TÜRKÇE SES METİNLERİNDEN CİNSİYET TANIMA İÇİN NÖRAL DEVRE POLİTİKALARI TABANLI YENİ BİR YÖNTEM

YÜKSEK LİSANS TEZİ

Sevda OLGUN

Danışman

Dr. Öğr. Üyesi Caner BALIM

BİLGİSAYAR ANABİLİM DALI

Temmuz 2026

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TÜRKÇE SES METİNLERİNDEN CİNSİYET TANIMA İÇİN NÖRAL DEVRE POLİTİKALARI TABANLI YENİ BİR YÖNTEM

Sevda OLGUN

Danışman
Dr. Öğr. Üyesi Caner BALIM

BİLGİSAYAR ANABİLİM DALI

Temmuz 2026

TEZ ONAY SAYFASI

Sevda OLGUN tarafından hazırlanan "Türkçe Ses Metinlerinden Cinsiyet Tanıma İçin Nöral Devre Politikaları Tabanlı Yeni Bir Yöntem" adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 13 / 07 / 2026 tarihinde aşağıdaki jüri tarafından **oy birliği** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Dr. Öğr. Üyesi Caner BALIM

Başkan : Dr. Öğr. Üyesi Emrah ÖZKAYNAK
Karabük Üniversitesi .....

Üye : Dr. Öğr. Üyesi Naim KARASEKRETER
Afyon Kocatepe Üniversitesi .....

Üye : Dr. Öğr. Üyesi Caner BALIM
Afyon Kocatepe Üniversitesi .....

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
..... /..... /..... tarih ve
……. sayılı kararıyla onaylanmıştır.

…….
Prof. Dr. Bekir YALÇIN
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

13 / 07 / 2026

Sevda OLGUN

ÖZET

Yüksek Lisans Tezi

TÜRKÇE SES METİNLERİNDEN CİNSİYET TANIMA İÇİN NÖRAL DEVRE POLİTİKALARI TABANLI YENİ BİR YÖNTEM

Sevda OLGUN
Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Caner BALIM

Konuşma hem dilsel hem de dil dışı bilgi taşıyan, insan iletişiminin en doğal ve etkili yollarından biridir. Cinsiyet sınıflandırması, biyometrik kimlik doğrulama ve kişiselleştirilmiş insan-bilgisayar etkileşimi gibi birçok uygulamada önemli rol oynamaktadır. Bu çalışma, Türkçe konuşma verileri üzerinde cinsiyet sınıflandırması yapmak amacıyla biyolojik sinir sistemlerinden esinlenen Nöral Devre Politikaları'na (NCP) dayalı hibrit bir derin öğrenme modeli önermektedir. Konuşma sinyallerinin spektral ve zamansal özelliklerini yakalamak için Mel-Frekans Kepstral Katsayıları (MFCC) ile log-Mel spektrogram öznitelikleri birleştirilmiş ve Conv1D katmanları aracılığıyla öğrenilmiştir. Uzun vadeli zamansal bağımlılıklar ise NCP mimarisi kapsamında tanımlanan Sıvı Zaman Sabiti (LTC) hücreleriyle modellenmiştir. Mozilla Common Voice 12.0 Türkçe veri kümesi üzerinde 5 katlı tabakalı çapraz doğrulama uygulanmıştır. Conv1D+NCP modeli %99,29 doğruluk ve %99,28 F1 puanı elde ederken UKSB (Uzun Kısa Süreli Bellek) tabanlı model daha düşük performans göstermiştir. NCP tabanlı model daha az parametre ile yüksek doğruluk ve hesaplama verimliliği sunarak gerçek zamanlı uygulamalar için güçlü bir alternatif olduğunu göstermektedir.

2026, ix + 59 Sayfa

Anahtar Kelimeler: Konuşma tabanlı cinsiyet sınıflandırılması, Derin öğrenme, Nöral Devre Politikaları.

ABSTRACT

M.Sc. Thesis

A NOVEL NEURAL CIRCUIT POLICIES-BASED APPROACH FOR GENDER RECOGNITION FROM TURKISH SPEECH TEXTS

Sevda OLGUN

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Computer

Supervisor: Asst. Prof. Caner BALIM

Speech is one of the most natural and effective forms of human communication, carrying both linguistic and non-linguistic information. Gender classification plays an important role in various applications such as biometric authentication and personalized human–computer interaction. This study proposes a hybrid deep learning model based on Neural Circuit Policies (NCP), inspired by biological neural systems, for gender classification using Turkish speech data. To capture both spectral and temporal characteristics of speech signals, Mel-Frequency Cepstral Coefficients (MFCC) and log-Mel spectrogram features are combined and learned through Conv1D layers. Long-term temporal dependencies are modeled using Liquid Time Constant (LTC) cells defined within the NCP architecture. A 5-fold stratified cross-validation is applied on the Mozilla Common Voice 12.0 Turkish dataset. The proposed Conv1D+NCP model achieves 99.29% accuracy and a 99.28% F1-score, while the UKSB (Long Short-Term Memory) based model shows lower performance. With fewer parameters, the NCP-based model demonstrates high accuracy and computational efficiency, making it a strong alternative for real-time applications.

2026, i**x + 59 pages**

Keywords: Speech-based gender classification, Deep learning, Neural Circuit Policies.

TEŞEKKÜR

Bu araştırmanın konusu, deneysel çalışmaların yönlendirilmesi, sonuçların değerlendirilmesi ve yazımı aşamasında yapmış olduğu katkılarından dolayı tez danışmanım Sayın Dr. Öğr. Üyesi Caner BALIM'a teşekkür ederim.

Bu araştırma boyunca maddi ve manevi desteklerinden dolayı aileme teşekkür ederim.

Sevda OLGUN

Afyonkarahisar 2026

İÇİNDEKİLER DİZİNİ

Sayfa

ÖZET i

ABSTRACT ii

TEŞEKKÜR iii

İÇİNDEKİLER DİZİNİ iv

SİMGELER ve KISALTMALAR DİZİNİ vi

ŞEKİLLER DİZİNİ viii

ÇİZELGELER DİZİNİ ix

1. GİRİŞ 1

1.1 Tezin Amacı ve Özgün Değeri 1

1.2 Tezin Organizasyonu 3

2. LİTERATÜR ÖZETİ 4

2.1 Literatür Taraması 4

3. MATERYAL ve METOT 9

3.1 Veri Seti 9

3.2 Metot 11

3.2.1 Ses ve Konuşma Sinyalleri 11

3.2.2 Ön İşleme Süreci 14

3.2.3 Mel Frekans Kepstral Katsayıları (MFCC) 14

3.2.4 Logaritmik Mel Spektrogram (log-Mel-Spectrogram) 15

3.2.5 Öznitelik Çıkarımı 16

3.2.6 Derin Öğrenme Yöntemleri 18

3.2.6.1 Evrişimsel Sinir Ağları (ESA) 19

3.2.6.2 Tekrarlayan Sinir Ağları (TSA) 22

3.2.6.3 Uzun Kısa Süreli Bellek (UKSB) 24

3.2.6.4 Nöral Devre Politikaları (NCP) 26
3.2.6.5 Tam Bağlantılı Katmanlar (Full Connected Layers) 28
3.2.6.6 Sigmoid Aktivasyon Fonksiyonu 29
3.3 Eğitim Süreci 29
3.4 Performans Ölçütleri 31
3.5 Tabakalı k-katlı (Stratified k-Fold) Çapraz Doğrulama ve Konuşmacı Bazlı Ayrık Çapraz Doğrulama (Leave-One-Speaker-Out, LOSO) 33
4. BULGULAR 36
4.1 Deneysel Kurulum 36
4.2 Elde Edilen Sonuçlar 36
5. TARTIŞMA ve SONUÇ 45
5.1 Genel Değerlendirme 45
5.2 Diğer Yöntemlerle Karşılaştırma 46
5.3 Katkılar ve Sınırlılıklar 50
5.4 Gelecek Çalışmalar İçin Öneriler 51
6. KAYNAKLAR 52
ÖZGEÇMİŞ 59

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler

%	Yüzde
b	Bias terimi
f	Aktivasyon fonksiyonu
h_t	Mevcut gizli durum
h_{t-1}	Önceki gizli durum
i, j	Giriş verisi üzerindeki satır ve sütun indisleri
K	Filtre matrisi
m, n	Filtre matrisi indisleri
W_{hh}	Gizli durum ağırlık matrisi
W_{xh}	Giriş-gizli durum ağırlık matrisi
x_t	Mevcut giriş verisi
γ	Özellik haritası

Kısaltmalar

ACO	Ant Colony Optimization (Karınca Koloni Optimizasyonu)
BOA	Butterfly Optimization Algorithm (Balina Optimizasyonu Algoritması)
DCT	Discrete Cosine Transform (Ayrık Kosinüs Dönüşümü)
DVM	Support Vector Machine (Destek Vektör Makineleri)
ESA	Evrişimsel Sinir Ağları
F1	F1-Score (F1-Skoru)
FFT	Fast Fourier Transform (Hızlı Fourier Dönüşümü)
GMM	Gaussian Mixture Model (Gaussian Karışım Modeli)
GPU	Graphics Processing Unit (Grafik İşlem Birimi)
GRU	Gated Recurrent Unit (Geçitli Tekrarlayan Birim)
HMM	Hidden Markov Model (Gizli Markov Modeli)
KNN	K-Nearest Neighbor (K En Yakın Komşu Algoritması)
LOSO	Leave One Speaker Out (Konuşmacı Bazlı Çapraz Doğrulama)
LTC	Liquid Time Constant (Sıvı Zaman Sabiti)

MFCC	Mel-Frequency Cepstral Coefficients (Mel-Frekans Kepstral Katsayıları)
MLP	Multi Layer Perceptron
MP3	MPEG Layer-3
NCP	Neural Circuit Policies (Nöral Devre Politikaları)
PSO	Particle Swarm Optimization
ReLU	Rectified Linear Unit
RF	Random Forest
ROC	Receiver Operating Characteristic
SGD	Stochastic Gradient Descent
SSA	Salp Swarm Algorithm
STFT	Short-Time Fourier Transform
TSA	Tekrarlayan Sinir Ağları
UKSB	Uzun Kısa Süreli Bellek
VQ	Vector Quantization
YSA	Yapay Sinir Ağları

ŞEKİLLER DİZİNİ

Sayfa

Şekil 3.1 Çalışmanın blok diyagramı 9

Şekil 3.3 Kadın ve erkeğe ait örnek metinlerin seslendirilmesi sonucu oluşan ses sinyalleri 11

Şekil 3.4 Konuşma üretim sürecinin genel yapısı 12

Şekil 3.5 Erkek ve kadın bireylerde ses tellerinin yapısal farklılıklarına bağlı olarak oluşan temel frekans değişiminin gösterimi 13

Şekil 3.6 MFCC özellik çıkarım süreci 15

Şekil 3.7 log-Mel spektrogram özelliklerinin çıkarılma süreci 16

Şekil 3.8 Kadın (a) ve erkek(b) konuşmacılara ait örnek ses sinyallerinin dalga formu, MFCC ve log-Mel spektrogram öznitelikleri 17

Şekil 3.9 Evrişimsel Sinir Ağı Mimarisi 20

Şekil 3.10 Tekrarlayan Sinir Ağları Mimarisi (Çelebi & Fidan, 2024) 24

Şekil 3.11 UKSB modeli blok diyagramı (Ayyüce Kızrak & Bolat, 2018)..... 25

Şekil 3.12 NCP mimarisinin temel aşamaları 27

Şekil 3.13 NCP tabanlı hibrit modelin genel yapısı 30

Şekil 3.14 5 katmanlı çapraz doğrulama tekniği 34

Şekil 4.1 1. katman için elde edilen sınıflandırma sonuçlarının karmaşıklık matrisi 40

ÇİZELGELER DİZİNİ

Sayfa

Tablo 3.1 Cinsiyete göre veri dağılımı 10

Tablo 3.2 Çalışmada kullanılan veri setindeki örnek metinler ve cinsiyetler..... 10

Tablo 3.3 Kullanılan akustik öznitelikler 17

Tablo 3.4 Model eğitim parametreleri 31

Tablo 3.5 Sınıflandırmada kullanılan performans metriklerinin hesaplaması..... 32

Tablo 4.1 NCP tabanlı model kullanılarak Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar..... 37

Tablo 4.2 UKSB tabanlı model kullanılarak Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar..... 37

Tablo 4.3 Conv1D+Tam Bağlantılı Katman hibrit modeli kullanılarak Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar..... 39

Tablo 4.4 TSA tabanlı hibrit modele ait Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar..... 40

Tablo 4.5 MFCC ve log-Mel öznitelikleri kullanılarak Conv1D tabanlı mimarilerin performans karşılaştırması..... 42

Tablo 4.6 Konuşmacıdan bağımsız (LOSO) doğrulama altında farklı Conv1D tabanlı mimarilerin sınıflandırma performansları 44

Tablo 5.1 Konuşmadan Cinsiyet Tanıma yöntemlerinin literatür tabanlı karşılaştırılması 47

1. GİRİŞ

1.1 Tezin Amacı ve Özgün Değeri

Konuşma sinyali zamana bağlı, yüksek boyutlu ve durağan olmayan bir yapıya sahiptir. Ses tellerinin titreşimi, konuşma hızı, vurgu, tonlama ve artikülasyon gibi özellikler zaman içerisinde sürekli değişim göstermektedir. Ayrıca konuşma sinyali frekans, enerji ve spektral yapı bakımından karmaşık örüntüler içermektedir. Bu örüntüler bireyler arasında farklılık gösterebilmektedir. Ortam gürültüsü, kayıt cihazı farklılıkları ve konuşmacının duygusal durumu gibi etkenler de sinyal yapısını doğrudan etkileyerek analiz sürecini daha karmaşık hâle getirmektedir. Bu nedenle konuşmadan cinsiyet tanıma problemi yalnızca basit bir sınıflandırma işlemi olarak değerlendirilemez. Konuşmacıya ait ayırt edici özelliklerin doğru biçimde belirlenebilmesi için konuşma sinyalinin hem frekans düzlemindeki spektral özelliklerinin hem de zaman boyutundaki zamansal bağımlılıklarının birlikte analiz edilmesi gerekmektedir. Dolayısıyla konuşmadan cinsiyet belirleme süreci çok boyutlu bir örüntü tanıma ve modelleme problemi olarak ele alınmaktadır.

Konuşmadan cinsiyet tanıma sistemleri; biyometrik güvenlik uygulamaları, konuşmacı ayrıştırma, çağrı merkezi otomasyonu ve kişiselleştirilmiş insan–bilgisayar etkileşimi gibi birçok alanda önemli bir rol oynamaktadır (Tranter ve Reynolds 2006, Alsulaiman vd. 2011, Alkhammash vd. 2022). Özellikle ses tabanlı kimlik doğrulama sistemlerinde cinsiyet bilgisinin önceden belirlenmesi arama alanını daraltarak doğrulama sürecinin daha hızlı ve daha verimli gerçekleştirilmesini sağlamaktadır (Alsulaiman vd. 2011). Bunun yanı sıra dijital asistan sistemlerinde ve içerik öneri platformlarında kullanıcıya özgü deneyim sunulabilmesi açısından da cinsiyet bilgisi önemli bir faktör olarak değerlendirilmektedir (Alkhammash vd. 2022).

Literatürde konuşmadan cinsiyet tanıma problemi için hem geleneksel makine öğrenmesi hem de derin öğrenme temelli yaklaşımlar kullanılmış ve yüksek doğruluk oranları rapor edilmiştir (Yücesoy ve Nabiyev 2013, Mohammed ve Al-Irhayim 2021, Hızlısoy vd. 2022, Younis vd. 2023). Geleneksel yöntemler çoğunlukla konuşma sinyalinden çıkarılan

özet akustik özelliklerin sınıflandırılmasına dayanmaktadır. Ancak bu yöntemler konuşma sinyalinin zamana bağlı dinamik doğasını yeterince temsil edememekte ve uzun vadeli zamansal ilişkileri modelleme konusunda sınırlı kalmaktadır. Derin öğrenme tabanlı yaklaşımlar ise konuşma sinyalindeki karmaşık örüntüleri ve zamansal bağımlılıkların öğrenilmesinde daha başarılı sonuçlar sunabilmektedir (Mohammed ve Al-Irhayim 2021a). Buna karşın Tekrarlayan Sinir Ağları (TSA), Uzun Kısa Süreli Bellek (UKSB) ve Geçitli Tekrarlayan Birimler (GRU) gibi derin öğrenme modelleri yüksek parametre sayısı, uzun eğitim süreleri ve yüksek hesaplama maliyeti gibi sınırlılıklara sahiptir (Purwins vd. 2019, Malashin vd. 2024). Bu durum özellikle gerçek zamanlı sistemlerde ve sınırlı donanım kaynaklarına sahip uygulamalarda önemli bir kısıt oluşturmaktadır.

Son yıllarda geliştirilen NCP (Nöral Devre Politikaları) mimarisi, zamansal verilerin sürekli zaman yaklaşımıyla modellenmesine imkân tanıması ve daha düşük parametre sayısı ile çalışabilmesi nedeniyle dikkat çekmektedir (Lechner vd. 2020). Ancak bu mimarinin konuşmadan cinsiyet tanıma probleminde kullanımına yönelik çalışmalar sınırlıdır. Bu bağlamda Türkçe konuşma verileri üzerinde NCP tabanlı hibrit bir modelin geliştirilmesi ve değerlendirilmesi literatüre özgün bir katkı sunmaktadır.

Çalışma kapsamında konuşma sinyalinden ayırt edici özellikler elde edebilmek amacıyla Mel Frekans Kepstral Katsayıları (MFCC) ve log-Mel spektrogram yapıları kullanılmıştır. Bu yapılar, insan işitme sisteminin frekans algısını yansıtan ve konuşma sinyalinin hem frekans hem de zaman boyutundaki özelliklerini ortaya koyan güçlü öznitelik çıkarım yöntemleridir. Elde edilen bu öznitelikler spektral örüntülerin öğrenilmesinde etkili olan Evrişimsel Sinir Ağları (ESA) ile işlenmiştir. Zamansal bağımlılıkların modellenmesi için ise sürekli zaman yaklaşımına dayanan ve düşük parametre sayısı ile öne çıkan NCP mimarisi kullanılmıştır. Böylece model hem yerel örüntüleri hem de uzun vadeli zamansal ilişkileri bütüncül bir şekilde öğrenebilen hibrit bir yapıya kavuşturulmuştur. Modelin performansı doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall) ve F1-skoru (F1-Score) gibi farklı değerlendirme ölçütleri kullanılarak analiz edilmiştir. Ayrıca modelin genelleme yeteneğini incelemek amacıyla farklı doğrulama stratejileri kullanılarak kapsamlı deneyler gerçekleştirilmiştir. Bu çalışma hem zamansal modelleme gücü yüksek hem de hesaplama açısından verimli bir yaklaşım önererek konuşma tabanlı

cinsiyet sınıflandırma alanındaki eksikliklere katkı sunmayı amaçlamaktadır.

1.2 Tezin Organizasyonu

Bu tez altı ana bölümden oluşmaktadır. Birinci bölümde çalışmanın amacı, önemi ve özgün değeri açıklanmıştır. İkinci bölümde konuşmadan cinsiyet tanıma alanına ilişkin literatürde yer alan çalışmalar incelenmiş ve mevcut yaklaşımlar değerlendirilmiştir. Üçüncü bölümde çalışmada kullanılan veri seti, ön işleme aşamaları, öznitelik çıkarım yöntemleri, derin öğrenme mimarileri ve önerilen Conv1D+NCP tabanlı hibrit model ayrıntılı olarak sunulmuştur. Dördüncü bölümde deneysel çalışmalar sonucunda elde edilen bulgular ve performans değerlendirmeleri sunulmuştur. Beşinci bölümde elde edilen sonuçlar tartışılmış, çalışmanın katkıları ve sınırlılıkları değerlendirilmiş ve gelecekte yapılabilecek çalışmalar için öneriler sunulmuştur. Altıncı bölümde ise tez kapsamında yararlanılan kaynaklara yer verilmiştir.

