T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

DOKTORA TEZİ

BEDEN DİLİNDEN ELDE EDİLEN MEKÂNSAL-ZAMANSAL VERİLER KULLANILARAK YAPAY ZEKÂ İLE DUYGU TESPİTİ

Abdulhalık OĞUZ

ARALIK-2024
BATMAN

T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

DOKTORA TEZİ

BEDEN DİLİNDEN ELDE EDİLEN MEKÂNSAL-ZAMANSAL VERİLER KULLANILARAK YAPAY ZEKÂ İLE DUYGU TESPİTİ

Abdulhalık OĞUZ

Danışman

Prof. Dr. Ömer Faruk ERTUĞRUL

Diğer Jüri Üyeleri

Prof. Dr. Fevzi HANSU Doç. Dr. Melih KUNCAN Doç. Dr. Yılmaz KAYA

Doç. Dr. Emrullah ACAR

ARALIK-2024
BATMAN

TEZ KABUL VE ONAYI

Abdulhalık OĞUZ tarafından hazırlanan "Beden Dilinden Elde Edilen Mekânsal-Zamansal Veriler Kullanılarak Yapay Zekâ ile Duygu Tespiti" adlı tez çalışması 30/12/2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Batman Üniversitesi Lisansüstü Eğitim Enstitüsü Elektrik-Elektronik Mühendisliği Ana Bilim Dalı'nda DOKTORA TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan	
Prof. Dr. Fevzi HANSU	……..
Danışman	
Prof. Dr. Ömer Faruk ERTUĞRUL	……..
Üye	
Doç. Dr. Melih KUNCAN	……..
Üye	
Doç. Dr. Yılmaz KAYA	……..
Üye	
Doç. Dr. Emrullah ACAR	……..

Yukarıdaki sonucu onaylarım.

Dr. Öğr. Üyesi Ömer Murat ÖTER

Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYANI

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını beyan eder, aksinin ortaya çıkması durumunda her türlü yasal sorumluluğu kabullendiğimi bildiririm.

ETHICAL DECLARATION

I declare that all the information in this thesis has been obtained within the framework of ethical behavior and academic rules, and that the source of any statements and information that do not belong to me in this study prepared in accordance with the thesis writing rules has been fully cited, and I declare that I accept all kinds of legal responsibility in case of any contrary situation.

Abdulhalık OĞUZ

30.12.2024

ÖZET

DOKTORA TEZİ

BEDEN DİLİNDEN ELDE EDİLEN MEKÂNSAL-ZAMANSAL VERİLER KULLANILARAK YAPAY ZEKÂ İLE DUYGU TESPİTİ

Abdulhalık OĞUZ

Batman Üniversitesi Lisansüstü Eğitim Enstitüsü

Elektrik-Elektronik Mühendisliği Ana Bilim Dalı

Danışman: Prof. Dr. Ömer Faruk ERTUĞRUL

2024, 97 Sayfa

Bu çalışma, beden hareketlerine dayalı duygu tanıma süreçlerinde mekânsal-zamansal verilerin ve çok boyutlu yaklaşımların etkinliğini kapsamlı bir şekilde incelemektedir. Kinematik ham veri seti ve video tabanlı DEMOS veri seti kullanılarak öfke, tiksinti, korku, mutluluk, nötr, üzüntü ve şaşkınlık gibi temel duyguların sınıflandırılmasına yönelik farklı yöntemlerin performansları karşılaştırılmıştır. Literatürde yüz ifadeleri ve ses tabanlı yöntemler ön planda yer alırken, bu çalışma, yüz ifadelerinin yetersiz kaldığı durumlarda beden hareketlerinden duygu tanımanın potansiyelini ortaya koymayı hedeflemiştir.

Kinematik veri analizlerinde, iskelet tabanlı ham pozisyon bilgileri hem doğrudan ham veri hem de öznitelik çıkarımı yapılarak değerlendirilmiştir. K-nearest neighbors, Random Forest, CatBoost ve XGBoost gibi makine öğrenimi algoritmalarının yanı sıra RegNetY, MobileNetV3, LSTM ve GRU gibi derin öğrenme yöntemleri test edilmiştir. Bu kapsamda, yedi duygu sınıfı için elde edilen en yüksek doğruluk oranı farklı pencereleme boyutları için %99'un üzerine kadar çıkmış ve bu durum ham kinematik sinyallerden duygu tanımanın yüksek doğrulukla mümkün olduğunu göstermiştir.

DEMOS video veri seti üzerinde yapılan çalışmalarda, altı duygu sınıfı için mekânsal ve zamansal verileri analize uygun modeller (SlowFast-R50, X3D-Medium, ResNet-3D-18 ve Attentive3D-CNN-LSTM gibi) derin öğrenme yöntemleriyle test edilmiştir. Tüm açılardan alınan video verileriyle, altı duygu sınıfı için en yüksek dengeli doğruluk oranı tüm test verisinde %60 olarak elde edilmiştir.

Sonuçlar, ham kinematik verilerin sağladığı yüksek doğruluğun çok sınıflı duygu sınıflandırma süreçlerinde kullanılabileceğini göstermiştir. Ayrıca, iskelet tabanlı video verilerinin bağlamsal zenginliğiyle birleştirildiği çok modelli yaklaşımların, duygu tanıma süreçlerini geliştirme potansiyeline işaret etmektedir. Çalışma, insan-makine etkileşimi, güvenlik, sağlık ve eğitim gibi farklı alanlarda geniş bir uygulama potansiyeli sunmaktadır. Bununla birlikte, sinyal işleme teknikleri, öznitelik çıkarımı, veri artırma ve transfer öğrenme gibi yöntemlerin, duygu tanıma süreçlerinde verimliliği artırmada etkili olabileceği vurgulanmıştır.

Kinematik ve video tabanlı veri setlerini karşılaştırmalı olarak analiz eden bu çalışma, duygu tanıma alanında farklı veri sistemlerin geliştirilmesine yönelik yenilikçi bir çerçeve sunmaktadır. Çalışma, duygu tanıma sistemlerinin geliştirilmesine yönelik bir temel oluşturmuş ve gelecekteki araştırmalar için metodolojik ve uygulamalı öneriler sunarak literatüre katkıda bulunmuştur.

Anahtar Kelimeler: Duygu Tanıma, Makine Öğrenimi, Derin Öğrenme, Spatio-Temporal Modeller, Çok Modaliteli Yaklaşımlar, Kinematik Veriler, Video Analizi, Vücut Duruşu

ABSTRACT

DOCTORAL THESIS

EMOTION DETECTION USING ARTIFICIAL INTELLIGENCE WITH SPATIOTEMPORAL DATA OBTAINED FROM BODY LANGUAGE

Abdulhalık OĞUZ

Batman University Graduate Education Institute

Department of Electrical and Electronics Engineering

Advisor: Prof. Dr. Ömer Faruk ERTUĞRUL

2024, 97 Pages

This study provides a comprehensive evaluation of the effectiveness of spatiotemporal data and multidimensional approaches in recognizing emotions through body movements. By utilizing a kinematic raw dataset and the video-based DEMOS dataset, it compares the performance of various methods for classifying fundamental emotions, including anger, disgust, fear, happiness, neutrality, sadness, and surprise. While methods based on facial expressions and voice dominate the literature, this study highlights the potential of body movement-based emotion recognition, particularly in scenarios where facial expressions are insufficient.

In the analysis of kinematic data, raw skeletal position information was assessed both as unprocessed data and after feature extraction. The study tested a range of machine learning algorithms, including K-nearest Neighbors, Random Forest, CatBoost, and XGBoost, alongside deep learning models such as RegNetY, MobileNetV3, LSTM, and GRU. For the seven emotion classes, the highest accuracy rate exceeded 99% across different windowing sizes, demonstrating that emotion recognition from raw kinematic signals is highly feasible with remarkable precision.

Experiments on the DEMOS video dataset tested spatiotemporal data for six emotion classes using deep learning methods (such as SlowFast-R50, X3D-Medium, ResNet-3D-18, and Attentive3D-CNN-LSTM).With video data captured from all angles, the highest balanced accuracy rate for the six emotion classes reached 60% across all test data.

The results show that raw kinematic data, with its high accuracy, can be effectively used in multi-class emotion classification. Additionally, combining skeleton-based video data with its contextual richness in multimodal approaches holds significant promise for improving emotion recognition. The study highlights broad application potential in fields such as human-machine interaction, security, healthcare, and education. Moreover, it emphasizes that techniques like signal processing, feature extraction, data augmentation, and transfer learning could substantially enhance the efficiency of emotion recognition processes.

This study compares kinematic and video-based datasets, presenting an innovative framework for the development of diverse data systems in emotion recognition. It establishes a solid foundation for advancing emotion recognition technologies and makes a valuable contribution to the literature by offering both methodological and practical recommendations for future research.

Keywords: Emotion Recognition, Machine Learning, Deep Learning, Spatiotemporal Models, Multimodal Approaches, Kinematic Data, Video Analysis, Body Language

ÖN SÖZ

Doktora öğrenim sürecimin her aşamasında rehberliğiyle beni yönlendiren, bilgi ve tecrübelerini paylaşarak akademik ve çalışma hayatı gelişimime büyük katkılarda bulunan, hoşgörüsünü benden hiçbir zaman esirgemeyen ve beni cesaretlendiren değerli danışman hocam Sayın Prof. Dr. Ömer Faruk Ertuğrul'a en içten teşekkürlerimi ve saygılarımı sunarım.

Tez izleme sürecinde yapıcı eleştirileri ve değerli yönlendirmeleriyle çalışmama önemli katkılarda bulunan, aynı zamanda motivasyonumu hep yüksek tutmamı sağlayan Tez İzleme Komitesi üyeleri Sayın Doç. Dr. Yılmaz Kaya ve Sayın Doç. Dr. Emrullah Acar'a gönülden teşekkür ederim.

Hayatım boyunca her zaman yanımda olan, sevgileri ve dualarıyla beni destekleyen kıymetli anne ve babama ve çok sevdiğim kardeşlerime; en zor zamanlarımda bile hep yanımda olup bana güç veren, varlığıyla kendimi her zaman şanslı hissettiğim güzel eşime ve sevgileriyle ve varlıklarıyla bu sürecin en büyük motivasyonu olan canım çocuklarıma minnettarım. Bu tezi, hayatımda yer alan tüm bu güzel insanlara ithaf ediyorum.

Abdulhalık OĞUZ

BATMAN-2024

İÇİNDEKİLER

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

SİMGELER VE KISALTMALAR

Kısaltmalar

2D	: İki Boyutlu
3D	: Üç Boyutlu
AS-LSTM	: Attention-Supervised Long Short-Term Memory
BMSNN	: Bi-Modular Sequential Neural Network
BP4D+	: BP4D+ Veri Seti
CatBoost	: Categorical Boosting
CMU	: Carnegie Mellon University- Carnegie Mellon Üniversitesi
CNN	: Convolutional Neural Network
CWT	: Continuous Wavelet Transformation
DEMOS	: Dalian Emotional Movement Open-source Set
DL	: Deep Learning- Derin Öğrenme
EEG	: Elektroensefalografi
FE	: Feature Extraction- Veriden Özellik Çıkarma
FN	: Yanlış Negatif- False Negative
FP	: Yanlış Pozitif- False Positive
GCN	: Graph Convolutional Networks
GRU	: Gated Recurrent Unit
GZSL	: Generalized zero-shot learning
JND	: Joint Neighborhood Distance
KNN	: K-Nearest Neighbors- K-en Yakın Komşu
LSTM	: Long Short-Term Memory
MBConv	: Mobile Inverted Bottleneck Layer
ML	: Machine Learning- Makine Öğrenmesi
MLLM	: Multimodal Large Language Models
MoCap	: Motion Capture
NTU RGB+D	: NTU RGB+D Veri Seti
ResNet	: Residual Network
RF	: Random Forest
RNN	: Recurrent Neural Network
SE	: Squeeze-and-Excitation
ST-CNN	: Spatio-Temporal CNN
TN	: Doğru Negatif -True Negative
TP	: Doğru Pozitif -True Positive
UnSkEm	: Unobtrusive Skeletal-based Emotion Recognition
X2D	: Xception-based 2D Model
X3D	: Xception-based 3D Model
XGBoost	: Extreme Gradient Boosting- Aşırı Gradyan Artırma

1. GİRİŞ

Duygu tanıma, bireylerin duygusal durumlarını belirleme ve analiz etme sürecidir. Duygu tanıma; metin, ses, sinyal ve görüntü gibi çeşitli veri türlerini işleyerek insanların öfke, mutluluk, üzüntü, şaşkınlık, korku, tiksinme ve nötr gibi temel duygusal durumlarını algılayabilmeyi hedefler. Bu süreç, insanların söyledikleri (metin), nasıl söyledikleri (ses) ve yüz ifadeleri ya da beden dili gibi görsel ipuçları (görüntü) üzerinden analiz edilebilir.

Duygu tanıma, insan-bilgisayar etkileşimini daha doğal ve etkili hale getirmek için kritik bir rol oynamaktadır. Özellikle sağlık alanında, bireylerin duygusal durumlarının doğru bir şekilde anlaşılması, psikolojik bozuklukların erken teşhis edilmesine ve tedavi süreçlerinin iyileştirilmesine olanak tanır. Örneğin, depresyon ve anksiyete gibi ruhsal bozuklukların erken belirtilerini algılayabilen bir duygu tanıma sistemi, sağlık profesyonellerine daha hızlı ve etkili müdahale fırsatı sunabilir (Moghe vd., 2024). Bu tür sistemler, bireylerin duygusal durumlarını sürekli olarak izleyerek, gerektiğinde destek veya müdahale önerileri sağlayabilir. Bu sayede, sağlık hizmetleri daha kişiselleştirilmiş ve etkili hale gelebilir.

Duygu analizi, insan-bilgisayar etkileşimini geliştirme, hizmetleri kişiselleştirme, erken teşhis yapma ve verimliliği artırma gibi birçok avantaja sahiptir (Hassan vd., 2021). Cihazların kullanıcılarla daha insancıl bir şekilde etkileşim kurmasını sağlayarak, kullanıcı deneyimini iyileştirebilir. Kullanıcıların duygusal durumlarına göre özelleştirilmiş hizmetler sunarak, bireysel ihtiyaçlara daha uygun çözümler sağlayabilir. Özellikle sağlık alanında, psikolojik rahatsızlıkların belirlenmesinde etkili bir araç olarak kullanılabilir. Duygu analizi sistemleri, veri gizliliği ve etik konularında önemli endişeleri beraberinde getirebilir. Özellikle kişisel verilerin işlenmesi, mahremiyet ihlallerine yol açarak kullanıcıların gizlilik haklarının ihlali ve etik sorunlar yaratma potansiyeline sahiptir (Taddeo ve Floridi, 2019). Ayrıca, farklı kültürel bağlamlarda duyguların ifade biçimleri değişebileceğinden analizlerde hata payı artabilir. Mevcut teknolojiler, duyguların karmaşıklığını anlamada henüz yeterince etkili değildir. Bunun yanı sıra, sistemlerin bireylerin duygularını yanlış sınıflandırması ciddi sonuçlara yol açabilir ve yanlış kararların alınmasına neden olabilir.

İnsanların duygusal durumlarını algılamak ve anlamak, özellikle yapay zekâ ve makine öğrenimi teknolojilerinin gelişimiyle birlikte, çeşitli uygulama alanlarında büyük bir önem kazanmıştır. Yapay zekâ, duygu tanıma süreçlerinde metin, ses, görüntü ve video

gibi farklı veri türlerini işleyerek bu analizi mümkün kılmaktadır. Bu veri türlerinin her biri, duygu analizi için kendine özgü yöntemler ve avantajlar sunar. Metin tabanlı analiz, yazılı içeriklerin duygu etiketleme teknikleriyle değerlendirilmesi üzerine odaklanır. Örneğin, sosyal medya gönderilerindeki duygusal tonun belirlenmesi bu yöntemin yaygın bir uygulamasıdır. Ses tabanlı analiz ise ses tonundaki değişiklikler, konuşma ritmi ve diğer akustik özellikler aracılığıyla bireylerin duygusal durumlarını tespit etmeye odaklanır. Görüntü tabanlı analiz, yüz ifadeleri, beden hareketleri ve mikro ifadeler gibi görsel ipuçları üzerinden duyguların algılanmasını sağlar.

Video tabanlı analiz, her bir yöntemi içerebilmekte ve tüm bu yöntemleri bir araya getirerek daha kapsamlı bir yaklaşım sunabilmektedir. Video tabanlı analiz, yüz ifadeleri, beden hareketleri, jest ve mimikler gibi görsel unsurların yanı sıra ses tonundaki değişiklikleri ve konuşma ritmini değerlendirerek duyguların daha doğru bir şekilde tespit edilmesine olanak tanıyabilir. Özellikle spatial (mekânsal, uzamsal) ve temporal (zamansal) modellerle desteklenen bu yaklaşım, duyguların dinamik ve bağlamsal bir şekilde analiz edilmesini sağlar. Farklı veri türlerinin bir arada ya da ayrı ayrı kullanılması, duygu tanımanın doğruluğunu ve etkinliğini artırabilmektedir.

Bu çalışmada beden hareketlerine dayalı duygu tanıma alanında ham kinematik veriler ve kinematik hareketlerin görselleştirildiği ve ayırt edici yüz görüntüsü içermeyen video tabanlı görsel içerikler gibi iki farklı veri türü incelenerek, bu alandaki yaklaşımlara yeni bir perspektif sunulmaktadır. Çalışmanın temel amacı, farklı veri tiplerinin duygu tanıma süreçlerindeki etkilerini karşılaştırmalı olarak analiz etmek ve detaylı analizler içeren her bir yaklaşımın bu süreçlere sağlayabileceği katkıları değerlendirmektir. Beden dilinin, yüz ifadelerinin yetersiz kaldığı durumlarda bile duygusal durumların doğru bir şekilde anlaşılmasına olanak sağladığına dair güçlü bir sav öne sürülmektedir.

Tezde, öncelikli olarak ham kinematik veriler üzerinden analizler gerçekleştirilmiştir. Bu aşamada, insan vücudunun iskelet temelli pozisyon ve hareket bilgileri kullanılarak duygusal durumların tanımlanabileceği gösterilmiştir. Bu çalışmalar, yüz ve ses ifadelerine ihtiyaç duyulmadan, yalnızca beden duruşu ve hareketlerinden duygu tanımlamanın mümkün olduğunu ortaya koymuştur. Ayrıca, ham kinematik verilere dayalı öznitelik çıkarımları yüksek doğruluk sağlamış ve bu tür verilerin duygu tanıma süreçlerine etkin bir şekilde entegre edilebileceğini göstermiştir. Diğer yandan, video tabanlı analizler, mekânsal-zamansal özellikleri bir arada değerlendirerek, duygu tanıma sürecini daha kapsamlı bir boyuta taşımıştır. Çalışmada kullanılan video veri setleri, beden hareketlerinin dinamik ve bağlamsal özelliklerini

sunarak, görsel boyutun duygu tanıma üzerindeki etkisini detaylandırmıştır. Bununla birlikte, video tabanlı analizlerin işlem maliyeti ve aşırı öğrenme eğilimi gibi zorluklarının, veri işleme teknikleri ve model optimizasyonu ile ele alınması gerektiği vurgulanmaktadır. Tezin bulguları, kinematik verilerin yüksek doğruluğunu ve video verilerinin sunduğu bağlamsal zenginliği bir araya getiren çok modaliteli bir sistemin, duygu tanıma süreçlerine önemli katkılar sağlayabileceğini göstermektedir. Böyle bir yaklaşım, insan-makine etkileşimi, güvenlik, sağlık ve eğitim gibi çeşitli alanlarda geniş uygulama potansiyeline sahiptir. Bu bağlamda, çok modaliteli bir duygu tanıma sisteminin geliştirilmesi hem teorik hem de pratik açıdan yenilikçi bir çözüm sunmaktadır.

Bu doktora tez çalışması, çok açılı veri toplama yöntemleri ve gelişmiş poz tahmini teknikleri ile desteklenmiş iki veri seti üzerinden duygu tanıma sürecini ele alarak, beden dili aracılığıyla duygusal durumların doğru ve güvenilir şekilde tanınması için yenilikçi bir çerçeve sunmaktadır. Çalışma, kinematik ve video tabanlı verilerle ayrı ayrı analizler yaparak, duygu tanıma alanındaki çok yönlü veri işleme yeteneklerini ortaya koymayı amaçlamaktadır. Ham kinematik veriler, insan vücudunun iskelet temelli pozisyon ve hareket bilgilerini içerirken, video tabanlı analizler mekânsal-zamansal özellikleri değerlendirme imkânı sunmaktadır. Bu iki yaklaşımın birlikte ele alınması, duygu tanıma süreçlerinin doğruluğunu artırmaya yönelik çok modaliteli bir çerçeve geliştirilmesine olanak sağlamaktadır. Aynı zamanda, video tabanlı analizlerde mekânsal-zamansal özelliklerin değerlendirilmesiyle bağlamsal zenginlik sağlanmış ve her iki veri tipinin duygu tanıma sürecine katkısı karşılaştırılmıştır.

Bu tez, beden hareketleri üzerinden duyguların yüz ifadelerine ihtiyaç duyulmadan tanımlanabileceğini ortaya koyarak literatüre katkı sağlamayı amaçlamaktadır. Çalışmanın özgün yönlerinden biri, ham kinematik verilerden öznitelik çıkarımı yapılması ve bu verilerin video formatına dönüştürülmüş halinin analiz edilmesidir. Ayrıca, kinematik ve video tabanlı veri türlerinin karşılaştırmalı olarak incelenmesi, bu iki veri tipinin duygu tanıma süreçlerindeki etkilerini çok modaliteli bir çerçevede ele almıştır. Bununla birlikte, geniş bir literatür taraması ve yöntemlerin karşılaştırmalı analizi, tezin teorik ve pratik katkılarının altını çizmektedir.

Çalışmanın bir diğer özgün yönü, kinematik ve video tabanlı veri türlerinin duygu tanıma sürecindeki etkilerini karşılaştırmalı bir analizle ele almasıdır. Bu iki veri tipinin bağlamsal zenginlik, doğruluk ve işlem maliyeti gibi açılardan incelenmesi, duygu tanıma sürecine ilişkin daha derin bir anlayış geliştirilmesine katkı sağlamaktadır. Ayrıca, bu

analizler geniş bir literatür taraması ile desteklenmiş ve mevcut çalışmalarla karşılaştırmalı bir değerlendirme yapılmıştır. Bu kapsamlı literatür incelemesi, tezin hem teorik hem de uygulamalı katkılarını güçlendiren bir diğer unsurdur.

