KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

AÇI GENLİK DÖNÜŞÜMÜ KULLANARAK KONUŞMACI TANIMLAMA

YÜKSEK LİSANS TEZİ

Büşra ORAN

EKİM 2025
TRABZON

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

AÇI GENLİK DÖNÜŞÜMÜ KULLANARAK KONUŞMACI TANIMLAMA

Büşra ORAN
ORCID : - - -

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

Tezin Enstitüye Verildiği Tarih : 15 / 09 /2025
Tezin Savunma Tarihi : 16 / 10 /2025

Tez Danışmanı : Prof. Dr. Cemal KÖSE
ORCID : - - -

Trabzon 2025

ÖNSÖZ

“Açı Genlik Dönüşümü Kullanarak Konuşmacı Tanımlama ” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Programı’nda hazırlanmıştır.

Tez çalışmam boyunca hiçbir desteğini esirgemeyen ve bilim insanı olma yolunda kıymetli bilgileriyle ve tecrübesiyle bana yol gösteren danışmanım Sayın Prof. Dr. Cemal KÖSE’ye teşekkürü bir borç bilirim. Ayrıca önerileri ile tezime büyük katkıda bulunan saygıdeğer hocalarım Sayın Dr. Öğr. Üyesi Ercüment ÖZTÜRK’e ve Sayın Dr. Öğr. Üyesi Bahar HATİPOĞLU YILMAZ’a , eğitim öğretim hayatımda katkısı olan tüm hocalarıma teşekkür ederim.

Son olarak, doğduğum günden beri yanımda olan, aldığım her kararda arkamda duran, maddi ve manevi desteklerini hiçbir zaman esirgemeyen annem Samiye ORAN ve babam Murat ORAN’a, kardeşlerime, arkadaşlarıma ve kıymetli yol arkadaşım Muhammed Emin ALTINIŞIK’a teşekkür eder, minnettarlığımı sunarım.

Bu tez çalışmasının bundan sonraki çalışmalara katkıda bulunmasını temenni ederim.

Büşra ORAN

Trabzon 2025

TEZ ETİK BEYANNAMESİ

Yüksek lisans tezi olarak sunduğum, “Açı Genlik Dönüşümü Kullanarak Konuşmacı Tanımlama” başlıklı bu çalışmayı baştan sona kadar danışmanım Prof. Dr. Cemal KÖSE’ nin rehberliğinde tamamladığımı, tezimde yer alan ve referans gösterilen tüm bilgileri metin içerisinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecim boyunca bilimsel araştırma ve etik kurallara uygun davrandığımı ve bu ilkelerin ihlal edildiğinin kanıtlanması durumunda üzerime düşen her türlü sorumluluğu kabul ettiğimi beyan ederim.
16/10/2025

Büşra ORAN

İÇİNDEKİLER

Sayfa No

Yüksek Lisans Tezi

ÖZET

AÇI GENLİK DÖNÜŞÜMÜ KULLANARAK KONUŞMACI TANIMLAMA

Büşra ORAN
Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Cemal KÖSE
2025, 66 Sayfa

Konuşmacı tanımlama, belirli bir ses kaydından konuşmayı gerçekleştiren kişiyi belirleme süreci olup güvenlik sistemleri gibi birçok alanda kullanılmaktadır. Bu çalışmada konuşmacı tanıma sürecinde kullanılan özellik çıkarım yöntemlerinin karşılaştırmalı analizi yapılmıştır. Literatürde yaygın olarak kullanılan MFCC (Mel-Frekans Kepstrum Katsayıları) ve LPCC (Doğrusal Öngörülü Kepstral Katsayılar) yöntemleri incelenmiş, bunlara ek olarak ilk kez önerilen Açı Genlik Dönüşümü (AGD) yöntemi uygulanmıştır. AGD, ses sinyallerindeki tepe ve çukur noktalarından açı ve genlik değerlerini hesaplayarak iki boyutlu görüntüler oluşturur. Bu görüntülerden Yönlendirilmiş Gradyanların Histogramı (HOG) yöntemiyle özellikler çıkarılmış ve K-En Yakın Komşu (KNN) ile Destek Vektör Makineleri (SVM) algoritmalarında sınıflandırma yapılmıştır. Ayrıca, görüntüler doğrudan Konvolüsyonel Sinir Ağı (CNN) ile sınıflandırılmıştır. Çalışmada LibriSpeech ve VoxCeleb1 veri setleri kullanılmış; veriler eğitim ve test gruplarına ayrılmıştır. Deneysel sonuçlar, birinci ve ikinci derece AGD ile CNN kullanıldığında LibriSpeech veri setinde sırasıyla %93.2 ve %91, VoxCeleb1 veri setinde ise %75.5 ve %71 doğruluk oranına ulaşıldığını göstermektedir. Ayrıca, MFCC ve birinci derece AGD yöntemlerinin hibrit CNN yaklaşımıyla birlikte kullanılması LibriSpeech'te %97.6, VoxCeleb1'de ise %84.3 doğruluk oranı sağlamıştır. Elde edilen bulgular, AGD yönteminin konuşmacı tanıma alanında alternatif ve etkili bir yaklaşım sunduğunu ortaya koymaktadır.

Anahtar Kelimeler: Konuşmacı tanımlama, özellik çıkarımı, açı genlik dönüşümü, sınıflandırma algoritmaları

Master Thesis

SUMMARY

SPEAKER IDENTIFICATION USING ANGLE AMPLITUDE TRANSFORM

Büşra ORAN
Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Computer Engineering Graduate Program
Supervisor: Prof. Dr. Cemal KÖSE
2025, 66 Pages

Speaker identification is the process of determining the person who performed the speech from a given audio recording, and it is widely used in various fields such as security systems. In this study, a comparative analysis of feature extraction methods used in the speaker recognition process was conducted. The commonly used methods in the literature, MFCC (Mel-Frequency Cepstral Coefficients) and LPCC (Linear Predictive Cepstral Coefficients), were examined, and in addition, a newly proposed method, Angle-Amplitude Transform (AAT), was applied for the first time. AAT computes angle and amplitude values from the peaks and troughs of speech signals and generates two-dimensional images. Features were extracted from these images using the Histogram of Oriented Gradients (HOG) method, and classification was performed with the K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) algorithms. Furthermore, the images were directly classified using a Convolutional Neural Network (CNN). The LibriSpeech and VoxCeleb1 datasets were used in the study, and the data was divided into training and testing sets. Experimental results show that with first and second order AAT combined with CNN, accuracy rates of 93.2% and 91% were achieved on the LibriSpeech dataset, while 75.5% and 71% were obtained on the VoxCeleb1 dataset, respectively. Moreover, the hybrid CNN approach combining MFCC and first order AAT achieved accuracy rates of 97.6% on LibriSpeech and 84.3% on VoxCeleb1. The findings demonstrate that the AAT method provides an alternative and effective approach in the field of speaker recognition.

Keywords: Speaker identification, feature extraction, angle amplitude transform, classification algorithms

ŞEKİLLER DİZİNİ

Sayfa No

TABLOLAR DİZİNİ

Sayfa No

SEMBOLLER VE KISALTMALAR DİZİNİ

MFCC	: Mel- Frekanslı Kepstral Katsayılar
LPCC	: Doğrusal Öngörülü Kepstral Katsayılar
AGD	: Açı Genlik Dönüşümü
HOG	: Yönlendirilmiş Gradyanların Histogramı
FFT	: Hızlı Fourier Dönüşümü
IFFT	: Ters Hızlı Fourier Dönüşümü
KNN	: K- En Yakın Komşu
SVM	: Destek Vektör Makineleri
CNN	: Konvolüsyonel Sinir Ağları

1. GENEL BİLGİLER

1.1. Giriş

İnsan sesi, retina ve parmak izi gibi kişiye özgü biyometrik özellikler taşır. Bu özellikler, insan sesini biyometrik bir tanımlayıcı haline getirir ve konuşmacı tanımlama sistemleri de bu biyometrik bilgiyi kullanarak kişilerin kimliğini belirlemeyi hedefler (Günerhan, 2023). Konuşmacı tanımlama güvenlik sistemleri, bankacılık uygulamaları gibi çeşitli alanlarda kimlik doğrulama amacıyla yaygın olarak kullanılmaktadır. Ancak sesin biyometrik yapısı çevresel gürültü, duygusal durum, sağlık koşulları gibi etkenlerden olumsuz etkilenebilmektedir (Jahangir vd., 2020).

Konuşmacı tanımlama süreci genellikle iki aşamadan oluşur: Eğitim aşamasında bireylerin sesleri kaydedilerek her bireye ait bir model oluşturulur. Test aşamasında ise gelen ses bu modellerle karşılaştırılarak kimlik tespiti yapılır. Bu süreçte üç temel adım öne çıkar: özellik çıkarımı, sınıflandırma ve başarı metriklerinin hesaplanması.

Özellik çıkarımı, ses sinyalinin karakteristik bilgilerini temsil eden özniteliklerin elde edilmesini sağlar. Literatürde en yaygın kullanılan özellik çıkarım yöntemleri arasında Mel-Frekans Kepstral Katsayıları (MFCC) ve Doğrusal Öngörülü Kepstral Katsayılar (LPCC) yer almaktadır. Elde edilen bu öznitelikler, sınıflandırma algoritmalarına giriş olarak verilir. Sınıflandırma sonucunda sistemin başarısı doğruluk oranı (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi metriklerle değerlendirilir.

Konuşmacı tanımlama alanında yaygın olarak LibriSpeech ve VoxCeleb1 veri setleri kullanılmaktadır. Ayrıca farklı çalışmalarda özel olarak toplanan ses kayıtlarıyla da veri setleri oluşturulmuştur (Kop & Bayındır, 2021; Fasounaki vd., 2021). Literatürde bu veri setleri üzerinde çeşitli özellik çıkarım ve sınıflandırma yöntemleri denenmiştir.

Bu çalışmada, konuşmacı tanıma alanına yenilikçi bir yaklaşım sunulmaktadır. Geleneksel öznitelik çıkarım yöntemlerinden farklı olarak, ses sinyalleri üzerinde ilk kez uygulanan Açı Genlik Dönüşümü (AGD) yöntemi ile konuşmacıya özgü yapısal özellikler görsel olarak temsil edilmiştir.

Çalışma kapsamında, ham ses sinyali öncelikle belirli uzunluklardaki zaman dilimlerine bölünerek çerçeveleme (frame) işlemi uygulanmıştır. Her çerçevenin başında ve sonunda oluşabilecek bozulmaları azaltmak için Hamming penceresi uygulanmış, böylece sinyalin geçişleri daha yumuşak hâle getirilmiştir. Pencerelenmiş her bir çerçeve üzerinden yerel değişim noktaları olan tepe (maksimum) ve çukur (minimum) noktalar tespit edilmiştir. Bu ekstremum noktalar arasında kurulan geometrik ilişkiler yardımıyla her bir merkez nokta için açı ve genlik değerleri hesaplanmıştır. Bu değerler, iki boyutlu bir düzlem üzerinde görselleştirilerek her çerçeve için özgün bir Açı Genlik Dönüşümü Görüntüsü oluşturulmuştur. Elde edilen bu görüntülerden Yönlendirilmiş Gradyanların Histogramı (Histogram of Oriented Gradients – HOG) yöntemiyle öznitelikler çıkarılmış; bu öznitelikler, K-En Yakın Komşu (KNN) ve Destek Vektör Makineleri (SVM) gibi makine öğrenmesi algoritmalarıyla sınıflandırılmıştır. Ayrıca açı genlik görselleri doğrudan Konvolüsyonel Sinir Ağı (CNN) mimarisine giriş olarak verilmiş ve bu sayede derin öğrenme temelli sınıflandırma gerçekleştirilmiştir.

Çalışmanın ikinci aşamasında ise, ses sinyalinin spektral özelliklerini çıkarmak amacıyla yaygın kullanılan MFCC yöntemi uygulanmıştır. MFCC hesaplamasında da benzer şekilde, sinyal önce kısa zamanlı çerçevelere bölünmüş ve her çerçeveye Hamming penceresi uygulanarak daha kararlı frekans temsilleri elde edilmiştir. Ardından Hızlı Fourier dönüşümü (FFT), Mel filtre bankası uygulaması ve Ters Hızlı Fourier Dönüşümü (IFFT) adımlarıyla öznitelik vektörleri elde edilmiştir. Daha sonra, AGD ve MFCC yöntemleri hibrit bir sistem altında birleştirilmiş ve çoklu temsile dayalı konuşmacı tanıma yaklaşımı geliştirilmiştir. Bu hibrit yapı kapsamında, AGD'den elde edilen görsel öznitelikler ile MFCC'den türetilen spektral öznitelikler hem özellik birleştirme yöntemi ile klasik sınıflayıcılara (KNN, SVM) verilmiş, hem de çok girişli CNN mimarisi kullanılarak her iki yöntemden doğrudan öğrenme gerçekleştirilmiştir.

Geliştirilen bu yöntemlerin değerlendirilmesinde, konuşmacı tanıma alanında literatürde sıklıkla kullanılan ve farklı koşullarda hazırlanmış olan LibriSpeech ve VoxCeleb1 veri setleri kullanılmıştır. LibriSpeech, sessiz ve stüdyo kalitesinde kayıtlar içerirken; VoxCeleb1, doğal ortamlardan alınan, gürültü içeren ve gerçek hayattan ses örnekleri sunmaktadır. Bu iki farklı veri seti üzerinden yapılan deneysel analizler, hem AGD hem de MFCC yöntemlerinin konuşmacıya özgü bilgileri etkin biçimde temsil

ettiğini, ayrıca hibrit biçimde kullanılan bu iki yöntemle tanıma performansının belirgin şekilde arttığını ortaya koymaktadır.

Bu yönüyle çalışma, geleneksel ses özniteliklerini yenilikçi bir görsel temsil ile bütünleştirerek konuşmacı tanıma sistemlerinde doğruluk ve genelleme başarısını artıran yeni bir yöntem önermektedir.

1.2. Literatür Taraması

Günerhan (2023), yaptığı çalışmalarda konuşmacı tanıma alanında yaygın kullanılan MFCC özellik çıkarım yöntemini Librispeech ve Voxceleb1 veri setlerine uygulamıştır. Bu veri setlerinden çıkarılan özniteliklerle Yapay sinir ağları (ANN), konvolüsyonel sinir ağları (CNN), Uzun-Kısa Süreli Bellek Ağları (LSTM) eğitilmiş ve karşılaştırılmıştır. Genel olarak uzun kısa süreli bellek ağlarının %95.2 sonuç verdiği tespit edilmiştir. Bu sonuçları geliştirmek adına hiperparametre eklenmesiyle sonuçların %99.5' e yükseldiği gözlenmiştir.

Jahangir vd. (2020), konuşmacı tanımlamada temel MFCC özellik çıkarım yöntemi yerine, zaman tabanlı MFCCT özellik çıkarım yöntemi kullanılması önerilmiştir. Librispeech veri setinden elde edilen öznitelikler, giriş olarak Rastgele Orman (RF), Naive Bayes (NB), KNN, J48, SVM gibi beş farklı makine algoritmalarına ve derin öğrenme tabanlı Derin Sinir Ağları (DNN) eğitim algoritmasına verilmiştir. Ayrıca DNN'ye giriş verilen MFCCT yöntemi özelliklerinin, temel MFCC'den daha etkili olduğu ve %92 sınıflandırma doğruluğuna ulaştığı görülmüştür.

Yao vd. (2023), tarafından konuşmacı kimliği tanımlamaya düşmanca saldırıya karşı Simetrik Belirginlik Temelli Kodlayıcı Çözücü (Symetric Saliency based Encoder Decoder- SSED) adlı yöntem önerilmiştir. Bu yöntem Voxceleb1 ve Voxceleb2 veri setinde gerçekleştirilerek kayıtlı konuşmacılardan rastgele birini seçip sahte konuşmacı olup olmadığı hakkında seçim yapmaktadır. Sistem test verilerinden % 93.15'lik doğruluk oranına ulaşmıştır.

Fasounaki vd. (2021), tarafından kısa konuşma segmentlerine dayalı olarak konuşmacı tanıma işlemi gerçekleştirmek amacıyla CNN mimarisi önerilmiştir. Özellik çıkarımı için MFCC yöntemi kullanılmış olup, sistem hem Librispeech hem de VoxCeleb1 veri setleri üzerinde test edilmiştir. Eğitim verileri 2 saniyelik ses parçalarından oluşmaktadır ve modelin girdi verisi olarak MFCC spektogram görselleri kullanılmaktadır. Sınıflandırma işlemi için önerilen CNN mimarisi ile Librispeech veri setinde %99.83, VoxCeleb1 veri setinde ise %97.1 doğruluk oranına ulaşılmıştır. Bu sonuçlar, özellikle kısa süreli ses örnekleriyle yapılan konuşmacı tanıma işlemlerinde CNN temelli mimarilerin etkili olabileceğini ortaya koymaktadır.

Nandan vd. (2022), metin bağımlı konuşmalarda fiziksel olarak değiştirilmiş ses varyasyonlarını dikkate alan bir tanıma sistemi önermişlerdir. Çalışmada, konuşmacıların sesleri normal hallerinin yanı sıra hızlı, yavaş konuşma, perde değişimi, ağız içinde nesne bulundurma gibi on farklı fiziksel değişim koşulu altında kaydedilmiştir. Bu farklı ses koşullarına ait örnekler üzerinde, MFCC ile birlikte birinci (delta MFCC) ve ikinci (delta delta MFCC) türevler çıkarılmış; elde edilen öznitelikler üzerinden ortalama ve korelasyon katsayılarına dayalı istatistiksel analizler gerçekleştirilmiştir. Tanıma sürecinde, istatistiksel özellikler kullanılarak Karar Ağacı (DT) algoritması ile sınıflandırma işlemi uygulanmıştır. Önerilen yöntem sayesinde, konuşmacının ağzında maske bulunduğu durumda %92,5 doğruluk oranı elde edilirken; ağızda nesne bulunduğu senaryoda ise %96,29 oranında tanıma başarısına ulaşılmıştır.

Ye & Yang (2021), konuşmacı tanıma için 2 boyutlu CNN ile Gated Recurrent Unit (GRU) tabanlı katmanları birleştiren bir derin sinir ağı (DNN) modeli önermiştir. Önerilen model, CNN katmanları ile ses spektrogramlarından özellikleri çıkarmakta, ardından ardışık GRU katmanları ile zamansal bağımlılıkları öğrenmektedir. Bu mimari, hem spektral hem de zamansal ilişkileri dikkate alarak konuşmacıya özgü ses özelliklerini üretmektedir. Modelin eğitimi, yüksek kaliteli Mandarin konuşmalarından oluşan AISHELL-1 veri seti üzerinde gerçekleştirilmiştir. Veri ön işleme sürecinde Hamming penceresi, Kısa Süreli Fourier dönüşümü (STFT) ve Mel filtre bankasından çıkarılan öznitelikler kullanılmıştır. Ayrıca birinci ve ikinci türev delta MFCC öznitelikleri çıkarılmış ve bu öznitelikler CNN katmanına giriş olarak verilmiştir. Yapılan karşılaştırmalı deneylerde, önerilen CNN- GRU modeli, MFCC- GMM %86.29, CNN

%92.45, RNN %89.67 ve LSTM %96.25 gibi diğer yöntemlere kıyasla %98.96 doğruluk oranı ile en yüksek başarıyı elde etmiştir.

Ahmed vd. (2022), tarafından yapılan çalışmada TIMIT ve HTIMIT veri setleri üzerinde geliştirilen DNN tabanlı sistem, adaptif budama tekniğiyle optimize edilmiştir. DNN, hem öznitelik çıkarımı hem de doğrudan sınıflandırma amacıyla kullanılmakla birlikte düşük karmaşıklıklı yapılarda etkili sonuçlar vermektedir. Karmaşıklığı azaltılan model, test verisinde %95.23 doğruluk sağlamış ve gürültülü ortamlarda da dayanıklı bir performans göstermiştir.

Leng vd. (2022), VoxCeleb1 veri kümesinde yer alan cinsiyet ve milliyet bilgilerini, konuşmacı tanımada ayrıştırıcı yüksek seviyeli öznitelikler olarak değerlendirilmiştir. Önerilen AT_GN (Dikkat Mekanizmasına Dayalı Cinsiyet ve Milliyet Bilgisi) algoritması, cinsiyet ve milliyet özniteliklerini dikkat mekanizması yardımıyla spektrogram özellikleriyle birleştirerek tanıma performansını artırmaktadır. Yapılan deneyler, önerilen yapının mevcut CNN tabanlı mimarilerde (VGG-M, ResNet-34, vb.) doğruluk artışı sağladığını ortaya koymuştur. Elde edilen sonuçlarda, cinsiyet ve milliyet özellikleriyle akustik özelliklerinin birlikte kullanılması sonucu yüksek başarı elde edilmiştir.

Kacur (2016), yaptığı çalışmada konuşmacı tanıma sistemleri için KNN algoritmasına dayalı sınıflandırma yöntemi geliştirmiştir. Klasik KNN algoritması, ağırlıklı KNN (w- KNN) biçiminde yapılandırılmış ve çeşitli pencereleme yöntemleri kullanılmıştır. Ayrıca öznitelik dönüşümleri olarak PCA (Temel Bileşenler Analizi) ve LDA (Doğrusal Ayırt Edici Analiz) gibi denetimli ve denetimsiz boyut indirgeme teknikleri kullanılarak test edilmiştir. Çalışmada kullanılan veri seti, 14 erkek ve 7 kadından oluşan toplam 21 konuşmacının, iki farklı uzaklıkta (1 ve 3 metre) ofis ortamında kaydedilen seslerinden oluşmaktadır. MFCC öznitelikleriyle temsil edilen konuşma verisi, 22 kHz örnekleme frekansı ile toplanmıştır. Deneysel sonuçlar, Gaussian ağırlıklı w- KNN modelinin hem doğruluk hem de çevresel koşullara karşı dayanıklılık açısından yüksek performans gösterdiğini ortaya koymuştur.

Singh (2023) tarafından gerçekleştirilen çalışmada, MFCC, Gammatone Frekanslı Kepstral Katsayıları (GFCC) ve Güç Normalizasyonlu Kepstral Katsayılar (PNCC) olmak üzere üç farklı öznitelik çıkarım yöntemi kullanılmıştır. Elde edilen öznitelikler, Yapay Sinir Ağı (ANN), Tekrarlayan Sinir Ağı (RNN) ve CNN tabanlı sınıflandırma

algoritmalarıyla değerlendirilmiş ve sistem performansı karşılaştırmalı olarak incelenmiştir. Ayrıca, sınıflandırma doğruluğunu ve işlem verimliliğini artırmak amacıyla PCA ile boyut indirgeme işlemi uygulanmıştır. Çalışmada TIMIT ve VoxCeleb1 veri setleri kullanılarak gerçekleştirilen deneylerde, TIMIT veri setinin hem temiz hem de yapay olarak oluşturulan gürültülü versiyonları test edilmiştir. Elde edilen sonuçlara göre, MFCC öznitelikleri ile CNN sınıflayıcının birlikte kullanıldığı durumda %96.90'a kadar doğruluk oranı sağlanmış, bazı test senaryolarında ise bu oran %99.23'e ulaşmıştır. RNN modellerinde GRU yapısı kullanıldığında %90'ın üzerinde tanıma başarısı elde edilmiştir.

Kop & Bayındır (2021), tarafından gerçekleştirilen çalışmada, bebek ağlamalarının nedenlerine göre sınıflandırılması amacıyla makine öğrenmesi temelli bir model geliştirilmiştir. Çalışmada Baby Chillanto veri seti ve çevrim içi kaynaklardan derlenen özel bir veri seti kullanılmıştır. Özellik çıkarımı için MFCC ve LPCC yöntemleri karşılaştırılmış; MFCC'nin daha yüksek performans sunduğu belirlenmiştir. Sınıflandırma sürecinde KNN, MLP (Çok Katmanlı Algılayıcılar), Karar Ağacı ve Rastgele Orman algoritmaları denenmiş; MFCC öznitelik çıkarım yöntemiyle MLP algoritması kullanıldığında %93 doğruluk oranı elde edilmiştir.