2. LİTERATÜR ÖZETİ

2.1 Literatür Taraması

Son yıllarda konuşma sinyallerinden cinsiyet, yaş ve duygu gibi paralinguistik bilgilerin otomatik olarak tanımlanması, insan–bilgisayar etkileşiminden biyometrik güvenliğe kadar uzanan geniş uygulama alanları nedeniyle önemli bir araştırma konusu haline gelmiştir. Bu alandaki çalışmalar, kullanılan giriş temsilleri ve sınıflandırma yaklaşımları bakımından farklılık göstermektedir. Giriş temsili açısından bazı çalışmalarda MFCC, log-Mel spektrogram, temel frekans, enerji ve benzeri akustik öznitelikler kullanılırken bazı çalışmalarda ham konuşma sinyalleri doğrudan modele verilmektedir. Sınıflandırma yöntemi bakımından ise geleneksel makine öğrenmesi algoritmaları, derin öğrenme modelleri ve biyolojik sinir sistemlerinden esinlenen mimariler öne çıkmaktadır.

Geleneksel makine öğrenmesi yaklaşımları, konuşma sinyalinden önceden çıkarılan akustik özniteliklerin DVM, GMM, KNN, karar ağaçları ve benzeri algoritmalarla sınıflandırılmasına dayanmaktadır. Bu yöntemler, sınırlı büyüklükteki veri kümelerinde uygulanabilmeleri ve görece düşük hesaplama maliyetleri nedeniyle uzun süredir kullanılmaktadır. Bununla birlikte elde edilen başarım, seçilen özniteliklerin temsil gücüne önemli ölçüde bağlıdır. Ayrıca bu yöntemlerin konuşma sinyalindeki karmaşık örüntüleri ve uzun süreli zamansal ilişkileri modelleme yetenekleri sınırlı kalabilmektedir.

MFCC, geleneksel yaklaşımlarda en yaygın kullanılan özniteliklerden biridir. Younis ve diğerleri, TIMIT veri setinden elde ettikleri 280 adet ses kaydında MFCC tabanlı öznitelikleri DVM algoritması ile sınıflandırmış ve %96,45 oranında doğruluk elde etmişlerdir (Younis vd. 2023). Chaudhari ve Kagalkar, MFCC özelliklerini GMM ile modellemiş, ardından DVM sınıflandırıcısı ile %80,3 doğruluk elde etmiştir (Chaudhari ve Kagalkar 2015).Yücesoy ve Nabiyev ise MFCC ve GMM tabanlı, metinden bağımsız bir yaklaşımı TIMIT veri seti üzerinde uygulayarak %97,76'ya varan doğruluk elde etmişlerdir (Yücesoy ve Nabiyev 2013).

Djemili ve diğerleri, MFCC öznitelikleri kullanılarak GMM, MLP, VQ ve LVQ sınıflandırıcılarının performansı değerlendirilmiş ve IViE veri seti üzerinde konuşmacı başına yalnızca 1 saniyelik konuşma verisi ile %96,4 doğruluk oranı elde edilmiştir (Djemili vd. 2012). Ramdinmawii ve Mittal, temel frekans, sinyal enerjisi ve MFCC özniteliklerini DVM ile sınıflandırarak başarılı sonuçlar elde etmiştir (Ramdinmawii ve Mittal 2016). Archana ve Malleswari ise MFCC, enerji entropisi ve çerçeve enerjisi gibi akustik öznitelikleri YSA ve DVM modelleriyle değerlendirmiş; DVM tabanlı yaklaşımın daha başarılı olduğunu belirlemiştir (Archana ve Malleswari 2015). Bu çalışmalar, MFCC özniteliklerinin tek başına ya da farklı akustik göstergelerle birlikte kullanıldığında etkili sonuçlar sağlayabildiğini göstermektedir.

Farklı sınıflandırma algoritmalarının karşılaştırıldığı çalışmalarda veri setinin yapısı, öznitelik seçimi ve model türünün performans üzerinde belirleyici olduğu görülmektedir. Yücesoy, Mozilla Common Voice veri setinin Türkçe alt kümesinden elde edilen konuşma kayıtlarından MFCC tabanlı öznitelikler çıkarılmış ve DVM, KNN, RF ile ESA modellerinin konuşmadan cinsiyet tanıma performansları karşılaştırılmıştır. Deneysel sonuçlara göre, %98,67 doğruluk oranı ile en yüksek başarı ESA modeli tarafından elde edilmiştir (Yücesoy 2024a). Bakır, Almanca konuşma verileri kullanılarak farklı MFCC boyutlarının sınıflandırma performansına etkisi değerlendirilmiş en başarılı sonuçlar HMM yöntemiyle ve yüksek MFCC boyutlarında elde edilmiştir. Ayrıca konuşma örneklerindeki kelime sayısının artması tanıma başarısını olumlu yönde etkilemiştir (Bakır 2016).

Türkçe konuşma verileri üzerinde gerçekleştirilen bir başka çalışmada Hızlısoy ve diğerleri, MFCC ve spektrogram tabanlı öznitelikleri kullanmış ve 894 ses kaydı üzerinde en yüksek doğruluğu %89 ile lojistik regresyon modelinde elde etmiştir (Hızlısoy vd. 2022). Zaman ve diğerleri, konuşma sinyallerinden elde ettikleri 20 istatistiksel öznitelik ile çeşitli algoritmaları karşılaştırmış; CatBoost algoritmasıyla %96,4 doğruluk oranı elde etmişlerdir (Zaman vd. 2021). Munoli ve diğerleri ise MFCC tabanlı öznitelikleri CatBoost, XGBoost, SGD ve karar ağaçları ile karşılaştırmış ve %88–90 aralığında doğruluk değerleri elde etmiştir (Munoli vd. 2023a).

Öznitelik seçiminin etkisini inceleyen Özbay ve Özbay, metasezgisel optimizasyon algoritmalarına dayalı bir özellik seçimi yöntemi önermiştir. Bu kapsamda Parçacık Sürüsü Optimizasyonu (PSO), Karınca Koloni Optimizasyonu (ACO), Salp Sürüsü Algoritması (SSA) ve Balina Optimizasyonu (BOA) algoritmaları kullanılarak en uygun öznitelik alt kümeleri belirlenmiş ve elde edilen bu öznitelikler ile Naive Bayes ve Karar Ağacı sınıflandırıcılarının performansında belirgin bir artış sağlandığı gözlemlenmiştir (Özbay ve Özbay 2023). Calp, Kaggle veri seti kullanarak KNN, Karar Ağacı, Naive Bayes ve YSA algoritmalarını karşılaştırılmış; farklı eğitim-test oranları altında yapılan deneyler sonucunda en yüksek doğruluk oranı %98,96 ile YSA modeli kullanılarak elde etmiştir (Calp 2025). Yılmaz ve diğerleri, ses verileri üzerinde denetimli öğrenme algoritmaları kullanarak cinsiyet tahmini gerçekleştirmiş ve elde edilen sonuçlar doğrultusunda cinsiyetin %97'nin üzerinde doğruluk oranı ile belirlenebildiğini elde etmiştir (Yılmaz vd. 2022).

GMM süpervektörleri ve DVM tabanlı yaklaşımların değerlendirildiği çalışmalarda farklı sonuçlar elde edilmiştir. Yücesoy ve Nabiyev, GKM süpervektörlerine dayalı DVM yaklaşımını kullanarak sınıflandırma gerçekleştirmiş; MFCC öznitelikleri ile yapılan deneylerde en yüksek doğruluk oranı %92,42 olarak elde edilmiştir (Yücesoy ve Nabiyev 2016). Yücesoy tarafından gerçekleştirilen diğer çalışmalarda MFCC özniteliklerine logaritmik enerji ve delta parametrelerini ekleyerek GMM süpervektörleri ve DVM tabanlı bir yaklaşım önermiştir. Yapılan deneyler sonucunda en yüksek sınıflandırma doğruluğu %60,23 olarak elde edilmiştir (Yücesoy 2021). Ayrıca Yücesoy, MFCC ve delta özniteliklerini kullanarak GMM tabanlı modeller ile DVM sınıflandırıcısını birlikte kullanmış ve en yüksek doğruluk oranını %60,95 olarak rapor etmiştir (Yücesoy 2020).

Derin öğrenme yaklaşımları, önceden belirlenmiş özniteliklere olan bağımlılığı azaltarak konuşma verilerindeki ayırt edici örüntülerin doğrudan öğrenilmesine olanak sağlamaktadır. ESA modelleri yerel ve spektral örüntülerin çıkarılmasında etkili olurken TSA ve UKSB tabanlı yapılar konuşma sinyalindeki zamansal ilişkilerin modellenmesinde kullanılmaktadır. Bununla birlikte bu modellerin yüksek parametre sayısı, uzun eğitim süresi ve hesaplama gereksinimi, özellikle sınırlı donanım kaynaklarına sahip uygulamalarda önemli bir kısıt oluşturabilmektedir.

Yücesoy, konuşma sinyallerinden elde edilen MFCC tabanlı özellikleri kullanarak 1D ve 2D ESA modelleri geliştirmiştir. Mozilla Common Voice Türkçe veri seti üzerinde gerçekleştirilen deneylerde 2D ESA modeli ile %94,40 doğruluk oranı elde edilmiştir (Yücesoy 2023). Bunun yanı sıra Yücesoy, veri artırma yöntemlerinin cinsiyet sınıflandırma performansı üzerindeki etkisini incelemiş; gürültü ekleme, zaman uzatma ve perde kaydırma tekniklerini farklı kombinasyonlarda kullanarak çeşitli modeller geliştirmiştir. Yapılan deneyler sonucunda tüm veri artırma yöntemlerinin birlikte kullanıldığı model ile sınıflandırma doğruluğu %84,583'ten %87,523'e yükseltilerek yaklaşık %3 oranında performans artışı elde edilmiştir (Yücesoy 2024b).

Jorrin-Coz ve diğerleri, ESA ve UKSB tabanlı modelleri farklı konuşma veri kümelerinde değerlendirmiş ve Mozilla Common Voice veri setinde EfficientNet modeliyle %99,82 doğruluk elde etmiştir(Jorrin-Coz vd. 2025). Mohammed ve Al-Irhayim, farklı dillere ait konuşma verileri üzerinde Bayes, KNN, GMM, DVM ve derin öğrenme modellerini karşılaştırarak derin öğrenme yaklaşımlarının daha yüksek doğruluk sağladığını ortaya koymuştur (Mohammed ve Al-Irhayim 2021a). Kone ve diğerleri, beş gizli katmana sahip bir derin öğrenme mimarisi kullanarak %91 doğruluk oranı ile cinsiyet tahmini gerçekleştirmiştir (Kone vd. 2023). Tümen, ani sivri uçlu sinir ağları (spiking neural networks) tabanlı bir model önermiş ve yapılan deneyler sonucunda %98,84 doğruluk oranı elde etmiştir (Tümen 2022).

Literatürde sınıflandırma modelinin yanı sıra konuşma sinyalini temsil eden özniteliklerin seçimine de önem verilmektedir. MFCC'nin temel frekans, enerji, spektrogram ve diğer spektral göstergelerle birleştirilmesi konuşmacılar arasındaki akustik farklılıkların daha kapsamlı biçimde temsil edilmesini sağlayabilmektedir. Ayrıca konuşma süresi, frekans aralığı ve bant seçimi gibi deneysel değişkenler de sınıflandırma performansını etkileyebilmektedir. Safavi ve diğerleri, çocuk konuşmaları üzerine yaptıkları çalışmada farklı yöntemleri karşılaştırmış ve %79,18 doğruluk elde etmiş, en iyi sonuçların belirli frekans aralıklarında elde edildiğini belirtmiştir (Safavi vd. 2018).

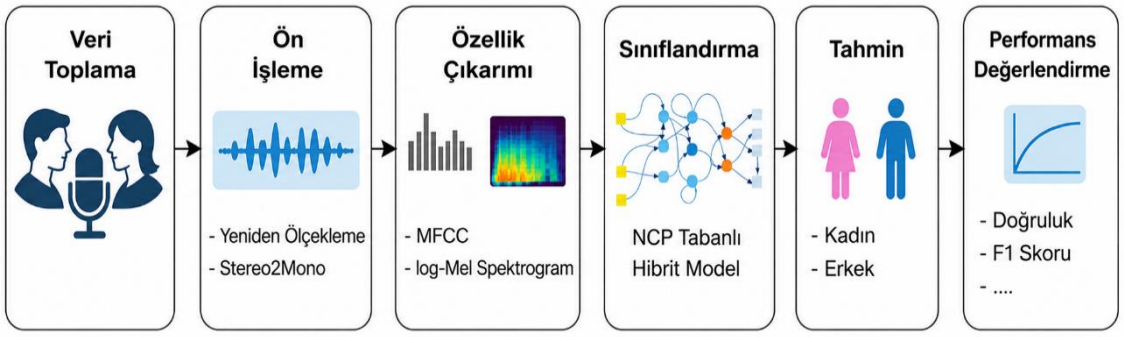
İncelenen çalışmalar, MFCC'nin konuşmadan cinsiyet sınıflandırmasında yaygın ve etkili bir öznitelik olduğunu; log-Mel spektrogram, temel frekans ve enerji gibi tamamlayıcı özelliklerle birlikte kullanılmasının temsil gücünü artırabildiğini göstermektedir. Derin öğrenme modelleri genellikle yüksek sınıflandırma performansı sunmasına rağmen model karmaşıklığı ve hesaplama maliyeti bakımından bazı sınırlılıklara sahiptir. Ayrıca biyolojik sinir sistemlerinden esinlenen NCP ve LTC tabanlı yapıların konuşmadan cinsiyet sınıflandırmasındaki kullanımı sınırlı düzeydedir.

Bu doğrultuda çalışmada, MFCC ve log-Mel spektrogram özniteliklerini birlikte değerlendiren Conv1D+NCP tabanlı hibrit bir model önerilmiştir. Conv1D katmanlarıyla yerel akustik örüntülerin, NCP yapısıyla ise zamansal ilişkilerin öğrenilmesi amaçlanmıştır. Önerilen yaklaşım, Türkçe konuşma verileri üzerinde yüksek sınıflandırma başarısı ile düşük model karmaşıklığını birlikte sağlamayı hedeflemektedir.

3. MATERYAL ve METOT

Bu çalışmada, konuşma sinyallerinden cinsiyet tanıma problemi için çok aşamalı bir model mimarisi tasarlanmıştır. Önerilen model veri toplama, ön işleme, özellik çıkarımı, sınıflandırma, tahmin ve performans değerlendirme adımlarını kapsayan bütüncül bir yapı sunmaktadır. Önerilen model mimarisinin genel yapısı Şekil 3.1'de sunulmaktadır.

Öncelikle konuşma sinyallerini öznitelik çıkarımına uygun hale getirmek için bir dizi işlem uygulanmıştır (Bölüm 3.2.2). Öznitelik çıkarımı aşamasında konuşma sinyalinin zamansal ve spektral özelliklerini temsil edebilen MFCC ve log-Mel spektrogram öznitelikleri elde edilmiştir (Bölüm 3.2.3, Bölüm 3.2.4). Çıkarılan bu öznitelikler, sınıflandırma aşamasında kullanılan NCP tabanlı hibrit derin öğrenme modeline giriş olarak verilmiştir (Bölüm 3.2.6.4). Model tarafından elde edilen çıktılar doğrultusunda konuşmacının cinsiyeti kadın veya erkek olarak tahmin edilmiştir. Son aşamada ise model performansı doğruluk, F1-skoru ve diğer değerlendirme ölçütleri kullanılarak analiz edilmiştir.



Şekil 3.1 Çalışmanın blok diyagramı

3.1 Veri Seti

Bu çalışmada, Mozilla Foundation tarafından oluşturulan ve açık kaynaklı bir konuşma verisi koleksiyonu olan Mozilla Common Voice 12.0 sürümüne ait Türkçe veri seti kullanılmıştır (Mozilla Foundation 2024). Mozilla Common Voice veri seti, farklı yaş ve cinsiyet gruplarından gönüllü katılımcılar tarafından kaydedilen konuşma örneklerinden oluşan geniş kapsamlı bir konuşma veri koleksiyonudur. Veri setindeki ses dosyaları MP3

formatında kaydedilmiş ve her bir kaydın meta verileri kayıt altına alınmıştır. Veri setindeki her bir ses kaydı yaş ve cinsiyet gibi demografik bilgileri içermektedir.
Mozilla Common Voice veri seti birçok farklı dili içermekte olup bu çalışmada yalnızca Türkçe konuşma kayıtları kullanılmıştır. Veri setindeki kayıtlar farklı ortam koşullarında ve farklı kayıt cihazları kullanılarak elde edildiği için gerçek dünya konuşma verilerinin çeşitliliğini yansıtmaktadır. Bu durum geliştirilen modelin farklı konuşma koşullarına karşı daha dayanıklı olmasına katkı sağlamaktadır.

Çalışmada güvenilirlik sağlamak amacıyla yalnızca cinsiyet bilgisi erkek veya kadın olarak açıkça etiketlenmiş kayıtlar dikkate alınmış; eksik, hatalı veya tutarsız etiket bilgisine sahip örnekler kapsam dışında bırakılmıştır. Bu işlem sonucunda 22.651 benzersiz ses kaydı kullanılmıştır. Bu kayıtların 11.142 erkek ve 11.509'i kadın konuşmacıya aittir. Cinsiyete göre veri dağılımı Tablo 3.1'de gösterilmiştir. Her bir konuşma kaydı gönüllü katılımcılar tarafından rastgele seçilen kısa ifadelerin seslendirilmesi ile elde edilmiştir.

Tablo 3.1 Cinsiyete göre veri dağılımı

Cinsiyet	**Konuşmacı Sayısı**	**Yüzde (%)**
Erkek	11.142	49.19
Kadın	11.509	50.81
Toplam	22.651	100

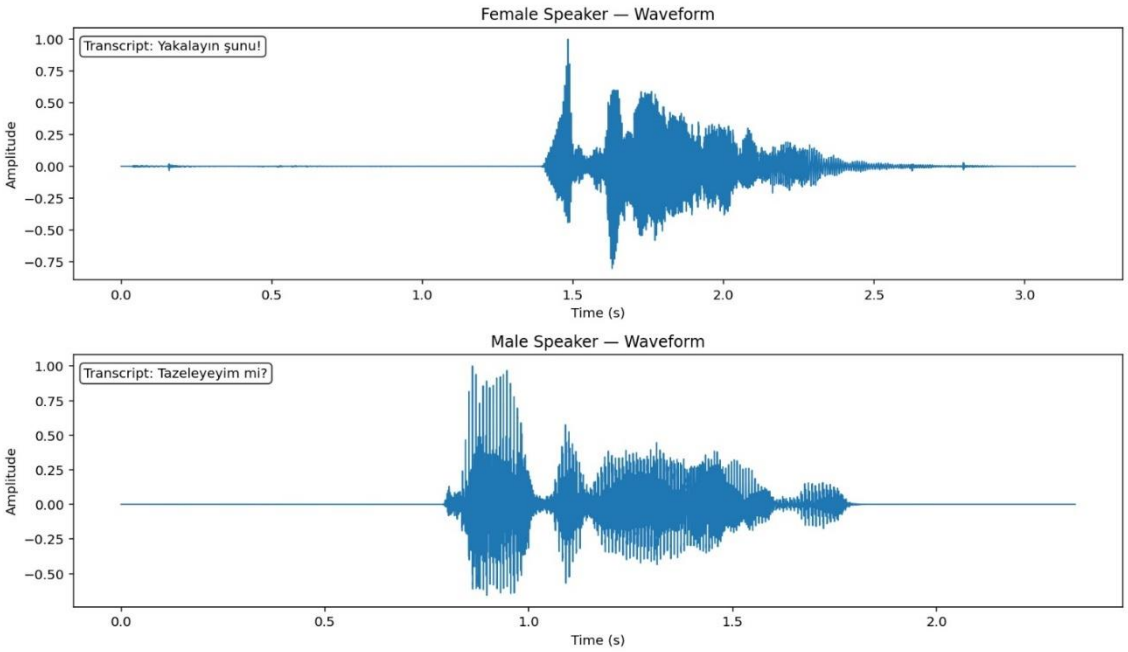
Bu çalışmada kullanılan veri seti farklı konuşmacılardan elde edilen ve dilsel açıdan çeşitli içeriklere sahip metinlerden oluşmaktadır. Veri setinin yapısını daha iyi açıklamak amacıyla örnek cümleler ve bunlara ait cinsiyet etiketleri Tablo 3.2'de sunulmaktadır.