Tezin özgün değerleri arasında ham kinematik verilere dayalı öznitelik çıkarımlarının geliştirilmesi, video formatına dönüştürülmüş aynı veride farklı yaklaşımların denenmesi, geniş bir literatür karşılaştırmasının yapılması ve elde edilen bulguların çok modaliteli bir duygu tanıma çerçevesi kapsamında sunulması yer almaktadır. Bu yönleriyle çalışma, duygu tanıma alanında hem metodolojik hem de uygulamalı açıdan bir yenilik sunmaktadır.

Tezde öncelikle literatürdeki ilgili çalışmalar incelenmiş, ardından kullanılan yöntemler detaylandırılmış ve elde edilen bulguların değerlendirilmesi ve sonuç bölümünde gelecekteki araştırmalara yönelik öneriler sunulmuştur.

2. LİTERATÜR ÇALIŞMALARI

Bu tez çalışmasında, duygu tanıma alanında yapılan araştırmalar üç temel odak noktasında ele alınarak kapsamlı bir literatür taraması yapılmıştır. Bibliyometrik analiz, kinematik verilere dayalı duygu tanıma ve video tabanlı duygu tanıma başlıkları altında gerçekleştirilmiş olan bu incelemede, tezde ele alınan yaklaşımların literatürdeki yerini belirlemek ve çalışmanın özgünlüğünü desteklemek amacıyla yapılandırılmıştır.

İlk olarak, bibliyometrik analiz ile literatürde kinematik ve video tabanlı duygu tanıma üzerine yapılan çalışmaların eğilimleri, öne çıkan konuları ve metodolojik yaklaşımları incelenmiştir. Bibliyometrik veriler, anahtar kelime analizi, kelime bulutları ve görselleştirme teknikleri (ör. TreeMap) kullanılarak değerlendirilmiştir.

İkinci olarak, kinematik veriler üzerine yapılan çalışmalar detaylı bir şekilde ele alınmıştır. Bu kapsamda, insan vücudunun hareket dinamiklerinden öznitelik çıkarımı, bu özniteliklerin sınıflandırma süreçlerine entegrasyonu ve duygu tanıma süreçlerine katkıları değerlendirilmiştir. Literatürde yüz ifadelerine veya ses tabanlı modalitelere bağımlı kalmadan, kinematik verilere dayalı duygu tanıma yöntemlerinin potansiyelini ortaya koyan çalışmalar taranmıştır. Özellikle ham kinematik verilere dayalı yenilikçi öznitelik çıkarım yaklaşımları ve özgün sınıflandırma algoritmaları üzerinde durulmuştur.

Son olarak, video tabanlı duygu tanıma alanında yapılan araştırmalar incelenmiştir ve video tabanlı analizlerin duygu tanıma süreçlerindeki başarısını değerlendiren çalışmalar detaylandırılmıştır. Görsel ve zamansal bilgilerin birleşimiyle elde edilen verilerin bağlamsal zenginliği ile kinematik verilerin doğruluk potansiyeli birleştirilerek çok modaliteli yaklaşımların katkıları analiz edilmiştir.

Bu literatür tarama süreci, tez çalışmasının metodolojik ve teorik temellerini güçlendirmeye, bilimsel boşlukları belirlemeye ve çalışmanın özgün katkılarını vurgulamaya yönelik bir zemin hazırlamıştır. Her bir başlık altında yapılan taramalar, tezde önerilen yaklaşımın literatürdeki yerini ortaya koymuştur.

2.1. Bibliyometrik Analiz

Duygu tanıma alanı, farklı modalitelerin kullanıldığı çok çeşitli sınıflandırma yöntemlerini içermektedir. Literatürde sıklıkla ses, yüz ifadeleri, EEG

(elektroensefalografi) gibi biyofizyolojik sinyaller ve metin tabanlı modaliteler öne çıksa da bu tez çalışması özellikle ham kinematik veriler ve yüz görüntüsü içermeyen video tabanlı sınıflandırmalara yoğunlaşmıştır. Bu kapsamda, bibliyometrik analiz, duygu tanıma alanındaki çalışmaların metodolojik çeşitliliğini anlamak ve tezde ele alınan veri türlerinin literatürdeki konumunu belirlemek için güçlü bir araç olarak kullanılmıştır.

Kinematik ham sinyal ve video tabanlı duygu tanıma görevlerinin bibliyometrik verileri ayrı analiz edilip görselleştirilmiştir. Elde edilen anahtar kelimelerin kelime bulutu (WordCloud) ve zenginleştirilmiş anahtar kelimelerden elde edilen TreeMap ile frekanslara göre trend topik kelimeler ile bu çalışma zenginleştirmiştir. Literatürde yapay zekâ ile duygu tanıma kullanımına dair binlerce çalışma bulunmaktadır. Bu çalışmaların tümü detaylı olarak analiz edilemediği için çalışmalara dair özün kavranabilmesi amacıyla "Scopus" veri tabanı incelenmiştir. Bu çalışmada Scopus veri tabanının tercih edilmesinin nedeni, belirlenen anahtar sorgular için "Web of Science" veri tabanına göre daha fazla çalışmaya erişilebilmesidir. Bu iki veri tabanını karşılaştıran kapsamlı çalışmalar yapılmıştır (Chadegani vd., 2013; Martín-Martín vd., 2018; Singh vd., 2021).

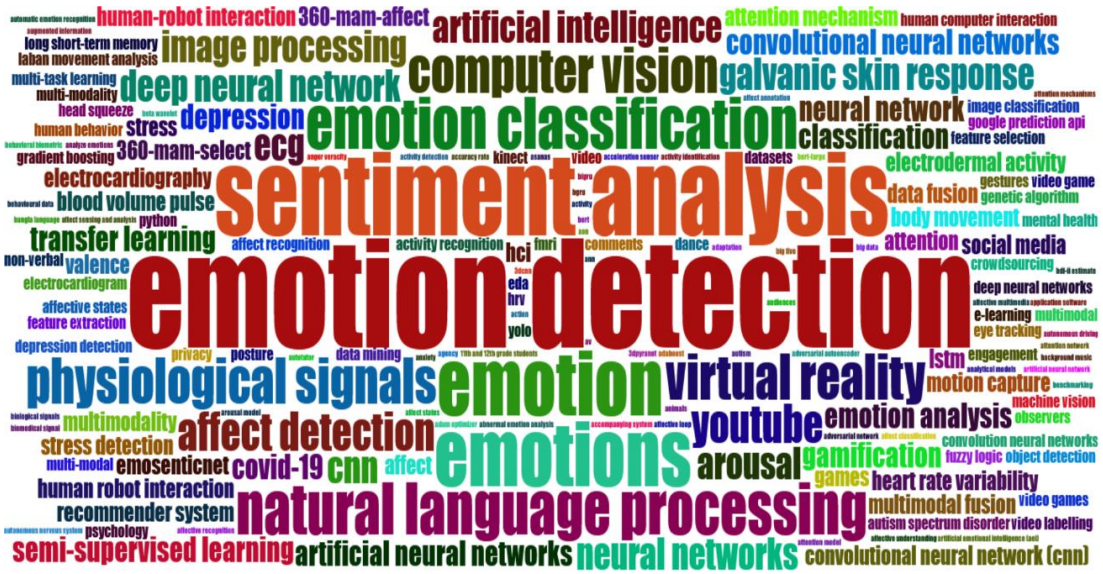
Scopusta analiz yapılırken *TITLE-ABS-KEY (("emotion recognition" OR "affective recognition" OR "emotion detection" OR "affective computing" OR "affect detection" OR "emotion analysis" OR "affective state identification" OR "emotional state recognition" OR "emotion classification") AND ("kinematic" OR "motion capture" OR "video" OR "visual data" OR "3D data" OR "motion analysis" OR "body signals" OR "posture" OR "body movement" OR "joint points") AND ("machine learning" OR "deep learning" OR "neural network*") AND NOT ("speech" OR "audio" OR "facial" OR "face" OR "eeg")) AND (EXCLUDE (PUBYEAR , 2004) OR EXCLUDE (PUBYEAR , 2006) OR EXCLUDE (PUBYEAR , 2008) OR EXCLUDE (PUBYEAR , 2009))) AND (LIMIT-TO (LANGUAGE , "English"))* sorgu kodunda yer alan anahtar kelimeler kullanılmıştır. Bu sorgu, yalnızca İngilizce dilindeki çalışmaları analiz etmiş ve 2010 sonrası çalışmaları analiz etmiştir. Burada duygu tanıma ve eş anlamlı ibareleri ile arama yapılmıştır. Arama sorgusunda literatürde sıklıkla kullanılan ses, yüz ve EEG ile yapılan duygu tanıma çalışmaları hariç tutulmuştur. Bu kriterlere Kasım 2024 itibarıyla uyan toplamda 313 çalışmanın bulunduğu tespit edilmiştir.

Kelime bulutu, kelime sıklığının görsel bir temsilidir. Analiz edilen metinde bir terim ne kadar sık geçiyorsa, bu terim görselde o kadar büyük görünmektedir. Bu çalışmada kullanılan kelime bulutu, R dilinde kurulabilir bir kütüphane olarak sunulan

açık kaynaklı ve çok işlevli bir bibliyometrik analiz yöntemiyle oluşturulmuştur (Aria ve Cuccurullo, 2017).

Kelime bulutu oluşturulurken en çok frekansa sahip " affective computing", " machine learning", " deep learning", “emotion recognition” ve “convolutional neural network” terimleri yüksek tekrar sıklıkları nedeniyle boyut açısından kelime bulutunda çok fazla yer kapladığından, diğer anahtar kelimelerin görünürlüğünü artırmak amacıyla hariç tutulmuştur.

Şekil 2.1'de oluşturulan kelime bulutu, konuların içeriği hakkında genel bir fikir vermede önemlidir. En sık kullanılan ilk 200 anahtar kelime öbeğinden ve yukarıda ayrıntılı olarak açıklananlardan oluşturulmuştur.

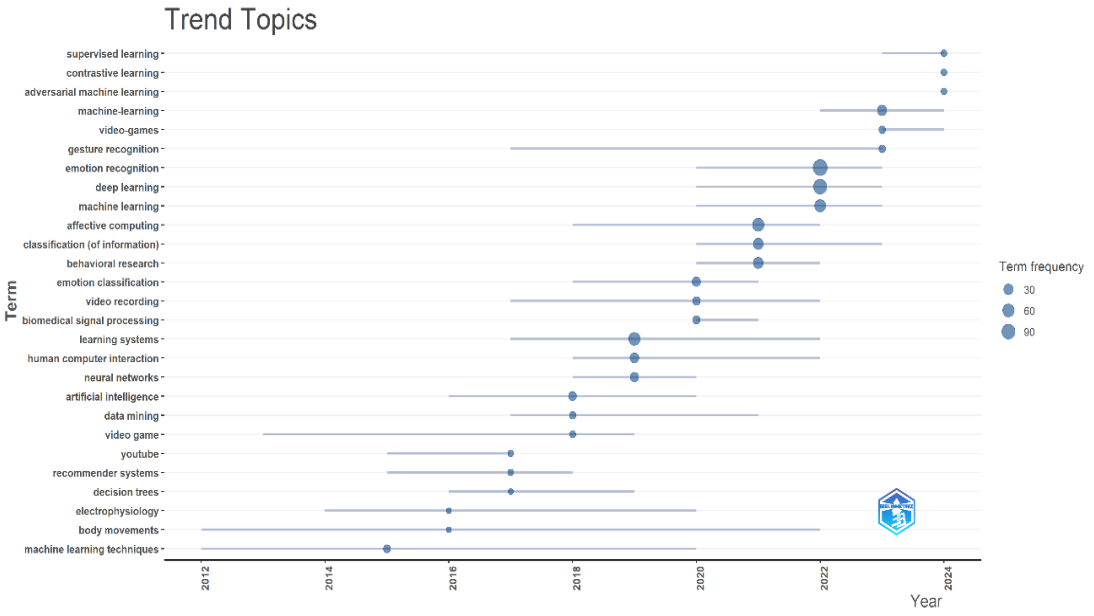


Şekil 2. 1. Kelime bulutu: ilk 200 anahtar kelime öbeği

Şekil 2.2'de, yine zenginleştirilmiş anahtar kelimelerden elde edilen TreeMap ile yıllar içinde yapılan çalışmalarda en sık kullanılan 75 kelimenin sayısal dağılımları verilmiştir.

Şekil 2. 2. TreeMap: yıllar içinde en sık kullanılan 75 kelimenin dağılımı

Şekil 2.3'te, her yıl en az 3 kez kullanılan ve minimum frekansı 4 olan anahtar kelimeler belirlenerek trend topic anahtar kelimeler listelenmiştir.



Şekil 2. 3. Yıllara göre trend topic anahtar kelimeler (frekans ≥ 4)

Bu görseller, duygu tanıma ve ilgili alanların yıllar içindeki gelişimini, popüler terimleri ve trendleri kapsamlı bir şekilde gözler önüne sermektedir. Görsellerden elde edilen bu bilgiler, çalışmanın literatürdeki ana eğilimleri ve odaklarını belirlemede katkı sağlamaktadır.

Elde edilen bulgular, duygu tanıma alanında kinematik veriler ve video tabanlı analizlerin (yüz ve ses içermeyen) henüz sınırlı sayıda çalışmada ele alındığını, ancak bu

yöntemlerin artan bir ilgiyle araştırılmaya devam ettiğini göstermektedir. Kelime bulutu ve TreeMap gibi görsel araçlarla belirlenen trendler, çalışmanızın odaklandığı spatio-temporal yöntemler ve çok modaliteli yaklaşımların literatürdeki önemini destekler niteliktedir. Bu analizler, literatürdeki eğilimleri ve boşlukları sistematik olarak inceleyerek tezinizin özgün değerini vurgulamakta ve duygu tanıma alanındaki metodolojik yaklaşımların zaman içindeki evrimini anlamaya yönelik bir rehberlik sunmaktadır. Bu veriler, çalışmanızın bilimsel katkılarını literatürle uyumlu bir şekilde ortaya koymak için uygun bir bağlam sağlamaktadır.

2.2. Kinematik Veriler ile Duygu Tanıma

Günümüzde, kinematik verilerden yararlanarak duygu tanıma çalışmaları, insan-makine etkileşiminde duygu odaklı sistemlerin geliştirilmesi açısından önemli bir araştırma alanı haline gelmiştir. İnsan vücudu, duygularını yalnızca yüz ifadeleriyle değil, aynı zamanda hareketleri ve beden duruşlarıyla da ifade etmektedir. Bu bağlamda, beden dili ve hareket analizi, duygu tanıma teknolojilerinin geliştirilmesinde zengin bir bilgi kaynağı sunmaktadır. Kinematik veriler, genellikle hareket yakalama cihazları, sensörler veya derin öğrenme algoritmaları ile işlenerek, duygusal durumların sınıflandırılmasına yönelik modeller oluşturulmasında kullanılmaktadır. Bu yöntemler, bireyin hareket dinamiklerini, hız, pozisyon, duruş değişiklikleri ve mekânsal-temporal özellikler gibi parametreler üzerinden analiz ederek duygu durumlarını anlamaya olanak tanımaktadır.

Son yıllarda, bu alanda yapılan çalışmalar, kinematik verilerin sıfır atış (zero-shot) öğrenimi, asenkron zaman serisi analizleri ve derin öğrenme yöntemleriyle birleştirilmesi gibi yenilikçi yaklaşımlarla genişletilmiştir. Ayrıca, hareket temelli duygu tanıma çalışmaları, yalnızca bireyin duygusal durumunu tanımakla kalmamakta, aynı zamanda insan davranışlarının ve sosyal etkileşimlerin daha derinlemesine anlaşılmasını sağlamaktadır. Bu literatürde yer alan çalışmalar, hareket yakalama sensörleri, iskelet temsilleri, derin sinir ağları ve zaman-temporal modellemeler gibi çeşitli yöntemlerin bu alandaki uygulamalarını inceleyerek, kinematik verilerin duygu tanımadaki potansiyelini ortaya koymuştur.

İskelet temsilleri ve kinematik özelliklerle duygu tanıma, kinematik verilerin, özellikle iskelet temsilleri veya hareket yakalama verileri gibi fiziksel özelliklerin analiziyle duyguların tanınmasını hedefler. Fourati ve Pelachaud'un çalışmasında, duyguların tanınması için çok seviyeli bir sınıflandırma çerçevesi geliştirilmiştir.

Çalışma, duygu sınıflandırmasında hem global postüral özellikleri hem de uzuv hareketleri gibi dinamik bileşenleri inceleyerek kapsamlı bir analiz sunmuştur. Veriler, geniş bir hareket yakalama veri tabanından elde edilmiştir ve farklı duygusal durumlara ait detaylı hareket bilgilerini içermektedir. Önerilen model, %85,6 genel doğruluk oranı ile beden hareketlerinden duygu tanıma alanında başarılı sonuçlar elde etmiştir. Özellikle belirgin duyguların, örneğin mutluluk ve üzüntünün sınıflandırılmasında %90 üzeri doğruluk sağlanırken, korku ve şaşkınlık gibi daha karmaşık duyguların tanınmasında %75-80 doğruluk oranlarına ulaşılmıştır. Çalışma, beden ifadelerinin çok katmanlı analizinin, duygu sınıflandırmasında doğruluk oranlarını artırmak için etkili bir yöntem olduğunu göstermiştir (Fourati ve Pelachaud, 2015).

Beden hareketlerinden duygu tanıma alanında farklı bir yaklaşım benimseyen Ahmed ve arkadaşlarının çalışmasında, insan beden hareketlerinden duygu tanımak için hareket dinamiklerini ve postüral özellikleri analiz eden detaylı bir yaklaşım önerilmiştir. İskelet verileri, eklem noktalarının koordinatlarından oluşturulmuş ve zaman serisi şeklinde modellenmiştir. Hareketlerin zamansal bağlamını öğrenmek için LSTM (long short-term memory) (uzun kısa süreli bellek) ağları kullanılmıştır. Çalışmada, mutluluk, üzüntü, öfke ve korku gibi temel duyguların tanınması hedeflenmiş ve CMU (Carnegie Mellon University), Motion Capture Database gibi açık kaynaklı veri setleri üzerinde test edilmiştir. Model, duygu sınıflandırmasında %87 genel doğruluk oranı elde etmiş, mutluluk gibi belirgin duygular %92 doğruluk oranıyla başarılı şekilde tanımlanmış, ancak düşük hareket içeren korku gibi duyguların tanınması %78 doğruluk seviyesinde kalmıştır (Ahmed vd., 2020). Zaman serisi verilerle çalışırken düzensiz örnekleme ve eksik veri gibi sorunların üstesinden gelmek için LSTM tabanlı modellerin etkili olduğu başka domainlerde de belirtilmektedir (Lipton vd., 2016)

Razzaq ve arkadaşlarının geliştirdiği bir başka yaklaşım, UnSkEm (Unobtrusive Skeletal-based Emotion Recognition) modelidir. Bu model, minimal veri işleme ve giyilebilir olmayan sistemler kullanılarak duygu tanıma amacıyla tasarlanmıştır. Model, yalnızca temel iskelet verilerini kullanarak duyguları analiz eder ve karmaşık sensörlere olan ihtiyacı ortadan kaldırır. İskelet verileri, Kinect gibi derinlik sensörlerinden toplanmış ve anahtar eklem noktalarının (örneğin, baş, omuz, dirsek, diz) konum bilgileri kullanılmıştır. Duyguların tanınmasında mekânsal ve zamansal ilişkileri öğrenmek için CNN (convolutional neural networks) (evrişimsel sinir ağları) ve RNN (recurrent neural networks) (yinelemeli sinir ağı) kombinasyonu içeren hafif bir mimari kullanılmıştır. Çalışmada, EMOTIC veri seti üzerinde test edilen model ile %84,5 doğruluk oranı elde

etmiştir. Mutluluk ve öfke gibi belirgin hareketlere sahip duygular %88 doğruluk oranıyla sınıflandırılmış, ancak üzüntü ve korku gibi düşük belirginlikteki hareketlerin tanınması %78-80 doğruluk seviyesinde kalmıştır. Modelin düşük işlem gücü gereksinimi sayesinde gerçek zamanlı uygulamalarda etkili şekilde kullanılabileceği belirtilmiştir. Bu çalışma, hem kullanıcı gizliliğini ön planda tutan hem de pratik uygulamalara yönelik bir duygu tanıma sistemi sunmaktadır (Razzaq vd., 2020). Yine insan hareketlerinden bir olguya varma konusu ve yüksek boyutluluk ve doğrusal olmayan dinamikler gibi zorluklarla mücadele için RNN tabanlı bir mimarinin etkili olduğunu gösterilmiştir (Martinez vd., 2017).

Benzer bir konuda yapılan bir diğer çalışma ise Wu ve arkadaşlarının çalışmasında, beden hareketlerinden duygu tanıma üzerine zero-shot öğrenme yöntemine dayalı yenilikçi bir model geliştirilmiştir. Bu yöntem, özellikle daha önce görülmemiş veya yeni duygu sınıflarını tanımlama konusunda güçlü bir yaklaşım sunmaktadır. Araştırmada, hem bilinen hem de bilinmeyen duygu sınıflarını ayırt edebilen GZSL (generalized zero-shot learning) (genelleştirilmiş sıfır atımlı öğrenme) modeli önerilmiştir. Model, hareket verilerinden anlamlı özellikler çıkarmak amacıyla dikkat mekanizmaları ve gömülü (embedding) temsil tekniklerini kullanmıştır. Çalışma, yaygın kullanılan BodyGesture-E veri seti üzerinde gerçekleştirilmiş ve veri setinde çeşitli duygularla ilişkilendirilen insan beden hareketleri analiz edilmiştir. Model, bilinmeyen duyguları tanımada %78,4, bilinen duyguları tanımada ise %92,6 doğruluk oranına ulaşmıştır. Bu bulgular, sıfır atış öğrenmenin, nadir veya sınırlı veriyle temsil edilen duyguları tanımada etkili bir yöntem olduğunu göstermektedir (Wu vd., 2022).

