BURSA TEKNİK ÜNİVERSİTESİ ❖ FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞLARI İLE KONUŞMACI DOĞRULAMA SİSTEMLERİ İÇİN SALDIRI TESPİTİ

YÜKSEK LİSANS TEZİ

BEKİR BAKAR

Elektrik-Elektronik Mühendisliği Anabilim Dalı

ARALIK 2018

BURSA TEKNİK ÜNİVERSİTESİ ❖ FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞLARI İLE KONUŞMACI TANIMA SİSTEMLERİ İÇİN SALDIRI TESPİTİ

YÜKSEK LİSANS TEZİ

Bekir BAKAR
(161082310)

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Cemal HANİLÇİ

ARALIK 2018

BTÜ, Fen Bilimleri Enstitüsü'nün 161082310 numaralı Yüksek Lisans Öğrencisi Bekir BAKAR, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "YAPAY SİNİR AĞLARI İLE KONUŞMACI TANIMA SİSTEMLERİ İÇİN SALDIRI TESPİTİ" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı: **Doç. Dr. Cemal HANİLÇİ**
Bursa Teknik Üniversitesi

Jüri Üyeleri: **Doç. Dr. Hakan GÜRKAN** .....
Bursa Teknik Üniversitesi

Doç. Dr. Ahmet Emir DİRİK .....
Uludağ Üniversitesi

Savunma Tarihi: 26 Aralık 2018

FBE Müdürü: **Doç. Dr. Murat ERTAŞ** .....
Bursa Teknik Üniversitesi **……/……./…….**

İNTİHAL BEYANI

Bu tezde görsel, işitsel ve yazılı biçimde sunulan tüm bilgi ve sonuçların akademik ve etik kurallara uyularak tarafımdan elde edildiğini, tez içinde yer alan ancak bu çalışmaya özgü olmayan tüm sonuç ve bilgileri tezde kaynak göstererek belgelediğimi, aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim.

Öğrencinin Adı Soyadı: Bekir BAKAR

İmzası:

Çok değerli aileme,

ÖNSÖZ

Bu tezde, Bursa Teknik Üniversitesi yüksek lisans programında yapmış olduğum, çalışmalarım sonucunda edindiğim bilgileri sizlere aktarmaktayım. Bu süreçte yardımlarını esirgemeyen başta Doç. Dr. Cemal HANİLÇİ olmak üzere Bursa Teknik Üniversitesinde görev yapan tüm akademisyenlere teşekkürü borç bilir, şükranlarımı iletmek isterim.

Bu çalışma, TÜBİTAK 115E916 numaralı proje tarafından desteklenmiştir.

Aralık 2018 Bekir BAKAR

İÇİNDEKİLER

Sayfa

KISALTMALAR

AFD : Ayrık Fourier Dönüşümü
ASVspoof : Automatic Speaker Verification Spoofing and Countermeasures Challenge (Otomatik Konuşmacı Doğrulama Yanıltma ve Karşı Önlemler Yarışması)
CPU : Central Processing Unit (Merkezi İşlem Birimi)
YSA : Yapay Sinir Ağları
DVM : Destek Vektör Makinaları
EHO : Eşit Hata Oranı
GKM : Gauss Karışım Modeli
GPU : Graphical Processing Unit (Grafik İşlem Birimi)
HFD : Hızlı Fourier Dönüşümü
KD : Konuşmacı Doğrulama
KSA : Konvolüsyonel Sinir Ağları
MFKK : Mel-frekansı Kepstrum Katsayıları
RAM : Random-access Memory (Rasgele Erişim Belleği)
SHÖ : Sezim Hata Ödünleşimi
SQKK : Sabit Q Kepstral Katsayıları
TO : Tekrar Oynatma
UDOS : Uzun Dönem Ortalama Spektrum

SEMBOLLER

α	**:** Ön Vurgulama Süzgeç Eğimi
kHz	**:** Kilo Hertz
s	**:** Saniye
T	**:** Transpoze
p	**:** Olasılık Değeri

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

YAPAY SİNİR AĞLARI İLE KONUŞMACI TANIMA SİSTEMLERİ İÇİN SALDIRI TESPİTİ

ÖZET

Kimlik kartları, manyetik kartlar gibi geleneksel ve eski şifreleme yöntemlerinin doğurduğu bazı olumsuz sonuçlar kişi veya kurumları alternatif şifreleme yöntemleri arayışına yöneltmiştir. Bir bireyin yüz, parmak izi, iris, avuç içi damar haritası, retina gibi fizyolojik veya yürüyüş biçimi, imza şekli, sesi gibi davranışsal özelliklerine biyometri denir. Biyometrik verileri kullanarak kimlik tespiti yapmayı hedefleyen sistemler ise, biyometrik doğrulama sistemleri olarak adlandırılır. Ses sinyalinin barındırdığı birçok kullanışlı bilgi ve bu bilgilerden elde edilen benzersiz çıktılar, verilen bir ses sinyalinin iddia edilen kişiye ait olup olmadığının tespit edilmesini hedefleyen, konuşmacı doğrulama (KD) sistemlerini cazip bir biyometrik sistem haline getirmiştir.

KD sistemlerinin kullanımının yaygınlaşması, sistemleri yanıltma çabalarını da beraberinde getirmiştir. Saldırgan diye adlandırılan bir kişi, KD sistemlerini yanıltmak için birçok yöntem kullanabilir. Bu yöntemlerden en basit, fakat bir o kadar da etkili olanı tekrar oynatma (TO) saldırısıdır. TO saldırısı daha önceden kaydedilmiş bir ses kaydının, KD sisteminin algılayıcı, mikrofon seviyesinden tekrar oynatılmasıdır. KD sistemlerinin güvenilirliğini ciddi bir şekilde tehdit ettiği yapılan çalışmalarda gösterilen TO saldırıları, bu tez çalışmanın temel konusudur.

Bu tez çalışmasında, KD sistemlerinde yaygın olarak kullanılan mel-frekansı kepstrum katsayıları, Otomatik Konuşmacı Doğrulama Yanıltma Saldırıları ve Karşı Önlemler (ASVspoof 2017) yarışmasında başarısı kanıtlanan sabit Q kepstral katsayıları ve uzun dönem ortalama spektrum (UDOS) öznitelikleri kullanılmıştır. Deneyler sırasında ASVspoof 2017 yarışması için hazırlanan veri tabanı kullanılmıştır. Sınıflandırıcı olarak ise, Gauss karışım modelinin yanı sıra; son yıllarda hemen her alanda kullanımı yaygınlaşan yapay sinir ağları (YSA) yaklaşımı esas alınmıştır. Bu tez çalışmasında, bahsedilen öznitelik ve sınıflandırıcılardan oluşan TO saldırı tespit sistemlerinin oluşturulması ve bu sistemlerin performanslarının detaylı olarak karşılaştırılması yapılmıştır.

UDOS özniteliklerinin, YSA yaklaşımı kullanılarak sınıflandırılması en iyi saldırı tespit sistemi olmuştur. Bu sistem geliştirme ve değerlendirme kümesi verileri için sırasıyla; %4,10 ve %20,77 eşit hata oranı üretmiştir. Ayrıca, alt-band frekans analizi yapılmış ve yüksek frekans bölgesinin tekrar saldırısı oluşturma senaryosunda en çok etkilenen bölge olduğu; yani yüksek frekans bölgesi analizinin TO saldırısı tespitinde önemli bir rol oynadığı gösterilmiştir. Ortalama ve varyans normalizasyonu işleminin ise etkili bir yöntem olup olmadığı da araştırılmış; etkili bir yöntem olmadığı gözlemlenmiştir.

Anahtar kelimeler: Konuşmacı doğrulama, yanıltma saldırıları, tekrar saldırısı tespiti, karşı önlemler, yapay sinir ağları

USING ARTIFICIAL NEURAL NETWORK ON ANTI-SPOOFING FOR SPEAKER VERIFICATION

SUMMARY

Some negative consequences of traditional and old authentication methods such as ID cards and magnetic cards have led people or institutions to search for alternative authentication methods. Physiological characteristics of an individual such as face, fingerprint, iris, palm vein map and retina or behavioral characteristics such as walking pattern, signature form, sound are called as biometrics. So systems that aim to identify of an indivudual using his/her biometrics are referred to as biometric verification systems. Since speech signal conveys many useful information which yields unique outputs from these information, automatic speaker verification systems (ASV) which aims to determine whether a given speech signal belongs to the claimed person or not, have become an attractive biometric system.

The widespread use of the ASV system has led to efforts to spoof the systems. A person called a fraud may use many methods to spoof ASV systems. The simplest of these methods yet the most effective is the replay attack. The replay attack can be described as re-playing a previously recorded speech of a target/genuine speaker to the ASV system from the sensor, the microphone level. The main focus of this dissertation is replay attack which is proved in many studies that have seriously threatened the reliability of ASV systems.

In this dissertation study, the mel frerquency cepstral coefficients (MFCC) widely used in ASV systems, constant Q cepstral coefficients (CQCC) that is shown its succes in Automatic Speaker verification Spoofing and Countermasures Challenge (ASVspoof 2017) competition and long-term average spectrum (LTAS) are used. The experiments were conducted using the database prepared for ASVspoof 2017 competition. As a classifier, besides Gaussian mixture model (GMM); artificial neural network (ANN) approach has been used since it is in almost all areas in recent years. In this thesis, replay attack detection system that consist of classifers and features that mention are developed and compared in detail.

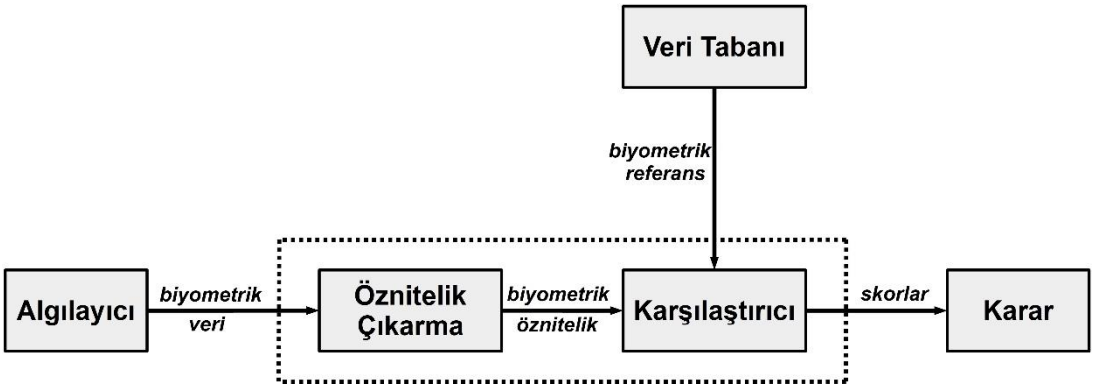
The best anti-spoofing system has been LTAS features when classified using ANN. This system has produced 4,10% and 20,77% equal error rate (EER) for the development and the evaluation dataset, respectively. In addition, sub-band frequency analysis is performed and found that the high frequency region is the most affected region in the relay attack scenario; it has been shown that high frequency region analysis plays an important role in the detection of replay attack. It was also investigated whether the mean and variance normalization procedure were effective methods during experimental studies and was observed that there was no effective method.

Keywords: Speaker verification, spoofing attacks, replay attack detection, countermeasures, artificial neural networks

1. GİRİŞ

Kişilere ait hassas verilerin korunması, çeşitli hizmetlere erişim kontrolü, banka işlemleri, tesis veya bina giriş kontrolü gibi birçok alanda güvenliğin sağlanması gerekmektedir. Yalnızca yetkili kişiye erişim izni verilmesi gereken bu ve benzeri durumlarda, çeşitli şifreleme yöntemleri (kimlik kartı, parola, manyetik kart vb.) kullanılmaktadır. Geleneksel/eski şifreleme yöntemlerinin kolay aldatılabilmesi, yavaş çalışması, maliyetli olması, çalınmasının kolay olması veya kolay unutulması gibi kullanıcıyı zor durumunda bırakabilecek dezavantajlar, alternatif şifreleme yöntemleri arayışını doğurmuştur. Bu arayışın doğal bir sonucu olarak, biyometrik doğrulama sistemleri günlük hayatta daha fazla yer almaya başlamıştır [1].

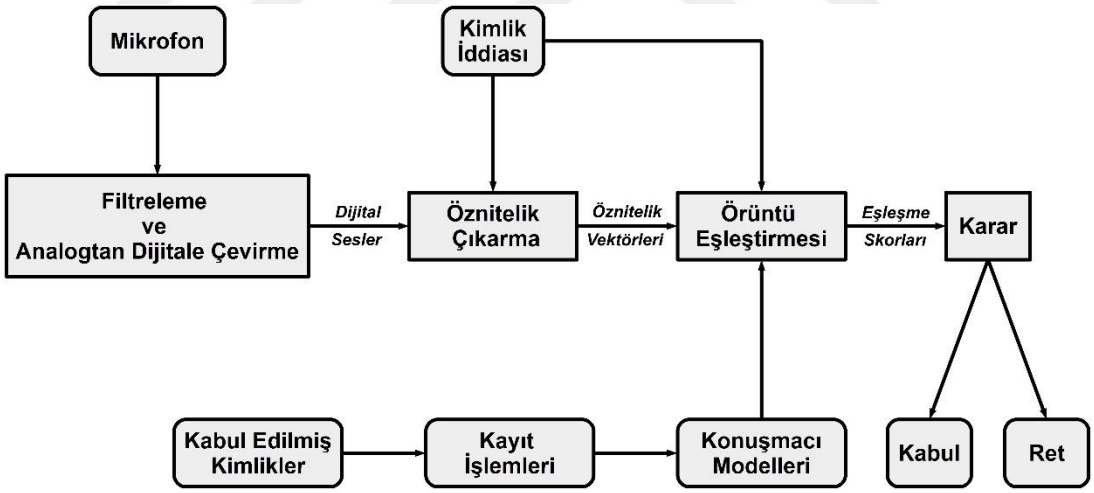
Bir bireyin yüz, parmak izi, iris, avuç içi damar haritası, retina gibi fizyolojik veya yürüyüş biçimi, imza şekli, sesi gibi davranışsal özelliklerine biyometri denir [2]. Bireyin ölçülebilir ve ayırt edici biyometrik özelliklerini kullanarak elde ettiği veriyi mevcut kayıtlar ile karşılaştırarak kimlik tespiti yapmayı hedefleyen yarı otomatik sistemlere ise biyometrik doğrulama sistemleri denilmektedir [1]. Burada artık şifreler, kimlik kartları, manyetik kartlar ve diğer bütün yöntemler yerini fizyolojik veya davranışsal ayırt edici özelliklere bırakmıştır. Bir biyometrik sistemin çalışma prensibi Şekil 1.1'de gösterilmiştir. Şekil 1.1'deki algılayıcıdan alınan fotoğraf, ses, titreşim gibi bilgilerden çıkarılan kullanışlı öznitelikler, önceden veri tabanına kaydedilmiş verilerle karşılaştırılır ve karar verilir. Kişinin kimlik iddiası kabul veya reddedilir.



