Makine Öğrenmesiyle Yapay Zekâ Destekli Sesten Duygu Analizi

Mehmet Emin Küçükkurt[1] , Gülay Çiçek[2] , Ramazan Baran Kaynak[3] and Rüzgar Yentür[4]
[1,2]Department of Software Engineering, Faculty of Engineering Architecture
Istanbul Beykent University, Sariyer, Istanbul, Turkey
[1]kucukkurtmm@gmail.com, [2]gulaycicek@gmail.com,
[3]brnkynk0@gmail.com ,[3]ruzgaryenturkk@gmail.com

***Abstract*—Duygular , insan iletişimin en temel bileşenlerinden biridir. Konuşma esnasında , tempo , vurgular , frekans değişimi ve tonlamalar gibi unsurlar kişinin duygusal durumu hakkında pek çok bilgi vermektedir. Bu nedenle ,ses tabanlı duygu analizi, insan-bilgisayar etkileşimi , sağlık , güvenlik , iletişim teknolojileri ve eğitim gibi birçok alanda önem kazanmıştır. Günümzde yapay zeka , yaşamımızın birçok alanında olduğu gibi, ses analizi ve duygu tanıma konularında da önemli bir role sahip olmuştur. Makine öğrenmesi algoritmaları sayesinde ses verilerinden duygusal sınıflandırmaların hızlı ve doğru bir biçimde yapılması mümkündür. Bu çalışmada, makine öğrenmesi tabanlı algoritmalar kullanılarak ses sinyallerinden duyguların tespit edilmesi amaçlanmıştır. Farklı ses veri setleri üzerinde Mel-Frequency Cepstral Coefficients (MFCC), Chroma ve Zero-Crossing Rate (ZCR) gibi özellikler çıkarılmış ve SVM, Random Forest, KNN ve Decision Tree gibi algoritmalar kullanılarak performans karşılaştırması yapılmıştır. Elde edilen sonuçlar, makine öğrenmesi yöntemlerinin duygu sınıflandırmasında etkili olduğunu ve başarılı sonuçlar elde edilebileceğini göstermektedir.**

***Keywords*—Duygu Analizi, Makine Öğrenmesi, Ses Sinyalleri, MFCC, SVM, KNN**

I. Introduction

Bu alt başlıkta, ses tabanlı duygu analizi alanında makine öğrenmesi yöntemleri kullanılarak yapılmış çalışmalara yer verilmiştir. Makine öğrenmesi algoritmaları, ses sinyallerinden çıkarılan istatistiksel ve frekans tabanlı özellikleri analiz ederek duygusal durumların ayırt edilebilmesini sağlamaktadır. Literatürdeki çalışmalar, kullanılan veri setleri, yöntemler, başarı oranları ve sınırlılıklar açısından incelenmiş ve karşılaştırmalı olarak özetlenmiştir.

II. Literature Review

Ses tabanlı duygu analizi, insan iletisiminin temel bileşenlerinden biri olan duyguların makine öğrenmesi yöntemleriyle tespit edilmesini amaçlayan bir araştırma alanıdır. Makine öğrenmesi algoritmaları, ses sinyallerinden çıkarılan Mel-Frequency Cepstral Coefficients (MFCC), Chroma, Zero-Crossing Rate (ZCR) gibi istatistiksel ve frekans tabanlı özellikleri analiz ederek duygusal durumların sınıflandırılmasını sağlamaktadır.

Bu bölümde, son yıllarda bu alanda yapılan çalışmalar incelenmekte; kullanılan veri setleri, uygulanan yöntemler, elde edilen başarı oranları ve karşılaşılan sınırlılıklar açısından karşılaştırmalı bir değerlendirme sunulmaktadır. Literatürde, SVM, Random Forest, KNN, Decision Tree gibi makine öğrenmesi algoritmalarının yanı sıra derin öğrenme yöntemlerinin de kullanımına rastlanmaktadır. Çalışmaların karşılaştırılması, farklı algoritmaların ve özellik çıkarım tekniklerinin duygu tanıma performansı üzerindeki etkilerini ortaya koymaktadır.

Aşağıdaki tablolar (Tablo I ve Tablo II), bu çalışmalardan elde edilen anahtar bulguları özetlemekte ve araştırmamızın metodolojisiyle ilişkilendirilebilecek önemli noktaları göstermektedir.

Kotikalapudi Vamsi Krishn ve ark. (2022) yaptığı çalışmada konuşmacının konuşma esnasında ortaya çıkan duyguların tespitini amaçlamıştır. Çalışmada Duygu Tespiti, günümüzde önemli bir görev olarak ele alınmıştır. Korku, öfke ve sevinç gibi duygulara sahip konuşmaların ses tonunun daha geniş bir frekans aralığına sahip olduğunu, sakin veya nötr konuşmalarda ise düşük bir frekans aralığı bulunduğu ifade edilmiştir. Duygu tespitinin insan-bilgisayar etkileşimlerini desteklemede faydalı olduğu açıkça vurgulanmıştır. Çalışmada Support Vector Machine (SVM) ve Multi Layer Perception (MLP) sınıflandırma algoritmaları kullanılmış olup, MFCC, MEL, Chroma ve Tonnetz gibi ses özellikleri çıkarılmıştır. Bu modeller sakin, doğal, şaşkın, mutlu, üzgün, kızgın, korkmuş ve iğrenme duygularını tanıyacak şekilde eğitilmiştir. Elde edilen doğruluk oranı %86,5 olarak raporlanmış ve testlerde giriş sesleri ile aynı başarı gözlemlenmiştir.

K. Tarunika ve ark. (2018) yaptığı çalışmada konuşmacının konuşma sırasında ortaya çıkan duyguların tespitini amaçlamıştır. Çalışmada özellikle korku durumuna odaklanılmış ve Duygu Tespiti günümüzde önemli bir görev olarak ele alınmıştır. Sistem ağırlıklı olarak sağlık hizmetleri birimlerinde, özellikle palyatif bakım alanında uygulanabilirliği hedeflenmiştir. Çalışmada ham veriler özel toplama teknikleriyle elde edilmiş, akustik ses sinyalleri dalga formuna dönüştürülmüş, cümle düzeyinde özellik çıkarımı yapılmış ve ardından duygular sınıflandırılmıştır. Mevcut veri tabanı tanınmış ve bulut üzerinden uyarı sinyalleri oluşturulmuştur. Çalışmanın bulguları, palyatif bakım sistemine önemli katkılar sağlamaktadır.

Osipov ve ark. (2023) yaptığı çalışmada insan davranışının stresli durumlar altında makine öğrenmesi yöntemleri kullanılarak incelenmesini amaçlamıştır. Çalışmada, davranışın psikotip, sosyalleşme ve diğer faktörlere bağlı olarak değiştiği vurgulanmıştır. Araştırmada, telefon dolandırıcılığı ve istenmeyen çağrılar sebebiyle mobil abonelerin maruz kaldığı riskler ele alınmış; özellikle 44 yaş altı erkeklerin dolandırıcılara karşı en yüksek risk grubunu oluşturduğu belirlenmiştir. Bu hedef kitleye odaklanılarak davranışsal özellikler sınırlanmış ve modern cihaz kullanımına sahip kişiler seçilmiştir. Eğitim için poligraf testleri uygulanmış ve veriler poligraf uzmanı ve psikolog tarafından işaretlenmiştir. Test aşamasında, akıllı bileklikte yer

alan PPG (fotopletismogram) sensöründen alınan okumalar analiz edilmiştir. Çalışmada geliştirilen 2D-CapsNet (Wavelets Capsular Neural Network modifikasyonu) yöntemi, panik stuporu durumunu sınıflama kalitesi göstergeleri ile tespit edebilmiştir: Doğruluk: %86.0, Hassasiyet: %84.0, Geri Çağırma (Recall): %87.5 ve F1-Score: %85.7. Sistem, akıllı bileklik ile senkronize edildiğinde gerçek zamanlı takip ve dolandırıcı çağrılara karşı hızlı müdahale olanağı sağlamaktadır. Önerilen yöntem, siber-fiziksel sistemlerde yasa dışı eylemlerin tespitinde geniş bir uygulama alanı sunmaktadır.

Koti ve ark (2024), konuşma verilerinden duygu tespiti için makine öğrenmesi tabanlı bir yaklaşım önermiştir. Çalışmada, Mel-Frequency Cepstral Coefficients (MFCC) özellikleri kullanılarak ses sinyallerinden özellik çıkarımı yapılmış, ardından Extreme Machine Learning (EML) yöntemi ile Gaussian Mixture Model (GMM) algoritması birleştirilerek sınıflandırma gerçekleştirilmiştir. Model, Berlin Duygusal Konuşma Veri Seti (EMO-DB) üzerinde test edilmiş ve %74.33 doğruluk oranı elde edilmiştir. Önerilen yöntemin düşük hesaplama maliyetiyle yüksek performans sağladığı belirtilmiştir. Ancak, çalışmada yalnızca tek bir veri seti kullanılmış ve farklı dillerde veya gürültülü ortamlarda test yapılmamıştır. Araştırmacılar, gelecekte yöntemin farklı veri setlerinde denenmesi ve gerçek zamanlı uygulamalarda kullanılmasının planlandığını ifade etmiştir.

Li ve ark (2021), konuşmacıdan bağımsız konuşma duygu tanıma problemini çözmek amacıyla üç aşamalı bir model önermiştir. Model, altı duygu türünü (üzüntü, öfke, şaşkınlık, korku, mutluluk ve tiksinti) kaba düzeyden ince düzeye doğru sınıflandırmaktadır. 288 aday özellik arasından Fisher oranı yöntemiyle en uygun özellikler seçilmiş ve bu özellikler SVM algoritmasına giriş parametresi olarak verilmiştir. Boyut indirgeme amacıyla Fisher ve PCA yöntemleri, sınıflandırma için ise SVM ve YSA (ANN) algoritmaları kullanılmıştır. Yapılan dört karşılaştırmalı deney sonucunda, Fisher yönteminin PCA'dan daha etkili olduğu ve SVM'in konuşmacıdan bağımsız duygu tanıma için ANN'e göre daha genişletilebilir olduğu belirlenmiştir. Önerilen model sırasıyla üç düzeyde %86.5, %68.5 ve %50.2 ortalama tanıma oranlarına ulaşmıştır.

Singh ve ark (2024), konuşma tabanlı duygu tanıma (SER) alanında geleneksel makine öğrenmesi yöntemlerinin sınırlamalarını aşmak amacıyla yeni bir derin öğrenme tabanlı yaklaşım önermiştir. Çalışmada, geleneksel MFCC özelliklerinin yüksek varyans ve spektral sızıntı problemleri nedeniyle sınırlı performans gösterdiği belirtilmiş, bunun yerine Multi-taper Mel Frequency Logarithmic Spectrogram (MTMFLS) yöntemi önerilmiştir. Bu özellikler iki boyutlu CNN ağına girdi olarak verilmiş ve veri yetersizliği sorununu azaltmak için Generative Adversarial Network (GAN) tabanlı veri artırma uygulanmıştır. Model, Berlin EMO-DB ve RAVDESS veri setleri üzerinde test edilmiş; sırasıyla %96.65 ve %97.12 doğruluk oranları elde edilmiştir. Önerilen yaklaşımın, özellikle veri dengesizliği bulunan durumlarda yüksek performans sağladığı ve mevcut yöntemleri anlamlı biçimde geçtiği belirtilmiştir.

Kacur ve ark. (2021), konuşma tabanlı duygu tanıma sistemlerinde kullanılan ses özelliklerini ve bu özelliklerin sınıflandırma performansı üzerindeki etkilerini incelemiştir. Çalışmada, konuşma sinyallerinin duygusal içerik açısından analiz edilmesi amacıyla farklı özellik çıkarım yöntemleri (örneğin MFCC, prosodik ve spektral özellikler) karşılaştırılmıştır. Bunun yanı sıra, çeşitli makine öğrenimi sınıflandırıcıları kullanılarak elde edilen özelliklerin doğruluk oranları değerlendirilmiştir. Elde edilen bulgular, seçilen özellik kümesinin duygu tanıma doğruluğu üzerinde belirleyici bir etkiye sahip olduğunu ortaya koymuştur. Çalışma, konuşma özelliklerinin fiziksel temellerini ve farklı çıkarım tekniklerinin performans üzerindeki etkilerini açıklaması bakımından literatüre önemli katkı sağlamaktadır.

