVSpoof 2017 geliştirme setinde 10.70 EER skoru almıştır.



Şekil 13. Dokuz farklı akustik özelliklerin Gauss Karışım Modellemesi ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Font ve ark., 2017).

Yang ve ark. (2018) tarafından önerilen çalışmada giriş sesinden STSSI, OPI ve FPI akustik özellikleri çıkarılmıştır. STSSI farklı CQT-spektral kutular üzerinden birinci ve ikinci dereceden istatistiklerin elde edildiği çerçeve düzeyinde istatistik bilgilerini içerirken OPI oktav bölümleme ve

ayrık kosinüs dönüşümünün (DCT) uygulandığı oktav bilgisini taşımaktadır. FPI ise Sabit-Q-Dönüşümü spektrumundan tam bant ilke bilgisini formüle eder. Son olarak, sahtecilik tespiti için bir özellik olarak delta ve hızlanma (acceleration) katsayılarını oluşturmak üzere üç alt özellik birleştirilmiştir. Önerilen özellik Sabit-Q İstatistikleri-Artı-Asıl Bilgi Katsayısı (Constant-Q-Statistics-Plus-Principal Information Coefficient, CQSPIC) olarak adlandırılmıştır. Önerilen özelliğe ek olarak Mel Frekans Kepstral Katsayıları ve Sabit-Q-Kepstral Katsayıları akustik özellikleri de çıkarılmıştır. Bu özelliklerin sınıflandırılması Derin Sinir Ağı kullanılarak yapılmış ve çalışmada ASVSpoof 2015 veriseti kullanılmıştır. ASVSpoof 2015 değerlendirme setinde CQSPIC hızlandırma özelliği ile ortalama 0.038 EER değerleri raporlanmış, ASVSpoof 2017 değerlendirme verisetinde OPI, FPI ve STSSI'nin ortalamasının birleştirilmesiyle 11.09 ortalama EER değeri elde edilmiştir.



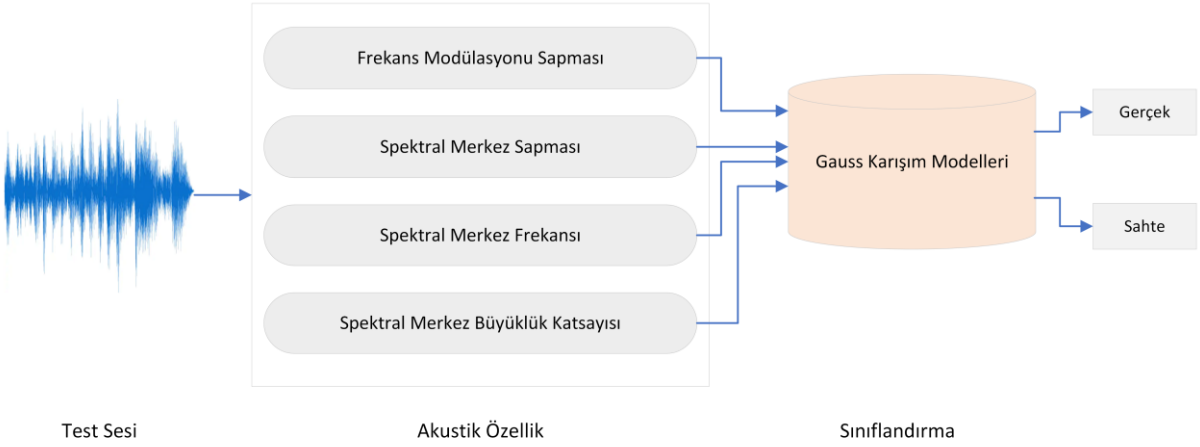
Şekil 14. Altı farklı akustik özelliklerin Derin Sinir Ağları ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Font ve ark., 2018).

Suthokumar ve ark. (2018) tarafından önerilen çalışmada giriş sesinden MCF özelliği çıkarılır. MCF kosinüs katsayıları (MCF-CC) olarak adlandırılan özelliği oluşturmak için Modülasyon Ağırlık Merkezi Frekansı boyunca DCT uygulanmakta ve Akustik frekanslar boyunca normalleştirilmiş modülasyon spektrumunun (Normalized Modulation Spectrum) sıfırıncı modülasyon kutusu enerjileri MSE olarak adlandırılmaktadır. DCT, MSE'nin boyutsallığını azaltmak için gerçekleştirilmiş ve kompakt MSE Kepstral Katsayısı (MSE Cepstral Coefficient, MSE-CC) özellikleri çıkarılmıştır. Bu akustik özelliklere ek olarak STCC özelliği de elde edilmiştir. Bu akustik özelliklerin sınıflandırması için GMM kullanılmıştır. Çalışmada ASVSpoof 2017 versiyon değerlendirme setinde üç akustik özelliğin füzyonunda 6.54 EER, versiyon iki verisetinde üç özelliğin füzyonunda 6.32 EER elde edildiği görülmüştür.

Şekil 15. Üç farklı akustik özelliklerin Gauss Karışım Modellemesi ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Suthokumar ve ark., 2018).

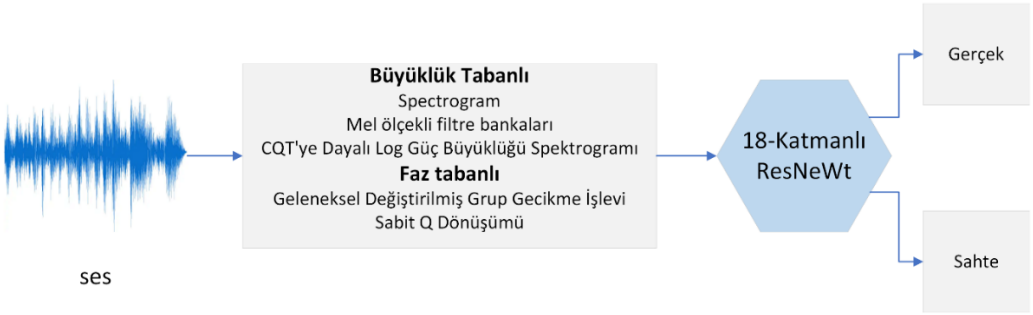
Gunendradasan ve ark. (2018) tarafından önerilen çalışmada giriş sesinden FMD, SCD, SCF ve SCMC özellikleri çıkarılmıştır. Bu özelliklerin sınıflandırılmasında yazarlar GMM'den faydalanmış ve önerilen modelin eğitim/testi için ASVSpoof 2017 verisetini kullanmıştır. Değerlendirme setinde

Spektral Merkez Büyüklük Katsayısı, Spektral Merkez Sapması ve Spektral Merkez Frekansı skor düzeyinde füzyon işlemi gerçekleştirilmiş ve 9.20 EER metrik değerinin elde edildiği görülmüştür.



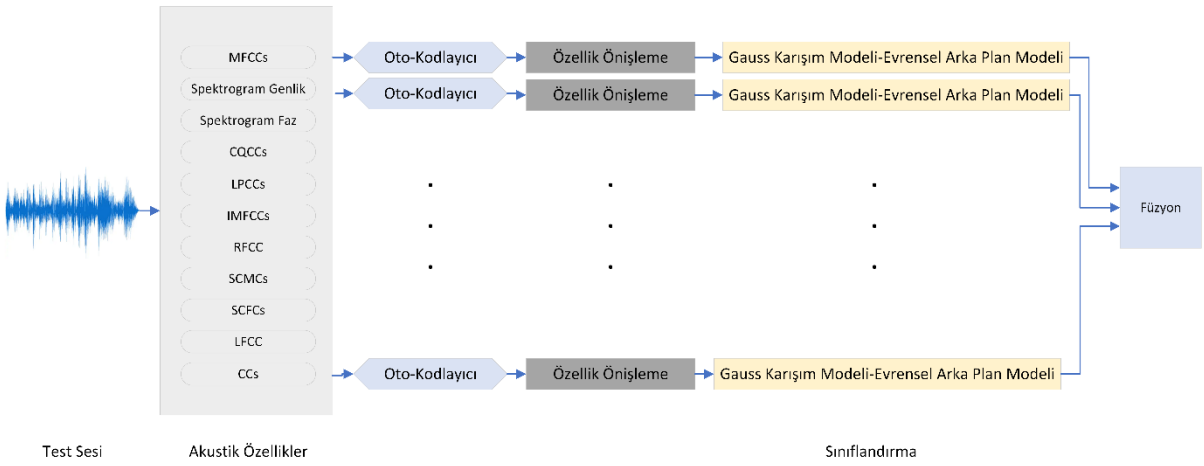
Şekil 16. Dört farklı akustik özelliklerin Gauss Karışım Modellemesi ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Gunendradasan ve ark., 2018).

Cheng ve ark. (2019) tarafından önerilen çalışmada giriş sesinin magnitüd tabanlı ve faz tabanlı zaman-frekans temsilleri çıkarılmıştır. Kullanılan magnitüd tabanlı özellikler Spectrogram, Mel-ölçekli filtre bankaları (Mel scale filter banks, MelFbanks), CQT'ye dayalı log güç büyüklüğü spektrogramı 'dır. Faz tabanlı özellikler ise MGD ve CQTMGD'dir. Özelliklerin eğitimi için 18-katmanlı ResNet (ResNet18) yapısı referans alınarak 18-katmanlı bir ResNeWt (ResNeWt18) mimarisi oluşturulmuştur. Çalışmada ASVSpoof2019 PA seti üzerinde deneyler yapılmıştır. Büyüklük tabanlı özellikler ve Faz tabanlı özelliklerle alınan sonuçlar füzyon yapılmış ve geliştirme setinde 0.20 EER ve 0.0049 t-DCF, değerlendirme setinde 0.39 EER ve 0.0096 t-DCF elde edilmiştir.



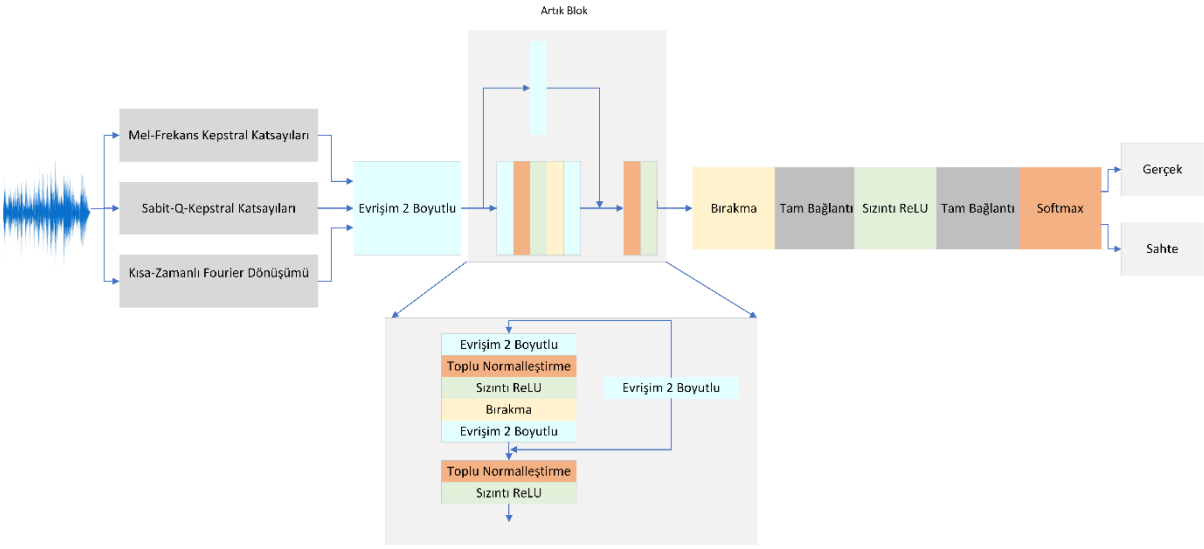
Şekil 17. Büyüklük ve Faz tabanlı özellikler ile 18 katmanlı ResNeWt mimarisi ile derin sahte ses tespiti (Cheng ve ark., 2019).

Balamurali ve ark. (2019) tarafından önerilen çalışmada giriş sesi ön işlem sonrası çerçevelere ayrılmıştır. Çerçeveleme işlemi %50 örtüşen hamming penceresi ile gerçekleştirilmiş ve her çerçeveden 11 farklı özellik çıkarılmıştır. Bu özellikler Mel-Frekans Kepstral Katsayıları, spektrogram genlik, spektrogram faz, Sabit-Q-Kepstral Katsayılar, LPCC, IMFCC, RFCC, LFCC, SCMC ve CC'dir. Özelliklerin sınıflandırması için GMM-UBM kullanılmıştır. Modelin eğitimi ve testinde ASVSpoof2017 verisetindeki yeniden oynatma ataklarının yer aldığı veriseti grubu kullanılmıştır. Deneyler her bir özellik için ayrı ayrı ve tüm özelliklerin füzyonu için yapılmıştır. Yapılan testlerde değerlendirme setinde en iyi başarının, tüm özellikler ve oto-kodlayıcıyla elde edilen özelliklerin füzyon işlemine alınması ile elde edildiği gösterilmiş ve 10.8 EER'ye ulaşılmıştır.



Şekil 18. 11 farklı özelliğin Gauss Karışım modeli ve evrensel arka plan modeli ile sınıflandırıldığı derin sahte ses tespiti yöntemi (Balamurali ve ark., 2019).

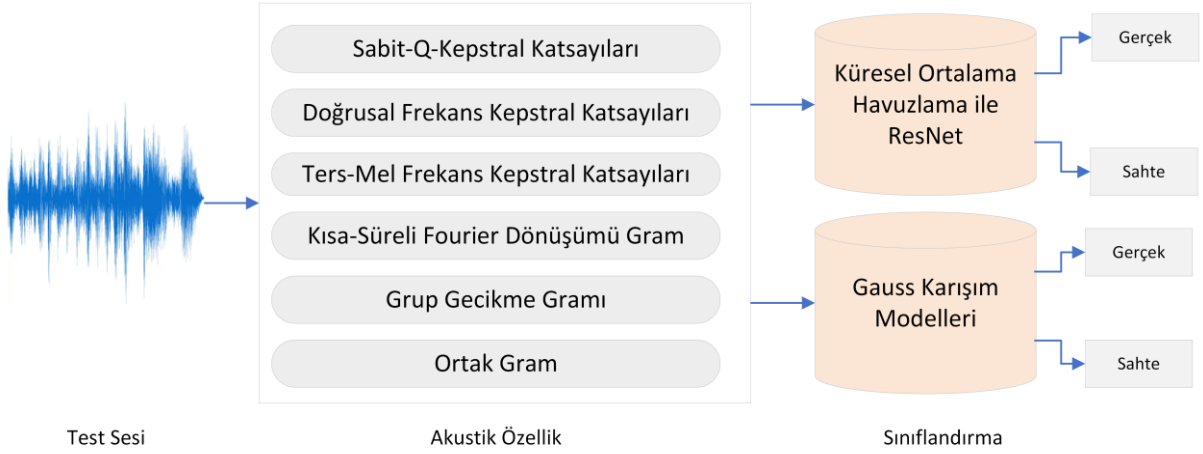
Alzantot ve ark. (2019) tarafından önerilen çalışmada giriş sesinden Mel-Frekans Kepstral Katsayıları, Sabit-Q-Kepstral Katsayıları ve STFT'nin Logaritmik Büyüklüğü akustik özellikleri çıkarılmıştır. Mel-Frekans Kepstral Katsayıları çıkarılırken, Mel-Frekans Kepstral Katsayı türevlerinin kepstral katsayı dinamiklerini yakalaması için MFCC'yi, birinci dereceden MFCC ve MFCC'nin ikinci türeviyle birleştirilmiştir. Artık Konvolüsyonel Sinir Ağı ile elde edilen özellikler sınıflandırılmıştır. Kullanılan mimariye Şekil 19'da yer verilmiştir. Yöntemin eğitim ve testi için ASVSpoof 2019 veriseti kullanılmıştır. Üç akustik özelliğin Artık Konvolüsyonel Sinir Ağı ile sınıflandırılmasının skor seviyesinde füzyon işlemi sonucunda LA geliştirme setinde 0.0 t-DCF ve 0.0 EER, değerlendirme setinde ise 0.1569 t-DCF ve 6.02 EER alınmıştır. PA geliştirme setinde 0.0581 t-DCF ve 2.65 EER, değerlendirme setinde 0.0693 t-DCF ve 2.78 EER değerleri raporlanmıştır.



Şekil 19. Üç farklı akustik özelliklerin önerilen ResNet mimarisi ile sınıflandırılmasına dayanan derin sahte ses tespiti yöntemi (Alzantot ve ark., 2019).