Şekil 1.1: Biyometrik doğrulama sistemleri için işlem akış şeması [3]

Biyometrik verilerin benzersiz, kalıcı, ölçülebilir ve evrensel olması gibi avantajlarını kullanan biyometrik doğrulama sistemleri, makinanın bireyi otomatik doğrulaması veya kişinin tepkilerine yanıt vermesini kolay ve güvenli bir şekilde sağlaması nedeni ile güvenlik sistemlerinin birçok alanına uygulanabilir.

Ses sinyali birçok bilgi barındırmakta ve bu bilgilerden önemli ayırt edici çıktılar elde edilmektedir. Bu durum konuşmacı doğrulama sistemlerini cazip bir biyometrik sistem haline getirmiştir. Konuşmacı doğrulama (KD), verilen bir ses sinyalinin iddia edilen kişiye ait olup olmadığının tespit edilmesini hedefleyen kimlik kabul veya reddetme işlemidir [4]. KD sistemleri oluşturulurken ilk olarak, bilinen kimliklerin kayıt işlemleri yapılır ve bu kayıtlardan çeşitli yöntemler kullanılarak bazı modeller oluşturulur. Gerçek zamanlı çalışması esnasında ise; mikrofondan alınan ses sinyalleri, filtreleme işleminden geçirildikten sonra dijital veriye dönüştürülür. Daha sonra bu sinyallerden çıkarılan öznitelikler, önceden oluşturulmuş model kullanılarak örüntü eşleştirmesi işlemine tabi tutulur. Son olarak eşleştirme skorlarından yola çıkılarak kimlik iddiasının kabulü veya reddedilmesi işlemi gerçekleştirilir. KD sisteminin çalışma prensibini gösteren genel bir işlem akış diyagramı Şekil 1.2'de gösterilmiştir.



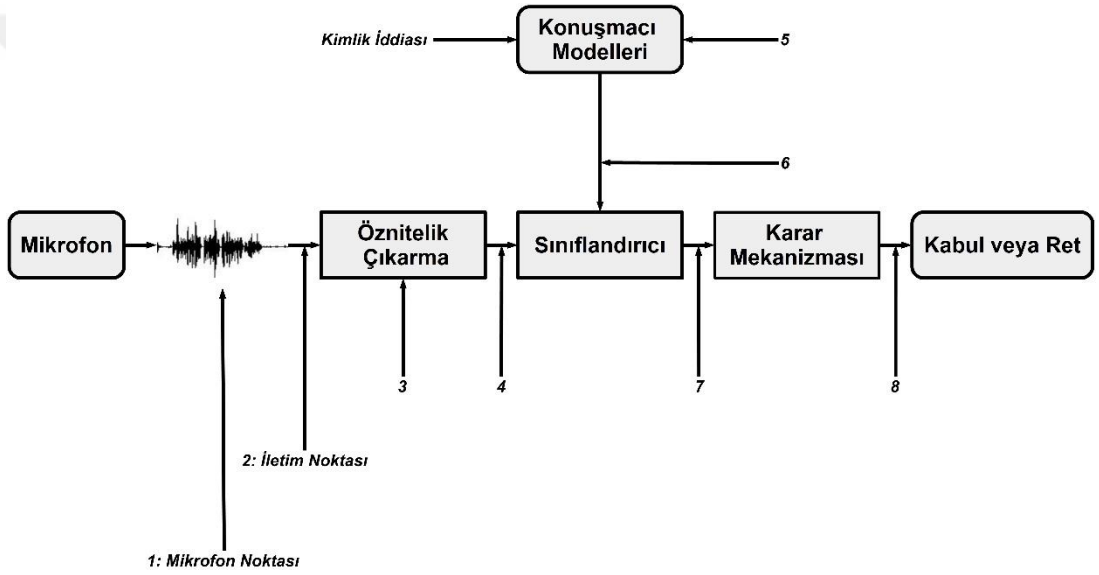
Şekil 1.2: Konuşmacı doğrulama sistemi için genel bir işlem akış diyagramı [5]

KD sistemlerinin mevcut sistemlere kolay adapte edilebilmesi, maliyetinin görece düşük olması, diğer biyometrik sistemlere göre aldatılmasının daha zor olması gibi durumlar konuya olan ilgiyi artırmıştır. Telekomünikasyon operatörleri, banka işlemleri, çağrı merkezleri, kişisel verilerin güvenliği, çeşitli sistemlere erişim kontrolü ve daha birçok alanda kullanılan sistemler önemli bir ekonomik altyapı barındırmaktadır. Artan bu ilgi ve yaygın kullanım, KD sistemlerini aldatmaya yönelik

saldırı girişimlerini de beraberinde getirmiştir. En gelişmiş teknolojik imkanlardan faydalanılarak oluşturulan KD sistemleri de dahil olmak üzere, yanıltma saldırıları KD sistemleri için önemli bir tehdit olup, KD sistemlerinin bu saldırılara karşı savunmasız olduğu açıkça gösterilmiştir [6, 7, 8, 9].

KD sistemlerinin ilk amacı, kimlik iddiasında bulunan kişiyi doğru bir şekilde kabul etmesi veya reddetmesidir. Bu işlem sırasında yanlış kişinin kabul edilmesi veya doğru kişinin reddedilmesi şeklinde iki çeşit hata oluşabilir.

Herhangi bir biyometrik doğrulama sistemine saldırı noktaları diye adlandırılan farklı kısımlardan ve çeşitli yöntemler kullanarak saldırılar yapılabilir [9, 10]. KD sistemlerine yapılabilecek muhtemel saldırı noktaları Şekil 1.3'de gösterilmiştir.



Şekil 1.3: Konuşmacı doğrulama sistemlerinde yanıltma saldırısı noktaları [9]

Herhangi bir biyometrik doğrulama sistemine yapılabilecek saldırılar doğrudan ve dolaylı saldırılar olmak üzere iki kısımda incelenir. Şekil 1.3'de belirtilen 1 ve 2 numaraları saldırı noktaları doğrudan yapılan saldırılar olarak nitelendirilmekte olup bu saldırılara sunum saldırısı da denilmektedir [9]. 1 ve 2 numaralı saldırı noktaları haricindeki noktalardan yapılan saldırılar ise dolaylı saldırılardır [9]. Doğrudan saldırılar mikrofon/algılayıcı kısmından basitçe amatör bir kişi tarafından yapılabilmektedir; buna karşın dolaylı saldırılar sistemin çalışma prensibi ile ilgili detayların bilinmesi ve saldırganın bu konuda uzmanlaşmış bir kişi olmasını gerektirmektedir. Bu sebeple, doğrudan yapılan saldırılar dolaylı saldırılara nazaran daha ciddiye alınması gereken saldırı yöntemleridir.

1.1 Tezin Amacı

Saldırgan, KD sistemlerini aldatmak için mikrofon veya algılayıcı seviyesinden gerçekleştireceği sunum saldırılarında birçok etkili yöntem kullanılabilmektedir. KD sistemlerine mikrofon seviyesinden yapılması olası saldırı türleri dört alt kategoriye indirgenmiştir [11]. Bunlar, konuşmacının sesinin sentezlenmesiyle oluşturulan ses sentezleme [12], herhangi bir sesin hedef kişinin sesine dönüştürüldüğü ses dönüştürme [13], hedef kişinin sesinin taklit edilmesi [14] ve hedef konuşmacının önceden kaydedilmiş ses kayıtlarının kullanılmasıyla oluşturulan tekrar oynatma (TO) [15] saldırılarıdır.

Bu saldırı türlerinden, taklit saldırısı özel bir yetenek gerektirdiğinden karşılaşılması güç saldırı türüdür ve önemli bir tehdit olarak algılanmamaktadır. Ses sentezleme ve ses dönüştürme saldırılarının ise KD sistemlerini aldatma oranı yüksektir, fakat bu alanda yapılmış çalışmalarla önemli ölçüde ilerleme kaydedilmiştir. TO saldırıları ise hemen herkesin sahip olduğu sıradan bir cep telefonu ile basitçe yapılabilmekte ve sistemleri aldatma oranı oldukça yüksek olmaktadır. Bu yüzden en güncel ve en fazla ciddiye alınması gereken saldırı türü olarak dikkat çekmektedir.

Bu tez çalışmasının temel amacı, KD sistemlerinin güvenirliğini ciddi bir şekilde tehdit eden TO saldırılarının tespit edilmesidir. Çalışmayla birlikte TO saldırıları için yapılmış araştırmalar incelenecek, yöntemler yeniden test edilecek ve yapay sinir ağları (YSA) yaklaşımının güvenilirliği araştırılacaktır.

1.2 Literatür Araştırması

1.2.1 Tekrar oynatma saldırısının konuşmacı doğrulama sistemlerine etkisi

KD sistemleri için başarılı olarak bilinen öznitelik çıkarma, model oluşturma ve performans ölçme yöntemlerinin tekrar saldırıları karşısında pek de başarılı olmadığı Villalba ve Llelida tarafından yapılan çalışmada gösterilmiştir [6]. Söz konusu çalışma kapsamında beş adet konuşmacıdan oluşan dört ayrı kategoride veri tabanı oluşturulmuştur. Bunlardan birincisi, konuşmacılardan mikrofona yakın bir mesafeden kayıt alınıp daha sonra bu kayıtların farklı telefon kanallarıyla iletildiği orijinal kayıtlardır. Bu grupta bir adet eğitim ve yedi adet test sinyali bulunmaktadır. Dijital olarak iletilen seslerden bir eğitim ve üç test, analog kabloyla ve kablosuz yolla iletilen seslerden ikişer adet test sinyali oluşturulmuştur. İkinci veri tabanı orijinal kayıtlarla

aynı anda uzak mesafeli bir mikrofonla kayıt altına alınan seslerden oluşmaktadır. Üçüncüsü, mikrofon test kayıtlarının TO saldırısı oluşturmak için telefon ahizesiyle tekrar kaydedildikten sonra analog kanal aracılığıyla iletildiği analog tekrar kayıtlarıdır. Sonuncusu ise, mikrofon test kayıtlarının tekrar kaydedilip dijital kanal aracılığıyla iletildiği dijital tekrar kayıtlarıdır. Orijinal kayıtlarla sistem test edildiğinde %0,71 eşit hata oranı (EHO) elde edilirken; analog ve dijital tekrar kayıtları ile elde edilen en düşük EHO değerleri sırasıyla %2,42 ve %2,30 olarak belirtilmiş, yanlış kabul oranı %68 olmuştur. Bahsi geçen çalışma TO saldırıları için yapılan ilk çalışmalardan olup, veri tabanı oluşturma konusunda yol göstericidir [6].

TO saldırısı için kayıt oluşturma esnasında, ses sinyalinde yankılanma ve gürültü oluşacağı ve bunun sonucu olarak sinyalin spektrumunda bir düzleşmeyle beraber modülasyon indekslerinde kayıp oluşacağı tespit edilmiş; bu gürültü ve yankılanmanın tespit edilmesinin saldırıların önlenmesindeki en önemli etkenlerden biri olacağı Villalba ve Lleida tarafından 2011 yılında yapılan çalışmada öne sürülmüştür [16]. Bahsedilen bu iki durumun ayrımı için dört adet öznitelik çıkarım yöntemi kullanılmış, sınıflandırıcı olarak ise destek vektör makinaları (DVM) eğitilmiştir. DVM'yi eğitmek için özel olarak oluşturulan veri tabanı kullanılmış; eğitim aşamasında saldırı seslerinin de kullanılması gerektiği önemle vurgulanmıştır. Çalışmanın sonucunda başarılı bir saldırı önleme sistemi oluşturulmuş ve yanlış kabul oranı sıfır olarak belirtilmiştir. Ayrıca EHO %2 ile %9 arasında kalmıştır. Bahsedilen çalışma, TO saldırısı tespiti için öznitelik çıkarım ve eğitim aşamasında kullanılması gereken veri tabanına dair önemli bir yol göstericidir [16].

Bir diğer çalışmada ise, KD sistemlerinde tekrar saldırılarının önlenmesi için kanal gürültüsünün esas alınması gerektiği vurgulanmıştır [17]. Kanal gürültüsünün inceleme altına alınan her bir kayıt için benzersiz kimlik işlevi gördüğü belirtilmiştir. Ayrımları yapabilecek öznitelikleri çıkarmak için, gürültü kaldırıcı filtreler ve istatiksel çerçeveler yaklaşımı uygulanmış ve 6 adet Legendre katsayıları ve istatistiksel öznitelikler çıkarılmıştır. Daha sonra bu özniteliklerle DVM sistemi eğitilmiş ve önerilen sistem EHO'yu mevcut sistemlere göre %30 oranında iyileştirmiştir. Yanlış kabul ve yanlış ret oranlarının ise %50 oranında düştüğü belirtilmiştir. Bahsedilen çalışma öznitelik çıkarımında dikkate alınması gereken hususlar hakkında bilgi sunmaktadır.

KD sistemlerine yapılması olası saldırı türlerini kategorize eden, saldırı noktalarına dikkat çeken, geçmişte yapılan çalışmaların bir incelemesini sunan ve mevcut sistemlerin karşılaştırılmasının yapıldığı bir çalışma 2015 yılında yapılmıştır [9]. Birleşik etmen analizi ve Gauss karışım modeli (GKM) tabanlı mevcut KD sistemleri TO saldırısı olması durumunda test edilmiş ve EHO'nun yaklaşık %30 oranında yükseldiği ortaya çıkmıştır. Bahsedilen çalışma, TO saldırısının yanı sıra ses sentezleme, ses dönüştürme ve taklit gibi diğer saldırı yöntemlerini de incelemiştir. Özellikle TO saldırısının yanıltma başarısına dikkat çekerek mevcut sistemlerin yetersizliği de ortaya koyulmuştur. Araştırmacıların daha güvenli saldırı yöntemlerinin geliştirilmesi ve performanslarının karşılaştırılabilmesi için ortak bir veri tabanı oluşturulması gerektiğinin de altı çizilmiştir.

Teknik bilgi gerektiren ve literatürde çalışmaları mevcut olan ses dönüştürme, ses sentezleme gibi saldırı yöntemleri yerine, yapılması daha basit ve üzerine pek çalışma yapılmamış TO saldırısı üzerine daha fazla çalışma yapılması gerekliliğini vurgulayan bir çalışma 2011 yılında yapılmıştır [7]. Çalışmada, i-vektör-olasılıksal doğrusal ayırt edici analiz, GKM modeli de dahil olmak üzere altı adet KD sistemi test edilmiştir. Oluşturulan en iyi sistemin dahi TO saldırısı karşısında savunmasız olduğu gösterilmiştir. Geliştirilen en iyi saldırı tespit sisteminin %10'nun altında yanlış kabul oranına sahip olasılıksal doğrusal ayırt edici analiz tabanlı sistem olduğu tespit edilmiş olup, bu durum TO saldırısının zor bir konu olduğunu göstermektedir. Çalışma TO saldırısı tespitinin önemini ve zorluğunu pekiştirmiştir.