Derin öğrenme tabanlı duygu tanıma modelleri, kinematik verilerden duygu tanımayı amaçlayan ve genellikle CNN, LSTM veya GCN (graph convolutional networks) (grafik evrişimli ağlar) gibi yapıları kullanan çalışmaları içerir. Bu bağlamda, H. Zhang ve ekibi, vücut hareketlerinden duygu tanımada doğruluğu artırmak için Attention-Supervised LSTM (AS-LSTM) modelini geliştirmiştir. BP4D+ veri setini kullanarak 3D hareket verilerindeki zamansal bağımlılıkları öğrenen bu model, dikkat mekanizmasıyla duygu sınıflandırması için kritik hareketleri vurgulamıştır. AS-LSTM, geleneksel LSTM yöntemlerine kıyasla doğruluk oranını %85'e kadar yükselterek %5-8 oranında iyileşme sağlamış ve özellikle uzun hareket dizilerindeki yanlış sınıflandırmaları azaltmıştır. Bu çalışma, hareket verilerinin önemli noktalarını seçerek daha odaklı ve başarılı bir sınıflandırma sunmuş ve derin öğrenme yöntemleriyle

kinematik verilerden duygu tanımada dikkat mekanizmasının etkinliğini ortaya koymuştur (Zhang vd., 2021).

Diğer yandan Ghaleb ve ekibinin çalışmasında ise iskelet tabanlı hareket verilerini grafik yapılarla temsil eden ve GCN kullanarak analiz eden bir duygu tanıma modeli önermiştir. Çalışmada, Kinect cihazından elde edilen NTU RGB+D veri seti kullanılmıştır. Bu veri seti, geniş bir duygu ve hareket varyasyonu içeren iskelet verileri sağlamıştır. Model, duygu sınıflandırmasında %85 doğruluk oranına ulaşırken açıklanabilirlik açısından da etkili bir performans göstermiştir. Örneğin, hangi eklem hareketlerinin spesifik bir duyguyu ifade ettiği, modelin içsel yapısı sayesinde anlaşılabilmiştir. Bu yenilikçi yaklaşım hem sınıflandırma başarısı hem de modelin insan davranışını anlama yeteneği açısından güçlü bir çözüm sunmuştur (Ghaleb vd., 2021).

H. Zacharatos ve ekibinin çalışması, 3D hareket yakalama verilerini kullanarak CNN'ler ile duygu tanımada kayda değer başarılar elde etmiştir. Araştırma, CMU gibi hareket verisi içeren bir veri seti üzerinde gerçekleştirilmiştir. 3D hareket verileri, duygu tanıma için zamansal ve mekânsal özellikleri içerecek şekilde işlenmiştir. Çalışmada kullanılan model, video benzeri hareket dizilerini analiz eden derin CNN tabanlı bir yapıdır. Deney sonuçları, mutlu, üzgün ve nötr gibi temel duyguların %88 doğruluk oranıyla sınıflandırıldığını göstermiştir. Bu sonuçlar, 3D hareket verilerinin CNN tabanlı modellerle analiz edilmesinin, duygu tanımada güçlü bir yöntem olduğunu kanıtlamıştır (Zacharatos vd., 2021).

Ancak, H. Zhang ve ekibinin çalışmasından farklı olarak, Y. Bhatia ve ekibi, hareket verilerindeki uzamsal ve zamansal özellikleri bağımsız modüllerle analiz eden BMSNN (Bi-modular sequential neural network) modelini geliştirmiştir. Model, hareketlerin uzamsal ve zamansal özelliklerini iki bağımsız modül aracılığıyla analiz etmektedir. Uzamsal modül, iskelet noktaları arasındaki mesafe ve açı ilişkileri gibi pozisyonel özellikleri işlerken, zamansal modül, hareketlerin zaman içindeki hız, ivme ve paternlerini anlamaya odaklanmıştır. Bu iki modülden elde edilen özellikler birleştirilerek duygu sınıflandırması gerçekleştirilmiştir. BP4D+ veri setiyle test edilen model, %87 doğruluk oranına ulaşmış ve geleneksel yaklaşımlara kıyasla %7-10 arasında bir iyileşme sağlamıştır. Çalışma, uzamsal ve zamansal özelliklerin ayrı ayrı işlenmesinin duygu tanıma doğruluğunu artırdığını göstermiştir (Bhatia vd., 2022).

Duygu tanıma alanında yenilikçi yaklaşımlar, sıra dışı yöntemler ve farklı veri türlerini birleştiren multimodal sistemlerle yeni boyutlar kazanmaktadır. Bu kategorideki çalışmalar, duygu tanımada farklı modalitelerin (örneğin görüntü, hareket ve ses) birlikte

kullanılmasının, analizin derinliğini ve doğruluğunu artırabileceğini göstermektedir. S. Saha ve ekibinin çalışması, Microsoft Kinect sensörüyle beden hareketlerinden duyguların tanımlanmasını ele almıştır. Çalışmada, Kinect'in sağladığı 3D iskelet verileri ve RGB görüntüler bir arada analiz edilmiştir. Derin öğrenme tabanlı modellerin kullanıldığı bu yaklaşım, beden hareketlerinin mekânsal ve zamansal özelliklerini ayrıntılı şekilde incelemiştir. Model, mutlu, üzgün ve öfkeli gibi temel duyguları sınıflandırmada %82 doğruluk oranına ulaşarak multimodal veri kullanımının başarısını ortaya koymuştur. Kinect gibi sensörlerin sağladığı zengin veri çeşitliliği sayesinde, beden hareketleri üzerinden duygu tanımada önemli ilerlemeler sağlanmıştır (Saha vd., 2014).

Bu tür yenilikçi yöntemlerin bir başka örneği, L. Farinelli'nin, aktör-robot tiyatro ortamında sahne hareketlerini ve duyguları sınıflandırmak için tasarladığı çok modelli çerçevedir. Çalışma, video, ses ve metin gibi farklı modaliteleri birleştirerek otonom bir sistemin insan hareketlerini anlamasına ve sınıflandırmasına olanak sağlamıştır. Derin öğrenme tabanlı algoritmalarla desteklenen bu framework, multimodal verilerin eşzamanlı analizini yaparak %90'ın üzerinde sınıflandırma doğruluğu elde etmiştir. Bu sistem, insan-robot etkileşimlerinin yaratıcı uygulamalardaki potansiyelini ortaya koyarken, multimodalite yaklaşımının duygu tanıma alanındaki katkılarını da net bir şekilde göstermiştir (Farinelli, 2022).

Kinematik yöntemler, duygu tanıma sistemlerinde beden hareketlerinin zamansal ve mekânsal analizini sağlayarak duyguların belirlenmesinde kullanılır. Literatürde bu yöntemler, geometrik temsiller ve derin öğrenme tabanlı analizlerle ele alınmaktadır. M. Daoudi ve ekibinin çalışmasında ele alınan beden hareketlerini simetrik pozitif kesin matrislerle matematiksel olarak modelleyen bir yöntem önermektedir. Bu yaklaşım, hareketlerin manifold geometrisi üzerinde temsil edilmesiyle duygu durumlarının sınıflandırılmasında geleneksel yöntemlere kıyasla daha yüksek doğruluk oranı sağlamıştır. Çalışmada, hareketlerin zamansal değişimleri Riemannian metrikler ve manifold tabanlı sınıflandırıcılar kullanılarak analiz edilmiştir. CMU veri seti üzerinde gerçekleştirilen bu araştırmada, temel duygular %87 doğruluk oranıyla sınıflandırılmıştır. Bu yöntem, hareketlerin mekânsal ve zamansal bileşenlerini geometrik bir perspektifle ele alarak duygu tanıma sistemlerinde yenilikçi bir çözüm sunmaktadır (Daoudi vd., 2017).

Bir başka çalışma olan D. Avola ve ekibinin çalışması, derin öğrenme tabanlı temporal analiz yöntemlerini kullanarak doğaçlama olmayan beden hareketlerini

incelemeye odaklanmıştır. Çalışmada, hareketlerin zamansal dinamiklerini anlamak için LSTM ve GRU (gated recurrent unit) (geçitli yineleme birimi) modelleri uygulanmıştır. FABO veri setiyle yapılan bu araştırmada, mutlu, üzgün ve öfkeli gibi temel duyguların sınıflandırılmasında %88 doğruluk oranına ulaşılmıştır. Temporal analizlerin sunduğu zamansal ilişki modelleme kapasitesi, bu çalışmada duygu tanıma sistemlerinin doğruluğunu artırmada güçlü bir araç olarak öne çıkmıştır (Avola vd., 2022).

2.3. Video Verilerinde Duygu Analizi ve Spatio-Temporal Yöntemler

Dinamik video verileri, duyguların zaman ve mekân içinde nasıl ifade edildiğini analiz etmek için zengin bir kaynak sunar. İnsan duygu durumları genellikle yüz ifadeleri, vücut hareketleri ve duruş değişiklikleri gibi birden fazla modalitede eş zamanlı olarak ortaya çıkar. Bu tür verilerin analizi, geleneksel durağan görüntü işleme yöntemlerinin ötesine geçerek, zaman içindeki değişimleri ve hareket kalıplarını yakalamayı gerektirir. Spatio-temporal yöntemler, mekânsal özelliklerin (örneğin, yüz ve vücut yapıları) zamansal düzenlemelerle (örneğin, hareket sekansları ve hız değişiklikleri) birleştirilmesini sağlayarak, dinamik duygu tanıma sistemlerinde kritik bir rol oynar (Cao vd., 2021). Bu bağlamda, derin öğrenme tabanlı yaklaşımlar, video dizilerindeki karmaşık duygu geçişlerini daha hassas bir şekilde anlamak için güçlü araçlar sunmaktadır. eMotions gibi geniş ölçekli veri setleri, kısa video analizleri için etiketlenmiş veri sağlayarak duygu tanıma araştırmalarının ilerlemesine önemli katkılarda bulunmuştur (Wu vd., 2023). Bu bölümde, dinamik video verilerinin işlenmesinde kullanılan spatio-temporal yöntemler ve bu yöntemlerin duygu tanıma alanındaki katkıları incelenmiştir.

Spatio-temporal (zaman ve mekân) temelli analizler, video tabanlı duygu tanıma süreçlerinde zamansal ve mekânsal bilgilerin bir arada işlenmesini hedefler. Bu konuda Zhang ve arkadaşları CNN ve RNN tabanlı hibrit bir model önererek, zamansal ve mekânsal bilgilerin birleştirilmesini sağlamıştır. Zhang ve arkadaşlarının çalışmasında, 3D-CNN kullanılarak video karelerinden mekânsal özellikler çıkarılmıştır. Zamansal bilgiler ise LSTM ve GRU ile analiz edilmiştir. Mekânsal ve zamansal özellikler, dikkat mekanizmalarıyla birleştirilerek duygu sınıflandırması gerçekleştirilmiştir. Çalışma hem düşük hem de yüksek çözünürlüklü video verilerinde etkili sonuçlar vermiştir. Ayrıca, domain adaptation teknikleriyle modelin farklı veri setlerindeki başarımı artırılmıştır (Zhang vd., 2018).

Zhang ve arkadaşlarının (2018) çalışmasında önerilen hibrit modelin performansı, EmotiW2018 veri seti üzerinde test edilmiştir. Bu testlerde, modelin doğruluk oranı %53,6 olarak rapor edilmiştir. Bu sonuç, modelin kompleks video tabanlı duygu sınıflandırmada düşük de olsa etkili olduğunu göstermektedir. Wang ve arkadaşları ise ST-CNN (spatio-temporal CNN) kullanarak zaman-mekân tabanlı özelliklerin çıkarımını gerçekleştirmiştir. Yöntemde, videolardan kareler çıkarılarak ön işleme tabi tutulmuş ve optik akış yöntemleriyle hareket bilgisi elde edilmiştir. Daha sonra, görüntü karelerinden mekânsal özellikler CNN ile çıkarılmış, zamansal bilgiler ise LSTM gibi modellerle analiz edilmiştir. Özellikler birleştirilerek, dikkat mekanizmaları yardımıyla önemli zaman dilimlerine ağırlık verilmiştir. Çıkarılan birleşik özellikler, fully connected layer üzerinden duygu sınıflarına atanmıştır. Bu yaklaşım, optik akış ile yüksek seviyeli görsel özelliklerin bir araya getirilmesi sayesinde, özellikle beden hareketleri üzerinden duygu tanıma performansını artırmıştır. Ancak, Wang ve arkadaşlarının çalışmasında önerilen ST-CNN modelinin performansına ilişkin spesifik başarı oranları belirtilmemiş olup, metodun zamansal ve mekânsal bilgileri birleştirerek duygu tanıma süreçlerinde daha yüksek doğruluk sağladığı ifade edilmiştir (Wang vd., 2021).

Multimodal yaklaşımlar, video tabanlı duygu tanımada birden fazla veri türünü (örneğin, yüz ifadeleri, vücut hareketleri, ses) birleştirerek daha kapsamlı ve doğru analizler yapmayı amaçlar. Chen ve arkadaşları çalışmalarında, yüz ifadeleri ve vücut hareketleri gibi farklı modaliteleri birleştirerek duygu tanıma performansını artırmayı hedefleyen bir sistem önerilmiştir. Önerilen sistem, her bir modaliteden özellikleri çıkararak bunları birleştirir ve daha sonra duygu sınıflandırması yapar. Yüz ifadelerinden mekânsal özellikler CNN kullanılarak çıkarılmış, vücut hareketlerinden ise optik akış yöntemleriyle zamansal özellikler elde edilmiştir. Daha sonra çıkarılan mekânsal ve zamansal özellikler, dikkat mekanizmaları ile birleştirilerek duygu tanıma için önemli özelliklere ağırlık verilmiştir. Son olarak, bu özellikler fully connected layer üzerinden duygu sınıflarına atanmıştır. Farklı modalitelerden elde edilen bilgilerin birleştirilmesi ve dikkat mekanizmalarının kullanımı, sistemin duygu tanıma performansını önemli ölçüde artırmıştır. Çalışmada önerilen sistem, farklı veri setlerinde test edilmiş ve yüz ifadeleri ile vücut hareketlerini birleştirerek duygu tanıma doğruluğunu %85,4'e kadar çıkarmayı başarmıştır. Bu sonuç, yalnızca yüz ifadelerini veya yalnızca vücut hareketlerini kullanan sistemlere kıyasla önemli bir gelişme göstermiştir. Çalışmada önerilen sistem ile özellikle, multimodal veri kullanımının duygu tanıma performansını artırdığı ve dikkat

mekanizmasının modelin önemli özelliklere odaklanmasında kritik bir rol oynadığı belirtilmiştir (Chen vd., 2023).

Vaiani ve arkadaşlarının çalışmasında ise, MLLM (multimodal large language models) (multimodal büyük dil modelleri) kullanılarak farklı veri türlerinin (metin, görüntü, ses vb.) bir arada analiz edilmesi hedeflenmiştir. Büyük dil modellerinin kapasitesini kullanarak multimodal verilerden duygu tanıma yapılmıştır. Bu sistemde, farklı modalitelerden veriler toplanmış ve ön işleme tabi tutulmuştur. Büyük dil modelleri, multimodal verilerle eğitilerek farklı veri türleri arasındaki ilişkileri öğrenmiştir. Eğitilen model, yeni multimodal veriler üzerinde duygu tanıma gerçekleştirmiştir. Büyük dil modellerinin kapasitesi sayesinde multimodal verilerin daha etkili bir şekilde analiz edilmesi sağlanmış, farklı veri türleri arasındaki ilişkilerin öğrenilmesiyle duygu tanıma performansı artırılmıştır ve testlerde %88,7 doğruluk oranı elde edilmiştir. Bu sonuç, metin, görüntü ve ses gibi birden fazla veri türünün bir arada analiz edilmesinin model performansını artırdığını göstermektedir. Özellikle, farklı modaliteler arasındaki ilişkilerin öğrenilmesi, duygu tanıma görevlerinde daha yüksek doğruluk sağlamıştır. Çalışma, MLLM'lerin multimodal duygu analizi için güçlü bir araç olduğunu ortaya koymuştur (Vaiani vd., 2024).

Kinematik hareketlere dayalı video tabanlı duygu tanıma, yüz ifadelerinden bağımsız olarak vücut hareketleri ve hareket dinamikleriyle duyguların çıkarımını hedefler. Bu yöntemler, özellikle yüz ifadelerinin yetersiz kaldığı durumlarda (örneğin, maskeli yüzler, uzak mesafedeki bireyler) önemli bir avantaj sunar. Kahou ve arkadaşları, çalışmalarında, gerçek hayattaki video verilerinde hareket tabanlı duygu tanımaya odaklanmıştır. Çalışmada, kinematik hareketlerden çıkarılan özellikler derin öğrenme modelleriyle analiz edilmiştir. Önerilen yöntem, özellikle insan gruplarındaki kolektif hareketlerin (örneğin, topluluklarda öfke veya mutluluk yayılımı) duygu analizi üzerindeki etkisini incelemiştir. Modelde, 3D hareket analizi için skeleton tabanlı bir veri çıkarımı yapılmış ve ardından bu veriler, RNN ile işlenmiştir. Çalışmanın sonuçları, modelin %78,9 doğruluk oranına ulaştığını ve gerçek hayatta kinematik verilere dayalı duygu çıkarımının uygulanabilirliğini kanıtladığını göstermiştir (Kahou vd., 2019).

Xu ve arkadaşları çalışmalarında ise duygu tanıma için kavramsal hareket modellerine dayalı bir yöntem geliştirmiştir. Bu yaklaşım, kinematik verilerden hareket dinamiklerini çıkararak bu bilgileri duygu analizi için kullanmayı hedefler. Çalışma, hareketlerin hız, ivme ve mekânsal koordinatlar gibi kinematik özelliklerini analiz etmiştir. Özellikler, CNN ile çıkarılmış ve LSTM ağlarıyla zamansal bağlamda

değerlendirilmiştir. Model, insan hareketlerinin duygusal anlamlarını çözümlemekte başarılı sonuçlar elde etmiştir. Testlerde, önerilen yöntemle %82,5 doğruluk oranı sağlanmış ve kinematik hareketlerin duygu tanımadaki etkinliği ortaya konulmuştur (Xu vd., 2021).

Zhang ve arkadaşları tarafından yapılan bir çalışmada, kinematik ve görsel verilerden elde edilen anahtar karelere dayalı bir duygu analizi yöntemi geliştirilmiştir. Bu model, video sekanslarından önemli anları (anahtar kareler) seçerek analiz yapar. Anahtar kare seçimi için optik akış algoritmaları kullanılmış ve hareketlerin yoğun olduğu zaman dilimleri belirlenmiştir. Bu karelerden çıkarılan mekânsal özellikler CNN ile analiz edilmiştir. Zamansal bağlam ise LSTM kullanılarak değerlendirilmiş ve tüm bilgiler birleştirilerek duygu sınıflandırması yapılmıştır. Testlerde, modelin %84,2 doğruluk oranı elde ettiği ve özellikle videolardaki hareket dinamiklerinin analizinde etkili olduğu belirtilmiştir (Zhang vd., 2021).

Zhang, Zhao ve Li ekibinin bir başka çalışmasında, gerçek zamanlı vücut hareketlerinden duygu tanıma amacıyla derin öğrenme ve takviyeli öğrenme tekniklerini birleştiren bir model geliştirilmiştir. Video tabanlı skeletal analiz yöntemiyle vücut hareketlerinin hız, ivme ve pozisyon gibi kinematik özellikleri çıkarılarak ön işleme tabi tutulmuştur. Model, mekânsal bilgileri analiz etmek için CNN, zamansal dinamikleri anlamak için ise LSTM ağlarını kullanmıştır. Takviyeli öğrenme yaklaşımı, modelin sınıflandırma performansını optimize etmek ve gerçek zamanlı işleme kapasitesini artırmak için uygulanmıştır. %86,3 doğruluk oranı elde eden sistem, bireysel ve grup tabanlı duygu tanıma görevlerinde etkili sonuçlar vermiştir. Bu çalışmada, yüz ifadelerine bağımlı olmayan kinematik analizlerin doğruluk ve uygulanabilirlik açısından güçlü bir alternatif olarak sunulduğu ortaya konulmaktadır (Zhang vd., 2022).

Bir başka benzer çalışmada, derin öğrenme tabanlı duygu tanıma yöntemlerinin kapsamlı bir incelemesi sunulmuş ve farklı yaklaşımların performanslarını değerlendirmiştir. Çalışmada, CNN tabanlı modellerin görsel duygu tanımada %85'in üzerinde doğruluk oranlarına ulaştığı belirtilmiştir. Zamansal analiz gerektiren görevlerde RNN ve LSTM modellerinin daha etkili olduğu, özellikle zamansal bağımlılıkları anlamada bu modellerin başarılı sonuçlar verdiği ifade edilmiştir. Ayrıca, Transformer tabanlı modellerin, multimodal veri analizi ve çapraz-modalite görevlerde en iyi sonuçları verdiği ve doğruluk oranlarını diğer yöntemlere göre önemli ölçüde artırdığı rapor edilmiştir. Çalışma genel bir karşılaştırma sunduğu için kesin bir doğruluk oranı belirtmek yerine, her yöntemin farklı veri kümelerinde sağladığı tipik başarı seviyelerini

özetlemiştir. Bu kapsamda, derin öğrenme tabanlı yöntemlerin duygu tanıma süreçlerinde yüksek potansiyel sunduğu vurgulanmıştır (Abdurrahman vd., 2022).

Burada sunulan literatür analizi, dürtüsel duygusal senaryolarda veri yönetimine odaklanmanın önemini ve duygu tanıma süreçlerinde kullanılan özellik çıkarma yöntemleri ile sınıflandırma algoritmalarının geliştirilmesi gerektiğini açıkça ortaya koymuştur. Bu bağlamda, beden dili ve duygusal durum arasındaki ilişkiyi derinlemesine anlamak ve duygusal durum tanıma sistemlerinin gerçek dünya koşullarında uygulanabilirliğini değerlendirmek için daha kapsamlı araştırmalara ihtiyaç duyulmaktadır. Bu çalışmada, iskelet tabanlı kinematik veri setlerinin hem ham veriler hem de işlenip video formatına dönüştürülmüş halleri üzerinden beden duruşuna dayalı otomatik duygu tanıma analizi gerçekleştirilmiş ve gerçek dünya senaryolarında gerçek zamanlı duygu tanıma için yenilikçi ve uygulanabilir bir yaklaşım önerilmiştir.