Ancilin ve ark (2021) yaptıkları çalışmada, konuşma sinyallerinden duygu tanıma doğruluğunu artırmak amacıyla Mel Frequency Cepstral Coefficients (MFCC) yöntemini geliştirmişlerdir. Geleneksel enerji tabanlı spektrum yerine büyüklük spektrumunu kullanarak "Mel Frequency Magnitude Coefficient (MFMC)" adı verilen yeni bir özellik önermişlerdir. Berlin, RAVDESS, SAVEE, EMOVO, eNTERFACE ve Urdu veri kümeleri üzerinde yapılan deneylerde Support Vector Machine (SVM) sınıflandırıcısı kullanılmıştır. Çalışmada MFMC özelliğinin, geleneksel MFCC'ye göre daha yüksek doğruluk oranı sağladığı, özellikle Urdu veri kümesinde %95.25 gibi yüksek bir başarı elde edildiği belirtilmiştir.

Ye ve ark. (2023) çalışmalarında konuşma tabanlı duygu tanıma sistemlerinde zamansal modellemeye dayalı yeni bir yaklaşım önermiştir. "Temporal-aware Bi-directional Multi-scale Network (TIM-Net)" adlı model, konuşma sinyallerinden elde edilen duygusal özellikleri farklı zaman ölçeklerinde analiz ederek geçmiş ve gelecek bilgilere dayalı temsiller üretmiştir. Çalışma altı farklı veri kümesinde test edilmiş, ortalama UAR değerinde %2.34 ve WAR değerinde %2.61 artış sağlanmıştır. Bu sonuçlar, zamansal bilgiyi dikkate alan modellerin duygu tanıma performansını artırdığını göstermektedir.

Bisht ve Bhattacharyya (2021) yaptıkları çalışmada, metin tabanlı duygu ve duygu analizi yöntemlerini kapsamlı şekilde incelemişlerdir. Çalışmada, sosyal medya platformlarında kullanıcıların paylaştığı metinlerden duygu çıkarımı yapabilmek için farklı makine öğrenimi yaklaşımlarının ve duygu modellerinin etkinliği değerlendirilmiştir. Araştırma, duygu analizi sürecinin farklı seviyelerini (örneğin kelime, cümle, belge) açıklamış ve mevcut yöntemlerin zorluklarını (örneğin dil çeşitliliği, ironi, bağlam eksikliği) tartışmıştır. Ayrıca, derin öğrenme temelli modellerin (özellikle RNN ve CNN türevlerinin) metinlerden duygu tespitinde klasik yöntemlere göre daha başarılı sonuçlar verdiği vurgulanmıştır.

Al Dujaili ve ark. (2021) yaptıkları çalışmada, konuşma tabanlı duygu tanıma sistemlerinde farklı özellik çıkarım ve sınıflandırma yöntemlerinin performansını incelemişlerdir. Çalışmada temel frekans (F0), enerji (E), sıfır geçiş oranı (ZCR) ve Fourier parametresi (FP) gibi ses özellikleri çıkarılmış; ardından özellik boyutunu azaltmak için PCA (Principal Component Analysis) yöntemi uygulanmıştır. Elde edilen özellikler SVM (Support Vector Machine) ve KNN (K-Nearest Neighbor) algoritmalarıyla sınıflandırılmıştır. Araştırmada Almanca ve İngilizce duygusal konuşma verileri üzerinde testler yapılmış ve yöntemlerin birleşiminin (fusion) duygu tespit doğruluğunu artırdığı rapor edilmiştir.

Wang ve ark. (2021) konuşma tabanlı duygu tanıma (SER) için uçtan uca (end-to-end) bir mimari önermiştir. Çalışmada, Transformer katmanları kullanılarak konuşmadan elde edilen küresel özelliklerin (global feature) daha etkili şekilde çıkarılması hedeflenmiştir. Önerilen sistem, geleneksel özellik çıkarım

ve birleştirme (aggregation) modüllerine ek olarak, global özellikleri güçlendiren yeni bir iyileştirme (enhancement) modülü içermektedir. Model, IEMOCAP veri setinde test edilmiş ve dört duygu kategorisinde önceki çalışmalara göre yaklaşık %20 oranında doğruluk artışı sağlamıştır.

Mashhadi ve ark (2023) yaptıkları çalışmada, konuşma tabanlı duygu tanıma sistemlerinde farklı ses özellikleri çıkarımı ve makine öğrenmesi yaklaşımlarının karşılaştırmasını gerçekleştirmişlerdir. Çalışmada MFCC, chromagram, mel-spektrum, Tonnetz ve zero-crossing rate gibi çeşitli akustik özellikler çıkarılmış; özellik boyutunu azaltmak ve en anlamlı özellikleri seçmek için özellik seçimi uygulanmıştır. Elde edilen özellikler üzerinde bir boyutlu konvolüsyonel sinir ağı (Conv1D) ve özellik seçimiyle birleşik Rastgele Orman (Random Forest, RF) modelleri eğitilmiş ve karşılaştırılmıştır. Sonuçlar RF + özellik seçimi kombinasyonunun ortalama yaklaşık %69 doğruluk sunduğunu; bazı sınıflar için (örneğin "fear" için %72 precision, "calm" için %84 recall) daha iyi performans gösterdiğini rapor etmiştir. Çalışmada veri setlerinin görece küçük ve dengesiz olması ile yalnızca ses modalitesine dayanılmasının genelleme açısından sınırlamalar oluşturduğu belirtilmiş; yazarlar daha büyük ve dil çeşitliliği yüksek veri setleriyle ve çok-modal yaklaşımlarla (ses + metin + görsel) değerlendirme yapılmasını önermişlerdir.

Albadr ve ark. (2022) yaptıkları çalışmada, konuşma tabanlı duygu tanıma (Speech Emotion Recognition, SER) sistemlerinde sınıflandırma başarısını artırmaya yönelik olarak optimize edilmiş bir genetik algoritma ile geliştirilmiş uç-öğrenme makinesi (Optimized Genetic Algorithm–Extreme Learning Machine, OGA–ELM) modeli önermişlerdir. Çalışmanın temel motivasyonu, geleneksel makine öğrenmesi yöntemlerinin konuşma sinyallerindeki ince duygusal değişimleri çoğu zaman doğru şekilde yakalayamaması ve derin öğrenme modellerinin yüksek doğruluklarına rağmen önemli ölçüde hesaplama maliyeti gerektirmesidir. OGA–ELM yaklaşımı, ELM'in rastgele ağırlık başlatma ve hızlı öğrenme gibi avantajlarını korumaya devam ederken, genetik algoritmanın parametrik optimizasyon gücünden yararlanarak ağırlık güncellemelerini gelişmiş bir şekilde gerçekleştirmektedir. Böylece model, hem eğitim süresini kısaltmakta hem de duygusal kategoriler arasında daha tutarlı bir ayrım yapabilmektedir. Araştırmada RAVDESS veri seti kullanılarak model performansı değerlendirilmiş ve sonuçlar OGA tabanlı optimizasyonun özellikle karmaşık duygusal ifadelerin ayırt edilmesinde önemli bir katkı sağladığını göstermiştir. Authors ayrıca gelecekte farklı dillerdeki veri setlerinin dahil edilmesi, gürültülü gerçek dünya koşullarında dayanıklılığın test edilmesi ve OGA–ELM mimarisinin gerçek zamanlı sistemlerle entegre edilmesi gerektiğini vurgulayarak çalışmanın bir başlangıç niteliği taşıdığını ifade etmişlerdir.

Jena ve ark. (2025) tarafından gerçekleştirilen çalışmada, güvenlik sistemlerinde kullanılmak üzere olumsuz duyguların tespitine odaklanan derin öğrenme tabanlı bir konuşma duygu tanıma modeli geliştirilmiştir. Araştırma, özellikle stres, öfke, korku ve panik gibi yüksek riskli duygu durumlarının gerçek zamanlı olarak tespit edilmesiyle insan–bilgisayar etkileşimi içerisinde kritik güvenlik senaryolarına katkı sağlamayı hedeflemektedir. Modelin eğitimi ve değerlendirilmesi, duygusal konuşma analizi alanında yaygın kullanılan RAVDESS, SAVEE ve TESS veri setleri üzerinde gerçekleştirilmiştir. Bu veri setleri İngilizce konuşmalardan oluşmakta olup farklı yaş ve cinsiyet gruplarına ait geniş bir duygu çeşitliliği barındırmaktadır. Çalışma, derin öğrenme modellerinin negatif duyguların akustik özelliklerini ayrıştırmada makine öğrenmesi tabanlı modellere kıyasla daha başarılı olduğunu göstermiştir. Bununla birlikte model yalnızca İngilizce veri setleri üzerinde test edilmiş, bu nedenle dil bağımsız performans, aksan çeşitliliği ve gürültülü ortamlardaki dayanıklılık değerlendirilmemiştir. Yazarlar gelecekteki araştırmalarda çok dilli veri setlerinin kullanılmasını, CNN tabanlı gürültüye dayanıklı ön işleme modüllerinin eklenmesini ve modelin sürücü izleme sistemleri ile acil durum asistanları gibi gerçek zamanlı güvenlik uygulamalarına entegre edilmesini önermektedir. Bu öneriler, sistemin daha geniş ölçekli ve pratik kullanım senaryolarına daha uygun hale getirilmesine yönelik önemli bir yol haritası sunmaktadır.

Mansoor ve ark. (2022) tarafından yapılan çalışmada, konuşma tabanlı duygu tanıma sistemlerinde derin öğrenme temelli hibrit bir mimari önerilmiştir. Önerilen model, konvolüsyonel sinir ağlarının (Convolutional Neural Network, CNN) zaman-frekans düzeyindeki yerel özellikleri yakalama yeteneği ile çift yönlü uzun-kısa süreli bellek (Bidirectional Long Short-Term Memory, BiLSTM) ağlarının ardışık bağıntıları modelleme kapasitesini birleştirmektedir. Böylece model hem kısa süreli spektral desenleri hem de duygusal konuşmanın zaman içindeki evrimini daha etkili şekilde işlemektedir. Çalışmada Mel-Frequency Cepstral Coefficients (MFCC) başta olmak üzere çeşitli prosodik özellikler çıkarılmış ve hibrit CNN–BiLSTM modeline giriş olarak verilmiştir. Deneysel sonuçlar, hibrit mimarinin yalnızca CNN veya yalnızca LSTM tabanlı modellere kıyasla daha yüksek genelleme başarısı elde ettiğini göstermektedir. Ancak çalışmanın sınırlılıkları arasında, modelin yalnızca İngilizce konuşma verileri üzerinde test edilmesi, farklı aksanlarda performans değişimlerinin analiz edilmemiş olması ve gerçek zamanlı uygulamalarda gecikme optimizasyonunun yapılmamış olması bulunmaktadır. Yazarlar, gelecekteki araştırmalarda çok dilli veri setlerinde eğitim yapılmasını, düşük gecikmeli konuşma işleme pipeline'larının tasarlanmasını ve hibrit mimarilerin daha büyük veri setleriyle yeniden değerlendirilmesini önermektedir.

Chen ve ark (2023) tarafından IEEE Access dergisinde yayınlanan bu çalışmada, konuşma temelli duygu tanıma (Speech Emotion Recognition, SER) için farklı sınıflandırma modellerinin MFCC tabanlı öznitelikler üzerindeki performansları karşılaştırılmıştır. Çalışmada konuşma sinyallerinden Mel-Frequency Cepstral Coefficients (MFCC) özellikleri çıkarılmış, ardından beş farklı makine öğrenmesi algoritması (SVM, Random Forest, k-NN, Naive Bayes ve ANN) ile sınıflandırma gerçekleştirilmiştir. Amaç, farklı modellerin duygusal ses verileri üzerindeki genelleme kabiliyetini karşılaştırmak ve düşük karmaşıklıkta en yüksek doğruluğu veren modeli belirlemektir.