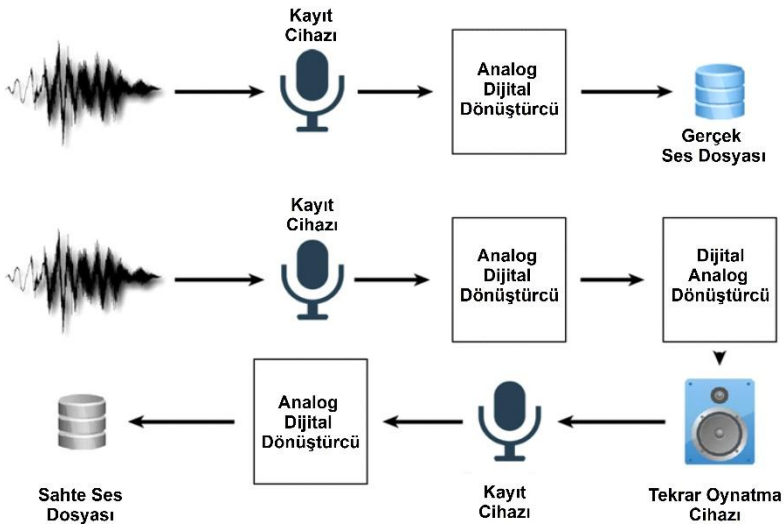
1.2.2 Tekrar oynatma saldırısı tespitinde özniteliklerin rolü

Yanıltma saldırılarının tespitinde kullanılan birçok öznitelik türü vardır. Bunlar uzun dönem spektral istatistik, faz ve güç spektrumu tabanlı öznitelikler olarak üç kategoride ele alınmaktadır [18]. Konuşma sinyalinin tamamının ayrık Fourier dönüşümü (AFD) alınarak birinci derece ve ikinci derece istatistiklerinin ortalama ve standart sapmasının tek bir vektörde birleştirilmesiyle elde edilen öznitelikler, uzun dönem spektral istatistik tabanlı özniteliklere bir örnektir [19]. Tipik faz spektrum tabanlı öznitelik örnekleri ise normalize faz özniteliği, grup gecikmesi, anlık frekans ve anlık frekans kosinüs katsayılarıdır [20]. Güç spektrumu tabanlı özniteliklerin en bilinen versiyonu ise sabit Q kepstral katsayılarıdır (SQKK) [21]. SQKK gerçek ve sahte sesin ayırt edilmesinde en yoğun kullanılan öznitelik türü olarak karşımıza

çıkmaktadır [22, 23, 24, 25]. KD sistemlerine yapılabilecek TO saldırısını tespit etmek amacı ile gerçek seslerin sahte seslerden ayırt edilmesi üzerine 2017 yılında düzenlenen Otomatik Konuşmacı Doğrulama Yanıltma ve Saldırıları Tespiti (ASVspoof 2017) yarışmasında SQKK özniteliğinin TO saldırısı tespitinde diğer öznitelik çıkarım yöntemlerinde üstün olduğu kanıtlanmış, SQKK yarışma komitesi tarafından referans öznitelik çıkarım yöntemi olarak ilan edilmiştir.

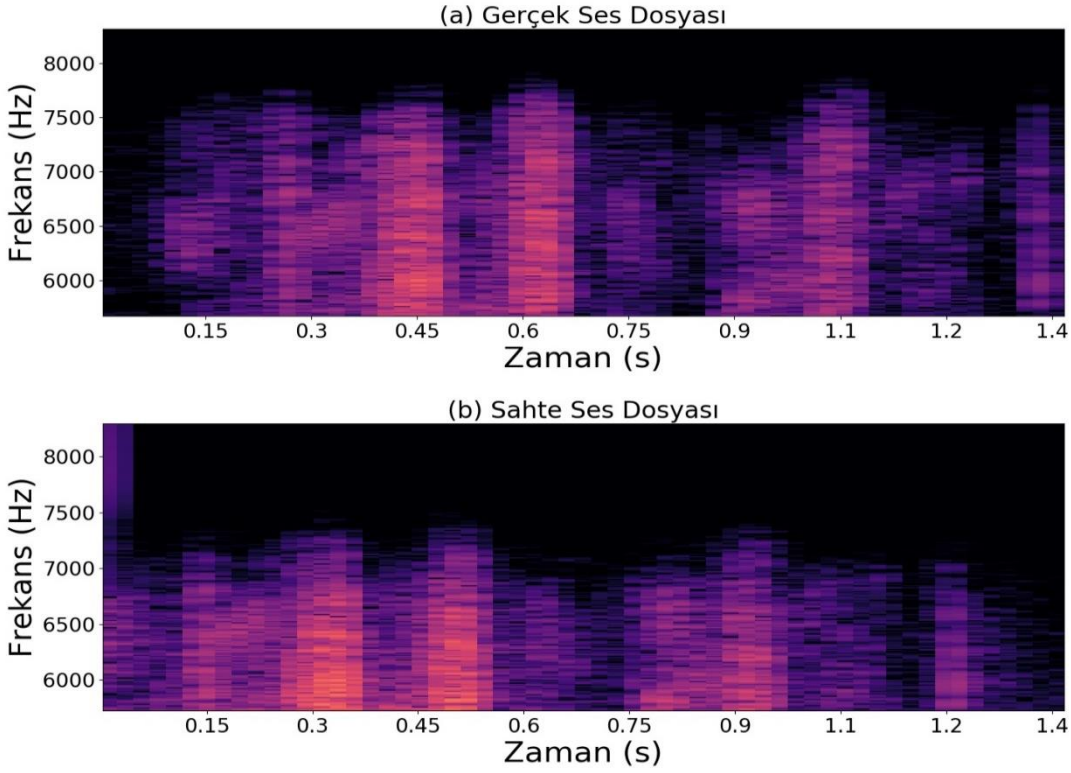
Font ve Espin tarafından 2017 yılında yapılan bir çalışmada, öznitelik çıkarma yöntemleri test edilmiş ve alt-band spektral merkez büyüklüğü katsayıları özniteliğinin iyi performans gösterdiği tespit edilmiştir [26]. Ayrıca, konuşmanın olmadığı çerçeveleri tespit ederek ses sinyalinden ayırt etmek için kullanılan ses aktivitesi tespiti yönteminin sistem performansını düşürdüğü, sessiz bölgelerin de bilgi taşıdığı gösterilmiştir. Kepstral ortalama normalizasyonunun performansı iyileştirdiği fakat ortalama ve varyans nozmalizasyonu ve kayan pencere ortalama ve varyans normalizasyonu işlemlerinin olumsuz etki yaptığı belirtilmiştir. Bahsi geçen çalışmada iki farklı veri tabanı kullanılmış ve çapraz eğitim ve doğrulama yapılmış olmasına rağmen %10,52 EHO gibi iyi bir sonuç elde edilmiştir. Söz konusu çalışmayı [26] destekler nitelikte, kepstral ortalama normalizasyon işleminin performansı iyileştirdiği bir diğer çalışmada da değinilmiştir [27]. Ortalama ve varyans normalizasyonu işleminin TO saldırısı tespit sistemi performansını düşürdüğü yapılan diğer çalışmalarda da belirtilmiştir [28, 29].

Witkowski ve arkadaşları tarafından 2017 yılında yapılan çalışmada, Şekil 1.4'de gösterilen TO saldırısı oluşturma senaryosu incelenmiştir [30].



Şekil 1.4: Gerçek ve sahte ses dosyası oluşturma senaryosu [30]

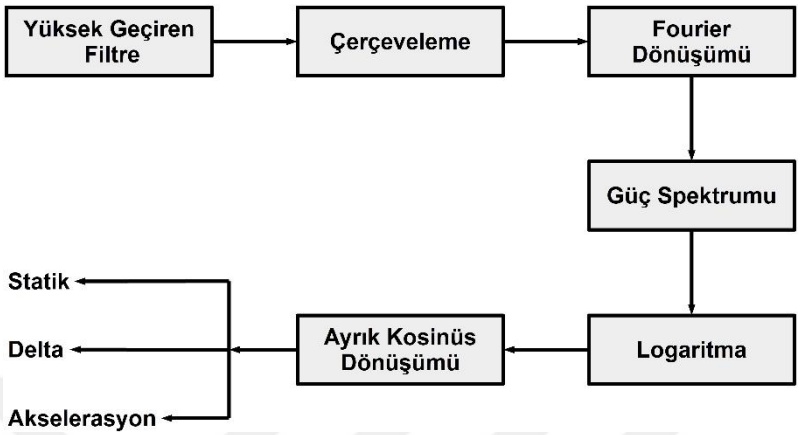
Şekil 1.4'de gösterildiği gibi, TO saldırısı oluşturma senaryosunda bir ses sinyali birden fazla analogdan dijitale ve dijitalden analoğa çevirme işlemine maruz kalmaktadır [31]. Bu durum en çok yüksek frekans bölgesini etkilediği için saldırı tespitinde yüksek frekans bölgesi analizi en önemli faktörlerden biridir [31]. Tekrar kayıt işleminin bir ses örneğine etkisi Şekil 1.5'te, 6-8 kHz bandındaki spektogramlardan gözlemlenebilir. Bahsi geçen çalışma, frekans bölgelerini alt bandlara ayırmış ve ayrı ayrı incelemiş olup, Şekil 1.4'te gösterilen çoklu analogdan dijitale çevrimlerin meydana getirdiği yapay kısımları tespit edilmesi üzerine durmuştur. Beş adet öznitelik çıkarma yöntemi ve GKM modellemesi sonucunda 4-8 kHz ve 6-8 kHz bandları kullanıldığında en iyi sonuçlar elde edilmiştir. Bu yaklaşım ASVspoof 2017 referans sistemine göre EHO'yu geliştirme verisinde %70, değerlendirme verisinde ise %30 oranında iyileştirmiştir [30].



Şekil 1.5: Gerçek (a) ve sahte (b) ses dosyalarının 6-8 kHz bandından çıkarılan spektogramları

TO saldırısı tespiti üzerine 2017 yılında yapılan bir diğer çalışmada, yeni bir öznitelik türü tanımlanmıştır [32]. Bu öznitelik yöntemi spektrumun konuşmanın olmadığı bölgedeki kanal karakteristiklerinin yakalanmasına yardımcı olan yüksek frekans kepstral katsayıları özniteliğidir. Ayrıca yüksek frekans kepstral katsayıları özniteliği, iyi performans gösterdiği bilinen sabit Q kepstral katsayıları (SQKK) özniteliği ile

birleşik olarak da kullanılmış ve iyi sonuçlar elde edilmiştir. Bu özniteliğin çıkarılma algoritması Şekil 1.6'da gösterilmiştir. Daha önceki çalışmalarda üzerinde durulan yüksek frekans bölgesi analizi, bahsi geçen çalışmada da incelenmiş olup 6-8 kHz frekans bandından çıkarılan özniteliklerin daha etkili olduğu gösterilmiştir.



Şekil 1.6: Yüksek frekans kepstral katsayıları özniteliği çıkarma algoritması [32]

1.2.3 Tekrar oynatma saldırısı tespitinde sınıflandırıcıların rolü

GKM ve YSA sınıflandırıcıları, KD sistemlerinde yaygın kullanılan sınıflandırıcılar olarak bilinmektedir. Bunlardan en yaygın kullanılan ve güvenilirliği yüksek olarak bilinen GKM yöntemidir [31, 33, 34]. GKM, ASVspoof 2017 yarışma komitesi tarafından referans sınıflandırıcı olarak ilan edilmiştir.

GKM yönteminin, TO saldırısının tespitindeki eksik yönlerine ise yapılan bir çalışmada değinilmiştir [29]. GKM'nin bazı dezavantajları ve YSA'nın çeşitli alanlarda başarısının kanıtlanması, çoğu araştırmacıyı YSA yaklaşımına yöneltmiştir veya araştırmalar gelecek çalışmalarında YSA sınıflandırıcısına daha fazla yer verileceğini belirtmişlerdir [29, 31].

GKM modellemesinin eksik yanlarından birisi 2017 yılında yapılan çalışmada belirtilmiştir [32]. Bahsi geçen bu çalışmada, GKM sınıflandırıcısının özniteliklerin Gauss karışımı olarak dağıldığını varsaydığı fakat pratikte bunun her zaman geçerli olmadığı iddia edilmiştir. Bunun yerine karmaşık sistemleri modellemede iyi olduğu kanıtlanan YSA yaklaşımının kullanılması önerilmiştir [32]. YSA'nın bir diğer avantajı ise YSA'nın bir kez eğitildiğinde karar verme ve skor hesaplama gibi işlemlerin GKM'ye göre çok daha hızlı yapılabilmesidir. Witkowski ve arkadaşları ise YSA yaklaşımının yeni yeni kullanılmaya başlandığını ve iyi sonuçlar vermesine rağmen güvenilirliğinin halen şüpheli olduğunu belirtmişlerdir [30].

KD sistemleri için, YSA'nın hem sınıflandırıcı hem de hem de öznitelik çıkarma olarak kullanıldığı çalışmalar da literatürde mevcuttur [34, 35]. Yapılan çalışmalar incelendiğinde görülmektedir ki; YSA öznitelik çıkarımı olarak kullanılıp, GKM'nin sınıflandırıcı olarak kullanıldığı sistemlerin [34] yanında, çeşitli YSA yöntemlerinin hem öznitelik çıkarma hem de sınıflandırıcı olarak birleştirildiği sistemler de bulunmaktadır [35].

TO saldırısı tespitinde YSA'nın kullanımını öneren bir çalışma 2017 yılında yapılmıştır [32]. Çalışmada YSA'yı eğitirken iki yöntem düşünülmüştür. Birincisi, YSA'yı ikili sınıflandırıcı olarak kullanıp gerçek ve sahte sesin birbirinden ayırt edilmesidir. İkicisi ise çok sınıflı sınıflandırıcı kullanıp, ses kayıtlarının kayıt, tekrar ve ortam koşulları olmak üzere üç kanal durumunun ayrımı için kullanılmasıdır. Deneysel olarak, ikinci yaklaşımın bilinen ve bilinmeyen verilerde daha iyi performans verdiği gözlemlenmiş ve ikinci yöntem tercih edilmiştir. Belirlenen YSA yapısı üç adet konvolüsyonel katmandan oluşan ve bir adet maksimum havuzlama katmanı ve üç adet tam bağlı katmandan oluşmaktadır. Bahsi geçen çalışma YSA'nın GKM'ye göre birçok avantajının olduğunu göstermiştir [32].

Konvolüsyonel sinir ağları (KSA) yaklaşımının TO saldırılarının tespitinde etkinliği Lavrentyeva ve arkadaşları tarafından araştırılmıştır [29]. KSA girdisi olarak sabit Q dönüşümünden elde edilen normalize logaritmik güç spektrumu ve hızlı Fourier dönüşümü (HFD) akustik özniteliklerini kullanılmıştır. Birleşik zaman-frekans şeklindeki öznitelikleri elde etmek için iki teknik düşünülmüştür. Bunlardan ilki, zaman ekseni boyunca spektrumu kırpmaktır. Bu yöntem sırasında gerekli uzunluğa ulaşmak için kısa dosyalar genişletilmiştir. İkincisi ise, sabit pencere sayılı kayan pencere yaklaşımı kullanmaktır. Kayan pencere yaklaşımı, budama yöntemine göre kötü sonuç göstermiştir. Bunun sebebi olarak budama yönteminde çoğu zaman her bir sesin bütün kısmının kullanılması olduğu belirtilmiştir. Tekrarlayan sinir ağları ile KSA kombinasyonu ise yalnızca hafif konvolüsyonel sinir ağları kullanıldığı duruma göre daha kötü performans vermiştir. Performans düşümüne getirilen açıklama ise; spektrum yaklaşımında frekans çözünürlüğünün azalmasıdır [29]. Çalışmanın tekrar saldırısı tespit probleminde sunduğu temel sistem: KSA+HFD+DVM+i-vektör ve KSA+HFD+geri beslemeli sinir ağları sistemlerinin skor birleşimidir. Önerilen sistem geliştirme kümesinde %3,95 ve değerlendirme kümesinde %6,73 EHO üretmiştir. Bu EHO değeri literatürde yer alan en iyi değerler arasındadır.