3. MATERYAL VE YÖNTEM

Çalışmanın amacı, beden hareketlerine dayalı duygu tanıma alanında iki farklı veri tipini kullanarak elde edilen bulguları ayrı ayrı değerlendirmektir. Beden hareketleriyle duyguların tanımlanması, hem kinematik verilere dayalı ham hareket analizleri hem de video tabanlı görsel tanımlamalar gerektiren geniş kapsamlı bir süreçtir. Çalışmamızda, her iki veri setiyle de bağımsız analizler yapılmış ve bu veri tiplerinin duygu tanıma üzerindeki etkileri ayrı ayrı incelenmiştir. İlk veri seti, oyuncuların anatomik düğüm noktalarından elde edilen kinematik hareket verileri ile duygu ifadelerini iskelet tabanlı olarak analiz etmeyi amaçlamaktadır. Bu kinematik veriler, duygu durumlarının vücut hareketleri üzerinden ölçülmesi ve sınıflandırılması için detaylı pozisyon ve rotasyon bilgileri sağlar. Diğer yandan, Dalian Emotional Movement Open-source Set (DEMOS) video veri seti, duygusal ifadelerin görsel olarak analiz edilmesine odaklanarak, video tabanlı sınıflandırma yöntemleriyle çalışılmasına imkân tanır. Bu iki veri seti ile yürütülen ayrı çalışmalar, duygu durumlarının bedensel ifadeler aracılığıyla tanınabilirliğini farklı yöntemlerle değerlendirmeyi amaçlamaktadır. Böylece, her bir veri seti kendi metodolojik yaklaşımlarıyla duygu tanıma sürecine özgün katkılarda bulunmakta ve bu alanın farklı perspektiflerden ele alınmasını sağlamaktadır.

3.1. Materyal

Kinematik veri seti ve DEMOS veri seti, aynı araştırma grubu tarafından geliştirilmiş olmaları nedeniyle birbiriyle ilişkili olmakla birlikte, çalışma kapsamı ve hedefleri açısından belirgin farklılıklar taşımaktadır. İlk olarak, kinematik veri seti yedi temel duygu kategorisini (öfke, tiksinti, korku, mutluluk, nötr, üzüntü ve şaşkınlık) içermekteyken, DEMOS veri seti yalnızca altı temel duyguya (öfke, tiksinti, korku, mutluluk, nötr ve üzüntü) odaklanmıştır. Bu durum, surprise (şaşkınlık) kategorisinin kinematik veri setinde yer almasına rağmen DEMOS veri setinde bulunmamasıyla ortaya çıkmaktadır. DEMOS veri setinin daha dengeli bir duygu dağılımı sunmayı ve duygular arasındaki ayrımı daha hassas bir şekilde incelemeyi amaçlayan bir tasarıma sahip olduğunu göstermektedir. Ayrıca, DEMOS veri setinde duygu kategorileri için eşit sayıda ve farklı açılardan (0°, 45° ve 90°) çekim yapılması, veri dengesi ve çeşitlilik açısından önemli bir avantaj sağlamaktadır.

Kinematik veri seti, duygu ifadelerinin temel hareket kalıplarını yüksek zaman çözünürlüğüyle (125 Hz) ve ham pozisyon verileri üzerinden incelemeye olanak tanırken, DEMOS veri seti, duygu ifadelerini video formatında ve farklı perspektiflerden görselleştiren bir yapı sunmaktadır. Bu, kinematik veri setinin özellikle hareketlerin detaylı anatomik analizi için uygun olduğunu, DEMOS veri setinin ise görsel algılama ve farklı açılardan duygu tanıma üzerine çalışmalara olanak sağladığını ortaya koymuştur. DEMOS veri seti ayrıca, oyuncuların hareketlerini belirli senaryolar üzerinden yönlendirilmiş şekilde kaydetmesiyle, kinematik veri setine kıyasla daha yapılandırılmış bir veri toplama sürecine sahiptir.

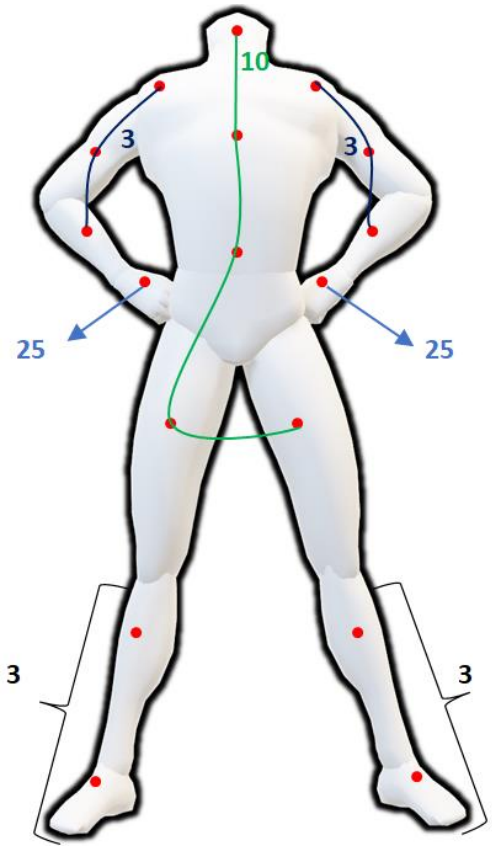
Her iki veri setinin aynı yazar grubu tarafından oluşturulmuş olması, birbirlerini tamamlayan ve farklı araştırma ihtiyaçlarına cevap verebilen bir yapı oluşturmasını sağlamıştır. Kinematik veri seti, duygu ifadelerinin iskelet tabanlı mikro düzeyde analizi için uygun bir zemin sunarken, DEMOS veri seti, bu ifadelerin makro düzeyde algısal ve görsel özelliklerini değerlendirmek için geniş bir çerçeve sunmaktadır. Dolayısıyla, bu iki veri setinin kombinasyonu, duygu tanıma çalışmalarında çok boyutlu bir analiz yapılmasını mümkün kılarak hem teknik hem de uygulamalı araştırmalarda kapsamlı bir veri kaynağı sağlamaktadır.

3.1.1. Ham kinematik veri seti

Çalışmada kullanılan kinematik veri seti, 22 yarı profesyonel oyuncunun yedi adet duyguyu (öfke, tiksinti, korku, mutluluk, nötr, üzüntü ve şaşkınlık) (anger, disgust, fear, happiness, neutral, sadness ve surprise) ayakta durarak sahnelediği ve PhysioNet platformunda 2020 yılında paylaşılmış zengin ve güncel bir kinematik veri setidir (Zhang vd., 2020; Goldberger vd., 2000). Veriler vücudun tamamına yakın olan 72 anatomik düğümü kapsamakta olup, bu düğümlerin x, y, z eksenindeki koordinatları için ham pozisyon ve rotasyon bilgilerini içerir.

Ham verilere ait kinematik düğümlerin büyük çoğunluğu el ve parmak noktalarında olup, eklem noktaları ve notasyonunun ayrıntılı detayları ilgili veri setinin paylaşıldığı çalışmadan elde edilebilir (Zhang vd., 2020). Bir elin her bir parmağından 5'er düğüm olmak üzere bir el için 25 ve iki el için toplamda 50 adet kinematik düğüm bilgisi alınabilmektedir. Bu da veri setinde kullanılan 72 kinematik düğümün üçte ikisinden fazlasının ellerde bulunan sensörlerden elde edildiğini göstermektedir. Şekil 3.1, aktörlerin baş, omurga, kalça, kol, el, bacak ve ayaklarına yerleştirilen 17 sensörün

yaklaşık konumlarını (kırmızı daireler) ve bu sensörlere karşılık gelen anatomik düğüm numaralarını göstermektedir.



Şekil 3. 1. Aktörlere yerleştirilen sensörlerin yaklaşık anatomik konumları

Verileri yakalamak için kullanılan kablosuz MoCap sistemi Perception Neuron adında bir ürün olup, iskelet tabanlı çalışmalarda etkin olarak kullanılabilen bir üründür (Robert-Lachaine vd., 2020). Oyuncular daha önce duygusal yoğunluğu test edilmiş senaryoya dayalı performansın yanı sıra duygu ifadesini anlayıp doğal olarak gerçekleştirdikleri performanslar da gerçekleştirmişlerdir.

Tablo 3. 1. Kullanılan ham kinematik veri setine ait bazı istatistiksel bilgiler

	Duygusal Durum						
	Öfke	Tiksinti	Korku	Mutluluk	Nötr	Üzüntü	Şaşkınlık
Dosya sayısı	200	210	217	216	145	202	212
Ortalama kare sayısı	891,2	924,9	855,7	836,1	903,7	1064,5	849,4
Ortalama kayıt uzunluğu (saniye)	7,13	7,39	6,84	6,68	7,23	8,51	6,79

Tablo 3.1'de kullanılan veri setine ait her bir duygu için ham dosya sayısı, dosya başına ortalama çerçeve ve saniye cinsinden süre bilgileri verilmiştir. Toplanan orijinal

kayıt sayısı 1402 adet olup, veriler 125 Hz'de örneklenmiş ve Biovision Hierarchy (.bvh) formatında sunulmuştur. Bu çalışmada veri setinde sunulmuş olan eklem rotasyon bilgileri kullanılmamış, sadece anatomik düğümlere ait ham pozisyon verileri kullanılarak denemeler gerçekleştirilmiştir.

3.1.2. DEMOS video veri seti

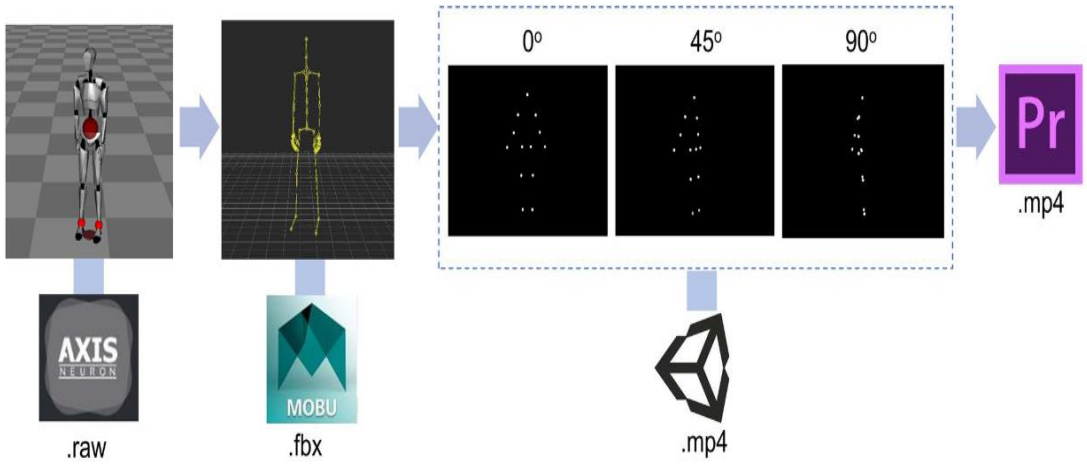
Çalışmada kullanılan DEMOS veri seti, beden hareketleri aracılığıyla duygu tanıma çalışmalarında kullanılan geniş kapsamlı bir veri setidir (Zhang vd., 2023). Veri seti, 22 yarı profesyonel oyuncunun gerçekleştirdiği altı temel duyguya (öfke, tiksinti, korku, mutluluk, nötr, üzüntü) (anger, disgust, fear, happiness, neutral, sadness) ait hareketlerden oluşmaktadır.

Tablo 3. 2. Kullanılan DEMOS veri setine ait bazı istatistiksel bilgiler

Kullanılan Görüntü Açıları	Duygular						
	Mutluluk	Üzüntü	Öfke	Korku	Nötr	Tiksinti	Toplam
0°	156	147	151	169	113	152	888
45°	156	147	151	169	113	152	888
90°	156	147	151	169	113	152	888
Toplam	468	441	453	507	339	456	2664

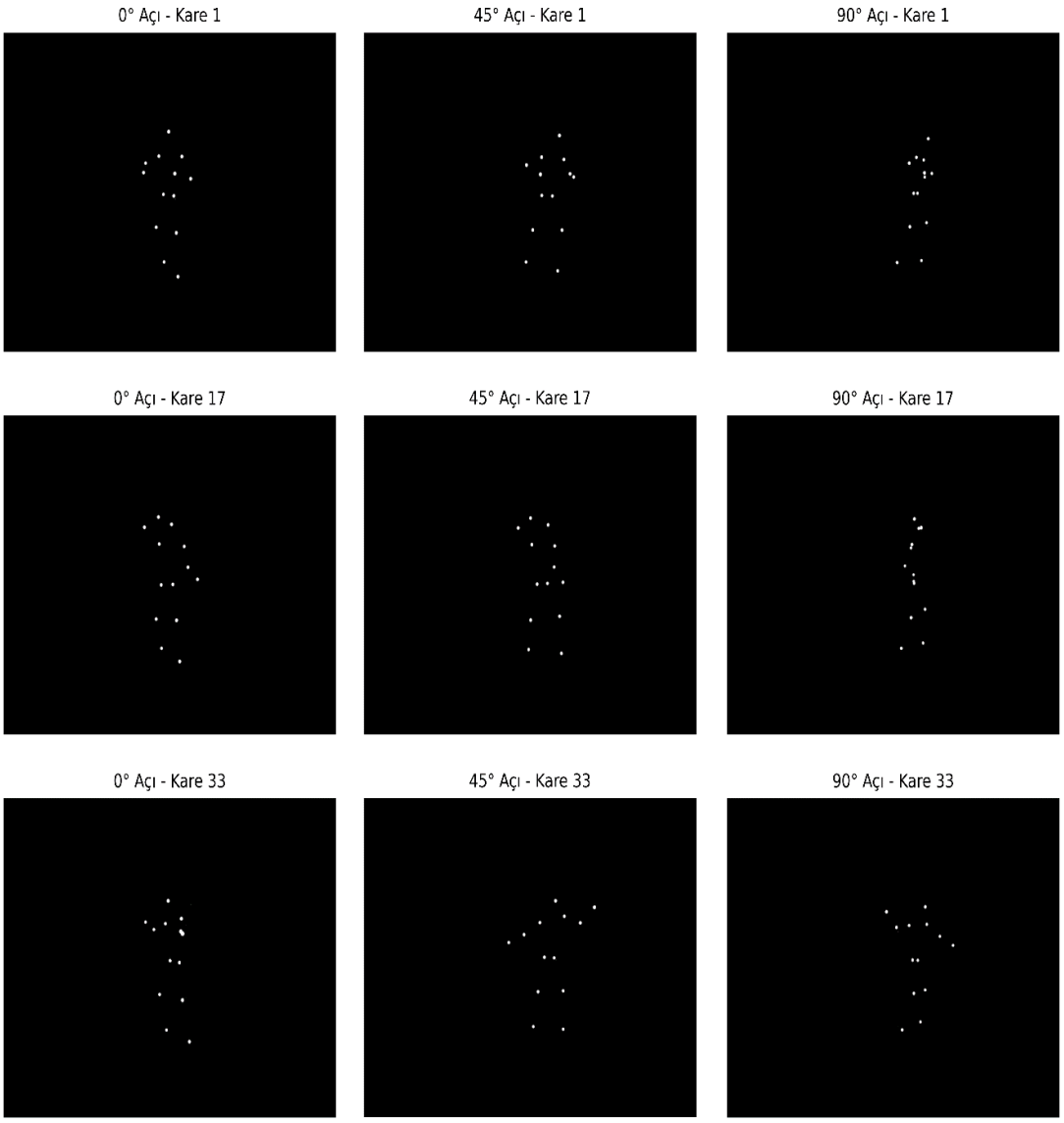
Her bir duygu, Tablo 3.2'de gösterilen üç farklı açıdan (0°, 45° ve 90°) kaydedilmiş toplam 2664 video ile temsil edilmektedir. Videolar; baş, omuzlar, kollar, kalçalar, dizler, ayaklar ve eller gibi 13 ana düğüm noktasına yerleştirilen beyaz ışık noktalarıyla siyah bir arka plan üzerinde gösterilmektedir. Bu düğüm noktaları, oyuncuların beden hareketlerinin ayrıntılı bir şekilde izlenebilmesini sağlar.

Her video iki saniye uzunluğunda olup, 720x540 piksel çözünürlükte ve saniyede 25 kare hızında MP4 formatında kaydedilmiştir. Veri setinin hazırlanmasında, oyuncular günlük yaşantıdan seçilen senaryolar üzerinden duygusal yoğunluğu yüksek hareketler sergilemek üzere yönlendirilmiştir. Bu kayıtlar, oyuncuların duygu durumlarını ifade eden beden hareketlerini yansıtmak amacıyla makale yazarları tarafından, MotionBuilder ve Axis Neuron yazılımları kullanılarak ham formatta alınmış, ardından FBX formatına dönüştürülerek işlenmiştir (Shi vd., 2024). Bu süreç, Şekil 3.2'de gösterildiği gibi gerçekleşmiştir.



Şekil 3. 2. DEMOS veri seti kayıt ve işleme sürecinin adımları

DEMOS veri seti, farklı açılardan kaydedilmiş duygu hareketleri sayesinde, duygu tanıma çalışmalarında farklı görüş açılarının etkilerini inceleme fırsatı sunmaktadır. Bu süreç, Şekil 3.3'te örnek olarak aynı hastaya ait 'öfke' duygusunun performansının, farklı kareler ve üç farklı açıdan görselleştirildiği bir şekilde sergilenmiştir.



Şekil 3. 3. Öfke duygusunun üç farklı açıdaki temsili

Bu yapı, araştırmacıların beden hareketlerinin farklı açılardan nasıl algılandığını değerlendirmelerine olanak tanır. Bu veri seti duygusal biyolojik hareketlerin analizi, sosyal biliş, psikiyatri ve duygusal hesaplama gibi alanlarda geniş bir kullanım potansiyeline sahip olup, duygu tanıma çalışmalarına özgün ve kapsamlı bir veri kaynağı sağlamaktadır.

3.2. Yöntem

3.2.1. Makine öğrenmesi mimarisi

Makine öğrenmesi, veriden anlam çıkarma ve tahmin yapma amacıyla çeşitli algoritmalar ve yöntemler kullanan bir yapay zekâ disiplinidir. Geleneksel programlamadan farklı olarak, makine öğrenmesi, veri setinden kendi kendine öğrenmeyi sağlar ve belirli kurallar belirlemeksizin görevleri çözmeyi hedefler (Caruana vd., 2006). Bu bağlamda, makine öğrenmesi modelleri sınıflandırma, regresyon, kümeleme ve boyut indirgeme gibi farklı görevlerde etkin bir şekilde kullanılmaktadır. Bu çalışmada ise beden hareketlerine dayalı duygu tanıma amacıyla makine öğrenmesi yöntemlerinin uygulanabilirliği değerlendirilmiştir.

Makine öğrenmesi mimarisi, veri işleme ve model eğitimi aşamalarını kapsayan katmanlı bir yapıya sahiptir ve dört ana aşamadan oluşur: veri ön işleme, özellik çıkarma, model eğitimi ve model değerlendirme (Abro vd., 2021). İlk aşama olan veri ön işleme, ham verinin analize uygun hale getirilmesi sürecini içerir. Eksik veya tutarsız verilerin temizlenmesi, verinin belirli bir ölçeğe getirilmesi, normalizasyon ve ölçekleme işlemleri bu aşamada gerçekleştirilir. Böylece, veri daha sağlıklı bir analiz ortamına hazırlanarak modelin eğitimi sırasında her bir veri noktasının eşit katkı sağlaması amaçlanır. Bu işlem, modelin doğruluğunu artırırken, eğitimin hızını da olumlu yönde etkiler.

İkinci aşama olan özellik çıkarma, ham verilerden anlamlı ve bilgi sağlayan özelliklerin elde edilmesini ifade eder. Bu aşamada, veriyi anlamak ve modelin doğru tahminler yapabilmesini sağlamak amacıyla veriye ait önemli nitelikler seçilir. Örneğin, görüntü işleme problemlerinde renk, doku ve kenar gibi özellikler belirlenirken; metin analizinde kelime sıklığı veya n-gram yapıları gibi metinsel özellikler çıkarılabilir. Özellik seçimi ile modelin sadece en önemli nitelikleri öğrenmesi sağlanarak gereksiz hesaplamalardan kaçınılır, bu da modelin performansını artırır ve aşırı öğrenmenin önüne geçilmesini sağlar.

Model eğitimi aşamasında, makine öğrenmesi algoritmaları veriden örüntüleri öğrenir ve model parametrelerini optimize eder. Bu aşamada üç temel öğrenme stratejisi öne çıkar: denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme (Penney vd., 2019). Denetimli öğrenme, etiketli veri üzerinde modelin hatalarını minimize etmeye çalışırken, denetimsiz öğrenme, etiketsiz veriyle çalışarak veriyi benzerliklere göre gruplar ve örüntüler çıkarır. Pekiştirmeli öğrenmede ise, model, ödül ve ceza mekanizması ile kendisini geliştirir. Örneğin, bir oyun veya robot kontrol uygulamasında, model başarılı adımlarında ödül alırken, hatalı adımlarında cezalandırılır ve bu sayede süreç boyunca optimal stratejiyi öğrenir (Penney vd., 2019).

Model değerlendirme aşamasında ise, modelin doğruluğu ve genelleme yeteneği ölçülür. Doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılarak, modelin performansı test edilir. Aşırı öğrenme gibi istenmeyen durumları tespit etmek için eğitim ve test veri kümeleri üzerindeki performans farkları gözlemlenir ve gerekirse model iyileştirmeleri yapılır. Bu aşamada, modelin tahmin yeteneğinin artırılması amacıyla düzenleme yöntemleri veya daha fazla veri kullanılarak model optimize edilebilir (Kaynar vd., 2016).

Makine öğrenmesi, çözmek istenilen problem türüne göre çeşitli algoritmalardan yararlanır. Doğrusal modeller, özellikle doğrusal ilişkilerin olduğu veri setlerinde etkili olurken; karar ağaçları ve birden fazla modeli bir araya getirerek daha güçlü tahminler yapan yöntemler, karmaşık örüntülerde yüksek doğruluk sunar. Bu model birleştirme yöntemlerinde, birden fazla makine öğrenmesi modeli kullanılarak bir model grubu oluşturulur. Her bir model, problemi çözmek için kendi tahminini yapar ve sonunda tüm modellerin tahminleri birleştirilerek nihai sonuca ulaşılır. Bu sayede, modelin genelleme kapasitesi artar ve karmaşık veri setlerinde daha başarılı sonuçlar elde edilir. En yaygın model birleştirme yöntemleri arasında Rastgele Ormanlar (Random Forests) (Breiman vd., 2001), Gradyan Artırma (Gradient Boosting) (Natekin vd., 2013) ve Bagging (Bootstrap Aggregation) (Lee vd., 2019) gibi teknikler yer alır. Bu yöntemler, makine öğrenmesi modellerinde farklı bakış açılarını bir araya getirerek daha kararlı ve doğru sonuçlar elde edilmesini sağlar.