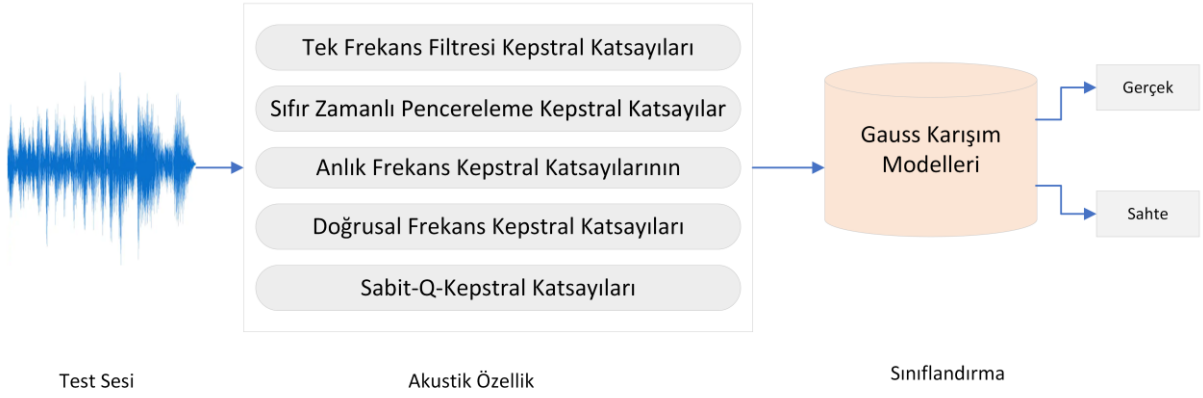
Cai ve ark. (2019) tarafından önerilen çalışmada giriş sesinden Sabit-Q-Kepstral Katsayıları, LFCC, IMFCC, STFT Gram, GD gram ve Ortak Gram akustik özellikleri çıkarılır. Ayrıca, eğitim verilerinin miktarını artırmak için ham dalga formuna hız pertürbasyonunu uygulayarak basit ama etkili bir veri arttırma stratejisi uygulanmıştır. Sınıflandırma yapmak amacıyla GAP katmanının da dahil olduğu ResNet ve GMM kullanılmıştır. Çalışmada ASVSpoof 2019 yeniden oynatma verisetinden deneylerde faydalanılmıştır. LFCC'nin, IMFCC'nin, veri arttırılmış STFT gramın, GD-gram ve veri

arttırılmış GD-gramın ve veri arttırım uygulanmış Joint Gramın skor seviyesinde füzyonu ile geliştirme setinde 0.24 EER, 0.0064 min t-DCF, değerlendirme setinde ise 0.66 EER, 0.0168 min t-DCF başarısı vardır.



Şekil 20. Altı farklı akustik özelliklerin önerilen ResNet mimarisi ve Gauss Karışım Modelleri yaklaşımlarıyla sınıflandırılmasına dayanan derin sahte ses tespiti yöntemi (Cai ve ark., 2019).

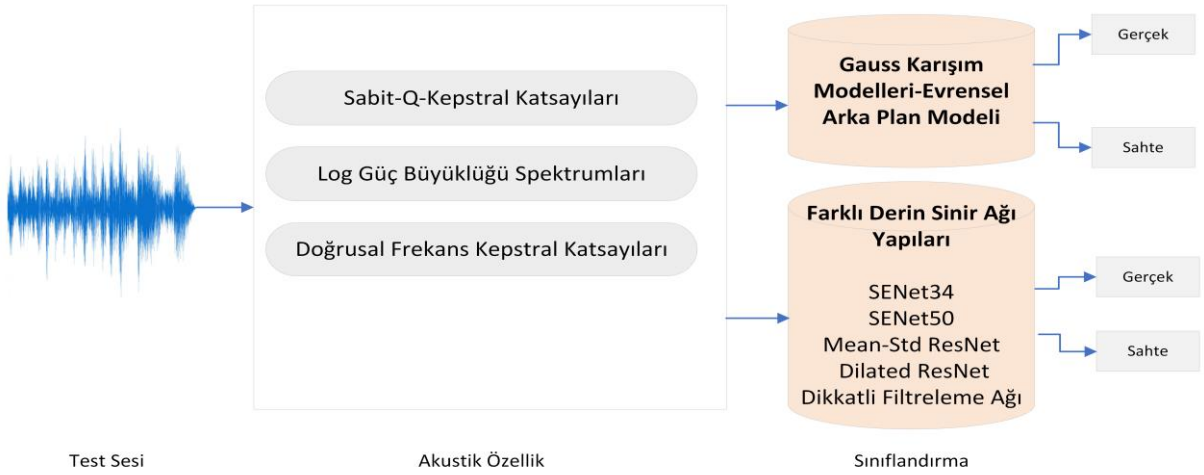
Alluri & Vuppala (2019) tarafından önerilen çalışmada giriş sesinden SFFCC, ZTWCC ve IFCC'den yanı sıra LFCC ve Sabit-Q-Kepstral Katsayıları akustik özellikleri çıkarılmıştır. Elde edilen özelliklerin sınıflandırılması için GMM'nin kullanılması önerilmiştir. Çalışmada ASVSpoof 2019 verisetinden faydalanılmış ve LA geliştirme setinde ZTWCC ve CQCC özelliklerinin birleştirilmesiyle de 0.0 EER, 0.0 t-DCF değerlerinin elde edildiği görülmüştür. Değerlendirme setinde 4.92 EER, 0.1239 t-DCF değerleri elde edilmiştir. PA geliştirme setinde ZTWCC kullanıldığında 10.11 EER, 0.2169 t-DCF başarısı alınmıştır. Değerlendirme setinde ZTWCC kullanıldığında 12.20 EER, 0.2810 t-DCF sonuçları elde edilmiştir. Çalışmada atak bazlı sonuçlar da verilmiştir.



Şekil 21. Beş farklı akustik özelliklerin Gauss Karışım Modellemesi ile sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Alluri & Vuppala, 2019).

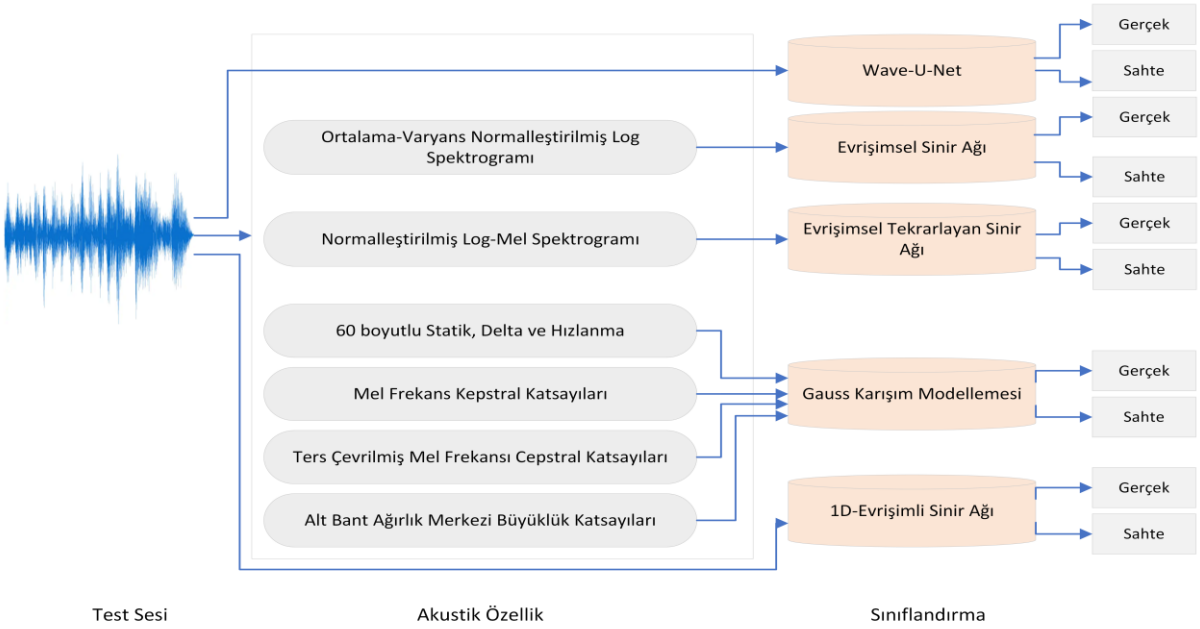
Lai ve ark. (2019) tarafından önerilen çalışmada giriş sesinden Sabit-Q-Kepstral, Log Güç Büyüklüğü Spektrumları (Log Power Magnitude Spectra, Logspec) ve LFCC akustik özellikleri çıkarılmıştır. Sınıflandırma yapmak amacıyla Gauss Karışım Modelleri-Evrensel Arka Plan Modeli (Gaussian Mixture Models-Universal Background Model, GMM), Gauss Karışım Modelleri ve farklı Derin Sinir Ağı yapıları kullanılmıştır. Sıkıştırma-Uyarma Ağlarının çeşidi olan ResNet34 omurgasına sahip SENet34 ve ResNet50 omurgasına sahip SENet50 mimarisi Şekil 22'de görselleştirildiği gibi önerilmiştir. Bunlara ek olarak Mean-Std ResNet, Dilated ResNet ve Dikkatli Filtreleme Ağı (Attentive-

Filtering Network) da çalışmada kullanılmıştır. ASVSpoof 2019 veriseti ile ilgili model eğitilmiş ve test sonuçların elde edilmesi gerçekleştirilmiştir. Log Güç Büyüklüğü Spektrumlarının SENet34, SENet50, Mean-Std ResNet ve Dilated ResNet ile sınıflandırılmasıyla CQCC'nin Mean-Std ResNet ile sınıflandırılması füzyon işlemine alındığında PA geliştirme setinde 0.003 EER ve 0.129 t-DCF başarısı, değerlendirme setinde ise 0.59 EER ve 0.016 t-DCF başarısı mevcuttur. Bununla birlikte LA geliştirme setinde 0.0 EER ve 0.0 t-DCF, değerlendirme setinde 6.70 EER ve 0.155 t-DCF başarısı bulunmaktadır.



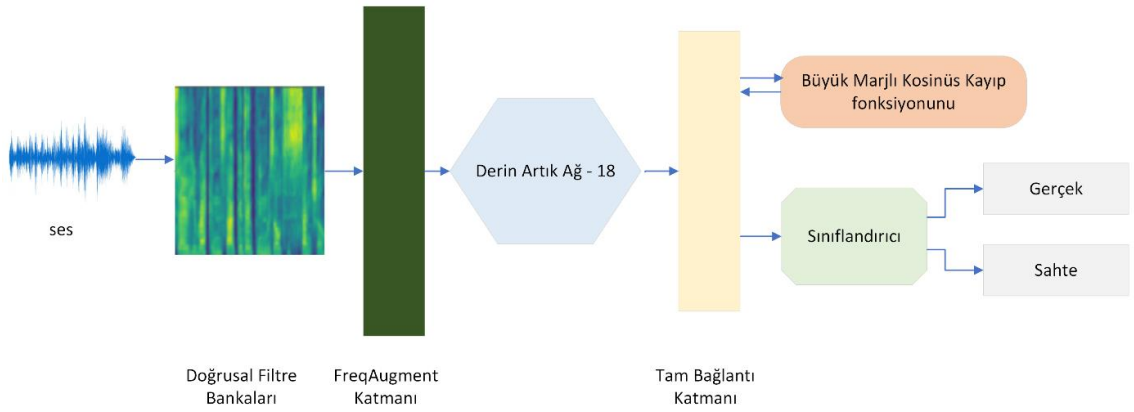
Şekil 22. Üç farklı akustik özelliklerin Gauss Karışım Modelleri-Evrensel Arka Plan Modeli ve Farklı derin sinir ağı yapıları kullanılarak sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Lai ve ark., 2019).

Chettri ve ark. (2019) tarafından giriş sesinin ham ve zaman-frekans (time-frequency) temsillerinin kullanıldığı Şekil 23'de verilen mimari önerilmiştir. Yazarlar eğitim ve doğrulama sırasında aynı tür atakları kullanmanın bilinmeyen atakları tespit etmede başarısızlığa sebep olabileceğini göz önünde bulundurularak eğitim ve geliştirme setini alt sete ayırmıştır. Evrişimli Sinir Ağı mimarisinde her ses örneğinin ilk ve son 4 saniyesinden Ortalama-Varyans Normalleştirilmiş Log Spektrogramı (Mean-Variance Normalized Log Spectrogram) çıkarılarak iki farklı modelin eğitimi yapılır. Normalleştirilmiş Log-Mel Spektrogramı CRNN mimarisi ile sınıflandırılmıştır. 1D-Evrişimli Sinir Ağı mimarisi girdi olarak ham ses alır. Wave-U-Net yapısı ses girişlerinin aynı uzunlukta olmasını sağlamak için tüm kayıtları 12.23 saniye olacak şekilde doldurur.60 boyutlu statik, delta ve hızlanma (Static, Delta and Acceleration, SDA), Mel Frekans Kepstral Katsayıları,IMFCC, SCMC akustik özellikleri Gauss Karışım Modellemesi kullanılarak sınıflandırılmıştır. I-Vektörleri ve Uzun Vadeli Ortalama Spektrum (Long-Term-Average-Spectrum, LTAS) Destek Vektör Makinesi ile sınıflandırılmıştır. Çalışmada 10 farklı eğitim yapılmış ve bu yapılan sınıflandırmalar 3 topluluk modelleri (ensemble models) ile birleştirilmiştir. LA senaryosunda sınıflandırma olarak Ortalama-Varyans Normalleştirilmiş Log Spektrogramı özelliği ile Evrişimli Sinir Ağı mimarisi, CRNN mimarisi, Gauss Karışım Modellemesi kullanılarak oluşturulan topluluk modelinde Geliştirme setinde 0.0 t-DCF, 0.0 EER, değerlendirme setinde ise 0.0755 t-DCF, 2.64 EER başarısı vardır. PA senaryosunda girdi olarak ham sesin kullanıldığı 1D-Evrişimli Sinir Ağı mimarisi hariç tutularak oluşturulan topluluk modelinde Geliştirme setinde 0.0354 t-DCF, %1.33 EER başarısı vardır. Değerlendirme setinde ise en yüksek başarı Ortalama-Varyans Normalleştirilmiş Log Spektrogramı özelliği ile Evrişimli Sinir Ağı mimarisi kullanılarak 0.1465 t-DCF, 5.43 EER ile alınmıştır.



Şekil 23. Altı farklı akustik özelliğin ve ham sesin kullanıldığı Gauss Karışım Modelleri ve Farklı derin sinir ağı yapıları kullanılarak sınıflandırılmasına dayalı derin sahte ses tespiti yöntemi (Chettri ve ark., 2019).

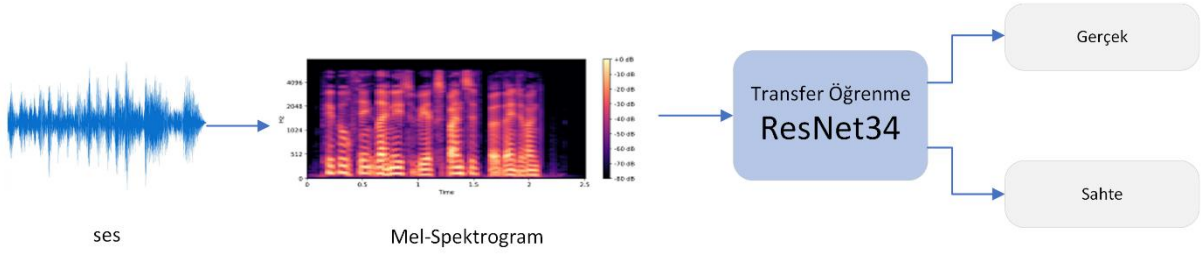
Chen ve ark. (2020) tarafından önerilen çalışmada var olan verisetlerine ek olarak daha gerçekçi senaryoları simüle etmek için ücretsiz olarak erişilebilen mevcut filmler, TV şovları, müzik kayıtlarından elde edilen sesler eğitim setine eklenmiştir. Yeniden oynatma saldırılarını simüle etmek için de ASVspoof 2019 verisetinde sesler VoIP kanalı aracılığıyla mantıksal olarak yeniden oynatılarak eğitim setine eklenmiştir. Girdi olarak alınan ses sinyallerinden 10ms çerçeve kaydırmalı (frame shift), 30ms pencereler kullanılarak 60 boyutlu lineer filtre bankaları (LFB) çıkarılmıştır. Sınıflandırma aşamasında ResNet-18 kullanılması önerilmiştir. Yöntemin genelleştirme yeteneğini arttırma amacıyla, LMCL kullanılmış ve eğitim sırasında bitişik frekans kanallarını rastgele maskeleyen bir katman olan random frekans maskeleme (FreqAugment) kullanılması önerilmiştir. Şekil 24 önerilen yöntemin genel akışını vermektedir. Performans değerlendirmeleri ASVSpoof2019 LA veriseti üzerinde yapılmıştır. ASVSpoof2019 LA değerlendirme setinde 1.26 EER elde edilmiştir.



Şekil 24. Doğrusal filtre bankaları ve Derin artık ağ-18 mimarisi kullanılarak derin sahte ses tespiti yöntemi (Chen ve ark., 2020).