1.3 Hipotez

Bu tez çalışmasında KD sistemlerinin güvenirliğini ciddi bir şekilde tehdit ettiği yapılan çalışmalarla kanıtlanan ve mevcut saldırı türleri arasında en etkilisi olan TO saldırıları ele alınacaktır. Mel-frekansı kepstrum katsayıları (MFKK), sabit Q kepstral katsayıları (SQKK) ve uzun dönem ortalama spektrum (UDOS) yöntemi gibi literatürde çalışmaları mevcut öznitelik yöntemleri kullanılarak öznitelikler çıkarılacaktır. Daha sonra bu özniteliklerin sınıflandırma ve karar işlemlerinin yapılmasıyla, TO saldırısının tespiti çalışmalarına katkıda bulunulması ana hedeftir. Çalışmanın TO saldırısı tespitine yaklaşımı YSA kullanılmasıdır. YSA resim, ses, yazı gibi verilerin daha anlamlı bir hale getirilmesine ve yorumlanmasına yardımcı olan çok katmanlı gösterim ve soyutlama algoritmasıdır. KD sistemlerinde YSA yeni yeni kullanılmaya başlanmıştır. Bu çalışmanın hipotezi YSA'yı TO saldırısı tespitinde kullanmaktır.

2. MATERYAL VE YÖNTEM

KD sistemlerine yapılması olası TO saldırısının tespiti, bir sesin gerçek, orijinal ses veya sahte ses diye adlandırılan daha önceden kaydedilip tekrardan oynatılan ses kaydı olup olmadığına karar verildiği, iki sınıflı bir örüntü tanıma problemidir.

Konuşmacıdan alınan ses örneklerinden benzersiz öznitelikleri çıkarma, bu özniteliklerle bir model oluşturma ve karar verme bu saldırıların önlenmesindeki üç temel aşamadır. Bu işlemler için öncelikli olarak herkes tarafından kabul görmüş, üzerinde çalışmalar yapılmış ve kayıt ortamı, kayıt cihazı vb. ortam koşullarının sağlandığı bir veri tabanı gerekmektedir. Bu şartları sağlayan ve erişime açık bir veri tabanı literatürde yapılan çalışmalar ile karşılaştırma imkânı sunacak ve karşı önlemlerin güvenilirliğini belirgin hale getirecektir.

Bu tez çalışmasında, TO saldırılarının tespitine özel olarak düzenlenen ASVspoof 2017 yarışması için hazırlanan ve yarışma ile aynı ismi taşıyan ASVspoof 2017 veri tabanı kullanılmıştır. Ayrıca, ASVspoof 2017 yarışması katılımcılarının çalışmaları referans alınmış, yapılan çalışmalar detaylı olarak incelenmiştir.

Saldırı tespit sistemi geliştirilirken ilk olarak, ASVspoof 2017 veri tabanındaki ses kayıtlarından mel-frekansı kepstrum katsayıları (MFKK), uzun dönem ortalama spektrum (UDOS) ve sabit Q kepstral katsayıları (SQKK) yöntemleriyle öznitelikler elde edilmiştir. Daha sonra bu öznitelikler kullanılarak, YSA ve GKM yöntemleriyle modeller oluşturulmuştur. Karar aşamasında ise logaritmik olabilirlik oranı referans alınmıştır.

Oluşturulan saldırı tespit sisteminin performansının belirlenmesi ve diğer çalışmalarla karşılaştırılması için ise, bir performans ölçüm kriteri gerekmektedir. Bu çalışmada, KD sistemleri üzerine yapılan çalışmalar için genel kabul görmüş EHO metriği ile birlikte sezim hata ödünleşimi (SHÖ) eğrileri kullanılmıştır. Bu bölümde kullanılan veri tabanının özellikleri, öznitelik çıkarma ve model oluşturma yöntemleri, karar aşaması ve performans metriği detaylı bir şekilde açıklanacaktır.

2.1 Veri Tabanı

KD sistemlerinde, TO saldırılarının tespiti amacıyla yapılan çalışmaların karşılaştırılabilmesi ve karşı önlemlerin iyileştirilebilmesi için ortak bir veri tabanı gereklidir. 2015 yılına kadar, bu gereksinimi sağlayacak bir veri tabanı oluşturulamamıştır. Bu kapsamda KD sistemlerine yapılan çeşitli saldırıların önlenmesi amacıyla 2015 yılında, ASVspoof 2015 yarışması düzenlenmiş ve yarışma ile aynı ismi taşıyan ASVspoof 2015 veri tabanı oluşturulmuştur [36, 37].

ASVspoof 2015 yarışmasının devamında ise, TO saldırısına özel olarak ASVspoof 2017 yarışması ve yine bu yarışmaya özel ASVspoof 2017 veri tabanı oluşturulmuştur [38, 39]. ASVspoof 2017 yarışmasının devamında ise 2017 ve 2018 yıllarında düzenlenen Interspeech konferanslarında özel oturumlar düzenlenmiş ve ASVspoof 2017 veri tabanı üzerinden birçok çalışma yapılmıştır.

Bu çalışmada yalnızca ASVspoof 2017 veri tabanında yer alan ses kayıtları kullanılmıştır. Veri tabanı 16 kHz'de örneklenmiş, 16 bit çözünürlüğünde standart wav formatında kaydedilmiş ses dosyalarından oluşmaktadır. Bu ses dosyaları eğitim, geliştirme ve değerlendirme olmak üzere birbiri ile örtüşmeyen üç alt kümeye ayrılmıştır. Çizelge 2.1'de gösterildiği gibi; 3014 adet eğitim, 1710 adet geliştirme ve 13306 adet değerlendirme alt kümelerine ait ses kayıtları mevcuttur.

Çizelge 2.1: ASVspoof 2017 veri tabanı

Alt Küme	Konuşmacı Sayısı	Gerçek Kayıt Sayısı	Sahte Kayıt Sayısı
Eğitim	10	1507	1507
Geliştirme	8	760	950
Değerlendirme	24	1298	12008

Eğitim, geliştirme ve değerlendirme alt küme verileri sırasıyla gerçek ve sahte seslere ait akustik modellerin oluşturulması, model oluşturma esnasında sistem parametrelerinin optimize edilmesi ve oluşturulan modellerin performansının genelleştirilmesi amacıyla kullanılmıştır.

Veri tabanında her bir alt kümeye ait protokol dosyası sağlanmıştır. Bu protokol dosyalarında sırasıyla dosya adı, ses türü (tekrar veya orijinal olduğu), konuşmacı kimliği, konuşulan ifade, kayıt cihazı, kayıt ortamı, sahte ses ve ilave olarak; tekrar oynatma cihazı ve tekrar kayıt cihazı bilgileri verilmiştir.

ASVspoof 2017 veri tabanının bazı düzeltmelerinin yer aldığı versiyonu ise daha sonra yayımlanmıştır [40]. Yapılan düzeltmeler boş ses kayıtlarının içeren ses dosyalarının kaldırılması ve protokol dosyalarında yer alan kayıt ortam bilgilerinin düzeltilmesi şeklindedir. Bu tez çalışması için yapılan deneyler esnasında veri tabanının güncel versiyonu kullanılmıştır.

2.2 Öznitelik Çıkarma

Ses sinyali, hava basıncındaki değişimi voltaj değişimine çevirmeye yarayan, mikrofon aracılığıyla elde edile hava partiküllerinin analog dalga formu olarak ifade edilebilir. Ses sinyali, ses işleme uygulamalarında kullanılmak için bir örnekleme periyodunda analogdan-dijitale dönüştürücüler kullanılarak, dijital verilere dönüştürülür.

Öznitelik çıkarma ise, ses sinyallerinin sınıflandırılması amacıyla alınan ölçümlerin daha özlü, gereksiz bilgilerden arındırılmış, daha az gürültülü ve genelde daha düşük boyutlu fakat ayırt edici bilgilerin vurgulandığı bir işlemdir. Bir KD sistemi için öznitelik çıkarmada genel kabul görmüş birçok yöntem kullanılmaktadır. Bunlar, ön vurgulama, çerçeveleme, pencereleme ve HFD yöntemleridir.

Bu bölümde ilk olarak bahsi geçen öznitelik çıkarma yöntemleri, daha sonra da bu yöntemler ve bunlara ilave olarak bazı yöntemlerle çıkarılan öznitelikler alt başlıklar halinde incelenecektir.

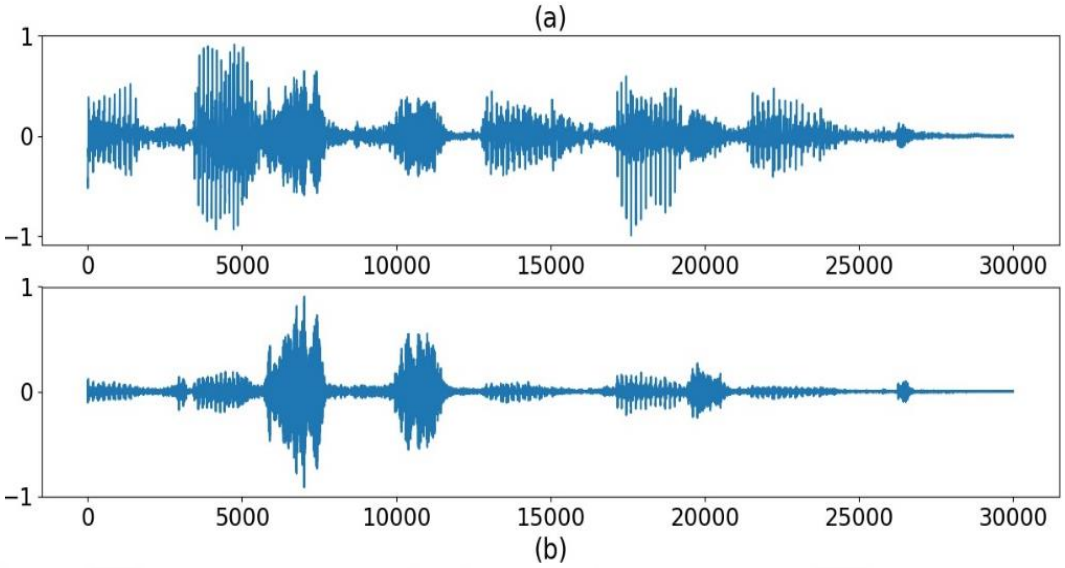
2.2.1 Ses işleme yöntemleri

2.2.1.1 Ön vurgulama

Ön vurgulama, bir sinyalin yüksek frekans bileşenlerinin baskın hale getirilmesini sağlayan ve genellikle diğer bütün işlemelerden önce uygulanan bir yöntemdir [41]. Şekil 2.1'de ses sinyalinin ön vurgulama yapılmadan önceki ve yapıldıktan sonraki davranışı gösterilmiştir.

En yaygın kullanılan ön vurgulama filtresinin transfer fonksiyonu denklem 2.1'de verilmiştir. Burada α süzgecin eğimini belirlemektedir ve bu tez çalışmasında 0,97 olarak seçilmiştir.

$$H(z) = 1 - \alpha z^{-1} \tag{2.1}$$

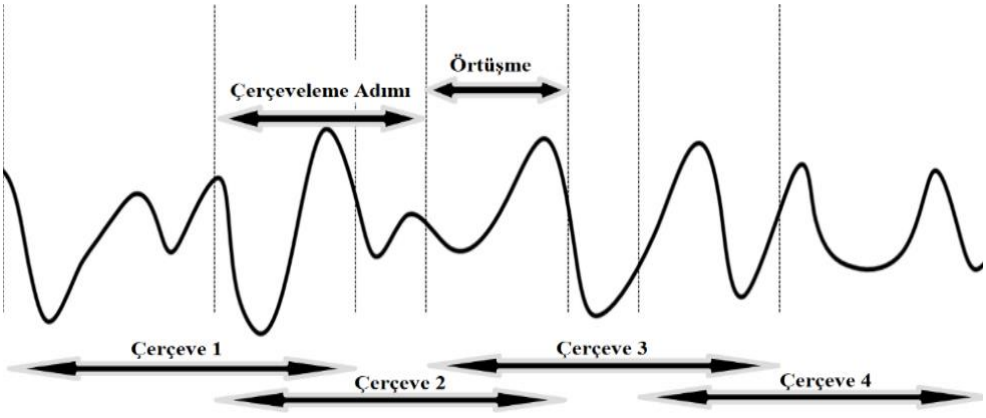


Şekil 2.1: Ses sinyali (a), ön vurgulamanın ses sinyaline etkisi (b)

2.2.1.2 Çerçeveleme

Sinyal işlemede, bir sinyalin durağan olması önemli bir avantaj sağlamaktadır. Gırtlak yapısındaki ani değişimlerden dolayı ses sinyali durağan değildir. Ancak çerçeveleme diye adlandırılan basit bir işlem ile sinyal durağan hale getirilmektedir. Çerçeveleme işlemi genel olarak, Şekil 2.2'de gösterildiği gibi M adet örnekten oluşan sinyalin N örnekli kısımları örtüşecek şekilde sinyalin birbiri üzerine eklenmesi şeklinde ifade edilir. Çerçeveleme işleminde çerçeve uzunluğu her durum için örtüşme uzunluğundan büyük olmalıdır (N<M). Şekil 2.2'de bir ses dört adet çerçeveye ayrılma işlemi gösterilmiştir.

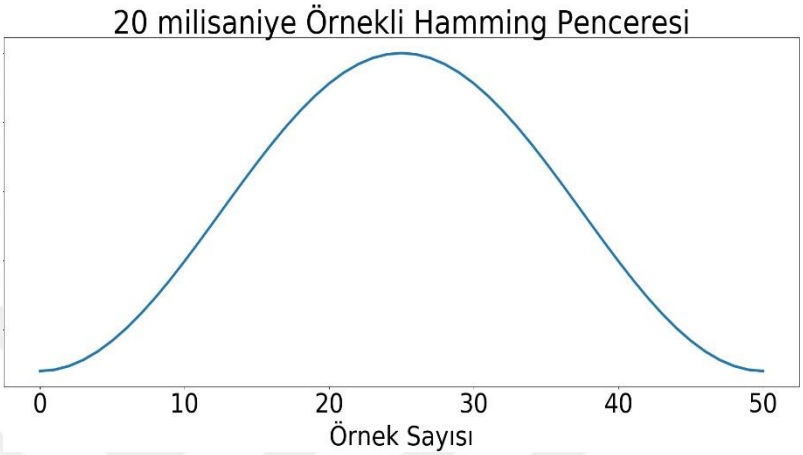
Bu tez çalışmasında kullanılan bütün çerçeveleme işlemleri için çerçeve uzunluğu 20 milisaniye, örtüşme uzunluğu ise 10 milisaniye olarak seçilmiştir.