Destek vektör makineleri (SVM) (Srivastava vd., 2010) ve K-en yakın komşu (KNN) (Sun vd., 2018) algoritmaları, sınıflandırma görevlerinde başarılıdır. Ayrıca, Bayes sınıflandırıcıları, veri noktalarının belirli bir sınıfa ait olma olasılıklarını hesaplar ve basit ama güçlü bir sınıflandırma yöntemi sunar.

Makine öğrenmesi ve derin öğrenme arasındaki temel farklardan biri, makine öğrenmesinde özellik çıkarma işleminin genellikle manuel yapılması ve modelin daha küçük veri kümeleri ile de etkili çalışabilmesidir. Derin öğrenme, çok katmanlı ve yüksek hesaplama gerektiren yapılarla çalışırken, makine öğrenmesi daha az karmaşık algoritmalarla hızlı çözümler sunar.

3.2.2. Derin öğrenme mimarisi

Derin öğrenme, karmaşık veri örüntülerini otomatik olarak öğrenebilen çok katmanlı yapılarla çalışan bir makine öğrenmesi yaklaşımıdır. Büyük veri setleri ve güçlü

hesaplama yeteneklerinin birleşimiyle, derin öğrenme modelleri, veriden anlam çıkarma ve sonuçları tahmin etme süreçlerinde yüksek doğruluk sunmaktadır. Bu mimariler, özellikle çok katmanlı yapay sinir ağlarına dayanmakta olup üç ana katmandan oluşur: giriş katmanı, gizli katmanlar ve çıkış katmanı. Giriş katmanı, modele işlenecek veriyi sağlar; gizli katmanlar, bu veriyi çok adımlı bir işlem sürecinden geçirerek anlamlı temsilciler çıkarır; çıkış katmanı ise nihai tahminleri sunar (LeCun vd., 2015).

Her bir katmanda, nöronlar birbirleriyle bağlantı halindedir ve belirli bir ağırlıklandırma işlemi yapılır. Giriş verileri, belirli ağırlıklarla (wi) çarpılarak bir nörona iletilir, burada toplanır ve aktivasyon fonksiyonları yardımıyla işlenir. Bu süreç, modelin farklı katmanlardan geçerek verideki karmaşık ilişkileri öğrenmesine olanak tanır. Derin öğrenme, Şekil 3.2'de gösterildiği gibi, ileri yayılım ve geri yayılım süreçlerine dayanır. İleri yayılım, modelin tahmin yapması için gerekli olan ilk adımdır; geri yayılım ise hataların düzeltilmesi amacıyla ağırlıkların güncellenmesini sağlar (LeCun vd., 2015).

Derin öğrenme modellerinin eğitiminde, üç temel öğrenme stratejisi öne çıkar. İlki olan denetimli öğrenme, etiketli veriler kullanılarak modelin öğrenme sürecini içerir. Bu yöntemde, her bir veri örneğine karşılık gelen etiketle modelin tahmini karşılaştırılır ve ortaya çıkan tahmin hataları hesaplanır. Bu hatalar doğrultusunda, modelin nöron ağırlıkları optimize edilerek doğruluğu artırılır. Denetimli öğrenme, sınıflandırma veya regresyon gibi belirli hedef sonuçları içeren problemlerde en sık kullanılan yaklaşımdır. Bu yöntemde, modele doğru etiketlerle birlikte verilen verilerle istenen çıktıya ulaşmak hedeflenir. İkinci bir yöntem olan yarı denetimli öğrenme ise, az miktarda etiketli veri ve bol miktarda etiketsiz veri olduğunda devreye girer. Burada model, az sayıdaki etiketli veriyle eğitilirken, etiketsiz veriler modelin genelleme becerisini geliştirmek için kullanılır. Bu yaklaşım, etiketlenmiş veri ihtiyacını azaltarak maliyetleri düşürür ve geniş veri kümelerinde daha verimli bir öğrenme sağlar.

Denetimsiz öğrenme ise yalnızca etiketsiz verilerle yapılır ve modelin verileri, aralarındaki benzerliklere göre gruplaması hedeflenir. Bu yöntemde, veriler arasındaki ilişkiler model tarafından otomatik olarak öğrenilir ve belirli kümeler veya yapılar oluşturulur. Denetimsiz öğrenme, özellikle örüntü çıkarımı, özellik keşfi ve veri sıkıştırma gibi işlemler için tercih edilir ve etiketsiz veri ile modelin içsel bir yapı kurmasını sağlar (Doğan vd., 2019). Bu üç öğrenme stratejisi, farklı veri türleri ve problemlere göre esnek bir derin öğrenme modeli yapısı sunarak modelin performansını artırır ve duygu tanıma gibi karmaşık görevlerde yüksek başarı elde edilmesini sağlar.

3.2.2.1. Evrişimsel sinir ağı

CNN ve RNN, yapay sinir ağları ailesinin iki temel temsilcisi olup, farklı veri türlerine yönelik özel olarak geliştirilmiş mimarilerdir (Doğan vd., 2019). Yapay sinir ağları, genel anlamda biyolojik sinir sistemlerinden ilham alarak tasarlanmış matematiksel modellerdir. Bu modeller, bilgisayar sistemlerinde öğrenme yoluyla çeşitli görevleri yerine getirebilen yapay nöronlardan oluşur. Büyük veri setleri ile çalışabilme yetenekleri sayesinde, sinir ağları görüntü işleme, ses tanıma ve dil işleme gibi karmaşık problemlerde etkili çözümler sunmaktadır.

RNN, sıralı veri türlerini işlemek için geliştirilmiş yapay sinir ağı modelleridir. Bu mimari, önceki adımlardan elde edilen bilgileri bellekte tutarak ve bu bilgileri sonraki adımlarda kullanarak geçmiş ve gelecekteki veriler arasındaki ilişkileri öğrenme kapasitesine sahiptir. Dil işleme, çeviri ve konuşma tanıma gibi sıralı yapıya sahip veri türlerinde oldukça etkili olan RNN, bu alanlarda geçmiş bilgileri kullanarak gelecekteki tahminleri doğrudan etkileyebilir. RNN'lerin bu hafıza özelliği, onları zaman serisi analizleri, metin sıralaması ve dil modelleme gibi sıralı verilerle çalışmada oldukça başarılı kılar.

Öte yandan, CNN, görsel verilerdeki desenleri tanımlamak amacıyla geliştirilmiştir. Görüntü verilerindeki özellikleri çıkarmak ve sınıflandırmak için evrişimsel filtreler kullanır ve bu filtreler yardımıyla büyük görsel veri setlerinde etkili şekilde çalışır. Örneğin, ilk katmanlar görüntüdeki temel yapıları (kenarlar, dokular) tanımlarken, sonraki katmanlar daha karmaşık desenleri (nesneler, yüzler) öğrenir. Bu özellikleri sayesinde CNN'ler, görüntü tanıma, nesne algılama ve yüz tanıma gibi görsel örüntü tanıma görevlerinde oldukça başarılı sonuçlar verir.

RNN ve CNN mimarileri, veri türlerine özgü avantajları sayesinde farklı uygulama alanlarında yüksek performans gösterir. RNN'ler sıralı verilerle (dil ve zaman serisi gibi) çalışmada daha başarılı iken, CNN yapısı özellikle görüntü verisini işleme konusunda güçlüdür. Bu iki sinir ağı türü çoğu zaman birbirini tamamlayacak şekilde kullanılır; örneğin, bir videodaki nesnelerin zaman içindeki hareketini analiz etmek için her iki mimarinin birleşimiyle bir model oluşturulabilir. Bu tür birleşik modeller, hareketli görsellerdeki nesnelerin hem uzamsal (CNN) hem de zamansal (RNN) özelliklerini öğrenerek daha doğru tahminlerde bulunur. Böyle bir entegrasyon, bilgisayar görüsü, doğal dil işleme ve daha birçok alanda önemli ilerlemelere yol açarak modern uygulamaların temelini oluşturur (Shrestha vd., 2019).

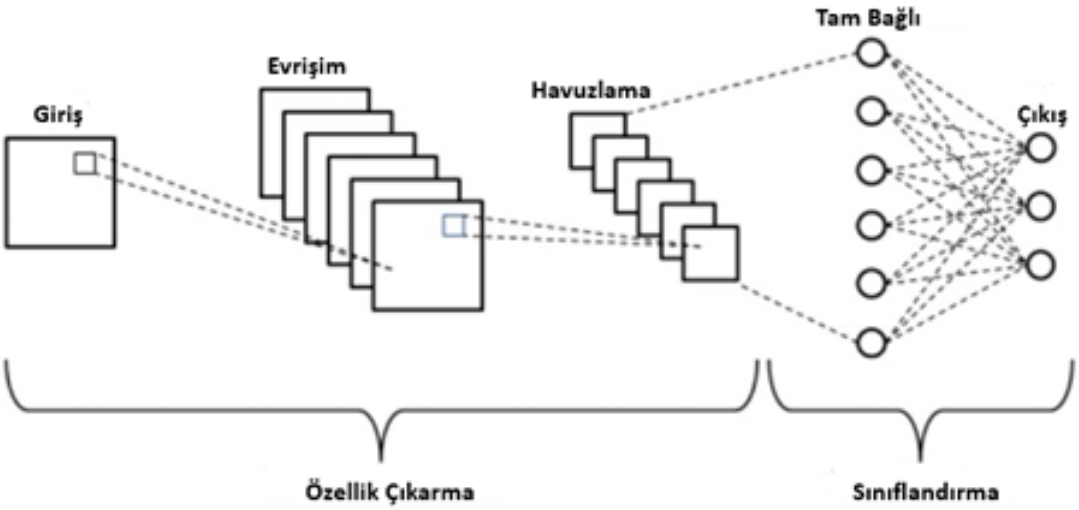
CNN mimarisi, insan beynindeki görsel işlemleme bölgesine benzer bir hiyerarşik düzene sahiptir. İnsan beynindeki nöronlar gibi, CNN yapısında da belirli bir düzen içinde organize edilmiş yapay nöronlar bulunur. Bu düzen, görsel uyaranları işleyen beyin yapısına benzer bir organizasyon sağlayarak, veriyi parça parça incelemeden bütünsel olarak analiz etme olanağı sunar. Geleneksel sinir ağlarına kıyasla, CNN'ler görsel girdilerle daha verimli çalışır ve ses ya da konuşma gibi farklı veri türlerinde de başarılı sonuçlar verir. Bu esneklik, onları çoklu veri türleri ile çalışabilecek çok yönlü bir yapı haline getirir.

Bu ağların sahip olduğu evrişimsel filtreler, görüntü işleme dışında ses ve konuşma sinyalleri gibi veri türlerinde de etkili çalışmasını sağlar. Bu çok yönlülük, CNN mimarisini geniş bir veri yelpazesinde uygulanabilir hale getirir ve derin öğrenme projelerinde önemli bir seçenek oluşturur. Örneğin, sesli komut tanıma, tıbbi görüntü analizleri ve video sınıflandırma gibi görevlerde CNN'in güçlü özellik çıkarma kapasitesinden yararlanılır. Bu mimari yapı sayesinde evrişimsel sinir ağları, yalnızca görüntü verisi için değil, aynı zamanda ses, metin ve diğer sinyalleri de işleyebilen güçlü bir araç haline gelir. Böylelikle, evrişimsel sinir ağları çok modlu verilerle çalışan uygulamalarda sıklıkla tercih edilmekte ve geniş kapsamlı bir veri işleme yeteneği sunmaktadır (Shrestha vd., 2019).

Sonuç olarak, CNN ve RNN gibi yapay sinir ağı modelleri, veri türüne göre optimize edilmiş yapıları sayesinde birçok farklı alanda başarıyla kullanılmaktadır. CNN'ler görsel veri işleme ve desen tanıma alanında öne çıkarken, RNN'ler sıralı verileri işleme konusunda üstün performans gösterir. Bu modeller, derin öğrenme uygulamalarının kapsamını genişletmiş ve günümüzde birçok modern teknolojiye güç sağlamıştır.

3.2.2.2. Evrişimsel sinir ağı katmanları

CNN derin öğrenme modelleri içinde kullanılan önemli yapı taşlarından biridir. CNN'ler, genel olarak Şekil 3.4'te gösterildiği gibi üç ana katmandan oluşur: evrişim katmanı, havuzlama katmanı ve tam bağlantılı katman.



Şekil 3. 4. CNN mimarisi (Mathew vd., 2023)

CNN'in ilk katmanı olan evrişim katmanı (Convolutional Layer), görüntüden özellik çıkarma işlevini üstlenir. Bu katman, görüntü üzerinde belirli desenleri ve yapıları algılamak için filtreler kullanır. Her bir filtre, görüntü üzerinde kayarak belirli bir alandaki bilgileri çıkarır ve bu bilgiyi özellik haritalarına dönüştürür. Bu sayede model, kenarlar, dokular ya da renk geçişleri gibi düşük seviyeli özellikleri belirgin hale getirir. Evrişim işlemi sırasında kullanılan bu filtreler, modelin daha karmaşık yapıları anlamasına temel oluşturur.

İkinci katman olan havuzlama katmanı (Pooling Layer), evrişim katmanından elde edilen özellik haritalarının boyutunu küçültmek ve belirgin özellikleri öne çıkarmak amacıyla devreye girer. Genellikle maksimum havuzlama veya ortalama havuzlama yöntemleri kullanılarak boyut azaltılır. Bu işlemler sayesinde hem hesaplama maliyeti düşer hem de öğrenilen özellikler daha genelleştirilmiş hale gelir. Havuzlama katmanı, modelin girdilerdeki küçük değişikliklere daha dayanıklı olmasını da sağlar ve veriye karşı esnekliği artırır.

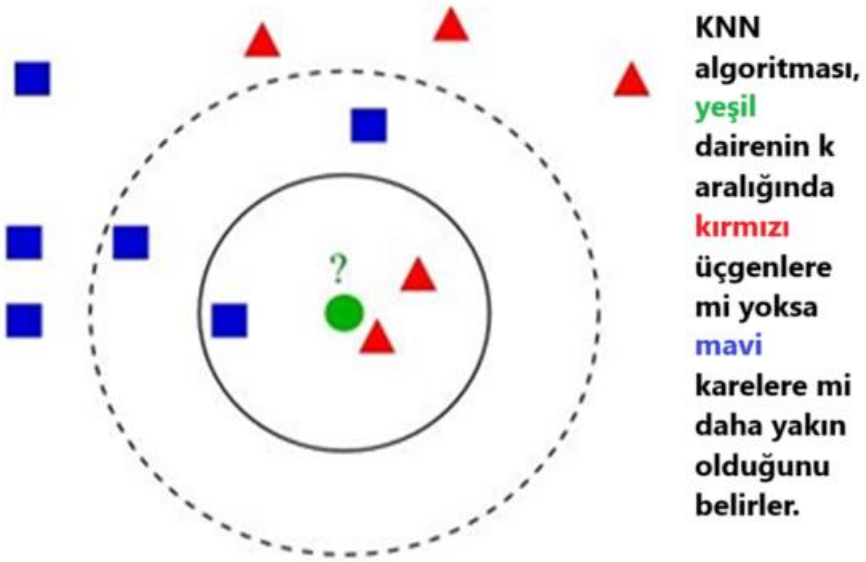
CNN'in son katmanı olan tam bağlantılı katman (Fully Connected Layer) ise, önceki katmanlarda çıkarılan özellikleri kullanarak nihai sınıflandırma veya tahmin görevini yerine getirir. Bu katmanda özellik vektörleri, tüm nöronlarla bağlantı kurarak ağırlıklarla çarpılır ve böylece belirli bir sınıf veya sonuç elde edilir. Tam bağlantılı katman, modelin öğrendiği özelliklerin bütüncül bir değerlendirmeye tabi tutulduğu son aşamadır.

Bu üç katman arasında ilerledikçe CNN'in karmaşıklığı artar ve model daha büyük ve karmaşık özellikleri öğrenebilir hale gelir (Alzubaidi vd., 2021). İlk katmanlarda daha basit yapılar tespit edilirken, ilerleyen katmanlarda bu yapılar birleşerek daha soyut ve yüksek seviyeli özelliklere dönüşür. Sonuç olarak, CNN, nesneyi veya görüntüyü bütünüyle tanımlayacak bir temsili öğrenmiş olur. Bu yapısıyla CNN'ler, görüntü tanıma ve sınıflandırma gibi görevlerde yüksek başarı oranlarıyla kullanılmaktadır.

3.2.5. Kullanılan makine öğrenmesi yaklaşımları

3.2.5.1. K-nearest neighbors (knn)

K-Nearest Neighbors (KNN), uzaklık ve komşuluk sayısı parametrelerinin önemli olduğu, hızlı ama eğitim anlamında "tembel" olarak nitelendirilebilecek temel makine öğrenmesi algoritmalarından biridir (Razzaq vd., 2020). Bu algoritma, tahmin edilecek noktaların diğer noktalara uzaklığını belirli uzaklık fonksiyonları (örneğin, Euclidean, Manhattan) yardımıyla hesaplar ve belirlenen k komşuluk değerinin büyüklüğüne göre sınıflandırma yapar. Şekil 3.5'te gösterilen kNN, basitliği ve etkili sonuçlar üretmesi nedeniyle birçok farklı veri kümesinde uygulanabilir bir yöntemdir.

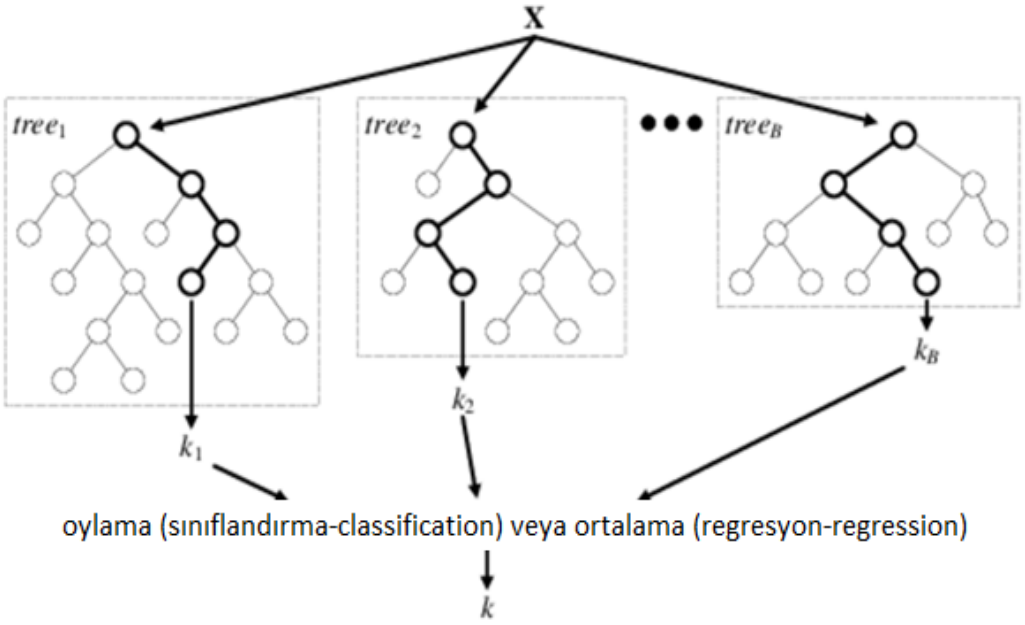


Şekil 3. 5. kNN algoritmasının uzaklık ve komşuluk ilişkisi

Bu çalışmada, kNN için uzaklık fonksiyonları ve k komşuluk parametresi, grid search yöntemiyle optimize edilmiştir. Ancak, farklı parametre değerleri arasındaki performans farklarının önemsiz düzeyde olması nedeniyle, Python kütüphanesi varsayılan değerleri kullanılmaya uygun görülmüştür. KNN, duygu tanıma problemlerinde temel bir sınıflandırıcı olarak değerlendirilmiştir.

3.2.5.2. Random forest (rf)

Random Forest (RF), karar ağaçlarının topluluk öğrenimi ("bagging") yöntemine dayalı bir model olarak tanımlanabilir. RF, tüm karar verici ağaçların birbirinden bağımsız olarak öğrendiği ve birçok karar verici ağaçtan oluşan bir ormanda, en fazla oy verilen çıktının seçildiği bir öğrenme modeli olarak öne çıkmaktadır (Abdulkareem vd., 2021; Breiman vd., 1996). Şekil 3.6'da gösterilen RF modeli, bilgi kazanımına en fazla katkısı olan ağaçların rastgele belirlenmesiyle, birçok DT (Decision Tree) sınıflandırıcısının yaptığı işi kendi içinde gerçekleştirir.



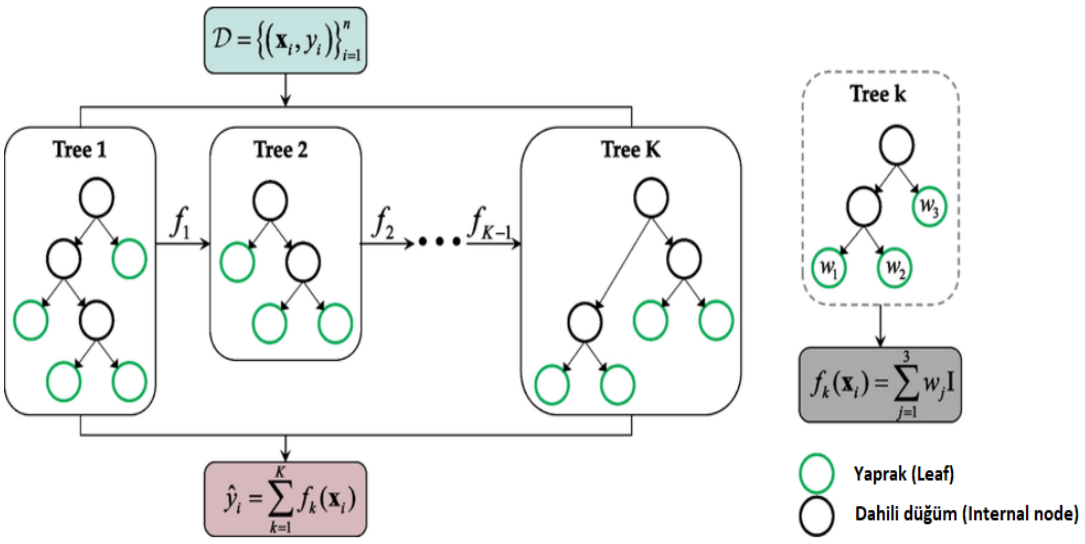
Şekil 3. 6. RF algoritması

RF, hiper parametre kestirim ihtiyacının azlığı hem regresyon hem de sınıflandırma problemlerine etkin bir şekilde uygunluğu ve özniteliklerin önemini belirleyebilme özellikleri ile oldukça popüler bir yöntemdir. Bu çalışmada RF için grid search yöntemi kullanılarak ağaç sayısı ve bölünme kriteri gibi parametreler optimize

edilmiş, ancak varsayılan parametrelerin tatmin edici sonuçlar verdiği gözlemlenmiştir. RF, bu çalışmada, özellikle verilerin dengesiz olduğu durumlarda güçlü bir performans sergileyen bir model olarak tercih edilmiştir.