Tablo 3.2 Çalışmada kullanılan veri setindeki örnek metinler ve cinsiyetler

No	**Metin**	**Cinsiyet**
1	Sen ne olacaksın?	Kadın
2	Yakalayın şunu!	Kadın
3	Yüzü kıpkırmızı olmuştu.	Kadın
4	"Mustafa:" "Baş üstüne beyim…" dedi."	Kadın
5	Yüzlerce kiloluk bir ağırlık taşıyormuş gibi aşağıya çekilen elini	Kadın

	uzattı.	
6	Param!	Erkek
7	Tazeleyeyim mi?	Erkek
8	Onu tanıyor musun?	Erkek
9	“Bugün sizin gününüz değil, ne diye kaldınız?”	Erkek
10	Almadan vermek Allah’a mahsustur.	Erkek

Tablo 3.2’de kadın konuşmacı tarafından seslendirilen “Yakalayın şunu!” metnine ait ses sinyali ile erkek konuşmacı tarafından seslendirilen “Tazeleyeyim mi?” metnine ait ses sinyali Şekil 3.3’te verilmiştir.



Şekil 3.2 Kadın ve erkeğe ait örnek metinlerin seslendirilmesi sonucu oluşan ses sinyalleri

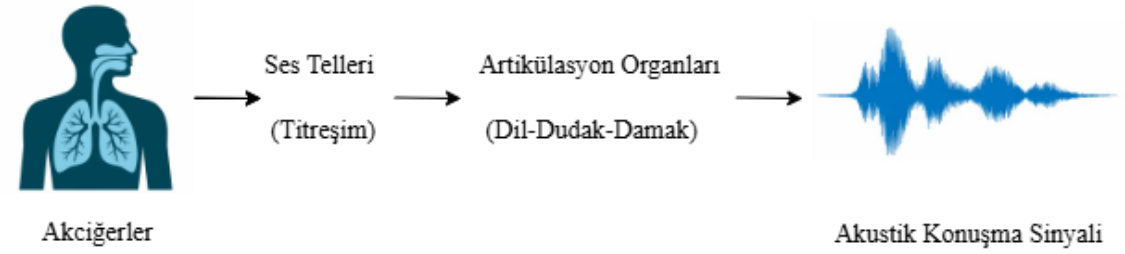
3.2 Metot

3.2.1 Ses ve Konuşma Sinyalleri

Konuşma, insanların düşüncelerini ve duygularını ifade etmelerini sağlayan en doğal iletişim biçimlerinden biridir. İnsan sesi üretimi solunum sistemi, ses telleri ve artikülasyon organlarının birlikte çalışmasıyla gerçekleşen karmaşık bir fizyolojik sürecin sonucudur. Konuşma sırasında akciğerlerden çıkan hava, gırtlakta bulunan ses tellerinin titreşmesine

neden olur. Oluşan bu titreşim dil, dudak, diş ve damak gibi artikülasyon organlarının hareketleriyle şekillendirilerek akustik bir sinyal meydana getirir. Bu süreç sonucunda ortaya çıkan konuşma sinyali, zaman ve frekans alanlarında değişim gösteren dinamik ve çok boyutlu bir yapıya sahiptir (Safavi vd. 2018, Altunbey ve Özbay 2022).

Konuşma üretim süreci genel olarak üç temel aşamada gerçekleşmektedir. İlk aşamada solunum sistemi tarafından oluşturulan hava akımı ses tellerine ulaşır ve burada titreşim oluşur. İkinci aşamada bu titreşim artikülasyon organları tarafından şekillendirilir. Son aşamada ise ortaya çıkan akustik dalgalar çevreye yayılarak konuşma sinyalini oluşturur. Bu sürecin genel yapısı Şekil 3.4'te gösterilmektedir.



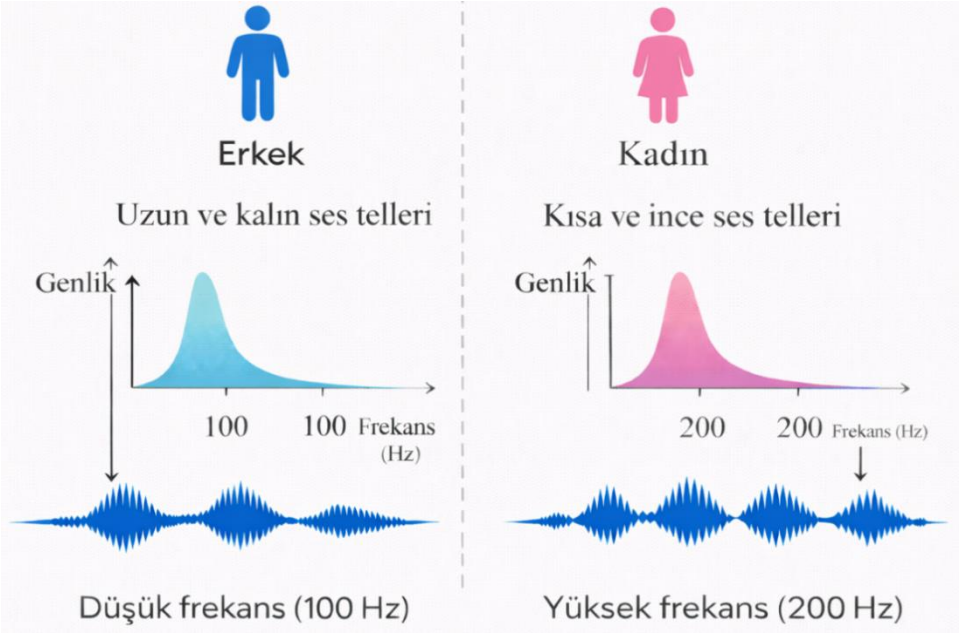
Şekil 3.3 Konuşma üretim sürecinin genel yapısı

Şekil 3.4'te görüldüğü gibi konuşma üretimi, solunum sistemiyle başlayan ve artikülasyon organları ile şekillenen bir süreçtir. Bu süreç sonucunda oluşan konuşma sinyali, çeşitli sinyal işleme teknikleri kullanılarak analiz edilmektedir.

Konuşma sinyali yalnızca dilsel bilgiyi ileten bir yapı değildir. Aynı zamanda konuşmacıya ait çeşitli biyometrik ve paralinguistik özellikleri de içinde barındırır. Konuşmacının cinsiyeti, yaşı, konuşma tarzı, aksanı ve duygusal durumu gibi özellikler konuşma sinyalinde belirli akustik farklılıklar oluşturmaktadır. Bu nedenle konuşma sinyalleri konuşmacı tanıma, cinsiyet belirleme ve duygu analizi gibi birçok konuşma işleme uygulamasında önemli bir veri kaynağı olarak kullanılmaktadır (Chaudhary ve Sharma 2018, Sánchez-Hevia vd. 2022).

Konuşmadan cinsiyet tanıma, konuşma üretim sürecinde ortaya çıkan fizyolojik farklılıkların konuşma sinyaline yansıması temeline dayanmaktadır. Erkek ve kadın bireylerin ses

telleri yapı bakımından farklı özelliklere sahiptir. Erkek bireylerde ses telleri genellikle daha uzun ve kalın bir yapıya sahip olduğundan titreşim frekansı daha düşük aralıklarda oluşmaktadır. Kadın bireylerde ise ses telleri daha kısa ve ince olduğu için titreşim frekansı daha yüksek aralıklarda gerçekleşmektedir (Cider Şener ve Gökçer 2005). Bu durum konuşma sinyalinin temel frekans değerlerinde belirgin farklılıkların oluşmasına neden olmaktadır. Erkek ve kadın bireylerde ses tellerinin yapısal farklılıkları ve buna bağlı olarak oluşan temel frekans değişimi Şekil 3.5'te gösterilmektedir.



Şekil 3.4 Erkek ve kadın bireylerde ses tellerinin yapısal farklılıklarına bağlı olarak oluşan temel frekans değişiminin gösterimi

Konuşma sinyallerinin analizi genellikle zaman alanı ve frekans alanı temelli yöntemlerle gerçekleştirilmektedir. Zaman alanı analizinde sinyalin genlik değişimleri incelenirken, frekans alanı analizinde sinyalin farklı frekans bileşenleri ve spektral dağılımı değerlendirilir. Ancak konuşma sinyalleri durağan olmayan bir yapıya sahip olduğu için analiz işlemleri çoğu zaman kısa zaman aralıklarında gerçekleştirilmektedir.

Konuşma işleme çalışmalarında kısa zamanlı Fourier dönüşümü (STFT), log-Mel spektrogram ve MFCC gibi yöntemler sıklıkla kullanılmaktadır. Bu yöntemler konuşma sinyalinin zaman-frekans özelliklerini ortaya çıkararak sinyalin akustik yapısının daha anlamlı bir şekilde temsil edilmesini sağlar. Elde edilen bu özellikler, makine öğrenmesi ve

derin öğrenme modellerine giriş verisi olarak kullanılarak konuşma tabanlı sınıflandırma problemlerinin çözümünde önemli bir rol oynamaktadır.

3.2.2 Ön İşleme Süreci

Mozilla Common Voice 12 Türkçe konuşma veri kümesinin ham hali, farklı kayıt koşulları ve kayıt cihazlarından kaynaklanan çeşitlilik nedeniyle doğrudan sınıflandırma işlemleri için uygun değildir. Bu nedenle veri setine öncelikle çeşitli ön işleme adımları uygulanmıştır.

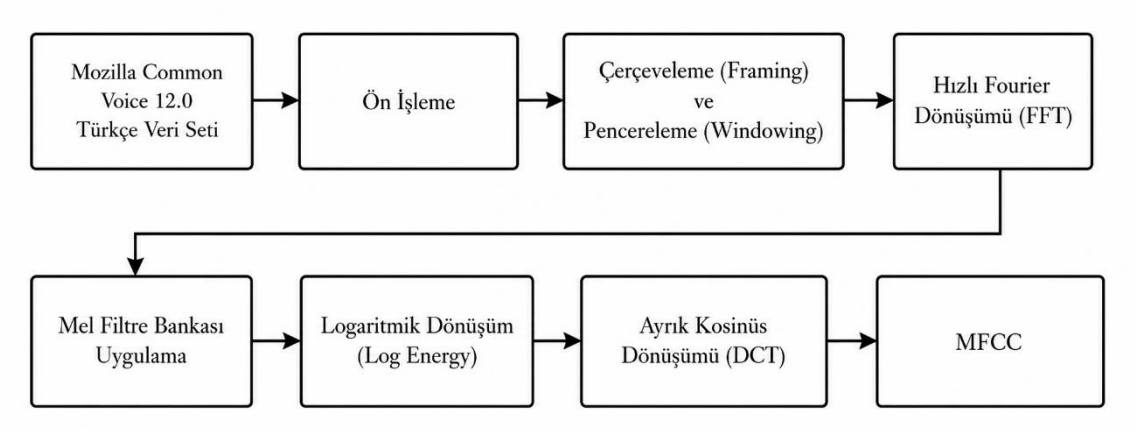
Bu kapsamda tüm ses kayıtları tek kanallı (mono) formata dönüştürülmüş ve örnekleme frekansı 16 kHz değerine yeniden ölçeklendirilmiştir. Böylece farklı cihazlardan elde edilen konuşma kayıtlarının zaman çözünürlüğü standart hale getirilmiş ve modelin öğrenme sürecinde daha tutarlı verilerle çalışması sağlanmıştır. Ayrıca bu standartlaştırma işlemi, veri setindeki farklı kayıt koşullarından kaynaklanan değişkenliklerin azaltılmasına katkı sağlamıştır. Ön işleme adımları sonrasında elde edilen konuşma sinyalleri kullanılarak MFCC ve log-Mel spektrogram öznitelikleri çıkarılmıştır.

3.2.3 Mel Frekans Kepstral Katsayıları (MFCC)

MFCC, insan işitme sisteminin logaritmik frekans algısını modelleyen Mel ölçekli filtre bankaları aracılığıyla ayırt edici akustik özelliklerin çıkarılmasını sağlamaktadır. Bu yöntem, konuşma sinyalindeki belirgin frekans bölgeleri, sesin tınısı ve konuşmacıya özgü spektral zarf gibi önemli akustik özellikleri temsil etmektedir (Zhao ve Shu 2023).

MFCC hesaplama sürecinde konuşma sinyali öncelikle ön işleme adımlarından geçirilerek analiz için uygun hale getirilmektedir. Bu aşamada sinyalin standart bir forma dönüştürülmesi sağlanarak sonraki işlemler için tutarlı bir veri yapısı elde edilmektedir. Daha sonra sinyal, zamana bağlı değişken yapısından dolayı kısa zamanlı analiz yaklaşımıyla belirli uzunluktaki çerçevelere ayrılmakta ve her bir çerçeveye pencereleme işlemi uygulanmaktadır. Bu işlem, çerçeve sınırlarında oluşabilecek süreksizlikleri azaltarak daha sağlıklı bir frekans analizi yapılmasına olanak tanımaktadır. Ardından her bir çerçeveye

Hızlı Fourier Dönüşümü (FFT) uygulanarak sinyalin frekans bileşenleri elde edilmektedir. Elde edilen frekans spektrumu, insan işitme sisteminin doğrusal olmayan frekans algısını yansıtacak şekilde Mel frekans filtre bankaları aracılığıyla yeniden ölçeklendirilmektedir. Bu sayede frekans bileşenleri, insan kulağının algısına daha uygun bir temsil kazanmaktadır. Mel filtre bankası uygulanmış sinyal üzerine logaritmik dönüşüm gerçekleştirilerek genlik bilgisi insan işitme sistemine daha uyumlu hâle getirilmektedir. Son aşamada ise Ayrık Kosinüs Dönüşümü (DCT) uygulanarak özellikler sıkıştırılmakta ve konuşma sinyalinin spektral yapısını temsil eden MFCC katsayıları elde edilmektedir. MFCC özellik çıkarım sürecinin genel aşamaları Şekil 3.6'da gösterilmektedir.



Şekil 3.5 MFCC özellik çıkarım süreci

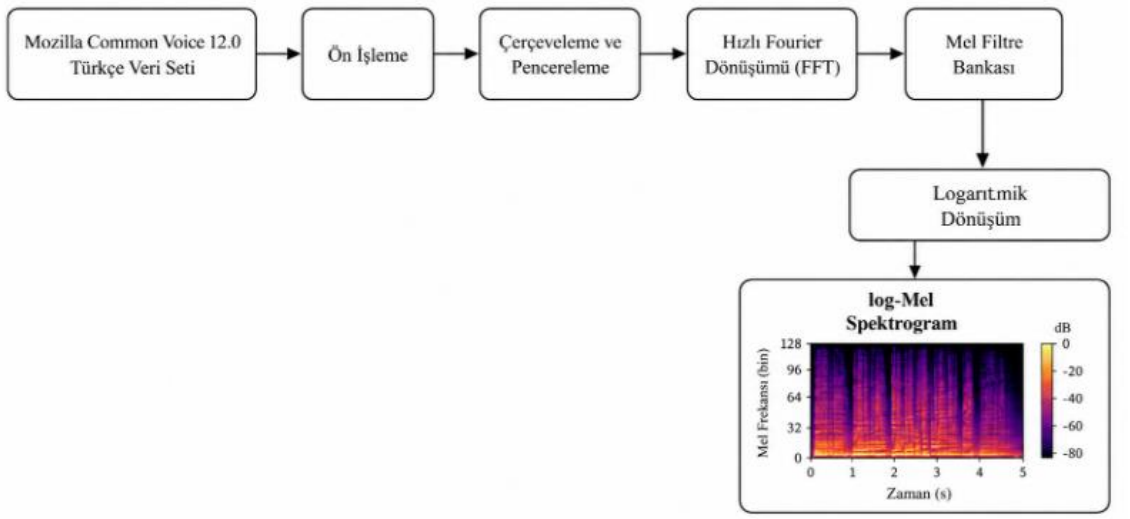
3.2.4 Logaritmik Mel Spektrogram (log-Mel-Spectrogram)

Konuşma sinyalinin zaman–frekans alanındaki enerji dağılımını temsil edebilmek amacıyla log-Mel spektrogram özellikleri kullanılmaktadır. Log-Mel spektrogram, frekans ekseninin insan işitme sistemine daha uygun olan Mel ölçeğine dönüştürülmesi ile elde edilmektedir. Ardından elde edilen güç spektrumunun logaritması alınarak nihai temsil oluşturulmaktadır. Bu dönüşüm sayesinde insan kulağının ses şiddetini logaritmik olarak algılama özelliği modellenmekte ve düşük enerjili bileşenlerin daha belirgin hâle gelmesi sağlanmaktadır.

log-Mel spektrogramlar, konuşma sinyalinin hem zamansal hem de frekanssal değişimlerini etkili bir şekilde temsil edebilen güçlü özniteliklerdir. Bu özellikleri sayesinde özellikle konuşma sinyallerinde yer alan ince frekans detaylarının ortaya çıkarılmasında

önemli avantaj sağlamaktadır. Ayrıca ESA tabanlı derin öğrenme modelleri için uygun bir giriş verisi oluşturarak konuşma sinyaline ait karakteristik örüntülerin daha başarılı bir şekilde öğrenilmesine katkıda bulunmaktadır.

Konuşma sinyali analizinde ayrıca log-Mel spektrogramlar cinsiyet, duygu durumu, konuşma hızı ve tonlama gibi biyometrik ve konuşmacının ritimsel özelliklerinin çıkarılmasında yaygın olarak kullanılmaktadır. (Donuk ve Hanbay 2022). Bu nedenle bu çalışmada önerilen hibrit derin öğrenme modelinin performansını artırmak amacıyla temel özniteliklerden biri olarak kullanılmıştır. Log-Mel spektrogram özelliklerinin elde edilme süreci Şekil 3.7'de gösterilmektedir.



Şekil 3.6 log-Mel spektrogram özelliklerinin çıkarılma süreci

3.2.5 Öznitelik Çıkarımı

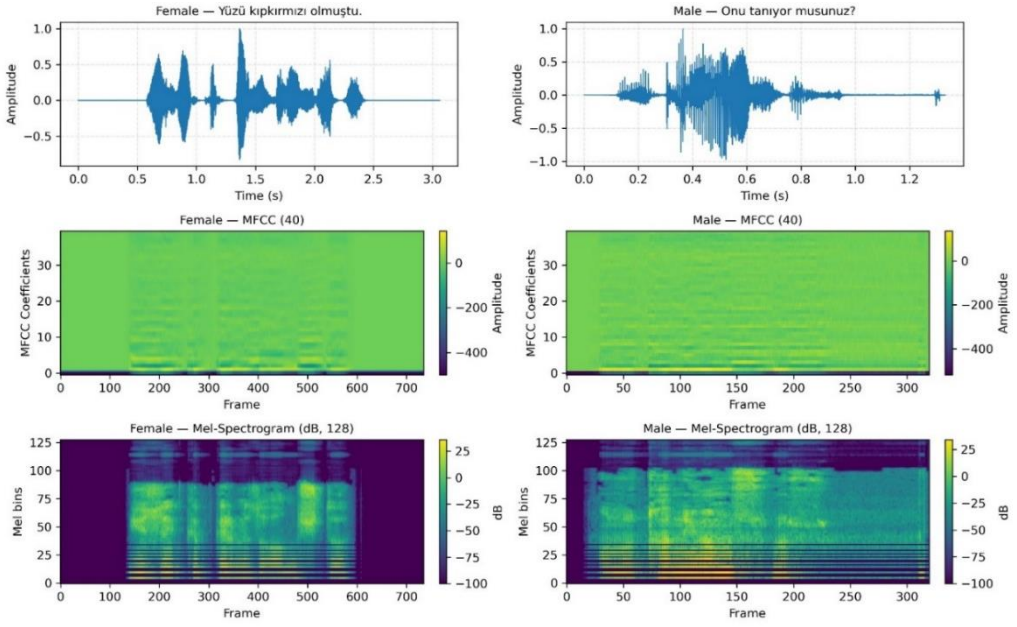
Konuşma sinyallerinin doğrudan sınıflandırma algoritmalarına uygulanması yüksek boyutluluk ve gürültüye duyarlılık nedeniyle modelin genelleme başarısını olumsuz etkileyebilmektedir. Bu nedenle ham ses kayıtlarından ayırt edici akustik bilgilerin elde edilmesi amacıyla MFCC ve log-Mel spektrogram öznitelikleri çıkarılmıştır. MFCC katsayıları sinyalin spektral yapısını özetlerken log-Mel spektrogram, enerjinin zaman ve frekans düzlemindeki dağılımını temsil etmektedir. Böylece derin öğrenme modellerinin konuşma verilerindeki ayırt edici örüntüleri daha etkili biçimde öğrenmesi amaçlanmıştır.