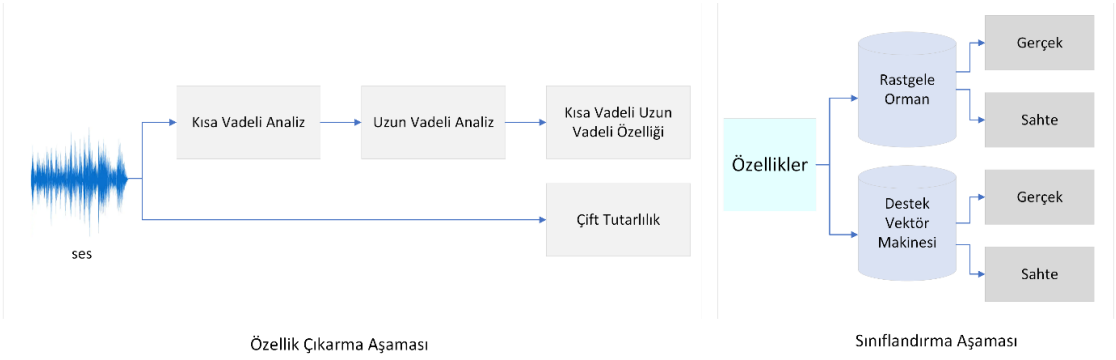
Rahul ve ark. (2020) tarafından önerilen çalışmada giriş sesinden Mel-Spektrogram çıkartılarak Şekil 25'deki gibi iki boyutlu uzaya haritalanmıştır. Transfer öğrenmeye dayalı Derin Artık Ağ (ResNet34) mimarisi ile Mel-Spektrogramın sınıflandırılması gerçekleştirilerek Şekil 22'deki gibi

sahte/gerçek etiketlemesi yapılmıştır. Modelin eğitimi ve testi için ASVSpoof 2019 veriseti kullanılmıştır. LA geliştirme setinde 0.9056 EER sonucu rapor edilirken, t-DCF metrik sonucuna dair raporlanma yapılmamıştır. Değerlendirme setinde ise 5.32 EER ve 0.1514 t-DCF değerlerine ulaşıldığı ifade edilmiştir. PA geliştirme setinde %5.57 EER metrik sonucu verilirken, t-DCF başarısı belirtilmemiş ve değerlendirme setinde ise 5.74 EER ile 0.1351 t-DCF sonuçları sunulmuştur.



Şekil 25. Mel-Spektrogram ve Transfer öğrenmeye dayalı ResNet mimarisi kullanılarak derin sahte ses tespiti yöntemi (Rahul ve ark., 2020).

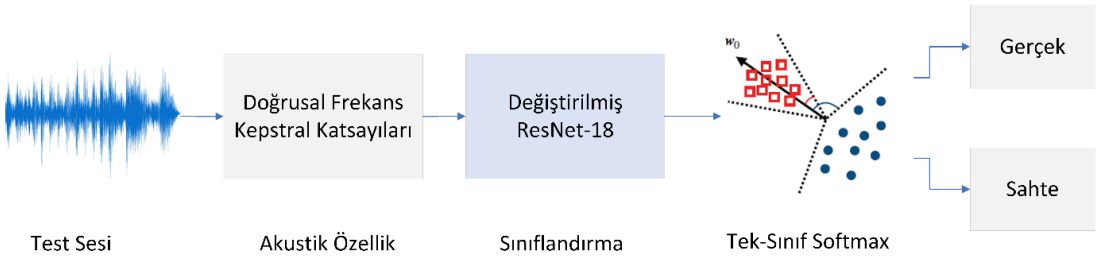
Borrelli ve ark. (2021) tarafından önerilen yöntem giriş sesini tanımlama amacıyla Şekil 26'daki gibi iki farklı öznitelik çıkarma yaklaşımına dayanmaktadır. Bunlardan ilkinin elde edilmesinde öncelikle sesin Kısa Vadeli (Short-Term, ST) analizi yapılmış ve ardından Uzun Vadeli (Long-Term, LT) analizi yapılarak Kısa Vadeli Uzun Vadeli (Short Term Long Term, STLT) özniteliği çıkarılmıştır. İkinci öznitelik vektörlerinin eldesinde ise Çift Tutarlılık (Bicoherence) yaklaşımından faydalanılmıştır. Elde edilen öznitelik vektörlerini kullanacak modelin eğitiminde üç farklı senaryo izlenmiştir. Birinci senaryo ikili (binary) senaryodur, sesin gerçek veya sahte olduğunu tespit etmeye odaklanır. İkinci senaryoda kapalı-set (closed-set) senaryosundan faydalanılmıştır. Bu senaryo sesin sadece sahte/orijinal olarak sınıflandırılmasına değil sahte ise hangi atak ile oluşturulduğunu da tespit etmeye odaklanır. Son olarak açık-set (open-set) senaryosuna dayalı yaklaşımdan faydalanılmıştır. Bu yaklaşım da ikinci senaryoya ek olarak, daha önce görmediği atak türlerini de tespit etmeye odaklanır. Çıkarılan özellikler bu senaryoları gerçekleyecek şekilde Rastgele Orman ve Destek Vektör Makinesi sınıflandırıcı yaklaşımlarıyla eğitilmiştir. Deneyler ASVSpoof2019 LA veriseti üzerinde yapılmıştır. İkili senaryoda en iyi başarı, geliştirme setinde 0.94; değerlendirme setinde 0.74 doğruluk değerini Bicoherence ve STLT füzyonu ile sağlanmıştır. Kapalı Set senaryosu ile yapılan eğitimde geliştirme setinde en iyi başarı (0.93 doğruluk) Bicoherence ve STLT füzyonu ile sağlanmıştır. Değerlendirme setinin %80'i eğitimde ve %20'si testte kullanılmıştır. Açık set senaryosunda daha önce görülmeyen atak türlerinin %49'u orijinal olarak algılanmıştır.



Şekil 26. Kısa Vadeli Uzun Vadeli (Short Term Long Term, STLT) ve Çift Tutarlılık özellikleri kullanan sahtecilik tespit yöntemi (Borrellli ve ark., 2021).

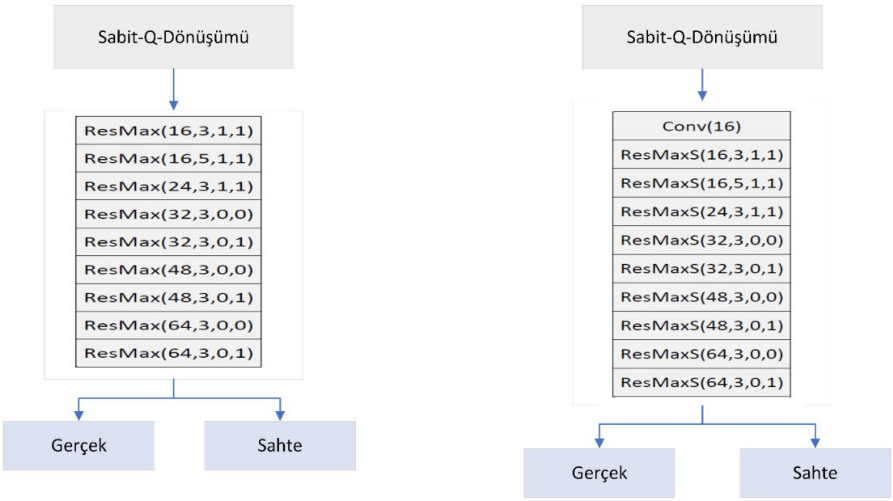
Zhang ve ark. (2021a) tarafından önerilen çalışmada girişten alınan sesten 60 boyutlu LFCC çıkarılmıştır. Bu katsayıların sınıflandırılmasında ise Küresel Ortalama Havuzlama Katmanının, Dikkatli Zamansal Havuzlama ile değiştirildiği ResNet-18 mimarisi ile Şekil 23'teki gibi

gerçekleştirilmiştir. Sahte konuşma verilerinin gerçek verilerden belirli bir marjla uzak tutulduğu, gerçek konuşma özelliklerinin kompakt bir sınıra sahip olduğu bir özellik uzayını öğrenmek için Tek Sınıf Softmax (One Class-Softmax) adlı bir kayıp fonksiyonun kullanılması önerilmiştir. Şekil 27 yöntemin genel akışını vermektedir. Çalışmada ASVSpoof2019 LA veriseti kullanılmıştır. Geliştirme setinde 0.20 EER ve 0.006 min t-DCF, değerlendirme setinde ise 2.19 EER ve 0.059 min-t-DCF başarısına ulaşılmıştır.



Şekil 27. Doğrusal frekans kepstral katsayıları ile değiştirilmiş ResNet-18 mimarisi ile derin sahte ses tespiti yöntemi (Zhang ve ark., 2021a).

Kwak ve ark. (2023) tarafından önerilen çalışmada ilk olarak girdi olarak alınan ses sinyallerinin sürelerinin dokuzar saniye şeklinde olacak şekilde sabitlenmesi gerçeklenmiştir. Sesin süresinin dokuz saniyeden fazla olması durumunda geri kalan bölümünün kesilmesi, az olması durumunda ise aynı ses dosyasının başına ekleme yapılarak genişletilmesi yapılmıştır. Ses sinyallerinin dönüşüm tabanlı yaklaşımlarından olan CQT yöntemi kullanılarak elde edilen özellik vektörleri iki boyutta haritalanmıştır. İki boyutta temsil edilen özellik vektörleri LCNN'den Maksimum özellik haritası konseptini ve ResNet'den bağlantı atlama konseptini birleştirerek Maksimum Özellik Haritasına sahip Artık Ağ (ResMax) blokları ile oluşturulan derin ağ mimarisi ile sınıflandırma gerçekleştirmiştir. Şekil 28'de önerilen bu yöntemde kullanılan mimari detayları verilmiştir. Derin sahte ses tespiti için kullanılan mimariler genellikle sadece sahteciliği tespit etmeye odaklandığı ve hesaplama maliyetini göz ardı ettiği için gerçek zamanlı kullanıma uygunluğu tartışılabilir bir mevzudur. Yazarlar bu durumu göz önünde bulundurup ResMax mimarisini daha hafif hale getirmek için, evrişim katmanları derinlemesine ayrılabilir evrişim katmanlarına (Depthwise Separable Convolution) dönüştürerek kullanmışlardır. Çalışmada ASVSpoof2019 veriseti üzerinde değerlendirmeler yapılmıştır. ResMax mimarisinin kullanımı durumunda PA'da oluşturulan alt veriseti için, değerlendirme setinde 0.30 EER ve geliştirme setinde 0.16 EER elde ederken LA değerlendirme setinde 2.19 EER ve geliştirme setinde ise 0.56 elde etmiştir. ResMaxSep mimarisi ile elde edilen sonuçlarda ise PA setinde ortalama EER 0.36 ve LA setinde ortalama EER 2.55 dir.

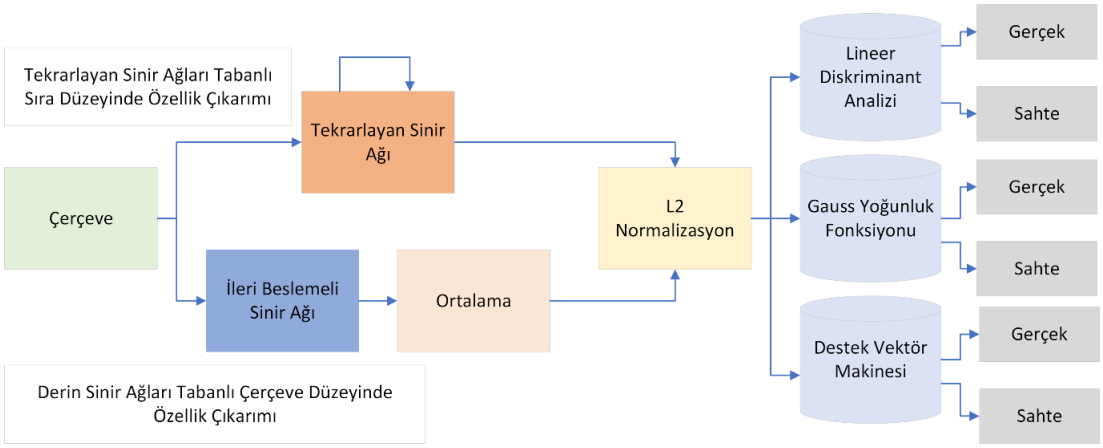


Şekil 28. ResMax ile derin sahte ses tespiti yönteminde kullanılan mimariler (Kwak ve ark., 2023).

4.2. Derin öğrenmeye dayalı özellikler kullanan çalışmaların detaylı analizi

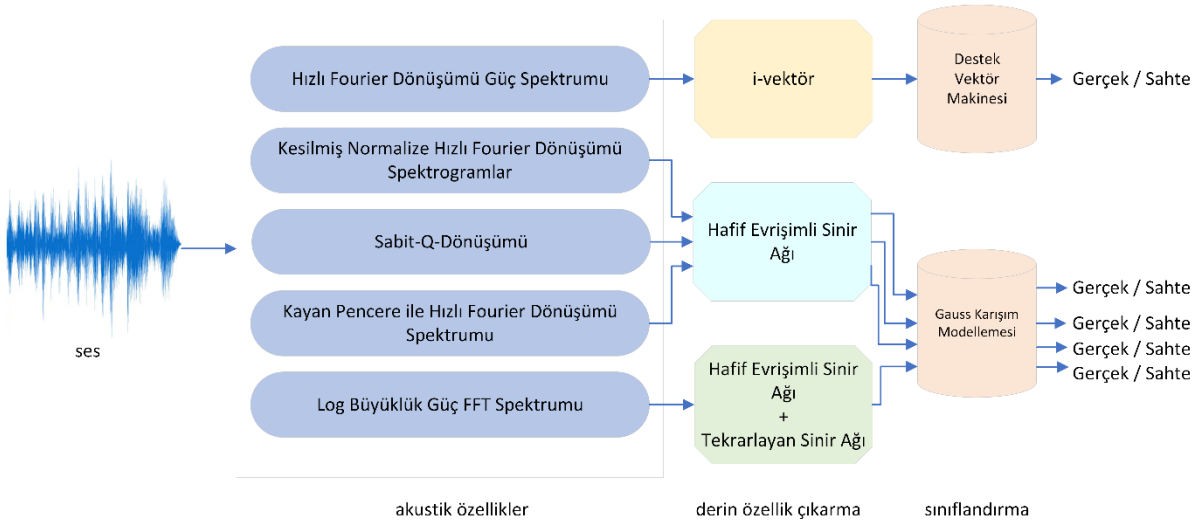
Bu bölümde değerlendirilen çalışmalarda girdi sesinden akustik özellikler çıkarılmadan doğrudan derin öğrenme yaklaşımları ile hem derin özelliklerin çıkarılması hem de sınıflandırılması gerçekleştirilmiştir.

Ele alınan çalışmalardan ilki Qian ve ark. (2016) tarafından önerilmiştir. Bu çalışmada Derin Sinir Ağları Tabanlı Çerçeve Düzeyinde Özellik Çıkarımı ve RNN Tabanlı Sıra Düzeyinde Özellik Çıkarımı yaklaşımları kullanılarak iki farklı derin özellik elde edilmiştir. DNN tabanlı özelliklerin elde edilmesi için, Yığınlı Otomatik Kodlayıcılar, Yanıltma-Ayırt Edici Derin Sinir Ağları ve Çok Görevli Ortak Öğrenilen Derin Sinir Ağları geliştirilmiştir. RNN sistemi ve LSTM–RNN mimarilerinin kullanılması Şekil 29'de özetlendiği gibi önerilmiştir. Önerilen modellerin eğitimi ve testi için ASVSpoof2015 verisetinden faydalanılmıştır. Derin özelliklerin, derin ağlarla sınıflandırılması durumunda performans değerlendirmesi için Accuracy metriği kullanılmıştır. Yanıltma-Ayırt Edici Derin Sinir Ağları kullanıldığında %84.9, Çok Görevli Ortak Öğrenilen Derin Sinir Ağları kullanıldığında %85.4 ve Uzun Kısa Süreli Bellek kullanıldığında %97.02 doğruluk değerine ulaşılmıştır. Çıkarılan derin özellikler ayrıca LDA, GDF ve Destek Vektör Makinesi yaklaşımları kullanılarak ayrı ayrı da gerçekleştirilmiştir. En yüksek başarı DNN ve RNN tabanlı derin özellikler birleştirilmesiyle 1.1 EER olarak raporlanmıştır.



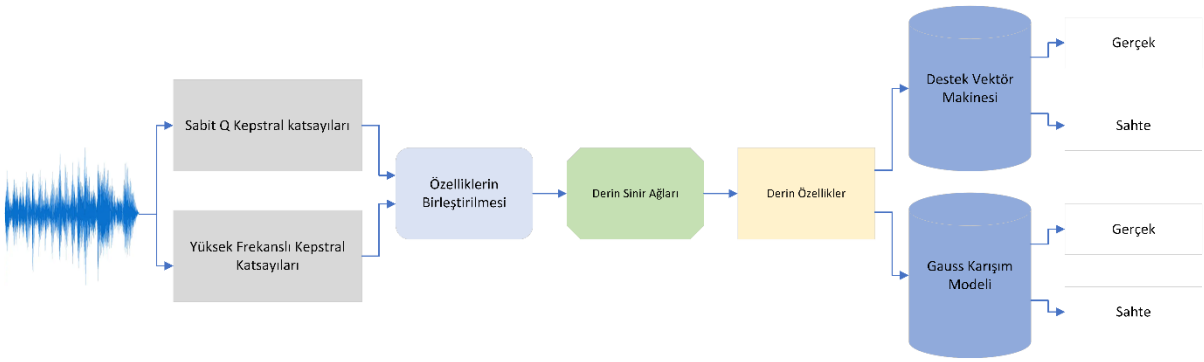
Şekil 29. Derin sahte ses tespiti için Qian ve ark. (2016) tarafından önerilen derin özelliklerin farklı sınıflandırıcılar ile sınıflandırılmasına ilişkin yaklaşım.