3.2.5.3. Xgboost

XGBoost (Extreme Gradient Boosting), boosting algoritmalarının güncel bir temsilcisi olup, ağaç yapılarına dayalı öğrenme tekniklerinde yüksek performansı ve esnekliği ile dikkat çekmektedir (Chen vd., 2016). Şekil 3.7'de gösterilen XGBoost, asimetrik ve seviye bazında ağaç büyümesi yapısı, gradyan inişi (gradient descent) optimizasyonu ve regularization (düzenleme) özellikleri ile öne çıkar. Bu özellikler, aşırı öğrenmeyi engellemeye yönelik güçlü bir yapı sunmaktadır.



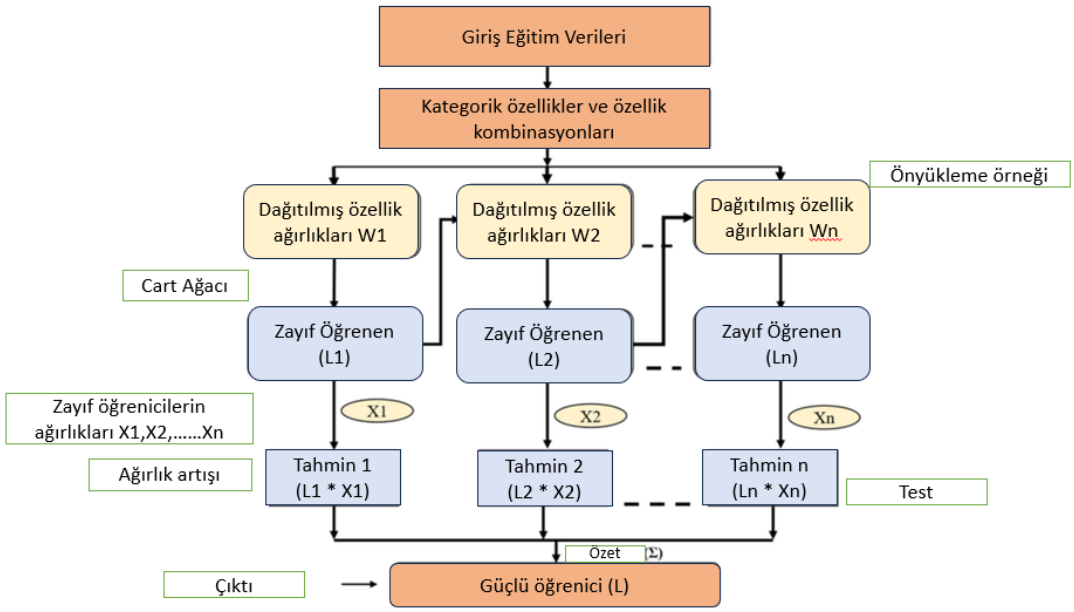
Şekil 3. 7. XGBoost mimarisi (Fazily vd., 2023)

Çalışmamızda, XGBoost için öğrenme hızı, maksimum ağaç derinliği ve ağaç sayısı gibi hiper parametreler grid search yöntemiyle optimize edilmiştir. Bununla birlikte, varsayılan parametreler ile optimize edilmiş değerler arasında kayda değer bir fark bulunmamış ve varsayılan ayarlar tercih edilmiştir. XGBoost, duygu tanıma çalışmalarında doğruluk ve hız açısından üstün bir model olarak bu çalışmada yer almıştır.

3.2.5.4. Catboost

CatBoost, boosting algoritmalarının bir başka modern temsilcisi olup, özellikle kategorik verilerle etkin çalışabilme yeteneğiyle öne çıkmaktadır (Prokhorenkova vd., 2018). Simetrik ağaç yapısı ve kategorik özniteliklerin dinamik bir şekilde işlenmesi, CatBoost'un dikkat çeken özelliklerindendir. Ayrıca, regularization, overfitting önleme ve paralel işleme gibi özellikler, bu algoritmayı birçok problemde etkili bir seçenek haline getirmiştir.

Bu çalışmada, CatBoost için öğrenme hızı, ağaç derinliği ve diğer parametreler grid search yöntemiyle optimize edilmiş, ancak varsayılan ayarların yeterli performans sağladığı gözlemlenmiştir. Şekil 3.8'de gösterilen CatBoost, özellikle kategorik özniteliklerin bulunduğu veri kümelerinde başarıyla uygulanmış ve bu çalışmada boosting tabanlı yöntemlerden biri olarak değerlendirilmiştir.



Şekil 3. 8. CatBoost model mimarisi (Sapkota vd., 2023)

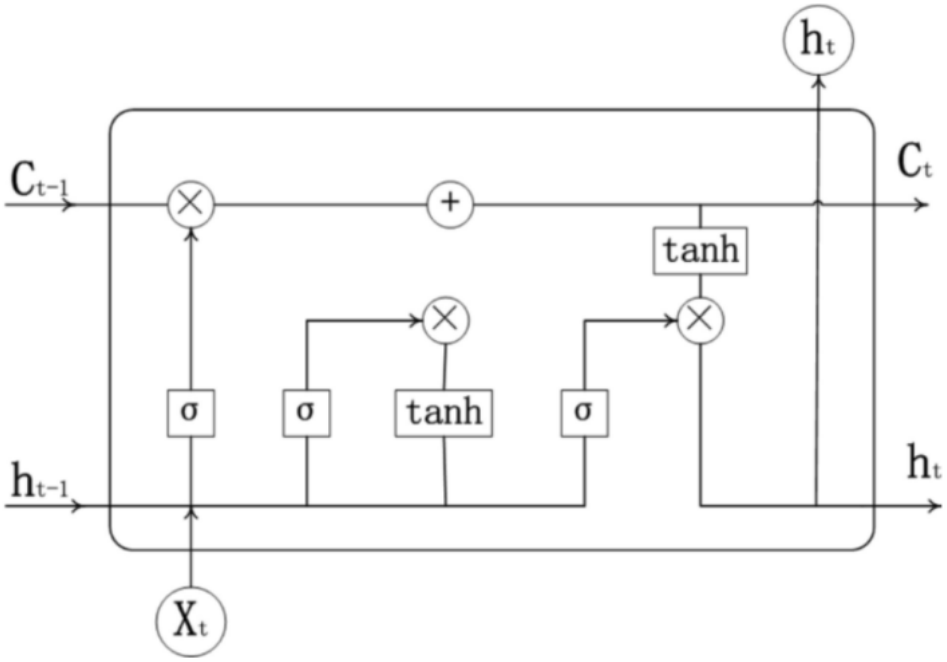
KNN, RF, XGBoost ve CatBoost modelleri hem bağımsız analizlerde hem de karşılaştırmalı değerlendirmelerde bu çalışmanın temel sınıflandırıcıları olarak yer almıştır.

3.2.6. Kullanılan derin öğrenme yaklaşımları

3.2.6.1. Long short-term memory (lstm)

RNN yapısı bir önceki veriye bağlı olarak bir sonraki adımı tahmin etmek için hatırlama özelliği barındıran ve sıralı veriler için ideal olan bir derin öğrenme algoritmasıdır (Bentéjac vd., 2021). Yani çıktı, gelen bilginin sadece ileri doğru işlendiği ileri besleme (feedforward) yapının aksine diğer andaki girişlere de bağlı olarak çıkarılarak geçmişten yararlanılır. RNN algoritmasının kullanım alanlarının genişliği bir avantaj iken yavaş işlem kapasitesi, uzun zaman önceki bilgiye erişme zorluğu ve gradient vanishing / exploding problemleri nedeniyle kararsız bir hal alması başlıca dezavantajlarıdır (Rumelhart vd., 1986; Pascanu vd., 2013). Yapılarına eklenmiş olan bilgi akışını düzenleyebilen kapılar içeren mimarileri sayesinde ilgili sorunları çözebilen Şekil 3.9'da gösterilen LSTM (Hochreiter vd., 1997) algoritması bu çalışmada sınıflandırma için kullanılmıştır.

Bir LSTM temelde üç kapı kullanarak işlemlerini gerçekleştirir: (i) unutma kapısı (forget gate) ile hangi bölümünün hatırlanmaya değer olduğunu belirler, (ii) giriş kapısı (input gate) ile ağ üzerindeki veri akışını sağlayan "hücre durumu (cell state)" in durumunu günceller, (iii) çıkış kapısı (output gate) ile bir sonraki katmana iletilecek bilgiye karar verir.



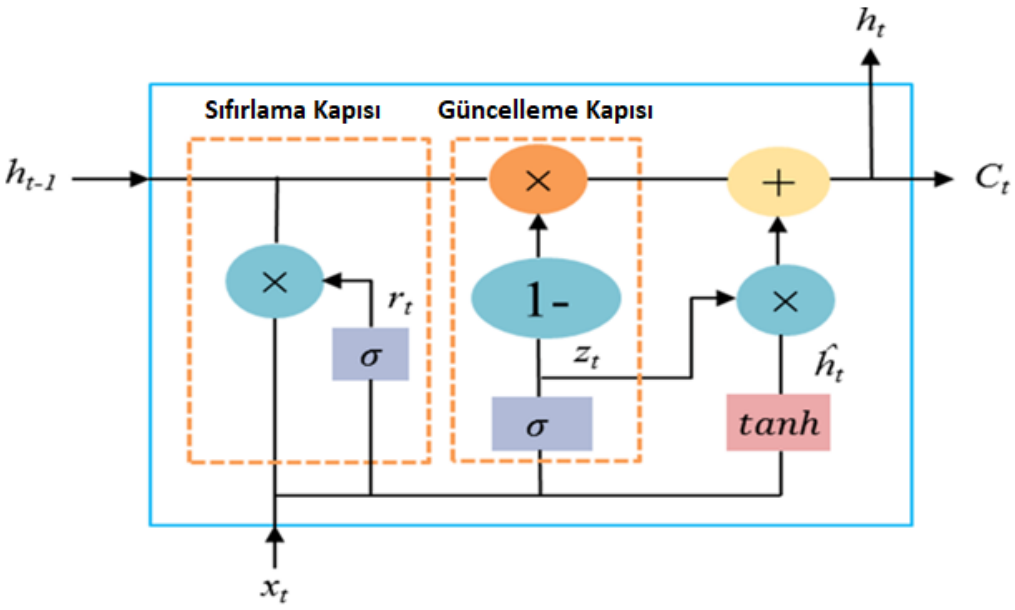
Şekil 3. 9. LSTM bellek hücresinin mimarisi (Wang vd., 2021)

Çalışmada ham kinematik veriler için yapılan denemelerde sıra (sequence) boyutu sensör sayısı olan 72 olup, sequence olarak verilen değerler temporal olmak yerine eklem bazlı sensörlerin ilişkileri olmuştur. Girdi olarak da o eklem bilgisine ait belirlenen pencere boyutundaki ham pozisyon bilgileri veya elde edilmiş öznitelik bilgileri verilmiştir. Ayrıca LSTM için çalışmada dropout için de denemeler yapılmış ama sonuçlarda olumlu bir etkisi görülmediğinden nihai parametre seçiminde devre dışı bırakılmıştır. Yine LSTM için birçok deneme yanılma sonrası karar kılınan son katman olan sınıflandırma katmanındaki modifikasyonlar da detaylandırılmıştır.

3.2.6.2. Gated recurrent unit (gru)

GRU algoritması, temelde LSTM ağlarının özellikle büyük veri setleri ile eğitimi sırasında hız performanslarını iyileştirmek için tasarlanmış algoritmalardan biridir. Parametre sayıları LSTM'den daha azdır. Şekil 3.10'da gösterilen GRU'da temelde iki kapı kullanılarak işlemler gerçekleştirilir: (i) güncelleme kapısı (update gate) ile bir sonraki duruma geçmesi gereken önceki bilgi miktarı belirlenirken, (ii) sıfırlama kapısı (reset gate) ile önceki bilgilerin ne kadarının unutulacağına karar verilmektedir.

Çalışmada GRU için de dropout mekanizması test edilmiş, ancak olumlu etkisi gözlemlenmediğinden devre dışı bırakılmıştır. Girdi olarak, eklem bazlı sensörlerin ilişkileri üzerine pencere boyutundaki ham pozisyon bilgileri veya elde edilmiş öznitelik bilgileri kullanılmıştır. LSTM ile karşılaştırıldığında GRU, daha az parametre içerdiği için hızlı bir alternatif olarak değerlendirilmiştir.



Şekil 3. 10. GRU algoritmasının mimarisi (Cho vd., 2014).

Genel olarak LSTM ve GRU için hız performansı karşılaştırmalarında GRU bir miktar daha fazla başarılı olur iken, doğruluk performansı için net bir şey söylenememektedir (Chung vd., 2014).

3.2.6.3. Attentive3d-cnn-lstm

Attentive3D-CNN-LSTM modeli, tarafımızca denemeler sonucu oluşturulan ve zaman serisi ve uzamsal-zamansal verilerin işlenmesi için geliştirilmiş bir derin öğrenme algoritmasıdır. Model, özellikle yüksek boyutlu 3D verilerdeki özelliklerin etkili bir şekilde çıkarılması ve işlenmesi hedeflenerek tasarlanmıştır. Bu mimari, 3D CNN katmanları ile LSTM katmanlarının birleşiminden oluşur ve bu iki yapı, dikkat mekanizması (attention mechanism) ile güçlendirilmiştir.

Modelin temel yapı taşlarından biri olan 3D CNN katmanları, giriş verisindeki uzamsal ilişkileri öğrenmek için kullanılır. Bu katmanlar, derin öğrenme modellerinde sıkça karşılaşılan gradyan kaybolması problemini minimize etmek için rezidüel bloklar (Residual Blocks) ile zenginleştirilmiştir. Rezidüel bloklar, ağın daha derin katmanlardan daha iyi öğrenme yapmasını sağlarken, gradyanların iletilmesini kolaylaştırır ve verimli bir öğrenme süreci sunar.

Zamansal ilişkileri modellemek için kullanılan bidirectional LSTM katmanları, verideki ardışık ilişkileri her iki yönde (geçmiş ve gelecek) analiz edebilir. Bu, özellikle

zaman serileri veya video analizlerinde daha anlamlı çıkarımlar yapmayı mümkün kılar. LSTM katmanlarının oluşturduğu çıktılar, modelin dikkat mekanizması (attention) ile birleştirilerek en kritik bilgi bölgelerinin seçilmesi sağlanır.

Modelin dikkat mekanizması, Google'ın Transformer modelinde kullanılan Query-Key-Value prensibini benimser. Bu mekanizma, verideki önemli özelliklerin vurgulanmasını sağlayarak, gereksiz bilgilerin ayıklanmasını mümkün kılar. Böylece modelin genel performansı ve doğruluğu artırılmış olur. Model aşağıdaki adımlarda girdiyi alır:

Modelin ilk aşamasında girdi katmanı, başlangıçta girdileri (örneğin, batch_size, channels, depth, height, width boyutlarındaki verileri) modelin 3D CNN katmanlarına aktarır. Bu katmanlarda, önce uzamsal özellikler çıkarılır, ardından rezidüel bloklar kullanılarak derin katmanlarda gradyan kaybı önlenir ve max pooling işlemleri ile veriler sıkıştırılarak özetlenir. Bu işlemler sonucunda, CNN katmanlarından çıkan özellikler boyut olarak sıkıştırılmış bir formatta (örneğin, batch_size, reduced_frames, reduced_features) çıkarılır.

İkinci aşamada, CNN çıktıları zaman serisi olarak bidirectional LSTM katmanlarına aktarılır. Bu katmanlar hem geçmiş hem de geleceğe yönelik zamansal ilişkileri öğrenir. LSTM'nin gizli katmanlarından elde edilen çıktılar (örneğin, batch_size, frames, lstm_hidden_size * 2 boyutlarında) daha sonra dikkat mekanizmasına aktarılır.

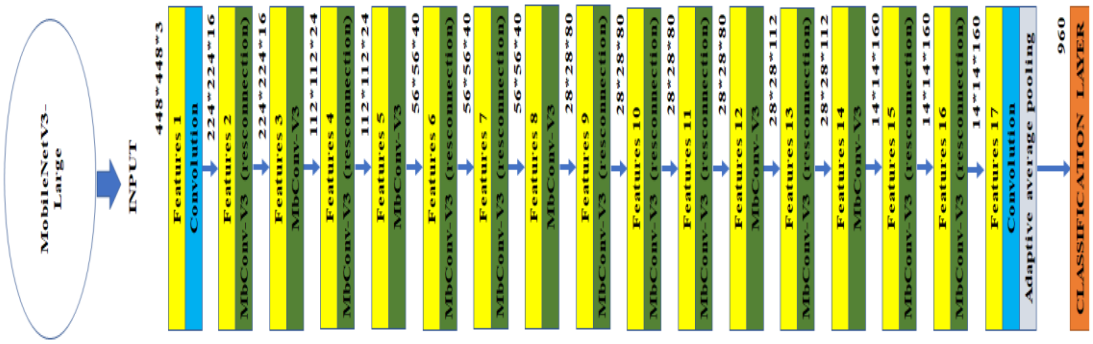
Dikkat mekanizması aşamasında, Query, Key ve Value vektörleri hesaplanarak softmax fonksiyonu ile dikkat skorları oluşturulur. Bu dikkat skorları, en kritik bilgi bölgelerini vurgulamak için kullanılır ve nihai olarak bir context vector (özet bilgi) üretilir. Bu vector, tam bağlantılı (fully connected) bir katmana aktarılır ve burada sınıflandırma işlemi gerçekleştirilir. Son aşamada, modelin tahmin ettiği sınıflar ve çıktı değerleri elde edilir.

Yukarıdaki akış çerçevesinde, Attentive3D-CNN-LSTM modeli hem uzamsal hem de zamansal özellikleri birleştirerek yüksek performans sunmakta ve zaman serisi veya video temelli analizlerde en sık kullanılan mimarilerden biri olarak öne çıkmaktadır. Bu çalışmada kullanılan model, yapay zekâ algoritmalarının çok boyutlu veri analizi için geliştirilmiş bir örneğidir.

3.2.7. Kullanılan transfer öğrenme yöntemleri

3.2.7.1. Mobilenetv3

MobileNetV3 modeli, Google ekibi tarafından kullanılacak algoritmanın düşük parametreli olması ve özellikle Android cihazlardaki performansının artırılması hedeflenerek oluşturulmuştur (Howard vd., 2019). Optimizasyon gecikmesini dikkate alarak ağ parametrelerini etkili bir şekilde azaltabilen bir yapıya sahip olan algoritma, temelde MobileNetV2 algoritmasının blok yapısını kullanır (Sandler vd., 2018). Örnek alınan blok yapı mobile inverted bottleneck layer (MBConv) olup, çok etkili ve az maliyetli matematiksel operasyonlar içerir. Şekil 3.11'de gösterilen MobileNetV3 modeli, MBConv blok yapısındaki (sıkma ve uyarma) squeeze-and-excitation (SE) katmanının konumsal bağlantı değişikliklerinin yanı sıra, sinir mimarisi araması (NAS) ve NetAdapt gibi yarı otomatik ağ araştırma optimizasyonları yardımıyla hızlı ve düşük donanım maliyeti sunan bir derin öğrenme algoritması olarak öne çıkmaktadır.



Şekil 3. 11. MobileNetV3 mimarisi

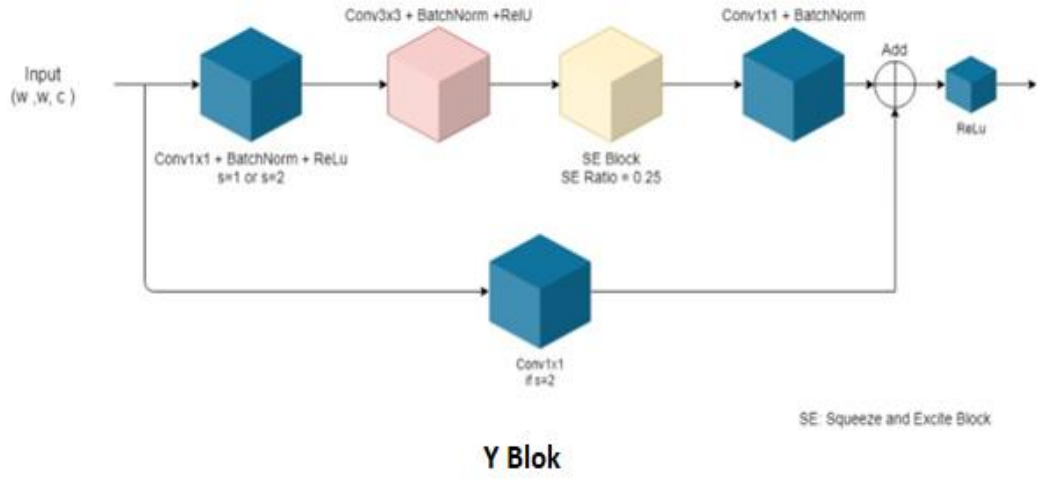
Bu çalışmada kullanılan MobileNetV3-Large modelinin toplam parametre sayısı 3,46 milyon olup, güncel derin öğrenme modelleri içinde en düşük parametreye sahip olan en başarılı algoritmalardan biridir.

3.2.7.2. Regnety-800mf

Facebook AI Araştırma ekibinden Radosavovic ve arkadaşlarının sunduğu RegNet modeli, verimlilik veya en iyi başarı isteğine göre modüle edilebilen farklı mimarileri içeren esnek bir ağ tasarım havuzu olarak geliştirilmiştir (Radosavovic vd., 2020). RegNet ile tasarım arayışı, ilgilenilen problem için farklı mimarileri denemek ve en uygun olanını bulmaktan farklıdır. RegNet, MobileNetV3 modelinde de kullanılan

NAS optimizasyonun donanımsal efektiflik arayışı kısıtlılığını kaldırarak, donanımsal efektifliğin yanı sıra istenirse en iyi başarıyı sağlayabilecek zengin model yapıları sunabilmektedir.