Nakagawa vd. (2012), tarafından gerçekleştirilen çalışmada, konuşmacı tanıma ve doğrulama amacıyla MFCC ile faz bilgisi birleştirilmiştir. Çalışmada, MFCC'nin sadece genlik spektrumunu dikkate alması nedeniyle konuşmacıya özgü faz bilgisinin ihmal edildiği vurgulanmış ve bu eksikliğe yönelik olarak çerçeve konumuna bağlı faz normalizasyon yöntemi önerilmiştir. Özellik çıkarımı için MFCC ve normalize edilmiş faz bilgisi kullanılarak Gaussian Karışım Modelleri (GMM) eğitilmiş ve bu modeller doğrusal olarak birleştirilmiştir. Deneyler Japonca NTT ve JNAS veri setleri üzerinde gerçekleştirilmiştir. Sonuçlar, MFCC ile faz bilgisinin birleşiminin MFCC'ye kıyasla konuşmacı tanıma doğruluğunu %97.4'ten %98.8'e çıkardığını göstermiştir. Bu bulgular, faz bilgisinin MFCC ile tamamlayıcı nitelikte olduğunu ve tanıma performansını anlamlı biçimde artırdığını ortaya koymaktadır.

Hamsa vd. (2023), tarafından yürütülen çalışmada, duygusal ve gürültülü konuşma ortamlarında konuşmacı tanıma problemi ele alınmıştır. Bu bağlamda, konuşma ayrıştırma ve konuşmacı tanıma amacıyla önceden eğitilmiş derin sinir ağı tabanlı bir ses ayrıştırma maskesi ile VGG-16 mimarisi kullanılarak uçtan uca bir model önerilmiştir. Öznitelik çıkarımı sürecinde MFCC, Genlik Modülasyon Spektrumu (AMS) ve RASTA-PLP gibi

özellikler kullanılmış; bunlar, Dalgacık Paket Dönüşümü (WPT) temelli filtre bankası yardımıyla elde edilmiştir. Karar verme aşamasında ise transfer öğrenmeye dayalı SpeechVGG mimarisi, modifiye edilmiş kayıp fonksiyonu ile eğitilmiştir. Modelin başarısı; RAVDESS, SUSAS ve Emirati-accented Speech Dataset (ESD) veri setleri üzerinde sırasıyla %86.16, %87.03 ve %86.6 ortalama konuşmacı tanıma doğrulukları ile değerlendirilmiş ve literatürdeki GMM-DNN, SVM ve MLP tabanlı yaklaşımlardan üstün sonuçlar elde edilmiştir.

Vuppala & Rao (2012), gürültülü ortamlarda konuşmacı tanıma performansını artırmak amacıyla, sesli harf (vowel) bölgelerinden elde edilen özniteliklerin etkinliğini araştırmıştır. Çalışmada, sesli harflerin yüksek enerjili ve periyodik yapısı nedeniyle konuşmacıya özgü bilgileri daha sağlam taşıdığı varsayılmıştır. Gürültü azaltımı için CTSP (Birleşik Zamansal ve Spektral İşleme) yöntemi kullanılmış; öznitelik çıkarımında MFCC, delta ve delta-delta katsayıları tercih edilmiştir. Sınıflandırmada GMM-UBM yaklaşımı uygulanmıştır. TIMIT veri seti ile yapılan deneylerde, kararlı sesli harf bölgelerinden çıkarılan özniteliklerin, düşük sinyal- gürültü oranlarında dahi %93'e kadar doğruluk sağladığı gösterilmiştir.

Liu vd. (2017), erişim kontrol sistemlerinde konuşmacı tanıma uygulamalarına yönelik, metinden bağımsız ve MFCC tabanlı bir sistem geliştirmiştir. Çalışmada, öznitelik çıkarımı için MFCC kullanılmış; bu özellikler konuşmacıya özgü spektral bilgileri temsil etmiş ve sistemin güvenilirliğini artırmıştır. Elde edilen öznitelikler, sınıflandırma aşamasında GMM ve KNN algoritmaları ile değerlendirilmiştir. Modelin başarısını değerlendirmek için TIMIT veri seti kullanılmış ve farklı konuşmacılar üzerinden gerçekleştirilen deneylerde ortalama doğruluk oranı %92,8 olarak raporlanmıştır. Bu sonuçlar, MFCC tabanlı sistemlerin metinden bağımsız senaryolarda etkin bir şekilde kullanılabileceğini göstermiştir.

González vd. (2019), tarafından yapılan çalışmada, stres koşulları altındaki konuşmalar üzerinden konuşmacı tanıma performansını artırmak amacıyla veri çoğaltma (Data Augmentation) yöntemleri geliştirilmiştir. Özellikle cinsiyete dayalı şiddet durumlarında kullanılmak üzere giyilebilir cihazlar için kişiselleştirilmiş bir konuşmacı tanıma sistemi önerilmiştir. Sistem, doğal stres verisi içermeyen durumlar için nötr konuşma verileri üzerinden yapay stresli örnekler üretmekte ve eğitim verisini bu şekilde zenginleştirmektedir. Öznitelik çıkarım aşamasında; ortalama ve standart sapma değerleri

alınmış, MFCC, formant frekansları, perde (pitch) ve ilk üç formant bileşenleri kullanılmıştır. 1 saniyelik segmentler üzerinden 34 boyutlu bir öznitelik vektörü elde edilmiştir. Veri çoğaltma sürecinde ise sesin perde ve hız değerleri değiştirilerek nötr verilerden yapay stresli konuşmalar türetilmiştir. Sınıflandırma aşamasında GMM, SVM ve MLP algoritmaları karşılaştırılmış, yüksek başarı oranı nedeniyle MLP tercih edilmiştir. Çalışmalarda VOCE veri seti kullanılmış ve kalp atış hızına dayalı otomatik stres etiketleme yöntemi uygulanmıştır. Deneysel sonuçlara göre, sadece nötr verilerle eğitilmiş bir sistemin stresli konuşmaları tanımada %79.2 doğruluk sağladığı; ancak veri çoğaltma teknikleriyle eğitilmiş sistemin doğruluk oranını %99.45'e kadar yükselttiği gözlemlenmiştir. Böylece, yapay üretilen stresli verilerle eğitilen sistemlerin stres koşulları altındaki konuşmacı tanımada oldukça başarılı sonuçlar verdiği ortaya konmuştur.

Geravanchizadeh & Ghalamiosgouei (2020), uzak mesafeden yapılan konuşmacı tanıma sistemlerinde ortaya çıkan gürültü ve yankı kaynaklı uyumsuzluk sorununu çözmek amacıyla yeni bir yöntem önermiştir. Bu çalışmada, insan işitsel sisteminden ilham alan iki kulaklı bir konuşmacı tanıma sistemi geliştirilmiştir. Sistemde, gelen ses sinyali üzerinde Eşitleme-İptal (Equalization-Cancelation, EC) işlemi uygulanarak ortam seslerinin etkisi azaltılmıştır. Özellik çıkarımı için MFCC ve GFCC yöntemleri kullanılmış, konuşmacı modellemesi için GMM-UBM (Gaussian Mixture Model- Universal Background Model) yapısı tercih edilmiştir. Sistem, Grid veri seti kullanılarak farklı gürültü türleri (White, Babble, Factory, SSN) ve yankılı ortamlarda test edilmiştir. Elde edilen sonuçlar, önerilen yöntemin hem sessiz hem de gürültülü ortamlarda diğer geleneksel yöntemlere göre daha başarılı olduğunu göstermiştir. Özellikle düşük Sinyal-Gürültü Oranlarında (SNR) ve yankılı koşullarda sistemin doğruluk oranı önemli ölçüde artmıştır.

Herbig vd. (2012), yaptığı çalışmalarda konuşmacıyı otomatik tanıyan ve her konuşmacıya özel olarak kendini geliştiren bir konuşma tanıma sistemi (SILAS) geliştirmiştir. Bu sistem, kullanıcılardan çok kısa bir ses kaydı alarak konuşmacı profili oluşturmakta ve sonraki konuşmalarla bu profili otomatik olarak güncelleyerek konuşma tanıma doğruluğunu artırmaktadır. Sistem, ses verilerinden öznitelik çıkarmak için MFCC, enerji değeri, kepstral ortalama çıkarımı ve LDA (Doğrusal Ayırt Edici Analiz) yöntemlerini kullanmaktadır. Konuşma ve konuşmacı modellemesinde ise GMM tabanlı yarı-sürekli HMM (Gizli Markov Modelleri) kullanılmıştır. Konuşmacıya özel uyarlama için MAP, MLLR ve Eigenvoice yöntemleri uygulanmıştır. SPEECON veri seti ile yapılan

deneylerde, sistem konuşmacı kimliği bilindiğinde konuşma tanıma hatasını %25, konuşmacı kimliği bilinmediğinde ise %20 oranında azaltmıştır. Ayrıca, sistem konuşmacıları %94'ün üzerinde başarıyla tanımlamıştır.

Chen vd. (2020), konuşmacı tanıma problemine yönelik olarak SpeakerGAN adlı yeni bir model önermiştir. Bu model, Koşullu Üretici- Çekişmeli Ağ (Conditional Generative Adversarial Network- CGAN) yapısını temel almakta olup, yalnızca sahte gerçek ayrımı değil, aynı zamanda konuşmacının kimliğine ilişkin sınıflandırma işlemini de eş zamanlı olarak gerçekleştirebilmektedir. Modelin mimarisi iki ana bileşenden oluşmaktadır. Üretici ağ (Generator), ses verilerini modellemek amacıyla Kontrollü CNN (Gated CNN) yapısı kullanılarak tasarlanmıştır. Ayrıştırıcı ağ (Discriminator) ise, daha güçlü ayrım yeteneği sunan modifiye edilmiş Residual bloklara (ResNet) dayalıdır. Öğrenme sürecinde kullanılan kayıp fonksiyonu ise üç temel bileşeni içermektedir: Sınıflandırma kaybı, Adversaryal kayıp ve Huber kaybı. Bu çok bileşenli kayıp yapısı, hem üretilen örneklerin çeşitliliğini artırmakta hem de modelin öğrenme sürecinde daha kararlı bir yapı sergilemesini sağlamaktadır. Modelin başarısı, LibriSpeech veri seti üzerinde yapılan deneylerle değerlendirilmiştir. SpeakerGAN, yapılan testlerde %98.20 doğruluk oranı ile üstün bir performans sergilemiştir.

Jing vd. (2014), tarafından gerçekleştirilen bu çalışmada, konuşmacı tanıma sistemlerinin doğruluk oranını artırmak ve hesaplama karmaşıklığını azaltmak amacıyla hibrit bir öznitelik çıkarım ve sınıflandırma yaklaşımı sunulmuştur. Çalışmada, konuşma sinyalinden elde edilen LPCC ve MFCC öznitelikleri ile bu özniteliklerin birinci dereceden türevleri (delta LPCC, delta MFCC) birleştirilerek zengin bir öznitelik kümesi oluşturulmuştur. Oluşturulan öznitelik vektörleri, boyut indirgeme amacıyla PCA yöntemi ile dönüştürülmüş; ardından çerçeve sayısını azaltmak amacıyla K-means kümeleme algoritması uygulanmıştır. Sınıflandırma işlemi ise VQ (Vector Quantization) yöntemiyle gerçekleştirilmiştir. Çalışmada kullanılan veri seti, araştırmacılar tarafından laboratuvar ortamında oluşturulmuş özel bir ses kümesidir. Veri seti, farklı dil içeriklerinde (İngilizce, Mandarin, Guangxi lehçesi) kaydedilen cümleleri içermekte olup iki ayrı test grubundan oluşmaktadır. Elde edilen deneysel sonuçlara göre, MFCC öznitelikleri LPCC'ye kıyasla daha yüksek tanıma doğruluğu sağlamış; özniteliklerin birleştirilmesi ve PCA ile sadeleştirilmesi sonucunda doğruluk oranı %96,35'e ulaşmıştır. Ayrıca K-means sonrası

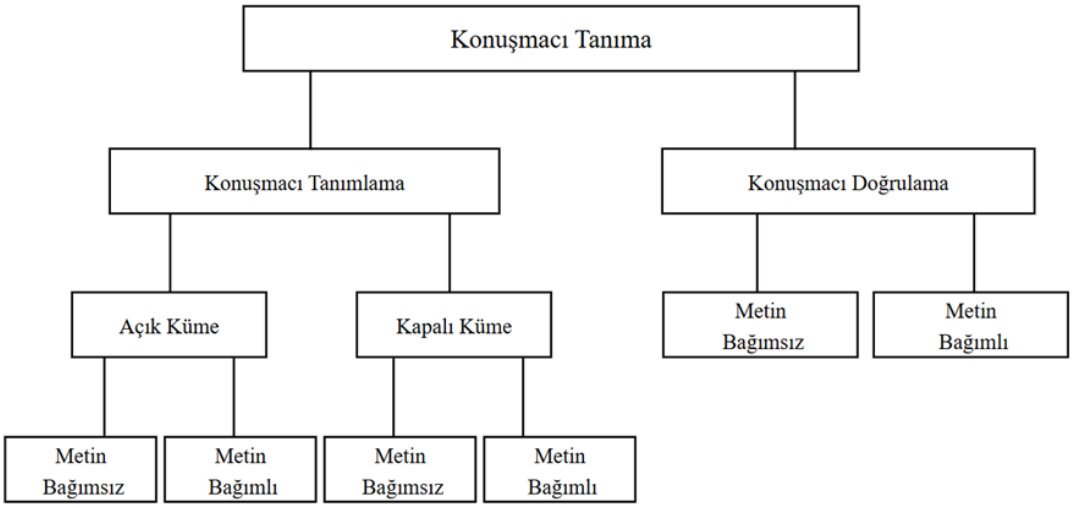
hesaplama yükü azaltılmış ve sistemin gerçek zamanlı uygulamalara uygunluğu artırılmıştır.

1.3. Konuşmacı Tanıma Sistemleri

Günümüzde sesli verilerle çalışan teknolojilerin yaygınlaşması, konuşmacı tanıma sistemlerini hem araştırma hem de uygulama açısından önemli bir alan hâline getirmiştir. Konuşmacı tanıma, bir ses kaydının hangi kişiye ait olduğunu belirlemeyi amaçlayan biyometrik bir tanıma yöntemidir. Bu sistemler, kişinin sesine özgü fiziksel özelliklerle tanımayı gerçekleştirir. Konuşmacı tanıma sistemleri genel olarak iki şekilde ele alınır: Konuşmacı Doğrulama (Speaker Verification) ve Konuşmacı Tanımlama (Speaker Identification) (Karasartova, 2011; Günerhan, 2023). Konuşmacı doğrulama, bir kişinin ses örneğinin, o kişiye ait olup olmadığını belirlemeye yönelik işlem olmakla birlikte, konuşmacı tanımlama ise bir ses örneğinin önceden kayıtlı bir konuşmacı veri setinde yer alan kişilerden hangisine ait olduğunu tespit etmeyi amaçlar.

Konuşmacı tanıma sistemleri, konuşmanın içeriğine bağlılık durumuna göre metinden bağımlı (text dependent) ve metinden bağımsız (text independent) olmak üzere iki alt kategoriye ayrılmaktadır. Metinden bağımlı sistemlerde, konuşmacıdan önceden belirlenmiş belirli bir ifadeyi söylemesi beklenir. Bu durum, sistemin daha az belirsizlikle çalışmasını sağladığından, tanıma doğruluğu genellikle daha yüksektir. Buna karşın, metinden bağımsız sistemlerde konuşmacı, herhangi bir metni veya doğal konuşma içeriğini kullanabilir. Bu yapı, sistemin daha esnek ve gerçek hayata daha uygun olmasını sağlasa da, tanıma sürecinin karmaşıklığını artırmaktadır.

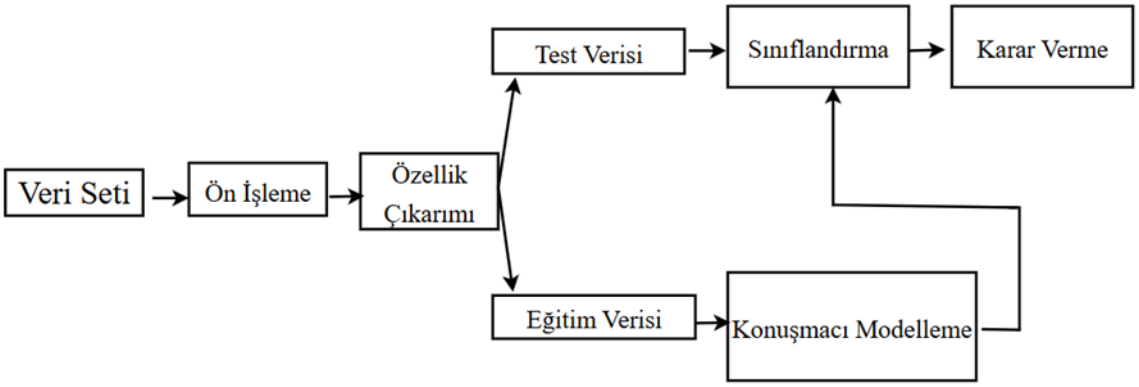
Öte yandan, tanımlanacak bireyin sistemde kayıtlı olup olmaması da sistemin yapısını doğrudan etkilemektedir. Bu bağlamda, konuşmacı tanıma sistemleri Kapalı küme (Closed set) ve Açık küme (Open set) olmak üzere iki şekilde sınıflandırılabilir. Kapalı küme sistemlerde, sesin mutlaka veri setinde kayıtlı bireylerden birine ait olduğu varsayılır ve sınıflandırma buna göre gerçekleştirilir. Açık küme sistemlerde ise, tanımlanmak istenen sesin veri setinde bulunmama olasılığı göz önünde bulundurulur. Bu durumda sistemin, hem eşleşen bir konuşmacıyı tespit edebilmesi hem de sesi bilinmeyen kişi olarak sınıflandırabilmesi beklenmektedir (Karasartova, 2011). Şekil 1’ de konuşmacı tanıma sistemlerinin genel yapısı verilmiştir.



Şekil 1. Konuşmacı tanıma sistemi genel yapısı

Konuşmacı tanıma sistemlerinin başarısı büyük ölçüde, ses sinyalinden elde edilen özniteliklerin doğruluğuna ve bu özniteliklerin karşılaştırılma yöntemine bağlıdır (Karasartova, 2011; Günerhan, 2023). Bu bağlamda, sistemin ilk aşamasında yer alan öznitelik çıkarımı (feature extraction) süreci, konuşmacıya özgü bilgilerin yoğun olarak bulunduğu kısa zamanlı akustik özelliklerin belirlenmesini amaçlar. Bu aşamada en yaygın kullanılan yöntemler arasında MFCC, LPCC ve Gammatone Frekanslı Kepstral Katsayılar (GFCC) tabanlı öznitelikler yer almaktadır (Karasartova, 2011; Günerhan, 2023; Kop & Bayındır, 2021; Geravanchizadeh & Ghalamiosgouei, 2020).

Elde edilen bu öznitelikler, konuşmacının ses özelliklerini sayısal bir vektör uzayında temsil eder. İkinci aşamada ise bu öznitelik vektörleri, sistemde kayıtlı konuşmacı modelleriyle karşılaştırılır. Böylece, tanınmak istenen ses örneği ile kayıtlı konuşmacılar arasında benzerlik ölçümü yapılarak sınıflandırma işlemi gerçekleştirilir. Konuşmacı tanıma sisteminin blok diyagramı Şekil 2'de gösterilmektedir.



Şekil 2. Konuşmacı tanımlama sistemi blok diyagramı

Konuşmacı tanıma sistemlerinde öznitelik çıkarımı ve modelleme aşamaları, sistemin genel performansı açısından kritik öneme sahiptir. Öznitelik çıkarımı aşamasında, ses sinyali üzerinde çeşitli işlemler gerçekleştirilerek konuşmacıya özgü ayırt edici özellikler elde edilir. Bu özellikler, sinyalin yapısal ve spektral karakteristiğini yansıtarak sınıflandırma süreci için temel veri kaynağını oluşturur. Modelleme aşamasında ise, çıkarılan öznitelikler makine öğrenmesi veya derin öğrenme algoritmaları aracılığıyla işlenerek sınıflandırma işlemi gerçekleştirilir. Bu iki aşamanın etkinliği, sistemin doğruluk oranı ile konuşmacılar arasındaki ayırt edicilik düzeyini doğrudan belirlemektedir.

1.4. Ön İşlemler

1.4.1. Çerçeveleme

Çerçeveleme (framing), konuşma sinyallerinin işlenmesinde temel bir adımdır ve özellikle öznitelik çıkarımı aşamasında kritik bir rol üstlenmektedir. Zamanla değişen yapıya sahip olan ses sinyali, doğrudan analiz edilemeyecek kadar karmaşık olduğundan, belirli uzunluklardaki kısa zaman aralıklarına bölünerek her bir parça ayrı ayrı analiz edilir. Bu işlem sonucunda, sinyal kısa süreli yapılar hâline getirilir ve her çerçeveden bağımsız olarak öznitelik çıkarımı gerçekleştirilebilir. Ayrıca, çerçeveler arasında belirli oranda örtüşme (overlap) uygulanması, sinyal geçişlerinin daha yumuşak bir şekilde işlenmesine olanak tanır (Başaran, 2007; Karasartova, 2011). Örneğin, ses sinyali çerçevelere ayrıldığında ilk çerçevenin N örnekten oluştuğu varsayılırsa, bir sonraki çerçevenin M<N

örnek kadar bir kısmı bir önceki çerçeveyle örtüşecek şekilde alınmalıdır (Kop , 2022). Bu yapı sayesinde hem zaman hem de frekans boyutunda daha kararlı bir temsil elde edilmekte ve böylece konuşma sinyalinin ayrıntılı özellikleri daha doğru şekilde yakalanabilmektedir.

1.4.2. Pencereleme

Pencereleme (windowing), çerçeveleme işleminden sonra uygulanan ve her bir çerçeve içerisindeki örneklerin uç kısımlarında meydana gelebilecek keskin geçişleri yumuşatmayı amaçlayan bir işlemdir. Çerçeveleme ile elde edilen her kısa süreli sinyal segmenti doğrudan işlenirse, çerçeve sınırlarında ani değişimler oluşabilir ve bu durum frekans analizinde spektral sızıntı olarak bilinen istenmeyen etkilere yol açabilir. Bu problemi azaltmak için her çerçeve, bir pencere fonksiyonu ile çarpılır ve böylece çerçeve başı ve sonu yumuşatılarak geçişlerin daha düzgün olması sağlanır (Karasartova, 2011; Sarı, 2017; Leu & Lin, 2017) . Pencereleme fonksiyonları genellikle sinyalin ortasını daha yüksek ağırlıkla alırken, uç kısımlara doğru bu ağırlık azalmaktadır. En yaygın kullanılan pencere fonksiyonlarından biri Hamming penceresi olup matematiksel ifadesi aşağıdaki formüldeki gibidir.

$$w(n) = 0.54 - 0.46\,cos\left(\frac{2\pi n}{N-1}\right),\ 0 \leq n \leq\ N-1 \qquad (1)$$

Hamming penceresi dışında Hanning, Blackman ve Rectangular (düz pencere) gibi farklı pencere türleri de yer almaktadır (Karasartova, 2011). Özellikle konuşmacı tanıma gibi alanlarda Hamming penceresi, frekans çözünürlüğü ile sızıntı arasında iyi bir denge sağlaması nedeniyle sıklıkla tercih edilmektedir.

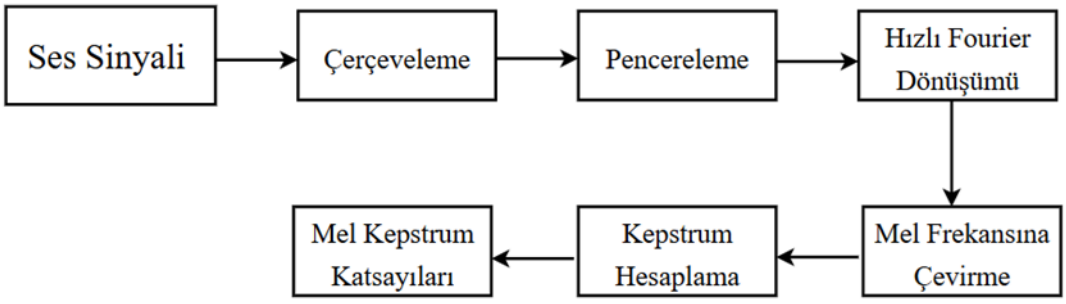
1.5. Öznitelik Çıkarımı

Öznitelik çıkarımı (Feature extraction) konuşmacı tanıma sistemlerinin temel yapı taşlarından biridir ve sistemin doğruluğunu doğrudan etkileyen kritik bir adımdır. Bu aşamada, ham ses sinyali belirli işlem adımlarından geçirilerek konuşmacıya özgü karakteristik bilgileri temsil eden öznitelikler elde edilir. Amaç, ses sinyalinden anlamsız bilgileri ayıklayarak, sınıflandırma algoritmaları için ayırt edici ve anlamlı bir temsil oluşturmaktır (Eskidere, 2007; Kop & Bayındır, 2021). Öznitelik çıkarımında sıklıkla kullanılan yöntemler arasında MFCC, LPCC, Perde (Pitch), formant frekansları ve enerji temelli ölçümler yer almaktadır. Bu öznitelikler, konuşmanın zamansal ve frekanssal

özelliklerini yansıtarak modelleme aşamasında sınıflandırma algoritmalarına girdi olarak sunulur. Başarılı bir öznitelik çıkarımı süreci, sistemin hem doğruluk oranını artırmakta hem de konuşmacılar arasındaki ayrım gücünü güçlendirmektedir.