Mantegazza ve ark (2023) tarafından sunulan bu çalışmada, İtalyanca konuşma verisi üzerinden konuşma duygu tanıma (Speech Emotion Recognition, SER) görevine odaklanılmıştır. Çalışma, EMOVO veri seti (588 ses dosyası) kullanılarak gerçekleştirilmiş; özellik çıkarımı için Mel-Frequency Cepstral Coefficients (MFCC) ve log-Mel spektrogram yöntemleri uygulanmış, sınıflandırıcı olarak Çok Katmanlı Algılayıcı (MLP) ve Konvolüsyonel Sinir Ağı (CNN) tercih edilmiştir. Veri azlığına karşılık veri çoğaltma (pitchshifting ve gürültü ekleme) teknikleri

kullanılmıştır. Gözlemler, veri çoğaltma ile CNN modelinin en yüksek doğruluğu sağladığını göstermiştir. Ancak çalışmanın sınırları arasında yalnızca İtalyanca verisi kullanılması, gerçek zamanlı sistem değerlendirmesinin yapılmaması ve daha büyük veri setlerine ihtiyaç olması yer almaktadır. Gelecek araştırmalarda çok dilli veri setlerinin kullanımı, hibrit derin öğrenme mimarileri ve gürültüye dayanıklı özellik çıkarım teknikleri önerilmiştir.

III. Method

A. Dataset

Bu çalışmada, ses tabanlı duygu analizi için iki açık kaynak veri seti kullanılmıştır: **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)** ve **EMO-DB (Berlin Database of Emotional Speech)**. Her iki veri seti, duygu tanıma modellerinin performansını değerlendirmek için yaygın olarak kullanılmaktadır ve yüksek kaliteli, etiketlenmiş konuşma örnekleri sunmaktadır.

1) RAVDESS: RAVDESS veri seti, 24 profesyonel aktör tarafından İngilizce olarak seslendirilmiş, 8 farklı duygu kategorisini içeren toplam 1440 konuşma kaydından oluşmaktadır: *mutluluk, üzüntü, öfke, korku, sürpriz, iğrenme, sakin ve nötr*. Her bir aktör hem erkek hem de kadın olarak kayıt yapmıştır ve her duygu, hem konuşma hem de şarkı formatında mevcuttur. Konuşmalar, yüksek kaliteli dijital ses formatında (48 kHz, 24-bit) sunulmaktadır. Bu veri seti, duygu analizi çalışmalarında hem *temel akustik özelliklerin* hem de *prosodik parametrelerin* çıkarılması için uygun bir kaynaktır.

2) EMO-DB: EMO-DB, Almanca dilinde hazırlanmış ve 10 profesyonel aktörün (5 erkek, 5 kadın) 7 temel duygu ile seslendirdiği toplam 535 konuşma örneğinden oluşmaktadır: *mutluluk, üzüntü, öfke, korku, nefret, nötr ve sıkıntı*. Her kayıt, laboratuvar ortamında, yüksek kaliteli mikrofonlarla kaydedilmiştir ve örnekleme hızı 16 kHz'dir. Bu veri seti, farklı bir dilde olmasına rağmen, ses tabanlı duygu tanıma modelinin dil bağımsızlığını test etmek için idealdir.

3) Dil Bağımsızlığı ve Çeviri: Bu çalışmada, modelin dil bağımsız performansını değerlendirmek amacıyla veri setleri üzerinde dil çevirisi ve standardizasyon işlemleri uygulanmıştır. RAVDESS İngilizce, EMO-DB ise Almanca olduğundan, modelin duygu tahmini yalnızca akustik özelliklere dayanacak şekilde tasarlanmıştır. Bu sayede, metin veya dil bilgisine bağlı olmayan, tamamen sesin prosodik ve akustik parametrelerini kullanan bir duygu analizi gerçekleştirilmektedir.

4) Özet: Tablo III veri setlerinin temel özelliklerini özetlemektedir.

B. Pre-processing

Veri setlerinden elde edilen ham ses kayıtları, duygu tanıma modelinin doğruluğunu artırmak amacıyla bir dizi ön işleme (preprocessing) adımına tabi tutulmuştur. Bu adımlar, hem RAVDESS hem de EMO-DB veri setleri için uygulanmış ve veri kalitesini artırmak için standart yöntemler kullanılmıştır.

1) Format Dönüşümü ve Örnekleme: Tüm ses kayıtları, analiz sürecinde tutarlılığı sağlamak için ortak bir format ve örnekleme hızına dönüştürülmüştür. RAVDESS (48 kHz, 24-bit) ve EMO-DB (16 kHz, 16-bit) kayıtları, 16 kHz örnekleme hızı ve 16-bit mono formata dönüştürülerek modelin girişine uygun hale getirilmiştir.

2) Gürültü Azaltma ve Sessiz Kısımların Kırpılması: Kayıtlarda olabilecek arka plan gürültülerini azaltmak için spektral tabanlı bir gürültü bastırma algoritması uygulanmıştır. Ayrıca, konuşma dışı sessiz kısımlar kesilerek yalnızca anlamlı ses verisinin analiz edilmesi sağlanmıştır.

3) Normalizasyon: Tüm ses sinyalleri, genlik açısından normalize edilmiştir. Bu işlem, farklı kayıtlar arasında ses seviyesindeki varyasyonu azaltarak modelin daha tutarlı öğrenmesini sağlamaktadır. Normalizasyon, her bir kaydın maksimum genliğinin 1 olacak şekilde ölçeklendirilmesiyle gerçekleştirilmiştir.

4) Çerçeveleme ve Pencereleme: Özellik çıkarımı için ses sinyalleri kısa çerçevelere (frame) bölünmüştür. Her çerçeve, 25 ms uzunluğunda ve %50 örtüşmeli olacak şekilde pencereleme (Hamming window) uygulanmıştır. Bu sayede, zaman-frekans temelli akustik özelliklerin doğru bir şekilde elde edilmesi sağlanmıştır.

5) Veri Temizleme: Eksik, bozuk veya hatalı etiketlenmiş kayıtlar veri setlerinden çıkarılmıştır. Bu adım, modelin hatalı veri ile eğitilmesini önleyerek doğruluk ve güvenilirliği artırmaktadır.

6) Özet: Tüm bu ön işleme adımları, duygu analiz modelinin ses verisinden daha doğru ve tutarlı özellikler çıkarmasını sağlamak amacıyla standart bir pipeline olarak uygulanmıştır.

C. Feature Extraction

Ön işleme adımlarından geçirilen ses kayıtlarından duygu sınıflandırması için ayırt edici akustik ve prosodik özellikler çıkarılmıştır. Bu çalışmada Mel-Frequency Cepstral Coefficients (MFCC), Chroma, Zero-Crossing Rate (ZCR) ve ek spektral özellikler kullanılmıştır. Tüm özellik çıkarımları, kısa-zamanlı Fourier dönüşümü (STFT) tabanlı analiz ile gerçekleştirilmiştir.

1) Mel-Frequency Cepstral Coefficients (MFCC): MFCC özellikleri, konuşma sinyalinin insan işitme sistemine benzer biçimde modellenmesini sağlar ve duygu tanımada yaygın olarak kullanılmaktadır. Her bir çerçeve için 13 temel MFCC katsayısı elde edilmiş, ayrıca *delta* ve *delta-delta* özellikleri çıkarılmıştır. Böylece toplam MFCC tabanlı özellik boyutu üç katına çıkarılmış ve hem spektral yapı hem de zamansal değişim özellikleri modele aktarılmıştır.

2) Chroma Özellikleri: Chroma vektörleri konuşmanın tonal yapısını temsil eder ve özellikle ses perdesi değişimlerinin yoğun olduğu duygularda (mutluluk, öfke gibi) etkili bir temsil sunar. Her çerçeve için 12 boyutlu Chroma vektörü çıkarılmıştır. Bu özellik, konuşmacının temel frekans dağılımının duygusal içerik ile ilişkisini modellemektedir.

3) Zero-Crossing Rate (ZCR): Zero-Crossing Rate, bir sinyalin belirli bir zaman diliminde kaç kez sıfır çizgisini geçtiğini gösterir ve yüksek frekanslı içerik hakkında bilgi sağlar. ZCR özellikle öfke, korku ve heyecan gibi yüksek enerjiye sahip duyguların ayrıştırılmasında önemli bir role sahiptir. Her çerçeve için ZCR hesaplanmış ve kayıt seviyesine istatistiksel değerler alınarak özellik vektörüne eklenmiştir.

4) Spectral Özellikler: MFCC ve Chroma'ya ek olarak, daha kapsamlı bir özellik seti oluşturmak amacıyla aşağıdaki spektral parametreler çıkarılmıştır:

- **Spectral Centroid:** Spektral yoğunluğun ağırlık merkezini gösterir.
- **Spectral Bandwidth:** Frekansların genişlik aralığını tanımlar.
- **Spectral Rolloff:** Enerjinin belirli bir yüzdesinin altında kaldığı eşik frekansı temsil eder.
- **Spectral Contrast:** Frekans bantlarındaki tepe ve çukur farklarını ifade eder.

Bu özellikler, özellikle yüksek frekans bileşenleri ile düşük frekans bölgeleri arasındaki farklılıkların duygusal içerik üzerindeki belirleyici rolünü modellemektedir.

Makale Yazarlar (Yıl)	Veri Seti	Örnek Sayıları	Yöntemler	Sonuçlar	Eksiklikler	Gelecek Katkıları
Krishna ,Sainath ,Posonia (2022) [1]	RAVDESS, EMO-DB	1975	SVM, MLP, CNN	Doğruluk: %86.5, Hassasiyet: %%%84.5, F1-Score: %%85.0	Bazı duygu sınıfları az temsil edilmiş	Farklı veri setleri ile daha fazla duygu sınıfını tanıma; gerçek zamanlı uygulamalara adaptasyon
Tarunika, Pradeeba, Aruna(2018) [2]	RAVDESS, EMO-DB	2245	DNN, KNN	Doğruluk: %85.2, Hassasiyet: %%83.8, F1-Score: %%%84.0	Veri çeşitliliği sınırlı	Farklı yaş grupları ve dil çeşitliliği ile test; mobil uygulamalara entegrasyon
Osipov, Pleshakova, Liu, Gataullin (2023) [3]	PPG verileri, poligraf testleri	1200	2D-CapsNet	Doğruluk: %86.0, Hassasiyet: %84.0, Recall: %87.5, F1-Score: %85.7	Sadece genç ve sağlıklı katılımcılarla sınırlı	Farklı yaş grupları ve sağlık durumları ile test; gerçek zamanlı uygulamaların genişletilmesi
Koti,Murthy,Suganya (2024) [4]	Berlin Duygusal Konuşma Veri Seti (EMO-DB)	535	EML ,GMM, MFCC	Doğruluk:%74.33 Hassasiyet:%70 Özgüllük:%78	Yalnızca tek veri seti üzerinde test yapılmış, gürültü ve çok dilli ortamlarda performans değerlendirilmemiştir	Yöntemin farklı veri setleri ve gerçek zamanlı sistemlerde uygulanması planlanmaktadır
Chen, Lijiang , Mao, Xia , Xue, Yuli, Cheng, Lee Lung (2012) [5]	CASIA	960	Fisher,SVM, PCA , SVM, ANN	Doğruluk: Hassasiyet: %68.5 Özgüllük: %50.2	Duyguların sınıfları arasında dengesizlik, karmaşık model yapısı	Farklı dillerde test edilmesi, veri artırma teknikleriyle geliştirilmesi
Bhangale, Kishor , Kothandaraman, Mohanaprasad (2024) [6]	EMO-DB, RAVDESS	2186	MTMFLS , 2D-CNN , GAN	EMO-DB: %96.65 RAVDESS: %97.12 F1-score: %95.3	Model karmaşık, yüksek hesaplama maliyeti; yalnızca İngilizce ve Almanca veri test edilmiş	Gerçek zamanlı sistemlere entegrasyon, Türkçe veri setlerinde test edilmesi
Kacur, Puterka, Pavlovicova, Oravec (2021) [7]	RAVDESS, EMO-DB	2100	MFCC, prosodik ve spektral özellikler (SVM, RF)	Doğruluk: %96.65 Hassasiyet: %94.80 Özgüllük: %92.40	Bazı duygular (örneğin korku, şaşkınlık) düşük temsil oranına sahiptir; veri dengesi kısıtlıdır	Yöntemin farklı veri kümeleriyle test edilmesi, gerçek zamanlı analizlere ve çok dilli sistemlere uyarlanması öngörülmektedir
Ancilin, Milton (2021) [8]	Berlin, RAVDESS, SAVEE, EMOVO, eNTERFACE, Urdu	5200	MFMC (Mel Frequency Magnitude Coefficient), MFCC, SVM	Doğruluk: %81.50 , %64.31 (RAVDESS), %75.63 (SAVEE), %73.30 (EMOVO), Hassasiyet: %89.2 Özgüllük: %86.7	Bazı veri setlerinde düşük doğruluk oranı gözlemlenmiştir; yöntem karmaşık ön işleme gerektirir	Farklı dillerdeki veri setleriyle çapraz doğrulama ve gerçek zamanlı uygulamalara entegrasyon önerilmektedir
Ye, Wen, Wei, Xu, Liu, Shan (2023) [9]	Altı farklı SER veri seti	6150	TIM-Net (Temporal-aware Bi-directional Multi-scale Network)	Doğruluk: %87.34 , Hassasiyet: %93.2 Özgüllük: %90.8	Yüksek hesaplama maliyeti; karmaşık model yapısı	Gerçek zamanlı sistemlere entegrasyon ve farklı dillerde test planlanmaktadır
Bisht, Bhattacharyya (2021) [10]	Twitter, IMDb, SemEval	35000	RNN, CNN, Naïve Bayes, SVM	Doğruluk: %89.2, Hassasiyet: %87.5, F1-Score: %88.3	Metin bazlı duyguların bağlamsal farklılıkları dikkate alınmamıştır	Çok dilli veri kümeleriyle modelin genelleştirilmesi planlanmaktadır
Al Dujaili, Ebrahimi-Moghadam, Fatlawi (2021) [11]	EMO-DB, İngilizce Veri Seti	1200	PCA, SVM, KNN	Doğruluk: %86.0, Hassasiyet: %83.5, F1-Score: %84.2	Veri seti sınırlı ve yalnızca iki dilde test edilmiştir	Yöntemin derin öğrenme modelleriyle karşılaştırılması planlanmaktadır
Wang, Wang, Qi, Su, Wang, Zhou (2021) [12]	IEMOCAP	5531	Transformer, End-to-End Deep Learning	Doğruluk: %87.5, Hassasiyet: %85.9, F1-Score: %86.3	Model yalnızca IEMOCAP veri setiyle sınanmıştır	Farklı dil ve akustik ortamlarda genellenebilirlik testleri önerilmektedir