Şekil 3.12'de gösterilen RegNet ağ tasarımında, ağ derinliği, eğim parametresi, niceleme parametresi, darboğaz oranı ve SE oranı gibi giriş parametreleri ile yapılandırma değerleri değişerek modeller çeşitlilik kazanmaktadır. Bu çalışmada 6,05 milyon parametreye sahip oldukça etkili RegNetY-800MF modeli tercih edilmiştir. Bu model, RegNetY ailesinin kendi seviyesindeki diğer modellerine kıyasla daha yüksek bir doğruluk sağlarken, eğitim süresi ve hesaplama maliyetlerinde azalma sağlamaktadır.

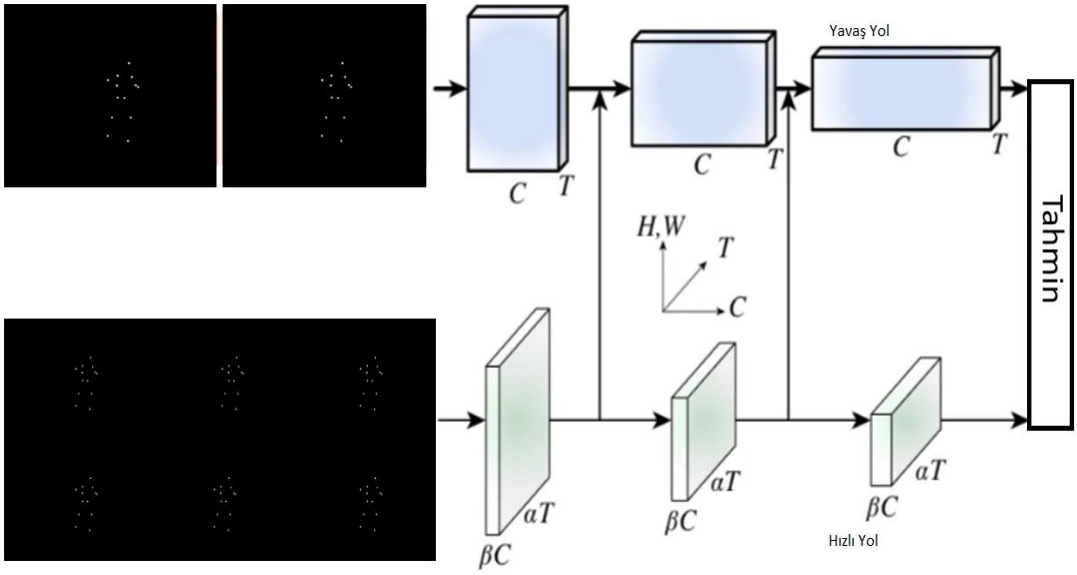


Şekil 3. 12. RegNetY mimarisi (Radosavovic vd., 2020)

3.2.7.3. SlowFast-R50

SlowFast R50, video işleme ve hareket tanıma gibi zaman içindeki hızlı ve yavaş değişimlerin olduğu verilerde etkili bir derin öğrenme modelidir (Feichtenhofer vd., 2019). Bu model, 3D evrişimsel sinir ağı (3D CNN) yapısını kullanır ve özellikle hızlı hareketlerin ve yavaş değişimlerin aynı anda işlendiği durumlarda yüksek performans sağlar. SlowFast modeli, hem düşük kare hızında (yavaş) hem de yüksek kare hızında (hızlı) iki paralel yol kullanarak, farklı hızlarda gerçekleşen özellikleri öğrenir. Bu yapı, modelin hareketleri daha doğru bir şekilde algılamasına olanak tanır. (i) Slow Path (Yavaş Yol), videonun yavaş hareket eden ve geniş zamansal özelliklerini çıkarır. Yavaş yol, düşük kare hızında çalışarak daha uzun süreli ve geniş bağlamlı bilgilere odaklanır. Bu

sayede model, sahnede yavaş gerçekleşen değişimlerin detaylarını daha iyi öğrenir. (ii) Fast Path (Hızlı Yol) ise videonun daha kısa süreli ve hızlı hareket eden bileşenlerini öğrenmek için yüksek kare hızında çalışır. Bu yol, zamansal çözünürlüğü yüksek olan bir veri akışını işler, böylece hızlı hareket eden nesne veya olayları daha hassas şekilde algılar. Slow ve Fast yolları arasında bilgi paylaşımı gerçekleşir. Bu bilgi alışverişi, hızlı yolun elde ettiği kısa süreli hareket bilgilerinin yavaş yolda kullanılan daha uzun süreli özelliklerle birleşmesini sağlar. Bu birleşim, modelin daha bütüncül bir hareket tanıma kapasitesi elde etmesine katkıda bulunur (Fan vd., 2021). Şekil 3.13'de gösterilen SlowFast R50 modeli, her iki yolda da ResNet-50 tabanlı bir mimari kullanır.



Şekil 3. 13. SlowFast R50 ağ mimarisi (Feichtenhofer vd., 2019)

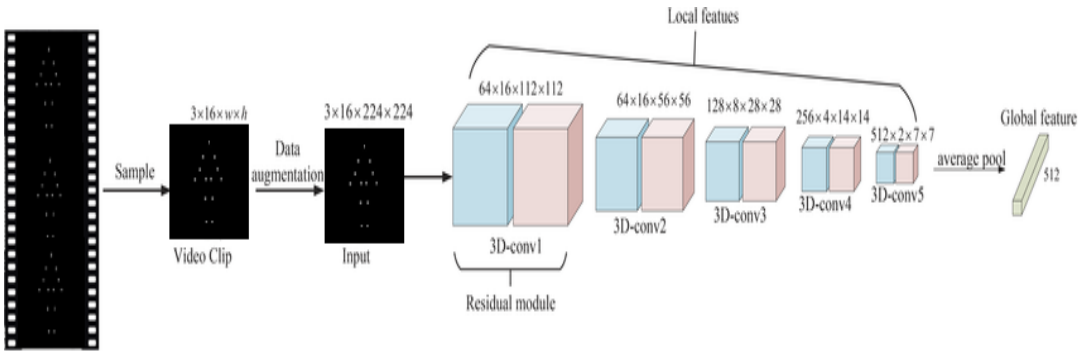
Bu omurga, derin yapısının avantajları sayesinde yüksek doğruluk oranına katkıda bulunur. ResNet-50'nin residual blok yapısı, katmanlar arasındaki bilgi kaybını azaltarak daha verimli bir öğrenme süreci sağlar (Fan vd., 2021).

3.2.7.4. Resnet 3d 18

ResNet-3D-18, görüntü ve video işleme gibi üç boyutlu veriler üzerinde çalışan bir derin öğrenme modelidir (Al-Khater vd., 2024). Bu model, 2D ResNet'in video veya ardışık görüntüler gibi zaman içindeki değişimleri de kapsayan 3D verilere uyarlanmış bir versiyonudur. ResNet, temel olarak derin sinir ağlarında katman sayısı arttıkça ortaya çıkan gradyan kaybolması (vanishing gradient) problemini çözmek için geliştirilmiştir.

ResNet-3D-18, bu avantajları 3D veri üzerinde uygulayarak hem verimli hem de etkili bir öğrenme sağlar (Hara vd., 2018).

ResNet-3D-18, toplamda 18 katmandan oluşan bir yapıdır ve her bir katmanda 3D evrişim (convolution) işlemi gerçekleştirir. Geleneksel evrişimsel sinir ağlarında kullanılan 2D filtreler yerine, 3D ResNet'te her bir filtre 3D olarak tanımlanır; bu sayede hem uzamsal (x ve y) hem de zamansal (t) bilgiyi aynı anda işler. Bu özellik, videolardaki hareketleri ve nesnelerin zaman içerisindeki değişimlerini anlamada büyük bir avantaj sağlar. Model, 3D evrişimler sayesinde görüntülerin ardışık karelerindeki ilişkiyi yakalayarak zaman boyutunda güçlü bir özellik çıkarımı yapar (Hara vd., 2018). Şekil 3.14'te görülen ResNet-3D-18'in temel yapı taşı olan artık bloklar (residual blocks), modeli derinleştirirken öğrenmeyi kolaylaştıran önemli bir bileşendir.



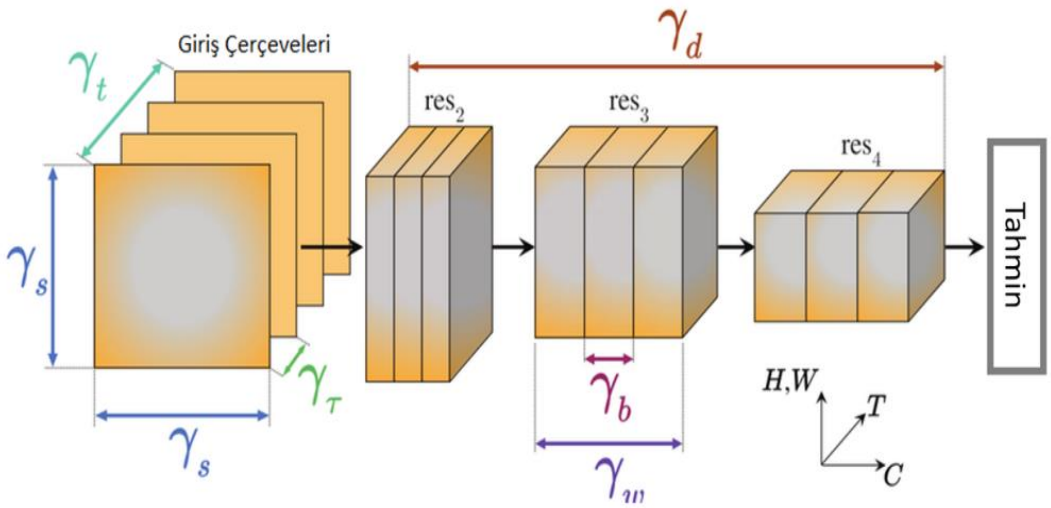
Şekil 3. 14. 3D-ResNet-18 ağ mimarisi (Xue vd., 2020)

Her blok, girdiyi bir sonraki katmandaki çıkışa doğrudan aktarır (skip connection), bu da modelin eğitimi sırasında bilgi kaybını azaltır. Bu yapılar sayesinde, modelin daha derin katmanlarında bile gradyan sinyalleri güçlü kalır ve böylece derin yapıların eğitilmesi kolaylaşır (He vd., 2016).

3.2.7.5. X3D-Medium

X3D modeli, video tanıma görevleri için optimize edilmiş ve çeşitli eksenlerde kademeli genişleme yöntemine dayanan verimli bir derin öğrenme mimarisidir (Feichtenhofer vd., 2020). Başlangıç olarak küçük ve sade bir 2D görüntü sınıflandırma ağından yola çıkarak, bu modeli uzaysal ve zamansal boyutlara doğru genişletir ve böylece 3D bir yapıya ulaşır. X3D mimarisinin temel amacı, doğruluk ve hesaplama maliyeti arasında optimal bir denge sağlamaktır; bu amaçla her adımda yalnızca tek bir

ekseni genişletir ve minimum hesaplama maliyeti ile yüksek doğruluk elde eder. Genişleme süreci, belirli eksenlerde aşamalı genişletme ve gerektiğinde daraltma adımlarına dayanır. Bu süreçte ağ yapısı, sırayla zamansal uzunluk (γt), kare hızı (γτ), uzaysal çözünürlük (γs), genişlik (γw), dar alan genişliği (γb) ve derinlik (γd) gibi faktörler üzerinden genişletilir. İlk aşamada model, çok düşük hesaplama maliyeti ile başlatılır ve ardından belirli bir doğruluk-hedef hesaplama bütçesi için eksenler sırayla genişletilir; her genişletme adımında doğruluk ve hesaplama maliyetine göre test edilir ve en uygun eksen seçilir (Feichtenhofer vd., 2020). Şekil 3.15'te gösterilen X3D, basit bir 2D ResNet tabanlı model olan X2D'den başlar; başlangıç modeli video verisi yerine tek bir görüntü çerçevesini kullanarak işlem yapar ve böylece bir video tanıma görevine uygun olacak şekilde genişletilmeye başlanır.



Şekil 3. 15. X3D-medium ağlarının çerçevesi (Feichtenhofer vd., 2020)

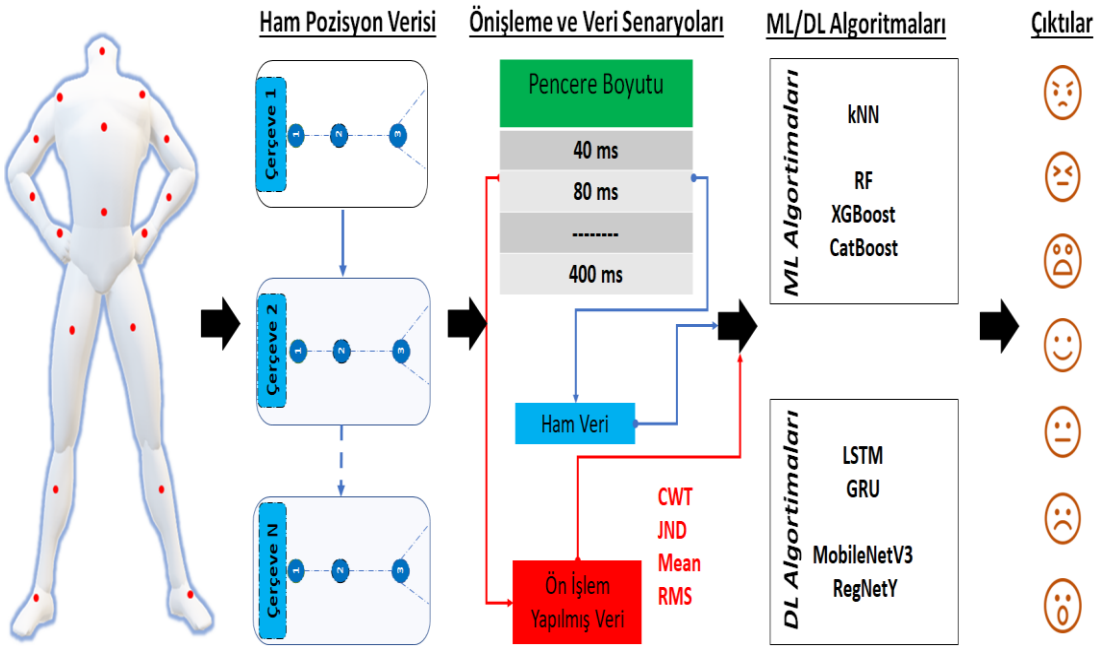
İlk adımlarda, ağın zamansal uzunluğu ve kare hızı artırılarak zamansal çözünürlük geliştirilir, bu sayede model hareket bilgilerini daha detaylı işleyebilir hale gelir. Ardından, uzaysal çözünürlük artırılır ve daha yüksek uzaysal ayrıntı yakalanır. Kanal genişliği ve dar alan (bottleneck) genişliği artırılarak kanal boyutu ve veri aktarım kapasitesi genişletilir; bu genişlemeler, özellikle hesaplama maliyetini düşük tutmak için etkili olur. Son olarak, ağın derinliği artırılarak daha karmaşık özellikleri işleyebilmesi sağlanır ve bu adım genellikle diğer eksenlerin genişletilmesinden sonra gerçekleştirilir. X3D modelinde genişletme işlemi, ileri yönlü genişletme ve gerektiğinde geri yönlü daraltma adımlarıyla gerçekleştirilir; hedef hesaplama bütçesine ulaşana kadar genişletme yapılır ve model bütçeyi aşarsa, hesaplama maliyetini düşürmek amacıyla geri yönlü

daraltma adımları uygulanır. Bu, özellikle düşük maliyetli uygulamalar için önem taşır. Model, ResNet yapı taşlarını kullanır ve her genişleme adımında test edilir. Kanal başına ayrık evrişimler (channel-wise separable convolutions) kullanılarak, 2D mobil görüntü sınıflandırma ağlarına benzer şekilde hafif bir yapıya sahip olması sağlanır. Bu yöntem, her bir eksen genişletmesi için yalnızca tek bir modelin eğitilmesini ve her genişletme adımında hesaplama maliyetinin minimumda tutulmasını sağlar (Feichtenhofer vd., 2020).

3.3. Önerilen Yaklaşımlar

3.3.1. Ham kinematik veri seti için önerilen yaklaşım

Bu veri seti duygusal olarak etiketlenmiş en büyük kinematik veri setlerinden biri olup, yapay olarak sentezlenmemiş ve gürültüsüz olduğu için değişken çerçeveli pencerelere bölündüğünde çok sayıda anlamlı örneğe sahip olabilmektedir. Bu nedenle bu çalışmada ham veriler, belli sayılarda pencerelere bölünerek veri artırımı yoluna gidilmiştir. Çerçeve içeren pencerelerin kullanım nedeni sinyalin en küçük tutarlı kısımlarından faydalanarak, zaman içinde değişen istatistiksel özellikleri daha iyi yakalayabilmektir (Sherstinsky vd., 2020; Diwan vd., 2023). Önerilen metoda ait işlem süreci Şekil 3.16'da özetlenmiştir.



Şekil 3. 16. Ham kinematik veri seti için uygulanan adımların diyagramı

Toplamda 5, 10, 20, 30, 40 ve 50 çerçeveden oluşan pencereler oluşturularak altı adet yeni veri seti alt kümesi elde edilmiştir.

Tablo 3. 3. Yeni pencerelerle elde edilen veri seti kümeleri

Veri kümesi bölme senaryosu	Yeni pencere çerçeve boyutu	Yeni pencere boyutu (milisaniye cinsinden)	Elde edilen yeni veri setindeki dosyalar
1	5	40	252.454
2	10	80	125.949
3	20	160	62.635
4	30	240	41.521
5	40	320	30.951
6	50	400	24.626

Denenen çerçeve sayıları ve elde edilen yeni veri setlerinin bilgisi Tablo 3.3'te paylaşılmıştır. Bvh uzantılı kinematik dosyalarını işlemek için Python programlama dilinde kullanılmak üzere geliştirilmiş olan bvhtoolbox modülü çalışmaya uygun bir şekilde modifiye edilerek kullanılmıştır [1].

3.3.1.1. Veri kullanım senaryoları

Bu çalışmada verilerin hem ham hali hem de öznitelik uygulandıktan sonraki hali denemelerde kullanılmıştır. Veriler gürültüsüz olduğu için sadece veri aralığını ölçeklendirmek adına ML sınıflandırıcılarında normalizasyon gerçekleştirilmiş, DL sınıflandırıcılarında ise ufak ta olsa bir performans kaybından dolayı normalizasyon tercih edilmemiştir. ML tabanlı sınıflandırıcılar için tercih edilen normalizasyon min-max normalizasyon yöntemi olmuştur.

3.3.1.2. Ham kinematik verisi

Ham veriler eşit sayıda çerçeveden oluşur hale getirildiğinde girdilerin aynı boyutta olma sorunu ortadan kalktığı için sorunsuz bir şekilde ML ve DL algoritmalarında kullanılabilmiştir. Örneğin 5 çerçeve kullanılarak oluşturulan her bir pencere için 72

[1] https://pypi.org/project/bvhtoolbox/

düğümden 3 koordinat için pozisyon bilgisi elde edildiğinden oluşan matrisin boyutu 72*5*3 şeklindedir. Bu matris kullanılan ML algoritmalarında çarpılarak tek boyuta (1*1080) indirgenirken, DL algoritmalarında sekans ve sütun bilgisine (72*15) dönüştürülmektedir. Burada sekans sayısı olarak normalde zamana bağımlı çerçeve sayısı yerine kinematik düğüm sayısı olan 72 olarak tercih edildiğinde daha başarılı sonuçlar alınmıştır.

3.3.1.3. Veriden özellik çıkarma

Dosyaların ham halini sınıflandırmada kullanmak bazı durumlarda tercih edilmemektedir. Örneğin ham veriler, bazen boyutsal maliyetleri nedeniyle zaman karmaşıklıkları baş edilemez hale gelebilmekte bazen de istenilen doğruluk hedefi için yetersiz olabilmektedirler. Hem maliyeti azaltmak hem de başarıyı artırmak için duygu tanıma literatüründe öz nitelik çıkarımı için birçok zaman, frekans ve istatistiksel tabanlı parametre kullanımı mevcuttur (Oğuz ve Ertuğrul, 2022; Ahmed vd., 2020). Bu çalışma için en kazançlı ve maliyetsiz özniteliklerin belirlenmesi adına RF ve son derece rastgele ağaçlar (extremely randomized trees) algoritmaları yardımıyla onlarca parametre içinden özellik seçimi (feature selection) işlemleri gerçekleştirilmiştir (Saganowski vd., 2022; Geurts vd., 2006). Özellik önemi sıralaması ile, sınıflandırmada hangi özelliklerin daha önemli olduğu ve hangilerine daha fazla odaklanılacağı belirlenmiştir. Karar verilen öznitelik parametreleri ortalama (mean), kök ortalama kare (RMS), sürekli dalgacık dönüşümü (continuous wavelet transformation, CWT) ve joint neighborhood distance (JND) dörtlüsünün kombinasyonuyla elde edilmiştir. Her bir eksendeki her bir koordinat için çıkarılan bu öznitelikler, üç boyutsal koordinat düzleminde (x, y, z) toplamda 12 tane olacak şekilde elde edilmiştir.

Birinci öznitelik, basitçe her bir penceredeki çerçevelerin ortalamaları alınarak elde edilmekte olup, Denklem 3.1'de gösterilmiştir.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{3.1}$$

Burada N çerçeve sayısı ve $x = \{x_1, x_2, \ldots, x_n\}$, çerçeve içindeki elemanlar olmak üzere ortalama $\bar{x}$ ile sembolize edilir. İkinci öznitelik yine ortalamayla oldukça

ilişkili ama en az onun kadar değerli ve bağımsız fayda sağlayan RMS yöntemi olup, notasyonu Denklem 3.2'de gösterilmiştir.

$$RMS = \sqrt{\frac{\sum_{i=1}^{N} (x_i)^2}{N}} \tag{3.2}$$

Üçüncü öznitelik yine sabit olmayan sinyallerden anlamlı spektral ve zamansal bilgileri çıkarma yeteneğine sahip bir algoritma olan CWT dönüşümüdür (Jovic vd., 2015). Ana bir dalgacık fonksiyonu seçilerek, sinyalin seçilen pencere aralığında ölçek (scale) ve kayma (shift) işlemleri gerçekleştirilerek anlamlı uyuşmalardan faydalanılan bu dönüşüm, Denklem 3.3'te gösterilmiştir.