1.6. Mel Frekans Kepstral Katsayılar

Mel Frekans Kepstral Katsayıları (MFCC), konuşma ve ses işleme alanında en yaygın kullanılan öznitelik çıkarım yöntemlerinden biridir. MFCC'nin temel amacı, ses sinyalinin spektral özelliklerini insan kulağının algısına daha yakın bir biçimde modelleyerek konuşmacıya özgü ayırt edici özellikleri ortaya çıkarmaktır (Eskidere, 2007; Karasartova, 2011). Bu yöntem, insan işitme sisteminin frekansları doğrusal değil, logaritmik olarak algıladığı bilgisine dayanır ve bu nedenle frekans eksenini Mel ölçeği olarak bilinen algısal bir ölçekte yeniden düzenler. Şekil 3' te MFCC yönteminin blok diyagramı gösterilmektedir.



Şekil 3. MFCC yöntemi blok diyagramı

Şekil 3'te de gösterildiği üzere, MFCC öznitelikleri çıkarılırken ilk olarak ham ses sinyali belirli zaman dilimlerine ayrılır. Her bir çerçevenin başında ve sonunda oluşabilecek ani geçişleri yumuşatmak ve spektral sızıntıyı azaltmak amacıyla pencereleme işlemi uygulanır. Ardından, her bir çerçevenin frekans bileşenlerini elde etmek için Hızlı Fourier Dönüşümü (FFT) uygulanır ve böylece zamana bağlı sinyal, frekans alanına dönüştürülmüş olur. Elde edilen frekans spektrumu, insan işitme sisteminin frekansa duyarlılığını modelleyen Mel filtre bankası üçgen şeklinde düzenlenmiş filtrelerle işlenir. Bu filtrelerden elde edilen çıktılara logaritmik ölçek uygulanarak sesin enerji dağılımı

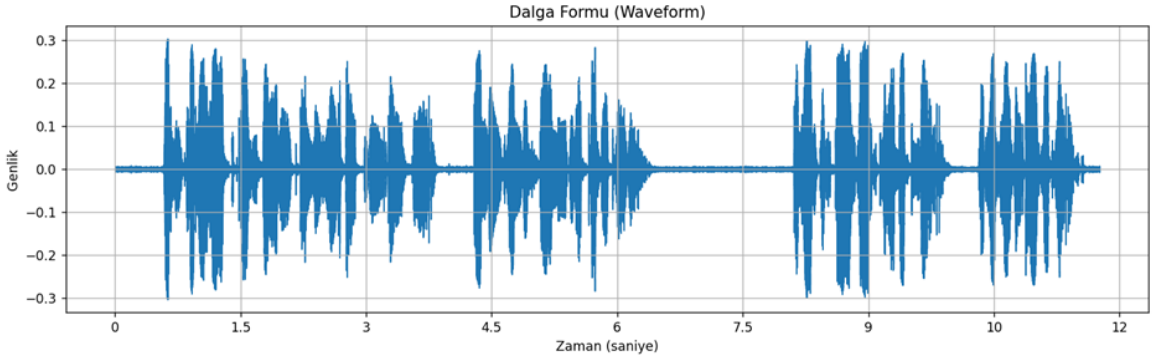
logaritmik ölçekte temsil edilir ve bu adım, insan kulağının ses şiddetine verdiği logaritmik tepkiyi yansıtmaktadır. Son aşamada ise, Log- Mel spektrumuna Ters Hızlı Fourier Dönüşümü (IFFT) uygulanarak zaman düzleminde özetlenmiş kepstral katsayılar, elde edilir. Bu katsayılar, ses sinyalinin spektral yapısını etkili bir biçimde temsil ederek konuşmacıya özgü bilgilerin çıkarılmasını sağlar (Karasartova, 2011; Günerhan, 2023).

1.6.1. Hızlı Fourier Dönüşümü

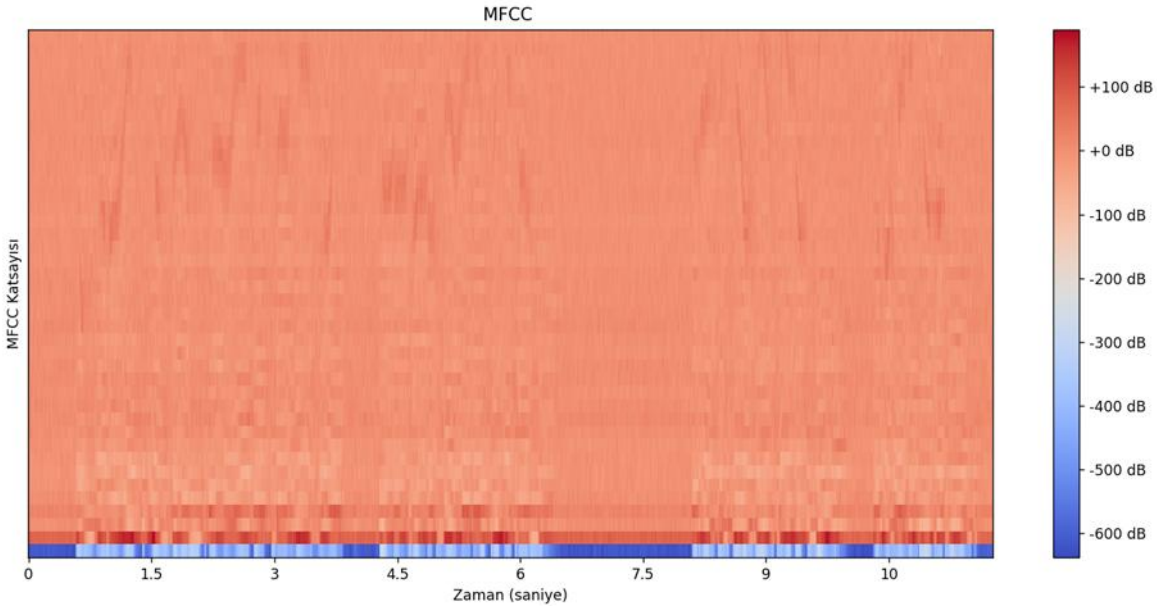
Hızlı Fourier Dönüşümü, zaman domainindeki bir sinyalin frekans domainine dönüşümünü sağlar (Fasounaki, 2021). Bu dönüşüm, sinyalin her bir çerçevesinin frekans bileşenlerini analiz ederek, özellikle ses ve konuşma sinyalleri üzerinde önemli bilgiler elde edilmesini sağlar. FFT'nin temel amacı, bir sinyalin zamanla değişen frekans bileşenlerini ortaya koymaktır ve N tane örnek içeren bir set için FFT aşağıdaki gibi hesaplanır (Başaran, 2007).

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi_j kn/N} \quad , n = 0,1,2 \dots \dots . N-1 \tag{2}$$

Frekans domainine geçiş, sinyalin farklı bileşenlerini daha iyi anlamamıza yardımcı olur. Genellikle, logaritmik büyüklük alınarak spektrum elde edilir. Bu spektrum, ses sinyallerinin zaman içindeki frekans yoğunluğunu görsel olarak gösterir ve genellikle spektrogram adı verilen üç boyutlu bir grafikle temsil edilir (Fasounaki, 2021). Bu spektrogram, konuşma tanıma ve sesle ilgili diğer görevlerde önemli bilgiler sunar. Aşağıda Şekil 4'te örnek ses sinyali ve Şekil 5'te bu sinyale ait spektogram görüntüsü verilmiştir.



Şekil 4. Örnek ses sinyali dalga formu



Şekil 5. Örnek ses sinyalinin spektogram görüntüsü

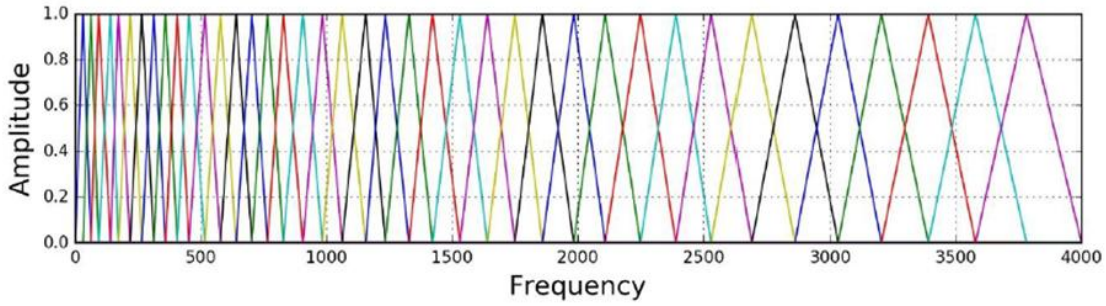
1.6.2. Mel Frekansına Çevirme

Mel frekansına çevirme aşaması, ses sinyali işleme sürecinde, frekansların insan kulağının duyma algısına daha yakın bir şekilde temsil edilmesi için kullanılır (Karasartova, 2011; Yujin vd., 2010). Mel frekansları, doğrusal frekans ölçeğinden farklı olarak, düşük frekansları daha hassas bir şekilde, yüksek frekansları ise daha az hassas bir şekilde temsil eder. Bu dönüşüm, frekans domainindeki bilgiyi, insan kulağının daha iyi algılayabileceği bir formatta sunmak için uygulanır. Mel frekansı dönüşümü aşağıdaki gibi hesaplanmaktadır.

$$f_{mel} = 2595^{*}\, log(1 + f\ /\ 700) \quad (3)$$

Mel frekansına dönüşüm, klasik frekans ölçeğinden Mel frekans ölçeğine geçişi sağlayan önemli bir adımdır. Mel frekansına dönüştürülmüş frekanslar, insan kulağının algılama şekline daha yakın bir temsil sunduğundan, ses işleme ve konuşma tanıma sistemlerinde yaygın olarak kullanılır . Bu dönüşüm, genellikle logaritmik bir büyüklük alarak, frekans bileşenlerinin daha anlamlı bir şekilde ifade edilmesini sağlar. Ardından, kepstral katsayılar (MFCC) hesaplamadan önce, Mel frekans ölçeğinde yerleştirilmiş filtre bankaları kullanılarak bu dönüşüm gerçekleştirilir.

Mel filtre bankası, Mel frekans ölçeğinde yerleştirilmiş ve her biri üçgen bant geçiren filtrelerden oluşan bir yapıdır. Bu filtre bankası, Mel frekansı aralığına bağlı olarak sabit aralıklarla yerleştirilmiş filtreler sunar. Filtreler, %50 oranında örtüşen ve sabit bant genişliğine sahip olan üçgen formlarını içerir. Aşağıda Şekil 6’ da Mel frekans filtre bankasının görseli verilmiştir.



Şekil 6. Mel frekans filtre bankası (Günerhan, 2023)

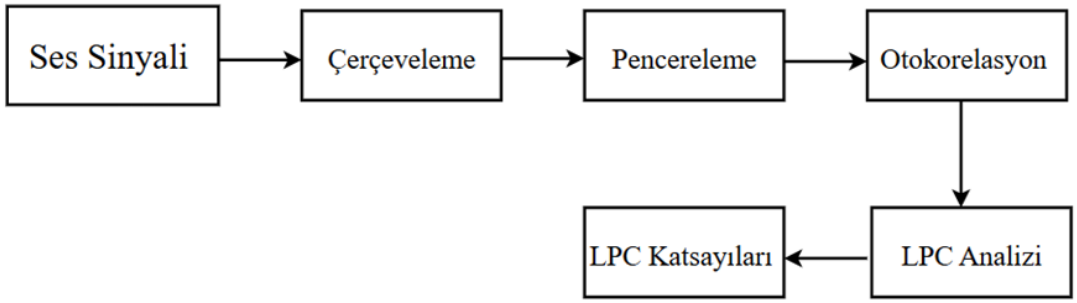
Her bir frekans bileşeni Mel ölçeğinde daha hassas bir şekilde temsil edilerek düşük frekanslı bileşenleri, yüksek frekanslı bileşenlerden ayırır. Bu işlem, ses sinyalinin daha doğru bir şekilde algılanmasını ve anlamlı özelliklerin çıkarılmasını sağlar. Bu filtre bankası, logaritmik büyüklükle birleştirilerek, sesin kepstral özelliklerinin elde edilmesine olanak tanır.

1.6.3. Kepstral Katsayıları Hesaplama

MFCC hesaplama sürecinde, Mel frekans filtresi üzerinden elde edilen logaritmik spektrum, zaman alanında kullanılabilmesi için Ters Hızlı Fourier Dönüşümü (IFFT) uygulanır (Kop, 2022; Karasartova, 2011). Dönüşüm sonucu oluşan katsayılara kepstral katsayıları denir ve bu katsayılarla öznitelik vektörü oluşturulur.

1.7. Doğrusal Öngörülü Kepstral Katsayılar

Doğrusal Öngörülü Kepstral Katsayılar, ses sinyalleri üzerinde öznitelik çıkarımı yapmak için kullanılan bir yöntemdir. LPCC, ses dalgasının zaman domainindeki verilerinden, belirli bir aralıkta doğrusal olarak tahmin edilen kepstral katsayıları çıkarır. Bu katsayılar, sesin akustik özelliklerini temsil eder ve genellikle konuşmacı tanımlama gibi uygulamalarda kullanılır. Yöntem temel olarak, Doğrusal Öngörülü Kodlama (LPC) katsayılarına Fourier dönüşümü uygulanmasıyla kepstral katsayıları elde etme işlemine dayanır (Kop, 2022). LPCC yöntemi çerçeveleme, pencereleme, otokorelasyon ve LPC analizi aşamalarından oluşur ve Şekil 7’ de LPCC yönteminin blok diyagramı verilmiştir.



Şekil 7. LPCC yöntemi blok diyagramı

LPCC yöntemi uygulanırken Şekil 7’de gösterildiği gibi öncelikle ham ses sinyaline çerçeveleme ve pencereleme işlemi uygulanır. Ardından pencerelenmiş her ses sinyaline otokorelasyon analizi yapılır. Daha sonra LPCC'nin temel bileşeni olan LPC, ses sinyalinin spektrumunu modellemek için doğrusal öngörülü analiz kullanır (Labied & Belangour, 2021). Burada, her çerçeve için doğrusal bir model ile sinyalin gelecekteki örnekleri tahmin edilir. LPC, genellikle sesin formant frekanslarını modellemek için kullanılır. LPC

ile elde edilen doğrusal öngörülü katsayılar Fourier dönüşümü yapılarak kepstral katsayılar elde edilir (Kop, 2022).

1.7.1. Otokorelasyon Analizi

Otokorelasyon, bir zaman serisinin kendisiyle olan ilişkisini ölçen bir yöntemdir ve özellikle sinyal işleme ile zaman serisi analizlerinde yaygın olarak kullanılır. Bu analiz, özellikle periyodik yapıları ve sinyaldeki düzenli tekrarlamaları tespit etmek için kullanılır. Pencerelenmiş ses sinyalinin her çerçevesine otokorelasyon analizi yapılır ve aşağıda verilen denklemdeki gibi hesaplanır (Karasartova, 2011).

$$r_l(m) = \sum_{n=0}^{N-1-m} \hat{x}_l(n) \cdot \hat{x}_l(n+m)\,, m = 0{,}1 \ldots., p \tag{4}$$

Denklemde verilen p değeri LPC analizinin derecesi olup genellikle 8 ile 12 arasında değer seçilerek oluşturulur (Kop , 2022).

1.7.2. Doğrusal Öngörülü Kodlama Analizi

Doğrusal Öngörülü Kodlama Analizi, konuşma sinyalinin n. örneğinin, kendinden önceki p tane örnekle doğrusal kombinasyonu şeklinde hesaplanmasıdır (Karasartova, 2011; Kawakami vd., 2014). Aşağıda, her çerçeveye ait otokorelasyondan LPC parametresi hesaplaması için kullanılan denklem verilmiştir.

$$\hat{s}(n) = \sum_{i=1}^{p} a_i \cdot s(n-i) \tag{5}$$

1.7.3. Doğrusal Öngörülü Kepstral Katsayılar Oluşturma

LPCC oluşturulurken, LPC katsayılarının kepstral katsayılara dönüştürülmesi işlemi gerçekleştirilir ve bu katsayılar aşağıdaki denklemdeki gibi hesaplanır (Karasartova, 2011).

$$c_n = a_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a_k c_{n-k} \ , 1 < n \leq p \tag{6}$$

Bu dönüşüm sayesinde, LPC katsayılarından doğrusal öngörülü özelliklerinin kepstral domaininde daha anlamlı bir şekilde temsil edilmesini sağlar. LPCC'ler, genellikle sesin temel özelliklerini modellemede kullanılır ve özellikle konuşmacı tanıma gibi alanlarda faydalıdır. Ancak, bu yöntem gürültülü ortamlarda daha az verimli olabilir ve bu nedenle, LPCC'nin performansını iyileştirmek için gürültü azaltma teknikleri veya daha gelişmiş modelleme yaklaşımları kullanılabilir (Labied & Belangour, 2021).

1.8. Makine Öğrenmesi

1.8.1. Denetimli Öğrenme

Denetimli öğrenme (Supervised Learning), makine öğrenmesinin en yaygın kullanılan ve en iyi bilinen yöntemlerinden biridir. Bu öğrenme türü, her bir giriş örneğiyle ilişkili olarak doğru çıktı etiketlerinin verildiği, etiketli verilerle yapılan öğrenme sürecini ifade eder. Yani, denetimli öğrenme algoritmaları, veri setindeki her örneği ve bu örneklerin doğru sonuçlarını kullanarak modelini eğitir. Eğitim sürecinde model, her bir veriyi ve onun doğru çıktısını öğrenerek, öğrenilen fonksiyonu kullanarak yeni, daha önce görmediği verileri doğru bir şekilde tahmin etmeye çalışır (Somvanshi vd., 2016; Turan & Polat, 2024). Bu tür öğrenme, özellikle sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılır (Pandey vd., 2019). Sınıflandırma, bir örneği belirli bir kategoriye atamak için kullanılırken; regresyon, bir sayısal değeri tahmin etmek için kullanılır. Denetimli öğrenme algoritmaları, verilen verilerden doğru çıkarımlar yaparak, yeni örneklerin sınıf etiketlerini veya sayısal değerlerini tahmin edebilir (Günerhan, 2023). KNN ve SVM, denetimli öğrenme için başlıca kullanılan algoritmalardandır.

Denetimli öğrenme algoritmalarının en önemli hedeflerinden biri, eğitim verileriyle sınırlı kalmayıp, eğitim sürecinden öğrenilen bilgiyi yeni veriler üzerinde de doğru bir şekilde uygulayabilmektir ve bu yetenek genelleme olarak bilinir. Genelleme kabiliyeti, modelin yalnızca eğitim verilerine değil, daha önce karşılaşmadığı verilere de doğru tahminlerde bulunabilmesi anlamına gelir (Somvanshi vd., 2016). Bu, denetimli

öğrenmenin en güçlü yönlerinden biridir, çünkü doğru bir şekilde genellenmiş bir model, çok geniş bir veri setinde etkili olabilir. Denetimli öğrenme algoritmaları, sağlık, güvenlik, finansal ve konuşmacı tanıma gibi alanda sıklıkla kullanılmaktadır (Günerhan, 2023).

1.8.2. Denetimsiz Öğrenme

Denetimsiz öğrenme (Unsupervised Learning), makine öğrenmesinin bir diğer önemli dalıdır ve etiketlenmemiş verilerle çalışarak verilerdeki gizli yapıları keşfetmeye yönelik bir yaklaşım sunar. Bu tür öğrenmede, modelin doğru sonuçları öğrenebilmesi için verilerdeki benzerlikleri, kalıpları veya yapıları bulmak için verilerin kendisini analiz eder. Denetimsiz öğrenme, genellikle verilerin kümeleme (clustering) gibi analiz yöntemlerini içeren bir süreçtir (Turan & Polat, 2024; Somvanshi vd., 2016). Kümeleme verileri benzer özelliklerine göre gruplamak için kullanılan bir tekniktir. Bu yöntem, genellikle pazarlama stratejileri gibi alanlarda kullanılır. Denetimsiz öğrenme, ayrıca Anomalik tespit (Anomaly detection) alanında da yaygın olarak kullanılır. Burada, algoritmalar normal veri örüntülerini öğrenerek, bu örüntülere uymayan anormal durumları tespit eder. Bu teknik, finansal dolandırıcılık tespiti, siber güvenlikte anormal etkinliklerin izlenmesi veya sağlık alanında anormal hastalık bulgularının erken tespiti gibi uygulamalarda etkilidir (Karakuş, 2021).

Bir diğer önemli kullanım alanı da özellik çıkarımı (feature extraction) alanıdır. Verilerdeki önemli özellikler, denetimsiz öğrenme teknikleri ile keşfedilir ve bu özellikler, veri setini daha iyi anlamak için veya başka algoritmalar için giriş olarak kullanılabilir (Günerhan, 2023). K-Means Kümeleme, GMM, PCA gibi algoritmalar denetimsiz öğrenme alanında sıklıkla kullanılmaktadır.

1.8.3. Yarı Denetimli Öğrenme

Yarı denetimli öğrenme (Semi Supervised Learning), makine öğrenmesinin önemli bir dalı olup, hem etiketli hem de etiketlenmemiş verilerle eğitim yapmayı amaçlayan bir öğrenme türüdür. Bu yaklaşım, etiketlenmiş verilerin sınırlı olduğu ancak etiketlenmemiş büyük miktarda verinin bulunduğu durumlarda kullanılır. Yarı denetimli öğrenmede model, sınırlı sayıdaki etiketlenmiş verilerden öğrenmeyi gerçekleştirerek, etiketlenmemiş verilerin yapılarını keşfedip, öğrenmeye devam etmektedir (Turan & Polat, 2024).

Etiketlenmemiş verilerin kullanımı, modelin genelleme yeteneğini artırarak, daha geniş veri kümesine dayalı doğru tahminler yapılmasını sağlar. Yarı denetimli öğrenme, özellikle büyük veri analizi ve derin öğrenme alanlarında büyük bir öneme sahiptir (Günerhan, 2023).

1.9. Sınıflandırma Yöntemleri

1.9.1. K- En Yakın Komşu Algoritması

K- En Yakın Komşu algoritması, denetimli öğrenme algoritmalarından biri olup, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılmaktadır (Turan & Polat, 2024). Bu algoritma, temel olarak, verilen bir veri noktasının sınıfını, en yakınındaki k komşusunun sınıfı ile belirler (Sugiharto vd., 2022; Kop & Bayındır, 2021). KNN, basit yapısı ve kolay uygulanabilirliği ile özellikle veri analizi ve makine öğrenmesi çalışmalarında tercih edilmektedir (Keskin, 2018).

KNN algoritmasının çalışma prensibi, öncelikle her bir veri noktasını, eğitim veri setindeki diğer noktalarla karşılaştırarak mesafe ölçümleri yapmaktır. Bu mesafe, genellikle Öklidyen, Manhattan, Minkowski mesafesi gibi metriklerle hesaplanır ve sırasıyla aşağıdaki denklemlerdeki gibidir (Hatipoğlu, 2017; Meral, 2008).