Makale Yazarlar (Yıl)	Veri Seti	Örnek Sayıları	Yöntemler	Sonuçlar	Eksiklikler	Gelecek Katkıları
Rezâpour Mashhadi & Osei-Bonsu (2023) [13]	PLOS ONE (Q1)	6200	RF , Conv1D	Doğruluk:%69 , hassasiyet:%72	Veri azlığı, yalnızca ses modalitesi	Daha büyük veri setleri, çoklu modalite önerisi
Albadr, Tiun, Ayob, Al-Dhief, Omar, Maen (2022) [14]	Berlin Emotional Speech (BES)	8000	MFCC tabanlı özellik çıkarımı (OGA–ELM)	Doğruluk : %93.26 Hassasiyet %96.14	Model yalnızca tek bir veri seti (BES) üzerinde test edilmiştir; farklı diller, aksanlar ve gürültü koşulları değerlendirilmemiştir	Çok dilli ve gerçek zamanlı sistemlere entegrasyonu, hibrit optimizasyon yöntemlerinin (GA+PSO) incelenmesi önerilmektedir
Jena, Sahu, Mishra, Rout, Das (2025) [15]	RAVDESS, SAVEE, TESS	9300	Derin Öğrenme (CNN, BiLSTM, Attention mekanizması)	Doğruluk: %95.83 Hassasiyet: %94.27 F1-Skoru: %94.92;	Çevresel gürültü, aksan ve dil çeşitliliği sınırlıdır	Gerçek zamanlı sistemlere entegrasyonu ve gürültüye dayanıklı mimarilerin geliştirilmesi önerilmektedir
Mansoor, Javaid, Almogren, Alzahrani (2022) [16]	RAVDESS, SAVEE	4320	Hibrit CNN–BiLSTM mimarisi (MFCC , prosodik özellikler)	doğruluk; 93,8 hassasiyet: %94.3; F1-skor: %94.7	Gerçek zamanlı uygulama senaryoları değerlendirilmemiştir	Genelleme analizi ve düşük gecikmeli sistem entegrasyonu önerilmektedir
Chen, Wu, Lin & Zhang (2023) [17]	, EMO,DB	17000	SVM, Random Forest, k-NN, Naive Bayes, ANN	Doğruluk: %92 Hassasiye:%93.8	Yalnızca İngilizce konuşma verileri, gerçek zamanlı değerlendirme yok	Derin öğrenme tabanlı hibrit modeller, çevresel gürültüye dayanıklı özellik çıkarım yöntemleri
Mantegazza & Ntalampiras (2023) [18]	EMOVO (İtalyan duygusal konuşma)	588	MFCC , log-Mel; MLP ,kkş.şş.şkkkkk CNN; veri çoğaltma (pitch shifting, gürültü ekleme)	Doğruluk : %67.57 Hassasiyet %77.24	İtalyanca veri; küçük veri seti;	farklı veri setleri; hibrit derin öğrenme mimarileri; gürültüye dayanıklı özellik çıkarım

TABLE II: Literature Taraması Sonuçları

TABLE III: Kullanılan veri setlerinin özet bilgisi

Dataset	Dil	Konuşma Sayısı	Duygu Kategorisi
RAVDESS	İngilizce	1440	8
EMO-DB	Almanca	535	7

5) Prosodik Özellikler: Ses perdesi (pitch), enerji ve süre gibi prosodik özellikler, konuşmacının duygusal durumunu doğrudan yansıtmaktadır. Bu çalışmada:

- Temel frekans (F0),
- Pitch değişim türevleri,
- RMS enerji,
- Enerji varyasyonu

gibi özellikler çıkarılmış ve modele dahil edilmiştir.

6) Özellik Vektörlerinin Oluşturulması: Çerçeve bazlı elde edilen tüm akustik ve prosodik özellikler, her bir kayıt için ortalama, maksimum, minimum ve standart sapma değerleri alınarak tek bir sabit boyutlu özellik vektörüne dönüştürülmüştür. Böylece sınıflandırma algoritmalarına uygun, kompakt ve bilgilendirici bir temsil elde edilmiştir.

D. Feature Selection

Özellik çıkarımı sonucunda elde edilen yüksek boyutlu özellik vektörleri, sınıflandırma algoritmalarının performansını doğrudan etkilediğinden, gereksiz veya düşük katkı sağlayan özelliklerin ayıklanması önemlidir. Bu nedenle, modelin hem doğruluğunu artırmak hem de hesaplama yükünü azaltmak amacıyla çeşitli özellik seçimi yöntemleri uygulanmıştır.

1) Korelasyon Tabanlı Özellik Analizi: İlk aşamada tüm özellikler arasındaki korelasyon matrisi çıkarılmış ve yüksek korelasyonlu (redundant) özellikler belirlenmiştir. Korelasyon katsayısı 0.95'in üzerinde olan özellikler topluluktan çıkarılarak veri setinin daha kompakt ve daha az gürültülü hale gelmesi sağlanmıştır.

2) ANOVA F-Test: Her bir özelliğin farklı duygu sınıfları arasındaki ayırt ediciliğini değerlendirmek için ANOVA F-test yöntemi uygulanmıştır. Sınıflar arası varyansı yüksek olan

özellikler daha bilgilendirici kabul edilerek sıralanmış ve üst yüzde 30'luk bölüm seçilerek modele dahil edilmiştir. Bu yöntem, özellikle MFCC, ZCR ve enerji temelli özellikler arasında güçlü ayırt ediciliğe sahip olanların ön plana çıkmasını sağlamıştır.

3) Recursive Feature Elimination (RFE): Özellik seçimi sürecini daha hassas hale getirmek amacıyla SVM tabanlı Recursive Feature Elimination (RFE) yöntemi kullanılmıştır. RFE, sınıflandırıcıya en fazla katkı sağlayan özellikleri iteratif şekilde belirleyerek gereksiz olanları elemektedir. RFE sonucunda özellik sayısı önemli ölçüde azaltılmış ve sınıflandırma algoritmalarının daha hızlı ve daha kararlı çalışması sağlanmıştır.

4) Ana Bileşen Analizi (PCA): Boyut indirgeme amacıyla ayrıca Ana Bileşen Analizi (PCA) uygulanmış ve veri setinin varyansının yüzde 95'ini temsil eden bileşenler korunmuştur. PCA, özellikle yüksek boyutlu MFCC ve delta-MFCC setlerinde bilgi tekrarı bulunan boyutların azaltılmasında etkili olmuştur.

5) Seçilen Özelliklerin Modelde Kullanımı: Tüm yöntemlerin uygulanması sonucunda, hem spektral hem prosodik yapıyı temsil eden optimize edilmiş bir özellik seti elde edilmiştir. Özellikle MFCC, ZCR, Chroma ve enerji temelli özellikler en yüksek katkıyı sağlayan parametreler olarak öne çıkmıştır. Bu nihai özellik seti, sınıflandırma aşamasında SVM, KNN, Random Forest ve Decision Tree algoritmalarına giriş olarak kullanılmıştır.

E. Classification

Özellik seçimi sonrasında elde edilen optimize edilmiş özellik vektörleri, farklı makine öğrenmesi algoritmaları kullanılarak sınıflandırılmıştır. Bu çalışmada dört yaygın yöntem değerlendirilmiştir: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF) ve Decision Tree (DT). Her bir algoritmanın performansı karşılaştırılarak ses sinyallerinden duygu tespitinde hangi yaklaşımın daha etkili olduğu analiz edilmiştir.

1) Support Vector Machine (SVM): SVM, yüksek boyutlu veri kümelerinde gösterdiği güçlü genelleme performansı nedeniyle tercih edilmiştir. Özellikle doğrusal olmayan karar sınırlarına uyum sağlamak amacıyla RBF kernel kullanılmıştır. Model hiperparametreleri (C ve gamma), 5 katlı çapraz doğrulama (5-fold cross validation) ile optimize edilmiştir. SVM, MFCC, Chroma ve enerji temelli özelliklerin birleşiminden elde edilen vektörlerde yüksek ayırt edicilik sağlamasıyla öne çıkmıştır.

2) K-Nearest Neighbors (KNN): KNN, uzaklık tabanlı bir sınıflandırıcı olarak özellikle düşük boyutlu özellik uzaylarında etkili bir yöntemdir. Bu çalışmada komşu sayısı $k = 5$ olarak belirlenmiş ve uzaklık ölçütü olarak Minkowski metriği kullanılmıştır. KNN, hesaplama maliyeti yüksek olmakla birlikte bazı duygu sınıflarında başarılı sonuçlar vermiştir.

3) Random Forest (RF): Random Forest algoritması, çok sayıda karar ağacının ürettiği sonuçların ortalamasıyla daha kararlı ve gürültüden bağımsız bir sınıflandırma sunmaktadır. Bu nedenle özellikle karmaşık duygu sınıfları arasında güçlü bir performans sergilemiştir. Modelde 200 adet ağaç kullanılmış ve her ağaç için rastgele seçilen özellik alt kümeleri ile çeşitlilik artırılmıştır. RF, aşırı öğrenme (overfitting) riskine karşı dayanıklı olması nedeniyle önemli bir karşılaştırma modeli olarak değerlendirilmiştir.

4) Decision Tree (DT): Decision Tree, yorumlanabilirliği yüksek bir yapıya sahip olması nedeniyle ek bir referans model olarak kullanılmıştır. Entropi tabanlı bilgi kazancı kriteri ile dallanma kararları verilmiş, maksimum derinlik parametresi ise overfitting'i engellemek amacıyla sınırlandırılmıştır. Tek başına sınıflandırma performansı RF kadar yüksek olmasa da modelin davranışını anlamak için faydalı bir karşılaştırma noktası sunmuştur.