Bu çalışmada her bir ses kaydından elde edilen (40,1) boyutlu MFCC ve (128,1) boyutlu log-Mel öznitelikleri birleştirilerek (168,1) boyutlu bir özellik vektörü oluşturulmuştur. Bu bütünleştirme sayesinde farklı akustik temsillerden elde edilen tamamlayıcı bilgilerin aynı model içerisinde değerlendirilmesi sağlanmıştır. Kullanılan özniteliklerin boyutları ve açıklamaları Tablo 3.3'te sunulmaktadır.

Tablo 3.3 Kullanılan akustik öznitelikler

Öznitelik	**Boyut**	**Açıklama**
MFCC	40	Konuşma sinyalinin spektral özelliklerini temsil eder.
log-Mel spektrogram	128	Konuşma sinyalinin zaman-frekans enerji dağılımını gösterir

Kullanılan özniteliklerin konuşma sinyali üzerindeki etkisini daha iyi gösterebilmek amacıyla kadın ve erkek konuşmacılara ait örnek sinyallerin dalga formu, MFCC ve log-Mel spektrogram temsilleri Şekil 3.8'de sunulmuştur.



(a) (b)

Şekil 3.7 Kadın (a) ve erkek(b) konuşmacılara ait örnek ses sinyallerinin dalga formu, MFCC ve log-Mel spektrogram öznitelikleri

Şekil 3.8 incelendiğinde, dalga formu üzerinden konuşma sinyalinin zaman içerisindeki genlik değişimleri gözlemlenebilmektedir. Kadın ve erkek konuşmacılara ait sinyaller arasında genlik yapısı ve süre açısından belirgin farklılıklar bulunmaktadır. Erkek konuşmacıya ait sinyalin daha yoğun ve düşük frekans bileşenleri içerdiği görülürken kadın konuşmacıya ait sinyalin daha yüksek frekanslı bileşenler barındırdığı dikkat çekmektedir.

MFCC öznitelikleri incelendiğinde, konuşma sinyalinin spektral özelliklerini temsil eden katsayıların zaman içerisindeki değişimi görülmektedir. Kadın ve erkek konuşmacılara ait MFCC dağılımlarının farklı örüntüler oluşturduğu ve bu örüntülerin sınıflandırma sürecinde ayırt edici bilgiler sunduğu anlaşılmaktadır.

log-Mel spektrogram öznitelikleri değerlendirildiğinde ise konuşma sinyalinin zaman–frekans düzlemindeki enerji dağılımı açıkça görülmektedir. Erkek konuşmacıya ait spektrogramda enerjinin daha çok düşük frekans bantlarında yoğunlaştığı, kadın konuşmacıya ait spektrogramda ise enerjinin daha geniş bir frekans aralığına yayıldığı gözlemlenmektedir. Bu durum konuşma üretim sürecindeki fizyolojik farklılıkların akustik temsillere doğrudan yansıdığını ortaya koymaktadır.

3.2.6 Derin Öğrenme Yöntemleri

Derin öğrenme çok katmanlı yapay sinir ağları aracılığıyla verilerden anlamlı özelliklerin otomatik olarak öğrenilmesini sağlayan bir makine öğrenmesi yaklaşımıdır (Purwins vd. 2019). Geleneksel yöntemlerde öznitelik çıkarımı çoğunlukla insan müdahalesine dayanmaktadır. Derin öğrenme yaklaşımlarında ise bu süreç otomatikleştirilmekte ve modelin veri içerisindeki ilişkileri doğrudan öğrenmesine olanak tanımaktadır.

Derin öğrenme yöntemleri özellikle görüntü işleme, doğal dil işleme ve konuşma işleme gibi alanlarda önemli başarılar elde etmiştir. Artan hesaplama kapasitesi ve büyük ölçekli veri kümelerinin yaygınlaşması ile bu yöntemler birçok araştırma alanında temel yaklaşımlardan biri haline gelmiştir (Duru vd. 2021, Yilmaz ve Poli 2022).

Konuşma sinyalleri hem zaman hem de frekans boyutunda değişkenlik gösteren dinamik yapılardır. Bu nedenle bu tür verilerin analizinde yalnızca spektral özelliklerin değil aynı zamanda zamansal bağımlılıkların da dikkate alınması gerekmektedir. Derin öğrenme tabanlı yöntemler bu iki bileşeni birlikte modelleyebilme yetenekleri sayesinde konuşma işleme problemlerinde yaygın olarak kullanılmaktadır (Mohammed ve Al-Irhayim 2021b).

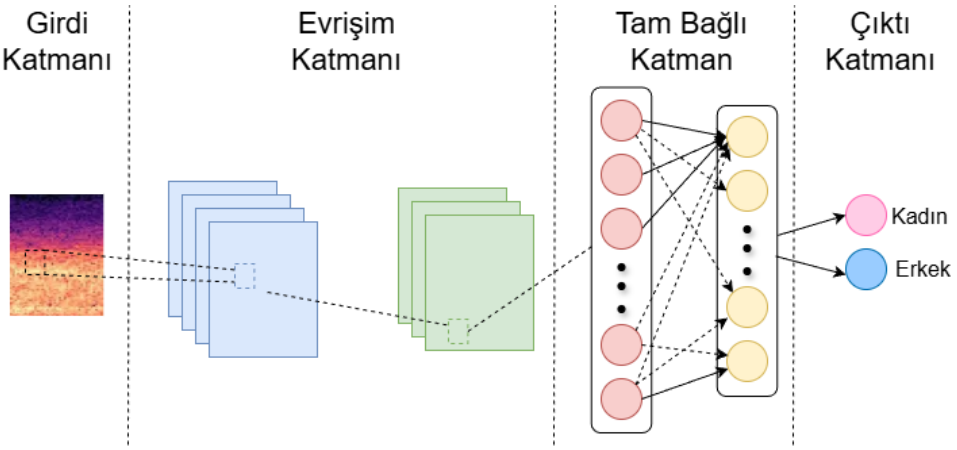
Bu çalışmada konuşma sinyallerinden elde edilen MFCC ve log-Mel spektrogram özniteliklerinin işlenmesi amacıyla farklı derin öğrenme yaklaşımları ele alınmıştır. Bu kapsamda ESA, TSA ve UKSB yapıları incelenmiş ve zamansal modelleme açısından karşılaştırılmıştır.

3.2.6.1 Evrişimsel Sinir Ağları (ESA)

ESA, çok katmanlı derin öğrenme mimarileri içerisinde özellikle görüntü, ses ve zaman-frekans temsiline sahip verilerde yüksek başarı gösteren yapay sinir ağı modelleridir. ESA mimarileri, biyolojik görme korteksinden esinlenerek geliştirilmiş olup giriş verisi üzerinde yerel bağlantılar ve ağırlık paylaşımı prensibi ile çalışmaktadır (LeCun vd. 2015a). Geleneksel tam bağlantılı sinir ağlarından farklı olarak ESA modellerinde her nöron giriş verisinin yalnızca belirli bir bölgesine bağlanmaktadır. Bu yapı sayesinde model, verideki yerel örüntüleri daha etkili bir şekilde öğrenebilmektedir. Ayrıca parametre sayısını önemli ölçüde azaltarak daha verimli bir öğrenme süreci sunmaktadır.

Konuşma sinyalleri zaman ve frekans eksenlerinde anlamlı örüntüler içeren karmaşık yapılardır. Özellikle MFCC ve log-Mel spektrogram gibi zaman-frekans temsilleri, görüntü benzeri yapılar oluşturduğu için ESA modelleri tarafından etkili biçimde işlenebilmektedir. ESA katmanları konuşma sinyalindeki kısa süreli frekans değişimlerini, harmonik yapıları ve spektral geçişleri otomatik olarak öğrenerek sınıflandırma sürecinde ayırt edici öznitelikler elde edilmesini sağlamaktadır.

ESA mimarisi genel olarak evrişim katmanı, aktivasyon katmanı, havuzlama katmanı ve tam bağlantılı katmanlardan oluşmaktadır. Bu mimarinin genel yapısı Şekil 3.9'da gösterilmektedir.



Şekil 3.8 Evrişimsel Sinir Ağı Mimarisi

Aşağıda ESA mimarisinde kullanılan evrişim yapıları başlıklar halinde açıklanmıştır.

Evrişim Katmanı (Convolution Layer): Evrişim katmanı ESA mimarisinin temel bileşenidir. Bu katmanda giriş verisi üzerinde belirli boyutlardaki filtreler (kernel) kaydırılarak yerel özellikler çıkarılmaktadır. Her filtre giriş verisinin farklı örüntülerini öğrenmektedir. Örneğin konuşma sinyallerinde bazı filtreler düşük frekans bileşenlerini öğrenirken bazıları yüksek frekans geçişlerini veya belirgin frekans yapılarını öğrenebilmektedir. Evrişim işlemi matematiksel olarak Denklem 3.1'de ifade edilmektedir.

$$Y(i,j) = \sum_{m}\sum_{n} X(i-m, j-n).K(m,n) \tag{3.1}$$

Denklem 3.1'de X giriş verisini, K filtre matrisini ve Y ise evrişim işlemi sonucunda elde edilen özellik haritasını temsil etmektedir. Denklem 3.1'de verilen işlem evrişimsel sinir ağlarında kullanılan iki boyutlu evrişim (convolution) işlemini ifade etmektedir. Bu denklemde i ve j giriş verisi üzerindeki satır ve sütun konumlarını gösterirken m ve n filtre matrisi içerisindeki indisleri temsil etmektedir. Evrişim işlemi sırasında filtre matrisi giriş verisi üzerinde belirli adımlarla kaydırılmakta ve her konumda karşılık gelen elemanlar

arasında çarpma ve toplama işlemleri gerçekleştirilmektedir. Bu işlem sonucunda giriş verisi içerisindeki kenar, şekil ve yerel örüntü gibi anlamlı özellikler ortaya çıkarılmakta ve modelin ayırt edici öznitelikleri öğrenmesi sağlanmaktadır.

Konuşma sinyallerinde kullanılan Conv1D yapıları özellikle zaman ekseni boyunca ilerleyen akustik örüntülerin öğrenilmesinde etkili olmaktadır. Bu nedenle çalışmada kullanılan hibrit modelde Conv1D katmanları tercih edilmiştir. Conv1D yapısı, konuşma sinyalinin zamansal akışını koruyarak kısa süreli bağımlılıkların öğrenilmesini sağlamaktadır.

Aktivasyon Katmanı: Evrişim işlemi doğrusal bir işlem olduğundan modelin karmaşık örüntüleri öğrenebilmesi için doğrusal olmayan aktivasyon fonksiyonlarına ihtiyaç duyulmaktadır. ESA modellerinde en yaygın kullanılan aktivasyon fonksiyonlarından biri ReLU (Rectified Linear Unit) fonksiyonudur. ReLU fonksiyonu Denklem 3.2'de gösterilmektedir.

$$f(x) = \max(0, x) \tag{3.2}$$

Denklem 3.2'de yer alan $f(x)$ ifadesi aktivasyon fonksiyonunun ürettiği çıktıyı, x ise katmana giren giriş değerini temsil etmektedir. ReLU aktivasyon fonksiyonu, giriş değerini sıfır ile karşılaştırarak büyük olan değeri çıktı olarak vermektedir. Bu nedenle fonksiyon matematiksel olarak denklem 3.2' deki gibi ifade edilir. Bu sayede negatif değerler sıfıra indirgenirken pozitif girişler korunmakta böylece modelin öğrenme süreci daha kararlı ve verimli hâle gelmektedir. (Nair ve Hinton 2010). Ayrıca ReLU aktivasyon fonksiyonu hesaplama açısından düşük maliyetli olduğu için derin öğrenme modellerinde yaygın olarak tercih edilmektedir.

Havuzlama Katmanı (Pooling Layer): Havuzlama katmanı, evrişim katmanlarından elde edilen özellik haritalarının boyutunu küçültmek amacıyla kullanılmaktadır. Bu işlem modelin hesaplama maliyetini azaltmakta ve aşırı öğrenme riskini düşürmektedir. Aynı zamanda önemli özelliklerin korunmasını sağlayarak modelin genelleme yeteneğini artırmaktadır (Goodfellow 2016).

En yaygın kullanılan havuzlama yöntemi maksimum havuzlama yöntemidir. Bu yöntemde belirli bir bölgedeki en yüksek değer seçilerek yeni özellik haritası oluşturulmaktadır. Böylece ilgili bölgedeki en güçlü özellik korunmaktadır. Konuşma sinyallerinde maksimum havuzlama (max pooling) işlemi, önemli frekans örüntülerinin korunmasına yardımcı olmakta ve gürültü etkisinin azaltılmasına katkı sağlamaktadır.

Yığın Normalizasyonu ve Seyreltme: ESA mimarisinde modelin daha kararlı öğrenebilmesi amacıyla yığın normalizasyonu (batch normalization) katmanları kullanılmaktadır. Yığın Normalizasyonu işlemi, katman girişlerini normalize ederek öğrenme sürecini hızlandırmakta ve modelin daha stabil çalışmasına katkı sağlamaktadır.

Seyreltme (Dropout) yöntemi ise aşırı öğrenmeyi önlemek amacıyla kullanılan bir düzenlileştirme tekniğidir. Eğitim sırasında belirli nöronlar rastgele devre dışı bırakılarak modelin belirli bağlantılara aşırı bağımlı hale gelmesi engellenmektedir. Bu yöntem modelin farklı örüntüleri öğrenmesini teşvik ederek genelleme performansını artırmaktadır.

ESA mimarisini oluşturan evrişim, aktivasyon, havuzlama ve tam bağlantılı katmanlar birlikte değerlendirildiğinde modelin güçlü bir özellik çıkarım yeteneğine sahip olduğu görülmektedir. ESA modelleri, konuşma sinyali içerisindeki yerel örüntüleri otomatik olarak öğrenebilmekte ve manuel öznitelik çıkarımına duyulan ihtiyacı azaltmaktadır. Ayrıca ağırlık paylaşımı mekanizması sayesinde daha düşük hesaplama maliyeti ile çalışabilmektedir. Gürültülü veriler üzerinde daha dayanıklı bir yapı sunmasının yanı sıra MFCC ve log-Mel spektrogram gibi zaman-frekans temsillerini etkili biçimde işleyebilmektedir. Bununla birlikte büyük veri kümelerinde yüksek sınıflandırma performansı sağlayarak konuşma işleme problemlerinde başarılı sonuçlar elde edilmesine katkı sağlamaktadır (Yu ve Deng 2011, Abdel-Hamid vd. 2014) .

3.2.6.2 Tekrarlayan Sinir Ağları (TSA)

TSA, zamansal bağımlılık içeren ardışık verilerin modellenmesi amacıyla geliştirilmiş derin öğrenme mimarileridir (Goodfellow 2016). Bu ağ yapısı, önceki zaman adımlarına

ait bilgileri saklayarak mevcut girdinin yorumlanmasına olanak sağlamaktadır. Bu özelliği sayesinde TSA'lar, konuşma sinyali gibi zaman içerisinde değişen ve birbirine bağlı verilerin analizinde etkin bir şekilde kullanılmaktadır.

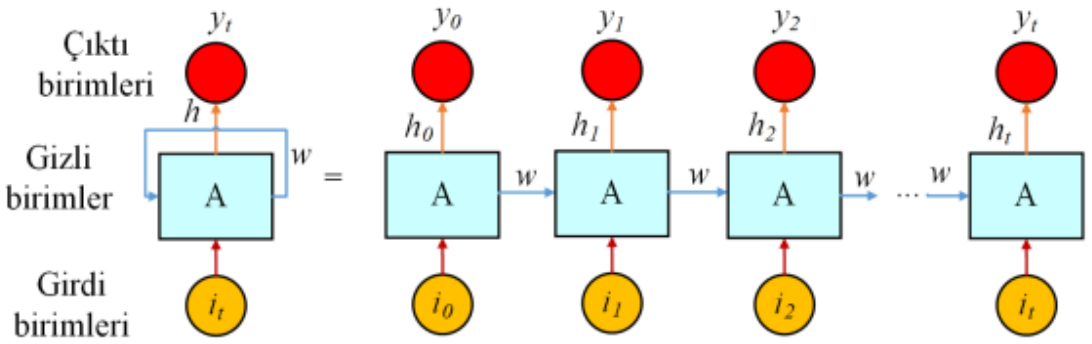
TSA mimarisinin temelinde, her zaman adımında güncellenen bir gizli durum (hidden state) yer almaktadır. Bu gizli durum, önceki zaman adımından gelen bilgiyi mevcut giriş ile birleştirerek yeni bir temsil oluşturmaktadır. Her zaman adımında gizli durum hem mevcut giriş hem de bir önceki gizli durum kullanılarak güncellenmektedir. Bu ilişki matematiksel olarak Denklem 3.3'te ifade edilmektedir (Goodfellow 2016).

$$h_t = f\left(W_{\{xh\}x_t} + W_{\{hh\}h_{\{t-1\}}} + b\right) \tag{3.3}$$

Denklem 3.3'te x_t mevcut giriş verisini, h_{t-1} bir önceki gizli durumu, W_{xh} ve W_{hh} ağırlık matrislerini, b bias terimini ve f aktivasyon fonksiyonunu temsil etmektedir. Bu yapı sayesinde model, zaman boyunca bilgi birikimi oluşturarak ardışık verilerdeki ilişkileri öğrenebilmektedir. Bu sayede model yalnızca anlık veriye değil aynı zamanda geçmiş verilere de bağlı olarak çıktı üretebilmektedir. Bu durum konuşma sinyallerinde zamansal örüntülerin ve ardışık ilişkilerin öğrenilmesini mümkün kılmaktadır.

TSA yapısında, ardışık veri dizisinin tüm zaman adımlarında aynı ağırlıkların (parametrelerin) kullanılması, modelin parametre sayısını azaltmakta ve daha verimli bir öğrenme süreci sağlamaktadır. Bu özellik klasik ileri beslemeli sinir ağlarından farklı olarak TSA'ların değişken uzunluktaki veri dizileri üzerinde etkili bir şekilde çalışmasına imkân tanımaktadır.

Şekil 3.10'da TSA mimarisinin genel yapısı gösterilmektedir. Bu yapıda giriş verileri ($i_0, i_1, i_2 \ldots i_t$) sırasıyla işlenmekte ve her adımda elde edilen gizli durumlar ($h_0, h_1, h_2 \ldots h_t$) bir sonraki adıma aktarılmaktadır. Aynı zamanda her zaman adımında bir çıktı ($y_0, y_1, y_2 \ldots y_t$) üretilmektedir. Bu yapı, modelin geçmiş bilgileri kullanarak zamansal bağımlılıkları öğrenmesini sağlamaktadır.



Şekil 3.9 Tekrarlayan Sinir Ağları Mimarisi (Çelebi ve Fidan 2024)

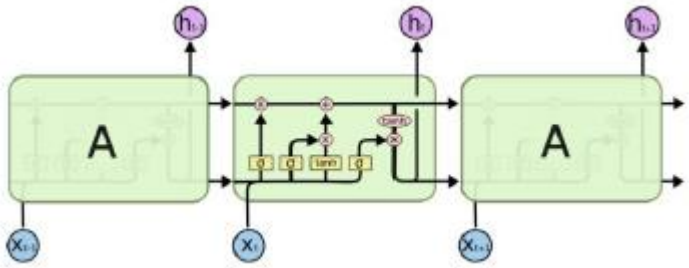
TSA yapıları dil modelleme, zaman serisi analizi, konuşma işleme ve makine çevirisi gibi ardışık bağımlılıkların önemli olduğu birçok alanda yaygın olarak kullanılmaktadır. Bu çalışmada ise TSA yapıları, ESA katmanları ile elde edilen MFCC ve log-Mel spektrogram tabanlı özniteliklerin zamansal bağımlılıklarını modellemek amacıyla kullanılmıştır. Böylece konuşma sinyalinin hem frekans hem de zaman boyutundaki özelliklerinin birlikte değerlendirilmesi hedeflenmiştir.