Şekil 2.2: Dört adet çerçeveye bölünmüş ses sinyali

2.2.1.3 Pencereleme

Pencereleme işlemi, sinyalin barındırdığı bilgi içermeyen bölgeleri bastırarak spektral bozulmaları engellemek amacıyla kullanılmaktadır. Bu çalışmada en yaygın kullanılan pencere fonksiyonlarından olan Hamming pencere fonksiyonu kullanılmıştır. Hamming pencere fonksiyonunun grafiği Şekil 2.3'de verilmiştir.

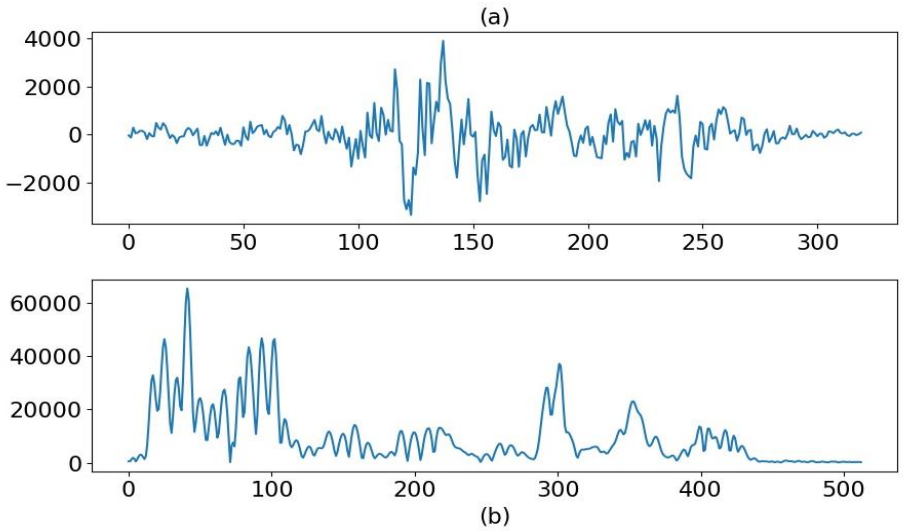


Şekil 2.3: 320 örnek (20 milisaniye) uzunluğunda Hamming pencere fonksiyonu

2.2.1.4 Hızlı Fourier dönüşümü

Bir sinyalin zaman uzayından, frekans uzayına çevirmek için AFD kullanılır. *N* adet örnekten meydana gelen bir *s[n], n=0, 1, 2…, N-1* ayrık zamanlı sinyalinin AFD'si denklem 2.2'de gösterilmiştir.

$$X_k = \sum_{n=0}^{N-1} x_n \, e^{-i2\pi kn/N} \quad k = 0, 1, 2 \ldots, N-1 \tag{2.1}$$



Şekil 2.4: Konuşma çerçevesi (a) ve konuşma çerçevesinin genlik spektrumu (b)

Bu dönüşümün daha hızlı olması için geliştirilen algoritma hızlı Fourier dönüşümü (HFD) olarak bilinmektedir. Şekil 2.4'de bir konuşma çerçevesi ve çerçevenin HFD'si alınarak hesaplanan genlik spektrumu gösterilmiştir.

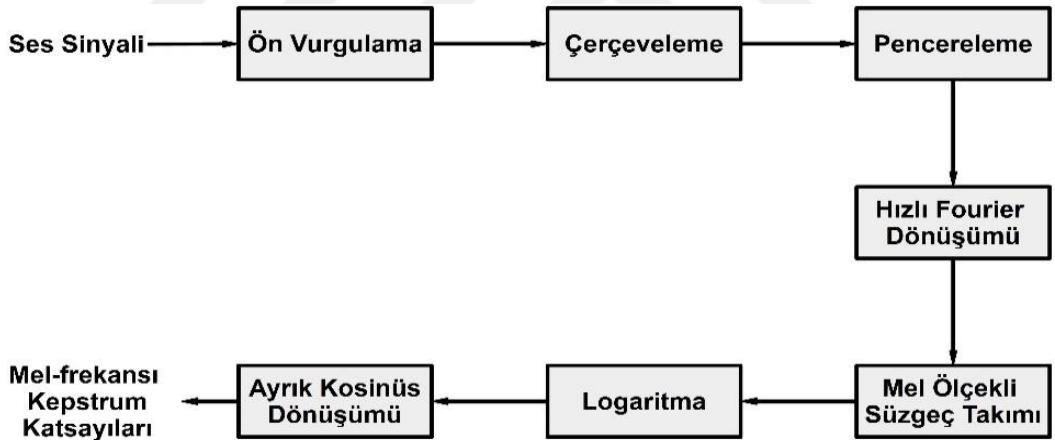
2.2.2 Öznitelikler

Bölüm 2.2.1'de anlatılan ses işleme yöntemleri ve genellikle ilave metotlar da kullanılarak çeşitli öznitelikler elde edilmektedir. Bu tez çalışmada kullanılan öznitelik yöntemleri bu kısımda alt başlıklar halinde açıklanacaktır.

2.2.2.1 Mel-frekansı kepstrum katsayıları

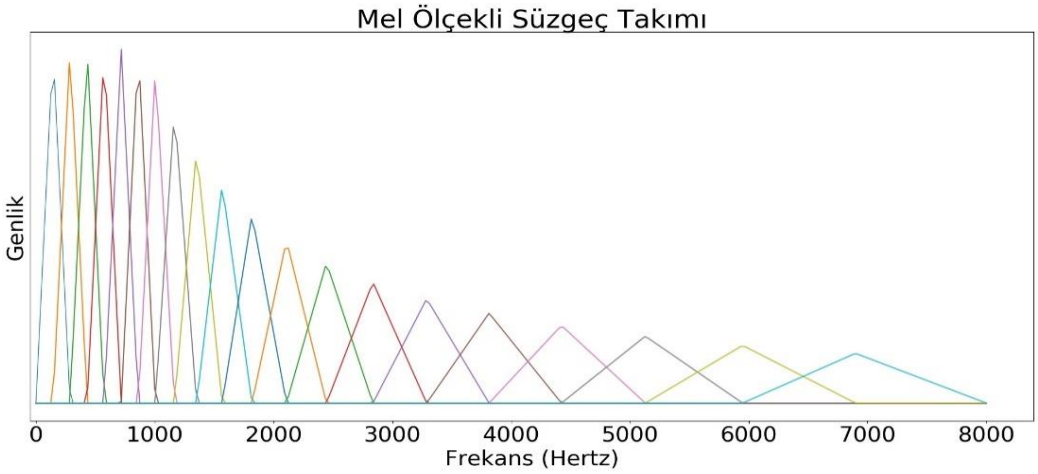
Mel-frekansı kepstrum katsayıları (MFKK) [42], konuşma veya konuşmacı analizi problemlerinde en yaygın kullanılan öznitelik çıkarma yöntemlerinden biridir.

MFKK özniteliklerini çıkarmak için Şekil 2.3'de gösterildiği gibi, ilk olarak ses sinyali ön vurgulama işleminden geçirilir ve daha sonra her biri 20 milisaniye uzunluğunda birbiri ile örtüşmeyen çerçevelere ayrılıp Hamming pencere fonksiyonu ile çarpılır.



Şekil 2.5: Mel-frekansı kepstrum katsayıları özniteliklerinin çıkarım algoritması

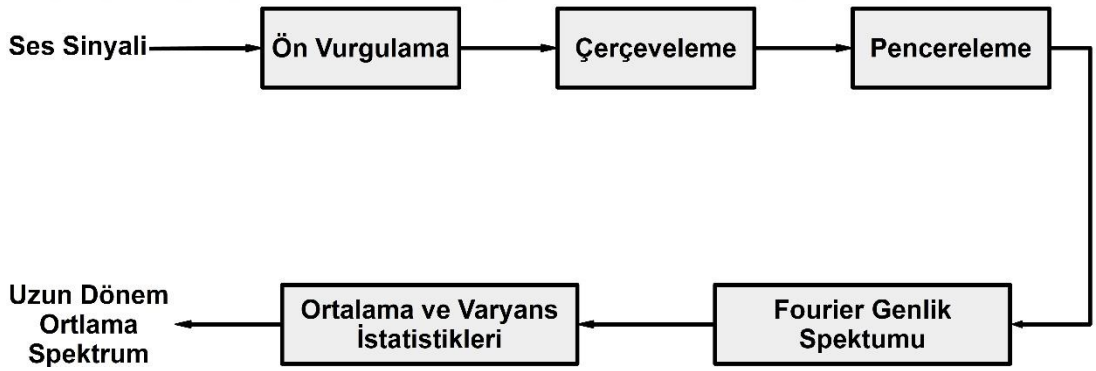
Pencereleme işlemi uygulanan çerçevelerin AFD'si alınarak genlik spektrumları elde edilir ve Şekil 2.6'da gösterilen üçgen süzgeçlerden oluşan süzgeç takımından geçirilir. Daha sonra, Logaritmik süzgeç çıkışlarının ayrık kosinüs dönüşümünün alınmasıyla MFKK öznitelikleri elde edilir. MFKK öznitelikleri yüksek frekansları bastırıp, alçak frekansları ön plana çıkarmaktadır. Bu durum TO saldırısı tespitinde faydalı bir sonuç elde etmeyi engellemektedir.



Şekil 2.6: Mel-ölçekli süzgeç takımı

2.2.2.2 Uzun dönem ortalama spektrum

Uzun dönem ortalama spektrum (UDOS), iki aşamadan oluşan bir öznitelik çıkarım yöntemidir [19]. Bu aşamalar, Fourier genlik spektrumunun hesaplanması ve Fourier frekans bileşenlerinin birince ve ikinci derece istatistiklerinin tek bir vektörde birleştirilmesidir. Fourier genlik spektrumunu hesaplamak için ilk olarak, Şekil 2.7'de gösterildiği gibi *s* ses sinyali ön vurgulama süzgecinden geçirilip, çerçeveleme ve pencereleme işlemelerine tabi tutulur.



Şekil 2.7: Uzun dönem ortalama spektrum özniteliği çıkarım algoritması

Daha sonra sinyalin N-noktalı AFD'si ve genlik spektrumu hesaplanır. Bir m, m ⊂ {1...M}: çerçevesi için bu işlem denklem 2.2'deki gibi ifade edilir.

$$S_m[k] = AFD(s_m[n]), \qquad n = 0, 1, 2 \ldots, N-1 \tag{2.2}$$

Burada çerçeve uzunluğu *w* olmak üzere $N = 2^{[log2(w)]}$ ve $k = 0 \ldots \frac{N}{2} - 1$ şeklindedir; çünkü sinyal $\frac{N}{2}$ etrafında simetriktir. Bu işlem her bir *m* çerçevesi için denklem 2.3'de gösterildiği gibi devam etmektedir.

$$S_m = [S_m[0] \ldots S_m[k] \ldots S_m[\frac{N}{2} - 1]]^T \quad (2.3)$$

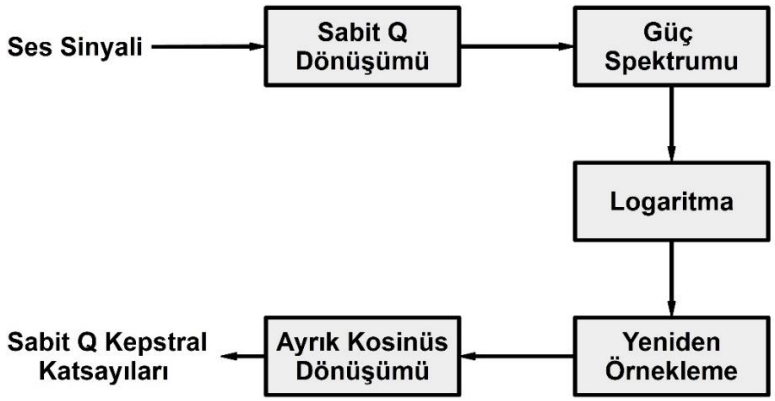
şeklinde devam etmektedir. İkinci aşama olarak, AFD vektörlerinin ($X_1 \ldots X_m \ldots X_M$) denklem 2.4'de gösterildiği gibi ortalaması ve denklem 2.5'de gösterildiği gibi standart sapması hesaplanır. Son olarak ortalama ve varyans istatistikleri birleştirilerek, UDOS öznitelik vektörleri elde edilir.

$$\mu[k] = \frac{1}{M} \sum_{m=1}^{M} log\,|X_m[k]| \quad (2.4)$$

$$\sigma^2 = \frac{1}{M} \sum_{m=1}^{M} (log|X_m[k] - \mu[k]|)^2 \quad (2.5)$$

2.2.2.3 Sabit Q kepstrum katsayıları

SQKK özniteliği, sabit Q dönüşümü (SQD) ve kepstral analizin bir kombinasyonu olan genlik tabanlı bir özniteliktir [21]. SQKK şekil 2.8'de gösterildiği gibi, pencereleme işlemi uygulanmış bir sinyalin güç spektrumundan elde edilmektedir. Burada standart kısa-dönem Fourier dönüşümü, güç spektrumu AFD yerine SQD kullanılarak hesaplanmaktadır.



Şekil 2.8: Sabit Q kepstral katsayıları özniteliği çıkarım algoritması

Bunun sebebi, AFD'nin sabit zaman ve çözünürlüğü sağlarken, SQD yüksek frekanslar için daha detaylı zaman çözünürlüğü, düşük frekanslar için ise daha detaylı frekans çözünürlüğü sağlamasıdır. Daha sonra logaritması alınan SQD yeniden örneklenip ayrık kosinüs dönüşümü alınarak SQKK öznitelikleri elde edilmiş olur. Ayrıca, SQKK ASVspoof 2017 yarışmasının referans öznitelik çıkarma yöntemi olarak önerilmiştir.

2.3 Sınıflandırıcılar

2.3.1 Gauss Karışım Modeli

GKM, KD sistemlerinde kullanılan en eski ve en güvenilir sınıflandırıcılardan biri olarak bilinmektedir [43]. Bu tez çalışmasında YSA yöntemi ile oluşturulan sistemlerin performansının karşılaştırılabilir olması amacı ile, benzer sistemler GKM yöntemi kullanılarak da oluşturulmuş ve karşılaştırılmalı analizi yapılmıştır.

GKM yönteminde, gerçek ve tekrar örüntü sınıfları denklem 2.6'da gösterildiği gibi *M* adet çok boyutlu Gauss yoğunluk fonksiyonunun ağırlıklandırılmış toplamı şeklinde ifade edilir.

$$p(\boldsymbol{X}|\lambda) = \sum_{i=1}^{M} w_i p_i(\boldsymbol{X}) \tag{2.6}$$

Denklem 2.6'daki $p_i(x)$ ifadesi, denklem 2,7'de ifade edildiği gibi *D* değişkenli, ortalaması μ_i ve ortak değişim matrisi $\sum_i$ olan, D öznitelik vektör boyutlu Gauss yoğunluk fonksiyonunu belirtmektedir. Bütün bir GKM denklemi ise denklem 2.8'de gösterildiği gibidir.

$$p_i(\boldsymbol{X}) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} e^{(-\frac{1}{2}(x-\mu_i)'\Sigma_i(x-\mu_i))} \tag{2.7}$$

$$\lambda = \{w_i, u_{i,} \sum i\}_{i=1}^{M} \tag{2.8}$$

GKM yönteminin eğitim aşamasında her bir sınıfın öznitelik vektörleri $X = \{x_1, x_2 \ldots, x_N\}$ eğitim öznitelikleri kullanılarak en büyük olabilirlik kriteri kullanılarak beklentinin maksimumlaştırılması algoritması ile GKM parametreleri (ağırlık, ortalama ve ortak değişinti) tahmin edilir.