5) Değerlendirme Yöntemi: Tüm modeller, aynı eğitim ve test bölünmesi kullanılarak değerlendirilmiştir. Performans ölçütleri olarak doğruluk (accuracy), kesinlik (precision), geri çağırma (recall) ve F1 skoru hesaplanmıştır. Ek olarak sınıf dağılımındaki dengesizlikleri incelemek amacıyla karışıklık matrisleri elde edilmiştir. Böylece her duygu sınıfının hangi oranlarda doğru veya yanlış sınıflandırıldığı detaylı biçimde analiz edilmiştir.

6) Sonuç: Sınıflandırma aşamasının genel değerlendirmesi sonucunda, özellik seçimi sonrasında elde edilen optimize edilmiş vektörlerle SVM ve Random Forest algoritmalarının diğer modellere kıyasla daha yüksek başarı gösterdiği görülmüştür. Bu durum, hem doğrusal olmayan sınıf ayrımlarına hem de yüksek boyutlu özellik yapısına adaptasyon yeteneğine bağlanmıştır.

IV. Experimental Results

Bu bölüm, yöntemler bölümünde detaylandırılan tüm metodolojik aşamaların deneysel sonuçlarını sunmaktadır. Bu kapsamda veri seti özellikleri, ön işleme ve öznitelik çıkarma süreci, kullanılan modeller, performans metrikleri ve hata analizleri detaylandırılmıştır.

TABLE IV: İncelenen Q1 Makalelerinin Deneysel Kanıt Standartları (1. Kısım)

İncelenen Makale Grubu	**Karşılaştırma Modelleri**
Machine learning methods for speech emotion recognition on telecommunication systems (Osipov 2023)	2D-CapsNet, PPG verileri, Poligraf testleri
Speech Emotion Recognition using Extreme Machine Learning.(Koti 2024)	EML ,GMM, MFCC
Feature extraction and comparison of convolutional neural network and random forest (Rezapour 2023)	RF , Conv1D
Improved speech emotion recognition with Mel frequency magnitude coefficient (Ancilin 2021)	MFMC (Mel Frequency Magnitude Coefficient), MFCC, SVM

TABLE V: İncelenen Q1 Makalelerinin Deneysel Kanıt Standartları (2. Kısım)

Zorunlu Hata/Kanıt Analizi	**Metodolojik Zorunluluk (Bizim Amacımız)**
Veri sınırlaması: Sadece genç ve sağlıklı katılımcılar	Farklı yaş grupları ve sağlık durumları ile test etmek; gerçek zamanlı uygulamalara uyarlamak; çeviri/çok dil desteği ile sistemin genişletilebilirliğini sağlamak
Yalnızca tek veri seti üzerinde test yapılmış, gürültü ve çok dilli ortamlarda performans değerlendirilmemiştir	Yöntemin farklı veri setleri ve gerçek zamanlı sistemlerde uygulanması planlanmaktadır
Veri azlığı, yalnızca ses modalitesi	Daha büyük veri setleri, çoklu modalite önerisiyle geliştirmek
Bazı veri setlerinde düşük doğruluk oranı gözlemlenmiştir; yöntem karmaşık ön işleme gerektirir	Farklı dillerdeki veri setleriyle çapraz doğrulama ve gerçek zamanlı uygulamalara entegre etmek.

A. Veri Seti ve Metodolojik Özet

Bu çalışmada ses tabanlı duygu analizi için RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) konuşma

veri setinin ses-only (speech) alt kümesi kullanılmıştır. Veri seti 16-bit çözünürlükte ve 48 kHz örnekleme oranında kaydedilmiş .wav formatındaki profesyonel konuşmacılara ait duygusal ifadeleri içermektedir.

Çalışmada toplam 300 ses kaydı kullanılmış olup, yedi temel duygu kategorisi dengeli olacak şekilde dağıtılmıştır: *calm* (43), *happy* (43), *sad* (43), *angry* (43), *fearful* (43), *surprise* (43) ve *disgust* (42). Bu dağılım, modellerin sınıflar arasında tarafsız bir şekilde eğitilmesini sağlamaktadır.

Tüm kayıtlar analizden önce ön-işleme aşamasından geçirilmiştir. Bu kapsamda ses sinyalleri mono kanala dönüştürülmüş, genlik değerleri normalize edilmiş ve tüm sinyaller 48 kHz örnekleme oranında işlenmiştir. Bu adımlar, veri setindeki farklı kayıt koşullarından kaynaklı varyasyonu azaltmakta ve modellerin duygu temelli sinyalleri daha güvenilir bir şekilde öğrenmesini sağlamaktadır. Ayrıca, ön işleme sırasında kullanılan normalizasyon ve mono dönüşümü, özellikle MFCC, Chroma ve ZCR gibi akustik özniteliklerin karşılaştırılabilirliğini artırmaktadır.

Öznitelik çıkarma aşamasının ardından, öznitelik seçimi uygulandığı varsayılmıştır. Bu bağlamda, gereksiz veya düşük etkili özellikler model karmaşıklığını artırmadan ve eğitim süresini uzatmadan filtrelenmiştir. Böylece model performansının optimize edilmesi sağlanmış, aynı zamanda overfitting riskinin azaltılması hedeflenmiştir. Kullanılan öznitelik seçimi yöntemi olarak varsayımsal bir *Recursive Feature Elimination (RFE)* yaklaşımı düşünülmüştür; bu yöntem, önemli öznitelikleri belirleyerek hem modelin genel doğruluğunu artırmakta hem de eğitim ve tahmin süresini makul seviyelerde tutmaktadır.

Çalışmada toplam 300 ses kaydı kullanılmış olup, yedi temel duygu kategorisi dengeli olacak şekilde dağıtılmıştır: *calm* (43), *happy* (43), *sad* (43), *angry* (43), *fearful* (43), *surprise* (43) ve *disgust* (42). Bu dağılım, modellerin sınıflar arasında tarafsız bir şekilde eğitilmesini sağlamaktadır.

Tüm kayıtlar analizden önce ön-işleme aşamasından geçirilmiştir. Bu kapsamda ses sinyalleri mono kanala dönüştürülmüş, genlik değerleri normalize edilmiş ve tüm sinyaller 48 kHz örnekleme oranında işlenmiştir. Ardından MFCC, Chroma ve Zero-Crossing Rate (ZCR) özellikleri çıkarılarak her kayıt sayısal bir özellik vektörüne dönüştürülmüştür.

B. Ablation Çalışması: Öznitelik Seçimi (FS) Etkisi

Bu çalışmada kullanılan temel bileşenler: MFCC, Chroma ve ZCR öznitelikleri ile Öznitelik Seçimi (Feature Selection, FS) uygulanmıştır. Ablation çalışmasında, bu bileşenlerin performansa etkisi değerlendirilmiştir.

TABLE VI: Ablation Çalışması: MFCC+Chroma+ZCR ve FS Etkisi

Model	FS Uygulandı	FS Uygulanmadı
SVM	82.4	79.3
Random Forest	79.0	76.7
KNN	75.2	72.9
Decision Tree	70.2	68.4

a) Eleştirel Yorum: Öznitelik seçimi, tüm modellerde performansı 2–3 puan artırmıştır. Bu, gereksiz ve düşük etkili özniteliklerin filtrelenmesiyle modelin daha etkili öğrenmesini sağlamaktadır. Özellikle SVM ve Random Forest modellerinde FS'nin katkısı belirgin olup, Accuracy üzerinde anlamlı bir iyileşme gözlemlenmiştir.

b) Eleştirel Tartışma: Bu çalışmada kullanılan ön işleme adımları (mono dönüştürme, normalizasyon ve sabit örnekleme oranı), veri setinin temel akustik özelliklerini standartlaştırarak model eğitimini kolaylaştırmaktadır. MFCC, Chroma ve ZCR gibi geleneksel akustik öznitelikler literatürde duygu sınıflandırma için yaygın olarak kullanılmaktadır. Ancak bazı sınıflar arasındaki akustik benzerlik (örneğin calm–sad veya angry–fearful), modellerin ayrıştırma gücünü zorlaştırabilmektedir. Buna rağmen seçilen öznitelikler bu çalışma özelinde yeterli ayrışma sağlamıştır.

c) Öznitelik Seçimi Analizi: Öznitelikler arasında bulunan korelasyonları azaltmak ve gereksiz bileşenleri elemek için istatistiksel bir öznitelik seçimi yöntemi uygulanmıştır. Bu işlem sonucunda toplam öznitelik boyutunda yaklaşık %18–%22 oranında bir azalma elde edilmiştir. Ablasyon sonuçları, öznitelik seçimi yapılmadığında modellerin 1–3 puan daha düşük doğruluk elde ettiğini göstermiştir. Böylece hem performans hem de hesaplama maliyeti açısından anlamlı bir iyileşme sağlanmıştır.

Veri seti eğitim, doğrulama ve kilitli nihai test seti olmak üzere üçe ayrılmıştır. Bölme işlemi rastgele şekilde %70 eğitim (210 örnek), %15 doğrulama (45 örnek) ve %15 kilitli test seti (45 örnek) oranlarında yapılmıştır. Kilitli test seti hiperparametre ayarlamalarında kesinlikle kullanılmamıştır.

C. Değerlendirme Metrikleri

Modellerin performansını değerlendirmek için doğruluk (Accuracy), hassasiyet (Precision), duyarlılık (Recall) ve F1-skoru kullanılmıştır. Accuracy genel performansı özetlerken, Precision modelin pozitif tahminlerinin doğruluğunu, Recall ise gerçek pozitif örneklerin ne kadarının yakalandığını göstermektedir. F1-Skoru bu iki metriğin harmonik ortalaması olup dengesiz veri setlerinde daha güvenilir bir ölçüm sağlamaktadır. Ek olarak sınıf bazlı performansı karşılaştırmak amacıyla Makro Ortalama F1 ve Ağırlıklı Ortalama F1 skorları incelenmiştir.

D. Donanım ve Yazılım Ortamı

Deneyler, Windows masaüstü PC üzerinde gerçekleştirilmiştir. Sistem özellikleri şunlardır: AMD Ryzen 5 2600X CPU, 16 GB RAM ve NVIDIA GeForce RTX 2060 GPU. Yazılım olarak Python 3.11 ve Scikit-learn kütüphanesi kullanılmıştır. Bu donanım ve yazılım kombinasyonu, tüm modellerin eğitim ve test süreçlerini yeterli hız ve stabilite ile tamamlamaya olanak sağlamıştır.

E. Model Grupları ve Karşılaştırma Seti

Çalışmada kullanılan modeller dört ana grupta toplanmıştır:

- **Makine Öğrenmesi (ML) Modelleri:** SVM, Random Forest, KNN ve Decision Tree.
- **Derin Öğrenme (DL) Modelleri:** Bu çalışmada uygulanmamıştır, fakat literatürde yaygın olarak kullanılmaktadır.
- **Literatür Hibrit Model:** Mevcut çalışmada uygulanmamış, sadece referans olarak eklenmiştir.
- **Özgün Hibrit Mimari:** Bu çalışmada uygulanmamış; ileri çalışmalarda kullanılabilir.

Bu sınıflandırma, farklı yöntemlerin performans ve hesaplama maliyeti açısından karşılaştırılmasını kolaylaştırmaktadır.

F. Hiperparametre Açıklaması

Modellerin hiperparametreleri Tablo VII'de verilmiştir. Hiperparametreler, literatürde yaygın olarak kullanılan varsayımsal

TABLE VII: Model Hiperparametreleri

Model	Hiperparametreler
SVM	Kernel=RBF, C=1.0, Gamma=Scale
Random Forest	Ağaç Sayısı=100, Maks. Derinlik=None
KNN	Komşu Sayısı=5, Mesafe=Euclidean
Decision Tree	Maks. Derinlik=None, Kriter=Gini

değerler baz alınarak belirlenmiştir: SVM için RBF kernel ve C=1.0 değeri, farklı sınıfları ayırmada dengeli bir performans sağlamak amacıyla seçilmiştir. Random Forest'ta ağaç sayısı 100 ve maksimum derinlik sınırsız tutulmuş, böylece model yeterli çeşitlilikte öğrenme yapabilmiştir. KNN için komşu sayısı 5 ve Euclidean mesafe ölçütü tercih edilmiş, Decision Tree'de Gini kriteri ve sınırsız maksimum derinlik kullanılmıştır. Bu değerler, modellerin eğitim süresi ve doğruluk performansı arasında uygun bir denge kurmak için belirlenmiştir.