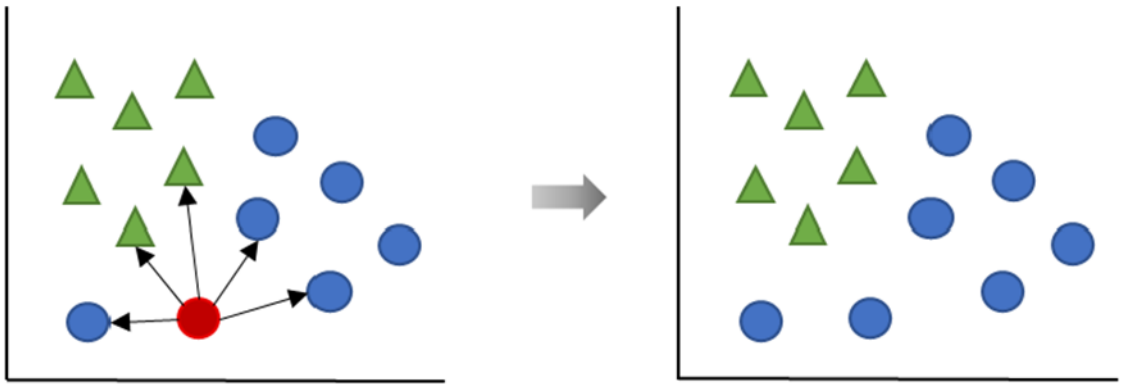
$$d_E = \sqrt{\sum_{i=0}^{N} \left(x_i - y_j\right)^2} \tag{7}$$

$$d_M = \sum_{i=0}^{N} |x_i - y_i| \tag{8}$$

$$d_m = \left(\sum_{i=0}^{n-1} |x_i - y_i|^p\right)^{\frac{1}{p}} \tag{9}$$

En yakın k komşu belirlendikten sonra, sınıflandırma problemlerinde, bu komşuların çoğunluk sınıfı, test verisinin sınıfını belirlemektedir. KNN algoritması, parametre olarak k

değerine bağlıdır. k değeri, komşuların sayısını belirler ve bu parametre, modelin başarısını doğrudan etkiler. k değeri küçük olduğunda model, gürültüye duyarlı hale gelebilir ve aşırı uyum (overfitting) yapabilir. Büyük bir k değeri ise, modelin daha genel olmasına neden olabilir, ancak doğru sınıflandırmayı zorlaştırabilir. Bu nedenle, k değerinin seçimi genellikle Çapraz doğrulama (Cross-validation) yöntemleriyle optimize edilir. KNN algoritmasının k, beş komşuluğu için örnek görseli Şekil 8'de verilmiştir.



Şekil 8. KNN en yakın beş komşuluk (Kop & Bayındır, 2021)

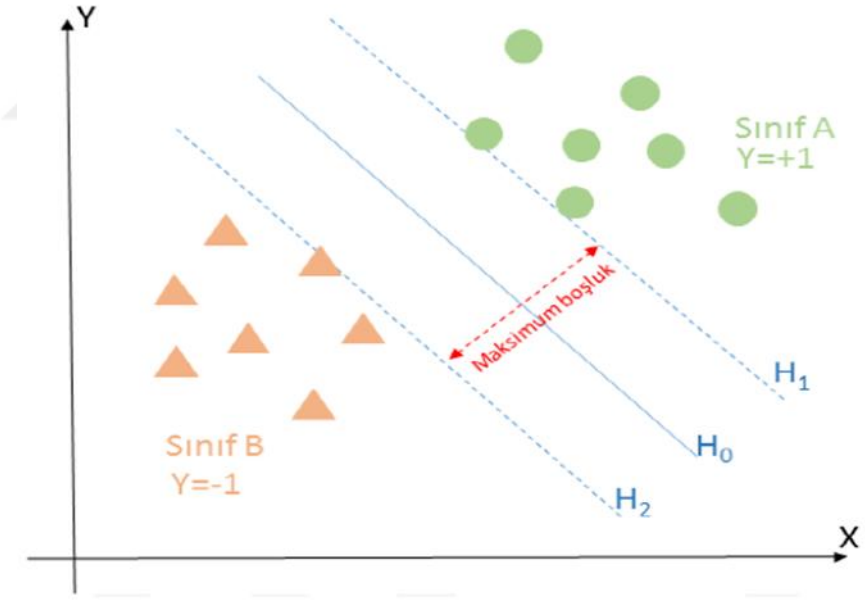
Algoritma Şekil 8'de de belirtildiği üzere, k için beş komşuluk seçildiğinde yeni gelen kırmızı örnek için diğer en yakın beş örnekle yakınlığı hesaplanır. Yeni gelen kırmızı örnek, hangi sınıf fazla ise o sınıfa atanır.

Eğitim aşamasında, algoritma yalnızca verileri saklar ve herhangi bir model parametrelerini öğrenmez. Ancak, test aşamasında her yeni örnek için eğitim veri setindeki her noktaya mesafe hesaplamak gereklidir. Bu durum, büyük veri setlerinde KNN'nin performansını önemli ölçüde düşürebilir (Akgün, 2024; Kacur, 2016). Bu yüzden, veri setinin büyüklüğü ile algoritmanın performansı arasında ters bir ilişki vardır ve bu problemi hafifletmek için çeşitli iyileştirmeler yapılabilir.

KNN algoritmasında, özetle öncelikle k adet yakın komşu belirlenir. Ardından yeni gelen örnekle, diğer örnekler arasındaki mesafeler hesaplanır ve uzaklıklar sıralanır. En yakın k tane komşu bu uzaklıklara göre belirlenir ve sınıf ataması gerçekleştirilir.

1.9.2. Destek Vektör Makineleri

Destek Vektör Makineleri, denetimli öğrenme algoritmalarından biri olup, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan güçlü bir yöntemdir (Mokgonyane vd., 2019). SVM' nin temel amacı, verileri sınıflandırırken iki sınıf arasındaki en geniş sınırı (margini) bulmaktır (Akgün, 2024; Turan & Polat, 2024). Bu algoritma, özellikle doğrusal olmayan problemlerde de etkili olabilen, yüksek boyutlu verilerde bile başarıyla çalışabilmektedir (Öztürk, 2024). SVM, verilerin n boyutlu bir uzayda, doğrusal ayrılabilen iki sınıfa ait örnekleri en iyi şekilde ayıran bir hiperdüzlem bulmaya çalışır. Bu hiperdüzlem, iki sınıfın arasındaki en geniş boşluğu (marjini) sağlamak üzere seçilir. Şekil 9' da doğrusal olarak ayrılabilen verilerde SVM' nin sınıflandırma grafiğini gösteren örnek verilmiştir ve iki sınıf arasındaki en geniş sınırı bulan fonksiyon H_0 hiper düzlem fonksiyonudur.



Şekil 9. Doğrusal ayrılabilen verilerde SVM sınıflandırma grafiği (Öztürk, 2024)

Sınır genişliğini maksimum yapan H_0 hiper düzlem fonksiyonu, n tane özellik için ve W ağırlık vektörüyle denklemdeki gibi hesaplanmaktadır.

$$H_0 = \sum_{i=1}^{n} W_i \cdot X_i + b = 0 \quad (10)$$

Sınıflardaki örnekler -1 ile +1 arasında etiketlenir ve H_1 ile H_2 üzerindeki örnekler destek vektörleri olarak adlandırılır. H_1 ve H_2 hiper düzlem denklemleri sırasıyla aşağıdaki gibidir.

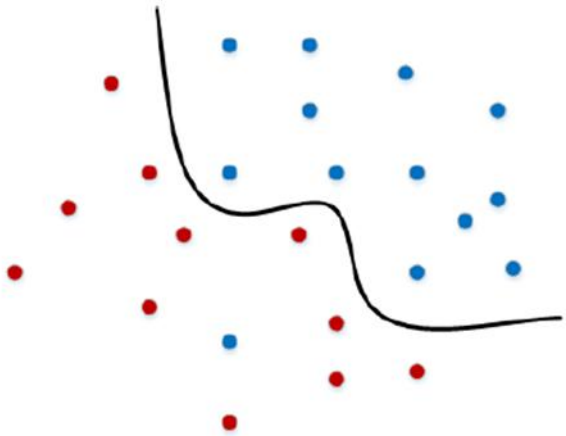
$$H_1 = W^T \cdot X + b = +1 \quad (11)$$

$$H_2 = W^T \cdot X + b = -1 \quad (12)$$

Destek vektörler arasındaki mesafeyi maksimum yapmak için W ağırlık vektörü değeri minimum seçilmelidir ve H_1 ile H_2 hiper düzlemleri arasındaki d mesafesi denklemdeki gibi elde edilmektedir.

$$d = \frac{|W^T \cdot X + b|}{\|w\|} \quad (13)$$

Veriler doğrusal olarak ayrılamıyorsa, SVM bu doğrusal olmayan durumları ele almak için çekirdek (kernel) fonksiyonları kullanır (Turan & Polat, 2024; Akgün, 2024). Çekirdek fonksiyonları, veriyi daha yüksek boyutlu uzaya taşıyarak doğrusal olmayan sınıflandırma problemlerini çözme yeteneği sağlar (Somvanshi vd., 2016). Şekil 10' da doğrusal ayrılamayan verilerin görsel örneği verilmektedir.



Şekil 10. Doğrusal ayrılmayan veri örneği (Hatipoğlu, 2017)

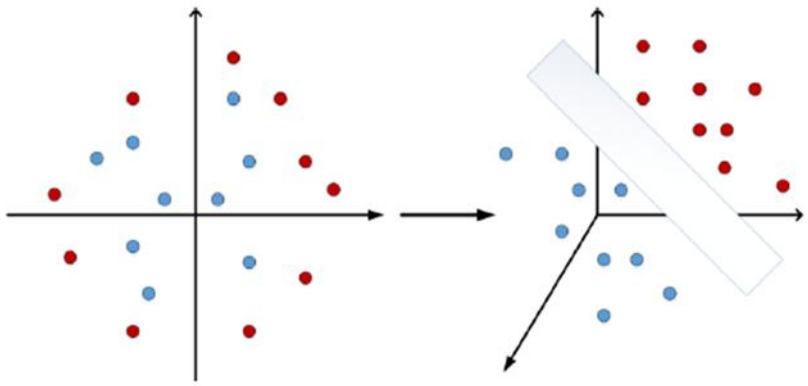
Literatürde Lineer, Polinomsal ve Radyal Tabanlı Çekirdek fonksiyonları en yaygın kullanılan çekirdek fonksiyonlarındandır ve sırasıyla aşağıdaki gibi elde edilmektedir (Eray, 2008).

$$K(u,v) = v^T \cdot u \tag{14}$$

$$K(u,v) = (v^T \cdot u + 1)^n \tag{15}$$

$$K(u,v) = exp\left(-\frac{\|u-v\|^2}{2\sigma^2}\right) \tag{16}$$

Çekirdek fonksiyonu ve uygun parametre seçimi büyük bir öneme sahip olup, yanlış çekirdek seçimi modelin başarısını olumsuz etkileyebilmektedir. Yüksek boyutlu uzaya taşınarak doğrusal hale getirilen verilerin görseli Şekil 11'de verilmiştir.



Şekil 11. Doğrusal ayrıştırılması sağlanan yüksek boyutlu uzay görseli (Hatipoğlu, 2017)

Sonuç olarak SVM, hem doğrusal hem de doğrusal olmayan sınıflandırma problemleri için etkili bir çözüm sunmaktadır. Özellikle yüksek boyutlu verilerle çalışırken, SVM'in genelleme kapasitesi ve çekirdek fonksiyonları, onu güçlü bir sınıflandırıcı yapmaktadır. Bununla birlikte, büyük veri setlerinde hesaplama maliyetleri ve parametre ayarları gibi zorluklar, uygulama süreçlerini karmaşık hale getirebilir. Bu

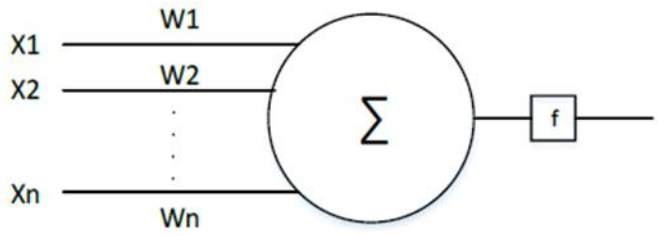
sebeple, doğru çekirdek fonksiyonu seçimi ve parametre iyileştirmesi, SVM uygulamalarının başarısı için kritik öneme sahiptir (Hatipoğlu, 2017).

Bu algoritma, özellikle konuşmacı tanıma, metin sınıflandırma ve yüz tanıma gibi alanlarda yüksek doğruluk oranları elde etmesiyle bilinir ve birçok uygulamada başarılı sonuçlar elde etmiştir (Somvanshi vd., 2016).

1.10. Yapay Sinir Ağları

Yapay sinir ağları (Artificial Neural Network), biyolojik sinir sistemlerinin çalışma prensiplerinden ilham alınarak geliştirilen bir yapay zeka modelidir. YSA, insan beynindeki sinir hücrelerinin (nöronlar) bir araya gelerek bilgi ilettiği yapıyı taklit eder (Karabina, 2017) (Caner & Üstün, 2006). Temelde, veri girişi ile bir dizi ağırlıklı bağlantılar üzerinden sinyallerin işlendiği, sonuçların ise çıkış katmanında elde edildiği bir yapıyı içerir. Bu ağlar, özellikle büyük veri setleri üzerinde öğrenme ve karar verme yetenekleri ile dikkat çekmektedir. YSA'nın temel bileşeni nöronlar olup, her biri bir giriş değeri alır, belirli bir işlem yapar ve bu işlemin sonucunu bir sonraki katmana aktarır.

Nöronlar arasındaki bağlantılar, genellikle ağırlıklı olarak belirlenir ve bu ağırlıklar ağın öğrenme sürecini optimize etmek için ayarlanır (Hanilçi, 2007; Karabina, 2017). Şekil 12'de n girişli bir nöronun genel gösterimi verilmiştir.



Şekil 12. n girişli nöronun genel gösterimi (Karabina, 2017)

Aktivasyon fonksiyonları, yapay sinir ağlarında, bir nöronun giriş sinyallerini çıkış sinyallerine dönüştürmek için kullanılan matematiksel fonksiyonlardır. Bu fonksiyonlar, ağın doğrusal olmayan ilişkileri öğrenmesini sağlamak ve modelin karmaşık problemlere çözümler geliştirebilmesini temin etmek için kritik öneme sahiptir. Yapay sinir ağlarında

kullanılan her nöron, gelen giriş sinyallerini belirli bir ağırlıkla çarpar, ardından bu toplamı aktivasyon fonksiyonundan geçirir. Aktivasyon fonksiyonu, nöronun çıktı değerini hesaplamak için bu toplamı dönüştürür (Günerhan, 2023). Aktivasyon fonksiyonları, ağın genel performansını, doğruluğunu ve öğrenme hızını doğrudan etkileyebilir. Literatürde sıkça kullanılan aktivasyon fonksiyonları Tablo 1' deki gibi listelenmiştir.

Tablo 1. Sıkça kullanılan aktivasyon fonksiyonları

Doğrusal Fonksiyon	$y = kx$
Sigmoid	$y = \frac{1}{1+e^{x}}$
Tanjant Hiperbolik	$y = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$
RELU	$y = \begin{cases} x, & x > 0 \\ O, & x \leq 0 \end{cases}$
Softmax	$y = \frac{e^{x_i}}{\Sigma_j e^{x_j}}$

Aktivasyon fonksiyonlarının özelliklerini inceleyecek olursak, Sigmoid fonksiyonu giriş değerini 0 ile 1 arasında bir değere döndürürken, Tanjant hiperbolik fonksiyonu çıkış değerini -1 ile 1 arasında sınırlar. RELU fonksiyonu, girişin sıfırdan büyük olması durumunda doğrusal bir çıktı üretirken, sıfırdan küçük olan değerleri sıfırlayan bir fonksiyondur. Softmax fonksiyonu ise, her bir sınıf için bir olasılık değerleri elde eder ve bu olasılıkların toplamı 1'dir (Karabina, 2017; Günerhan, 2023).

Sonuç olarak, aktivasyon fonksiyonları, sinir ağlarının öğrenme yeteneğini belirleyen önemli bir bileşendir. Her bir fonksiyonun avantajları ve dezavantajları bulunmakta olup, modelin eğitim sürecini optimize etmek için doğru seçimler yapılmalıdır.

Yapay sinir ağları, derin öğrenme alanında önemli bir rol oynamaktadır. Derin öğrenme, çok katmanlı sinir ağlarının kullanılması ile daha karmaşık özelliklerin ve ilişkilerin öğrenilmesini mümkün kılar. Bu, özellikle görüntü tanıma, ses tanıma ve doğal dil işleme gibi alanlarda büyük başarılar elde edilmesine olanak sağlamıştır. Derin öğrenme yöntemlerinde kullanılan ağlar arasında CNN ve Tekrarlayan Sinir Ağları (RNN) yer alır.

Yapay sinir ağları, genelleme yetenekleri sayesinde, önceki verilerde öğrenilen bilgiyi yeni, görülmemiş verilere uygulayabilme kapasitesine sahiptir. Bu da onları gerçek dünya problemleri için güçlü bir araç haline getirir. Ancak YSA'ların başarısı, verinin kalitesine, ağın mimarisine, eğitim verisinin çeşitliliğine ve doğru parametrelerin seçilmesine bağlıdır.

1.11. Derin Öğrenme

Derin öğrenme (Deep Learning), yapay zeka ve makine öğrenmesi alanında, insan beynine benzer şekilde verilerden otomatik olarak özellik çıkarımı yapabilen ve öğrenme süreçlerini derin katmanlar üzerinden gerçekleştiren bir tekniktir (Çolakoğlu vd., 2021). Derin öğrenme, özellikle büyük veri setleri ve güçlü işlem gücü kullanarak çok katmanlı yapılarla öğrenmeyi mümkün kılar. Bu yöntem, geleneksel makine öğrenmesi algoritmalarından daha derin ve karmaşık ilişkileri öğrenebilir ve genellikle daha yüksek başarı elde edilir. Derin öğrenme, adını sinir ağlarının çok sayıda gizli katmana sahip olmasından alır. Bu ağlar, daha karmaşık özellikleri ve verilerin soyut temsilini öğrenmek için çok sayıda ardışık katman kullanır (Kurbanov, 2018; Cengil & Çınar, 2016).

Derin öğrenme, birçok farklı mimariyle çeşitli problemlere çözüm olmaktadır. Çok Katmanlı Algılayıcılar, Konvolüsyonel Sinir Ağları ve Tekrarlayan Sinir Ağları mimarileri en yaygın kullanılan sinir ağı mimarileridir.

Derin öğrenme, veri biliminden yapay zekaya kadar birçok alanda devrim niteliğinde yenilikler sunmaktadır. Görüntü tanıma, sesli yanıt sistemleri, otonom araçlar gibi günlük hayatımıza kullandığımız birçok uygulamanın temelinde derin öğrenme yer alır. Bu teknoloji, özellikle büyük veri ve güçlü hesaplama altyapıları ile desteklendiğinde yüksek doğrulukla sonuçlar verebilir. Ancak, büyük veri setleri ve güçlü işlemciler gerektirdiği için bazı zorluklarla da karşı karşıya kalınmaktadır. Yine de, derin öğrenme hızla gelişmeye devam eden ve gelecekte daha geniş bir uygulama alanına sahip olacak bir teknolojidir (Kurbanov, 2018).

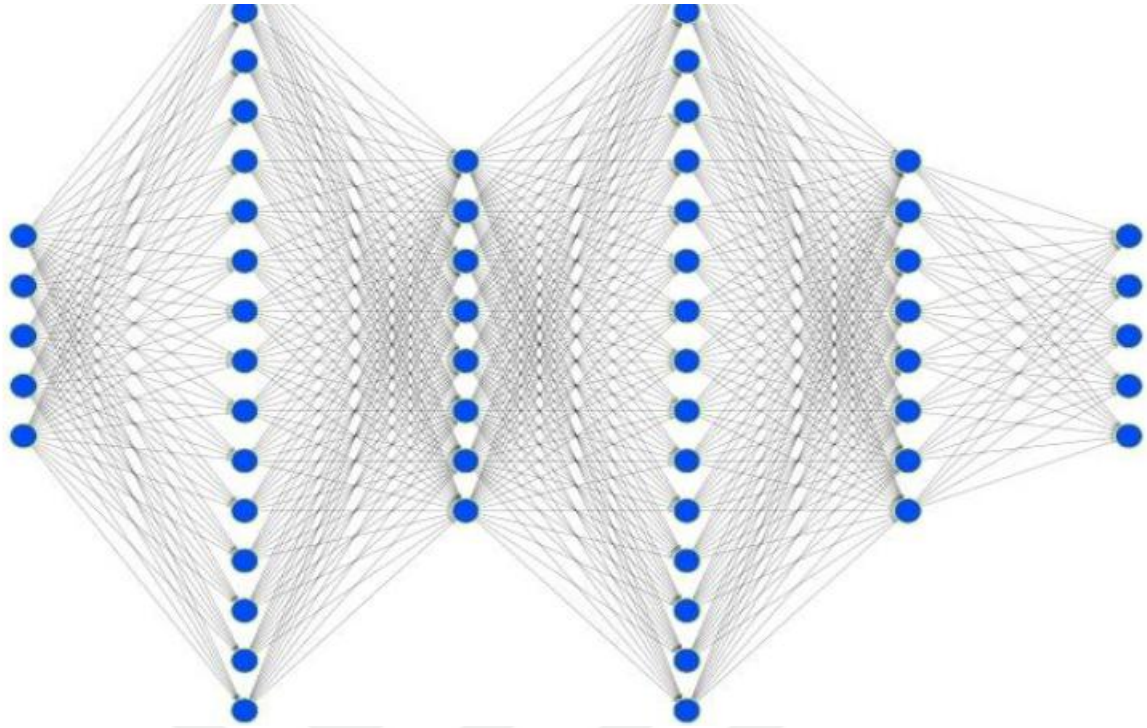
1.11.1. Konvolüsyonel Sinir Ağları

Konvolüsyonel Sinir Ağları, özellikle görsel verilerle ilgili problemlerde yüksek başarılar elde eden derin öğrenme mimarilerindendir. Görüntü işleme alanındaki başarıları, CNN' leri bu alanda en çok tercih edilen yöntemlerden biri yapmaktadır. CNN' ler, yapay sinir ağlarının bir üyesi olup, daha spesifik olarak görsel verilerdeki uzamsal bağımlılıkları ve özellikleri öğrenebilmek için tasarlanmış bir yapıya sahiptir (Çolakoğlu vd., 2021). Konvolüsyonel sinir ağları, büyük ve karmaşık görüntü verilerinin etkin bir şekilde işlenmesine olanak tanıyan çok katmanlı yapılarıyla dikkat çeker. Bu ağlar, özellikle görüntü işleme, nesne tanıma, yüz tanıma, tıbbi görüntü analizi, konuşmacı tanıma ve video analizleri gibi uygulamalarda yaygın olarak kullanılır (Ye & Yang, 2021).

CNN, genellikle görsel verilerle ilişkilendirilse de, ses verisinin işlenmesinde de etkili bir şekilde kullanılmaktadır. Ses sinyalleri, zamanla değişen özelliklere sahip sürekli bir veri kaynağı sunduğundan, CNN' ler bu verilerdeki karmaşık yapıları ve özellikleri öğrenmek için uygundur.

Ses sinyalleri genellikle zaman frekans alanında analiz edilir. Bu süreç, sesin frekans bileşenlerinin görselleştirilmesiyle gerçekleştirilir. Literatürde, ses sinyali ve konuşma tanıma için zaman frekans temsillerinin kullanımı yaygındır. Spektrogramlar olarak bilinen görsel temsiller, ses sinyalinin frekans bileşenlerini zamanla görselleştirir ve bu temsiller, sesle ilgili önemli bilgileri içerir (Zhang vd., 2022). CNN, ses verisini işleme ve konuşmacı tanıma gibi görevlerde yaygın bir şekilde kullanılmaktadır.

Literatürde, CNN'lerin konuşmacı tanıma sistemlerinde, sesin frekans zaman temsillerindeki yerel değişimleri öğrenerek yüksek performans gösterdiği belirtilmiştir (Hershey vd., 2017). Liu & Leu (2018), CNN'lerin ses verilerinde görsel temsillerle nasıl etkileşimde bulunduğunu inceleyerek, bu yöntemlerin konuşmacı tanıma sistemlerine uygulanmasının ne kadar önemli olduğunu vurgulamaktadır. Şekil 13' te 5 giriş, 5 çıkış düğümü ve 4 adet gizli katman bulunan CNN mimarisinin genel gösterimi verilmiştir.



Şekil 13. CNN mimarisi düğümler ve bağlantılar (Günerhan, 2023)

CNN'ler, farklı katmanlardan oluşan bir yapıya sahiptir. Bu katmanlar, görsel verinin farklı düzeylerde özelliklerini öğrenir ve işleme sürecinde verilerin özellik temsillerini oluşturur (Lukic vd., 2016; Ashar vd., 2020). CNN mimarisinin giriş katmanında, genellikle ses sinyallerinin önemli özelliklerinin görsel temsili olan spektogram görüntüleri bu katmanda giriş olarak kullanılır.

Ardından konvolüsyonel katmanda filtreler aracılığıyla giriş görüntüsündeki veya sesin spektogramındaki önemli özellikler çıkarılır. Ses sinyalinin zaman frekans temsilleri, genellikle sabit boyutlarda filtreler ile işlenir ve her filtre, sesin belirli bir kısmına odaklanarak, kenarları, doku desenlerini veya frekans bileşenlerini öğrenir.