Test aşamasında ise; test ses sinyallerinden elde dilen öznitelik vektörleri Y = {y1, y2, $y_3 \ldots y_T$} kullanılarak logaritmik olabilirlik oranı skoru denklem 2.9'da gösterildiği gibi hesaplanır.

$$LLR = log(Y|\lambda_{gerçek}) - log(Y|\lambda_{sahte}) \tag{2.9}$$

Burada $\lambda_{gerçek}$ ve λ_{sahte} sırası ile gerçek ve tekrar sınıflarına ait GKM modellerini temsil etmektedir.

Test öznitelik vektörleri için olabilirlik oranı ise denklem 2.10'da gösterildiği gibidir.

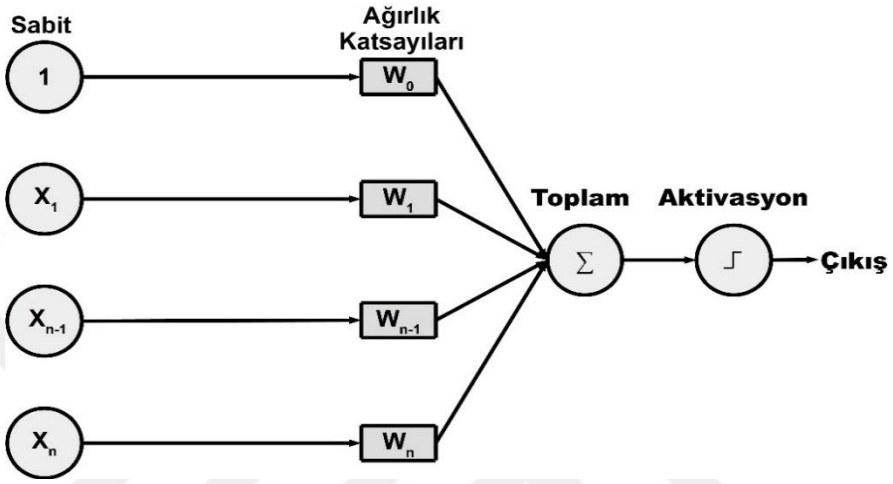
$$p(y|\lambda) = \prod_{t=1}^{T} \sum_{i=1}^{m} w_i p_i(y_i|\lambda) \tag{2.10}$$

2.3.2 Yapay Sinir Ağları

Yapay sinir ağları, makine öğrenmesinin temel hedeflerinden olan yapay zekayı gerçekleştirmeyi hedefleyen makine öğrenmesi araştırma alanıdır [45]. Daha detaylı olarak; resim, ses, yazı gibi verilerin daha anlamlı bir hale getirilmesi ve yorumlanmasına yardımcı olan çok katmanlı gösterim ve soyutlama algoritmasıdır. YSA çalışmaları yapılırken birçok programlama dili, geliştirme ortamı, kütüphaneler ve donanımlardan faydalanılmaktadır. Python, zamanla derin öğrenme çalışmalarında en çok kullanılan programlama dili olmuştur [46]. Haliyle python dili ile derin öğrenme için geliştirilen birçok derin öğrenme kütüphanesi geliştirilmiştir. Bunlardan bazıları theano, tensorflow ve bunları kullanarak daha üst seviye bir ara-yüz sağlayan keras modülleridir. Bu modüller hem merkezi işlem birimi (CPU) hem de grafik işlem birimi (GPU) üzerinden çalışabilen, hızlı sayısal hesaplamalar sağlamaktadır. CPU'lar daha genel hesaplama işlemleri için tasarlanmış yapılardır. GPU'lar ise daha az esnek yapıya sahip olmakla birlikte, CPU'nun yapacağı aynı işlemleri paralel olarak hesaplayabilmektedir. YSA, özdeş yapay nöronların binlerce ağındaki her bir katmanlar aynı hesaplamayı yapacak şekilde homojen bir yapıdadırlar. İşte bu yapı GPU kullanımına tam olarak uygundur. GPU'lar CPU'lara göre yavaştır fakat, GPU paralel hesaplama mimarisinden dolayı tek seferde okunan ve işlenen veri CPU'nun hızıyla kıyaslanamaz [47]. Sonuç olarak makina öğrenmesi çalışmaları için GPU'ların CPU'lara göre birçok avantajı vardır. Bunlardan en öne çıkanları daha fazla hesaplama birimine ve hafızadan veri okumak için yüksek bir band genişliğine sahip olmaktır. Bunun yanında GPU'ların daha düşük kapasitede rasgele erişilebilir hafızaya (RAM) sahip olması, saat hızının düşük olması ve GPU kullanımı için CPU'ya ihtiyaç duyulması ise zayıflıkları arasında gösterilebilir. Sonuç olarak, ileri seviye derin öğrenme çalışmalarında GPU kullanımı zorunludur. Bu çalışmada kapsamında matematiksel operasyonları ve çeşitli sinyal işleme yöntemlerini sağlayan numpy, scipy gibi python ve derin öğrenme deneylerini gerçekleştirmek için keras [48] modülü

kullanılmıştır. Bilgisayar donanımı ise, Intel i7 CPU ve NVIDIA 1080 TI GPU'dan oluşmaktadır.

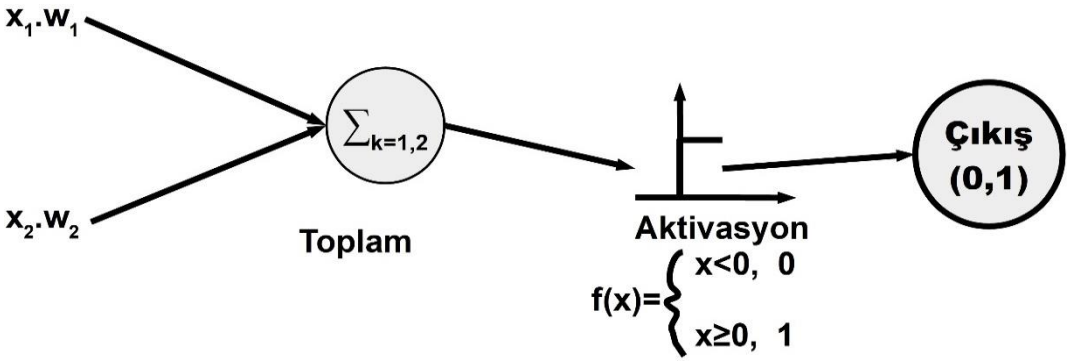
Bir YSA'nın en temel elemanı Şekil 2.8'de gösterilen algılayıcı yapısıdır. Algılayıcı, ikili sınıflandırıcıların denetimli öğrenimi için geliştirilmiş bir algoritmadır. Bu ikili sınıflandırıcı, sayı vektörleri şeklinde temsil edilen bir girdinin belirli bir sınıfa ait olup olmadığına karar vermektedir.



Şekil 2.8: Algılayıcı yapısı

Diğer bir ifade ile algılayıcı, öznitelik vektörleri ile ağırlık katsayılarını birleştiren ve doğrusal belirleyici bir fonksiyona dayanan bir lineer sınıflandırma algoritmasıdır. Bir algılayıcı öznitelik vektörü, ağırlık katsayıları, bias sabiti, toplama ağı ve aktivasyon fonksiyonundan meydana gelir. Algılayıcı algoritmasında ilk olarak, Şekil 2.9'da gösterildiği gibi her bir girdi ($X_1 \ldots X_{n-1} \ldots X_n$,) ağırlık katsayıları ile çarpılır.

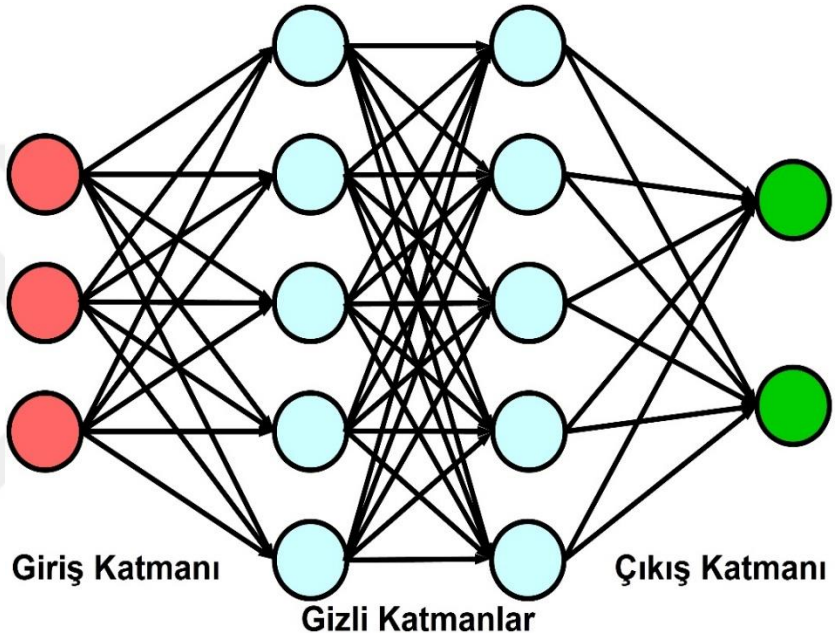
Daha sonra çarpımdan elde edilen sonuçlar toplanır ve buna ağırlıklandırılmış toplam denir. Son olarak aktivasyon fonksiyonu uygulanır ve karar aşamasına geçilir. Karar aşamasında aktivasyon fonksiyonu eşik değerine göre çıkış 0 veya 1 olur.



Şekil 2.9: Algılayıcı algoritması

Birden fazla algılayıcının bir araya gelerek oluşturduğu yapıya ise çok katmanlı algılayıcı denir. Çok katmanlı algılayıcı, girdilerin lineer olmayan bir aktivasyon fonksiyonuna sahip olan bir katmana verilmesidir [45]. Bu aktivasyon fonksiyonun genellikle tanh, sigmoid veya relu fonksiyonlarıdır. Katman sayısının birden fazla olması ise derin olarak adlandırılmaktadır.

Bu tez çalışmasında, Şekil 2.10'da gösterildiği gibi çok katmanlı algılayıcıların birbirine tam olarak bağlanmış, ileri yönlü, relu aktivasyon fonksiyonların her bir katmanın çıkışına yerleştirildiği bir derin sinir ağı yapısı kullanılmıştır.



Şekil 2.10: Çok katmanlı algılayıcı

YSA yapısı gerçek ve tekrar seslerini birbirinden ayırt etmek amacıyla eğitilmiş bir sınıflandırıcı olarak kullanılmıştır. YSA'nın giriş katmanındaki nöron sayısı, kullanılan özniteliklerin boyutu ile aynı olup, gizli katmanlarda farklı sayılarda nöron kullanılmıştır. KD sistemler için tekrar saldırı tespiti ikili sınıflandırma problemi olduğundan, çıkış katmanlarının nöron sayısı iki ve aktivasyon fonksiyonu softmax olarak belirlenmiştir. İki adet çıkış birimi sırası ile gerçek ve sahte sınıflara karşılık gelmektedir.

Her bir çıkış nöronu, ilgili sınıfın sonsal olasılığını temsil ettiğinden sonsal olasılıklar, logaritmik olabilirlik oranı skoruna denklem 2.11'de gösterildiği gibi dönüştürülmüştür.

$$LLR = log_{\rho}(C_1|x) - log_{p}(C_2|x) \tag{2.11}$$

YSA yapısında, hata fonksiyonunu optimize etmek için Stochastic Gradient Descent algoritması kullanılmıştır. Her öznitelik türü için kullanılan farklı sayılarda nöron ve gizli katman içeren YSA sınıflandırıcısı ile yapılan deneylerle ideal parametreler tespit edilmiştir. MFKK, SQKK ve UDOS öznitelik vektörlerinin boyutları, karakteristikleri yani davranışları benzer olmadığından, öznitelikler için tespit edilen ideal YSA yapıları farklıdır. Çizelge 2.2'de her bir özniteliğe ait YSA yapısı gösterilmiştir.

Çizelge 2.2: Yapay sinir ağları yapısı

	Girdi Boyutu	Gizli Katman Sayısı	Gizli Katman Birim Boyutu	Dropout Değeri
SQKK	18	3	256	0.2
MFKK	57	3	256	0.2
UDOS	514	5	1024	0.5

Deneyler sırasında aşırı eğitimi (over training) engellemek amacıyla, öğrenme oranı learning rate) düşük seçilmiş ve her bir katmandan dropout eklenmiştir [49, 50]. Öğrenme oranının düşük olması nedeniyle en iyi modele ulaşmak için devir (epoch) sayısı yüksek (10000) seçilmiştir. Diğer parametrelerin sisteme etkisini azaltmak ve hafıza problemlerini aşmak için; eğitim verisi gruplara ayrıştırılarak eğitilmiştir (batch training) [51]. Eğitilen modelin, geliştirme verisi kullanılarak belirli bir frekansta validasyonu yapılmış ve validasyon skoru bir önceki sonuçtan iyi olduğu tespit edilen model en iyi model olarak kaydedilmiştir. Eğitim zamanından tasarruf amacı ile, validasyon skoru azalmadığında veya artmaya başladığında 10000 epoch sayısı beklenmeden eğitim işlemi sonlandırılmıştır (early stopping).

2.4 Performans Kriteri

KD sistemlerde TO saldırıları için sistem geliştirme aşamasında iki tür sınama yöntemi vardır. Bunlardan biri ses örneğinin kimlik iddiasında bulunan kişiye ait olduğu yani gerçek/orijinal seslerin kullanıldığı gerçek sınama, diğeri ise sistemi yanıltmaya çalışan kişinin sahte/yapay ses örneklerini kullandığı sahte sınamadır.