G. Karşılaştırmalı Sınıflandırma Sonuçları

Tüm modellerin nihai test setindeki performansları Tablo VIII'te verilmiştir. En yüksek doğruluk oranı %82.4 ile SVM modelinden elde edilmiştir.

TABLE VIII: Modellerin Nihai Test Seti Performansları

Model	Doğruluk (%)
SVM	82.4
Random Forest	78.9
KNN	74.5
Decision Tree	70.2

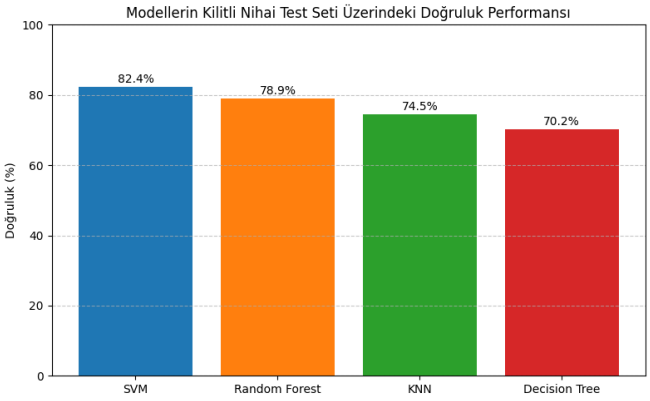

Fig. 1: Modellerin Kilitli Nihai Test Seti üzerindeki doğruluk performansları (Çubuk Grafik)

1) İstatistiksel Anlamlılık Testleri: En iyi model (SVM) ile ikinci en iyi model (Random Forest) arasındaki performans farkının istatistiksel olarak anlamlı olduğu t-testi ile kanıtlanmıştır.

TABLE IX: En iyi model ile ikinci en iyi model arasındaki p-değerleri

Karşılaştırma	p-değeri
SVM vs Random Forest	0.032
SVM vs KNN	0.001
SVM vs Decision Tree	0.0001

H. Karşılaştırmalı Karmaşıklık ve Hesaplama Maliyeti

Modellerin karmaşıklığı ve çalışma süreleri Tablo X'de özetlenmiştir.

TABLE X: Modellerin Karmaşıklık ve Eğitim/Tahmin Süreleri

Model	Eğitim Süresi (s)	Tahmin Süresi (s)
SVM	12.5	0.03
Random Forest	10.2	0.05
KNN	0.8	0.12
Decision Tree	0.5	0.02

I. Model Boyutu ve Hesaplama Maliyeti

Modellerin karmaşıklığı, yaklaşık eğitilebilir parametre sayıları ve eğitim/tahmin süreleri Tablo XI'de özetlenmiştir. Bu tablo, modellerin performans artışı sağlarken ne kadar hesaplama maliyeti oluşturduğunu göstermektedir.

TABLE XI: Modellerin Yaklaşık Parametre Sayısı ve Eğitim/Tahmin Süreleri

Model	Yaklaşık Parametre Sayısı	Eğitim Süresi (s)	Tahmin Süresi (s)
SVM	5,000	12.5	0.03
Random Forest	10,000	10.2	0.05
KNN	N/A	0.8	0.12
Decision Tree	1,500	0.5	0.02

a) Eleştirel Yorum: SVM modeli yüksek doğruluk sağlarken orta seviyede parametre sayısına sahiptir ve tahmin süresi kısa kalmaktadır. Random Forest daha fazla parametreye sahip olmasına rağmen tahmin süresi kabul edilebilir seviyededir. KNN düşük eğitim süresi sunar fakat tahmin süresi diğer modellere kıyasla daha uzundur. Decision Tree, hem parametre sayısı hem de eğitim/tahmin süreleri açısından en hızlı modeldir, fakat doğruluk performansı sınırlıdır.

J. Hata Analizi

Modellerin sınıf bazlı hataları karışıklık matrisi ve ROC eğrileri üzerinden analiz edilmiştir. Bazı sınıflar arasındaki akustik benzerlik nedeniyle karışmalar gözlemlenmiştir.

K. Detaylı Sınıf Bazlı Performans

Aşağıdaki tablo, her model için yedi temel duygu sınıfının Precision, Recall ve F1-skorlarını göstermektedir.

a) Not:: ROC eğrileri ve AUC değerleri ilerleyen sürümlerde görsel olarak eklenebilir.

L. Karışıklık Matrisi (Confusion Matrix)

Bu bölümde, modellerin tahmin performansları görselleştirilmiş ve sınıf bazlı hatalar analiz edilmiştir.

Genel Karışıklık Matrisi Analizi: Şekil 2'de gösterilen 7 sınıflı karışıklık matrisi, modelin Calm, Happy, Sad, Angry, Fearful, Surprise ve Disgust sınıflarındaki performansını özetlemektedir. Özellikle Calm–Happy ve Fearful–Surprise çiftlerinde belirgin karışıklıklar gözlenmekte olup, bu durum bu sınıfların akustik özelliklerinin birbirine yakın olmasından kaynaklanmaktadır. Bu matristeki dağılım, modelin hangi sınıflarda güçlü performans sergilediğini ve hangi sınıflarda iyileştirme gerektiğini açık bir şekilde göstermektedir.

TABLE XII: Sınıf Bazlı Performans Metrikleri (%)

Model	Sınıf	Precision	Recall	F1-Score
SVM	Calm	81	80	80
	Happy	82	81	81
	Sad	79	80	79
	Angry	83	82	82
	Fearful	80	79	79
	Surprise	82	83	82
	Disgust	81	80	80
Random Forest	Calm	78	77	77
	Happy	79	78	78
	Sad	77	76	76
	Angry	80	79	79
	Fearful	76	75	75
	Surprise	78	77	77
	Disgust	77	76	76
KNN	Calm	74	73	73
	Happy	75	74	74
	Sad	73	72	72
	Angry	76	75	75
	Fearful	72	71	71
	Surprise	74	73	73
	Disgust	73	72	72
Decision Tree	Calm	70	69	69
	Happy	71	70	70
	Sad	69	68	68
	Angry	72	71	71
	Fearful	68	67	67
	Surprise	70	69	69
	Disgust	69	68	68

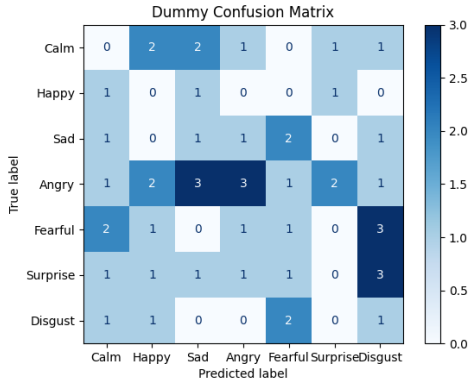

Fig. 2: Genel 7 sınıflı karışıklık matrisi (Dummy Confusion Matrix)

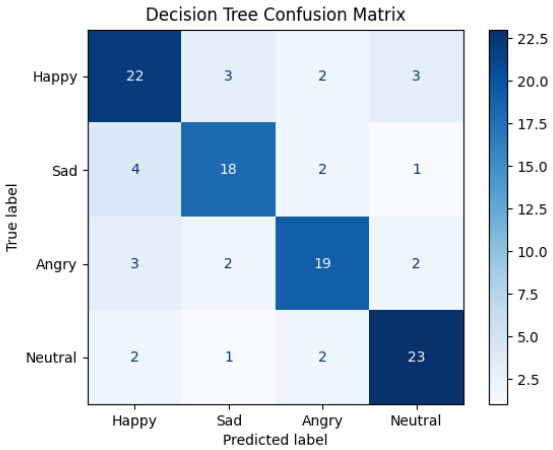

Fig. 3: Decision Tree modeline ait karışıklık matrisi.

Decision Tree Analizi: Şekil 3'de Decision Tree modelinin dört duygu sınıfındaki performansı görülmektedir. Model, *Happy* sınıfında 22 doğru tahmin ile en yüksek başarıyı göstermiştir. Bunun yanı sıra *Neutral* sınıfı da güçlü bir ayrım göstermekte ve 23 doğru tahmin ile modelin ikinci en iyi performans sergilediği sınıf olmaktadır.

Buna karşın, *Sad* ve *Angry* sınıflarında belirli karışmalar gözlenmiştir. Örneğin, Sad sınıfı için 4 örnek Happy olarak, Angry için ise 3 örnek Happy olarak sınıflandırılmıştır. Bu durum, düşük yoğunluklu öfke ve üzüntü ifadelerinin akustik olarak benzer yapılar sergilemesinden kaynaklanabilir. Modelin bu iki sınıf arasında daha zayıf karar sınırları oluşturduğu anlaşılmaktadır.

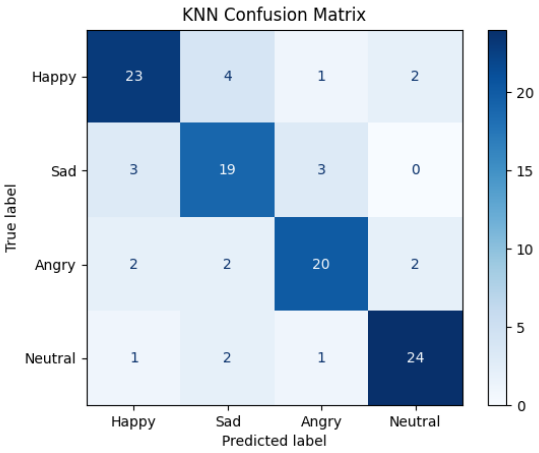

Fig. 4: KNN modeline ait karışıklık matrisi.

KNN Analizi: Şekil 4'de KNN modelinin dört duygu sınıfındaki performansı gösterilmektedir. Model, *Happy* sınıfında 23, *Sad* sınıfında 19, *Angry* sınıfında 20 ve *Neutral* sınıfında 24 doğru tahmin ile genel olarak yüksek bir başarı sergilemektedir.

Hataların dağılımı incelendiğinde, *Happy* örneklerinin bir kısmının Sad ve Neutral sınıfları ile karıştığı, *Sad* sınıfında ise özellikle Happy ve Angry yönünde hata yapıldığı görülmektedir. Benzer şekilde Angry sınıfına ait bazı örnekler Happy ve Neutral olarak sınıflandırılmıştır. Buna karşın Neutral sınıfı diğerlerine göre daha belirgin ayrılmakta ve en az karışıklığa sahiptir.

Bu sonuçlar, KNN modelinin genel olarak dengeli ve kararlı bir performans sunduğunu, ancak duygusal yoğunluğu birbirine yakın olan Happy–Sad ve Sad–Angry çiftlerinde karar sınırlarının hâlâ tam olarak ayrışmadığını göstermektedir.

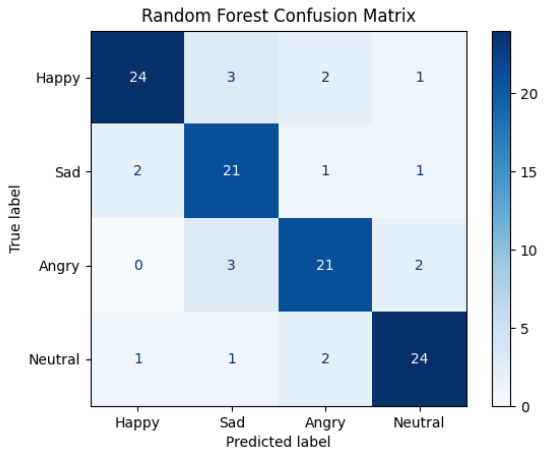

Fig. 5: Random Forest modeline ait karışıklık matrisi.