Aktivasyon katmanında, konvolüsyonel katmandan elde edilen çıktıya bir aktivasyon fonksiyonu uygulanır. CNN'lerde yaygın olarak genellikle RELU aktivasyon fonksiyonu kullanılır. RELU, negatif değerleri sıfıra çevirirken, pozitif değerleri olduğu gibi bırakır. Bu işlem, doğrusal olmayan özelliklerin öğrenilmesine olanak tanır ve öğrenme sürecinin hızlanmasını sağlar (Bunrit vd., 2019).

Havuzlama katmanı, veriyi boyut küçültmek ve özellikleri daha kompakt hale getirmek için kullanılır. Konvolüsyonel katmandan elde edilen Özellik haritaları (Feature

maps) üzerinde max pooling gibi işlemler uygulanır. Bu katman, her bölgede en yüksek değeri alarak, görüntüdeki önemli özellikleri korur ve ağın hesaplama maliyetini azaltır.

Normalizasyon katmanında, batch normalization gibi normalizasyon teknikleri kullanılır. Bu katman, ağın öğrenme sürecini stabilize eder ve daha hızlı eğitim sağlar. Normalizasyon katmanı, girişlerin belirli bir aralıkta olmasını sağlar, böylece gradyanların aşırı büyümesini veya küçülmesini engeller. Bu, ağın her katmanındaki aktivasyonların dengesiz hale gelmesini engeller ve modelin doğruluğunu artırır.

Konvolüsyonel ve havuzlama katmanları, sesin özellik haritalarını çıkarmaktan sorumludur. Ancak, bu çıkarılan özelliklerin konuşmacı tanıma işlemine uygun hale gelmesi için tam bağlantılı katmanlar kullanılır. Tam bağlantılı katmanlar, ağın öğrendiği özelliklerin belirli bir konuşmacıya ait olup olmadığını sınıflandırmak için kullanılır. Bu katmanlar, daha önce çıkarılmış olan özellikleri alır ve sınıf etiketlerine dönüştürür.

Son olarak, ağın çıktısı, sesin hangi konuşmacıya ait olduğuna dair tahmin yapan bir sınıf etiketidir. Konuşmacı tanıma problemlerinde, bu çıktı genellikle konuşmacı kimliği veya konuşmacı numarası gibi değerler olabilir. Softmax fonksiyonu kullanılarak, her sınıf için olasılık değeri elde edilir. Bu, modelin hangi konuşmacının daha olası olduğunu belirlemesini sağlar (Fasounaki vd., 2021; Tiwari & Verma, 2024).

1.12. Performans Değerlendirme Metrikleri

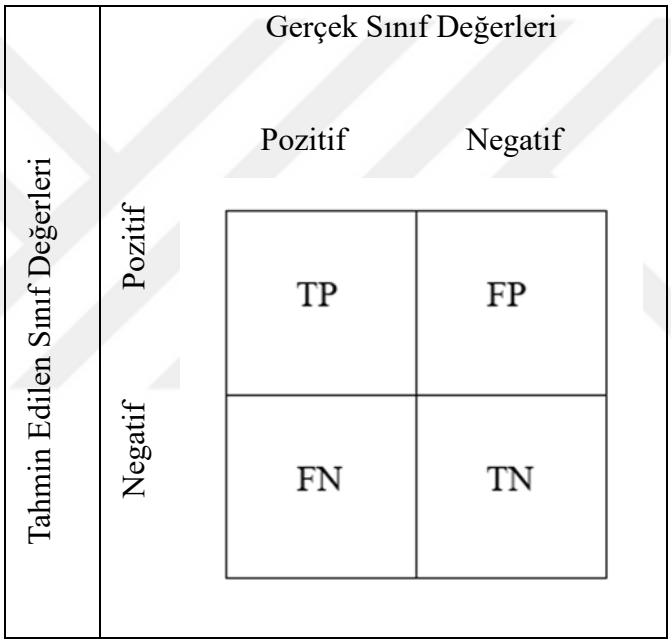
Performans değerlendirme metrikleri (Performance Evaluation Metrics), bir makine öğrenmesi modelinin performansını değerlendirmek için kullanılan ölçütlerdir. Bu metrikler, modelin doğruluğunu, güvenilirliğini ve verimliliğini analiz etmeye yardımcı olur. Konuşmacı tanıma, görüntü sınıflandırma veya regresyon gibi farklı türdeki problemler için kullanılabilecek çok sayıda başarı metriği bulunmaktadır. Her metriğin kendine özgü avantajları ile sınırlamaları vardır ve seçilecek başarı metrikleri, problemin hedeflerine göre değişir.

Literatürde yapılan deneylerin başarı performanslarını ölçmek için genellikle doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru (F1 score) gibi metrikler bulunmaktadır. Bu başarı metrikleri karmaşıklık matrisi (confusion matrix)

olarak adlandırılan yapılardan elde edilen değerlerle hesaplanmaktadır (Kop & Bayındır, 2021; Turan & Polat, 2024) .

Karmaşıklık matrisi, modelin performansını ayrıntılı bir şekilde ortaya koyar ve diğer performans metriklerinin de hesaplanmasına katkıda bulunur. Karmaşıklık matrisi hesaplanırken oluşturulan sınıflandırma sistemine verilen örneklerin ait olduğu sınıf ile sınıflandırıcının tahmin ettiği sınıfın karşılaştırılması yapılır (Turan & Polat, 2024). Tablo 2'de karmaşıklık matrisinin oluşturulması verilmiştir.

Tablo 2. Karmaşıklık matrisinin yapısı



Karmaşıklık matrisinde, TP (True Positive) ifadesi, tahmin edilen sınıfın pozitif ve doğru olduğu durumları belirtirken, TN (True Negative) ifadesi, tahmin edilen sınıfın negatif ve doğru olduğu durumları göstermektedir. Ayrıca, FP (False Positive) ifadesi, tahmin edilen sınıfın pozitif ve yanlış olduğu durumları belirtirken, FN (False Negative) ifadesi de tahmin edilen sınıfın negatif ve yanlış olduğu durumları göstermektedir (Kop & Bayındır, 2021).

Doğruluk (Accuracy), sınıflandırma problemlerinde en yaygın kullanılan başarı metriğidir. Bu metrik, modelin doğru tahmin ettiği örneklerin toplam örnek sayısına oranını ifade eder ve aşağıdaki denklemdeki gibi hesaplanır.

$$Doğruluk = \frac{TP+TN}{TP+FP+TN+FN} \quad (17)$$

Kesinlik (Precision), modelin doğru şekilde pozitif sınıfı tahmin etme oranını gösterir. Özellikle yanlış pozitif (false positive) tahminlerinin önemli olduğu durumlarda kullanılır ve aşağıdaki gibi hesaplanır.

$$Kesinlik = \frac{TP}{TP+FP} \quad (18)$$

Duyarlılık (Recall), modelin doğru şekilde pozitif sınıfı bulma oranını ölçer. Bu metrik, yanlış negatif (false negative) tahminlerinin önemli olduğu durumlarda kullanılır ve aşağıdaki şekilde hesaplanır.

$$Duyarlılık = \frac{TP}{TP+FN} \quad (19)$$

Son olarak F1 skoru (F1 Score), kesinlik ve duyarlılığın dengelenmesi gereken durumlarda kullanılır. Hem doğru pozitiflerin sayısını hem de yanlış pozitif ve yanlış negatiflerin etkilerini göz önünde bulundurur. F1 skoru, aşağıdaki gibi hesaplanır.

$$F1\ Skoru = 2 * \frac{Kesinlik*Duyarlılık}{Kesinlik+Duyarlılık} \quad (20)$$

2. YAPILAN ÇALIŞMALAR

Bu çalışmada, konuşmacı tanımlama alanına yönelik yenilikçi bir yöntem olarak Açı Genlik Dönüşümü (AGD) yaklaşımı geliştirilmiş ve ilk kez ses sinyallerine uygulanmıştır. Yöntemde, öncelikle konuşma sinyallerinin yerel değişim noktaları (tepe ve çukur noktaları) tespit edilmiştir. Her bir yerel ekstremum noktasından, belirli komşu noktalar kullanılarak açı ve genlik bilgileri hesaplanmış, bu bilgiler iki boyutlu bir düzleme aktarılarak AGD tabanlı görsel temsiller oluşturulmuştur.

Oluşturulan bu görsellerden, Yönlendirilmiş Gradyanların Histogramı (Histogram of Oriented Gradients) yöntemi ile öznitelikler çıkarılmış ve bu öznitelikler kullanılarak KNN ile SVM algoritmalarıyla sınıflandırma yapılmıştır. Ayrıca, elde edilen AGD görselleri doğrudan CNN mimarisine girdi olarak verilerek derin öğrenme tabanlı sınıflandırma da gerçekleştirilmiştir.

Yöntemin performansı, geniş çapta kullanılan LibriSpeech ve VoxCeleb1 veri setleri üzerinde test edilmiştir. Çalışmada AGD yöntemi, literatürde yaygın olarak kullanılan MFCC ve LPCC yöntemleri ile aynı koşullar altında karşılaştırılmış, aynı veri setleri ve sınıflandırma algoritmaları kullanılarak başarı metrikleri (doğruluk, kesinlik, duyarlılık, F1 skoru) değerlendirilmiştir.

Bunun yanı sıra, AGD yöntemi MFCC yöntemi ile hibrit olarak birleştirilmiş ve bu birleşik yaklaşımın performansı da analiz edilerek karşılaştırmalı sonuçlar elde edilmiştir. Yapılan deneysel çalışmalar, AGD yönteminin hem tek başına hem de hibrit biçimde kullanıldığında konuşmacı tanımlama alanında yüksek performans sergileyen, yenilikçi ve etkili bir alternatif olduğunu ortaya koymuştur.

2.1. Veri Setleri ve Özellikleri

2.1.1. LibriSpeech Veri Seti

LibriSpeech, konuşmacı tanımlama gibi alanlarda yaygın olarak kullanılan, açık kaynaklı ve yüksek kaliteli bir veri setidir. Kamuya açık LibriVox projesinden türetilen, İngilizce kitapların yüksek sesle okunması sonucu elde edilen uzun süreli konuşma

örneklerinden oluşmaktadır (Jahangir vd., 2020). LibriSpeech, konuşmacıların farklı konuşma tarzları, aksanları, konuşma hızları ve vurguları gibi bireysel özelliklerini içeren geniş bir veri yapısına sahiptir. Veri setinde yer alan konuşmacılar farklı yaş gruplarından ve cinsiyetlerden oluşmakta olup, bu çeşitlilik konuşmacı tanımlama sistemlerinin genelleme kabiliyetini değerlendirmek açısından önemli bir avantaj sunmaktadır.

Veri seti, 16 kHz örnekleme oranında, mono kanal formatında kaydedilmiş, .flac uzantılı seslerlerden oluşmakta ve ses dosyalarının yanı sıra, her kayda karşılık gelen metin transkripsiyonlarını da içermektedir. LibriSpeech'in açık erişimli, etiketli ve kapsamlı bir yapıya sahip olması, onu akademik araştırmalarda yaygın olarak kullanılan bir veri seti haline getirmiştir. Literatürde birçok çalışma, bu veri setini kullanarak konuşmacı tanımlama modellerinin başarısını değerlendirmiş ve farklı algoritmaların başarısını karşılaştırmalı olarak ortaya koymuştur (Jahangir vd., 2020; Fasounaki vd., 2021; Günerhan, 2023). LibriSpeech, konuşmacı tanıma modellemelerinin başarısını artırmak ve yöntemsel karşılaştırmalar yapmak açısından önemli bir kaynak niteliğinde olmaktadır.

2.1.2. VoxCeleb1 Veri Seti

VoxCeleb1, konuşmacı tanımlama görevleri için geliştirilmiş, çok konuşmacılı ve gerçek dünya koşullarında kaydedilmiş ses verilerinden oluşan büyük ölçekli bir veri setidir (Günerhan, 2023). YouTube platformundan otomatik olarak toplanarak konuşmacı doğrulama ve yüz tanıma gibi görüntü tabanlı algoritmalar kullanılarak oluşturulmuştur (Nagrani vd., 2017; 2020). VoxCeleb1, 1.251 farklı konuşmacıya ait 153.516 ses örneği içermekte olup, konuşmacılar farklı yaş, cinsiyet ve aksana sahip kişilerden oluşmaktadır. Bu çeşitlilik, konuşmacı tanıma modellerinin geniş bir yelpaze üzerinde değerlendirilmesini ve test edilmesini mümkün kılmaktadır (Günerhan, 2023; Fasounaki vd., 2021).

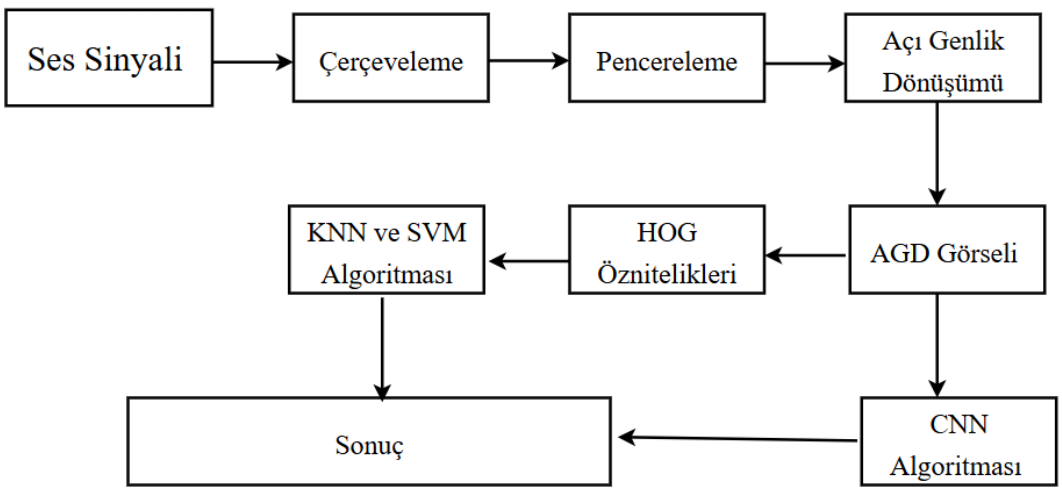
Veri seti, röportajlar, basın toplantıları ve stüdyo kayıtları gibi çeşitli ortamlardan elde edilmiş videolardan çıkarılan ses örneklerini içermektedir. Bu kayıtlar, arka plan gürültüsü, müzik, kahkahalar gibi gerçek dünya kaynaklı akustik durumları da bulundurmaktadır. VoxCeleb1 veri setindeki ses dosyaları .wav formatında, mono kanal olarak ve 16 kHz örnekleme oranıyla kaydedilmiştir. Ses örneklerinin süresi genellikle 4-15 saniye arasında değişmektedir.

Yapılan çalışmada, Librispeech ve Voxceleb1 veri setleri üzerinde deneyler gerçekleştirilmiştir. Librispeech veri setinde 20 kişinin yaklaşık yüz ses kaydı ele alınırken, Voxceleb1 veri setinde de 20 kişinin yaklaşık yirmi ile kırk arasında ses kayıtları kullanılmıştır. Her iki veri seti için de %80 eğitim, %20 test verisi olarak ayrılmıştır ve deneyler gerçekleştirilerek sınıflandırma sonuçları belirlenmiştir.

2.2. Açı Genlik Dönüşümü

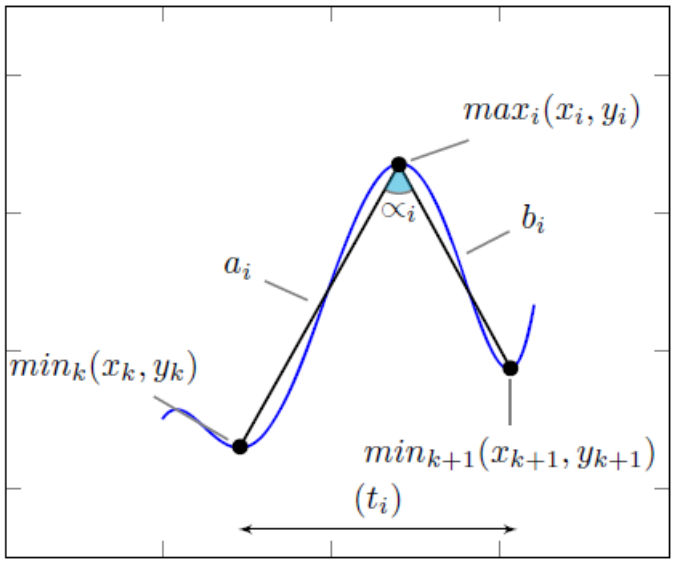
Bu çalışmada, literatürde yaygın olarak kullanılan MFCC ve LPCC gibi geleneksel öznitelik çıkarım yöntemleri yerine, ses sinyallerine açı genlik dönüşümü yöntemi uygulanmıştır. Sinyal üzerinde çerçeveleme ve pencereleme işlemi uygulanarak, her çerçeve için maksimum ve minimum noktalar arasındaki doğruların eğimleri ve bu doğruların uzunluklarına bağlı olarak hesaplanan açı ve normalize edilmiş genlik değerleri, açı genlik görüntüsü üzerinde konumlandırılmıştır.

Elde edilen bu görseller üzerinden öznitelik çıkarımı ve sınıflandırma işlemleri gerçekleştirilmiştir (Hatipoglu vd., 2019; 2020). AGD sayesinde, ses sinyalleri görsel uzaya dönüştürülerek, görüntü işleme yöntemleriyle etkili bir şekilde analiz edilebilir duruma gelmiştir. Şekil 14'te AGD yönteminin genel blok diyagramı verilmiştir.



Şekil 14. AGD yöntemi blok diyagramı

AGD yönteminde, öncelikle sinyalin tüm yerel maksimum ve minimum noktaları tespit edilmiştir. Daha sonra merkez nokta olarak seçilen bir maksimum ya da minimum nokta etrafında, bu merkeze zıt olmak üzere bir önceki ve bir sonraki değişim noktaları belirlenmiştir. Seçilen bu noktalar arasındaki öklidyen mesafeler hesaplanarak, bu mesafelerin oluşturduğu doğrular yardımıyla eğim değerleri elde edilmiştir. Doğrunun eğiminden yola çıkılarak açı değeri arctangent (arctan) fonksiyonu ile hesaplanmıştır. Aynı zamanda genlik değeri, merkez noktanın sol ve sağındaki mesafelerin karşılaştırılmasıyla normalleştirilmiş bir biçimde oluşturulmuştur. Böylece her değişim noktası için bir açı genlik çifti oluşturularak sinyalin görsel temsili elde edilmiştir. (Hatipoglu vd., 2019; 2020).



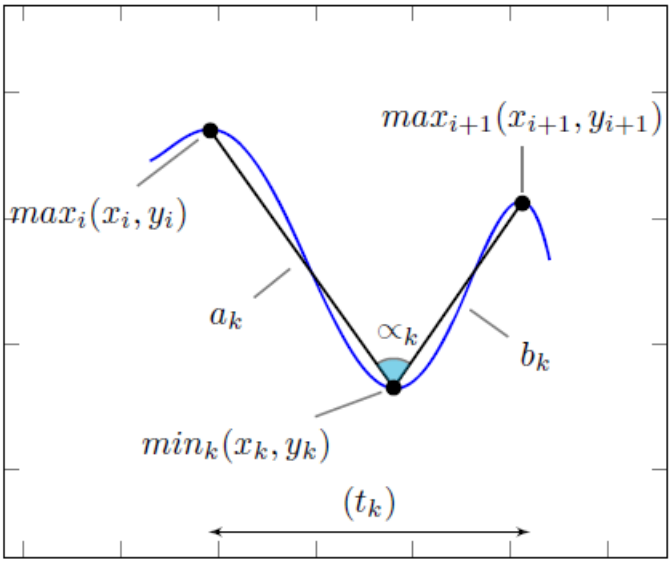
Şekil 15. Merkez seçilen maksimum nokta, bir önceki ve sonraki minimumlarla oluşturduğu doğrular (Hatipoglu vd., 2019)

Şekil 15'te verilen görselde, merkez seçilen max_i (x_i, y_i) noktası ile bu noktaya zıt bir önceki min_k (x_k, y_k) ve bir sonraki min_{k+1} (x_{k+1}, y_{k+1}) noktalarını ele alalım. t_i yerel maksimum bölgesinde min_k (x_k, y_k) ile max_i (x_i, y_i) noktaları arasında oluşan doğrunun uzunluğu a_i, max_i (x_i, y_i) ile min_{k+1} (x_{k+1}, y_{k+1}) noktaları arasında oluşan doğrunun uzunluğu b_i ve bu iki doğrunun arasındaki açı da α_i olsun. Bu merkez maksimum ve bir önceki minimum nokta arasındaki a_i öklidyen mesafesi denklemdeki gibi hesaplanmaktadır (Hatipoglu vd., 2019; 2020).

$$a_i(t_i) = \sqrt{\left((x_{i,k})^2 + (y_{i,k})^2\right)} \tag{21}$$

Merkez maksimum ve bir sonraki minimum nokta arasındaki b_i öklidyen mesafesi aşağıdaki denklemdeki gibidir.

$$b_i(t_i) = \sqrt{\left((x_{i,k+1})^2 + (y_{i,k+1})^2\right)} \tag{22}$$



Şekil 16. Merkez seçilen minimum nokta, bir önceki ve sonraki maksimumlarla oluşturduğu doğrular (Hatipoglu vd., 2019)

Benzer şekilde Şekil 16'da verilen görselde seçilen min_k (x_k, y_k) noktası ile bu noktaya zıt bir önceki max_i (x_i, y_i) ve bir sonraki max_{i+1} (x_{i+1}, y_{i+1}) noktalarını ele alalım. t_k yerel minimum bölgesinde max_i (x_i, y_i) ile min_k (x_k, y_k) noktaları arasında oluşan doğrunun uzunluğu a_k, min_k (x_k, y_k) ile max_{i+1} (x_{i+1}, y_{i+1}) noktaları arasında oluşan doğrunun uzunluğu b_k ve bu iki doğruların oluşturduğu açı da α_k olsun.

Merkez seçilen minimum ve bir önceki maksimum nokta arasındaki a_k öklidyen mesafesi denklemdeki gibi hesaplanmaktadır (Hatipoglu vd., 2019; 2020).

$$a_k(t_k) = \sqrt{\left((x_{k,i})^2 + (y_{k,i})^2\right)} \tag{23}$$

Merkez minimum ve bir sonraki maksimum nokta arasındaki b_k öklidyen mesafesinin hesaplanması aşağıdaki gibidir.

$$b_k(t_k) = \sqrt{\left(\left(x_{k,i+1}\right)^2 + \left(y_{k,i+1}\right)^2\right)} \tag{24}$$

Oluşan doğrular arasındaki açı değerleri doğru denkleminden eğimleri elde edilerek hesaplanmaktadır. $y_r = m_r x_r + c_r$ ve $y_{r+1} = m_{r+1}.xr_{+1} + c_{r+1}$ doğrularının eğimleri denklemdeki gibi elde edilmektedir.

$$m_r = \frac{y_{r+1} - y_r}{x_{r+1} - x_r} \tag{25}$$

Böylece elde edilen eğimlerden pozitif α açısı $(m_r - m_{r+1}) / (1 - m_r * m_{r+1})$ ifadesinin arctanjantı ile hesaplanır. Yine elde edilen negatif α açısı $(-1) * (m_r - m_{r+1}) / (1 - m_r * m_{r+1})$ ifadesinin arctanjantı ile hesaplanır. Bu şekilde pozitif ve negatif açı değerleri açısal formüldeki gibi belirlenir.

$$\alpha_k(t) = \begin{cases} (m_r - m_{r+1}) / \left(1 - (m_r * m_{r+1})\right), yerel\ maksimum\ ise \\ -(m_r - m_{r+1}) / \left(1 - (m_\Gamma * m_{r+1})\right), diğer\ durumlarda \end{cases} \tag{26}$$

Öte yandan normalleştirilmiş genlik hesabı için sol ve sağ mesafeleri kıyaslanır. Sinyal üzerinde her tepe veya çukur bölgeleri için büyük uzunluğa sahip olan kenar ile küçük uzunluğa sahip olan kenar değerleri oranlanmıştır. Bu şekilde normalleştirilmiş genlik değerlerinin hesaplanmasıyla hem her bölge için uzaklık değerleri belirlenmiş hem de bu değerler normalleştirilerek -1 ile +1 aralığında tutulmuştur (Hatipoglu, 2017).