Her ne kadar TSA yapıları başarılı sonuçlar üretse de klasik TSA yapıları uzun veri dizilerinde gradyan sönmesi (vanishing gradient) ve gradyan patlaması (exploding gradient) gibi problemlerle karşılaşabilmektedir. Bu nedenle pratik uygulamalarda genellikle UKSB ve Geçitli Tekrarlayan Birim (GRU) gibi geliştirilmiş TSA türevleri tercih edilmektedir. Bu yapılar, uzun vadeli bağımlılıkların daha etkin bir şekilde öğrenilmesine olanak sağlamaktadır (Graves vd. 2013).

3.2.6.3 Uzun Kısa Süreli Bellek (UKSB)

Tekrarlayan sinir ağlarının gelişmiş bir türü olan UKSB, klasik TSA yapılarında karşılaşılan uzun vadeli bağımlılıkları öğrenememe ve gradyan sönmesi gibi problemlerin giderilmesi amacıyla geliştirilmiştir. UKSB mimarisi, bilgi akışını kontrol eden kapı mekanizmaları sayesinde geçmiş bilgilerin korunmasını ve gerekli durumlarda güncellenmesini sağlamaktadır (Sakarya ve Yılmaz 2019).

Şekil 3.11'de UKSB modeline ait blok diyagramı gösterilmektedir. Şekilde görüldüğü üzere her bir UKSB hücresi, önceki gizli durum (h_{t-1}) ve mevcut giriş (x_t) bilgilerini kullanarak yeni bir gizli durum (h_t) üretmektedir. Bu süreçte hücre içerisinde yer alan kapı mekanizmaları bilgi akışını düzenlemektedir.



Şekil 3.10 UKSB modeli blok diyagramı (Ayyüce Kızrak ve Bolat 2018)

UKSB yapısı giriş kapısı, unutma kapısı ve çıkış kapısı olmak üzere üç temel bileşenden oluşmaktadır. Unutma kapısı, önceki hücre durumunda bulunan bilgilerin ne kadarının korunacağını belirleyerek gereksiz bilgilerin sistemden çıkarılmasını sağlamaktadır. Giriş kapısı, mevcut giriş verisi ve önceki gizli durumdan elde edilen bilgileri kullanarak hücreye eklenecek yeni bilgileri kontrol etmektedir. Bu aşamada aday hücre durumu oluşturularak mevcut bellek güncellenmektedir.

Şekil 3.11'de gösterilen işlem akışına göre unutma kapısından geçen bilgi ile giriş kapısından gelen yeni bilgiler birleştirilerek güncellenmiş hücre durumu elde edilmektedir. Bu yapı bilginin zaman boyunca bozulmadan taşınmasına olanak sağlamaktadır. Son olarak çıkış kapısı, hücre durumunda bulunan bilginin ne kadarının dışarı aktarılacağını belirleyerek yeni gizli durumu (h_t) oluşturmaktadır. Bu mekanizma sayesinde UKSB modeli, geçmiş bilgileri uzun süre boyunca koruyabilmekte ve zamansal bağımlılıkları daha etkin bir şekilde öğrenebilmektedir. Bu özellik, özellikle konuşma sinyali gibi zaman serisi verilerinin analizinde önemli avantajlar sağlamaktadır. Konuşma sinyallerinin zamana bağlı yapısı dikkate alındığında, UKSB mimarisi bu tür verilerdeki zamansal ilişkilerin modellenmesi için etkili bir yaklaşım sunmaktadır.

Bu çalışmada UKSB yapısı, konuşma sinyallerinden elde edilen MFCC ve log-Mel spektrogram öznitelikleri arasındaki zamansal bağımlılıkları modellemek amacıyla kullanıl-

mıştır. ESA katmanları tarafından çıkarılan yerel ve ayırt edici özellikler ardışık veri yapısını koruyacak şekilde UKSB katmanına aktarılmıştır. UKSB katmanı, giriş dizileri içerisindeki kısa ve uzun dönemli ilişkileri öğrenerek konuşma sinyalinin zamansal dinamiklerini temsil etmektedir. Böylece konuşma sırasında zaman boyunca ortaya çıkan akustik değişimlerin modellenmesi sağlanmıştır. Çalışmada UKSB mimarisi, konuşma sinyalindeki zamansal bağımlılıkların öğrenilmesinde referans model olarak kullanılmış ve elde edilen sonuçlar önerilen NCP tabanlı mimarinin performansı ile karşılaştırılmıştır.

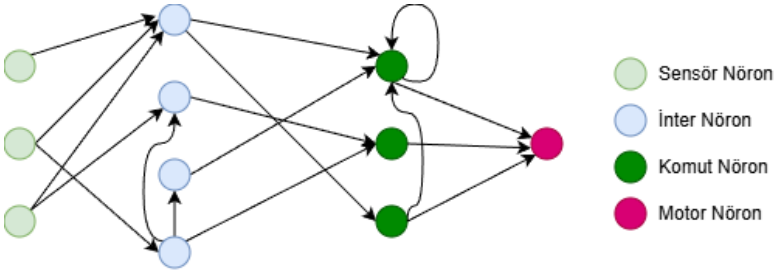
3.2.6.4 Nöral Devre Politikaları (NCP)

NCP, biyolojik sinir sistemlerinin yapısal organizasyonundan esinlenerek geliştirilen ve özellikle düşük parametreli, yüksek verimli öğrenme süreçlerini hedefleyen bir yapay sinir ağı yaklaşımıdır. Mimari, *Caenorhabditis elegans adlı* nematodun sinir sistemindeki bağlantı düzeninden ilham almakta ve sınırlı sayıdaki nöron ile sinaptik bağlantının görev odaklı biçimde düzenlenmesine dayanmaktadır(Lechner vd. 2020).

Klasik derin öğrenme modellerinde kullanılan yoğun ve tam bağlantılı yapılar, yüksek temsil kapasitesi sağlamakla birlikte parametre sayısını ve hesaplama gereksinimini artırabilmektedir. NCP mimarisinde ise nöronlar arasındaki bağlantılar daha seyrek ve düzenli bir yapı içerisinde oluşturulmaktadır. Bu özellik, model karmaşıklığının sınırlandırılmasına ve hesaplama kaynaklarının daha verimli kullanılmasına katkı sağlamaktadır (Olgun 2022). Düşük parametreli yapısı nedeniyle NCP modelleri, sınırlı hesaplama kaynaklarına sahip sistemler açısından uygun bir yaklaşım olarak değerlendirilebilmektedir.

NCP mimarisi temel olarak sensör nöronları, inter-nöronlar, komut nöronları ve motor nöronlar olmak üzere dört ana bileşenden oluşmaktadır. Sensör nöronları, dış ortamdan gelen verileri alarak sisteme giriş sağlamaktadır. Bu giriş verileri ara katmanda bulunan inter-nöronlar aracılığıyla işlenerek daha anlamlı temsillere dönüştürülmektedir. Elde edilen bilgiler komut nöronları tarafından değerlendirilmekte ve karar mekanizmasına aktarılmaktadır. Son aşamada ise motor nöronlar sistemin nihai çıktısını üretmektedir. Bu biyolojik esinli yapı, klasik yapay sinir ağlarından farklı olarak daha modüler ve açıklanabilir bir mimari sunmaktadır. Bu çalışmada motor nöronlar tarafından üretilen çıktı,

konuşma sinyaline dayalı olarak kadın veya erkek sınıflandırması şeklinde elde edilmektedir. NCP mimarisinin genel yapısı ve bilgi akışı Şekil 3.12'de gösterilmektedir.



Şekil 3.11 NCP mimarisinin temel aşamaları

ESA katmanlarından elde edilen öznitelikler, zamansal bağımlılıkların modellenmesi amacıyla LTC hücrelerine aktarılmıştır. LTC hücreleri, sürekli zaman dinamiklerini modelleyebilen ve diferansiyel denklem tabanlı bir yapı sunan özel bir tekrarlayan sinir ağı türüdür (Lechner vd. 2020). Bu yaklaşım klasik ayrık zamanlı modellerden farklı olarak zamanın sürekli bir değişken olarak ele alınmasını sağlamaktadır. Böylece model zaman içerisindeki değişimleri daha hassas bir şekilde takip edebilmekte ve özellikle konuşma sinyali gibi dinamik verilerde daha gerçekçi temsiller oluşturabilmektedir.

LTC hücrelerinin önemli avantajlarından biri parametre sayısının düşük olmasına rağmen güçlü zamansal modelleme yeteneğine sahip olmasıdır. Bu durum modelin hem hesaplama verimliliğini artırmakta hem de daha iyi genelleme performansı sunmasına katkı sağlamaktadır (Lechner vd. 2020). Ayrıca LTC yapısı, uyarlanabilir zaman sabitleri sayesinde farklı zaman ölçeklerindeki bağımlılıkları öğrenebilme kapasitesine sahiptir. Bu özellik konuşma sinyalindeki kısa ve uzun vadeli değişimlerin birlikte modellenmesine olanak tanımaktadır.

Bu çalışmada LTC yapısı, önerilen NCP tabanlı mimarinin zamansal işleme bileşeni olarak kullanılmıştır. ESA katmanları tarafından çıkarılan yerel ve ayırt edici özellikler, ardışık veri yapısını koruyacak şekilde LTC katmanına aktarılmıştır. LTC yapısı, sürekli zamanlı dinamikleri modelleyebilme yeteneği sayesinde konuşma sinyalindeki zamansal değişimleri etkili bir şekilde öğrenmektedir. Geleneksel TSA yapılarından farklı olarak

LTC, her nöronun zaman sabitini giriş verisine bağlı olarak dinamik biçimde güncelleyebilmektedir. Bu sayede model, konuşma sinyalinde yer alan kısa ve uzun dönemli bağımlılıkları daha esnek bir şekilde temsil edebilmektedir.

LTC katmanı tarafından oluşturulan zamansal öznitelik temsilleri, NCP mimarisinin seyrek ve biyolojik olarak esinlenilmiş bağlantı yapısı içerisinde işlenmektedir. Bu yapı, ağ içerisindeki bilgi akışını düzenleyerek önemli özelliklerin daha etkin bir şekilde öğrenilmesine katkı sağlamaktadır. Böylece konuşma sırasında zaman boyunca meydana gelen frekans, enerji ve tonlama değişimleri daha ayrıntılı biçimde modellenebilmektedir. NCP mimarisi içerisinde işlenen öznitelik temsilleri, tam bağlantılı katmanlar aracılığıyla sınıflandırılarak konuşmadan cinsiyet tanıma işlemi gerçekleştirilmektedir. Bu yaklaşım sayesinde konuşma sinyalindeki zamansal dinamiklerin etkin öğrenilmesi ve sınıflandırma performansının artırılması hedeflenmektedir.

3.2.6.5 Tam Bağlantılı Katmanlar (Full Connected Layers)

ESA mimarisinin son aşamasında tam bağlantılı katmanlar bulunmaktadır. Bu katmanlar, önceki katmanlardan elde edilen özellikleri kullanarak sınıflandırma işlemini gerçekleştirmektedir. Tam bağlantılı katmanlarda her nöron bir sonraki katmandaki tüm nöronlarla bağlantılıdır (LeCun vd. 2015b).

Bu çalışmada ESA katmanları tarafından çıkarılan spektral öznitelikler, zamansal bağımlılıkların modellenmesi amacıyla UKSB ve NCP tabanlı yapılara aktarılmıştır. UKSB yapısı, konuşma sinyalindeki uzun ve kısa dönemli bağımlılıkları kapı mekanizmaları aracılığıyla öğrenirken NCP mimarisi biyolojik olarak esinlenilmiş seyrek bağlantı yapısı sayesinde zamansal dinamikleri esnek bir şekilde modellemektedir. Bu yapılardan elde edilen öznitelik temsilleri tam bağlantılı katmanlara iletilerek sınıflandırma işlemi gerçekleştirilmektedir. Böylece konuşma sinyalinin hem spektral hem de zamansal özellikleri birlikte değerlendirilerek cinsiyet tanıma işlemi yapılmaktadır.

3.2.6.6 Sigmoid Aktivasyon Fonksiyonu

Bu çalışmada tam bağlantılı katmanlardan sonra sigmoid aktivasyon fonksiyonuna sahip bir çıkış katmanı kullanılmıştır. Sigmoid fonksiyonu, model çıktısını 0 ile 1 arasında bir aralığa sıkıştırarak olasılık değeri üretmektedir. Bu sayede her ses örneği için "kadın" veya "erkek" sınıfına ait olma olasılığı hesaplanmakta ve nihai sınıf kararı bu olasılık değeri üzerinden verilmektedir. Sigmoid fonksiyonu özellikle ikili sınıflandırma problemlerinde olasılık temelli ve yorumlanabilir çıktı üretmesi nedeniyle yaygın olarak tercih edilmektedir. Sigmoid aktivasyon fonksiyonu matematiksel olarak Denklem 3.4'te ifade edilmektedir.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$

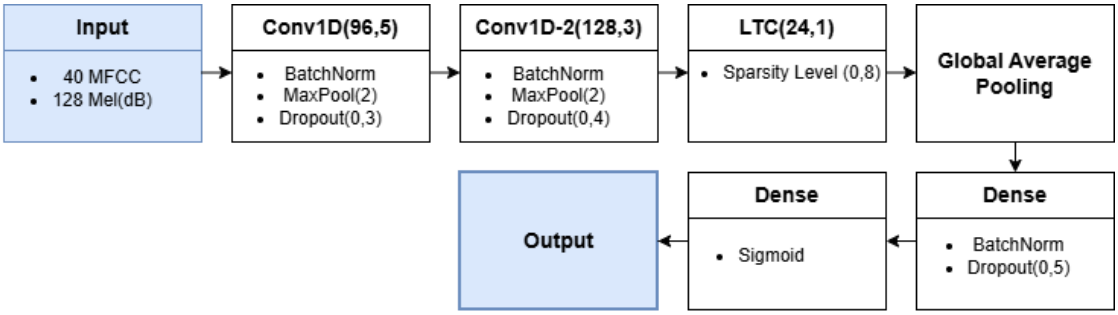
Denklem 3.4'te x, sinir ağına gelen giriş değerini; $\sigma(x)$ ise 0 ile 1 arasında normalize edilmiş çıktı değerini temsil etmektedir. Giriş değeri arttıkça fonksiyon çıktısı 1'e yaklaşırken, azaldıkça 0'a yaklaşmaktadır. Bu özellik, sigmoid fonksiyonunu özellikle olasılık tabanlı karar verme gerektiren ikili sınıflandırma problemleri için uygun hale getirmektedir.

3.3 Eğitim Süreci

Modelin giriş bölümünde, özniteliklerdeki kısa süreli yerel örüntülerin öğrenilmesi amacıyla ardışık bir boyutlu ESA katmanları kullanılmıştır. ESA katmanlarıyla birlikte yığın normalizasyonu, havuzlama ve seyreltme işlemlerine yer verilerek modelin eğitim sürecinin daha kararlı ve genellenebilir olması amaçlanmıştır.

ESA katmanlarından elde edilen öznitelik haritaları farklı model yapılarının sınıflandırma performanslarını karşılaştırmak amacıyla TSA, Tam Bağlı, NCP ve UKSB tabanlı katmanlara aktarılmıştır. Bu yapılardan elde edilen çıktılar Global Average Pooling işlemiyle tek boyutlu öznitelik vektörlerine dönüştürülmüştür. Ardından yığın normalizasyonu ve 0,5 oranında seyreltme uygulanmış, elde edilen öznitelikler tam bağlantılı kat-

manlar aracılığıyla sınıflandırma işlemine uygun yüksek seviyeli temsillere dönüştürülmüştür. Modelin son aşamasındaki sigmoid aktivasyon fonksiyonuna sahip çıkış katmanı her bir konuşma örneğinin kadın veya erkek sınıfına ait olma olasılığını üretmektedir. Önerilen NCP tabanlı hibrit modelin genel mimarisi Şekil 3.13'te gösterilmektedir.



Şekil 3.12 NCP tabanlı hibrit modelin genel yapısı

Model eğitimi öncesinde veri seti eğitim, doğrulama ve test kümelerine ayrılmıştır. Veri setinin %70'i eğitim, %10'u doğrulama ve %20'si test amacıyla kullanılmıştır. Eğitim kümesi model parametrelerinin öğrenilmesinde, doğrulama kümesi eğitim sürecinin izlenmesi ve uygun model ayarlarının belirlenmesinde, test kümesi ise modelin daha önce görmediği örnekler üzerindeki nihai performansının değerlendirilmesinde kullanılmıştır. Veri bölme işlemi sırasında kadın ve erkek sınıflarına ait örneklerin kümelerdeki dağılımlarının korunmasına dikkat edilmiştir.

Modelin farklı veri bölünmeleri karşısındaki performansını daha güvenilir biçimde değerlendirebilmek amacıyla eğitim verileri üzerinde Tabakalı 5-Katlı Çapraz Doğrulama yöntemi uygulanmıştır. Bu yöntemde eğitim verileri beş alt kümeye ayrılmış; her yinelemede alt kümelerden biri doğrulama ve test, kalanları ise eğitim amacıyla kullanılmıştır. Böylece model performansının yalnızca tek bir veri bölünmesine bağlı olarak değerlendirilmesinin önüne geçilmiştir. Ayrıca modelin konuşmacıdan bağımsız genelleme yeteneğini incelemek amacıyla LOSO çapraz doğrulama yöntemi kullanılmıştır.
Model eğitimi sırasında iki sınıflı sınıflandırma problemlerinde yaygın olarak kullanılan Binary Cross Entropy kayıp fonksiyonu tercih edilmiştir (Goodfellow 2016). Model parametrelerinin optimize edilmesi amacıyla Adam optimizasyon algoritması kullanılmıştır. Adam algoritması, öğrenme sürecinde her parametre için adaptif öğrenme oranı belirleyerek modelin daha hızlı ve kararlı bir şekilde yakınsamasını sağlayan gradyan tabanlı

bir optimizasyon yöntemidir (Chollet 2021).

Model toplam 20 epoch boyunca eğitilmiş ve her eğitim adımında veriler 32 örnekten oluşan mini veri kümeleri hâlinde modele verilmiştir. Eğitim sürecinde doğrulama veri kümesi üzerinde elde edilen doğruluk ve kayıp değerleri takip edilerek modelin öğrenme süreci analiz edilmiştir. Doğrulama performansının iyileşmemesi durumunda eğitimi sonlandırmak için erken durdurma yöntemi uygulanmıştır. Ayrıca doğrulama kaybındaki değişime bağlı olarak öğrenme oranını otomatik biçimde azaltan ReduceLROnPlateau stratejisi kullanılmıştır. Eğitim sonunda doğrulama kümesinde en iyi performansı sağlayan model ağırlıkları kaydedilmiş ve nihai değerlendirme test kümesi üzerinde gerçekleştirilmiştir. Eğitim sürecinde kullanılan temel model parametreleri Tablo 3.4'te sunulmaktadır.

Tablo 3.4 Model eğitim parametreleri

Parametre	**Değer**
Optimizasyon	Adam
Kayıp fonksiyonu	Binary Cross Entropy
Epoch sayısı	20
Batch size	32
Doğrulama yöntemi	Tabakalı 5-Katlı, LOSO

Modelin başarımı doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans ölçütleri kullanılarak analiz edilmiştir.

3.4 Performans Ölçütleri

Önerilen modelin performansını daha kapsamlı ve güvenilir bir şekilde değerlendirmek amacıyla farklı sınıflandırma performans ölçütleri kullanılmıştır. Bu ölçütler, modelin yalnızca genel doğruluk oranını değil aynı zamanda sınıflandırma sürecindeki farklı hata türlerini de analiz etmeye olanak sağlamaktadır. Böylece modelin veri üzerindeki davranışı daha ayrıntılı biçimde incelenebilmekte ve elde edilen sonuçların güvenilirliği daha sağlıklı bir şekilde değerlendirilebilmektedir.