KD sistemlere yanlış kişinin reddedilmesi (yanlış ret) ve yanlış kişinin kabul edilmesi (yanlış kabul) şeklinde iki hata oluşabilir. Yanlış kabul, ses örneğinin kimlik iddiasında bulunan kişiye ait olmamasına rağmen sistemin bu kişiyi kabul etmesidir. Yanlış ret ise, ses örneğinin kimlik iddiasında bulunan kişiye ait olmasına rağmen

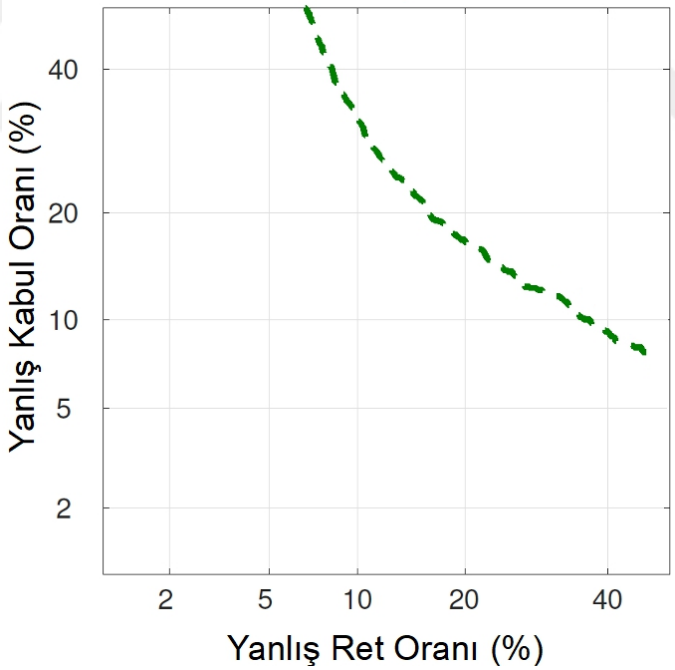
sistemin kişiyi reddetmesidir. Yanlış kabul ve yanlış ret oranları denklem 2.3 ve 2.4'de gösterildiği gibi hesaplanmaktadır.

$$\text{Yanlış Kabul Oranı} = \frac{\text{Kabul Edilen Yanlış Sınama Sayısı}}{\text{Toplam Yanlış Sınama Sayısı}} X\ 100 \tag{2.12}$$

$$\text{Yanlış Red Oranı} = \frac{\text{Kabul Edilen Yanlış Red Sayısı}}{\text{Toplam Doğru Sınama Sayısı}} X\ 100 \tag{2.13}$$

Yanlış kabul ve yanlış ret oranlarına, uygulama türüne göre belirlenen eşik değere göre karar verilir. Akademik çalışmalarda yanlış kabul oranının, yanlış ret oranına eşit olduğu değere denk gelen eşit hata oranı (EHO) yöntemi kullanılmaktadır. Bu çalışma elde edilen sonuçlar EHO kullanılarak belirlenmiştir.

Ayrıca, EHO metriğine ek olarak sezim hata ödünleşimi (SHÖ) eğrileri de verilmiştir. SHÖ eğrileri, her iki hata durumun birbirine göre değişimlerinin grafiksel olarak gösterilme yöntemidir [44]. Şekil 2.5'de bir SHÖ eğrisi verilmiştir. Bu eğride yanlış kabul ve yanlış ret oranlarının birbirine eşit olduğu, siyah nokta ile vurgulanmış nokta EHO noktasıdır.



Şekil 2.9: Sezim hata ödünleşimi grafiği

3. DENEYSEL SONUÇLAR

Deneysel çalışmalar sırasında ilk olarak, ASVspoof 2017 veri tabanının eğitim kümesi kullanılarak saldırı tespit sistemleri geliştirilmiştir. Sistem geliştirme aşamasında yöntemlerin ve parametrelerinin optimize edilmesi için geliştirme kümesinden faydalanılmıştır. Daha sonra, değerlendirme kümesi verileri kullanılarak sistemlerin performansları genelleştirilmiştir.

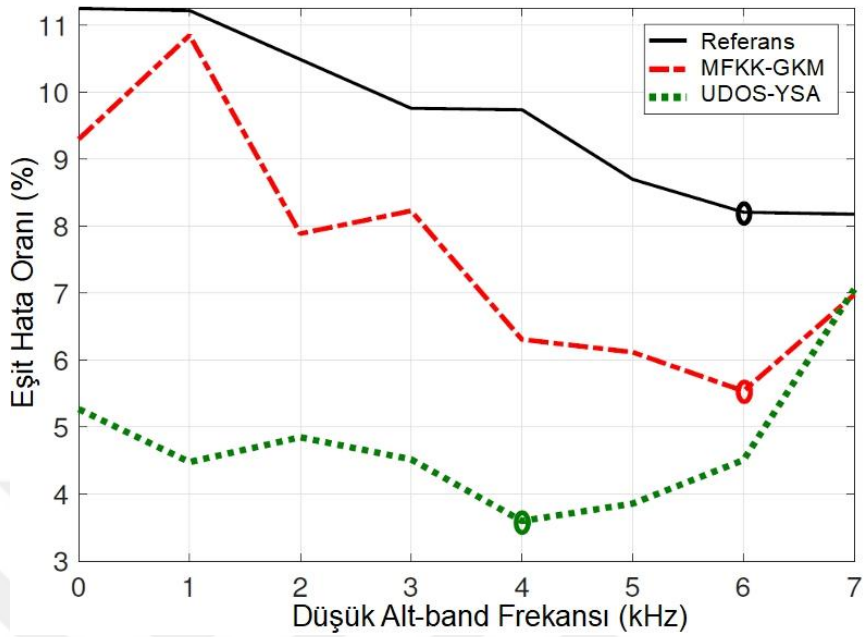
Bu kısımda ilk olarak, geliştirme kümesi üzerinden, daha önceki çalışmalarda önemi vurgulanan alt-band frekans analizi ve ortalama-varyans normalizasyonu işleminin etkisi incelenecektir. Daha sonra, alt-band frekans analizi ve normalizasyon deneylerinin sonuçları da dikkate alınarak, geliştirme kümesi üzerinden yapılan deneylere ait sistemler ve elde edilen sonuçlar EHO değerleri ve SHÖ eğrileri halinde verilecektir. Son olarak, geliştirme kümesi üzerinden yapılan deneyler sonucunda elde edilen en iyi modeller, değerlendirme kümesi kullanılarak test edilecek ve yine sonuçlar EHO değerleri ve SHÖ eğrisi halinde sunulacaktır.

3.1 Geliştirme Kümesi Sonuçları

Deneyler sırasında ilk olarak, etkili olduğu önceki çalışmalarda çeşitli öznitelikler üzerinden gösterilen alt-band frekans analizi yapılmıştır. MFKK-GKM ve UDOS-YSA sistemlerinin alt-band frekans analizi sonuçları, ASVspoof 2017 yarışmasında ilan edilen SQKK-GKM (Referans Sistem) sonuçları ile karşılaştırılmıştır. Bu karşılaştırmaya ait sonuçların grafiksel gösterimi Şekil 3.1'de verilmiştir. Frekans bölgesi analizi yapılırken, düşük frekans limiti $0 \leq f \leq 8$ kHz aralığında 0 kHz'den başlayarak 7 kHz'ye kadar artırılmıştır.

SQKK için alt-band frekans analizi, ilgili alt-band frekans değeri kesim frekansı olarak seçilen bir yüksek geçiren filtre uygulanarak yapılmıştır. Burada deneysel olarak 4. dereceden filtre daha uygun olduğu bulunmuş ve tüm kesim frekansları için bu derece kullanılmıştır. MFKK öznitelikleri için ise, mel filtrelerinin ilgili frekansa göre

değiştirilmesiyle yapılmıştır. UDOS için alt-band frekans analizi yapılırken, HFD frekans bileşenlerinin sınırlandırılması yöntemi uygulanmıştır.



Şekil 3.1: Alt-band frekans analizi sonuçları

Deneylerde göstermiştir ki, daha önceki çalışmalarda da bahsedildiği gibi belirli bir alt-band aralığından çıkarılan öznitelikler, öznitelik türünden veya sınıflandırıcıdan bağımsız olarak, 0-8 kHz frekans bandında çıkarılan özniteliklere göre daha iyi performans göstermektedir. UDOS-YSA kombinasyonu, diğer sistemlerden daha düşük EHO üretmiş ve tüm alt-band analizleri, 0-8 kHz bandından çıkarılan özniteliklere göre daha iyi performans göstermiştir. Genel olarak en iyi alt-band sonuçları 4-8 kHz aralığından elde edilmiştir. MFKK-GKM ve referans sisteme için 6-8 kHz bandı en iyi sonuç verirken, çalışmanın devamındaki bütün öznitelik çıkarma işlemelerinde ilgili alt-band frekans analizi sonuçları referans alınmıştır.

Çalışmanın devamında, ortalama-varyans normalizasyon işleminin etkisi araştırılmıştır. Çizelge 3.1'den incelenebilecek deneysel sonuçlara göre ortalama-varyans normalizasyon işlemi herhangi bir öznitelik veya sınıflandırıcı üzerinde performansı iyileştirici bir etkisi göstermemiştir. Çalışmanın bir sonraki adımında, ortalama-varyans normalizasyon işleminin etkisi araştırılmıştır. Çizelge 3.1'den incelenebilecek deneysel sonuçlara göre ortalama-varyans normalizasyon işlemi herhangi bir öznitelik veya sınıflandırıcı üzerinde performansı iyileştirici bir etkisi göstermemiştir.

Çizelge 3.1: Ortalama ve varyans normalizasyon işleminin etkisi

Sistem	EHO (%)	EHO (%)
SQKK-GKM	15,15	8,18
MFKK-GKM	13,40	5,54
SQKK-YSA	17,18	10,05
MFKK-YSA	12,51	6,64
UDOS-YSA	**6,05**	**4,1**

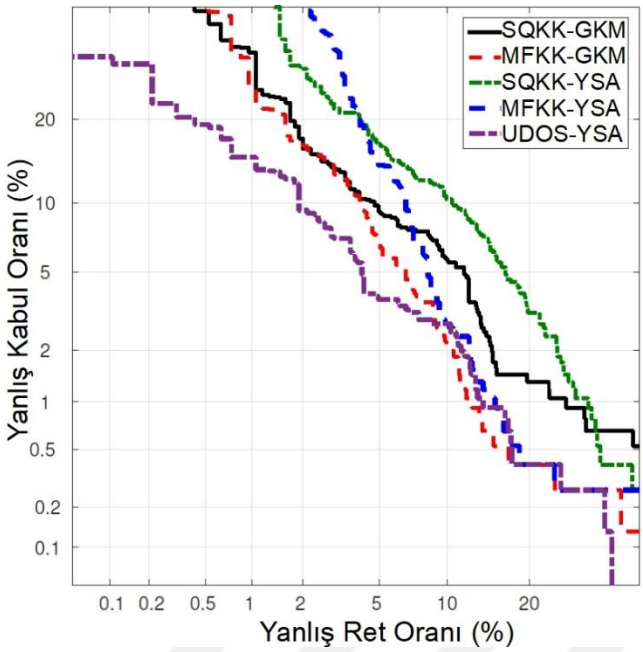
Daha sonra GKM ve YSA sınıflandırıcısının performansı incelenmiş ve deneylere ait sonuçlar, Çizelge 3.2'de verilmiştir. Burada UDOS özniteliklerine ait GKM sınıflandırıcısının sonuçlarının olmamasının nedeni, GKM sınıflandırıcısının çok boyutlu öznitelikleri vektörlerinin sınıflandırmak için uygun olduğu ve buna karşın UDSİ özniteliklerinin 514 uzunluğunda tek boyutlu bir vektörden oluşmasıdır. Sonuçlara göre, MFKK-GKM sistemi referans (SQKK-GKM) sisteme göre %48 daha iyi sonuç vermektedir; YSA sınıflandırıcısı devreye girdiğinde bu iki sistem arasındaki performans farkı daha da artmaktadır.

Çizelge 3.2: Geliştirme kümesi deney sonuçları

Sistem	EHO (%)
SQKK-GKM	8.18
MFKK-GKM	5.54
SQKK-YSA	10.05
MFKK-YSA	6.64
UDOS-YSA	**4.1**

Şekil 3.1 ve Çizelge 3.2'den görüleceği gibi kısa dönem öznitelikleri (MFKK ve SQKK), sınıflandırıcıdan bağımsız olarak uzun dönem özniteliklerine göre düşük performans göstermektedir. Bunun sebebi olarak, kısa-dönem özniteliklerinin konuşmacı kimliği, dil ve fonetik gibi KD sistemlerinde TO saldırılarının tespiti için önem arz etmeyen bilgileri de barındırmasıdır. Buna karşın, uzun dönem özniteliklerde bu gereksiz bilgiler, çerçevelerin ortalama ve standart sapmasının hesaplanması işlemi sırasında kaldırılmaktadır.

Deneylere ait SHÖ eğrileri Şekil 3.2'de verilmiştir. Bu eğrilerden görüldüğü üzere, UDOS öznitelikleri tüm operasyon noktalarında diğer sistemlere göre daha iyi performans göstermektedir. MFKK-GKM ve SQQK-GKM sistemleri operasyon noktalarına göre farklılık göstermektedir. Örneğin bu sistemler yanlış ret noktalarında benzer performans gösterirken, yanlış kabul noktalarında sistemler arasındaki performans farkı artmaktadır.



Şekil 3.2: Geliştirme kümesine ait SHÖ eğrileri

3.2 Değerlendirme Kümesi Sonuçları

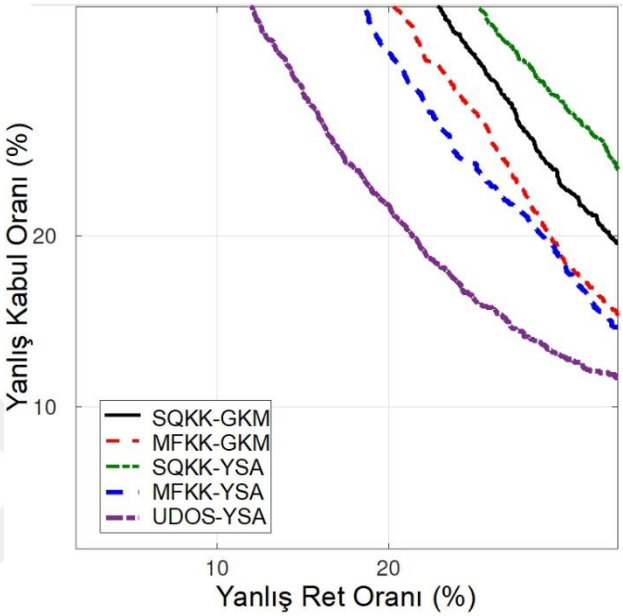
Geliştirme kümesi deneylerinde elde edilen çıkarımlara göre, değerlendirme kümesi öznitelikleri çıkarılırken alt-band frekans analizi uygulanmış, buna karşın ortalama ve varyans normalizasyon işleminin iyi yönde bir etki göstermediği tespit edildiği için uygulanmamıştır. Çizelge 3.3'de verilen gösterilen deney sonuçlarına göre en iyi karşı önlem sistemi %20,77'lik EHO değeriyle UDOS-YSA sistemi olmuştur. Buna karışı UDOS-YSAsistemi de dahil olmak üzere bütün sistemler, ürettikleri EHO değerlerine göre KD sistemlerinde TO saldırı tespitinde başarılı olmamıştır. Bunun nedeni olarak, değerlendirme kümesindeki ses kayıtlarının, eğitim ve geliştirme kümesinde yer alan ses kayıtlarından farklı konfigürasyonlarla oluşturulmuş ses kayıtları olmasıdır. Burada geliştirme kümesi kayıtları bilinmeyen türde saldırılar olarak isimlendirilir.