Lavrentyeva ve ark. (2017) tarafından önerilen çalışmada giriş sesinden Hızlı Fourier Dönüşümü güç, Kesilmiş Normalize Hızlı Fourier Dönüşümü Spektrogramlar, Sabit-Q-Dönüşümü, zaman ekseni boyunca kayan pencere uygulanarak Hızlı Fourier Dönüşümü spektrumu ve Log Büyüklük Güç FFT Spektrumu yaklaşımları ile ayrı ayrı akustik özellikler elde edilmiştir. Çalışmada akustik özelliklere ek olarak derin özelliklerin de elde edilmesi gerçekleştirilmiştir. Çıkarılan akustik özelliklere Şekil 30'daki gibi farklı yaklaşımların uygulanması ile derin özellikler eklenmiş ve iki farklı sınıflandırıcı kullanılarak performans değerlendirmesi yapılmıştır. Çalışmada ilk olarak Hızlı Fourier Dönüşümü güç spektrumu, i-vektör yapısı kullanılarak Destek Vektör Makinesi ile sınıflandırılmıştır. Kesilmiş Normalize Hızlı Fourier Dönüşümü Spektrogramlar, Sabit Q Dönüşümü ve kayan pencere kullanılarak çıkarılmış Hızlı Fourier Dönüşümü spektrum akustik özelliklerinden Light CNN mimarisi ile özellikler çıkarılmıştır. Bu mimariden üç farklı özellik elde edilerek her birinin sınıflandırılmasında GMM yaklaşımının performansı incelenmiştir. Son olarak Konvolüsyonel Sinir Ağları ve Tekrarlayan Sinir Ağı birleşimi ile Log Büyüklük Güç FFT Spektrumundan özellikler çıkarılmış ve Gauss Karışım Modellemesi ile sınıflandırılmıştır. Çalışmada ASVSpoof 2015 yeniden oynatma veriseti kullanılmıştır. En yüksek başarı, kayan pencere kullanılarak çıkarılan FFT dışında 4 tekli sistem skor seviyesinde füzyon yapıldığında geliştirme setinde 3.95 EER, değerlendirme setinde 6.73 EER ile olmuştur.



Şekil 30. Lavrentyeva ve ark. (2017) tarafından önerilen derin sahte ses tespiti için analiz edilen modeller.

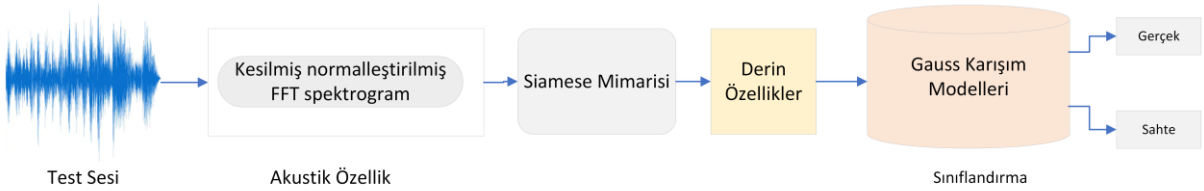
Nagarsheth ve ark. (2017) tarafından önerilen çalışmada giriş sesinden Sabit Q Kepstral katsayıları ve kayıt cihazında kanalı yakalaması için geliştirilen Yüksek Frekanslı Kepstral Katsayıları (High-Frequency Cepstral Coefficients, HFCC) çıkarılmıştır. Şekil 31 önerilen sisteme ait blok diyagramı vermektedir. Birinci ve ikinci türevleriyle birlikte sıfır ortalama ve birim varyans normalleştirilmiş 30 boyutlu CQCC'ler kullanılmıştır. Ortalamalar ve varyanslar eğitim verileri üzerinde hesaplanmıştır. Aynı normalizasyon HFCC özelliklerine de uygulanmıştır. Bu özelliklere ek olarak derin özellikler Derin Sinir Ağları ile elde edilerek Destek Vektör Makinesi ve Gauss Karışım Modeli ile sınıflandırılmıştır. Çalışmada ASVSpoof 2017 yeniden oynatma veriseti kullanılmıştır. Sabit Q Kepstral katsayıları ve Yüksek Frekanslı Kepstral Katsayılarının GMM ile eğitim sonuçları skor seviyesinde (score-level) füzyon işlemine tabi tutulması durumunda, geliştirme setinde 3.2 EER ve değerlendirme setinde 18.1 EER elde edilmiştir. Son olarak Sabit-Q-Kepstral katsayıları ve Yüksek Frekanslı Kepstral Katsayıları özellik düzeyinde füzyon işlemi yapılmış ve Derin Sinir Ağlarıyla derin özellik çıkarılmıştır. Çıkarılan derin özelliklerin Destek Vektör Makinesiyle sınıflandırılması sonucunda geliştirme setinde 7.6 EER ve değerlendirme setinde 11.5 EER alınmıştır.



Şekil 31. Nagarsheth ve ark. (2017) tarafından önerilen yaklaşım ile derin ve akustik özelliklerin kullanıldığı sahte ses tespiti.

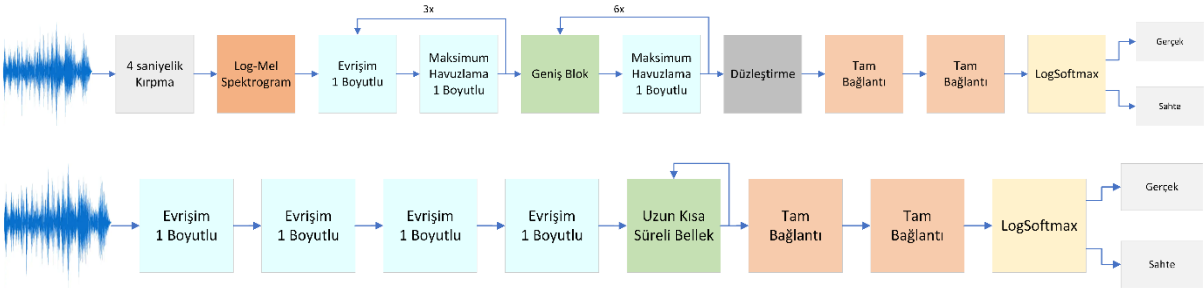
Sriskandaraja ve ark. (2018) tarafından önerilen çalışmada giriş sesinden Kesilmiş normalleştirilmiş FFT spektrogramlar yaklaşımı ile akustik özellikler çıkarılmıştır. Bu özelliklere ek olarak derin özelliklerin öğrenilmesi amacıyla Derin Sinir Ağı mimarilerinin kullanıldığı Siamese yapısının kullanılması önerilmiştir. Sınıflandırma aşamasında Gauss Karışım Modellemesinden faydalanılmıştır. Şekil 32'de genel akışı verilen modelin eğitim ve testi ASVSpoof 2017 veriseti ile gerçekleştirilmiştir. Performans sonuçları yalnızca değerlendirme setinde sunulmuş olup geliştirme

setindeki değerlendirmeye yer verilmemiştir. Değerlendirme setinde 6.40 EER metrik sonucunun raporlandığı görülmektedir.



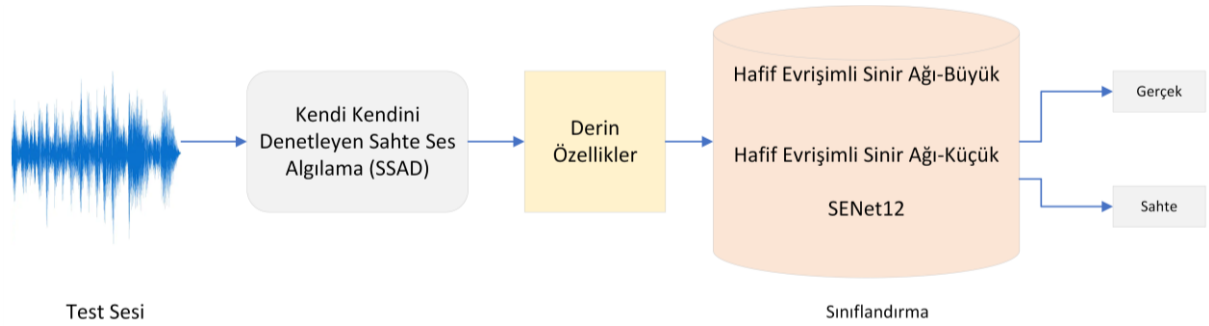
Şekil 32. Sriskandaraja ve ark. (2018) tarafından önerilen yaklaşım ile derin özelliklerin kullanıldığı sahte ses tespiti.

Chintha ve ark. (2020) tarafından önerilen çalışmada CRNNSpoof ve WIRENetSpoof olarak iki farklı mimari Şekil 33'deki gibi kullanılmıştır. CRNNSpoof mimarisinde ağın girişinde spektral özellikler yerine ham ses kullanılarak, ayırt edici özelliklerin mimari tarafından öğrenilmesi sağlanmıştır. WIRENetSpoof mimarisinin kullanıldığı yaklaşımda ise giriş sesi 4 saniyelik seslere kırpılmakta ve log-mel spektrogram dört saniyelik seslerden özellik olarak çıkarılmaktadır. Mimariye log-mel spektrogram giriş olarak verilmiştir. Çalışmada ASVSpoof2019 veriseti üzerinde değerlendirmeler yapılmıştır. CRNNSpoof mimarisi tek yönlü LSTM katmanı kullanıldığında değerlendirme setinde 4.02 EER ve 0.134 t-DCF elde edilmiştir.



Şekil 33. CRNNSpoof ve WIRENetSpoof mimarisi olmak üzere iki farklı mimarinin kullanıldığı derin sahte ses tespiti yaklaşımları (Chintha ve ark., 2020).

Jiang ve ark. (2020) tarafından önerilen çalışmada giriş sesi dalga formunda işleme alınmış ve Kendi Kendini Denetleyen Sahte Ses Algılama (Self-Supervised Spoofing Audio Detection, SSAD) olarak isimlendirdikleri mimari önerilmiştir. Şekil 34 bu mimariye ilişkin genel akışı vermektedir. SSAD'de ses sinyaline ilişkin özellik vektörünün elde edilmesi için 8 evrişimli bloğun kullanıldığı kodlayıcıdan faydalanılır. Özelliklerin zamansal niteliğini yakalamak için TCN veya GRU kullanılmıştır. Çalışmada ayrıca hiç zamansal özelliğin çıkarılmadığı yalnızca SSAD mimarisinin performansının da analizi gerçekleştirilmiştir. Üç regresyon çalışanı (regression workers) ve bir ikili çalışan (binary worker) sahte ses algılamada daha iyi performans elde etmek için tasarlanmıştır. Regresyon çalışanları hedef özellikleri tahmin etmeyi amaçlar. Bu hedef özellikler Log Güç Spektrumu, Log-Frekans Kepstral Katsayılar ve Sabit-Q-Kepstral katsayıları olmak üzere üç adettir. Kodlayıcının kullanılmasındaki amaç bu özellikler ile ağ tahminleri arasında Ortalama Karesel Hatayı (Mean Squared Error, MSE) en aza indirmektir. İkili çalışan sahte ses ile orijinal ses arasında ayrım yapmak amacıyla kullanılmıştır. Sınıflandırma aşamasında LCNN-big, LCNN-small ve SENet12 modelleri kullanılmıştır. Hafif Evrişimli Sinir Ağları Açısal Softmax (Angular Softmax, A-Softmax) kullanırken, SENet12 modeli orijinal softmax kullanır. Çalışmada ASVSpoof 2019 LA veriseti ile önerilen modellerin eğitim ve testi yapılmıştır. Geçitli Tekrarlayan Ünite'nin kullanıldığı SSAD ve SENet12 sınıflandırmasıyla, geliştirme setinde 0.15, değerlendirme setinde 7.24 EER başarısı elde edilmiştir. SSAD ve LCNN-big kullanımında geliştirme setinde 0.78, değerlendirme setinde 5.31 EER başarısı raporlanmıştır.



Şekil 34. Jiang ve ark. (2020) tarafından önerilen yaklaşım ile derin özelliklerin kullanıldığı sahte ses tespiti.

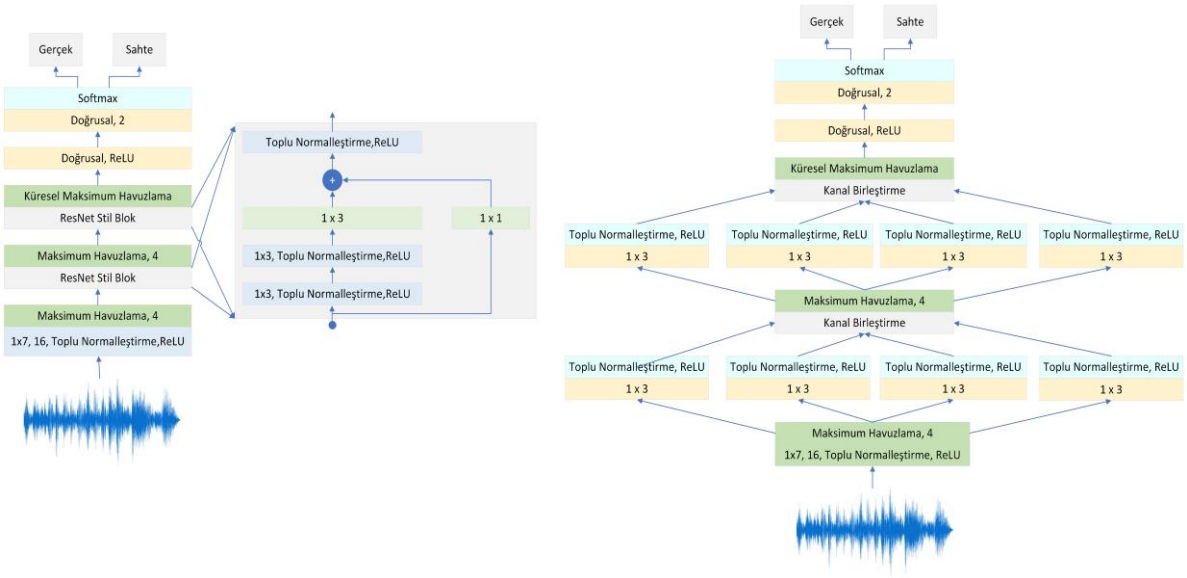
Tak ve ark. (2021) tarafından önerilen çalışma derin sahte ses tespiti alanında RawNet2'yi kullanan ilk çalışmadır. Girişten ham ses alınmış ve RawNet2 üç farklı konfigürasyonla eğitilmiştir. Bu konfigürasyonlar Sabit Mel-Ölçekli Sinc Filtreleri (Fixed Mel-Scaled Sinc Filters), Sabit Ters Mel-ölçekli Sinc Filtreleri (Fixed İnverse Mel-Scaled Sinc Filters) ve Sabit Lineer Ölçekli Sinc Filtreleri (Fixed Linear-Scale Sinc Filters) şeklinde belirlenmiştir. Çizelge 8'de kullanılması önerilen derin ağ mimarisine ilişkin katmanlardan ve bunların çıkışlarının boyutlarından bahsedilmiştir. Çalışmada ASVSpoof 2019 LA veriseti kullanılarak model eğitilmiş ve test edilmiştir. Sabit Ters Mel-ölçekli Sinc Filtreleri kullanılarak geliştirme setinde 0.285 t-DCF, 0.86 EER ve değerlendirme setinde 0.1175 t-DCF, 5.13 EER alınmıştır. Çalışmada daha çok ASVSpoof 2019 A17 sahteciliğini tespit etmeye odaklanılmıştır. Füzyon işlemi yapmadan A17 tespit etme konusunda en iyi başarı 0.1810 t-DCF ile Sabit Lineer Ölçekli Sinc Filtreleri kullanılarak alınmıştır. Çalışmada temel sınıflandırıcı olarak LFCC'nin GMM ile sınıflandırıldığı çalışma kullanılmıştır. Füzyon işlemi Destek Vektör Makinesi tabanlı füzyon yaklaşımı kullanılarak gerçekleştirilmiştir. Üç farklı RawNet2 konfigürasyonu ile temel sınıflandırıcı füzyon yapılmış ve değerlendirme setinde 0.0347 t-DCF, 1.14 EER alınmıştır. Füzyon işlemi sonucunda A17 bazında 0.0808 t-DCF başarısına ulaşılmıştır.