Bu çalışmada dengeli bir veri seti kullanıldığından, modelin genel sınıflandırma başarısını gösteren doğruluk temel performans ölçütlerinden biri olarak ele alınmıştır. Doğruluk, tüm örnekler içinde doğru sınıflandırılan örneklerin oranını ifade etmektedir. Bununla birlikte model performansının yalnızca doğruluk değeriyle değerlendirilmesi, yapılan hata türleri hakkında yeterli bilgi sunmamaktadır. Bu nedenle doğruluk ölçütünün yanı sıra kesinlik, duyarlılık ve F1 skoru da kullanılmıştır. Erkek sınıfının pozitif sınıf olarak kabul edildiği değerlendirmede kesinlik, model tarafından erkek olarak tahmin edilen örneklerin ne kadarının gerçekte erkek olduğunu; duyarlılık ise gerçek erkek örneklerin ne kadarının doğru sınıflandırıldığını göstermektedir. F1 skoru, kesinlik ve duyarlılık arasındaki dengeyi tek bir değerle ifade ederek model performansının daha kapsamlı biçimde değerlendirilmesini sağlamaktadır.

Çalışmada kullanılan performans ölçütleri ve bunlara ait matematiksel ifadeler Tablo 3.5'te verilmiştir. Tabloda yer alan formüllerde TP (True Positive), erkek ses örneklerinin doğru şekilde erkek olarak sınıflandırıldığı durumları; TN (True Negative), kadın ses örneklerinin doğru şekilde kadın olarak sınıflandırıldığı durumları ifade etmektedir. FP (False Positive), gerçekte kadın olan bir örneğin yanlışlıkla erkek olarak sınıflandırılmasını; FN (False Negative) ise gerçekte erkek olan bir örneğin yanlışlıkla kadın olarak sınıflandırılmasını göstermektedir.

Tablo 3.5 Sınıflandırmada kullanılan performans metriklerinin hesaplaması

Metrik	**Formül**
Doğruluk (Accuracy)	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
Kesinlik (Precision)	$Precision = \frac{TP}{TP + FP}$
Duyarlılık (Recall)	$Recall = \frac{TP}{TP + FN}$
F1 Skoru (F1 Score)	$F1 = 2.\frac{Precision.Recall}{Precision + Recall}$

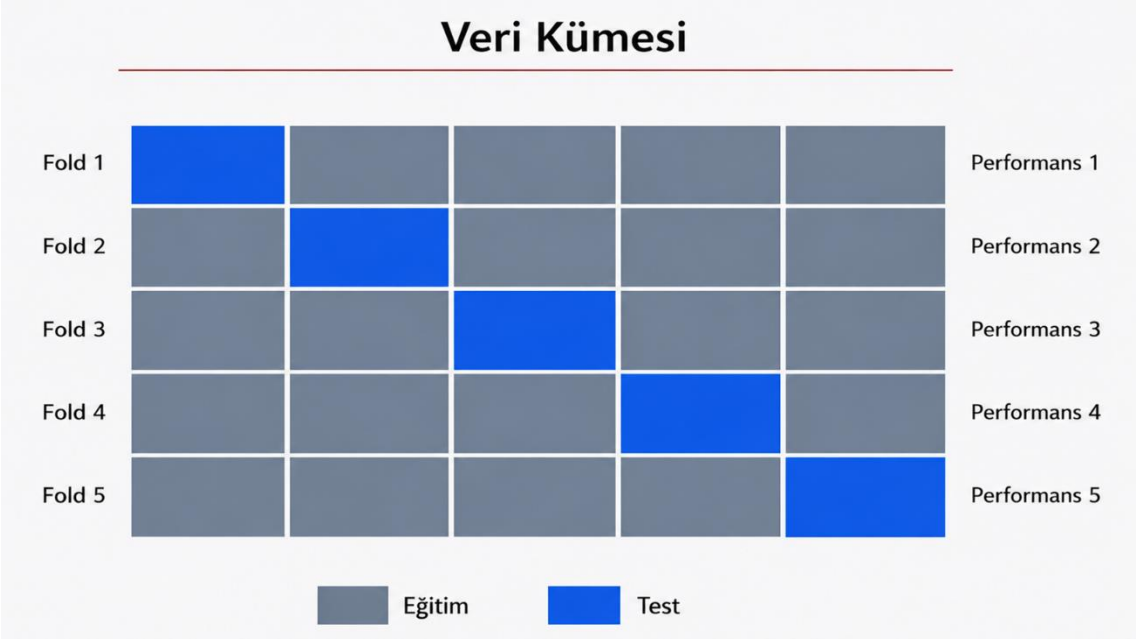
3.5 Tabakalı k-katlı (Stratified k-Fold) Çapraz Doğrulama ve Konuşmacı Bazlı Ayrık Çapraz Doğrulama (Leave-One-Speaker-Out, LOSO)

Bu çalışmada önerilen Conv1D+NCP tabanlı modelin genellenebilirliğini artırmak ve elde edilen performans sonuçlarını daha güvenilir biçimde değerlendirebilmek amacıyla tabakalı k-katlı çapraz doğrulama yöntemi kullanılmıştır. Tabakalı k-katlı çapraz doğrulama, makine öğrenmesi ve derin öğrenme modellerinin performansını daha nesnel bir şekilde değerlendirmek için yaygın olarak tercih edilen istatistiksel bir doğrulama tekniğidir (Kohavi 2001). Bu yöntem, modelin yalnızca belirli bir veri bölünmesine bağımlı kalmasını engelleyerek daha sağlam ve genellenebilir sonuçlar elde edilmesini sağlamaktadır.

Tabakalı k-katlı çapraz doğrulama yönteminde, veri seti rastgele ancak sınıf dağılımı korunacak şekilde k adet alt kümeye (fold) ayrılmaktadır. Her bir yinelemede bu alt kümelerden biri test verisi olarak kullanılırken, kalan $k-1$ alt küme eğitim verisi olarak kullanılmaktadır. Bu işlem k kez tekrarlanmakta ve her bir alt küme bir kez test verisi olarak değerlendirilmiş olmaktadır. Böylece model veri setinin tamamı üzerinde hem eğitim hem de test edilme fırsatı bulmaktadır. Ayrıca farklı veri bölünmelerine karşı gösterdiği performans kapsamlı bir şekilde analiz edilebilmektedir.

Her bir yinelemede modelin baştan eğitilmesi, modelin farklı veri dağılımlarına karşı nasıl tepki verdiğini ortaya koymaktadır. Bu durum aşırı öğrenme (overfitting) riskinin daha sağlıklı bir şekilde değerlendirilmesine olanak tanımaktadır. Aynı zamanda bu yöntem veri setinin belirli bir kısmına aşırı uyum sağlanması durumunda oluşabilecek yanlılıkları azaltarak daha dengeli performans ölçümleri elde edilmesini sağlamaktadır.

Tabakalı 5-katlı çapraz doğrulama yöntemi kullanılarak her bir katmanda sınıf dağılımının korunması sağlanmıştır. Böylece eğitim ve test kümeleri arasında dengesizlik oluşmasının önüne geçilmiştir. Bu durum özellikle sınıflandırma problemlerinde performans ölçümlerinin daha doğru ve tutarlı olmasına katkı sağlamaktadır. Şekil 3.14'te, uygulanan tabakalı 5-katlı çapraz doğrulama yöntemi şematik olarak gösterilmiştir.



Şekil 3.13 5 katmanlı çapraz doğrulama tekniği

Her bir katman için elde edilen performans değerleri ayrı ayrı hesaplanmış ve bu değerlerin ortalaması alınarak modelin genel performansı belirlenmiştir. Ortalama değerlerin yanı sıra katman bazlı performans sonuçlarının da incelenmesi modelin farklı veri bölünmelerindeki tutarlılığını değerlendirmek açısından önemli bir avantaj sağlamaktadır.

Bununla birlikte, konuşma verileri üzerinde gerçekleştirilen çalışmalarda aynı konuşmacıya ait birden fazla kaydın bulunması, tabakalı k-katlı çapraz doğrulama yönteminde bazı sınırlılıkları beraberinde getirebilmektedir. Aynı konuşmacıya ait kayıtların hem eğitim hem de test kümelerinde yer alabilmesi modelin konuşmacıya özgü özellikleri öğrenmesine ve performans sonuçlarının olduğundan yüksek görünmesine neden olabilmektedir. Bu nedenle modelin konuşmacı bağımsız genelleme yeteneğini daha gerçekçi bir şekilde değerlendirebilmek amacıyla LOSO doğrulama yöntemi de uygulanmıştır.

LOSO yönteminde her yinelemede bir konuşmacıya ait tüm kayıtlar test kümesi olarak ayrılırken diğer konuşmacılara ait kayıtlar eğitim kümesi olarak kullanılmaktadır. Böylece model, eğitim sürecinde hiç karşılaşmadığı bir konuşmacı üzerinde değerlendirilmekte ve konuşmacı bağımsız performansı ölçülebilmektedir. Veri setinde yer alan her konuşmacı bir kez test verisi olarak kullanıldıktan sonra elde edilen performans sonuçlarının ortalaması alınarak modelin genel başarımı hesaplanmaktadır.

LOSO doğrulama yöntemi özellikle konuşma işleme ve biyometrik sınıflandırma uygulamalarında yaygın olarak kullanılmaktadır. Bu yöntem, modelin daha önce görmediği bireyler üzerindeki performansını ortaya koyduğu için gerçek kullanım senaryolarını daha iyi temsil etmektedir. Bu bağlamda, tabakalı 5 katlı çapraz doğrulama yöntemi modelin farklı veri bölünmeleri üzerindeki genel performansını değerlendirirken LOSO doğrulama yöntemi modelin konuşmacı bağımsız genelleme yeteneğini ölçmektedir.

Sonuç olarak, her iki doğrulama yönteminin birlikte kullanılması sayesinde önerilen Conv1D+NCP tabanlı modelin performansı hem farklı veri bölünmeleri hem de daha önce görülmemiş konuşmacılar üzerindeki genellenebilirlik açısından kapsamlı ve güvenilir bir şekilde değerlendirilmiştir.

4. BULGULAR

4.1 Deneysel Kurulum

Bu çalışmada önerilen Conv1D-NCP tabanlı konuşmadan cinsiyet sınıflandırma modelinin performansını değerlendirmek amacıyla kapsamlı deneysel çalışmalar gerçekleştirilmiştir. Deneysel süreçte kullanılan yazılım ortamı, kütüphaneler ve donanım bileşenleri modelin geliştirilmesi ve eğitilmesi aşamalarında önemli rol oynamıştır.

Deneysel çalışmalar Python programlama dili kullanılarak gerçekleştirilmiştir. Model geliştirme ve eğitim sürecinde TensorFlow ve Keras derin öğrenme kütüphanelerinden yararlanılmıştır. Deneyler yüksek hesaplama gereksinimleri nedeniyle derin öğrenme uygulamalarını destekleyen Google Colab ortamında gerçekleştirilmiştir(Google Colab 2025). Model eğitimi sırasında işlemci (CPU) ve grafik işlem birimi (GPU) kaynaklarından yararlanılmıştır. GPU kullanımı, özellikle büyük veri kümeleri üzerinde model eğitimi süresinin azaltılmasına önemli katkı sağlamaktadır (Krizhevsky vd. 2012).

4.2 Elde Edilen Sonuçlar

Bu çalışmada önerilen hibrit modelin performansı, Mozilla Common Voice 12.0 Türkçe veri seti üzerinde gerçekleştirilen deneyler ile kapsamlı bir şekilde değerlendirilmiştir. Elde edilen sonuçlar daha önce yayınlanan çalışmada sunulmuştur (Olgun vd. 2026) . NCP tabanlı hibrit modele ait katman bazlı performans sonuçları Tablo 4.1'de sunulmaktadır. Tablo incelendiğinde modelin tüm katmanlarda yüksek doğruluk değerleri elde ettiği ve performansın veri bölünmelerine bağlı olarak önemli bir dalgalanma göstermediği dikkat çekmektedir. Model ortalama %99.29 doğruluk ve %99.28 F1 skoru ile oldukça başarılı bir sınıflandırma performansı sergilemiştir. Bu sonuçlar modelin yalnızca doğru sınıflandırma oranı açısından değil aynı zamanda kesinlik ve duyarlılık dengesini de başarılı bir şekilde koruyabildiğini göstermektedir.

Tablo 4.1 NCP tabanlı model kullanılarak Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar

Katman	Doğruluk (%)	F1-skoru (%)	Kesinlik (%)	Duyarlılık (%)
1	99.65	99.64	99.51	99.78
2	99.18	99.17	99.68	98.65
3	98.61	98.59	98.13	99.06
4	99.45	99.44	99.73	99.15
5	99.58	99.57	99.55	99.60
Ortalama	99.29	99.28	99.32	99.25

Katman bazında yapılan değerlendirmelerde doğruluk değerlerinin tüm katmanlarda %98'in üzerinde olduğu gözlenmiştir. En yüksek doğruluk değeri %99.65 ile 1. katmanda elde edilirken, en düşük doğruluk değeri %98.61 ile 3. katmanda gerçekleşmiştir. Bununla birlikte en düşük ve en yüksek değerler arasındaki farkın oldukça sınırlı olması modelin farklı veri alt kümeleri üzerinde kararlı bir performans sergilediğini ortaya koymaktadır. Bu durum modelin belirli bir veri dağılımına aşırı uyum sağlamadığını ve genellenebilirlik açısından güçlü bir yapıya sahip olduğunu göstermektedir.

UKSB tabanlı hibrit modele ait performans sonuçları ise Tablo 4.2'de ayrıntılı olarak sunulmaktadır. Elde edilen bulgular incelendiğinde, modelin ortalama doğruluk değerinin %99.20, ortalama F1 skorunun ise %99.19 olduğu görülmektedir. Bu yüksek performans değerleri modelin hem doğru sınıflandırma yapma kapasitesinin hem de sınıflar arasındaki dengeyi koruma başarısının oldukça güçlü olduğunu göstermektedir. Özellikle F1 skorunun doğruluk değeriyle neredeyse aynı seviyede olması modelin yalnızca çoğunluk sınıfa değil her iki sınıfa karşı da dengeli bir performans sergilediğini ortaya koymaktadır.

Tablo 4.2 UKSB tabanlı model kullanılarak Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar

Katman	Doğruluk (%)	F1-skoru (%)	Kesinlik (%)	Duyarlılık (%)
1	98.70	98.66	99.86	97.49
2	98.92	98.90	99.01	98.79
3	99.01	98.99	98.75	99.24
4	99.78	99.78	99.82	99.73

5	99.60	99.60	99.51	99.69
Ortalama	99.20	99.19	99.39	98.99

Katman bazlı değerlendirmeler incelendiğinde, tüm katmanlarda doğruluk değerlerinin %98'in üzerinde gerçekleştiği gözlemlenmiştir. Bu durum modelin farklı veri bölünmelerine karşı tutarlı sonuçlar ürettiğini ve genellenebilirliğinin yüksek olduğunu göstermektedir. En yüksek doğruluk değeri 4. katmanda %99.78 olarak elde edilirken, en düşük doğruluk değeri ise 1. katmanda %98.70 olarak kaydedilmiştir. Katmanlar arasındaki bu sınırlı performans farkı, modelin istikrarlı bir öğrenme süreci gerçekleştirdiğini ve veri dağılımındaki değişimlerden minimum düzeyde etkilendiğini ortaya koymaktadır. Dolayısıyla düşük varyanslı bu sonuçlar modelin güvenilirliğini destekleyen önemli bir göstergedir.

Öte yandan UKSB tabanlı modelin elde ettiği yüksek performansa rağmen NCP tabanlı hibrit model ile yapılan karşılaştırmalar dikkate alındığında NCP modelinin ortalama doğruluk ve F1 skoru açısından bir miktar daha üstün olduğu görülmektedir. Bu fark her ne kadar düşük olsa da NCP mimarisinin zamansal bağımlılıkları daha etkin modelleyebilme yeteneği sayesinde daha başarılı sonuçlar ürettiği söylenebilir. Sonuç olarak her iki model de oldukça yüksek performans sergilemekte olup UKSB tabanlı model güçlü bir alternatif sunarken NCP tabanlı model genel performans metrikleri açısından daha avantajlı bir yapı ortaya koymaktadır.

Tam Bağlantılı Katman tabanlı hibrit modele ait performans sonuçları ise Tablo 4.3'te ayrıntılı olarak sunulmaktadır. Elde edilen bulgular incelendiğinde, modelin ortalama doğruluk değerinin %98.96, ortalama F1 skorunun ise %98.94 olduğu görülmektedir. Bu sonuçlar modelin konuşma sinyallerinden cinsiyet sınıflandırma görevinde yüksek bir başarı sergilediğini göstermektedir. Ayrıca doğruluk ve F1 skoru değerlerinin birbirine oldukça yakın olması modelin sınıflar arasında dengeli bir performans ortaya koyduğunu ve yalnızca belirli bir sınıfa odaklanmadan etkili bir sınıflandırma gerçekleştirdiğini göstermektedir. Bu durum, modelin her iki sınıfı da etkili biçimde öğrenebildiğini ve sınıflandırma sürecinde taraflı bir yaklaşım sergilemediğini göstermektedir.

Tablo 4.3 Conv1D+Tam Bağlantılı Katman hibrit modeli kullanılarak Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar

Katman	Doğruluk (%)	F1-skoru (%)	Kesinlik (%)	Duyarlılık (%)
1	98.70	98.67	99.23	98.12
2	99.32	99.30	99.64	98.97
3	98.87	98.86	98.19	99.55
4	99.25	99.24	98.84	99.64
5	98.70	98.67	99.01	98.34
Ortalama	98.96	98.94	98.98	98.92

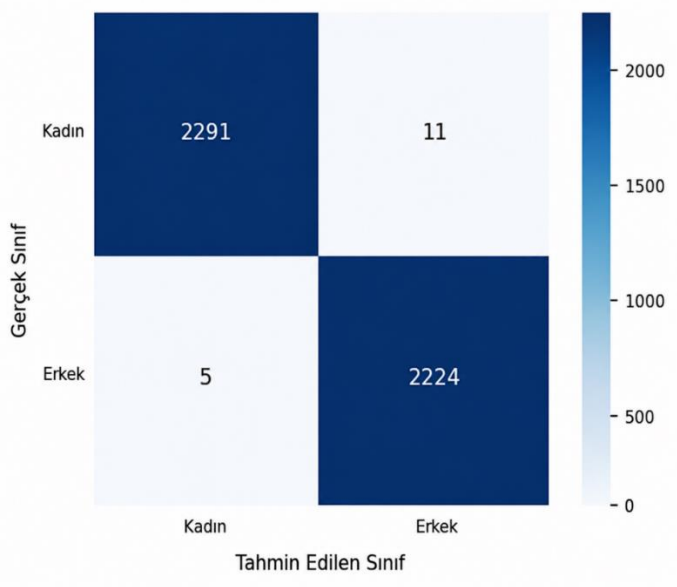
Katman bazlı değerlendirmeler incelendiğinde, tüm katlarda doğruluk değerlerinin %98'in üzerinde gerçekleştiği gözlenmiştir. En yüksek doğruluk değeri 2. katmanda %99.32 olarak elde edilirken, en düşük doğruluk değeri 1.ve 5. katmanlarda %98.70 olarak kaydedilmiştir. Katmanlar arasındaki performans farkının oldukça sınırlı olması, modelin istikrarlı bir öğrenme süreci gerçekleştirdiğini ve veri dağılımındaki değişimlerden düşük düzeyde etkilendiğini ortaya koymaktadır. Dolayısıyla düşük varyanslı bu sonuçlar Tam Bağlantılı Katman tabanlı hibrit modelin güvenilir ve kararlı bir sınıflandırma performansı sunduğunu desteklemektedir.