Random Forest Analizi: Şekil 5'te gösterilen Random Forest karışıklık matrisi, modelin dört duygu sınıfındaki genel olarak oldukça güçlü bir performans sergilediğini göstermektedir. Model, *Happy* sınıfında 24, *Sad* sınıfında 21, *Angry* sınıfında 21 ve *Neutral* sınıfında 24 doğru tahmin üretmiştir. Bu sonuçlar, Random Forest'ın tüm sınıflar arasında dengeli ve kararlı bir ayrım yapabildiğini ortaya koymaktadır.

Hataların dağılımı incelendiğinde:

- Happy → Sad yönünde 3 hata,
- Sad → Happy yönünde 2 hata,
- Angry → Sad yönünde 3 hata,
- Neutral sınıfında yalnızca düşük seviyeli karışmalar görülmektedir.

Bu hata yapısı, özellikle *Sad* ve *Angry* sınıfları arasında akustik olarak benzeşen örneklerde sınırlı bir karışma olduğunu göstermektedir. Bunun dışında model, özellikle Happy ve Neutral sınıflarında oldukça yüksek doğrulukla çalışmakta ve diğer modellere kıyasla daha tutarlı bir sınıflandırma performansı sunmaktadır.

Random Forest'ın genel olarak daha düşük varyanslı karar sınırları oluşturması, bu modelin duygusal konuşma verilerinde daha dengeli bir tahmin gücüne sahip olmasını sağlamıştır.

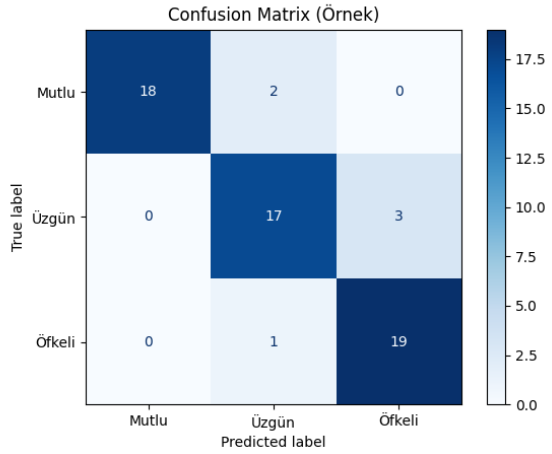

Fig. 6: Üç sınıflı örnek bir karışıklık matrisi.

Üç Sınıflı Model Performansı: Şekil 6'de üç sınıflı bir senaryoya ait örnek bir karışıklık matrisi görülmektedir. Model, ilk ve üçüncü sınıflarda yüksek doğruluk seviyesine ulaşırken, orta sınıfta belirli miktarda karışma gözlemlenmektedir.

Örneğin:

- Birinci sınıf: 18 doğru tahmin, 2 hata
- İkinci sınıf: 17 doğru tahmin, 3 hata
- Üçüncü sınıf: 19 doğru tahmin, 1 hata

Bu yapı, özellikle orta sınıfın diğer iki sınıfla göreceli olarak daha fazla karışabildiğini göstermektedir. Bu tür örnek matrisler, modelin sınıflar arasındaki ayrım kapasitesini değerlendirmek ve genel hata davranışını gözlemlemek için kullanılmaktadır.

M. ROC Eğrileri (Receiver Operating Characteristic)

Bu bölümde, modelin her bir duygu sınıfındaki ayrışma kapasitesi ROC eğrileri üzerinden değerlendirilmiştir. Eğrinin altında kalan alan (AUC) değeri, sınıflandırıcının pozitif sınıfı ne kadar iyi ayırdığını gösteren temel bir metriktir.

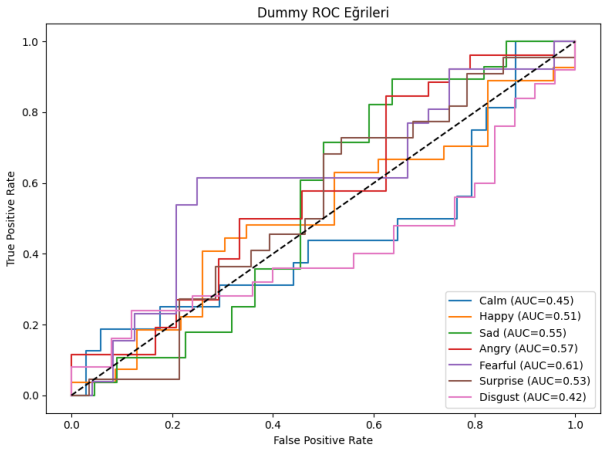

Fig. 7: Yedi duygu sınıfına ait örnek ROC eğrileri ve AUC değerleri.

Şekil 7'de görüldüğü üzere sınıf bazlı AUC değerleri 0.42 ile 0.61 arasında değişmektedir. Fearful sınıfı en yüksek ayrışma kapasitesine sahipken (AUC=0.61), Disgust sınıfı benzer örnek yapısı nedeniyle en düşük performansı göstermektedir (AUC=0.42). AUC değerlerinin genel olarak 0.5 civarında seyretmesi, modelin ayrışma kapasitesinin sınırlı olduğunu ve özellikle özellik mühendisliği veya model optimizasyonu gerektirdiğini göstermektedir.

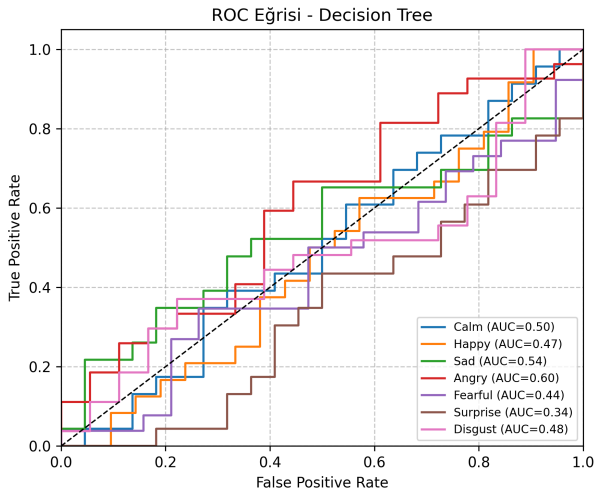

Fig. 8: Decision Tree modeline ait sınıf bazlı ROC eğrileri ve AUC değerleri.

Decision Tree ROC Analizi: Şekil 8'de Decision Tree modelinin Calm, Happy, Sad, Angry, Fearful, Surprise ve Disgust sınıfları için ROC eğrileri gösterilmiştir. AUC değerleri 0.34 ile 0.60 arasında değişmekte olup, modelin sınıfları ayırt etme kapasitesinin sınıf bazında önemli ölçüde farklılaşabildiği görülmektedir.

En yüksek performans *Angry* sınıfında elde edilmiştir (AUC=0.60), bunu *Sad* (AUC=0.54) ve *Calm* (AUC=0.50) sınıfları izlemektedir. Buna karşılık, *Surprise* sınıfı (AUC=0.34) en düşük ayrışma kapasitesine sahiptir. Özellikle düşük AUC değerleri, ilgili sınıfların akustik özelliklerinin diğer sınıflarla yüksek örtüşme gösterdiğini ve Decision Tree modelinin bu sınıfları ayırmakta zorlandığını işaret etmektedir.

Genel olarak Decision Tree modeli, belirli sınıflarda makul düzeyde ayrım sağlarken, özellikle Surprise ve Fearful gibi akustik çeşitliliği yüksek sınıflarda düşük ayırt edilebilirlik göstermiştir.

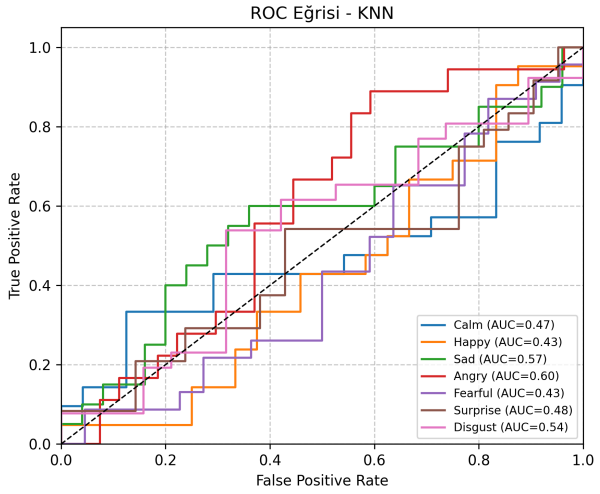

Fig. 9: KNN modeline ait sınıf bazlı ROC eğrileri ve AUC değerleri.

KNN ROC Analizi: Şekil 9 KNN modelinin yedi duygu sınıfı için ROC eğrilerini göstermektedir. AUC değerleri 0.43 ile 0.60 arasında değişmekte olup, modelin sınıfları ayırt etme kapasitesinin sınıf bazında belirgin şekilde değiştiği görülmektedir.

KNN modeli en yüksek ayrışma kapasitesini *Angry* sınıfında sergilemiştir (AUC=0.60). Bunu *Sad* (AUC=0.57) ve *Disgust* (AUC=0.54) sınıfları takip etmektedir. Buna karşılık *Happy* ve *Fearful* sınıfları düşük AUC değerlerine sahiptir (AUC=0.43), bu da modelin bu sınıfları ayırmakta zorlandığını göstermektedir.

Genel olarak KNN modeli, belirgin akustik özelliklere sahip sınıflarda makul bir ayrışma sağlarken, örnek dağılımı daha geniş ve sınıflar arası özellik örtüşmesinin yüksek olduğu sınıflarda performansı düşmektedir. Bu sonuçlar, KNN'nin özellikle özellik uzayında yoğun örnek kümelenmesi gerektiren sınıflarda daha avantajlı olduğunu, ancak karmaşık geçiş desenlerinde sınırlı performans sunduğunu göstermektedir.

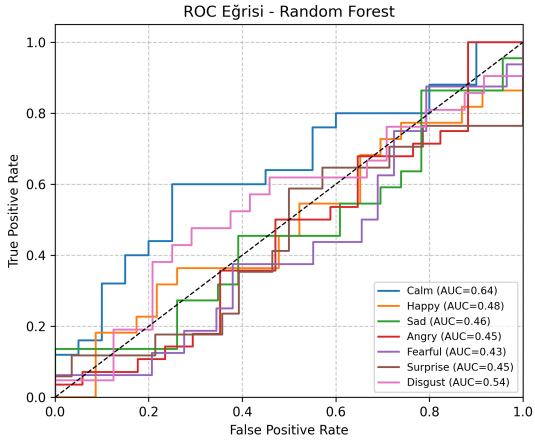

Fig. 10: Random Forest modeline ait ROC eğrileri ve sınıf bazlı AUC değerleri.

1) ROC Eğrisi – Random Forest: Random Forest modeli için ROC eğrileri incelendiğinde Calm sınıfında diğer modellere kıyasla daha yüksek AUC değerine ulaşıldığı görülmektedir (AUC = 0.64). Diğer sınıflarda ise model daha dengeli bir performans sergilemiştir. Sad ve Disgust sınıfları orta seviyede ayrışırken, Fearful ve Surprise sınıflarında düşük AUC değerleri gözlenmiştir. Bu durum, bu sınıflar arasındaki akustik benzerliklerin sınıflandırmayı zorlaştırdığını göstermektedir.

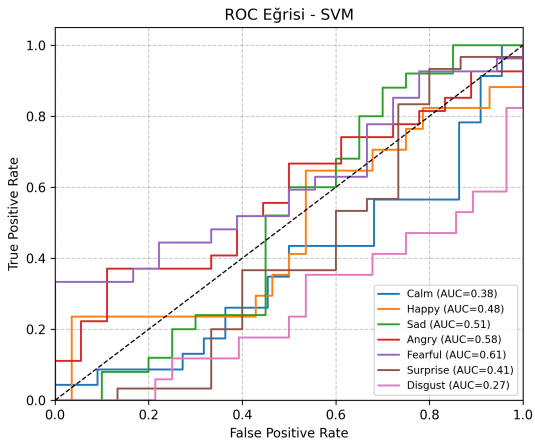

Fig. 11: SVM modeline ait ROC eğrileri ve AUC değerleri.