Çizelge 8. Tak ve ark. (2021) tarafından önerilen derin ağ mimarisinde kullanılan katmanlar ve boyutları

Katman	Giriş 64000	Çıkış Şekli
Sabit Sinc Filtreleri	Evrişim(129,1,128) Maksimum Havuzlama (3) Toplu Normalleştirme ve LeakyReLU	(21290, 128)
Kalıntı Blok x 2	Toplu Normalleştirme ve LeakyReLU Evrişim(3,1,128) Toplu Normalleştirme ve LeakyReLU Evrişim(3,1,128) Maksimum Havuzlama (3) Özellik Haritası Ölçeklendirme	(2365, 128)
Kalıntı Blok x 4	Toplu Normalleştirme ve LeakyReLU Evrişim(3,1,512) Toplu Normalleştirme ve LeakyReLU Evrişim(3,1,512) Maksimum Havuzlama (3) Özellik Haritası Ölçeklendirme	(29,512)
Geçitli tekrarlayan birim	Geçitli tekrarlayan birim (1024)	1024
Tam Bağlantı	1024	1024
Çıkış	1024	2

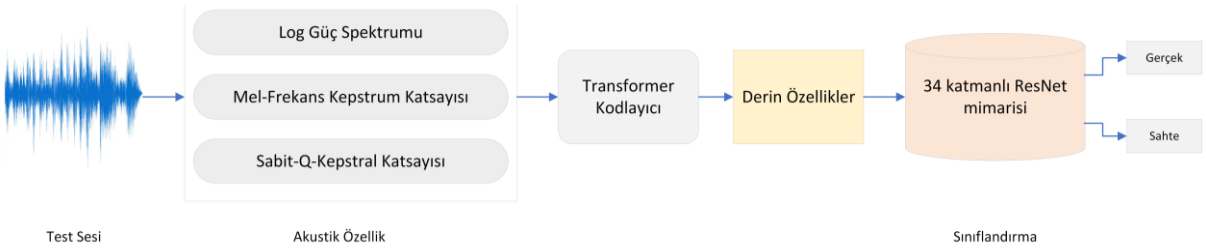
Hua ve ark. (2021) tarafından derin sahte ses tespiti için yapılan çalışmada girişten alınan sesin sahte/orijinal etiketlenmesi amacı 6 saniye ile sınırlandırılarak derin öğrenme modelleri Şekil 35'teki gibi gerçekleştirilmiştir. Önerilen modeller ResNet ve Inception tarzı yapılara sahip olan TSSDNet

mimarilerdir. Res-TSSDNet mimarisinde ResNet tarzı atlama bağlantısı kullanılmıştır. Inc-TSSDNet mimarisinde ise Inception stilinde paralel konvolüsyonlar bulunmaktadır. Çalışmada ASVSpoof 2019 LA ve ASVSpoof2015 verisetleri kullanılmıştır. Res-TSSDNet ile ASVSpoof 2019 LA geliştirme setinde 0.74 EER ve değerlendirme setinde 1.64 EER başarısına ulaşılmıştır. Ayrıca modeller ASVSpoof 2019 eğitim setinde eğitilmiş ve ASVSpoof 2015 geliştirme ve değerlendirme setlerinde test edilmiştir. Res-TSSDNet kullanıldığında ASVSpoof 2015 geliştirme setinde 0.71 EER ve değerlendirme setinde 1.95 EER değerine ulaşılmıştır.



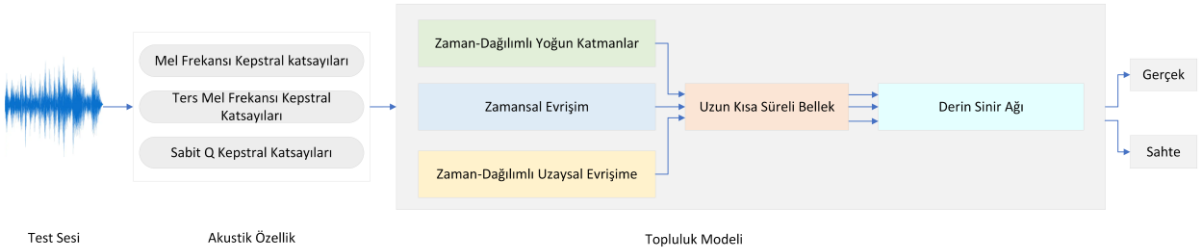
Şekil 35. Hua ve ark. (2021) tarafından derin sahte ses tespiti için önerilen mimariler.

Zhang ve ark. (2021b) tarafından önerilen çalışmada ise ilk olarak eğitim verisi Gauss Gürültü Ekleme (Gaussian Noise Addition, GNA), Sinyal-Gürültü Oranı Gürültü Ekleme (Signal-to-Noise Ratio Noise Addition, SnrNA), zaman değişimi (time shifting), perde değiştirme (pitch shifting) ve zaman uzatması (time stretching) arttırım yöntemleri kullanılarak 5 kat artırılmıştır. Sese ait akustik özelliklerin elde edilmesinde otomatik konuşmacı doğrulama sistemlerinde etkinliği ispat edilmiş LPS, MFCC ve Sabit-Q-Kepstral Katsayısı yöntemleri kullanılmıştır. Bu özelliklere ek olarak derin özelliklerin çıkarılması için Transformer Kodlayıcının kullanılması önerilmiş ve çıkarılan derin özelliklerin sınıflandırılması için 34 katmanlı ResNet mimarisi modifiye edilmiştir. Çalışmada ASVSpoof 2019 LA ve FoR-norm verisetleri kullanılmıştır. ASVSpoof 2019 LA verisetinde Log Güç Spektrumu kullanımında geliştirme setinde 0.11 EER, değerlendirme setinde 6.02 EER başarısına ulaşılmıştır. FoR-norm verisetinde Log Güç Spektrumu kullanımında geliştirme setinde 0.03 EER, değerlendirme setinde 4.38 EER başarısına ulaşılmıştır. Önerilen sistemin çapraz veri kümesi(cross-dataset) başarısı da incelenmiştir. FoR-norm ile eğitilmiş sistem ASVSpoof 2019 LA ile test edildiğinde değerlendirme setinde Log Güç Spektrumu kullanımı ile 19.23 EER değerini üretmiştir. ASVSpoof 2019 LA ile eğitilmiş sistem, FoR-norm ile test edildiğinde değerlendirme setinde Log Güç Spektrumu kullanımı ile 18.15 EER rapor etmiştir.



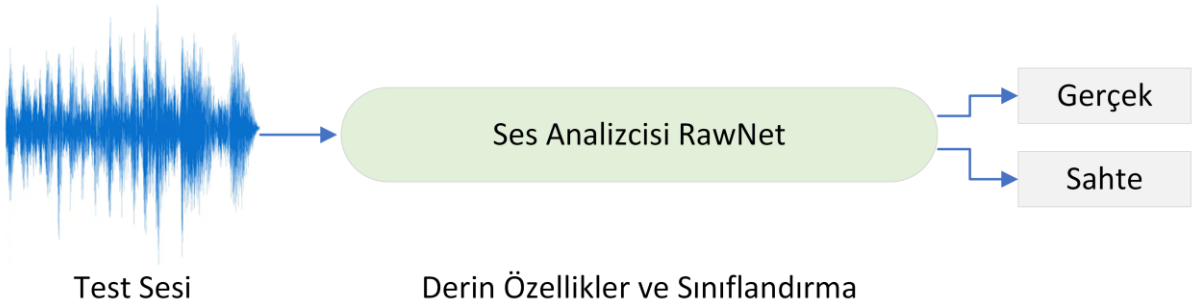
Şekil 36. Zhang ve ark. (2021b) tarafından derin sahte ses tespiti için önerilen üç farklı akustik özellik ve Transformer Kodlayıcının kullanıldığı mimari.

Dua ve ark. (2022) tarafından önerilen çalışmada giriş sesinden MFCC, IMFCC ve CQCC akustik özellikleri çıkarılmıştır. Derin özelliklerin çıkarılması için üç farklı model önerilmiştir. İlk model Zaman-Dağılımlı Yoğun Katmanlar ile LSTM katmanlarının birleşimidir. Diğer iki Derin Sinir Ağı modelleri Zamansal Evrişime ve Uzaysal Evrişme yaklaşımlarına dayanmaktadır. Son aşamada bu üç Derin Sinir Ağı modellerinden oluşan bir topluluk modeli kullanılmıştır. ASVspoof 2015 veri seti ile eğitilen sistemin performansı ASVspoof 2019 değerlendirme seti ile düşerken, ASVspoof 2019 veri seti ile eğitim yapıldığında aynı sistemin performansı artmaktadır.En iyi başarı CQCC akustik özelliği ile topluluk modeli kullanılarak alınmıştır. Eğitim seti olarak ASVspoof 2015 kullanıldığında ASVspoof 2019 değerlendirme setinde 24.8 EER elde edilirken, Eğitim ve değerlendirme seti olarak ASVspoof 2019 kullanıldığında 0.81 EER'ye ulaşılmıştır.



Şekil 37. Dua ve ark. (2022) tarafından derin sahte ses tespiti için önerilen üç farklı akustik özellik ve Topluluk Modelinin kullanıldığı mimari.

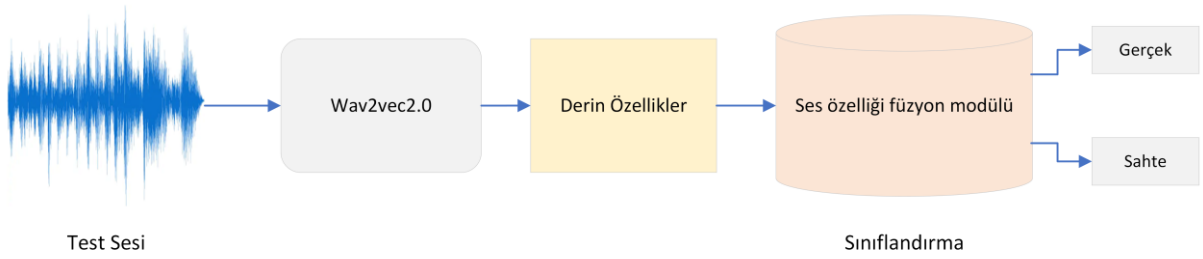
Zarish ve ark. (2022) tarafından önerilen çalışmada giriş sesinden derin özellik çıkarılması için Ses Analizcisi RawNet (Audio Examiner RawNet, AEXANet) mimarisi önerilmiştir. Önerilen mimaride MFM, LeakyRelu, SiLU, ve Relu gibi farklı aktivasyon fonksiyonları ile deneysel sonuçlarda bulunulmuştur. Çalışmada ASVspoof 2019 ve ASVSpoof 2015 verisetleri kullanılmıştır. Deneysel sonuçlar MFM aktivasyon fonksiyonunun en başarılı olduğunu göstermiştir. ASVSpoof 2019 veriseti LA senaryosunda 4.93 EER, 0.17 t-DCF alınırken PA senaryosunda ise 5.29 EER, 0.2061 t-DCF başarısına ulaşılmıştır. ASVSpoof 2019 TTS ataklarında 0.61 EER ve 0.08 t-DCF, VC ataklarında ise 12.10 EER ve 0.40 t-DCF raporlanmıştır. Çalışmada Cross-Dataset ile de sonuç verilmiştir. Eğitim seti olarak ASVSpoof 2015 eğitim ve geliştirme, test seti olarak ASVSpoof2019 LA değerlendirme seti kullanıldığında 37.60 EER ve 0.91 t-DCF, eğitim seti olarak ASVSpoof 2019 LA eğitim ve geliştirme, test seti olarak ASVSpoof2015 değerlendirme seti kullanıldığında 19.71 EER ve 0.58 t-DCF raporlanmıştır.



Şekil 38. Zarish ve ark. (2022) tarafından derin sahte ses tespiti için önerilen Ses Analizcisi RawNet'in kullanıldığı mimari.

Zhang ve ark. (2023) tarafından önerilen çalışma ham dalga formlarından özellikler çıkarmak için önceden eğitilmiş bir wav2vec2 modeli ve arka uç sınıflandırma için bir ses özelliği füzyon modülü kullanılmaktadır. Ses özelliği füzyon modülü, ön uçtan çıkarılan ses özelliklerini kullanarak, zaman dilimlerinden ve özellik boyutlarından gelen bilgileri birleştirir. Modelin performansını geliştirmek için veri artırma teknikleri de kullanılmıştır. Model ASVspoof 2019 LA eğitim ve geliştirme setleri üzerinde eğitilmiş ve ASVspoof 2021 mantıksal erişim (Logical Access, LA) ve Derin Sahte (Deep Fake,DF)

değerlendirme verisetleri üzerinde test edilmiştir. ASVspoof 2021 LA'da 1.18 EER ve ASVspoof 2021 DF'de 2,62 EER başarısına ulaşılmıştır.

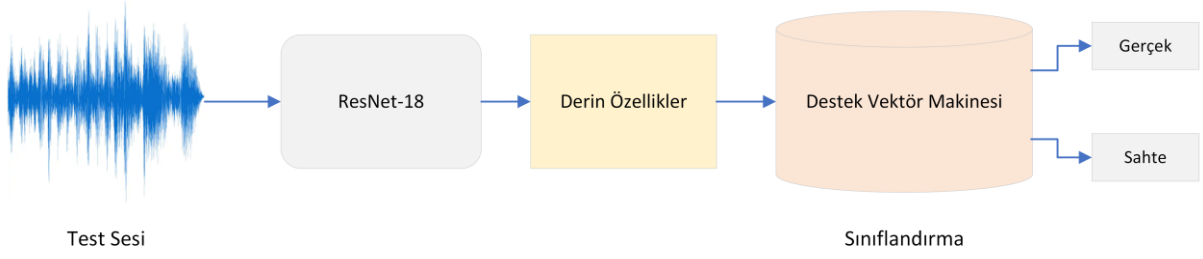


Şekil 39. Zhang ve ark. (2023) tarafından derin sahte ses tespiti için önerilen Wav2vec2.0 ve Ses özelliği füzyon modülünün kullanıldığı mimari.

Mewada ve ark. (2023) tarafından önerilen çalışmada giriş sesinden Gauss Filtresi Tabanlı MFCC, Gamaton Filtresi bazlı GTCC, Spektral Entropi (Spectral Entropy), Spektral Düzlük (Spectral Flatness) ve Pitch Bilgisi (Pitch Information) akustik özellikler olarak çıkarılmıştır. Akustik özelliklerin kombinasyonu Bayes algoritması tarafından optimize edilmiş BiLSTM ile derin özellikler çıkarılmış ve sınıflandırılmıştır. Şekil 40 yöntemin genel akışını vermektedir. Çalışmada ASVSpoof 2017 veriseti kullanılmıştır. Geliştirme setinde 1.02 EER, değerlendirme setinde ise 6.58 EER başarısına ulaşılmıştır.

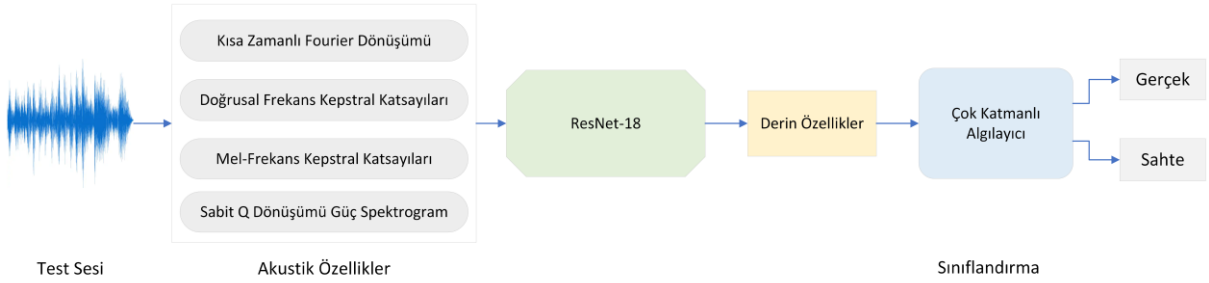
Şekil 40. Mewada ve ark. (2023) tarafından derin sahte ses tespiti için önerilen beş farklı akustik özellik ve Bayes tarafından optimize edilmiş BiLSTM'nin kullanıldığı mimari.

Tan ve ark. (2023) tarafından önerilen çalışmada giriş sesinden 60 boyutlu LFCC akustik özelliği çıkarılmıştır. Derin özelliğin çıkarılması amacıyla Küresel Ortalama Havuzlama Katmanının yerine Dikkatli Zamansal Havuzlamanın uygulandığı ResNet-18 mimarisi önerilmiştir. Sınıflandırma SVM kullanılarak yapılmıştır. SVM'nin Radyal Tabanlı Fonksiyon, Doğrusal, Polinom ve Sigmoid olmak üzere 4 farklı çekirdek ile etkinliği araştırılmıştır. Çalışmada ASVSpoof 2019 LA veriseti kullanılmıştır. En yüksek başarı Doğrusal çekirdekli SVM ile değerlendirme setinde 0.57 EER ile alınmış olmakla birlikte geliştirme setinde skor sonuçları raporlanmamıştır.



Şekil 41. Tan ve ark. (2023) tarafından derin sahte ses tespiti için önerilen ResNet-18 ve Destek Vektör Makinesinin kullanıldığı mimari.

Abdzadeh & Veisi (2023) tarafından önerilen çalışmada giriş sesinin 4 saniyelik eşit uzunlukta olması için kırpma veya doldurma işlemleri yapılmıştır. Akustik özellikler olarak STFT, LFCC, MFCC, CQT Power Spectrogram çıkarılmıştır. Akustik özelliklerin her biri, Kendine-Dikkat (Self-Attention) katmanına sahip ResNet-18 mimarisine verilmiş ve gömme vektörü çıkarılmıştır. Çıkarılan derin özellikler 3 katmanlı Çok Katmanlı Algılayıcı MLP ve Tek Sınıf-Softmax (One Class-Softmax, OC-Softmax) ile Şekil 41'de özetlendiği gibi sınıflandırılmıştır. Çalışmada ASVspoof 2019 LA veriseti kullanılmıştır. Akustik özellikler arasında en yüksek başarı Sabit-Q-Dönüşümü Güç Spektrogramı kullanılarak değerlendirme setinde 2.33 EER ve 0.120 t-DCF olarak raporlanmıştır.