TSA tabanlı hibrit modele ait performans sonuçları ise Tablo 4.4'te ayrıntılı olarak sunulmaktadır. Elde edilen bulgular incelendiğinde, modelin ortalama doğruluk değerinin %98.91, ortalama F1 skorunun ise %98.89 olduğu görülmektedir. Bu sonuçlar modelin konuşma sinyallerinden cinsiyet sınıflandırma görevinde başarılı bir performans sergilediğini göstermektedir. Özellikle ortalama kesinlik ve duyarlılık değerlerinin yüksek olması, modelin hem yanlış pozitif hem de yanlış negatif sınıflandırmaları düşük seviyede tutabildiğini ortaya koymaktadır. Ayrıca doğruluk ve F1 skoru değerlerinin birbirine yakın olması modelin sınıflar arasında dengeli bir ayrım gerçekleştirebildiğini ve güvenilir sınıflandırma sonuçları ürettiğini ortaya koymaktadır.

Tablo 4.4 TSA tabanlı hibrit modele ait Tabakalı 5-katlı Çapraz Doğrulama ile elde edilen başarımlar

Katman	Doğruluk (%)	F1-skoru (%)	Kesinlik (%)	Duyarlılık (%)
1	98.90	98.87	99.86	97.89
2	99.21	99.19	99.24	99.15
3	99.58	99.57	99.64	99.51
4	98.01	98.00	97.13	98.88
5	98.87	98.85	99.46	98.25
Ortalama	98.91	98.89	99.06	98.73

Katman bazlı değerlendirmeler incelendiğinde, doğruluk değerlerinin %98'in üzerinde gerçekleştiği gözlenmiştir. Bu durum modelin farklı veri bölünmelerine karşı tutarlı sonuçlar ürettiğini ve genellenebilirliğinin yüksek olduğunu göstermektedir. En yüksek doğruluk değeri 3. katmanda %99.58 olarak elde edilirken, en düşük doğruluk değeri 4. katmanda %98.01 olarak kaydedilmiştir. Katmanlar arasındaki performans farklılığının diğer modellere kıyasla bir miktar daha yüksek olduğu gözlenmekle birlikte tüm katlarda elde edilen sonuçların %98'in üzerinde olması modelin genel olarak istikrarlı bir performans sergilediğini göstermektedir. Dolayısıyla elde edilen sonuçlar, TSA tabanlı hibrit modelin farklı veri bölünmelerinde güvenilir ve kararlı bir sınıflandırma performansı sunduğunu destekleyen önemli göstergeler olarak değerlendirilebilir.



Şekil 4.1 1. katman için elde edilen sınıflandırma sonuçlarının karmaşıklık matrisi

Şekil 4.1'de önerilen NCP tabanlı modelin en başarılı sonuç verdiği katman 1'e ait karmaşıklık matrisi gösterilmiştir. Elde edilen sonuçlara göre model, 2.291 kadın örneğinden yalnızca 11'ini, 2.224 erkek örneğinden ise 5'ini yanlış sınıflandırmıştır. Bu oldukça düşük hata oranı modelin her iki sınıf için de yüksek doğrulukla ve dengeli bir şekilde çalıştığını göstermektedir.

Karmaşıklık matrisi incelendiğinde, doğru sınıflandırmaların baskın olduğu ve hataların oldukça sınırlı kaldığı görülmektedir. Bu durum modelin hem kadın hem de erkek sınıflarında benzer başarı seviyesine ulaşarak herhangi bir sınıfa yönelik belirgin bir yanlılık göstermediğini ortaya koymaktadır. Ayrıca düşük hata sayıları, modelin yüksek hassasiyet ve geri çağırma değerlerine sahip olduğunu dolaylı olarak desteklemektedir.

Yanlış sınıflandırılan örnekler niteliksel olarak değerlendirildiğinde ise hataların büyük ölçüde düşük sinyal–gürültü oranına sahip kayıtlar ile frekans özellikleri bakımından birbirine yakın olan sınır durum örneklerinden kaynaklandığı görülmüştür. Bu tür durumlarda konuşma sinyalinin ayırt edici özellikleri zayıflamakta ve modelin karar verme süreci zorlaşmaktadır. Sonuç olarak model genel anlamda oldukça yüksek bir performans sergilemektedir. Bu bulgular modelin büyük çoğunlukla yüksek doğrulukla çalıştığını göstermektedir. Ancak gürültülü ya da sınırda kalan örneklerde performansının kısmen düşebildiğini göstermektedir.

Zamansal modelleme bileşeninin ve akustik öznitelik türünün modele olan katkısını daha kapsamlı bir şekilde incelemek amacıyla ek deneyler gerçekleştirilmiştir. Bu kapsamda Conv1D tabanlı mimariler; TSA, Tam Bağlantılı Katman, UKSB ve önerilen NCP yapılarıyla birlikte değerlendirilmiştir. Böylece farklı mimari yaklaşımların konuşmadan cinsiyet sınıflandırma başarılarına katkıları sistematik olarak analiz edilmiştir.

Karşılaştırmaların adil ve güvenilir olabilmesi için tüm modeller aynı deneysel koşullar altında eğitilmiş ve test edilmiştir. Eğitim sürecinde kullanılan hiper parametreler, optimizasyon yöntemi, epoch sayısı ve doğrulama stratejileri sabit tutulmuştur. Böylece performans farklılıklarının yalnızca model mimarisi ve kullanılan özniteliklerden kaynaklanması sağlanmıştır. Değerlendirme aşamasında ise doğruluk ve F1 skoru gibi yaygın

performans metrikleri kullanılarak modellerin başarımı çok yönlü olarak incelenmiştir.

Öznitelik temsillerinin etkisini analiz edebilmek amacıyla MFCC ve log-Mel spektrogram öznitelikleri ayrı ayrı ve birlikte kullanılarak deneyler gerçekleştirilmiştir. Bu yaklaşım sayesinde her bir öznitelik türünün bireysel katkısı ile birleşik kullanımın sağladığı tamamlayıcı bilginin modele olan etkisi değerlendirilmiştir. Elde edilen sonuçlar, bu özniteliklerin birlikte kullanılmasıyla modelin veriyi daha kapsamlı bir şekilde temsil edebildiğini ve sınıflandırma performansının arttığını göstermektedir.

Gerçekleştirilen tüm deneylerin karşılaştırmalı sonuçları Tablo 4.5'te özetlenmiş olup farklı mimari seçimlerin ve öznitelik temsillerinin genel doğruluk ve F1 skoru üzerindeki etkileri açık bir şekilde ortaya konulmuştur. Bulgular incelendiğinde, UKSB ve özellikle NCP mimarilerinin basit TSA ve Dense yapılara kıyasla daha yüksek doğruluk ve F1 skorları elde ettiği görülmektedir. Bu durum zamansal bağımlılıkları modelleyebilen gelişmiş yapıların konuşma sinyali gibi dinamik veriler üzerinde daha etkili olduğunu ortaya koymaktadır.

MFCC ve log-Mel öznitelikleri tek başlarına da başarılı sonuçlar üretmekle birlikte, performansı belirleyen temel unsurun kullanılan mimari yapı olduğu görülmektedir. Özellikle NCP mimarisinin sınırlı parametre sayısına rağmen yüksek başarı göstermesi bu yapının hem hesaplama verimliliği hem de öğrenme kapasitesi açısından önemli avantajlar sunduğunu ortaya koymaktadır. Bu bulgular, konuşma tabanlı cinsiyet sınıflandırma problemlerinde mimari tasarımın kritik bir rol oynadığını göstermektedir.

Tablo 4.5 MFCC ve log-Mel öznitelikleri kullanılarak Conv1D tabanlı mimarilerin performans karşılaştırması

Model	**Öznitelikler**	**Doğruluk (%)**	**F1-skoru (%)**
Conv1D+ TSA	MFCC + log-mel	98.91	98.89
Conv1D+ Dense	MFCC + log-mel	98.97	98.95
Conv1D+ UKSB	MFCC	99.19	99.18
Conv1D+ UKSB	log-mel	99.20	99.19
Conv1D+ NCP	MFCC	98.79	98.76
Conv1D+ NCP	log-mel	98.82	98.81

Modelin daha önce eğitim sürecinde yer almayan konuşmacılar üzerindeki performansını değerlendirmek amacıyla farklı Conv1D tabanlı mimariler üzerinde konuşmacıdan bağımsız doğrulama deneyleri gerçekleştirilmiştir. Bu yaklaşımda her yinelemede bir konuşmacıya ait tüm veriler test kümesi olarak ayrılmış kalan konuşmacılara ait veriler ise eğitim sürecinde kullanılmıştır. Bu sayede modelin daha önce hiç görmediği konuşmacılar üzerindeki genelleme performansı gerçekçi bir şekilde değerlendirilmiştir. Veri seti analiz edildiğinde toplam 16 farklı konuşmacıya ait konuşma kayıtlarının bulunduğu görülmektedir. Konuşmacılar arasında dengeli bir değerlendirme sağlamak amacıyla her bir konuşmacıdan rastgele 139 konuşma örneği seçilmiştir. Ancak bir konuşmacı için yalnızca 21 örnek mevcut olduğundan bu doğal dağılım korunmuş diğer 15 konuşmacıya ait örnek sayıları bu dengeye göre düzenlenmiştir. Bu durum veri setindeki sınırlı örnek içeren konuşmacının modele etkisini daha gerçekçi bir şekilde yansıtmayı amaçlamaktadır.

Elde edilen sonuçlar Tablo 4.6'da sunulmaktadır. Tablo incelendiğinde, tüm modellerin %80'in üzerinde doğruluk değerine ulaştığı görülmektedir. Bu durum kullanılan mimarilerin konuşmacıdan bağımsız senaryoda dahi belirli bir başarı düzeyini koruyabildiğini göstermektedir. Bununla birlikte önerilen Conv1D+NCP tabanlı model %87.19 doğruluk oranına ulaşarak güçlü bir performans sergilemiştir. Öte yandan Conv1D+UKSB modeli en yüksek doğruluk değerini elde etmiş olsa da Conv1D+NCP modeli %92.50 F1 skoru ile daha dengeli bir sınıflandırma performansı ortaya koymuştur. F1 skorunun yüksek olması, modelin kesinlik ve duyarlılık arasındaki dengeyi başarılı biçimde koruduğunu ve sınıflandırma hatalarını dengeli biçimde azalttığını göstermektedir.

NCP tabanlı mimarinin özellikle konuşmacı çeşitliliğinin yüksek olduğu durumlarda daha güçlü bir genelleme yeteneğine sahip olduğu görülmektedir. Zamansal bağımlılıkları modelleme kapasitesinin yüksek olması sayesinde NCP yapısı, konuşmacıya özgü varyasyonları daha etkili bir şekilde yönetebilmekte ve farklı konuşmacılar arasında daha tutarlı sonuçlar üretebilmektedir. Buna karşılık Conv1D+Dense ve Conv1D+TSA modelleri de rekabetçi sonuçlar üretmekle birlikte daha sınırlı zamansal modelleme yetenekleri nedeniyle karmaşık konuşma örüntülerini öğrenmede diğer modellere göre daha düşük per-

formans sergilemektedir. Sonuç olarak konuşmacıdan bağımsız değerlendirme senaryolarında gelişmiş zamansal modelleme yaklaşımlarının belirgin bir avantaj sağladığı açıkça görülmektedir.

Tablo 4.6 Konuşmacıdan bağımsız (LOSO) doğrulama altında farklı Conv1D tabanlı mimarilerin sınıflandırma performansları

Model	**Doğruluk (%)**	**F1-skoru (%)**
Conv1D+ TSA	82.10	88.06
Conv1D+ Dense	82.28	87.91
Conv1D+ UKSB	88.76	91.80
Conv1D+ NCP	87.19	92.50

5. TARTIŞMA ve SONUÇ

5.1 Genel Değerlendirme

Bu çalışmada Türkçe konuşma verileri üzerinde cinsiyet sınıflandırması gerçekleştirmek amacıyla MFCC ve log-Mel spektrogram özniteliklerini birleştiren Conv1D tabanlı hibrit derin öğrenme mimarileri geliştirilmiş ve farklı zamansal modelleme yaklaşımlarının performansları karşılaştırılmıştır. Elde edilen sonuçlar, önerilen Conv1D+NCP tabanlı modelin konuşmadan cinsiyet tanıma probleminde yüksek doğruluk, güçlü ayırt edicilik ve hesaplama verimliliği açısından başarılı bir performans sergilediğini göstermektedir.

Tabakalı 5-katlı çapraz doğrulama deneylerinde Conv1D+NCP modeli ortalama %99,29 doğruluk ve %99,28 F1-skoru elde ederek tüm modeller arasında en yüksek genel performansı göstermiştir. Aynı sayıda birime sahip Conv1D+UKSB modeli ise %99,20 doğruluk ve %99,19 F1-skoru ile benzer düzeyde başarılı sonuçlar üretmiştir. Her iki modelin de tüm katmanlarda %98'in üzerinde doğruluk sağlaması, MFCC ve log-Mel spektrogram özniteliklerinin konuşma tabanlı cinsiyet sınıflandırma probleminde yüksek ayırt edicilik sunduğunu göstermektedir.

Karşılaştırmalı analizler, zamansal modelleme yeteneğine sahip mimarilerin zamansal olmayan yaklaşımlara göre daha başarılı olduğunu ortaya koymuştur. Tam Bağlantılı Katman tabanlı hibrit model %98,96 doğruluk ve %98,94 F1-skoru, TSA tabanlı hibrit model ise %98,91 doğruluk ve %98,89 F1-skoru elde etmiştir. Ancak her iki yaklaşım da NCP ve UKSB tabanlı modellerin gerisinde kalmıştır. Bu durum, konuşma sinyalindeki zamansal bağımlılıkların sınıflandırma başarısında önemli bir rol oynadığını ve bu bağımlılıkları etkili biçimde modelleyebilen yapıların daha yüksek performans sağlayabildiğini göstermektedir.

NCP tabanlı modelin dikkat çekici özelliklerinden biri, daha düşük parametre sayısına rağmen yüksek performansını koruyabilmesidir. Yaklaşık 133,995 parametre içeren NCP modeli, 135,233 parametreye sahip UKSB modeline kıyasla benzer hatta bazı ölçütlerde

daha yüksek başarı elde etmiştir. Bu sonuç, NCP mimarisinin konuşma sinyalindeki dinamik değişimleri etkili biçimde temsil edebildiğini ve yüksek doğruluk ile hesaplama verimliliğini birlikte sağlayabildiğini göstermektedir. Bu avantaj özellikle gerçek zamanlı sistemler, gömülü uygulamalar ve sınırlı donanım kaynaklarına sahip platformlar açısından önem taşımaktadır.

Modelin genellenebilirliğini değerlendirmek amacıyla gerçekleştirilen konuşmacıdan bağımsız doğrulama deneylerinde doğruluk oranlarının tabakalı çapraz doğrulama sonuçlarına kıyasla daha düşük olduğu gözlenmiştir. Bunun temel nedeni, modelin eğitim sürecinde hiç karşılaşmadığı konuşmacılara ait veriler üzerinde test edilmesidir. Bu daha zorlu değerlendirme senaryosunda Conv1D+NCP modeli %87,19 doğruluk ve %92,50 F1-skoru elde ederek en yüksek F1 performansını göstermiştir. Conv1D+UKSB modeli %88,76 doğruluk ile biraz daha yüksek bir doğruluk değerine ulaşsa da %91,80 F1-skorunda kalmıştır. Bu sonuçlar, NCP tabanlı modelin farklı konuşmacılar arasında daha dengeli bir sınıflandırma gerçekleştirdiğini ve güçlü bir genelleme yeteneğine sahip olduğunu ortaya koymaktadır.

Sonuç olarak, elde edilen bulgular önerilen Conv1D+NCP tabanlı hibrit modelin Türkçe konuşma verileri üzerinde cinsiyet sınıflandırması için etkili, güvenilir ve verimli bir çözüm sunduğunu göstermektedir. Modelin hem tabakalı çapraz doğrulama hem de konuşmacıdan bağımsız doğrulama değerlendirmelerinde başarılı sonuçlar elde etmesi gerçek dünya uygulamalarında kullanılabilirliğini desteklemektedir. Ayrıca daha düşük parametre sayısıyla yüksek performans sunabilmesi, biyolojik esinli zamansal modelleme yaklaşımlarının konuşma işleme alanındaki potansiyelini ortaya koymakta ve gelecekte gerçekleştirilecek çalışmalar için önemli bir temel oluşturmaktadır.

5.2 Diğer Yöntemlerle Karşılaştırma

Bu çalışmada önerilen Conv1D+NCP tabanlı hibrit modelin performansı literatürde yaygın olarak kullanılan geleneksel makine öğrenmesi ve derin öğrenme tabanlı yöntemler ile karşılaştırmalı olarak değerlendirilmiştir. Karşılaştırma kapsamında farklı yöntemlerin doğruluk oranları kullanılan öznitelikler ve model özellikleri dikkate alınmıştır.

Konuşmadan cinsiyet tanıma çalışmalarında geleneksel makine öğrenmesi yöntemleri yaygın olarak kullanılmaktadır. Özellikle MFCC özniteliklerinin DVM, GMM ve KNN gibi sınıflandırıcılarla birlikte kullanıldığı çalışmalar dikkat çekmektedir. Bu yöntemler düşük hesaplama maliyeti ve basit uygulama yapıları sayesinde avantaj sağlamakla birlikte konuşma sinyalinin zamana bağlı yapısını doğrudan modelleyememeleri nedeniyle performans açısından belirli sınırlılıklara sahiptir. Ayrıca bu yaklaşımlarda öznitelik çıkarım süreci manuel olarak gerçekleştirildiğinden model başarısı büyük ölçüde kullanılan özniteliklerin kalitesine bağlı kalmaktadır.

Derin öğrenme yaklaşımları, konuşma sinyallerindeki ayırt edici örüntülerin otomatik olarak öğrenilmesini sağlamaktadır. ESA tabanlı modeller yerel akustik özellikleri etkili biçimde çıkarırken uzun süreli zamansal ilişkilerin modellenmesi için UKSB gibi yapılarla birleştirilebilmektedir. Ancak bu hibrit modellerin artan parametre sayısı ve hesaplama maliyeti daha verimli mimarilere duyulan gereksinimi ortaya çıkarmaktadır.

Bu çalışmada önerilen Conv1D+NCP modeli ise hem spektral hem de zamansal özellikleri düşük parametre sayısı ile modelleyebilen alternatif bir yaklaşım sunmaktadır. NCP mimarisi, sürekli zaman yaklaşımı ve biyolojik esinli yapısı sayesinde zamansal bağımlılıkları daha etkin bir şekilde öğrenebilmekte ve daha az parametre ile rekabetçi performans elde edebilmektedir. Bu kapsamda literatürde yer alan yöntemler ile önerilen modelin karşılaştırılması Tablo 5.1'de sunulmuştur.

Tablo 5.1 Konuşmadan Cinsiyet Tanıma yöntemlerinin literatür tabanlı karşılaştırılması

Yöntem	**Veri Kümesi**	**Dil**	**Kullanılan Öznitelikler**	**Doğruluk (%)**	**Avantaj**	**Dezavantaj**
DVM (Younis vd. 2023)	TIMIT	İngilizce	MFCC	96,45	Düşük hesaplama maliyeti	Zamansal bağımlılık yok
GMM + DVM (Munoli vd. 2023b)	TIMIT	İngilizce	MFCC	80,30	Basit yapı	Genelleme sınırlı

GMM (Yücesoy ve Nabiyev 2013)	TIMIT	İngilizce	MFCC	97,76	İstatistiksel modelleme güçlü	Karmaşık veride sınırlı
GMM / MLP / VQ / LVQ (Djemili vd. 2012)	IViE Speech Database	İngilizce	MFCC	96,40	Farklı model karşılaştırması	Kısa veri bağımlı
Lojistik Regresyon (Hızlısoy vd. 2022)	58 farklı dizi ve filmden oluşturulan Türkçe ses veri seti	Türkçe	MFCC + Spektrogram	89,00	Basit model	Düşük doğruluk
ESA (Yücesoy 2024a)	Mozilla Common Voice-27.04.2022 sürümü	Türkçe	MFCC	98,67	Otomatik öznitelik öğrenme	Zamansal modelleme sınırlı
ESA (Yücesoy 2023)	Mozilla Common Voice-v11.0	Türkçe	MFCC	94,40	Spektral örüntü öğrenme	Zamansal bağımlılık yok
Spiking Sinir Ağı (SNN) (Tümen 2022)	Voice Gender Recognition Dataset	İngilizce	Konuşma sinyali	98,84	Biyolojik modelleme	Karmaşık yapı
EfficientNet (Derin ESA) (Jorrin-Coz vd. 2025)	Mozilla Common Voice v17.0, TIMIT, VoxCeleb1	Çok dilli	Spektrogram	99,82	Çok yüksek doğruluk	Çok yüksek maliyet
Conv1D + NCP (Önerilen)	**Mozilla Common Voice-v12.0**	**Türkçe**	**MFCC + Log-Mel**	**99,29**	**Yüksek doğruluk + düşük parametre**	**Gürültüye hassasiyet**

Tablo 5.1 incelendiğinde konuşmadan cinsiyet tanıma probleminde kullanılan yöntemlerin performanslarının büyük ölçüde model mimarisi ve kullanılan öznitelik türlerine bağlı olarak değiştiği görülmektedir. Geleneksel makine öğrenmesi yöntemleri genel olarak %80–98 doğruluk aralığında sonuçlar üretirken derin öğrenme tabanlı yöntemlerin daha yüksek başarı değerlerine ulaştığı dikkat çekmektedir. Özellikle ESA tabanlı derin modeller ve biyolojik esinli sinir ağı yaklaşımları konuşma sinyalinden ayırt edici özellikleri daha etkili bir şekilde öğrenebilmektedir.