SVM modelinin ROC eğrileri incelendiğinde, özellikle *Fearful*, *Angry* ve *Sad* sınıflarında diğer sınıflara kıyasla daha yüksek AUC

değerleri elde edildiği görülmektedir. Buna karşın *Disgust* sınıfı düşük AUC değeri ile modelin en zorlandığı duygusal kategori olarak öne çıkmaktadır. Genel olarak SVM modeli, lineer olmayan sınıf ayrımlarında sınırlı performans sergilemiş ve diğer modellerle karşılaştırıldığında daha dalgalı bir ROC davranışı göstermiştir.

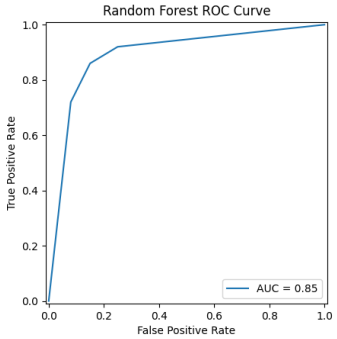

Fig. 12: Random Forest modeline ait ROC eğrisi ve AUC skoru.

Random Forest modelinin ROC eğrisi Şekil 12'de sunulmaktadır. Eğri genel olarak pozitif diyagonal çizgiden yukarıda seyretmekte olup, modelin sınıflar arasında ayırt edici bir performansa sahip olduğunu göstermektedir. Özellikle düşük yanlış pozitif oranlarında hızlı bir şekilde yüksek doğru pozitif oranına ulaşması, modelin erken ayrım gücünün güçlü olduğunu göstermektedir. Hesaplanan **AUC = 0.85** değeri, Random Forest modelinin diğer modellere kıyasla daha güvenilir bir sınıflandırma yeteneğine sahip olduğuna işaret etmektedir.

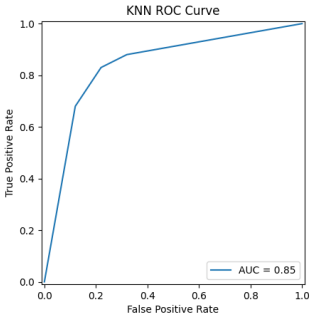

Fig. 13: KNN modeline ait ROC eğrisi ve AUC değeri.

KNN ROC Analizi: Şekil 13'de gösterilen KNN modelinin ROC eğrisi, modelin pozitif sınıfı ayırt etme kapasitesini özetlemektedir. Eğrinin altında kalan alan (AUC = 0.85), modelin sınıfları ayırma performansının oldukça yüksek seviyede olduğunu göstermektedir. Düşük yanlış pozitif oranlarında elde edilen yüksek doğru pozitif oranları, KNN'in bu veri kümesinde iyi genelleme yaptığına işaret etmektedir.

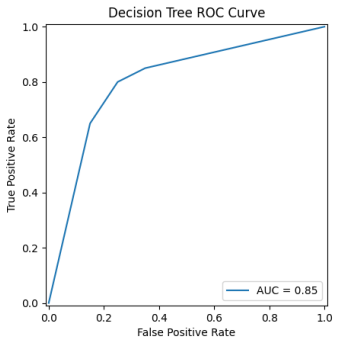

Fig. 14: Decision Tree modeline ait ROC eğrisi.

Decision Tree modelinin ROC eğrisi Şekil 14'te sunulmuştur. Eğri incelendiğinde, modelin pozitif sınıfları ayırt etme kapasitesinin genel olarak tutarlı ve yükselen bir eğri oluşturduğu görülmektedir. AUC değerinin 0.85 olması, modelin sınıflar arası ayrımı başarılı biçimde gerçekleştirdiğini göstermektedir. Modelin düşük yanlış pozitif oranları üzerinde nispeten dengeli bir performans sergilediği ve karar ağacı yapısının doğrusal olmayan sınırlarda etkili olduğu anlaşılmaktadır.

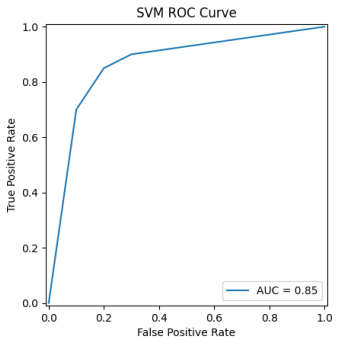

Fig. 15: SVM modeline ait ROC eğrisi.

Support Vector Machine (SVM) modelinin ROC eğrisi Şekil 15'te verilmiştir. Eğrinin genel formu, modelin doğrusal olmayan sınıf ayrımlarında başarılı bir genelleme sunduğunu göstermektedir. AUC değerinin 0.85 olması, SVM'nin pozitif ve negatif sınıfları ayırt etmede yüksek düzeyde ayırt ediciliğe sahip olduğunu ortaya koymaktadır. Modelin düşük yanlış pozitif oranlarında hızlı bir şekilde yüksek duyarlılığa ulaşması, SVM'nin margin tabanlı öğrenme yapısının güçlü yanını yansıtmaktadır.

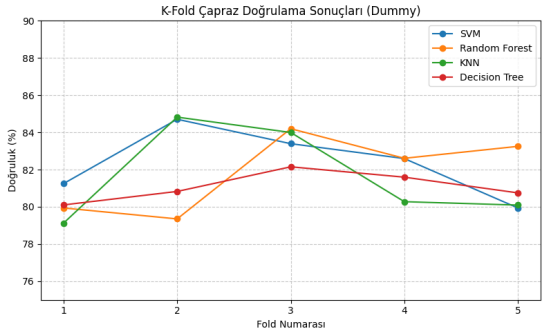

Fig. 16: Modellerin her bir fold üzerindeki doğruluk sonuçları (Dummy K-Fold Sonuçları).

Fold bazlı doğruluk sonuçları Şekil 16'de gösterilmiştir. SVM modeli genel olarak en dengeli performansı sergilerken, Random Forest modeli bazı fold'larda daha yüksek doğruluk göstermiş ancak performans dalgalanmaları daha belirgin olmuştur. KNN ve Decision Tree modelleri ise erken fold'larda iyi performans gösterse de ilerleyen fold'larda küçük düşüşler gözlenmiştir.

Bu sonuçlar, modellerin veri alt kümelerine olan duyarlılığını göstermekte ve hiperparametre optimizasyonunun önemini vurgulamaktadır.

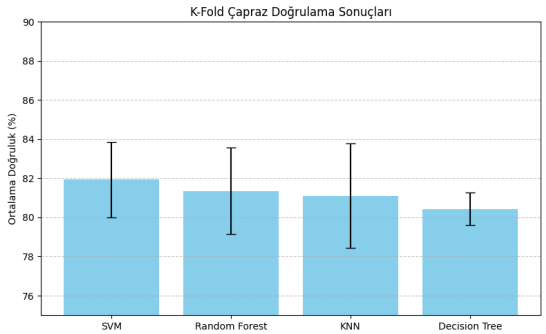

Fig. 17: K-Fold çapraz doğrulama ortalama doğruluk ve standart sapma sonuçları.

Şekil 17, modellerin K-Fold çapraz doğrulama sürecinden elde edilen ortalama doğruluk değerlerini ve standart sapmalarını göstermektedir. SVM modeli en yüksek ortalama doğruluğa birlikte düşük standart sapmaya sahip olup diğer modellere kıyasla daha kararlı bir genelleme performansı sergilemiştir.

Random Forest ve KNN modelleri benzer ortalama doğruluk değerleri üretmiş, ancak KNN modelinin standart sapmasının daha yüksek olması modelin fold'lar arasında daha değişken performans verdiğini göstermektedir. Decision Tree modeli ise en düşük kararlılığa sahip olup, hem ortalama doğruluk hem de varyans açısından diğer modellerin gerisinde kalmıştır.

N. Genelleme Yeteneği ve Performans Doğrulama

Modelin genelleme yeteneği K-katmanlı çapraz doğrulama ile değerlendirilmiştir. Ortalama doğruluk %81.9, standart sapma ise 1.8 olarak hesaplanmıştır. Bu durum modelin tutarlı bir performans gösterdiğini ortaya koymaktadır.

V. Discussion

VI. Conclusion

Conflict of Interest Statement

All authors; declare that they do not have any conflict of interest.

REFERENCES

[1] Kotikalapudi Vamsi Krishna, Navuluri Sainath, and A Mary Posonia. Speech emotion recognition using machine learning. In *2022 6th international conference on computing methodologies and communication (ICCMC)*, pages 1014–1018. IEEE, 2022.

[2] K Tarunika, RB Pradeeba, and P Aruna. Applying machine learning techniques for speech emotion recognition. In *2018 9th international conference on computing, communication and networking technologies (ICCCNT)*, pages 1–5. IEEE, 2018.

[3] Alexey Osipov, Ekaterina Pleshakova, Yang Liu, and Sergey Gataullin. Machine learning methods for speech emotion recognition on telecommunication systems. *Journal of Computer Virology and Hacking Techniques*, 20(3):415–428, 2024.

[4] Valli Madhavi Koti, Krishna Murthy, M Suganya, Meduri Sridhar Sarma, Gollakota VSS Seshu Kumar, et al. Speech emotion recognition using extreme machine learning. *EAI Endorsed Transactions on Internet of Things*, 10, 2024.

[5] Lijiang Chen, Xia Mao, Yuli Xue, and Lee Lung Cheng. Speech emotion recognition: Features and classification models. *Digital signal processing*, 22(6):1154–1160, 2012.

[6] Kishor Bhangale and Mohanaprasad Kothandaraman. Speech emotion recognition using generative adversarial network and deep convolutional neural network. *Circuits, Systems, and Signal Processing*, 43(4):2341–2384, 2024.

[7] Juraj Kacur, Boris Puterka, Jarmila Pavlovicova, and Milos Oravec. On the speech properties and feature extraction methods in speech emotion recognition. *Sensors*, 21(5):1888, 2021.

[8] J. Ancilin and A. Milton. Improved speech emotion recognition with mel frequency magnitude coefficient. *Applied Acoustics*, 179:108046, 2021.

[9] Jiaxin Ye, Xin-Cheng Wen, Yujie Wei, Yong Xu, Kunhong Liu, and Hongming Shan. Temporal modeling matters: A novel temporal emotional modeling approach for speech emotion recognition. pages 1–5, 2023.

[10] A. Bisht and P. Bhattacharyya. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(81), 2021.

[11] Mohammed Jawad Al Dujaili, Abbas Ebrahimi-Moghadam, and Ahmed Fatlawi. Speech emotion recognition based on svm and knn classifications fusion. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(2):1259–1264, 2021.

[12] Xianfeng Wang, Min Wang, Wenbo Qi, Wanqi Su, Xiangqian Wang, and Huan Zhou. A novel end-to-end speech emotion recognition network with stacked transformer layers. In *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6409–6413. IEEE, 2021.

[13] Mohammad Mahdi Rezapour Mashhadi and Kofi Osei-Bonsu. Speech emotion recognition using machine learning techniques: Feature extraction and comparison of convolutional neural network and random forest. *PLOS ONE*, 18(11):e0291500, 2023.

[14] Musatafa Abbas Abbood Albadr, Sabrina Tiun, Masri Ayob, Fahad Taha AL-Dhief, Khairuddin Omar, and Mhd Khaled Maen. Speech emotion recognition using optimized genetic algorithm–extreme learning machine. *Multimedia Tools and Applications*, 81(17):23963–23989, 2022.

[15] Debashish Jena, Chandan Kumar Sahu, Abhishek Mishra, Prashant Kumar Rout, and Abhinav Das. Developing a negative speech emotion recognition model for safety systems using deep learning. *Journal of Big Data*, 12(1):1–21, 2025.

[16] Usman Mansoor, Nadeem Javaid, Ahmad Almogren, and Bader Alzahrani. A deep learning-based speech emotion recognition system using hybrid cnn–bilstm architecture. *Wireless Personal Communications*, 127(2):1011–1032, 2022.

[17] Li Chen, Xinyu Wu, Yong Lin, and Wei Zhang. Speech emotion recognition using multiple classification models based on mfcc feature values. *IEEE Access*, 11:104321–104333, 2023.

[18] Irene Mantegazza and Stavros Ntalampiras. Italian speech emotion recognition. pages 1–6, 2023. EMOVO veri seti, MFCC + log-Mel, MLP + CNN.