Şekil 42. Abdzadeh & Veisi (2023) tarafından derin sahte ses tespiti için önerilen dört farklı akustik özellik, ResNet-18 ve Çok Katmanlı Algılayıcı'nın kullanıldığı mimari.

4.3. Yöntemlerin karşılaştırmalı performans değerlendirmesi

Yukarıdaki iki alt bölümde detayları sunulan yöntemlerin karşılaştırmalı performans değerlendirmelerine bu bölümde yer verilecektir. Önceki bölümlerde de bahsedildiği gibi yapılan çalışmalarda önerilen sistemlerin eğitilmesi ve test edilmesinde farklı verisetlerinden faydalanılmıştır. Bu sebeple yapılan değerlendirmeler veriseti bazında gerçekleştirilmiştir. Verisetlerindeki geliştirme ve değerlendirme setlerinde ayrı ayrı raporlanan sonuçlar tablolar halinde sunularak değerlendirmeleri yapılacaktır. Ele alınan çalışmalarda kullanılan EER ve t-DCF metrik değerleri performans karşılaştırması için irdelenmiştir ancak her çalışmada her iki metriğinde kullanılmadığı da görülmüştür.

Yapılan ilk değerlendirmede ASV2015 verisetinde eğitilen ve test edilen çalışmalara yer verilmiştir. Bu çalışmalara ilişkin metrik değerlendirme sonuçları Çizelge 9'da sunulmuştur. Bu çalışmalarda değerlendirme metriği olarak yalnızca EER metriğinin kullanıldığı için tabloda yöntemlerin geliştirme ve değerlendirme setinde EER sonuçları sunulmuştur. Bu çalışmalardan Wang ve ark. (2015) tarafından yapılan çalışmada geliştirme setinde en yüksek performansın önerilen üç farklı özellik çıkarma yaklaşımı ile sistemin eğilmesinin füzyon durumunda elde edildiği görülmektedir. Ancak yazarların değerlendirme setinde bazı ikili ve üçlü füzyon sonucunu rapor etmemesi çalışmanın eksikliği olarak göze çarpmaktadır. Xiao ve ark. (2015)'nın yaptığı değerlendirmeye göre tekil yaklaşımlara göre tüm özelliklerin ayrı ayrı kullanılıp füzyon edilmesiyle eğitilen sistem ile geliştirme setinde en düşük EER değeri elde edildiği söylenebilir. Yazarlar bu değerlendirmeden sonra değerlendirme setinde yalnızca füzyon EER sonucunu 2.62 olarak rapor etmişler. Geleneksel yöntemler ile özellik elde eden çalışmalar arasından değerlendirme setinde performansı en düşük olan yöntem Wang ve ark. (2015) tarafından yapılan çalışmada Değiştirilmiş Grup Gecikmesi Kepstral Katsayılarının kullanılması durumunda olduğu görülmektedir. Buna karşın değerlendirme setinde en iyi olan modelin Yang ve ark. (2018) tarafından önerilen 0.038 EER (%) değerinin raporlandığı Oktav-Bandı Temel Bilgileri, Tam Bant Temel Bilgiler, Kısa-Vadeli Spektral İstatistik Bilgileri varyansı özelliklerinin kullanılması durumundaki model olduğu söylenebilir. Derin özellikler kullanan ve ASV2015 geliştirme setinde sonuç rapor eden çalışmalar arasında en iyi performans Hua ve ark. (2021) tarafından gerçekleştirilen, ham sesin Res-TSSDNet mimarisi ile doğrudan eğitilmesi durumunda, %0.71 EER değerinin raporlandığı çalışma ile elde edilmiştir. Buna karşın bu yaklaşım değerlendirme setinde %1.95 EER değerine yükselmiştir. Değerlendirme setinde en düşük EER'nin elde edildiği ve dolayısıyla ASV2015 verisetinde en başarılı yöntem olarak yorumlanacak yöntemin Dua ve ark. (2022) tarafından Sabit-Q-Kepstral Katsayılarının Zaman-Dağılımlı Uzaysal Evrişim ve Uzun Kısa Süreli Bellek kullanımı durumundaki yöntem olduğu söylenebilir.

Çizelge 9. ASV2015 veriseti kullanan çalışmaların performans sonuçları

Geleneksel yöntemler ile özellik elde eden çalışmalar				
Referans çalışmalar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**	**Değerlendirme seti**
Değerlendirme metriği			**EER**	**EER**
Wang ve ark. (2015)	Mel-Frekans Kepstral (A)	Gauss Karışım Modeli	1.74	-
	Değiştirilmiş Grup Gecikmesi Kepstral Katsayıları (B)		0.83	3.958
	Fourier spektrumundan alınan Bağıl Faz (C)		0.013	3.925
	A, B Füzyon		0.256	-
	A, C Füzyon		0.004	-
	B, C Füzyon		0.004	3.726
	A, B, C Füzyon		0.002	-
Xiao ve ark. (2015)	Log Magnitüd Spektrumu	Çok Katmanlı Algılayıcı	0.543	-
	Artık Log Magnitüd Spektrumu		0.486	-
	Grup Gecikmesi		0.114	-
	Değiştirilmiş Grup Gecikmesi		1.572	-
	Anlık Frekans Türevi		0.428	-
	Temel Bant Faz Farkı		3.431	-
	Pitch Senkron Fazı		1.345	-
	Füzyon		**0.001**	2.62
Patel & Patil (2015)	Koklear Filtre Kepstral Katsayıları (A)	Gauss Karışım Modeli-Log Olasılık	1.37	-
	Koklear Filtre Kepstral Katsayıları Anlık Frekans (B)		1.4	-
	Mel-Frekans Kepstral Katsayıları (C)		1.6	-
	A, C Füzyon		0.89	-
	B, C Füzyon		0.83	1.211
Paul ve ark. (2017)	Mel-Frekans Kepstral Katsayıları	Gauss Karışım Modeli-Maksimum Olabilirlik	-	2.17
	Sabit-Q-Kepstral Katsayıları		-	0.44
Yu ve ark. (2017)	Derin Sinir Ağı Filtre Bankası Kepstral Katsayıları	Gauss Karışım Modeli-Maksimum Olasılık	0.09	0.56
Yang ve ark. (2018)	Oktav-Bandı Temel Bilgileri, Tam Bant Temel Bilgiler	Derin Sinir Ağı	-	0.042
	Oktav-Bandı Temel Bilgileri, Tam Bant Temel Bilgiler, Kısa-Vadeli Spektral İstatistik Bilgileri ortalama ve varyansı		-	0.045
	Oktav-Bandı Temel Bilgileri, Tam Bant Temel Bilgiler, Kısa-Vadeli Spektral İstatistik Bilgileri varyansı		-	**0.038**

Çizelge 9. ASV2015 veriseti kullanan çalışmaların performans sonuçları (devam)

Derin özellikler ile özellik elde eden çalışmalar				
Referans çalışmalar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**	**Değerlendirme seti**
Değerlendirme metriği			**EER**	**EER**
Qian ve ark. (2016)	Derin Sinir Ağları Tabanlı Çerçeve Düzeyinde Özellik Çıkarımı (A)	Lineer Diskriminant Analizi	-	2.6
		Gauss Yoğunluk Fonksiyonu	-	22.7
		İkili Destek Vektör Makinesi	-	3.9
		Tek-Sınıf Destek Vektör Makinesi	-	5.9
	Recurrent Neural Network (RNN) Tabanlı Sıra Düzeyinde Özellik Çıkarımı (B)	Lineer Diskriminant Analizi	-	1.6
		Gauss Yoğunluk Fonksiyonu	-	2.2
		İkili Destek Vektör Makinesi	-	1.4
		Tek-Sınıf Destek Vektör Makinesi	-	1.5
	A, B Füzyon		-	1.1
Lavrentyeva ve ark. (2017)	Hızlı Fourier Dönüşümü güç spektrumu ve i-vektörden (A)	Destek Vektör Makinesi	9.8	12.54
	Kesilmiş Normalize Hızlı Fourier Dönüşümü Spektrogramları ve Hafif CNN (B)	Gauss Karışım Modellemesi	4.53	7.37
	Sabit Q Dönüşümü ve Hafif CNN (C)		4.8	16.64
	Hızlı Fourier Dönüşümü ve Hafif CNN		5.25	11.81
	Log Büyüklük Güç FFT Spektrumu ve Konvolüsyonel Sinir Ağları, Tekrarlayan Sinir Ağı (D)		7.51	10.69
	A, B, C, D Füzyon		3.95	6.73
Hua ve ark. (2021)	Ham ses	Res-TSSDNet	**0.71**	1.95
		Inc-TSSDNet	1.31	1.96

Çizelge 9. ASV2015 veriseti kullanan çalışmaların performans sonuçları (devam)

Derin özellikler ile özellik elde eden çalışmalar				
Referans çalışmalar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**	**Değerlendirme seti**
Değerlendirme metriği			**EER**	**EER**
Dua ve ark. (2022)	Mel Frekansı Kepstral Katsayıları ve Zaman-Dağılımlı Yoğun Katmanlar, Uzun Kısa Süreli Bellek,	Derin Sinir Ağı	-	9.7
	Ters Mel Frekansı Kepstral Katsayıları ve Zaman-Dağılımlı Yoğun Katmanlar, Uzun Kısa Süreli Bellek		-	9.1
	Sabit Q Kepstral Katsayıları ve Zaman-Dağılımlı Yoğun Katmanlar, Uzun Kısa Süreli Bellek		-	7.1
	Mel Frekansı Kepstral Katsayıları ve Zaman-Dağılımlı Yoğun Katmanlar, Uzun Kısa Süreli Bellek		-	8.4
	Ters Mel Frekansı Kepstral Katsayıları ve Zaman-Dağılımlı Yoğun Katmanlar, Uzun Kısa Süreli Bellek		-	8.4
	Sabit Q Kepstral Katsayıları ve Zaman-Dağılımlı Yoğun Katmanlar, Uzun Kısa Süreli Bellek		-	3.9
	Mel Frekansı Kepstral Katsayıları ve Zaman-Dağılımlı Uzaysal Evrişim, Uzun Kısa Süreli Bellek		-	1.9
	Ters Mel Frekans Kepstral Katsayıları ve Zaman-Dağılımlı Uzaysal Evrişim, Uzun Kısa Süreli Bellek		-	1.3
	Sabit-Q-Kepstral Katsayıları ve Zaman-Dağılımlı Uzaysal Evrişim, Uzun Kısa Süreli Bellek		-	**0.7**
	Mel Frekansı Kepstral Katsayıları ve Topluluk Modeli		-	1.7
	Ters Mel Frekans Kepstral Katsayıları ve Topluluk Modeli		-	0.9

ASV2015 verisetinde ataklı sesler kullanarak test sonuçları rapor eden çalışmaların S1-S10 atakları için atak bazlı sonuçları Çizelge 10'da toplanmıştır. Atak bazlı en düşük EER sonuçları koyu font ile belirtilmiştir. Tabloda Wang ve ark. (2015) tarafından önerilen çalışmada Değiştirilmiş Grup

Gecikmesi Kepstral Katsayıları özelliği tabloda A, Fourier spektrumundan alınan Bağıl Faz özelliği B olarak temsil edilmiştir. Kullanılan ilk özellik ile sadece bilinen ve bilinmeyen ataklarda ortalama sonuçları rapor edilmiştir. B özeliği ve bu özelliklerinin füzyon durumunda her bir atak durumu için ayrı ayrı sonuçlar sunulmuştur. Bilinen ataklar için ortalama sonuçların B özelliği kullanıldığı durumda tüm çalışmalar arasında en iyi olduğu görülmektedir. Tabloda Paul ve ark. (2017) tarafından önerilen çalışma sonuçları için Mel-Frekans Kepstral Katsayıları özelliği A ile Sabit-Q-Kepstral Katsayıları özelliği için B ile gösterilmiştir. Sriskandaraja ve ark. (2017) tarafından önerilen çalışmada kullanılan Saçılma Kepstral Katsayıları özelliği kullanılarak elde edilen sonuçlar tabloda yer almaktadır. Bu çalışma ile S3 ve S4 ile EER değeri 0 olarak hesaplanmıştır. Patel & Patil (2015) tarafından Koklear Filtre Kepstral Katsayıları Anlık Frekans ve Mel-Frekans Kepstral Katsayıları Füzyonu özelliği (A) kullanılarak elde edilen sonuçlar yine tabloda yer almaktadır. Yang ve ark. (2018) tarafından önerilen çalışmada kullanılan özelliklerden Sabit-Q İstatistikleri-Artı-Asıl Bilgi Katsayısı-Delta özelliği A; Sabit-Q İstatistikleri-Artı-Asıl Bilgi Katsayısı-Delta ve Hızlanma özelliği B; Sabit-Q İstatistikleri-Artı-Asıl Bilgi Katsayısı ve Hızlanma özelliği B olarak temsil edilmiştir. Bu çalışmalar için atak bazlı en düşük EER sonuçları 0 olarak görülürken, ortalama sonuçlar arasında bilinen ataklar için ortalama EER sonuçlarının en düşüğü 0.005 iken bilinmeyen atak durumunda ortalama EER sonucu 0.33 olarak görülmüştür.

Çizelge 10. ASV2015 verisetinde ataklı sesler üzerinde değerlendirme yapan çalışmaların atak bazlı EER (%) sonuçları

		Atak türü											
		Bilinen ataklar						**Bilinmeyen ataklar**					
Ref	**Özellik**	**S1**	**S2**	**S3**	**S4**	**S5**	**Ort**	**S6**	**S7**	**S8**	**S9**	**S10**	**Ort**
Wang ve ark. (2015)	A	-	-	-	-	-	1.15	-	-	-	-	-	6.76
	B	**0**	0.03	**0**	**0**	0.015	**0.005**	0.28	0.01	1.18	**0**	37.7	7.84
	A, B Füzyon	**0**	0.09	**0**	**0**	0.015	0.05	0.09	0.01	0.08	**0**	37.1	7.45
Paul ve ark. (2017)	A	0.009	1.46	**0**	**0**	0.36	-	0.3	0.02	0.03	0.02	19.5	-
	B	0.02	0.31	0.01	0.03	0.27	-	0.25	0.12	2.29	0.15	0.94	-
Sriskandaraja ve ark. (2017)	A	0.01	0.12	**0**	**0**	0.02	0.02	0.01	0.01	0.03	0.01	3.94	**0.33**
Xiao ve ark. (2015)	Füzyon	**0**	**0**	**0**	**0**	0.01	-	0.01	**0**	**0**	**0**	26.1	-
Patel ve Patil (2015)	A	0.101	0.863	**0**	**0**	1.07	0.4	0.84	0.24	0.14	0.34	8.49	2.01
Yang ve ark. (2018)	A	**0**	0.004	**0**	**0**	0.024	-	0.02	0.004	0.01	**0**	0.86	-
	B	**0**	0	**0**	**0**	0.009	-	0.01	**0**	0.01	**0**	0.82	-
	C	**0**	0	**0**	**0**	**0.004**	-	0	**0**	0.01	**0**	0.36	-

Yapılan ikinci değerlendirmede ASV2017 verisetinde eğitilen ve test edilen çalışmalara yer verilmiştir. Bu çalışmalara ilişkin metrik değerlendirme sonuçları Çizelge 11'de sunulmuştur. Bu çalışmalarda da değerlendirme metriği olarak yalnızca EER metriğinin kullanıldığı görülmüş ve tabloda

bu sonuçlar sunulmuştur. Bu çalışmalardan geliştirme setinde en düşük EER sonucu Chen ve ark. (2017) tarafından üçlü füzyon yaklaşımında 2.58 olarak elde edildiği görülmüştür. Ancak geliştirme setinde bu metrik sonucu 13.3 olarak alınmıştır. Değerlendirme setinde ise en iyi sonucun Sriskandaraja ve ark. (2018) arkadaşlarının Kesilmiş normalleştirilmiş FFT spektrogramlar ve Siamese yapısı ile elde edilen özelliklerin Gauss Karışım Modeli ile sınıflandırılması şeklinde önerildikleri yaklaşım ile elde edildiği görülmüştür.