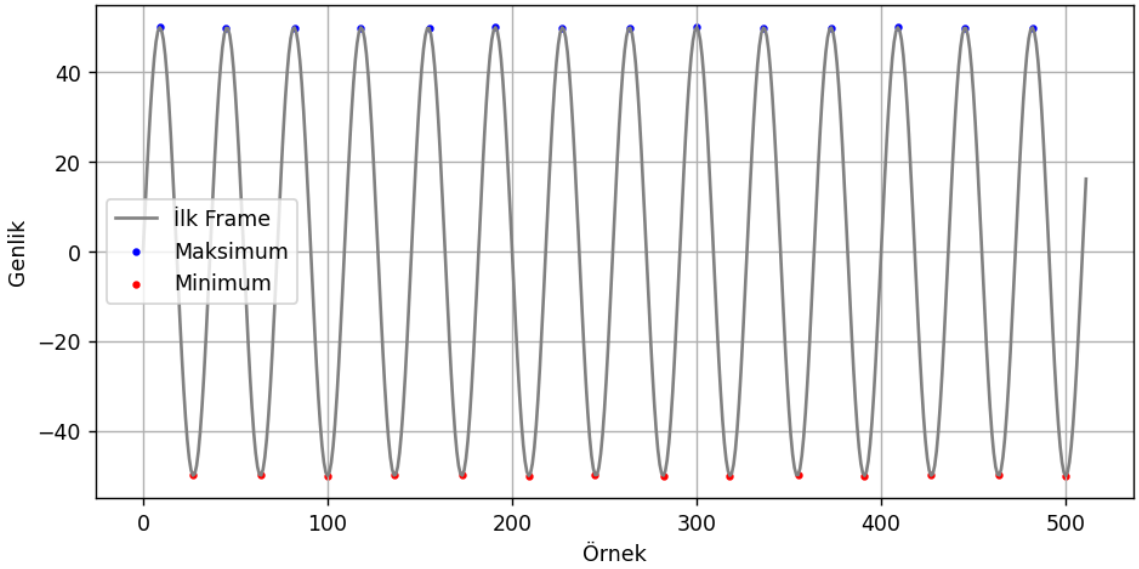
$$R_k(t) = \begin{cases} \frac{b_k(t)}{a_k(t)}, & b_k(t) < a_k(t) \\ \frac{-a_k(t)}{b_k(t)}, diğer\ durumlarda \end{cases} \tag{27}$$

Açı ve genlik değerlerinin hesaplanmasının ardından, elde edilen maksimum ve minimum noktalar açı genlik uzayında görselleştirilir. Bu görselleştirme sırasında, x ekseni normalleştirilmiş genlik değerlerini, y ekseni ise açı değerlerini temsil eder. Pozitif açı değerleri grafiğin üst kısmında, negatif açı değerleri ise alt kısmında yer alacak şekilde konumlandırılır. Benzer şekilde, pozitif genlik değerleri sağ tarafa, negatif genlik değerleri ise sol tarafa yerleştirilir (Hatipoglu vd, 2019; 2020).

İkinci dereceden açı genlik dönüşümünde de aynı işlem adımları ve görselleştirme süreci izlenmektedir. Ancak bu yöntemde, birinci derecede olduğu gibi merkez alınan

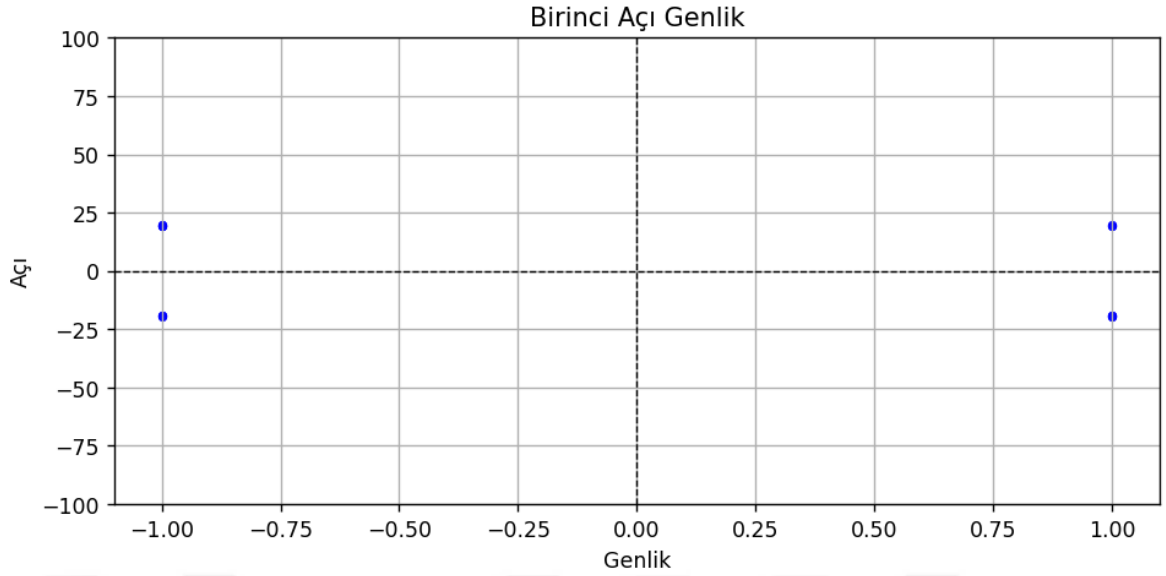
ekstremum (değişim) noktasının yalnızca bir önceki ve bir sonraki zıt türdeki ekstremum noktaları değil, iki önceki ve iki sonraki zıt ekstremum noktaları dikkate alınmaktadır (Hatipoglu vd, 2019; 2020). Böylece sinyalin yapısına ilişkin daha kapsamlı ve genelleştirilebilir özellikler elde edilmesi hedeflenmektedir.

Aşağıda Şekil 17'de örnek sinüs sinyali ile bu sinyalin maksimum ve minimum noktalarının gösterimi verilmiştir.



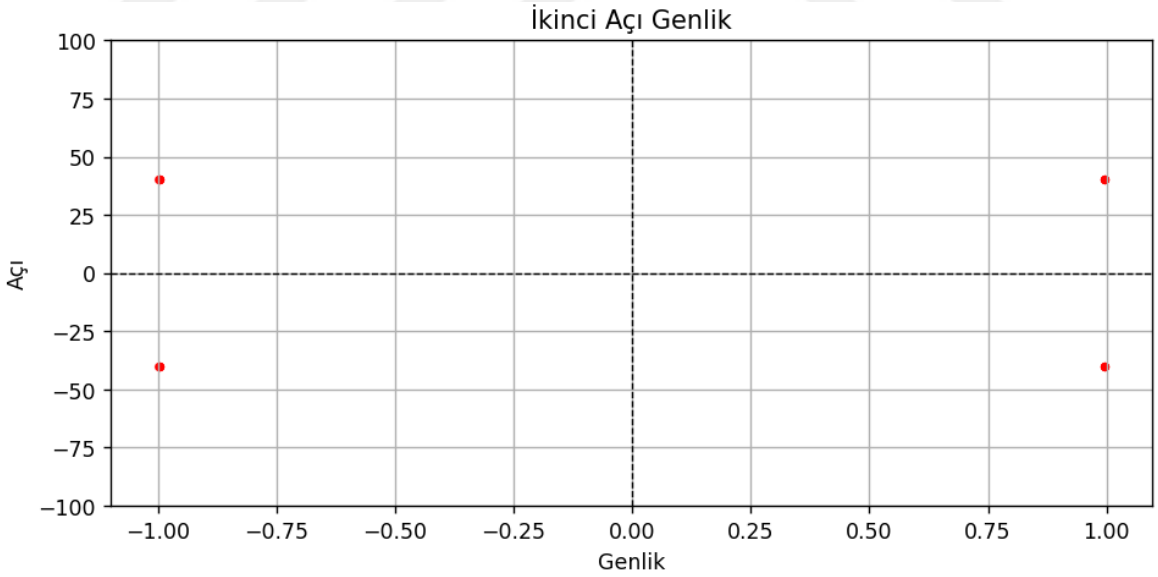
Şekil 17. Sinüs sinyali örneği ve maksimum- minimum noktalarının belirlenmesi

Sinüs sinyalindeki belirlenen maksimum ve minimum noktalarının, her ekstremum için birinci ve ikinci açı genlik değerleri hesaplanmıştır. Birinci açı genlik değerleri merkez seçilen ekstremum noktası için kendinden bir önceki ve bir sonraki zıt ekstremum noktasıyla elde edilmiştir. Sinüs sinyalinin birinci AGD görüntüsü Şekil 18' de verilmiştir.



Şekil 18. Sinüs sinyalinin birinci AGD görüntüsü

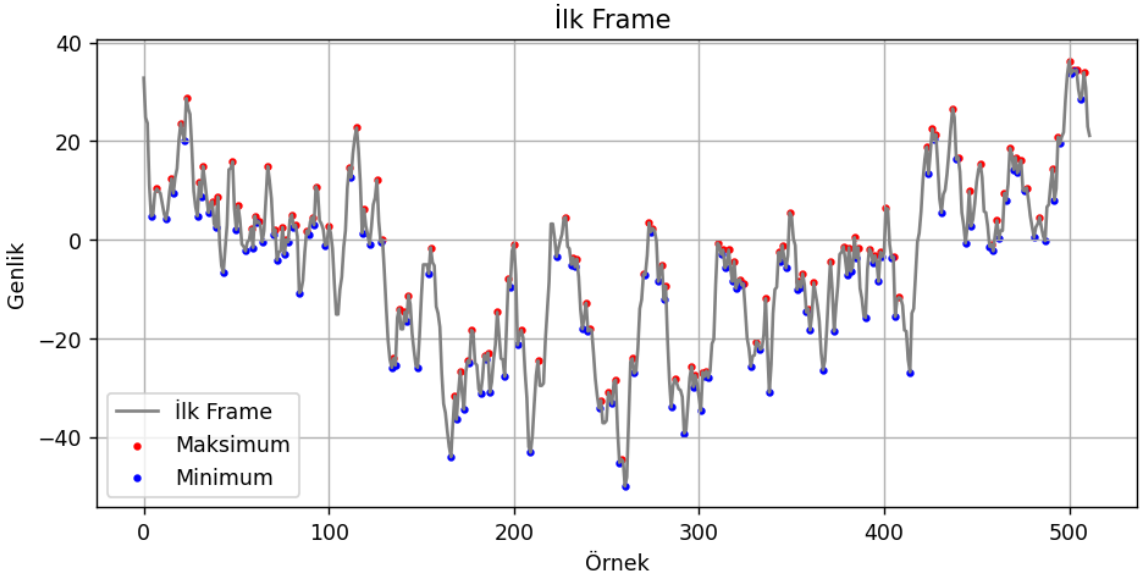
İkinci açı genlik değerleri de merkez seçilen ekstremum noktası için kendinden iki önceki ile iki sonraki zıt ekstremum noktalar ele alınarak elde edilir ve birinci açı genlik değerlerinde kullanılan işlem adımları uygulanır. Sinüs sinyalinin ikinci AGD görüntüsü de Şekil 19'da verildiği gibidir.



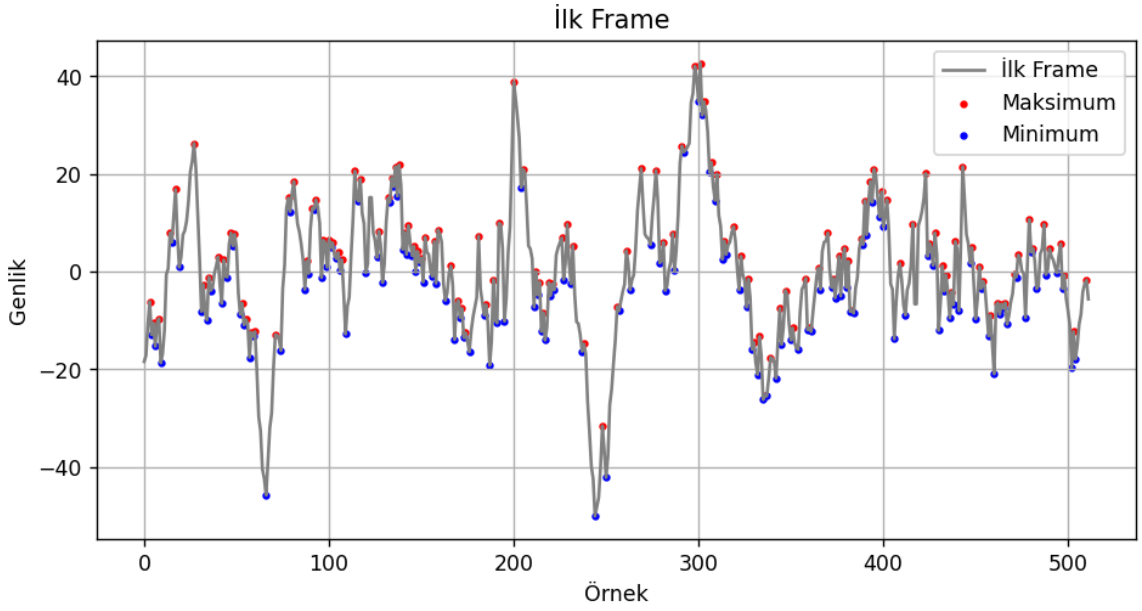
Şekil 19. Sinüs sinyalinin ikinci AGD görüntüsü

Şekil 18 ve Şekil 19' da görüldüğü üzere yerel maksimum noktaları x ekseninin üst kısmında yerel minimum noktaları da x ekseninin alt kısmına ve normalleştirilmiş genlik değerleriyle birlikte Açı Genlik Grafiğine yerleştirilmiştir.

LibriSpeech ve VoxCeleb1 veri setinden alınan birer örnek ses sinyalinin ilk çerçevesine ait görünümü ile bu çerçeve içerisindeki yerel maksimum ve minimum noktalarının konumları LibriSpeech için Şekil 20'de, VoxCeleb1 için Şekil 21'de görselleştirilmiştir. Ses sinyallerine Hamming penceresi uygulanmış olup, en iyi sonucu veren çerçeve boyutu (frame size) 512 seçilmiştir ve %25 örtüşme kullanılmıştır.

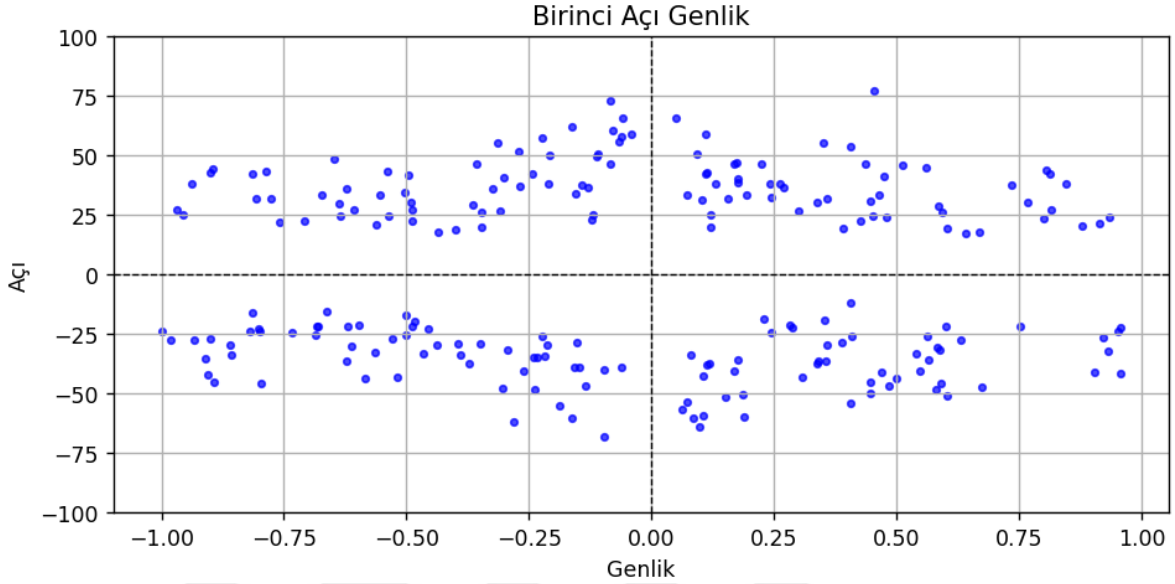


Şekil 20. LibriSpeech veri setinden örnek ses sinyali ve ekstremum noktaları bulunması

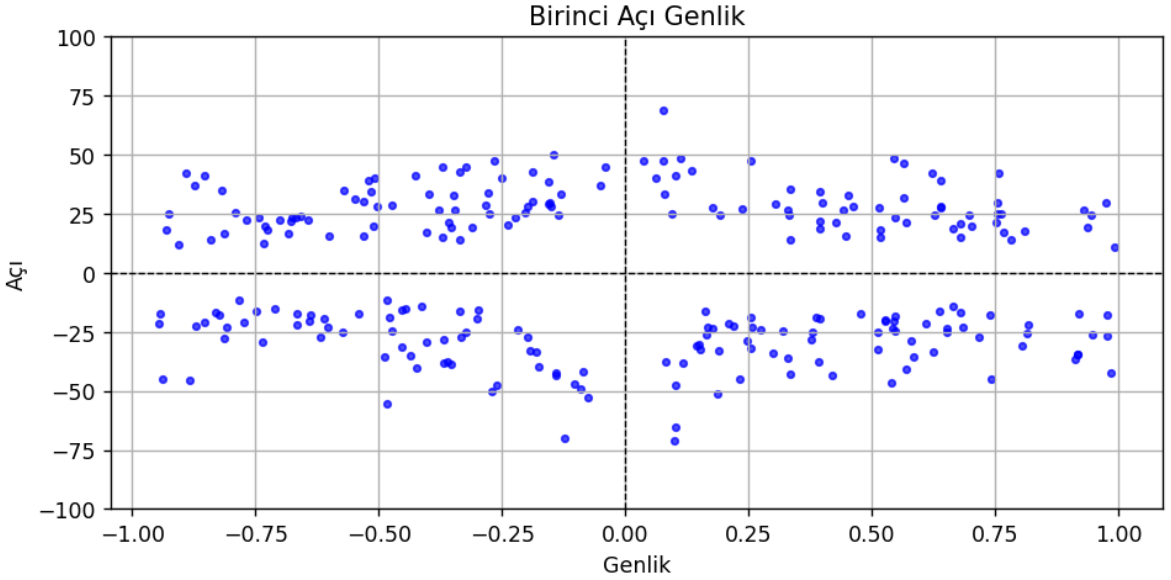


Şekil 21. VoxCeleb1 veri setinden örnek ses sinyali ve ekstremum noktaları bulunması

Şekil 22 ve Şekil 23'te ses örneklerinin ilk çerçevesine uygulanan birinci derece açı genlik dönüşümlerine karşılık gelen nokta dağılımı görüntüleri sunulmaktadır.

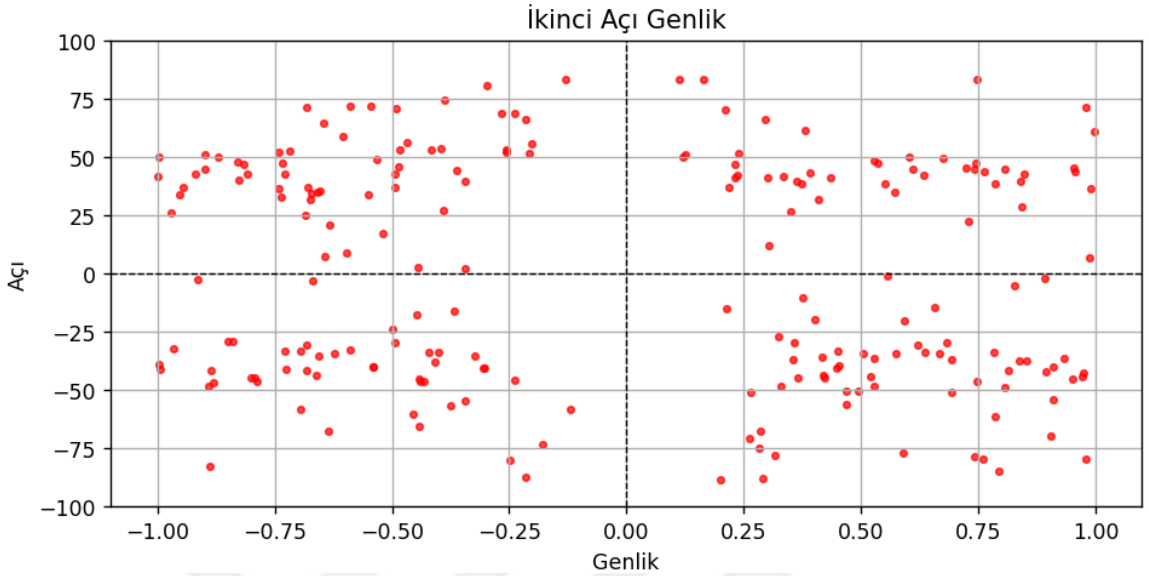


Şekil 22. LibriSpeech ses örneğinin ilk çerçevesinin birinci AGD görseli

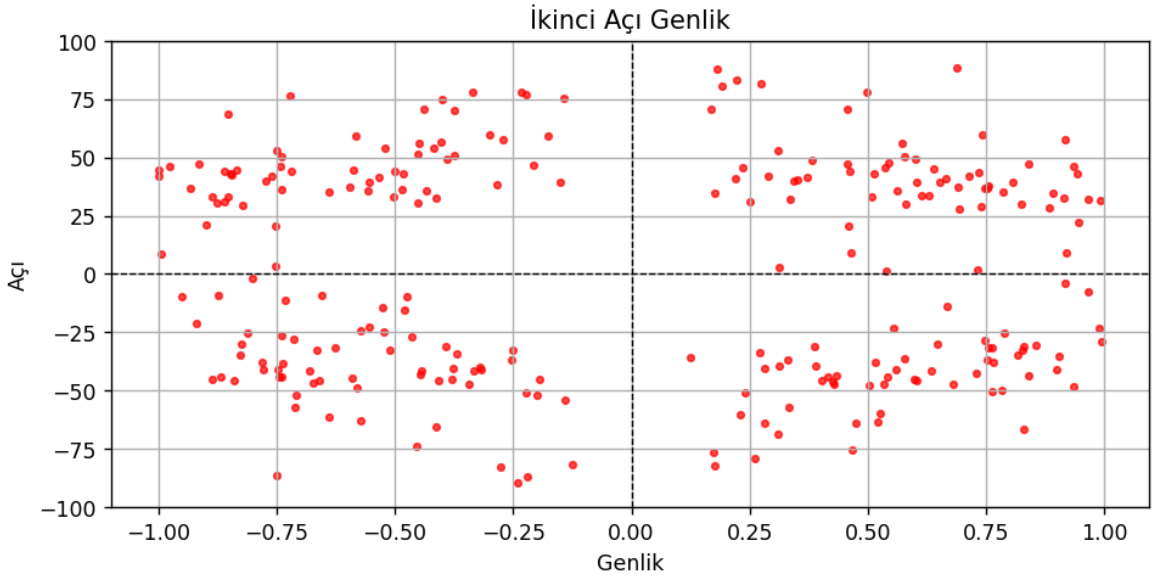


Şekil 23. VoxCeleb1 ses örneğinin ilk çerçevesinin birinci AGD görseli

Ayrıca ilk çerçeveye uygulanan ikinci derece açı genlik dönüşümlerine karşılık gelen nokta dağılımı görüntüleri Şekil 24 ve Şekil 25'te verilmiştir.



Şekil 24. LibriSpeech ses örneğinin ilk çerçevesinin ikinci AGD görseli



Şekil 25. VoxCeleb1 ses örneğinin ilk çerçevesinin ikinci AGD görseli

2.3. Yönlendirilmiş Gradyanların Histogramı

Bu çalışmada, elde edilen birinci ve ikinci açı genlik görsellerinden öznitelik çıkarımı yapmak amacıyla Yönlendirilmiş Gradyanların Histogramı yöntemi kullanılmıştır. HOG, bir görüntüdeki gradyan dağılımını yön bilgisiyle birlikte temsil eden bir tekniktir ve kenar, yapı gibi görsel örüntülerin tanımlanmasında oldukça etkilidir (Sugiharto vd., 2022). Gri tonlamalı olarak elde edilen birinci ve ikinci AGD görüntüleri üzerinde gradyan değeri (φ) ve yönü (θ) denklemlerdeki gibi hesaplanır.

$$\varphi = \sqrt{\left(G_x^2 + G_y^2\right)} \tag{28}$$

$$\theta = arctan\left(G_y \middle| G_x\right) \tag{29}$$

Elde edilen görüntü sabit boyutlarda hücrelere (nxn) bölünür ve her hücre içindeki piksellerin yönelimi, belirli yön aralıklarına göre gruplandırılarak histogramlar oluşturulur. Ardından birden fazla hücreden oluşan bloklar kaydırılarak görüntü üzerinde ilerletilir. Tüm bloklardan elde edilen normalize histogramlar uç uca eklenerek sabit boyutlu tek bir HOG öznitelik vektörü elde edilir ve KNN, SVM gibi sınıflandırma algoritmalarına giriş olarak verilir (Sugiharto vd, 2022).