Çizelge 3.3: Geliştirme kümesi deney sonuçları

Sistem	EHO (%)
SQKK-GKM	29.94
MFKK-GKM	27.74
SQKK-YSA	32.64
MFKK-YSA	25.34
UDOS-YSA	**20.77**

UDOS-YSA sistemi, referans sisteme göre EHO'yu %44 oranında azaltmıştır. SQKK-YSA ve UDOS-YSA sistemleri arasındaki performans farkı ise daha yüksektir. MFKK

öznitelikleri için ise YSA sınıflandırıcısı GKM sınıflandırıcısına göre daha başarılı olmuştur. GKM sınıflandırıcısının YSA'ya göre üstünlük gösterdiği tek öznitelik türü ise SQKK olmuştur. Son olarak, deneylere ait SHÖ eğrileri Şekil 3.3'de verilmiştir. Geliştirme kümesine benzer bir şekilde tüm operasyon noktalarında UDOS-YSA sistemi üstündür fakat gösterdiği performans yeterli değildir.



Şekil 3.3: Değerlendirme kümesine ait SHÖ eğrileri

4. TARTIŞMA VE ÖNERİLER

Bu çalışmada konuşmacı doğrulama sistemlerine yapılması olası saldırı türlerinden tekrar saldırıları, daha önceden kaydedilen sesin tekrar oynatılması, ele alınmıştır. Daha önceki çalışmalarda genellikle öznitelik çıkarma yöntemi olarak kullanılan derin sinir ağları yaklaşımı sınıflandırıcı olarak kullanılmış olup, sonuçlar konuşmacı doğrulama sistemlerinde yoğun bir şekilde kullanılan Gauss karışım modeli yöntemiyle karşılaştırılmıştır. Mel-frekansı kepstrum katsayıları, sabit Q kepstrum katsayıları ve uzun dönem ortalama spektrum yöntemleri kullanılarak, tekrar saldırılarına özel olarak 2017 yılında düzenlenen, Otomatik Konuşmacı Doğrulama ve Karşı Önlemler yarışması için hazırlanan veri tabanındaki ses kayıtlarından öznitelikler çıkarılmıştır. Derin sinir ağları ve Gauss karşım modeli yöntemleriyle sınıflandırılan bu özniteliklerden en iyi sonucu uzun dönem ortalama spektrum öznitelikleri ve derin sinir ağları sınıflandırıcısı yöntemi vermiştir. Elde edilen en iyi sonuçlar geliştirme kümesi ve değerlendirme kümesi için sırasıyla %4,10 ve %20,77 eşit hata oranlarıyla elde dilmiştir. Sonuç olarak uzun dönem özniteliklerini, kısa dönem özniteliklere nazaran daha iyi performans gösterdiği ve buna ek olarak derin sinir ağları yaklaşımının umut verici bir sınıflandırıcı olduğu çalışmanın çıktıları arasında not edilmiştir. Bunlara ek olarak, yüksek frekans bölgesinin daha fazla bilgi taşıdığı ve ortalama-varyans normalizasyon işleminin ise faydalı olmadığı da çalışmada gösterilmiştir.

Gelecek çalışmalar için yapılacak ilk öneri deneylerin farklı veri tabanları üzerinden yapılmasıdır. Bu durum geliştirilen sistem sonuçlarının genelleştirilmesini sağlayacaktır. Ayrıca, derin sinir ağları yaklaşımının öznitelik çıkarmada kullanılmasının yanı sıra, konvolüsyonel sinir ağları, zaman gecikmeli sinir ağları gibi çeşitli yapıların da test edilmesi önem arz etmektedir. Son olarak farklı öznitelik yöntemlerinin gelecek çalışmalarda test edilmesi de öneriler arasındadır.

KAYNAKLAR

[1] Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14 (1), 4-20.

[2] Jain, A. K., Ross, A., & Pankanti, S. (2006). Biometrics: a tool for information security. *IEEE transactions on information forensics and security*, 1 (2), 125-143.

[3] Hadid, A., Evans, N., Marcel, S., & Fierrez, J. (2015). Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32 (5), 20-30.

[4] Bimbot, F., Bonastre, J. F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., ... & Reynolds, D. A. (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, 2004 (4), 430-451.

[5] Campbell, J. P. (1997). Speaker recognition: A tutorial. *Proceedings of the IEEE*, 85 (9), 1437-1462.

[6] Villalba, J., & Lleida, E. (2010). Speaker verification performance degradation against spoofing and tampering attacks, *FALA wokshop*, (ss. 131-134). Vigo: İspanya, Kasım, 10-12.

[7] Alegre, F., Janicki, A., & Evans, N. (2014). Re-assessing the threat of replay spoofing attacks against automatic speaker verification. *International Conference of the Biometrics Special Interest Group (BIOSIG)*, (ss. 1-6). Darmstadt: Germany, 10-12 Eylül.

[8] Ergünay, S. K., Khoury, E., Lazaridis, A., & Marcel, S. (2015). On the vulnerability of speaker verification to realistic voice spoofing. *Theory, Applications and Systems (BTAS)*, (ss. 1-6), Arlington: Amekira Birleşik Devletleri, 8-11 Eylül.

[9] Wu, Z., Evans, N., Kinnunen, T., Yamagishi, J., Alegre, F., & Li, H. (2015). Spoofing and countermeasures for speaker verification: A survey. *Speech Communication*, 66, 130-153.

[10] Ratha, N. K., Connell, J. H., & Bolle, R. M. (2001). Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40, 614-634.

[11] Chen, N., Qian, Y., Dinkel, H., Chen, B., & Yu, K. (2015). Robust deep feature for spoofing detection—The SJTU system for ASVspoof 2015 challenge. *Interspeech*. Dresden: Almanya, 6-10 Eylül.

[12] De Leon, P. L., Pucher, M., Yamagishi, J., Hernaez, I., & Saratxaga, I. (2012). Evaluation of speaker verification security and detection of HMM-based synthetic speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 20, 2280-2290.

[13] Stylianou, Y. (2009). Voice Transformation: A survey. *IEEE International Conference on Acoustics, Speech and Signal Processing,* (ss. 3585-3588). Portland: Amerika Birleşik Devletleri, 13-17 Haziran.

[14] Hautamäki, R. G., Kinnunen, T., Hautamäki, V., & Laukkanen, A. M. (2015). Automatic versus human speaker verification: The case of voice mimicry. *Speech Communication,* 72, 13-31.

[15] Gałka, J., Grzywacz, M., & Samborski, R. (2015). Playback attack detection for text-dependent speaker verification over telephone channels. *Speech Communication,* 67, 143-153.

[16] Villalba, J., & Lleida, E. (2011). Detecting replay attacks from far-field recordings on speaker verification systems. *European Workshop on Biometrics and Identity Management,* (ss. 274-285). Brandenburg: Almanya, 8-10 Mart.

[17] Wang, Z., Wei, G., & He, Q. (2011). Channel pattern noise based playback attack detection algorithm for speaker recognition. *Machine Learning and Cybernetics,* 4, 1708-1713.

[18] Yang, J., You, C., & He, Q. (2018). Feature with Complementarity of Statistics and Principal Information for Spoofing Detection. *Interspeech,* (ss. 651-655). Hyderabad: Hindistan, 2-6 Eylül.

[19] Muckenhirn, H., Korshunov, P., Magimai-Doss, M., Marcel, S., Muckenhirn, H., Korshunov, P., ... & Marcel, S. (2017). Long-Term Spectral Statistics for Voice Presentation Attack Detection. *IEEE/ACM Transactions on Audio, Speech and Language Processing,* 25, 2098-2111.

[20] Xiao, X., Tian, X., Du, S., Xu, H., Chng, E. S., & Li, H. (2015). Spoofing speech detection using high dimensional magnitude and phase features: The NTU approach for ASVspoof 2015 challenge. *International Speech Communication Association,* (ss. 1-5). Dresden: Almanya, 6-10 Eylül.

[21] Todisco, M., Delgado, H., & Evans, N. (2016). A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients. *Speaker Odyssey Workshop,* (ss. 249-252). Bilbao: İspanya, 26-29 Haziran.

[22] Todisco, M., Delgado, H., & Evans, N. (2017). Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification. *Computer Speech & Language,* 45, 516-535.

[23] Chen, Z., Xie, Z., Zhang, W., & Xu, X. (2017). Resnet and model fusion for automatic spoofing detection. *Interspeech,* (ss. 102-106). Stockholm: İsveç, 20-24 Ağustos.

[24] Wang, X., Xiao, Y., & Zhu, X. (2017). Feature selection based on CQCCs for automatic speaker verification spoofing. *Interspeech,* (ss. 32-36). Stockholm: İsveç, 20-24 Ağustos.

[25] Kinnunen, T., Sahidullah, M., Delgado, H., Todisco, M., Evans, N., Yamagishi, J., & Lee, K. A. (2017). The asvspoof 2017 challenge: Assessing the limits of replay spoofing attack detection.

[26] Font, R., Espn, J. M., & Cano, M. J. (2017). Experimental analysis of features for replay attack detection results on the ASVspoof 2017 challenge. *Interspeech,* (ss. 7-11). Stockholm: İsveç, 20-24 Ağustos.

[27] Sailor, H., Kamble, M., & Patil, H. (2018). Auditory filterbank learning for temporal modulation features in replay spoof speech detection. *Interspeech*, (ss. 666-670). Hyderabad: Hindistan, 2-6 Eylül.

[28] Saranya, M. S., & Murthy, H. A. (2018). Decision-level feature switching as a paradigm for replay attack detection. *Interspeech,* (ss. 686-690). Hyderabad: Hindistan, 2-6 Eylül.

[29] Lavrentyeva, G., Novoselov, S., Malykh, E., Kozlov, A., Kudashev, O., & Shchemelinin, V. (2017). Audio replay attack detection with deep learning frameworks. *Interspeech,* (ss. 82-86). Stockholm: İsveç, 20-24 Ağustos.

[30] Witkowski, M., Kacprzak, S., Zelasko, P., Kowalczyk, K., & Gałka, J. (2017). Audio replay attack detection using high-frequency features. *Interspeech,* (ss. 27-31). Stockholm: İsveç, 20-24 Ağustos.

[31] Li, L., Chen, Y., Wang, D., & Zheng, T.F. (2017). A Study on Replay Attack and Anti-Spoofing for Automatic Speaker Verification. *Interspeech,* (ss. 1-5). Stockholm: İsveç, 20-24 Ağustos.

[32] Nagarsheth, P., Khoury, E., Patil, K., & Garland, M. (2017). Replay Attack Detection Using DNN for Channel Discrimination. *Interspeech,* (ss. 97-101). Stockholm: İsveç, 20-24 Ağustos.

[33] Hanilçi, C., Kinnunen, T., Sahidullah, M., & Sizov, A. (2015). Classifiers for synthetic speech detection: a comparison.

[34] Larcher, A., Lee, K., Ma, B., & Li, H. (2014). Text-dependent speaker verification: Classifiers, databases and RSR2015. *Speech Communication*, 60, 56-77.

[35] Gosztolya, G., Busa-Fekete, R., Grósz, T., & Tóth, L. (2017). Dnn-based feature extraction and classifier combination for child-directed speech, cold and snoring identification. *Interspeech*, (ss. 3522-3526). Stockholm: Sweden, 20-24 Ağustos.

[36] Wu, Z., Kinnunen, T., Evans, N., & Yamagishi, J. (2014). ASVspoof 2015: Automatic speaker verification spoofing and countermeasures challenge evaluation plan. *Training*, 15, 3750.

[37] Wu, Z., Kinnunen, T., Evans, N., Yamagishi, J., Hanilçi, C., Sahidullah, M., & Sizov, A. (2015). ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge. *International Speech Communication Association, (*ss 1-5). Dresden: Almanya, 6-10 Eylül.

[38] Kinnunen, T., Sahidullah, M., Delgado, H., Todisco, M., Evans, N., Yamagishi, J., & Lee, K. A. (2017). The asvspoof 2017 challenge: Assessing the limits of replay spoofing attack detection.

[39] Kinnunen, T., Sahidullah, M., Falcone, M., Costantini, L., Hautamäki, R. G., Thomsen, D., ... & Evans, N. (2017). Reddots replayed: A new replay spoofing attack corpus for text-dependent speaker verification research. *Acoustics, Speech and Signal Processing (ICASSP),* (ss. 5395-5399). New Orleans: Ameriak Birleşik Devletleri, 5-9 Mart.

[40] Delgado, H., Todisco, M., Sahidullah, M., Evans, N., Kinnunen, T., Lee, K. A., & Yamagishi, J. (2018). ASVspoof 2017 Version 2.0: meta-data analysis and baseline enhancements. *Odyssey 2018 The Speaker and Language Recognition Workshop*, (ss. 296-303). Les Sables: Fransa, 26-29 Haziran.

[41] Picone, J. W. (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 9, 1215-1247.

[42] Davis, S. B., & Mermelstein, P. (1990). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Readings in speech recognition*, 65-74.

[43] Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital signal processing*, 1-3, 19-41.

[44] Martin, A., Doddington, G., Kamm, T., Ordowski, M., & Przybocki, M. (1997). The DET curve in assessment of detection task performance. *Eurospeech*, (ss. 1895-1898). Rhodes: Yunanistan, 22-25 Eylül.

[45] LeCun, Y., & Ranzato, M. (2013). Deep learning tutorial. *Tutorials in International Conference on Machine Learning (ICML'13)*, (ss. 1-29). Atlanta: Amerika Birleşik Devletleri, 16-21 Haziran.

[46] Bergstra, J., Bastien, F., Breuleux, O., Lamblin, P., Pascanu, R., Delalleau, O., ... & Bengio, Y. (2011). Theano: Deep learning on gpus with python. *NIPS 2011, BigLearning Workshop*, (ss. 1-48). Nevada: İspanya, 16-17 Aralık.

[47] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., ... & Bengio, Y. (2010). Theano: A CPU and GPU math compiler in Python. *Python in Science*. Texas: Amerika Birleşik Devletleri, 2 Haziran-3 Temmuz.

[48] Chollet, F. (2018). Keras: The python deep learning library. *Astrophysics Source Code Library*.

[49] Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. *Neural networks: Tricks of the trade*, 7700, 437-478.

[50] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929-1958.

[51] Li, M., Zhang, T., Chen, Y., & Smola, A. J. (2014). Efficient mini-batch training for stochastic optimization. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, (ss. 661-670). New York: Amerika Birleşik Devletleri, 24-27 Ağustos.

ÖZGEÇMİŞ

Ad-Soyadı **: Bekir Bakar**

Doğum Tarihi ve Yeri **: 19/10/1992 Çatalpınar/ORDU**

E-posta **:** b.bakar@outlook.com

ÖĞRENİM DURUMU:

- **Lisans:** 2015, Uludağ Üniversitesi, Mühendislik Fakültesi, Elektronik Mühendisliği
- **Yüksek Lisans:** 2018, Bursa Teknik Üniversitesi, Elektrik-Elektronik Mühendisliği Anabilim Dalı

TEZDEN TÜRETİLEN ESERLER, SUNUMLAR VE PATENTLER:

- Bakar, B., & Hanilçi, C. (2018). Replay spoofing attack detection using deep neural networks. *26. Signal Processing and Communications Applications Conference (SIU)*. İzmir: Türkiye, 2-5 Mayıs.
- Bakar, B & Hanilçi, C. (2018). An Experimental Study on Audio Replay Attack Detection Using Deep Neural Network, *Spoken Language Technology Workshop (SLT)*. Atina: Yunanistan,18-21, Aralık.