Çizelge 11. ASVSpoof 2017 veriseti kullanan çalışmaların performans sonuçları

Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**	**Değerlendirme seti**
Değerlendirme metrikleri			**EER**	**EER**
Witkowski ve ark. (2017)	Sabit-Q-Kepstral Katsayıları	Gauss Karışım Modellemesi	5.13	17.31
	Kepstrum		3.38	22.24
	Mel-Frekans Kepstral Katsayıları		3.16	-
	Ters Mel-Frekans Kepstral Katsayıları		16.18	-
	Doğrusal Tahmin Kepstral Katsayıları Artık		6.22	27.61
Chen ve ark. (2017)	Sabit-Q-Kepstral Katsayıları (A)	Gauss Karışım Modelleri	10.83	28.46
	Sabit-Q-Kepstral Katsayıları (B)	Derin Sinir Ağları	5.18	19.41
	Sabit-Q-Kepstral Katsayıları (C)	Artık Sinir Ağı (ResNet)	5.05	18.79
	Mel-Frekans Kepstral Katsayıları (D)		10.95	16.26
	B, C Füzyon		5.05	18.79
	C, D Füzyon		3.45	14.88
	A, C, D Füzyon		**2.58**	13.3
	A, B, D Füzyon		2.76	13.44
Font ve ark. (2017)	Sabit-Q-Kepstral Katsayıları	Gauss Karışım Modeli	8.2	17.41
	Mel Frekans Kepstral Katsayıları		7.76	27.12
	Doğrusal Frekans Kepstral Katsayıları		5.61	26.27
	Ters Mel-Frekans Kepstral Katsayıları		3.85	30.91
	Dikdörtgen Filtre Kepstral Katsayıları		6.91	11.9
	Doğrusal Tahmin Kepstral Katsayıları		5.94	25.2
	Alt Bant Spektral Akı Katsayıları		24.5	24.83
	Alt Bant Spektral Merkezi Frekans Katsayıları		9.32	11.49
	Alt Bant Spektral Merkezi Büyüklük Katsayıları		12.8	22.38

Çizelge 11. ASVSpoof 2017 veriseti kullanan çalışmaların performans sonuçları (devam)

Referanslar	Kullanılan özellikler	Sınıflama yöntemi	Geliştirme seti	Değerlendirme seti
Değerlendirme metrikleri			**EER**	**EER**
Yang ve ark. (2018)	Oktav-Bandı Temel Bilgileri, Tam Bant Temel Bilgiler, Kısa-Vadeli Spektral İstatistik Bilgileri ortalama	Derin Sinir Ağı	-	11.09
	Oktav-Bandı Temel Bilgileri, Tam Bant Temel Bilgiler, Kısa-Vadeli Spektral İstatistik Bilgileri ortalama ve varyansı		-	11.19
				11.4
	Oktav-Bandı Temel Bilgileri, Sabit-Q-Kepstral Katsayıları, Kısa-Vadeli Spektral İstatistik Bilgileri ortalama ve varyansı		-	
Suthokumar ve ark. (2018)	MCF Kosinüs Katsayıları (A)	Gauss Karışım Modeli	-	12.92
	MSE Kepstral Katsayısı (B)		-	11.97
	Kısa Vadeli Kepstral Katsayılar özelliği (C)		-	11.27
	A, B Füzyon		-	7.2
	A, C Füzyon		-	9.21
	B, C Füzyon		-	8.25
	A, B, C Füzyon		-	6.54
Gunendradasan ve ark. (2018)	Frekans Modülasyonu Sapması	Gauss Karışım Modelinin	-	13.3
	Spektral Merkez Sapması (A)		-	11.45
	Spektral Merkez Frekansı (B)		-	12.34
	Spektral Merkez Büyüklük Katsayısı (C)		-	15.68
	A, B, C Füzyon		-	9.2
	Mel-Frekans Kepstral Katsayıları	Gauss Karışım Modeli-Evrensel Arka Plan Modeli	-	27.2
	Spektrogram Genlik		-	20.9
	Spektrogram Faz		-	26.8
	Sabit-Q-Kepstral Katsayılar		-	17.5
	Lineer Öngörü Kepstral Katsayıları		-	26.1
	Ters Mel-Frekans Kepstral Katsayıları		-	18.1
	Dikdörtgen Filtre Kepstral Katsayıları		-	20.3
	Lineer Filtre Kepstral Katsayıları		-	19
	Alt Bant Merkezi Büyüklük Katsayısı		-	21.1
	Alt Bant Merkezi Frekans Katsayısı		-	24
	Karmaşık Kepstral Katsayıları		-	25.9

Çizelge 11. ASVSpoof 2017 veriseti kullanan çalışmaların performans sonuçları (devam)

Referanslar	Kullanılan özellikler	Sınıflama yöntemi	Geliştirme seti	Değerlendirme seti
Değerlendirme metrikleri			**EER**	**EER**
Gunendradasan ve ark. (2018)	GMM'li özelliklerin Lojistik Regresyon Füzyonu (A)		-	12.2
	Mel-Frekans Kepstral Katsayıları	Otokodlayıcı	-	24.2
	Spektrogram Genlik		-	20.2
	Spektrogram Faz		-	26.5
	Sabit-Q-Kepstral Katsayılar		-	21.5
	Lineer Öngörü Kepstral Katsayıları		-	31.2
	Ters Mel-Frekans Kepstral Katsayıları		-	21.6
	Dikdörtgen Filtre Kepstral Katsayıları		-	27.9
	Lineer Filtre Kepstral Katsayıları		-	24
	Alt Bant Merkezi Büyüklük Katsayısı		-	23
	Alt Bant Merkezi Frekans Katsayısı		-	35.9
	Karmaşık Kepstral Katsayıları		-	32
	Otokodlayıcılı özelliklerin Lojistik Regresyon Füzyonu (B)		-	12.6
	A, B Füzyon		-	10.8
Derin öğrenmeye dayalı özellikler kullanan çalışmalar				
Nagarsheth ve ark. (2017)	Sabit Q Kepstral katsayıları (A)	Gauss Karışım Modeliyle	11	24.7
	Normalize edilmiş Sabit Q Kepstral katsayıları (B)		13.7	28.5
	Yüksek Frekanslı Kepstral Katsayıları (C)		5.9	23.9
	A, C Füzyon		3.2	18.1
	Sabit-Q-Kepstral katsayıları ve Derin Sinir Ağı	Destek Vektör Makinesi	6.9	16.5
	Sabit-Q-Kepstral katsayıları ve Yüksek Frekanslı Kepstral Katsayıları özellik düzeyinde füzyon ve Derin Sinir Ağı		7.6	11.5
Sriskandaraja ve ark. (2018)	Kesilmiş normalleştirilmiş FFT spektrogramlar ve Siamese yapısı	Gauss Karışım Modellemesi	-	6.4

ASV2019 LA verisetinde eğitim ve test işlemini gerçekleştiren yöntemlere ilişkin sonuçlar ise Çizelge 12'de toplanmıştır. Bu çalışmalardan EER metriğinin yanında t-DCF metrik sonuçları da geliştirme ve değerlendirme setleri için ayrı ayrı verilmiştir.

Çizelge 12. ASVSpoof 2019 LA veriseti kullanan çalışmaların performans sonuçları

Geleneksel yöntemler ile özellik elde eden çalışmalar						
Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**		**Değerlendirme seti**	
Değerlendirme metrikleri			**EER %**	**t-DCF**	**EER %**	**t-DCF**
Alzantot ve ark. (2019)	Mel-Frekans Kepstral Katsayıları (A)	Artık Konvolüsyonel Sinir Ağı	3.34	0.10	9.33	0.204
	Sabit-Q-Kepstral Katsayıları (B)		0.01	0.0002	7.69	0.217
	STFT'nin Logaritmik Büyüklüğü (C)		0.11	0.002	9.68	0.274
	A, B, C Füzyon		0	0	6.02	0.157
Alluri ve Vuppala (2019)	Tek Frekans Filtresi Kepstral Katsayıları (A)	Gauss Karışım Modeli	0.12	0.0034	5.22	0.129
	Sıfır Zamanlı Pencereleme Kepstral Katsayılar (B)		0.04	0.0005	6.13	0.141
	Anlık Frekans Kepstral Katsayıları		0.01	0.0002	10.21	0.289
	Sabit-Q-Kepstral Katsayıları (C)		0.43	0.012	9.57	0.236
	B, C Füzyon		0	0	4.92	0.124
	Sabit-Q-Kepstral	Gauss Karışım Modelleri-Evrensel Arka Plan Modeli	0.43	0.012	-	-
	Sabit-Q-Kepstral	Dilated ResNet	0	0	-	-
	Sabit-Q-Kepstral	Dikkatli Filtreleme Ağı	0	0	-	-
	Log Güç Büyüklüğü Spektrumları (F)	SENet34	0	0	11.75	0.216
	Log Güç Büyüklüğü Spektrumları (G)	SENet50	0	0	-	-
	Log Güç Büyüklüğü Spektrumları (H)	Mean-Std ResNet	0	0	-	-
	Log Güç Büyüklüğü Spektrumları (I)	Dilated ResNet	0	0	-	-
	Log Güç Büyüklüğü Spektrumları	Dikkatli Filtreleme Ağı	0	0	-	-
	Doğrusal Frekans Kepstral Katsayıları		2.71	0.066	-	-
	Sabit-Q-Kepstral	Gauss Karışım Modelleri	4.12	0.121	-	-
	Sabit-Q-Kepstral (J)	Mean-Std ResNet	0.04	0.001	-	-
	F, G, H, I, J Füzyon		0	0	6.7	0.155

Çizelge 12. ASVSpoof 2019 LA veriseti kullanan çalışmaların performans sonuçları (devam)

Geleneksel yöntemler ile özellik elde eden çalışmalar						
Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**		**Değerlendirme seti**	
Değerlendirme metrikleri			**EER %**	**t-DCF**	**EER %**	**t-DCF**
Chettri ve ark. (2019)	İlk 4 saniyesinden Ortalama-Varyans Normalleştirilmiş Log Spektrogramı (A)	Evrişimli Sinir Ağı	0.32	0.007	7.66	0.179
	Son 4 saniyesinden Ortalama-Varyans Normalleştirilmiş Log Spektrogramı (B)		0.27	0.004	5.98	0.167
	Normalleştirilmiş Log-Mel Spektrogramı (C)	Evrişimsel Tekrarlayan Sinir Ağı	5.65	0.17	-	-
	Ters Çevrilmiş Mel Frekansı Cepstral Katsayıları (G)	Gauss Karışım Modellemesi	1.73	0.044	-	-
	Alt Bant Ağırlık Merkezi Büyüklük Katsayıları (H)		-	-	-	-
	İ-Vektörleri (I)		0.16	0.005	-	-
	A, B, C, G, H, I Füzyon		0	0	2.64	0.076
Chen ve ark. (2020)	60 boyutlu lineer filtre bankaları	ResNet-18	-	-	1.26	-
Rahul ve ark. (2020)	Mel-Spektrogram	ResNet34	0.91	-	5.32	0.151
Zhang ve ark. (2021a)	60 boyutlu Doğrusal Frekans Kepstral Katsayıları	ResNet-18	0.2	0.006	2.19	0.059
Kwak ve ark. (2023)	Sabit-Q-Dönüşümü	ResMax	0,56	0.018	2,19	0.06
		ResMaxSep	-	-	2.55	-
Derin öğrenmeye dayalı özellikler kullanan çalışmalar						
Chintha ve ark. (2020)	Ham ses	CRNNSpoof	2.94	0.069	4.02	0.13
Jiang ve ark. (2020)	Ham ses ve Geçitli Tekrarlayan Ünite SSAD	SENet12	0.15	-	7.24	-
	Ham ses ve Zamansal Evrişimli Ağın kullanıldığı SSAD	SENet12	0.47	-	6.55	-
	Ham ses ve SSAD	LCNN-small	0.86	-	7.16	-
	Ham ses ve SSAD	LCNN-big	0.78	-	5.31	-
Tak ve ark. (2021)	Sabit Mel-Ölçekli Sinc Filtreleri RawNet2		1.36	0.46	5.64	0.13
	Sabit Ters Mel-ölçekli Sinc Filtreleri RawNet2		0.86	0.285	5.13	0.12
	Sabit Lineer Ölçekli Sinc Filtreleri RawNet2		1.25	0.4	4.66	0.13
	Füzyon		-	-	1.14	0.035

Çizelge 12. ASVSpoof 2019 LA veriseti kullanan çalışmaların performans sonuçları (devam)

Derin öğrenmeye dayalı özellikler kullanan çalışmalar						
Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**		**Değerlendirme seti**	
Değerlendirme metrikleri			**EER %**	**t-DCF**	**EER %**	**t-DCF**
Hua ve ark. (2021)	Res-TSSDNet		0.74	-	1.64	-
	Inc-TSSDNet		1.09	-	4.04	-
Zhang ve ark. (2021b)	Log Güç Spektrumu ve Transformer Kodlayıcı	34 katmanlı ResNet	0.11	-	6.02	-
	Mel-Frekans Kepstrum Katsayısı ve Transformer Kodlayıcı		0.21	-	6.54	-
	Sabit-Q-Kepstral Katsayısı ve Transformer Kodlayıcı		0.19	-	7.14	-
Zarish ve ark. (2022)	Ses Analizcisi RawNet, Maximum Özellik Haritalaması		-	-	4.93	0.18
	Ses Analizcisi RawNet, LeakyRelu		-	-	6.76	0.21
	Ses Analizcisi RawNet, SiLU		-	-	7.25	0.22
	Ses Analizcisi RawNet, Relu		-	-	9.39	0.32
Tan ve ark. (2023)	60 boyutlu Doğrusal Frekans Kepstral Katsayıları ve ResNet-18	Destek Vektör Makinesi, Radyal Tabanlı Fonksiyon	-	-	0.82	-
		Destek Vektör Makinesi, Doğrusal	-	-	0.57	-
		Destek Vektör Makinesi, Polinom	-	-	23.26	-
		Destek Vektör Makinesi, Sigmoid	-	-	0.94	-
Abdzadeh & Veisi (2023)	Kısa Zamanlı Fourier Dönüşümü ve ResNet-18	Çok Katmanlı Algılayıcı	-	-	-	-
	Doğrusal Frekans Kepstral Katsayıları ve ResNet-18		-	-	-	-
	Mel-Frekans Kepstral Katsayıları ve ResNet-18		-	-	-	-
	Sabit-Q-Dönüşümü Güç Spektrogram ve ResNet-18		-	-	2.33	0.12

ASVSpoof 2019 PA verisetinde sonuç rapor eden yöntemlere ilişkin sonuçlar ise Çizelge 13'de sunulmuştur. Geliştirme setinde en düşük EER sonucu Lai ve ark. (2019) tarafından önerilen yaklaşımların dörtlü füzyonunda elde edildiği görülmektedir. En düşük t-DCF değeri ise Kwak ve ark. (2023) arkadaşları tarafından önerilen Sabit-Q-Dönüşümü ile elde edilen özelliklerin ResMax ile sınıflandırılması şeklindeki yaklaşım ile elde edildiği söylenebilir. Çalışmalarda raporlanan sonuçlara göre özellik çıkarma aşamasında derin özelliklerin kullanılmasının da performans avantajı sağlamadığı şeklinde yorumlanabilir.

Çizelge 13. ASVSpoof 2019 PA veriseti kullanan çalışmaların performans sonuçları

Geleneksel yöntemler ile özellik elde eden çalışmalar						
Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**		**Değerlendirme seti**	
Değerlendirme metrikleri			**EER %**	**t-DCF**	**EER %**	**t-DCF**
Cheng ve ark. (2019)	Spectrogram (A)	18-katmanlı bir ResNeWt	3.15	0.088	-	-
	Mel-ölçekli filtre bankaları (B)		1.7	0.043	-	-
	CQT'ye dayalı log güç büyüklüğü spektrogramı (C)		0.39	0.011	-	-
	Geleneksel Değiştirilmiş Grup Gecikme İşlevi (D)		0.97	0.025	2.15	0.046
	Sabit-Q-Dönüşümü tabanlı CQTMGD (E)		0.54	0.015	0.94	0.025
	B, C Füzyon (F)		0.41	0.009	0.52	0.013
	D, F Füzyon		0.28	0.006	-	-
	E, F Füzyon		0.31	0.007	-	-
	A,B,C,D,E Füzyon		0.2	0.005	0.39	**0.009**
Alzantot ve ark. (2019)	Mel-Frekans Kepstral Katsayıları (D)	Artık Konvolüsyonel Sinir Ağı	15.91	0.377	-	-
	Sabit-Q-Kepstral Katsayıları (E)		4.3	0.103	4.43	0.107
	STFT'nin Logaritmik Büyüklüğü (F)		3.85	0.096	3.81	0.099
	D, E, F Füzyon		2.65	0.058	2.78	0.069
Cai ve ark. (2019)	Sabit-Q-Kepstral Katsayıları	Gauss Karışım Modelleri	9.87	0.195	11.04	0.245
	Doğrusal Frekans Kepstral Katsayıları		11.96	0.255	13.54	0.302
	Sabit-Q-Kepstral Katsayıları	ResNet	4.77	0.1127	-	-
	Doğrusal Frekans Kepstral Katsayıları (A)		2.62	0.0609	-	-
	Ters-Mel Frekans Kepstral Katsayıları (B)		3.66	0.089	-	-
	Kısa-Süreli Fourier Dönüşümü Gram		4.11	0.107	-	-
	Grup gecikme gramı (D)		1.81	0.0467	1.08	0.028
	Ortak Gram		1.14	0.0209	1.23	0.032
	Veri arttırma yapılarak STFT gram (C)		3.35	0.0904	-	-
	Veri arttırma yapılarak GD gram (E)		1.03	0.0265	1.79	0.044
	Veri arttırma yapılarak Joint gram (F)		1.14	0.0209	-	-
	A,B,C,D,E,F Füzyon		0.24	0.0064	0.66	0.017

Çizelge 13. ASVSpoof 2019 PA veriseti kullanan çalışmaların performans sonuçları (devam)