Önerilen Conv1D+NCP modeli %99,29 doğruluk oranı ile literatürde yer alan birçok yöntemden daha yüksek performans göstermiştir. Elde edilen bu sonuç, Conv1D katmanlarının sağladığı spektral temsil gücü ile NCP mimarisinin zamansal modelleme yeteneğinin birbirini tamamladığını göstermektedir. Ayrıca modelin daha düşük parametre sayısıyla çalışabilmesi hesaplama verimliliği açısından önemli bir avantaj sağlamaktadır.

Bu çalışmada gerçekleştirilen deneylerde NCP tabanlı model, UKSB tabanlı modellere kıyasla daha dengeli ve kararlı bir performans sergilemiştir. Bu bulgu, modelin farklı konuşmacılara ait veriler üzerindeki genelleme başarısını desteklemektedir.

Literatürde daha yüksek doğruluk oranlarına ulaşan çalışmalar da bulunmaktadır. Örneğin EfficientNet tabanlı derin ESA modeli %99,82 doğruluk oranı elde etmiştir. Ancak bu tür modeller yüksek hesaplama maliyeti ve büyük parametre sayısı gerektirmektedir. Buna karşın önerilen Conv1D+NCP modeli, doğruluk ve hesaplama verimliliği arasında daha dengeli bir yapı sunmaktadır.

Sonuç olarak elde edilen bulgular, önerilen Conv1D+NCP tabanlı yaklaşımın konuşmadan cinsiyet tanıma probleminde yüksek doğruluk sağlayabildiğini ve hesaplama maliyeti açısından avantaj sunduğunu göstermektedir. Bu yönüyle model özellikle gerçek zamanlı uygulamalar ve kaynak kısıtlı sistemler için etkili ve uygulanabilir bir alternatif olarak değerlendirilebilir.

5.3 Katkılar ve Sınırlılıklar

Bu çalışmanın temel katkısı, NCP mimarisinin Türkçe konuşma sinyallerinden cinsiyet sınıflandırma problemine uyarlanması ve farklı Conv1D tabanlı mimarilerle aynı deneysel koşullar altında karşılaştırılmasıdır. MFCC ve log-Mel spektrogram özniteliklerinin birlikte değerlendirilmesiyle konuşma sinyaline ait tamamlayıcı akustik bilgilerin tek bir model içerisinde kullanılması sağlanmıştır. Ayrıca TSA, Tam Bağlı, UKSB ve NCP tabanlı yapıların karşılaştırılması, farklı zamansal modelleme yaklaşımlarının sınıflandırma performansı üzerindeki etkisinin belirlenmesine olanak tanımıştır.

Çalışmanın bir diğer katkısı, model performansının yalnızca tek bir veri bölünmesi üzerinden değerlendirilmemesidir. Tabakalı 5-katlı çapraz doğrulamanın yanı sıra konuşmacıdan bağımsız doğrulamanın kullanılmasıyla modelin daha önce görülmeyen konuşmacılar üzerindeki başarısı da incelenmiştir. Böylece önerilen yaklaşımın sınıflandırma başarısı ve genelleme yeteneği farklı değerlendirme senaryoları altında ortaya konulmuştur.

Çalışmanın bazı sınırlılıkları bulunmaktadır. Deneylerin yalnızca Mozilla Common Voice 12.0 Türkçe veri seti üzerinde gerçekleştirilmiş olması elde edilen sonuçların farklı dillere ve veri kümelerine doğrudan genellenmesini sınırlandırmaktadır. Ayrıca sınıflandırma problemi, veri setinde yer alan etiketlere bağlı olarak yalnızca kadın ve erkek sınıfları üzerinden ele alınmıştır. Bu nedenle daha geniş demografik çeşitliliği kapsayan sınıflandırma senaryoları çalışma kapsamı dışında kalmıştır.

Yanlış sınıflandırmaların özellikle düşük sinyal-gürültü oranına sahip kayıtlar ile akustik özellikleri birbirine yakın örneklerde yoğunlaşması model performansının kayıt kalitesinden etkilenebildiğini göstermektedir. Ayrıca modelin gerçek zamanlı kullanılabilirliği parametre sayısı bakımından değerlendirilmiş; farklı donanımlardaki çıkarım süresi, bellek kullanımı ve enerji tüketimi doğrudan ölçülmemiştir. Bu durum, modelin uygulama ortamındaki hesaplama verimliliğinin ayrıca değerlendirilmesini gerekli kılmaktadır.

5.4 Gelecek Çalışmalar İçin Öneriler

Bu çalışmada önerilen Conv1D+NCP tabanlı hibrit model, konuşmadan cinsiyet sınıflandırma probleminde yüksek başarı göstermiştir. Gelecek çalışmalarda modelin farklı diller ve veri kümeleri üzerinde sınanması, yaklaşımın dil ve veri setinden bağımsız genelleme yeteneğinin değerlendirilmesine katkı sağlayacaktır.

Modelin farklı kayıt koşullarına karşı dayanıklılığını artırmak amacıyla gürültü ekleme, zaman kaydırma, hız değiştirme ve perde kaydırma gibi veri artırma yöntemlerinden yararlanılabilir. Ayrıca MFCC ve log-Mel spektrogram özniteliklerine temel frekans, formant, spektral kontrast, enerji, konuşma hızı ve tonlama gibi akustik ve prozodik özelliklerin eklenmesiyle daha kapsamlı temsiller elde edilebilir.

NCP mimarisinin farklı yapılandırmaları ve hiperparametrelerinin incelenmesi de model performansının geliştirilmesine katkı sağlayabilir. Özellikle LTC hücrelerine ait yapısal parametrelerin optimize edilmesi ve NCP mimarisinin dikkat mekanizmaları ya da Transformer tabanlı yapılarla birleştirilmesi araştırılabilir.

Önerilen modelin mobil cihazlar, gömülü sistemler ve uç bilişim ortamlarındaki çıkarım süresi, bellek kullanımı ve enerji tüketimi bakımından değerlendirilmesi, gerçek zamanlı uygulamalardaki kullanılabilirliğinin belirlenmesi açısından önemlidir. Bunun yanında yaş tahmini, duygu analizi ve konuşmacı tanıma gibi farklı paralinguistik görevlerin aynı yapı içerisinde ele alındığı çok görevli öğrenme modelleri geliştirilebilir.

6. KAYNAKLAR

Abdel-Hamid O, Mohamed A, Jiang H, Deng L, Penn G, Yu D, 2014, Convolutional Neural Networks for Speech Recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing 22, ss. 1533-1545. Erişim adresi: https://api.semanticscholar.org/CorpusID:206602362.

Alkhammash E, Hadjouni M, Elshewey A, 2022, A Hybrid Ensemble Stacking Model for Gender Voice Recognition Approach, Electronics 11, s. 1750. doi: 10.3390/electronics11111750.

Alsulaiman M, Ali Z, Muhammad G, 2011, Gender Classification with Voice Intensity, 2011 UKSim 5th European Symposium on Computer Modeling and Simulation, ss. 205-209. doi: 10.1109/EMS.2011.37.

Altunbey Özbay F, Özbay E, 2022, Ses verilerinden cinsiyet tespiti için yeni bir yaklaşım: Optimizasyon yöntemleri ile özellik seçimi, Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi 38(2), ss. 1179-1192. doi: 10.17341/gazimmfd.938294.

Archana GS, Malleswari M, 2015, Gender identification and performance analysis of speech signals, 2015 Global Conference on Communication Technologies (GCCT), ss. 483-489. doi: 10.1109/GCCT.2015.7342709.

Ayyüce Kızrak M, Bolat B, 2018, Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma, Bilişim Teknolojileri Dergisi 11(3), ss. 263-286. doi: 10.17671/gazibtd.419205.

Bakır Ç, 2016, Alman Dili Üzerinde Konuşmacı Cinsiyetinin Otomatik Olarak Belirlenmesi, Academic Platform - Journal of Engineering and Science 4(2). doi: 10.21541/apjes.49291.

Calp MH, 2025, Makine Öğrenmesi Yöntemleri Kullanılarak Ses Verilerinden Cinsiyet Tahmin Edilmesi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi 29(1), ss. 96-113. doi: 10.19113/sdufenbed.1590769.

Chaudhari SJ, Kagalkar RM, 2015, Methodology for gender identification, classification and recognition of human age, Int. J. Comput. Appl 975, s. 8887.

Chaudhary S, Sharma DK, 2018, Gender Identification based on Voice Signal Characteristics, 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), ss. 869-874. doi: 10.1109/ICACCCN.2018.8748676.

Chollet F, 2021, Deep learning with Python, simon and schuster.

Cider Şener A, Gökçer C, 2005, Doğru ve güzel konuşarak gelişen kimlik. Erişim adresi: http://hdl.handle.net/11693/18303.

Çelebi SB, Fidan Ş, 2024, RNN-Based Time Series Analysis for Wind Turbine Energy Forecasting, International Journal of Engineering and Innovative Research 6(1), ss. 15-28. doi: 10.47933/ijeir.1387314.

Djemili R, Bourouba H, Korba MCA, 2012, A speech signal based gender identification system using four classifiers, 2012 International Conference on Multimedia Computing and Systems, ss. 184-187. doi: 10.1109/ICMCS.2012.6320122.

Donuk K, Hanbay D, 2022, Konuşma Duygu Tanıma için Akustik Özelliklere Dayalı LSTM Tabanlı Bir Yaklaşım, Computer Science Vol:7(Issue:2), ss. 54-67. doi: 10.53070/bbd.1113379.

Duru I, Sunar AS, White S, Diri B, 2021, Deep Learning for Discussion-Based Cross-Domain Performance Prediction of MOOC Learners Grouped by Language on FutureLearn, Arabian Journal for Science and Engineering 46, ss. 3613-3629. Erişim adresi: https://api.semanticscholar.org/CorpusID:230773870.

Goodfellow I, 2016, Deep learning.

Google Colab, 2025, Google Colab. Erişim adresi: https://colab.research.google.com/ [Erişim: 30 Temmuz 2026].

Graves A, Mohamed A, Hinton GE, 2013, Speech recognition with deep recurrent neural networks, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, ss. 6645-6649. Erişim adresi: https://api.semanticscholar.org/CorpusID:206741496.

Hızlısoy S, Çolakoğlu E, Arslan RS, 2022, Speech-to-Gender Recognition Based on Machine Learning Algorithms, International Journal of Applied Mathematics Electronics and Computers 10(4), ss. 84-92. doi: 10.18100/ijamec.1221455.

Jorrin-Coz J, Nakano M, Perez-Meana H, Hernandez-Gonzalez L, 2025, Multi-Corpus Benchmarking of CNN and LSTM Models for Speaker Gender and Age Profiling, Computation 13(8). Erişim adresi: https://www.mdpi.com/2079-3197/13/8/177.

Kohavi R, 2001, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, 14.

Kone VS, Anagal A, Anegundi S, Jadhav P, Kulkarni U, M MS, 2023, Voice-based Gender and Age Recognition System, 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), ss. 74-80. doi: 10.1109/InCACCT57535.2023.10141801.

Krizhevsky A, Sutskever I, Hinton GE, 2012, ImageNet Classification with Deep Convolutional Neural Networks, Pereira F, Burges CJ, Bottou L, Weinberger K ed., Advances in Neural Information Processing Systems, Curran Associates, Inc. Erişim adresi: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Lechner M, Hasani R, Amini A, Henzinger T, Rus D, Grosu R, 2020, Neural circuit policies enabling auditable autonomy, Nature Machine Intelligence 2, ss. 642-652. doi: 10.1038/s42256-020-00237-3.

LeCun Y, Bengio Y, Hinton G, 2015a, Deep learning, nature 521(7553), ss. 436-444.

LeCun Y, Bengio Y, Hinton G, 2015b, Deep learning, nature 521(7553), ss. 436-444.

Malashin I, Tynchenko V, Gantimurov A, Nelyub V, Borodulin A, 2024, Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review, Polymers 16(18). Erişim adresi: https://www.mdpi.com/2073-4360/16/18/2607.

Mohammed AA, Al-Irhayim YF, 2021a, An overview for assessing a number of systems for estimating age and gender of speakers, Tikrit Journal of Pure Science 26(1), ss. 94-100.

Mohammed AA, Al-Irhayim YF, 2021b, Speaker age and gender estimation based on deep learning bidirectional long-short term memory (BiLSTM), Tikrit Journal of Pure Science 26(4), ss. 76-84.

Mozilla Foundation, 2024, Mozilla Common Voice.

Munoli BK, Jain KAK, Kumar P, PS AR, others, 2023a, Human voice analysis to determine age and gender, 2023 International conference on recent trends in electronics and communication (ICRTEC), ss. 1-4.

Munoli BK, Jain KAK, Kumar P, S ARP, Ashwini, 2023b, Human Voice Analysis to Determine Age and Gender, 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC), ss. 1-4. doi: 10.1109/ICRTEC56977.2023.10111890.

Nair V, Hinton GE, 2010, Rectified Linear Units Improve Restricted Boltzmann Machines, International Conference on Machine Learning. Erişim adresi: https://api.semanticscholar.org/CorpusID:15539264.

Olgun N, 2022, Lazer işaretleri ile yapay zeka temelli hedef analizi (Artificial intelligence based target analysis with laser signals), Frat University, Turkey.

Olgun S, Balım C, Olgun N, 2026, A CNN–NCP Based Hybrid Deep Learning Model for Speech-Driven Gender Classification, Bitlis Eren Üniversitesi Fen Bilimleri Dergisi 15(1), ss. 312-321. doi: 10.17798/bitlisfen.1803512.

Özbay FA, Özbay E, 2023, A new approach for gender detection from voice data: Feature selection with optimization methods, J Fac Eng Archit Gazi Univ 38(2), ss. 1179-1192.

Purwins H, Li B, Virtanen T, Schlüter J, Chang S, Sainath T, 2019, Deep Learning for Audio Signal Processing, IEEE Journal on Selected Topics in Signal Processing 13. doi: 10.1109/JSTSP.2019.2908700.

Ramdinmawii E, Mittal VK, 2016, Gender identification from speech signal by examining the speech production characteristics, 2016 International Conference on Signal Processing and Communication (ICSC), ss. 244-249. doi: 10.1109/ICSPCom.2016.7980584.

Safavi S, Russell M, Jančovič P, 2018, Automatic speaker, age-group and gender identification from children's speech, Computer Speech & Language 50, ss. 141-156. Erişim adresi: https://www.sciencedirect.com/science/article/pii/S088523081630136X.

Sakarya Ş, Yılmaz Ü, 2019, Derin Öğrenme Mimarisi Kullanarak BİST30 İndeksinin Tahmini, European Journal of Educational and Social Sciences 4(2), ss. 106-121. Erişim adresi: https://dergipark.org.tr/tr/pub/ejees/article/647547.

Sánchez-Hevia HA, Gil-Pita R, Utrilla-Manso M, Rosa-Zurera M, 2022, Age group classification and gender recognition from speech with temporal convolutional neural networks, Multimedia Tools and Applications 81(3), ss. 3535-3552. doi: 10.1007/s11042-021-11614-4.

Tranter SE, Reynolds DA, 2006, An overview of automatic speaker diarization systems, IEEE Transactions on Audio, Speech, and Language Processing 14(5), ss. 1557-1565. doi: 10.1109/TASL.2006.878256.

Tümen V, 2022, Gender Determination in Human Voice Signals using Synaptic Efficacy Function-based Leaky Integrate and Fire Neuron Model, Bitlis Eren Üniversitesi Fen Bilimleri Dergisi 11. doi: 10.17798/bitlisfen.1024236.

Yılmaz EN, Gönen S, Barışkan MA, Karacayılmaz G, Alhan B, Artuner H, 2022, Gender Detection Via Voice Using Artificial Intelligence Algorithms, Gazi Journal of Engineering Sciences 8(3), ss. 567-575. Erişim adresi: https://dergipark.org.tr/en/pub/gmbd/article/1169674.

Yilmaz A, Poli R, 2022, Successfully and efficiently training deep multi-layer perceptrons with logistic activation function simply requires initializing the weights with an appropriate negative mean, Neural Networks 153, ss. 87-103. Erişim adresi: https://www.sciencedirect.com/science/article/pii/S0893608022002040.

Younis HA vd., 2023, Multimodal age and gender estimation for adaptive human-robot interaction: A systematic literature review, Processes 11(5), s. 1488.

Yu D, Deng L, 2011, Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP], IEEE Signal Processing Magazine 28, ss. 145-154. Erişim adresi: https://api.semanticscholar.org/CorpusID:237593.

Yücesoy E, 2020, Konuşmacının Yaş ve Cinsiyetine Göre Sınıflandırılmasında DVM Çekirdeğinin Etkisi, El-Cezeri 7(3), ss. 970-982. doi: 10.31202/ecjse.707179.

Yücesoy E, 2021, MFKK Özniteliklerine Eklenen Logaritmik Enerji ve Delta Parametrelerinin Yaş ve Cinsiyet Sınıflandırma Üzerindeki Etkileri, Journal of the Institute of Science and Technology 11(1), ss. 32-43. doi: 10.21597/jist.772804.

Yücesoy E, 2023, Speaker age and gender recognition using 1D and 2D convolutional neural networks, Neural Computing and Applications 36, ss. 1-11. doi: 10.1007/s00521-023-09153-0.

Yücesoy E, 2024a, Gender Recognition from Speech Signal Using CNN, KNN, SVM and RF, Procedia Computer Science 235, ss. 2251-2257. doi: 10.1016/j.procs.2024.04.213.

Yücesoy E, 2024b, Konuşmacıları Kadın, Erkek ve Çocuk Olarak Sınıflandırmada Veri Artırmanın Performansa Etkisi, Journal of the Institute of Science and Technology 14(3), ss. 974-987. doi: 10.21597/jist.1505349.

Yücesoy E, Nabiyev VV, 2013, Gender identification of a speaker using MFCC and GMM, 2013 8th International Conference on Electrical and Electronics Engineering (ELECO), ss. 626-629. doi: 10.1109/ELECO.2013.6713922.

Yücesoy E, Nabiyev VV, 2016, Konuşmaci Yaş Ve Cinsiyetinin Gkm Süpervektörlerine Dayali Bir Dvm Siniflandiricisi İle Belirlenmesi, Gazi Üniversitesi Mühendislik Mimarlk Fakültesi Dergisi 31(3).

Zaman SR, Sadekeen D, Alfaz MA, Shahriyar R, 2021, One Source to Detect them All: Gender, Age, and Emotion Detection from Voice, 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), ss. 338-343. doi: 10.1109/COMPSAC51774.2021.00055.

Zhao Y, Shu X, 2023, Speech emotion analysis using convolutional neural network (CNN) and gamma classifier-based error correcting output codes (ECOC), Scientific Reports 13(1), s. 20398. doi: 10.1038/s41598-023-47118-4.

ÖZGEÇMİŞ