Çalışma kapsamında birinci ve ikinci AGD sonucu elde edilen görüntüler 300×300 piksel boyutunda oluşturulmuştur. Oluşan görüntü, kenar bilgilerini yeterli hassasiyetle elde etmek amacıyla 8×8 piksel boyutundaki hücrelere bölünmüş ardından, bu hücrelerden 2×2'lik bloklar oluşturularak bölgesel histogramlar birleştirilmiştir. Her hücredeki gradyan yönleri 12 farklı açı aralığına ayrılarak oluşan histogramlar uç uca eklenerek birleştirilmiş sabit boyutlu bir öznitelik vektörü elde edilmiştir.

2.4. Deneysel Çalışmalar

Bu çalışma kapsamında, konuşmacı tanıma sistemlerinin doğruluğunu artırmaya yönelik olarak farklı öznitelik çıkarım yöntemleri üzerinde kapsamlı bir analiz gerçekleştirilmiştir. Deneysel çalışmalar, hem kontrollü ve temiz kayıtlar içeren LibriSpeech veri kümesi hem de gürültü ve arka plan sesleri barındıran VoxCeleb1 veri kümesi üzerinde yürütülmüştür. Böylece, önerilen yöntemlerin hem ideal koşullarda hem de gerçekçi, zorlu ortamlarda gösterdiği performans değerlendirilmiştir.

İlk aşamada, literatürde yaygın olarak kullanılan MFCC yöntemi uygulanmıştır. MFCC hesaplamasında, ham ses sinyalleri üzerinde öncelikle 512 örnek (frame size) uzunluğunda çerçeveleme yapılmış, çerçeveler arasında %25 örtüşme uygulanmıştır. Her çerçeveye, kenar etkilerini azaltmak amacıyla Hamming penceresi uygulanmıştır. Ardından, her çerçeveden 40 adet MFCC özelliği çıkarılmıştır. MFCC, insan işitme sisteminin frekans duyarlılığına uygun olarak mel ölçeğinde filtreleme yapması ve ardından kepstral katsayıların elde edilmesi ile sesin spektral özelliklerini temsil etmektedir. MFCC yönteminde, KNN ve SVM algoritmaları ile sınıflandırma yapılmasının yanı sıra, Mel spektrogramı doğrudan CNN mimarisine giriş olarak verilmiş ve derin öğrenme tabanlı sınıflandırma gerçekleştirilmiştir.

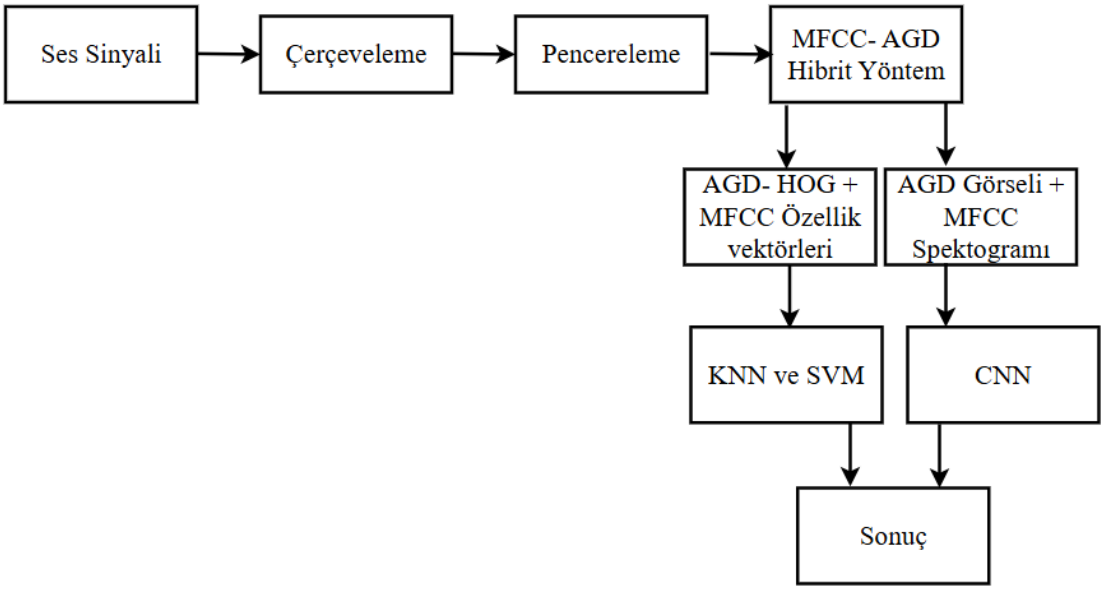
MFCC'nin yanı sıra, sesin üretim mekanizmasını modelleyen LPCC yöntemi de kullanılmıştır. LPCC hesaplamasında, doğrusal öngörülü kodlama analizi için order parametresi 16 seçilmiş ve elde edilen 16 adet LPCC katsayısı öznitelik vektörünü oluşturmuştur. LPCC yöntemi yalnızca KNN ve SVM algoritmaları ile sınıflandırılmış, öznitelikler sadece vektör olduğundan CNN tabanlı bir yaklaşım bu yöntem için uygulanmamıştır.

Bunun yanında, çalışmada literatüre yeni bir bakış açısı kazandıran AGD yöntemi uygulanmıştır. AGD, ses sinyali üzerindeki yerel maksimum ve minimum noktaların (ekstremum) belirlenmesi ve bu noktalar arasındaki ilişkilerin açı ve genlik parametreleri ile modellenmesine dayanmaktadır. Yöntemin birinci derecesinde, her bir ekstremum noktası için bir önceki ve bir sonraki zıt tipte ekstremumlar dikkate alınarak açı ve genlik değerleri hesaplanmıştır.

İkinci derecede ise, iki önceki ve iki sonraki zıt ekstremumlar üzerinden aynı hesaplama gerçekleştirilmiş ve böylece sinyalin daha geniş bağlamdaki yapısal ilişkileri yakalanmıştır. AGD ile elde edilen görüntüler üzerinde HOG yöntemi kullanılarak sayısal öznitelikler çıkarılmış, bu öznitelikler KNN ve SVM algoritmaları ile sınıflandırılmıştır. Ayrıca, AGD görüntüleri doğrudan CNN mimarisine verilerek ek öznitelik çıkarımı olmaksızın derin öğrenme tabanlı sınıflandırma yapılmıştır.

Son aşamada ise, kepstral tabanlı MFCC yöntemi ile sinyalin zamansal genlik ilişkilerini modelleyen birinci derece AGD yöntemi hibrit bir yapıda birleştirilmiştir.

Bu hibrit yaklaşımda, MFCC'den elde edilen 40 özellik ile birinci derece AGD görüntülerinden elde edilen özellikler, öznitelik seviyesinde birleştirilmiş ve KNN, SVM algoritmaları ile sınıflandırılmıştır. Ayrıca MFCC'nin spektogram görüntüsüyle, AGD görüntüsü CNN'e giriş olarak verilmiştir. Böylece, MFCC'nin spektral temsil gücü ile AGD'nin zaman genlik ilişkilerini ortaya çıkarma yeteneğinin birbirini tamamlaması sağlanmıştır. Hibrit yöntemin blok diyagramı Şekil 26'da verilmiştir.



Şekil 26. Hibrit yöntem blok diyagramı

Elde edilen sonuçlar, farklı veri setleri ve koşullar altında MFCC, LPCC, AGD ve hibrit MFCC-AGD yöntemlerinin performanslarını karşılaştırmalı olarak ortaya koymakla birlikte hibrit yaklaşımın, özellikle gürültülü ortamlarda ve karmaşık sinyal yapılarında belirgin bir performans artışı sağladığını göstermektedir. Deneysel çalışmalarda yapılan işlemler Tablo 3'te özetlenmiştir.

Tablo 3. Çalışma kapsamında yapılan işlemlerin özeti

Aşamalar	Ön İşlemler	Yöntemler	Veri Seti	Sınıflandırma
1	Çerçeveleme Pencereleme	MFCC	LibriSpeech VoxCeleb1	KNN SVM CNN
2	Çerçeveleme Pencereleme	LPCC	LibriSpeech VoxCeleb1	KNN SVM
3	Çerçeveleme Pencereleme	Birinci AGD Görseli+HOG İkinci AGD Görseli+HOG Birinci AGD Görseli İkinci AGD Görseli	LibriSpeech VoxCeleb1 LibriSpeech VoxCeleb1	KNN SVM CNN
4	Çerçeveleme Pencereleme	Hibrit Yöntem Hibrit Yöntem	LibriSpeech VoxCeleb1 LibriSpeech VoxCeleb1	KNN SVM CNN

3. BULGULAR VE TARTIŞMA

Bu çalışmada, konuşmacı tanıma alanına yönelik olarak farklı özellik çıkarım yöntemlerinin ve sınıflandırma algoritmalarının etkinliği karşılaştırmalı olarak incelenmiştir. Özellik çıkarımı aşamasında MFCC, LPCC, birinci derece ve ikinci derece Açı Genlik Dönüşümü (AGD) yöntemleri kullanılmıştır. Ayrıca MFCC ve AGD yöntemlerinin birleşimiyle oluşturulan hibrit bir öznitelik seti de değerlendirmeye alınmıştır.

Çalışma kapsamında, LibriSpeech ve VoxCeleb1 veri setlerindeki 20 kişinin hem temiz hem de gerçek dünya koşullarını yansıtan ses örnekleri üzerinde analizler gerçekleştirilmiştir. Tüm öznitelik çıkarımı yöntemleri için ham ses sinyalleri üzerinde 512 örneklik çerçeve boyutu, %25 örtüşme oranı (hop size 256) ve Hamming penceresi tercih edilmiştir. MFCC çıkarımında 40, LPCC yöntemi için ise 16 öznitelik vektörü oluşturulmuştur. AGD yönteminde ise ses sinyalinden elde edilen ekstremum noktaları üzerinden birinci ve ikinci derece açı ve genlik bilgileri hesaplanarak iki boyutlu görüntüler elde edilmiştir. Bu görüntülerden HOG yöntemiyle öznitelikler çıkarılmıştır.

Elde edilen bu öznitelik vektörleri, KNN, SVM algoritmaları ile sınıflandırılmıştır. KNN algoritmasında Manhattan mesafesi kullanılmış olup, deneme-yanılma yöntemiyle en iyi sonuç k=5 değeri için elde edilmiştir. SVM sınıflandırmasında ise lineer çekirdek fonksiyonu tercih edilmiştir. AGD görüntüleri CNN modellerine doğrudan giriş olarak verilmiş, MFCC için spektrogram temsilleri kullanılarak derin öğrenme tabanlı sınıflandırmalar yapılmıştır. LPCC yöntemi sadece geleneksel makine öğrenmesi algoritmaları olan KNN ve SVM ile değerlendirilmiştir.

Hibrit yapıda, MFCC öznitelikleri ile birinci derece AGD-HOG öznitelikleri birleştirilerek klasik sınıflandırıcılar üzerinde test edilmiştir. CNN mimarisi için ise MFCC ve AGD temsillerinden elde edilen görüntüler eş zamanlı olarak modele verilerek hibrit görsel sınıflandırma gerçekleştirilmiştir.

Deney sonuçları, özellikle AGD tabanlı görüntülerle eğitilen mimarilerin yüksek sonuçlara ulaştığını ve konuşmacı tanıma alanına yeni bir yaklaşım olacabileceğini göstermektedir. Ayrıca MFCC ve AGD'nin birlikte kullanıldığı hibrit yaklaşım, her iki

yöntemin güçlü yönlerini bir araya getirerek sınıflandırma başarısını önemli ölçüde artırmıştır.

3.1. Birinci ve İkinci AGD Yöntemi ile Sınıflandırma Sonuçları

Tablo 4. Birinci AGD sonuçları

Veri Seti	Sınıflandırma	Doğruluk oranı (%)	Kesinlik (%)	Duyarlılık (%)	F1 Skoru (%)
LibriSpeech	KNN	83.3	84.9	83.5	83.8
	SVM	91.4	92.6	91.4	91.8
	CNN	93.2	93.8	93.3	93.3
VoxCeleb1	KNN	52	51.2	52.4	50
	SVM	80.2	81.4	80.2	80
	CNN	75.5	81	75.5	75.6

Tablo 5. İkinci AGD sonuçları

Veri Seti	Sınıflandırma	Doğruluk oranı (%)	Kesinlik (%)	Duyarlılık (%)	F1 Skoru (%)
LibriSpeech	KNN	85.8	87	86.2	86.2
	SVM	87	88.6	87.3	87.5
	CNN	91	91.4	90.9	90.8
VoxCeleb1	KNN	47.1	50	47.1	46
	SVM	78.4	79.1	78.4	77.7
	CNN	71	73.2	71	70

LibriSpeech ve VoxCeleb1 veri setlerinden açı genlik dönüşümü sonrası oluşan birinci ve ikinci açı genlik görüntülerinin ayrı ayrı sınıflandırılması ve başarı metriklerinden elde edilen sonuçlar Tablo 4 ve Tablo 5'te gösterilmiştir. Tablolar birinci ve ikinci AGD görüntülerinden elde edilen doğruluk, kesinlik, duyarlılık ve F1 skoru ölçütlerinin sınıflandırma algoritmaları üzerinden gücünü ve sınırlılıklarını sunmaktadır.

Birinci derece AGD yöntemiyle LibriSpeech veri setinde en başarılı sonuç CNN algoritması ile elde edilmiş olup doğruluk oranı %93.2, F1 skoru ise %93.3'tür. SVM algoritması da oldukça başarılı sonuçlar göstermiştir (%91.4 doğruluk, %91.8 F1 skoru). KNN algoritması ise %83.3 doğruluk ve %83.8 F1 skoru ile daha düşük bir performans göstermektedir.

VoxCeleb1 veri seti ise içerdiği arka plan gürültüleri ve gerçek yaşamdan alınmış çeşitli ortam sesleri nedeniyle daha düşük başarı göstermiştir. En yüksek doğruluk değeri %80.2 ile SVM algoritmasına aittir, ardından %75.5 ile CNN ve %52.0 ile KNN gelmektedir. Bu fark, verinin yapısından kaynaklanan zorluklara dikkat çekmekte ve CNN'in stüdyo ortamı verilerinde daha etkili olduğunu ortaya koymaktadır.

İkinci derece AGD yönteminde de LibriSpeech veri seti üzerindeki başarı dikkat çekicidir. Özellikle CNN algoritması %91 doğruluk ve %90.8 F1 skoru ile birinci dereceye benzer yüksek performans göstermiştir. SVM algoritması da %87 doğruluk ve %87.5 F1 skoru ile ikinci sırada yer alırken, KNN %85.8 doğruluk ile daha sınırlı bir başarı sergilemiştir.

VoxCeleb1 veri setinde ise ikinci derece AGD yöntemi, genel olarak birinci dereceye kıyasla biraz daha düşük performans göstermiştir. CNN %71 doğruluk ve %70 F1 skoru ile iyi sonuç verirken, SVM %78.4 doğruluk ve %77.7 F1 skoru ile en yüksek başarı sağlamıştır. KNN ise %47.1 doğruluk ve %46 F1 skoru ile düşük bir performans sergilemiştir.

Birinci derece AGD yöntemi, özellikle CNN tabanlı sınıflandırmalarda her iki veri setinde de yüksek başarı göstermiştir. Ancak ikinci derece AGD ile daha geniş sinyal yapısı değerlendirildiğinden, belirli durumlarda sınıflandırma başarısında küçük düşüşler yaşanabilmektedir. Bu durum özellikle gürültülü veriler içeren VoxCeleb1 veri setinde belirginleşmektedir. Bununla birlikte, ikinci derece AGD, sinyalin daha geniş bağlamını dikkate alması sayesinde bazı sınıflandırıcılarda istikrarlı sonuçlar verebilmektedir.

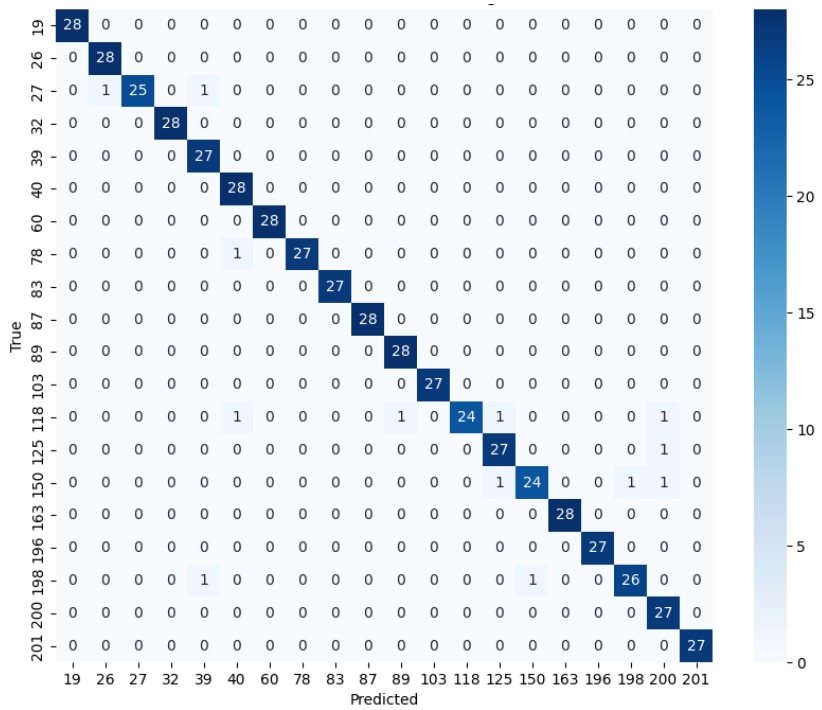
Bu analiz, açı genlik tabanlı görsel temsillerin CNN gibi derin öğrenme modelleri ile kullanıldığında, özellikle gürültüsüz veri setlerinde (LibriSpeech) yüksek doğrulukla konuşmacı tanıma sağlayabildiğini göstermektedir. Öte yandan, daha karmaşık ve doğal veri ortamlarında başarıyı sürdürebilmek için gürültü azaltma ve veri iyileştirme tekniklerinin de kullanılmasıyla birlikte başarı arttırabilir.

3.2. MFCC - AGD Hibrit Yapı ve Sınıflandırma Sonuçları

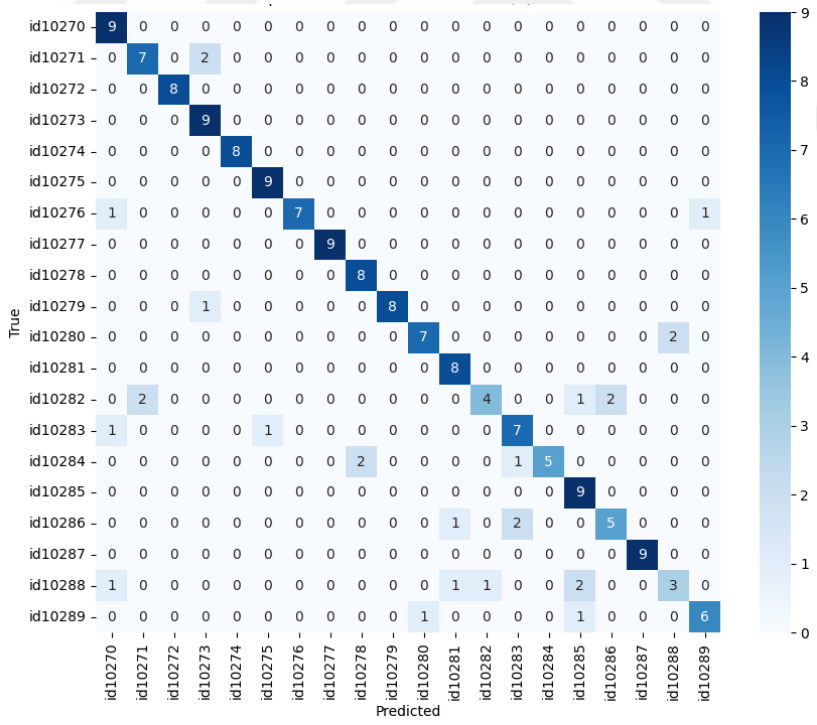
Tablo 6. Hibrit yapı sonuçları

Veri Seti	Sınıflandırma	Doğruluk oranı (%)	Kesinlik (%)	Duyarlılık (%)	F1 Skoru (%)
LibriSpeech	KNN	97.8	98	97.8	97.9
	SVM	100	100	100	100
	CNN	97.6	97.7	97.6	97.6
Voxceleb1	KNN	90.3	92.6	89.4	89.4
	SVM	95.6	96.3	94.8	95.1
	CNN	84.3	85	84	83.4

Aynı veri setlerinden elde edilen ve aynı koşullarla oluşturulan hibrit yapı için sınıflandırma algoritmalarının başarı metrikleri sonucu Tablo 6'da gösterilmiştir. Bu sonuçlarla, MFCC yöntemi ve AGD yönteminin güçlü özellikleri birleştirilerek sistem başarısının arttığı gözlemlenmiştir. Şekil 27 ve Şekil 28'de LibriSpeeech ve VoxCeleb1 veri setleri için Hibrit yapı ve CNN mimarisinin karmaşıklık matrisi sonucu verilmiştir.



Şekil 27. LibriSpeech veri setine uygulanan hibrit yapı- CNN karmaşıklık matrisi



Şekil 28. VoxCeleb1 veri setine uygulanan hibrit yapı- CNN karmaşıklık matrisi

MFCC ve AGD yöntemlerinin güçlü yönlerini birleştiren hibrit yaklaşım, her iki veri seti üzerinde KNN, SVM ve CNN sınıflandırıcıları ile test edilmiştir. LibriSpeech veri setinde, hibrit özniteliklerle en yüksek doğruluk oranı %100 ile SVM algoritmasında elde edilmiştir. KNN ve CNN algoritmaları da sırasıyla %97.8 ve %97.6 doğruluk oranları ile yüksek performans göstermiştir. Bu durum, temiz kayıt ortamına sahip olan LibriSpeech veri setinde hibrit özniteliklerin yüksek genelleme başarısına sahip olduğunu ortaya koymaktadır.

VoxCeleb1 veri setinde ise çevresel gürültü ve değişken kayıt koşulları sebebiyle başarı oranları nispeten daha düşük olmakla birlikte, hibrit yapı yine de güçlü sonuçlar vermiştir. SVM algoritması %95.6 doğruluk ile en yüksek performansı gösterirken, KNN %90.3 ve CNN %84.3 doğruluk oranı ile takip etmiştir. Bu sonuçlar, hibrit özniteliklerin gürültülü verilerde bile anlamlı katkı sağladığını göstermektedir.

Genel olarak, MFCC ve AGD özniteliklerinin hibrit yapıda birleştirilmesi, her iki veri setinde de sistemin genel başarısını artırmış ve özellikle SVM algoritması ile oldukça yüksek doğruluk, kesinlik ve F1 skoru değerleri elde edilmiştir. Bu durum, farklı özellik çıkarım yöntemlerinin bütünleştirilmesinin konuşmacı tanıma performansını artırabileceğini açıkça ortaya koymaktadır.

3.3. Klasik Öznitelik Çıkarım Yöntemleri ve Önerilen Metotların Karşılaştırılması

Tablo 7. Metotlar ve sonuçların karşılaştırılması

Veri Seti	Sınıflandırma	MFCC Doğruluk (%)	LPCC Doğruluk (%)	Birinci AGD Doğruluk (%)	İkinci AGD Doğruluk (%)	Hibrit Yöntem Doğruluk (%)
Librispeech	KNN	99.16	69.83	83.3	85.8	97.8
	SVM	98.73	73.71	91.4	87	100
	CNN	95.78		93.2	91	97.6
Voxceleb1	KNN	98.25	54.47	52	47.1	90.3
	SVM	92.11	50.88	80.2	78.4	95.6
	CNN	86.04		75.5	71	84.3

Tablodaki sonuçlar, klasik yöntemler (MFCC ve LPCC), önerilen birinci ve ikinci AGD yöntemleri ile hibrit yapının farklı veri setleri üzerinde ve tüm yöntemlerin aynı koşulları içeren durumlardaki performansını açıkça ortaya koymaktadır. MFCC, her iki veri setinde de yüksek doğruluk oranları sergileyerek klasik öznitelik çıkarım teknikleri arasında en başarılı yöntem olarak öne çıkmaktadır. Özellikle LibriSpeech veri setinde KNN ve SVM algoritmalarında sırasıyla %99.16 ve %98.73 doğruluk elde edilmiş, CNN ile ise %95.78'lik bir başarı sağlanmıştır. Buna karşın, LPCC yöntemi, MFCC'ye kıyasla daha düşük doğruluk değerleri göstermiş; LibriSpeech'te %69.83– %73.71 aralığında, VoxCeleb1'de ise %50.88– %54.47 seviyelerinde kalmıştır.