Geleneksel yöntemler ile özellik elde eden çalışmalar						
Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**		**Değerlendirme seti**	
Değerlendirme metrikleri			**EER %**	**t-DCF**	**EER %**	**t-DCF**
Alluri ve Vuppala (2019)	Tek Frekans Filtresi Kepstral Katsayıları	Gauss Karışım Modeli	11.09	0.24	13.97	0.323
	Sıfır Zamanlı Pencereleme Kepstral Katsayılar		10.11	0.22	12.2	0.281
	Anlık Frekans Kepstral Katsayıları		13.45	0.293	15.59	0.3573
	Sabit-Q-Kepstral Katsayıları		9.87	0.195	11.04	0.2454
Lai ve ark. (2019)	Sabit-Q-Kepstral	Gauss Karışım Modelleri	9.87	0.195	-	-
	Sabit-Q-Kepstral (E)	Mean-Std ResNet	1.43	0.041	-	-
	Sabit-Q-Kepstral	Dilated ResNet	0.78	0.024	-	-
	Sabit-Q-Kepstral	Dikkatli Filtreleme Ağı	0.74	0.021	-	-
	Log Güç Büyüklüğü Spektrumları (A)	SENet34	0.575	0.015	1.29	0.036
	Log Güç Büyüklüğü Spektrumları (B)	SENet50	0.631	0.017	-	-
	Log Güç Büyüklüğü Spektrumları (C)	Mean-Std ResNet	0.832	0.022	-	-
	Log Güç Büyüklüğü Spektrumları (D)	Dilated ResNet	1.072	0.029	-	-
	Log Güç Büyüklüğü Spektrumları	Dikkatli Filtreleme Ağı	1.057	0.027	-	-
	Doğrusal Frekans Kepstral Katsayıları		11.96	0.255	-	-
	Sabit-Q-Kepstral	Gauss Karışım Modelleri-Evrensel Arka Plan Modeli	12.52	0.322	-	-
	A,B,C,D,E Füzyon		**0.003**	0.129	0.59	0.016
Chettri ve ark. (2019)	İlk 4 saniyesinden Ortalama-Varyans Normalleştirilmiş Log Spektrogramı (A)	Evrişimli Sinir Ağı	10.77	0.28	-	-
	Son 4 saniyesinden Ortalama-Varyans Normalleştirilmiş Log Spektrogramı (B)		5.98	0.167	5.75	0.158
	A, B Füzyon		4.85	0.13	5.43	0.146
Kwak ve ark. (2023)	Sabit-Q-Dönüşümü	ResMax	0.16	**0.0042**	**0.3**	**0.009**
		ResMaxSep	-	-	0,36	-

Çizelge 13. ASVSpoof 2019 PA veriseti kullanan çalışmaların performans sonuçları (devam)

Derin öğrenmeye dayalı özellikler kullanan çalışmalar						
Referanslar	**Kullanılan özellikler**	**Sınıflama yöntemi**	**Geliştirme seti**		**Değerlendirme seti**	
Değerlendirme metrikleri			**EER %**	**t-DCF**	**EER %**	**t-DCF**
Zarish ve ark. (2022)	Ham ses	Ses Analizcisi RawNet, Maximum Özellik Haritalaması	-	-	5.29	0.21

5. Sonuç

Yapay zekâ alanındaki gelişmeler, internet ortamında kolaylıkla ulaşılabilinen işitsel içeriklerin manipülasyonunu kolaylaştırmış ve derin sahte seslerin ortaya çıkışını hızlandırmıştır. Hem toplumsal hem kişisel ciddi sorunlara sebep olabilecek şekilde bu seslerin kötü niyetli kullanımlara sebebiyet vermesi kaçınılmazdır. Bu tür sahte sesler, insan kulağı ile fark edilmesi mümkün olmayacak şekilde oluşturulmaktadır. Bu nedenle bir sesin bu sahtecilik türü ile oluşturulup oluşturulmadığını otomatik olarak tespit edebilecek algılama araçlarının ihtiyacı açıktır. Araştırmacıların konunun önemini fark ederek gerçekleştirdikleri çalışmalar mevcut olsa da derin sahte ses oluşturma tekniklerinde iyileştirmelerin de bir yandan devam etmesi ve bu seslerin sosyal medya gibi ortamlarda yayılma hızının artması kritik sorunu devam ettirmektedir. Bu çalışmada literatürde derin sahte ses tespiti için önerilen yaklaşımlar iki grup halinde incelenmiş, önerilen modellerin eğitimi için kullanılan verisetleri üzerinde performans analiz sonuçlarından bahsedilmiştir. Çalışmalar incelenirken göze çarpan en önemli durumun yöntemlerin genellikle tek bir saldırı türüne odaklandığı ve çalışmalarda tek bir veriseti üzerinde eğitim ve test yapılmasıdır. Bu da çalışmaların genelleştirme problemine sahip olduğu söylenebilir. Manipülasyonun gerçekleştirilmesinde kullanılan teknik önceden bilinmeyeceğinden, yeniden oynatma, metinden konuşmaya ve ses dönüşüm saldırılarının hepsini elen alan dayanıklı bir sistem ihtiyacının henüz tam olarak karşılanamadığı söylenebilir. Bu sebeple ele alınan problemin çözümüne ilişkin geliştirilen yöntemlerin yüksek performans ile genelleştirilebilir yöntemler olması gerekliliği kaçınılmazdır. Ayrıca derin sahte ses üretiminde yeni algoritmaların kullanılması ve yeni modellerin geliştirilmesinin devam edeceği öngörüsü ile bu sorunun güncelliğini koruması düşünülmektedir.

Teşekkür

Bu çalışma 1001 projesi kapsamında Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenmektedir (Proje No: 122E013).

Kaynakça

Abdzadeh, P., & Veisi, H. (2023). A comparison of CQT spectrogram with STFT-based acoustic features in Deep Learning-based synthetic speech detection. *Journal of AI and Data Mining*, 11(1), 119-129. doi:10.22044/jadm.2022.12373.2382

Alluri, K. N. R. K., & Vuppala, A. K. (2019, September). *IIIT-H spoofing countermeasures for automatic speaker verification spoofing and countermeasures challeng*. Interspeech 2019, Graz, Austria. doi:10.21437/Interspeech.2019-1623

Alzantot, M., Wang, Z., & Srivastava, M. B. (2019, September). *Deep residual neural networks for audio spoofing detection*. Interspeech 2019, Graz, Austria. doi:10.21437/Interspeech.2019-3174

Balamurali, B. T., Lin, K. W. E., Lui, S., Chen, J. M., & Herremans, D. (2019). Toward robust audio spoofing detection: a detailed comparison of traditional and learned features. *IEEE Access*, 7, 84229-84241. doi:10.1109/ACCESS.2019.2923806

Borrelli, C., Bestagini, P., Antonacci, F., Sarti, A., & Tubaro, S. (2021). Synthetic speech detection through short-term and long-term prediction traces. *EURASIP Journal on Information Security*, 2021, 2. doi:10.1186/s13635-021-00116-3

Cai, W., Wu, H., Cai, D., & Li, M. (2019, September). *The DKU replay detection system for the ASVspoof 2019 challenge: on data augmentation, feature representation, classification, and fusion*. Interspeech 2019, Graz, Austria. doi:10.21437/Interspeech.2019-1230

Chen, T., Kumar, A., Nagarsheth, P., Sivaraman, G., & Khoury, E. (2020, November). *Generalization of audio deepfake detection*. The Speaker and Language Recognition Workshop (Odyssey 2020), Tokyo, Japan. doi:10.21437/Odyssey.2020-19

Chen, Z., Xie, Z., Zhang, W., & Xu, X. (2017, August). *ResNet and model fusion for automatic spoofing detection*. Interspeech 2017, Stockholm, Sweeden. doi:10.21437/Interspeech.2017-1085

Cheng, X., Xu, M., & Zheng, T. F. (2019, March). *Replay detection using CQT-based modified group delay feature and ResNeWt network in ASVspoof 2019*. 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China. doi:10.1109/APSIPAASC47483.2019.9023158

Chettri, B., Stoller, D., Morfi, V., Ramírez, M. A. M., Benetos, E., Sturm, B. L. (2019, September). *Ensemble models for spoofing detection in automatic speaker verification*. Interspeech 2019, Graz, Austria. doi:10.21437/Interspeech.2019-2505

Chintha, A., Thai, B., Sohrawardi, S. J., Bhatt, K., Hickerson, A., Wright, M., & Ptucha, R. (2020). Recurrent convolutional structures for audio spoof and video deepfake detection. *Journal of Selected Topics in Signal Processing*, 14(5), 1024-1037. doi:10.1109/JSTSP.2020.2999185

Dua, M., Jain, C., & Kumar, S. (2022). LSTM and CNN based ensemble approach for spoof detection task in automatic speaker verification systems. *Journal of Ambient Intelligence and Humanized Computing*, 13, 1985-2000. doi:10.1007/s12652-021-02960-0

Font, R., Espín, J. M., & Cano, M. J. (2017, August). *Experimental analysis of features for replay attack detection — results on the ASVspoof 2017 challenge*. Interspeech 2017, Stockholm, Sweeden. doi:10.21437/Interspeech.2017-450

Gunendradasan, T., Wickramasinghe, B., Le, N. P., Ambikairajah, E., & Epps, J. (2018, September). *Detection of replay-spoofing attacks using frequency modulation features*. Interspeech 2018, Hyderabad, India. doi:10.21437/Interspeech.2018-1473

Hua, G., Teoh, A. B. J., & Zhang, H. (2021). Towards end-to-end synthetic speech detection. *IEEE Signal Processing Letters*, 28, 1265-1269. doi:10.1109/LSP.2021.3089437

Jiang, Z., Zhu, H., Peng, L., Ding, W., & Ren, Y. (2020, October). Self-supervised spoofing audio detection scheme. Interspeech 2020, Shangai, China. doi:10.21437/Interspeech.2020-1760

Kinnunen, T., Delgado, H., Evans, N., Lee, K.A., Vestman, V., Nautsch, A., …, & Reynolds, D. A. (2020). t-DCF: a detection cost function for the tandem assessment of spoofing countermeasures and automatic speaker verification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2195-2210. doi:10.1109/TASLP.2020.3009494

Korshunov, P., Marcel, S., Muckenhirn, H., Gonçalves, A. R., Souza Mello, A. G., Velloso, V. R. P., …, & Sahidullah, M. (2016, September). *Overview of BTAS 2016 speaker anti-spoofing competition*. 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, USA. doi:10.1109/BTAS.2016.7791200

Kwak, Y., Kwag, S., Lee, J., Jeon, Y., Hwang, J., Choi, H.J., …, & Yoon, J. W. (2023). Voice spoofing detection through residual network, max feature map, and depthwise separable convolution. *IEEE Access*, 11, 49140-49152. doi:10.1109/ACCESS.2023.3275790

Lai, CI, Chen, N., Villalba, J., & Dehak, N. (2019, September). *ASSERT: Anti-spoofing with squeeze-excitation and residual networks*. Interspeech 2019, Graz, Austria. doi:10.21437/Interspeech.2019-1794

Lavrentyeva, G., Novoselov, S., Malykh, E., Kozlov, A., Kudashev, O., & Shchemelinin, V. (2017, August). *Audio replay attack detection with deep learning frameworks*. Interspeech 2017, Stockholm, Sweden. doi:10.21437/Interspeech.2017-360

Mewada, H., Al-Asad, J. F., Almalki, F. A., Khan, A. H., Almujally, N. A., El-Nakla, S., & Naith, Q. (2023). Gaussian-filtered high-frequency-feature trained optimized BiLSTM network for spoofed-speech classification. *Sensors*, 23, 6637. doi:10.3390/s23146637

Nagarsheth, P., Khoury, E., Patil, K., & Garland, M. (2017, August). *Replay attack detection using DNN for channel discrimination*. Interspeech 2017, Stockholm, Sweden. doi:10.21437/Interspeech.2017-1377

Nautsch, A., Wang, X., Evans, N., Kinnunen, T. H., Vestman, V., Todisco, M., …, & Lee, K. A. (2021). ASVspoof 2019: Spoofing countermeasures for the detection of synthesized, converted and replayed speech. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2), 252-265. doi:10.1109/TBIOM.2021.3059479

Patel, T. B., & Patil, H. A. (2015, September). *Combining evidences from mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech*. Interspeech 2015, Dresden, Germany. doi:10.21437/Interspeech.2015-467

Paul, D., Sahidullah, M., & Saha, G. (2017, March). *Generalization of spoofing countermeasures: A case study with ASVspoof 2015 and BTAS 2016 corpora*. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA. doi:10.1109/ICASSP.2017.7952516

Qian, Y., Chen, N., & Yu, K. (2016). Deep features for automatic spoofing detection. *Speech Communication*, 85, 43-52. doi:10.1016/j.specom.2016.10.007

Rahul, T. P., Aravind, P. R., Ranjith, C., Usamath, N., & Paramparambath, N. (2020). Audio spoofing verification using deep convolutional neural networks by transfer learning. *ArXiv*, abs/2008.03464,2020. doi:10.48550/arXiv.2008.03464

Reimao, R., & Tzerpos, V. (2019, October). *FoR: A dataset for synthetic speech detection*. 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timisoara, Romania. doi:10.1109/SPED.2019.8906599

Suthokumar, G., Sethu, V., Wijenayake, C., & Ambikairajah, E. (2018, September). *Modulation dynamic features for the detection of replay attacks*. Interspeech 2018, Hyderabad, India. doi:10.21437/Interspeech.2018-1846

Sriskandaraja, K., Sethu, V., Ambikairajah, E., & Li, H. (2017). Front-end for antispoofing countermeasures in speaker verification: Scattering spectral decomposition. *IEEE Journal of Selected Topics in Signal Processing*, 11(4), 632-643. doi:10.1109/JSTSP.2016.2647202

Sriskandaraja, K., Sethu, V., & Ambikairajah, E. (2018, September). *Deep siamese architecture based replay detection for secure voice biometric*. Interspeech 2018, Hyderabad, India. doi:10.21437/Interspeech.2018-1819

Tak, H., Patino, J., Todisco, M., Nautsch, A., Evans, N., & Larcher, A. (2021, June). *End-to-End anti-spoofing with RawNet2*. ICASSP 2021- 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada. doi:10.1109/ICASSP39728.2021.9414234

Tan, C. B., Hijazi, M. H. A., & Nohuddin, P. N. E. (2023, September). *A hybrid classification approach for artificial speech detection*. 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia. doi:10.1109/IICAIET59451.2023.10291764

Xiao, X., Tian, X., Du, S., Xu, H., Siong, C. E., & Li, H. (2015, September). *Spoofing speech detection using high dimensional magnitude and phase features: the NTU approach for ASVspoof 2015 challenge*. Interspeech 2015, Dresden, Germany. doi:10.21437/Interspeech.2015-465

Wu, Z., Kinnunen, T., Evans, N., Yamagishi, J., Hanilçi, C., Sahidullah, M., & Sizov, A. (2015, September). *ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge*. Interspeech 2015, Dresden, Germany. doi:10.21437/Interspeech.2015-462

Wu, Z., Yamagishi, J., Kinnunen, T., Hanilçi, C., Sahidullah, M., Sizov, ..., & Delgado, H. (2017). ASVspoof: the automatic speaker verification spoofing and countermeasures challenge. *IEEE Journal of Selected Topics in Signal Processing*, 11(4), 588-604. doi:10.1109/JSTSP.2017.2671435

Witkowski, M., Kacprzak, S., Żelasko, P., Kowalczyk, K., & Gałka, J. (2017, August). *Audio replay attack detection using high-frequency features*. Interspeech 2017, Stockholm, Sweden. doi:10.21437/Interspeech.2017-776

Wang, L., Yoshida, Y., Kawakami, Y., & Nakagawa, S. (2015, September). *Relative phase information for detecting human speech and spoofed speech*. Interspeech 2015, Dresden, Germany. doi:10.21437/Interspeech.2015-473

Yamagishi, J., Wang, X., Todisco, M., Sahidullah, M., Patino, J., Nautsch, A., …, & Delgado H. (2021, September). *ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection*. 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge, France. doi:10.21437/ASVSPOOF.2021-8

Yang, J., You, C., & He, Q. (2018, September). *Feature with complementarity of statistics and principal information for spoofing detection*. Interspeech 2018, Hyderabad, India. doi:10.21437/Interspeech.2018-1693

Yu, H., Tan, Z. H., Zhang, Y., Ma, Z., Guo, J. (2017). DNN filter bank cepstral coefficients for spoofing detection. *IEEE Access*, 5, 4779-4787. doi:10.1109/ACCESS.2017.2687041

Zarish, A., Javed, A., & Khalid, M. (2022). *AEXANet: An end-to-end deep learning based voice anti-spoofing system*. Workshop on Artificial Intelligence for Multimedia Forensics and Disinformation Detection (AI4MFDD).

Zhang, Y., Jiang, F., & Duan, Z. (2021). One-class learning towards synthetic voice spoofing detection. *IEEE Signal Processing Letters*, 28, 937-941. https://doi.org/10.1109/LSP.2021.3076358

Zhang, Z., Yi, X., & Zhao, X. (2021, June). *Fake speech detection using residual network with transformer encoder*. Proceedings of the 2021 ACM workshop on information hiding and multimedia security, Belgium. doi:10.1145/3437880.3460408

Zhang, J., Tu, G., Liu, S., & Cai, Z. (2023). Audio anti-spoofing based on audio feature fusion. *Algorithms*, 16, 317. doi:10.3390/a16070317