Birinci ve ikinci derece AGD yöntemleri ise, özellikle temiz kayıt koşullarına sahip LibriSpeech veri setinde, LPCC'ye kıyasla anlamlı bir iyileşme sağlamıştır. Birinci derece AGD, KNN ile %83.3, SVM ile %91.4 ve CNN ile %93.2 doğruluk sunarken; ikinci derece AGD'nin daha geniş bağlam bilgisi kullanması, bazı sınıflandırıcılarda (örneğin KNN) ek katkı sağlamış, ancak her durumda birinci dereceye göre belirgin bir üstünlük göstermemiştir.

Hibrit yapı (MFCC - AGD), her iki veri setinde de tekil yöntemlere kıyasla dengeli ve yüksek başarı sağlamıştır. LibriSpeech'te SVM ile %100 doğruluk elde edilirken, diğer sınıflandırıcılarla da %97'nin üzerinde sonuçlara ulaşılmıştır. VoxCeleb1 veri setinde ise hibrit yapı, özellikle KNN (%90.3) ve SVM (%95.6) ile, tek başına MFCC veya AGD'ye kıyasla daha yüksek doğruluk sağlamış, CNN'de ise %84.3 ile makul bir performans sergilemiştir.

Genel olarak, bulgular hibrit yapının veri çeşitliliğine karşı daha dayanıklı olduğunu ve hem stüdyo kayıtlarında hem de gürültülü ortamlarda klasik yöntemlere kıyasla daha istikrarlı sonuçlar sunduğunu göstermektedir. Birinci ve ikinci AGD yöntemleri, özellikle temiz veri koşullarında LPCC'ye göre ciddi bir iyileşme sağlarken; hibrit yapı, MFCC'nin spektral avantajlarını, AGD'nin zamansal geometrik temsil gücüyle birleştirerek yüksek başarıya ulaşmıştır.

4. SONUÇLAR

Yapılan deneysel çalışmalar ve gerçekleştirilen analizler sonucunda, Açı Genlik Dönüşümü (AGD) yöntemi kullanılarak ses sinyalinin geometrik yapısı üzerinden konuşmacı tanımlama alanına yenilikçi bir yaklaşım sunulmuştur. Hem AGD görüntülerinden elde edilen özniteliklerle hem de hibrit yapılarla elde edilen sınıflandırma sonuçları, bu yöntemin alana katkı sağlayabilecek düzeyde başarılı olduğunu göstermektedir.

Tablo 8'de, literatürde konuşmacı tanımlama alanında LibriSpeech ve VoxCeleb1 veri setleriyle yapılmış çalışmalar ile bu tez kapsamında önerilen yöntemlerin karşılaştırmalı analizi sunulmuştur.

Tablo 8. Literatür çalışmaları ve önerilen yöntemlerin karşılaştırılması

Yazar,Yılı	Özellik çıkarma	Sınıflandırma	Veri seti	Doğruluk (%)
Günerhan (2023)	MFCC	Ek hiper parametreli LSTM	VoxCeleb1 LibriSpeech	99.5 97.8
Jahangir vd., (2020)	MFCCT	DNN	LibriSpeech	89
An vd., (2019)	MFCC	VGG like-CNN+ Self Attention ResNet-18+Self Attention	VoxCeleb VoxCeleb	93.8 96.5
Yao vd., (2023)	SSED Saliency Map Decoder Angular Loss Function	Fast ResNet-34	VoxCeleb1 VoxCeleb2	93.15

Tablo 8'in devamı

Yazar, Yılı	Özellik çıkarma	Sınıflandırma	Veri seti	Doğruluk (%)
Fasounaki vd., (2021)	MFCC	Derin CNN mimarisi	VoxCeleb1 LibriSpeech	97.1 99.8
Chen vd., (2020)	SpeakerGAN	Gated CNN	LibriSpeech	98.20
Önerilen Yöntem (1)	Birinci AGD Görseli + HOG	KNN	LibriSpeech VoxCeleb1	83.3 52
		SVM	LibriSpeech VoxCeleb1	91.4 80.2
	Birinci AGD Görseli + CNN	CNN	LibriSpeech VoxCeleb1	93.2 75.5
Önerilen Yöntem (2)	İkinci AGD Görseli+ HOG	KNN	LibriSpeech VoxCeleb1	85.8 47.1
		SVM	LibriSpeech VoxCeleb1	87 78.4
	İkinci AGD Görseli + CNN	CNN	LibriSpeech VoxCeleb1	91 71
Önerilen Yöntem (3) (Hibrit Yöntem)	AGD Görseli+ HOG + MFCC	KNN	LibriSpeech VoxCeleb1	97.8 90.3
		SVM	LibriSpeech VoxCeleb1	100 95.6
	AGD + MFCC Görselleri	CNN	LibriSpeech VoxCeleb1	97.6 84.3

Tablo 8'de konuşmacı tanıma sistemleri için geliştirilen açı genlik dönüşümüne dayalı yöntemlerin başarısı, LibriSpeech ve VoxCeleb1 veri setleri üzerinde yapılan deneylerle değerlendirilmiştir. Çalışmada yalnızca AGD temelli yaklaşımlar değil, aynı zamanda bu yöntemin MFCC gibi literatürde yaygın kullanılan öznitelik çıkarım yöntemleriyle birlikte kullanıldığı hibrit yapılar da test edilmiştir. Elde edilen sonuçlar, önerilen yöntemin hem klasik makine öğrenmesi algoritmaları (KNN, SVM) hem de derin öğrenme tabanlı CNN mimarileriyle entegre olarak etkili bir şekilde uygulanabildiğini ortaya koymuştur.

Yapılan deneysel analizlerde, Birinci Derece AGD ile çıkarılan görseller üzerinden elde edilen özniteliklerin, HOG yöntemi ile birlikte kullanılması durumunda LibriSpeech veri setinde KNN ile %83.3, SVM ile %91.4 doğruluk oranı sağladığı gözlemlenmiştir. Ancak VoxCeleb1 veri setinde aynı yapı KNN ile yalnızca %52 başarı göstermiş, SVM ile bu oran %80.2'ye kadar yükselmiştir. Bu durum, LibriSpeech veri setinin stüdyo ortamında ve düşük gürültü seviyelerinde kaydedilmiş olmasıyla açıklanabilirken, VoxCeleb1 veri setinde ise doğal ve gürültülü ortamlardan gelen kayıtların sınıflandırma başarısını olumsuz etkilediği değerlendirilmiştir.

Benzer şekilde, CNN mimarisi ile doğrudan AGD görselleri kullanılarak yapılan sınıflandırmalarda LibriSpeech veri setinde %93.2, VoxCeleb1 veri setinde ise %75.5 doğruluk oranına ulaşılmıştır. Bu sonuçlar, özellikle CNN gibi derin öğrenme yaklaşımlarının AGD tabanlı görsel temsilleri etkin biçimde sınıflandırabildiğini göstermektedir.

İkinci Derece AGD kullanılarak yapılan deneylerde ise sinyalin daha geniş bir yapısını temsil eden özniteliklerin kullanılması hedeflenmiştir. LibriSpeech veri setinde HOG öznitelikleriyle KNN ve SVM sınıflandırıcılar sırasıyla %85.8 ve %87 doğruluk sağlamış, CNN ise %91 doğruluk oranına ulaşmıştır. VoxCeleb1 veri setinde ise bu oranlar KNN için %47.1, SVM için %78.4 ve CNN için %71 olarak kalmıştır.

Çalışmanın önemli bulgularından biri de, AGD ve MFCC özniteliklerinin birlikte kullanıldığı hibrit yöntemde elde edilmiştir. Bu yapıda, KNN algoritmasıyla LibriSpeech üzerinde %97.8, VoxCeleb1 üzerinde %90.3 doğruluk sağlanmıştır. Aynı yapı SVM ile %100 ve %95.6; CNN ile %97.6 ve %84.3 doğruluk oranlarına ulaşmıştır. Bu sonuçlar, hibrit yapının, MFCC'nin zaman-frekans çözünürlüğü ile AGD'nin sinyalin geometrik

yapısına odaklanan güçlü yönlerini birleştirerek başarıyı anlamlı ölçüde artırdığını göstermektedir.

Literatürde yer alan diğer çalışmalara bakıldığında, MFCC tabanlı yaklaşımların genellikle yüksek doğruluk oranları sağladığı görülmektedir. Örneğin, Günerhan (2023) tarafından yapılan çalışmada LSTM tabanlı sınıflandırıcı ile %99.5 doğruluk elde edilmiştir. Benzer şekilde, Jahangir vd. (2020) çalışmasında DNN mimarisiyle %89; An vd. (2019) tarafından kullanılan VGG-like CNN + Self-Attention yapısıyla %93.8 doğruluk raporlanmıştır. Bu doğrultuda, önerilen AGD yöntemiyle ulaşılan doğruluk oranları, hem tek başına hem de hibrit yapılarla MFCC tabanlı sistemlerle rekabet edebilecek düzeyde olduğu sonucunu desteklemektedir.

Elde edilen bu sonuçlar, açı genlik dönüşümüne dayalı öznitelik çıkarımının konuşmacı tanıma sistemlerinde yenilikçi ve başarılı bir alternatif sunduğunu göstermektedir. Hibrit yapılarla entegre edildiğinde sistem başarısını anlamlı düzeyde artırabileceği sonucuna varılmıştır.

5. ÖNERİLER

Bu tez çalışmasında konuşmacı tanıma alanına yönelik olarak önerilen AGD yöntemi, hem klasik hem de derin öğrenme tabanlı sınıflandırma algoritmaları ile test edilmiş ve özellikle hibrit yapı kullanıldığında oldukça başarılı sonuçlar elde edilmiştir. Ancak çalışmanın daha da ileri taşınabilmesi için bazı öneriler yapılabilir.

Görsel temsillerin daha etkili çıkarılabilmesi adına, görsel iyileştirme teknikleri (örneğin kontrast artırma, gürültü filtreleme) kullanılarak sinyalin geometrik yapısı daha belirgin hale getirilebilir.

Çalışma kapsamında açı genlik dönüşümü uygulanırken birinci ve ikinci derecede değerlendirmeler ele alınmıştır. Bu değerlendirmelere ek, farklı varyasyonları uygulamak çeşitli sonuçlar elde edilmesine katkıda bulunabilir.

AGD görselleri üzerinden çıkarılan özniteliklerde sadece HOG değil, PCA (Temel Bileşenler Analizi), SIFT (Ölçekten Bağımsız Özellik Dönüşümü), Gabor filtreleri gibi diğer yöntemler de kullanılabilir. Bu sayede farklı açılardan sinyal karakteristikleri yakalanabilir.

Hibrit yapı kapsamında kullanılan MFCC parametreleri sabit tutulmuştur. MFCC parametrelerinin (kepstral katsayı sayısı, filtre sayısı vb.) optimize edilmesi, AGD ile birlikte çalışıldığında daha güçlü bir sonuç oluşturabilir.

Bu çalışmada yalnızca LibriSpeech ve VoxCeleb1 veri setleri değerlendirilmiştir. Gelecek çalışmalarda farklı dil ve kayıt koşullarına sahip veri setlerinin de dahil edilmesi, yöntemin genellenebilirliğini test etmek açısından faydalı olacaktır.

Sınıflandırma yöntemleri olarak farklı sınıflandırma algoritmaları denenebilir ve başarı metrikleri kıyaslanabilir.

Bu öneriler doğrultusunda yapılacak çalışmaların, açı genlik dönüşümünün konuşmacı tanıma alanındaki yerini daha da güçlendirebileceği ve literatürde yenilikçi bir katkı sunacağı düşünülmektedir.

6. KAYNAKLAR

Ahmed, S. R., Abbood, Z. A., Farhan, H. M., Yasen, B. T., Ahmed, M. R., & Duru, A. D. (2022). Speaker Identification Model Based On Deep Neural Networks. IJCSM, 3(1), 108- 113.

Akgun, K. (2024). İnsan Sesinden Hibrit Spektral Özniteliklerle Konuşmacı Özelliklerinin Tahmini. Kütahya Dumlupınar Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi, Kütahya.

An, N. N., Thanh, N. Q., & Liu, Y. (2019). Deep CNNs with self-attention for speaker identification. IEEE access, 7, 85327-85337.

Ashar, A., Bhatti, M. S., & Mushtaq, U. (2020). Speaker identification using a hybrid cnn-mfcc approach. In 2020 International conference on emerging trends in smart technologies (ICETST) IEEE, 1-4.

Başaran, S. (2007). Yapay Sinir Ağları Kullanarak Konuşmacı Tanıma. Uludağ Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Bursa.

Bunrit, S., Inkian, T., Kerdprasop, N., & Kerdprasop, K. (2019). Text-independent speaker identification using deep learning model of convolution neural network. International Journal of Machine Learning and Computing, 9(2), 143-148.

Caner, M., & Üstün, S. V. (2006). Yapay Sinir Ağları ile Konuşmacı Kimliğini Tanıma Uygulaması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 12(2), 279-284.

Cengil, E., & Çınar, A. (2016). A new approach for image classification: convolutional neural network. European Journal of Technique (EJT), 6(2), 96-103.

Chen, L., Liu, Y., Xiao, W., Wang, Y., & Xie, H. (2020). SpeakerGAN: Speaker identification with conditional generative adversarial network. Neurocomputing, 418, 211-220.

Çolakoğlu, E., Hızlısoy, S., & Arslan, R. S. (2021). Konuşmadan Duygu Tanıma Üzerine Detaylı bir İnceleme: Özellikler ve Sınıflandırma Metotları. Avrupa bilim ve teknoloji dergisi, 32, 471-483.

Eray, O. (2008). Destek Vektör Makineleri ile Ses Tanıma Uygulaması. Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Denizli.

Eskidere, Ö. (2007). İstatiksel Modelleme ile Konuşmacı Tanıma. Uludağ Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Bursa.

Fasounaki, M. (2021). CNN-Based Text-Independent Automatic Speaker Identification. İ.T.Ü., Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi, İstanbul.

Fasounaki, M., Yuce, E.B., Oncul, S., & Ince, G. (2021). A Comparative Assessment of Text independent Automatic Speaker Identification Methods Using Limited Data. Avrupa Bilim ve Teknoloji Dergisi, 26, 217- 222.

Fasounaki, M., Yuce, E.B., Oncul, S., & Ince, G. (2021). CNN-based Text-independent Automatic Speaker Identification Using Short Utterances. In 2021 6th international conference on computer science and engineering (UBMK) IEEE, 413-418.

Geravanchizadeh, M., & Ghalamiosgouei, S. (2020). Binaural speaker identification using the equalization-cancelation technique. EURASIP Journal on Audio, Speech, and Music Processing, 2020(1), 20.

Günerhan, E. (2023). Uzun Kısa Süreli Bellek Tipi Derin Sinir Ağları ile Konuşmacı Tanıma K.T.Ü., Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Trabzon.

Hanilçi, C. (2007). Konuşmacı Tanıma Yöntemlerinin Karşılaştırmalı Analizi. Uludağ Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Bursa.

Hamsa, S., Shahin, I., Iraqi, Y., Damiani, E., Nassif, A. B., & Werghi, N. (2023). Speaker identification from emotional and noisy speech using learned voice segregation and speech VGG. Expert Systems with Applications, 224, 119871.

Hatipoglu, B. (2017). İmleç Hareketinin Hayali Sırasında Kaydedilmiş EEG Sinyallerinin Sınıflandırılmasını Kolaylaştırmak Amacıyla Yeni Bir Dönüşüm Yöntemi. K.T.Ü., Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Trabzon.

Hatipoglu, B., Yilmaz, C. M., & Kose, C. (2019). A signal-to-image transformation approach for EEG and MEG signal classification. Signal, Image and Video Processing, 13(3), 483-490.

Herbig, T., Gerl, F., & Minker, W. (2012). Self-learning speaker identification for enhanced speech recognition. Computer Speech & Language, 26(3), 210-227.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., & Wilson, K. (2017). CNN architectures for large-scale audio classification. In 2017 ieee international conference on acoustics, speech and signal processing (icassp) IEEE, 131-135.

Jahangir, R., Teh, Y. W., Ali, I., Ishtiaq, U., Akhtar , M. Z., Memon, N. A., Mujtaba, G., & Zareei, M. (2020). Text-Independent Speaker Identification Through Feature Fusion and Deep Neural Network. IEEE, 8, 2973541, 32187-32202.

Jing, X., Ma, J., Zhao, J., & Yang, H. (2014). Speaker recognition based on principal component analysis of LPCC and MFCC. In 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 403-408.

Kacur, J. (2016). Modifications of KNN classifier for speaker identification system. In 2016 International Symposium ELMAR IEEE, 35-38.

Karabina, A. (2017). Konuşmacı Tanımada Makine Öğrenmesi Tekniklerinin Kullanımı. O.M.Ü., Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Samsun.

Karasartova, S. (2011). Metinden Bağımsız Konuşmacı Tanıma Sistemlerinin İncelenmesi ve Gerçekleştirilmesi. Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.

Karakus, C. (2024). Makine Öğrenmesi Algoritmaları , Ders Notları.

Kawakami, Y., Wang, L., Kai, A., & Nakagawa, S. (2014). Speaker identification by combining various vocal tract and vocal source features. In International conference on text, speech, and dialogue , Springer International Publishing, 382-389.

Keskin, A. K. (2018). Makine Öğrenmesi Sınıflandırma Algoritmalarının İncelenmesi. Sinop Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Sinop.

Kop, B. Ş. (2022). Makine Öğrenmesi Yöntemleri ile Ses Tanıma Uygulamaları. ATATÜRK Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Erzurum.

Kop, B. S., & Bayındır, L. (2021). Bebek Ağlamalarının Makine Öğrenmesi Yöntemleriyle Sınıflandırılması. Avrupa Bilim ve Teknoloji Dergisi, 27, 784-791.

Kurbanov, O. (2018). Derin Sinir Ağları Kullanarak Parmak İzi Tanımada Yeni Yaklaşımlar. Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri.

Labied, M., & Belangour, A. (2021). Automatic speech recognition features extraction techniques: A multi-criteria comparison. International Journal of Advanced Computer Science and Applications, 12(8).

Leng, Y., Tang, Y., Liu, C., Zhao, W., Yuan, Q., Li, D., Xu, H., Sun, J., Sun, C., & Wang, R. (2022). Attention based gender and nationality information exploration for speaker identification. Digital Signal Processing, 123, 103449.

Leu, F. Y., & Lin, G. L. (2017). An MFCC-based speaker identification system. In 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), 1055-1062.

Liu, J. C., Leu, F. Y., Lin, G. L., & Susanto, H. (2018). An MFCC-based text-independent speaker identification system for access control. Concurrency and Computation: Practice and Experience, 30(2), e4255.

Lukic, Y., Vogt, C., Dürr, O., & Stadelmann, T. (2016). Speaker identification and clustering using convolutional neural networks. In 2016 IEEE 26th international workshop on machine learning for signal processing (MLSP) IEEE, 1-6.

Meral, O. (2008). Doğrusal Öngörülü Kodlama ve Adaptif Algoritma Tabanlı Konuşmacı Tanıma. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.

Mokgonyane, T. B., Sefara, T. J., Modipa, T. I., Mogale, M. M., Manamela, M. J., & Manamela, P. J. (2019). Automatic speaker recognition system based on machine learning algorithms. In 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA) IEEE, 141-146).

Nagrani, A., Chung, J.S. & Zisserman, A. (2017). Voxceleb: A large- scale speaker identification dataset. Interspeech, 2616-2620.

Nagrani, A., Chung, J.S., Xie, W. & Zisserman, A. (2020). Voxceleb: A large- scale speaker verification in the wild. ElsevierLtd. Computer Speech and Language, 60.

Nakagawa, S., Wang, L., & Ohtsuka, S. (2011). Speaker identification and verification by combining MFCC and phase information. IEEE transactions on audio, speech, and language processing, 20(4), 1085-1095.

Nandan, D., Singh, M. K., Kumar, S., & Yadav, H. K. (2022). Speaker Identification Based on Physical Variation of Speech Signal. IIETA, 39(2), 711- 716.

Öztürk, K. (2024). Makine Öğrenme Algoritmalarıyla Konuşmacı Tanılaması. Doğuş Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi, İstanbul.

Pandey, D., Niwaria, K., & Chourasia, B. (2019). Machine learning algorithms: a review. IRJET, 6(2).

Rituerto-González, E., Mínguez-Sánchez, A., Gallardo-Antolín, A., & Peláez-Moreno, C. (2019). Data augmentation for speaker identification under stress conditions to combat gender-based violence. Applied sciences, 9(11), 2298.

Sari, Y. (2017). Ses Kayıtlarındaki İnsan Seslerinin Makine Öğrenmesi Yöntemleriyle Tanınması ve Sınıflandırılması. Karabük Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Karabük.

Singh, M. K. (2023). A text independent speaker identification system using ANN, RNN, and CNN classification technique. Multimedia Tools and Applications, 83(16), 48105-48117.

Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. In 2016 international conference on computing communication control and automation (ICCUBEA) IEEE, 1-7).

Sugiharto, A., Wirawan, P. W., Nugroho, F. A., Yudantomo, S. A., Pratama, R., Mahe, M. H., Nabila, A. M. & Aldira, M. (2022). Comparison of SVM, random forest and KNN classification by using HOG on traffic sign detection. In 2022 6th International Conference on Informatics and Computational Sciences (ICICoS) IEEE, 60-65.

Tiwari, M., & Verma, D. K. (2024). Enhanced text-independent speaker recognition using MFCC, Bi-LSTM, and CNN-based noise removal techniques. International Journal of Speech Technology, 27(4), 1013-1026.

Turan, A. K., & Polat, H. (2024). Yarı denetimli makine öğrenmesi yöntemini kullanarak müzik türlerinin tespiti. Gazi University Journal of Science Part C:Design and Technology, 1-1.

Vuppala, A. K., & Rao, K. S. (2013). Speaker identification under background noise using features extracted from steady vowel regions. International Journal of Adaptive Control and Signal Processing, 27(9), 781-792.

Yao, J., Chen, X., Zhang, X. L., Zhang, W. Q., & Yang, K. (2023). Symmetric Saliency-Based Adversarial Attack to Speaker Identification. IEEE Signal Processing Letters, 30, 1-5.

Ye, F., & Yang, J. (2021). A Deep Neural Network Model for Speaker Identification. Applied Sciences, 11(8), 3603.

Yilmaz, B. H. (2022). Sinyal Görüntü Dönüşümleri Yardımıyla Çok Modlu Duygu Tanıma Sisteminin Geliştirilmesi. K.T.Ü., Fen Bilimleri Enstitüsü, Doktora Tezi, Trabzon.

Yilmaz, B. H., Yilmaz, C. M., & Kose, C. (2020). Diversity in a signal-to-image transformation approach for EEG-based motor imagery task classification. Medical & biological engineering & computing, 58(2), 443-459.

Yujin, Y., Peihua, Z., & Qun, Z. (2010). Research of speaker recognition based on combination of LPCC and MFCC. In 2010 IEEE international conference on intelligent computing and intelligent systems, 3, 765-767.

Zhang, S., Zhao, X., & Tian, Q. (2022). Spontaneous speech emotion recognition using multiscale deep convolutional LSTM. IEEE Transactions on Affective Computing, 13(2), 680-688.

ÖZGEÇMİŞ

Büşra ORAN, İlk ve orta öğrenimini Osman Altıntaş İlköğretim Okulu'nda, Lise öğrenimini 88. Yıl Cumhuriyet Anadolu Lisesi'nde tamamladı. 2014- 2015 öğretim yılında Karadeniz Teknik Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü'nde lisans programına başladı ve 2021 yılında bu bölümden mezun oldu. 2021- 2022 öğretim yılında Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı'nda tezli yüksek lisans öğrenimine başladı.

Yayınlar

1. Oran, B. & Köse, C. (2025). Açı Genlik Dönüşümü Kullanarak Konuşmacı Tanımlama. International Multidisciplinary Scientific Research Congress, Bildiriler Kitabı, 1, 79-89.