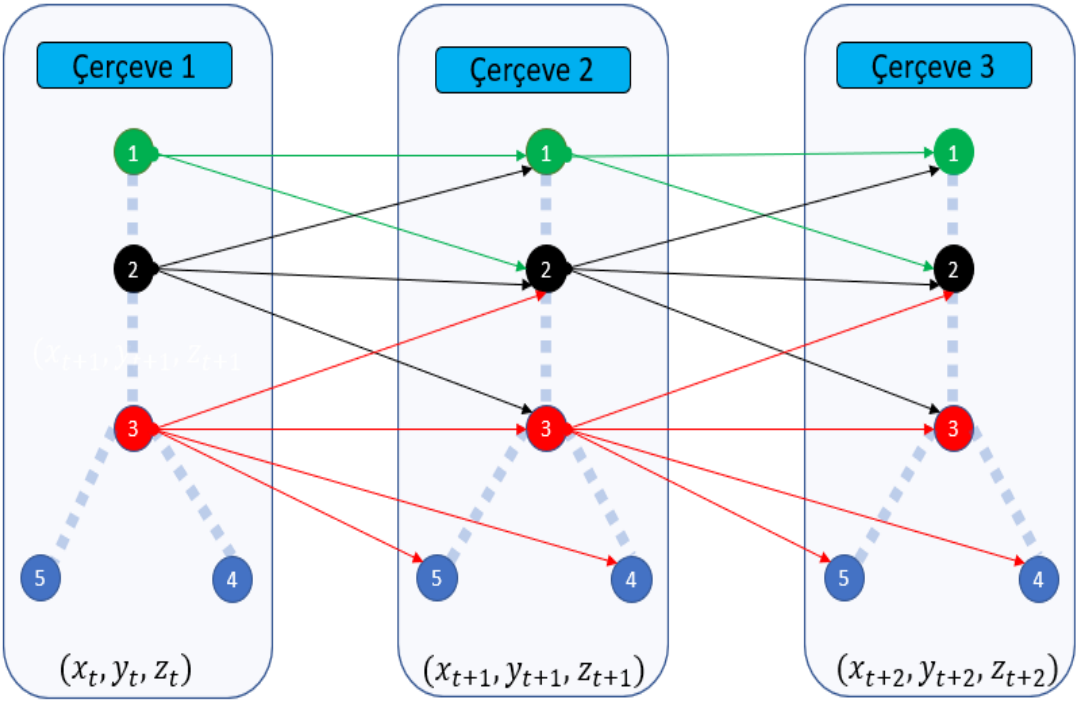
$$CWT\ (\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t)\ \psi^*(\frac{t-\tau}{s})dt \tag{3.3}$$

Burada τ shift parametresi, s scale değeri, ψ(t) ana dalgacık fonksiyonudur. Bu hesaplama yine Python'da "PyWavelets" kütüphanesi kullanılarak gerçekleştirilmiş, s değeri 3 ve dalgacık fonksiyonu olan ψ(t) değeri için de "morlet" seçilmiştir.

Son öznitelik değerimiz olan JND, her bir çerçeve içindeki her bir eklem değerinin diğer çerçevelerde kendisine ve kendisine komşu olan eklemlere olan uzaklıklarının ortalaması alınarak elde edilir. Bunun için her bir eklemin sensör bazlı komşuluk matrisinin listesi çıkarılmıştır. Bu çalışmada kullanılan JND yaklaşımının her bir eklem için hesaplanmasının matematiksel analizi Denklem 3.4'teki gibidir.

$$JND = \frac{1}{n} \sum_{i,\ t=1}^{N,\ \mathrm{Jn}} (|x_{i,\ t} - x_{k,\ t+1}|) \tag{3.4}$$

Denklem 3.4'te N ilgili penceredeki çerçeve sayısını ifade eder. $Jn = |x_1, x_2, \ldots, x_{\mathrm{Jn}}|$ verideki referans eklem düğümü noktasına ait komşuluk matrisinin uzunluğunu ifade eder ve komşuluk matrisinin boyutu her bir eklem için değişiklik gösterir. Denklemde x'e bağımlı i parametresi N'nin bir alt kümesi olup referans düğüm noktasını, k parametresi ise düğümün, k. komşusunu ifade etmektedir. Yine denklemde x'e bağımlı t parametresi Jn'nin bir alt kümesi olup referans düğüm noktasının bulunduğu zamanı, t+1 ise bir sonraki çerçeve (frame) zamanı gösterir. JND için örnek bir hesaplama süreci Şekil 3.17'de gösterilmiştir.



Şekil 3. 17. JND ile eklem düğümleri üzerinden özellik çıkarımı

Şekil 3.17'de bir pencerede bulunan ilk üç eklem düğümü için hesaplamalar görselleştirilmiştir. JND özniteliği hesaplanırken hem temporal hem mekânsal (spatial) anlamda her bir eklemin kendisinin ve komşusunun ilişkisini hesaba katılır. Yeşil ile gösterilmiş 1. düğümün iki eklem komşusu olup, komşu matrisi $Jn_1 = |x_1, x_2|$ den oluşmaktadır. Yine devamında siyah ile gösterilmiş 2. düğümün komşuluk matrisi $Jn_2 = |x_1, x_2, x_3|$ ve kırmızı ile gösterilmiş 3. düğümün komşuluk matrisi ise $Jn_3 = |x_2, x_3, x_4, x_5|$ ile gösterilmektedir. Her bir eklem için çerçeve bazında komşulukların ortalaması alınarak Denklem 3.4'teki gibi hesaplama işlemi yapılır.

JND yaklaşımına benzer hesaplamalar geçmişten beri çeşitli kombinasyonlarla denenmektedir (Chun-Lin vd., 2010). Eklem bilgisi fazla olan yüksek çözünürlüğe sahip veri setlerinde tüm eklemlerin birbiri ile olan ilişkilerinin hesaplanması çok maliyetli olduğundan, bu makalede önerilen özellik çıkarma yönteminde uzamsal anlamda sadece eklemin kendisi ve kendisine komşuları temporal anlamda dikkate alınmaktadır. Ayrıca JND bazında tüm eksenlerdeki koordinatların öklitsel hesaplaması yapılarak totalde tek öznitelik çıkarılması yerine her bir koordinattan alınan JND özniteliği tekil olarak birer parametre şeklinde alınmıştır. Tekil olarak alınan parametrelerin performans anlamında daha fazla kazanç sağladığı görülmüştür.

Tablo 3. 4. Belirli pencere boyutları için dosya başına öznitelik çıkarım maliyeti

Veri seti senaryosu	Örneklem büyüklüğü FE öncesi	Örneklem büyüklüğü FE'den sonra	FE için ortalama süre (milisaniye cinsinden)
1	72*15	72*12	63.83
2	72*30		65.19
3	72*60		68.38
4	72*90		72.1
5	72*120		75.15
6	72*150		79.29

Tablo 3.4'te belirli pencere boyutları için dosya başına öznitelik çıkarım maliyeti hesapları gösterilmiştir. En düşük boyut olan 5 çerçeveli ve 40 milisaniyelik ebata sahip senaryo ile en yüksek boyut olan 50 çerçeveli ve 400 milisaniyelik ebata sahip senaryolar arasında tüm eksenler hesaba katıldığında maliyet farkı 16 milisaniyeden daha azdır. Hesaplanan dosyaların ebatı arttıkça hesaplama maliyeti başlangıç maliyetine göre daha az maliyetli hale gelmektedir. Tabi bu çalışmada kullanılan işlemcinin gücü ve seçilen çerçeve boyutuna göre maliyet fayda analizi değişiklik gösterecektir.

3.3.1.3. Veri için kullanılan yöntemler ve parametreler

Kinematik veri setine uygulanan metotların farklı yapıdaki algoritmalarla sınıflandırılması ile çözülmek istenen görevin analizi daha kapsamlı gerçekleştirecek ve sonuçlar daha sağlıklı olacaktır. Bu nedenle ilk etapta basit ve etkili bir ML algoritması olan kNN ve ağaç tabanlı (tree-based) ML algoritmaları olan rastgele orman (RF), aşırı gradyan artırma (XGBoost) ve CatBoost algoritmaları ile testler yapılmıştır. Bir sonraki adımda state of art sayılabilecek, düşük parametreli ve oldukça hızlı DL tabanlı daha güçlü algoritmalar ile denemeler gerçekleştirilmiştir. RNN tabanlı LSTM ve GRU modellerinin yanı sıra CNN tabanlı MobileNetV3-Large ve RegNetY-800MF ile de bu çalışmada detaylı analizler gerçekleştirilmiştir. Seçilen DL algoritmalarının canlı tepki içeren sistemlere uyumlu olabilecek şekilde, mobil işlemcilerin kaldırabileceği efektiflikte olmasına öncelik verilmiştir.

Tablo 3. 5. Kullanılan ML algoritmaları için uygulanan parametreler

Algoritma Adı	Parametre adı ve değerler
kNN	n_neighbors=5, distance metric="Euclidean"
RF	n_estimators=100, criterion for split="Gini Impurity"
CatBoost	iterations: 20, learning_rate: 0,1, task_type="GPU"
XGBoost	n_estimators=1000, learning_rate: 0,1, tree_method='gpu_hist'

Tüm derin öğrenme ve transfer learning modellerde eğitim sırasında AdamW optimizasyon algoritması kullanılmış ve başlangıç öğrenme oranı 0.001 olarak belirlenmiştir. Öğrenme oranı, adım boyu (step size) 6 ve gamma değeri 0.9 olacak şekilde otomatik olarak azaltılmıştır. Batch boyutu tüm modeller için 512 olarak seçilmiş ve maksimum epoch sayısı 150 olarak uygulanmıştır. Bu ayarlar, modellerin eğitim sürecinin hem verimli hem de performans açısından optimize edilmesini sağlamıştır.

Tablo 3. 6. Kullanılan DL algoritmaları için uygulanan hiper parametreler

Parametre adı	GRU	LSTM	MobileNetV3-Large	RegNetY-800MF
Batch boyutu	512			
Epoch sayısı	150			
Giriş boyutu	(Ham veriler için = (çerçeve sayısı * 3)) VEYA (özellik çıkarılan veriler için = 12)			
Etiket yumuşatma	0,11			
Öğrenme oranı	0,001			
Öğrenme_oranı_azalması	Adım boyu= 6, gamma= 0.9			
Optimize edici	AdamW			
Sequence uzunluğu	72			
Yığılmış katman boyutu	2		–	
Gizli durum boyutu	512			
Değiştirilmiş sınıflandırma katmanı	[0]: SiLu, [1]: Linear1 (giriş = gizli durum boyutu, çıkış = 128), [2]: SiLu, [3]: Linear2 (giriş = 128, çıkış = 7)		[0]: Linear1 (giriş = (960 için "MobileNetV3", ve 784 için "RegNetY"), çıkış = 512), [1]: Hardswish, [2]: Dropout (p = 0.25), [3]: Linear2 (giriş = 512, çıkış = 7)	
Toplam Parametreler (milyon)	2,45 ile 2,66 arasında	3,25 ile 3,52 arasında	3,46	6,05

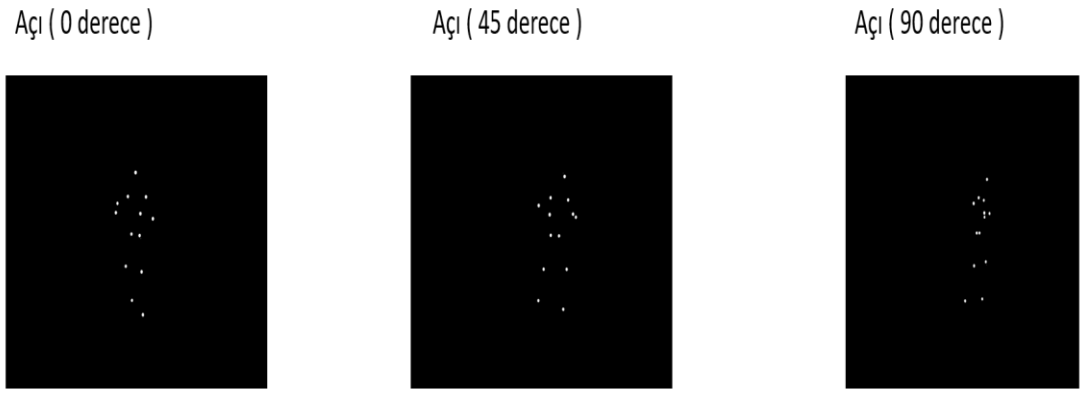
Bu çalışmadaki ham kinematik veri seti için kullanılan makine öğrenmesi ve derin öğrenme modellerinin eğitim süreçleri ve hiper parametre ayarları, Tablo 3.5'te ve Tablo 3.6'da detaylı bir şekilde sunulmuştur.

Giriş boyutları, kullanılan verilerin türüne göre ayarlanmıştır. Ham veriler için giriş boyutu çerçeve sayısına bağlı olarak (örneğin, çerçeve sayısı * 3) tanımlanırken, öznitelik çıkarılan verilerde bu boyut 12 olarak belirlenmiştir. Dizi uzunluğu (sequence length) LSTM ve GRU modelleri için 72 olarak seçilmiş ve bu, eklem bazlı sensörlerden elde edilen ardışık verilerin modellenmesini sağlamıştır. Ayrıca, LSTM ve GRU modelleri için gizli durum boyutu 512 olarak belirlenmiş ve bu modeller iki yığılmış katman (stacked layer) ile yapılandırılmıştır. Dropout mekanizması hem LSTM hem de GRU modelleri için test edilmiş, ancak sonuçlarda olumlu bir etkisi görülmediğinden nihai hiper parametrelerde devre dışı bırakılmıştır. Bu modellerin sınıflandırıcı katmanları SiLU aktivasyon fonksiyonu ile yapılandırılmış ve sınıflandırma katmanında giriş boyutu ile çıkış boyutu arasındaki bağlantılar optimize edilmiştir.

MobileNetV3 ve RegNetY gibi mobil tabanlı CNN modelleri için giriş boyutları optimize edilmiş ve modelin sınıflandırıcı katmanında farklı teknikler uygulanmıştır. Özellikle, MobileNetV3-Large modeli toplam 3,46 milyon parametreye sahipken, RegNetY-800MF modeli 6,05 milyon parametreye sahiptir. MobileNetV3'te, mobil inverted bottleneck layer (MBConv) ve squeeze-and-excitation (SE) katmanlarının yanı sıra, Hardswish aktivasyon fonksiyonu ve 0,25 oranında dropout mekanizması kullanılmıştır. Benzer şekilde, RegNetY modeli, SE oranı, darboğaz oranı ve ağ derinliği gibi parametrelerin optimize edilmesiyle eğitim sürecinde donanımsal verimliliği ve model başarısını artırmayı hedeflemiştir. Sınıflandırıcı katmanlarda, giriş boyutu ve aktivasyon fonksiyonları modellenerek son katman doğruluklarının optimize edilmesi sağlanmıştır. Ayrıca, tüm modellerde aşırı öğrenmeyi (overfitting) engellemek ve modelin daha genelleştirilebilir hale gelmesini sağlamak için 0,11 değeri ile etiket yumuşatma (label smoothing) uygulanmıştır. Bu, modelin sınıflandırma sırasında daha az kendine güvenli kararlar vermesine olanak tanımış ve dengesiz sınıf dağılımlarında daha dengeli bir performans sergilemesini sağlamıştır. Eğitim sürecinin sonunda, en iyi performans sağlayan hiper parametrelerle eğitilen model ağırlıkları kaydedilmiş ve nihai değerlendirmeler bu ağırlıklarla gerçekleştirilmiştir. Tüm bu hiper parametre ayarları, modellerin eğitim sürecinde istikrarlı bir öğrenme sağlamalarını ve karşılaştırmalı değerlendirmelerde adil bir kıyaslama yapılmasını mümkün kılmıştır.

3.3.2. DEMOS video veri seti için önerilen yaklaşım

Bu çalışmada, daha önce oluşturulmuş ve yayımlanmış olan video şeklindeki DEMOS veri seti üzerinde analizler gerçekleştirilmiş ve duygu tanıma süreci için özgün bir yöntem önerilmiştir. DEMOS veri seti, altı temel duygu (anger, disgust, fear, happiness, neutral, sadness) için Şekil 3.18'de gösterilen üç farklı açıdan (0°, 45°, 90°) çekilmiş toplam 2664 video klibi içermektedir. Her bir klip, iki saniye uzunluğunda olup 720x540 piksel çözünürlükte kaydedilmiştir. Veri setinde yer alan videolar, oyuncuların belirli duyguları yansıttığı beden hareketlerini temsil eden görsel kayıtlar sunmaktadır. Bu veri seti, yazarları tarafından kapsamlı bir biçimde hazırlanmış olup, veri toplama, işleme ve veri artırma (data augmentation) işlemleri makale kapsamında gerçekleştirilmiştir. Bu çalışmada, veri setine herhangi bir ek veri artırma işlemi uygulanmamış, yalnızca mevcut videolar analiz edilmiştir.

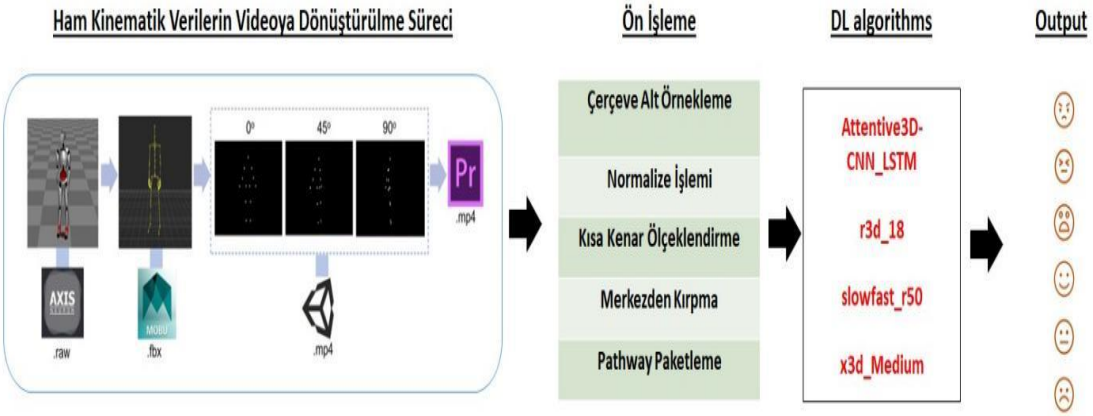


Şekil 3. 18. DEMOS veri setinin farklı açılardan görünümü

Önerilen yöntem, DEMOS veri setini doğrudan kullanarak, ham kinematik verilerin video haline dönüştürülmesi ve bu videolar üzerinde duygu tanıma gerçekleştirilmesi sürecini kapsamaktadır. Veri işleme adımlarında videolar belirli oranlarda alt örneklenmiş ve normalize edilmiştir. Görüntülerin kısa kenarı belirli bir boyuta ölçeklendirilmiş ve merkezden kırpma işlemi yapılarak gereksiz alanlar çıkarılmıştır. Ayrıca, modelin zamansal ilişkileri daha iyi anlaması için pathway paketleme yöntemi uygulanmıştır.

Analiz sürecinde, duygu tanıma amacıyla dört farklı derin öğrenme modeli kullanılmıştır. Bu modeller arasında uzamsal ve zamansal dikkat özelliklerini bir arada kullanan Attentive3D-CNN-LSTM, hareketlerin zamansal takibini gerçekleştiren 3 boyutlu ResNet tabanlı r3d_18, yavaş ve hızlı hareketleri paralel yollarla analiz eden slowfast_r50 ve farklı eksenlerde kademeli genişleme sağlayarak video tabanlı

görevlerde optimal performans sunan x3d_Medium modeli bulunmaktadır. Bu modeller aracılığıyla öfke, tiksinme, korku, mutluluk, nötr ve üzüntü gibi çeşitli duygusal durumların tanımlanması gerçekleştirilmiştir. Önerilen metoda ait işlem süreci Şekil 3.19'da özetlenmiştir.



Şekil 3. 19. Çalışmada uygulanan adımların diyagramı

3.3.2.1. Video verisi için ön işleme süreçleri

Videolar üzerinde gerçekleştirilen ön işleme süreci, verilerin modele uygun hale getirilmesi için ardışık bir dizi işlemden oluşmaktadır. İlk olarak, Çerçeve Alt Örnekleme (Temporal Subsampling) ile videolar belirli bir kare sayısına indirgenmektedir (UniformTemporalSubsample). Bu sayede, modelin işleyebileceği uygun boyuta getirilerek video verisinin karmaşıklığı azaltılmaktadır. Ardından, Merkezden Kırpma (CenterCropVideo) işlemi uygulanarak her video sabit bir boyuta getirilmekte ve bu işlem ile modelin girdi boyutları standartlaştırılmaktadır.

Normalizasyon aşamasında, videodaki her bir karede bulunan piksel değerleri 0 ile 1 arasında normalize edilmektedir (normalize(x) = x / 255.0). Bu normalizasyon, modelin daha hızlı ve verimli öğrenmesini sağlamak amacıyla veri dağılımını dengeler ve öğrenme sürecini kolaylaştırır.

Kısa Kenar Ölçeklendirme (ShortSideScale) adımı ise videoların kısa kenarının belirlenen bir boyuta ölçeklendirilmesini sağlamaktadır. Bu işlem, modelin işleyebileceği tutarlı bir boyut elde etmeye yardımcı olur ve veri çeşitliliğiyle başa çıkmayı kolaylaştırır.

Bu işlemleri takiben, Yol Paketleme (PackPathway) ile videolar SlowFast modeline uygun olacak şekilde iki farklı yol olarak ayrılmaktadır. Fast Pathway, videonun tamamını içererek yüksek çözünürlükte bilgi sunarken, Slow Pathway ise kareler

arasından belirli aralıklarla örneklem alarak daha düşük frekansta bilgi taşımaktadır. Bu sayede model hem hızlı hareketleri hem de genel bağlamı öğrenme kabiliyeti kazanır.

Son olarak, bu ön işleme adımları, Compose kullanılarak ardışık bir dizi halinde birleştirilmekte ve videoya uygulanmaktadır. ApplyTransformToKey işlemi ile bu ön işlemler belirli anahtarlar (örneğin, "video" anahtarı) üzerine uygulanmakta ve tüm işlemler verinin belirlenen aşamalarla hazırlanmasını sağlamaktadır. Bu kapsamlı ön işleme süreci, modelin veriyi doğru bir şekilde anlamasını ve verimli bir şekilde işleyebilmesini hedeflemektedir.

3.3.2.2. Video verisi için kullanılan yöntemler ve parametreler

DEMOS video veri seti üzerinde daha kapsamlı analizler ve güvenilir sonuçlar elde etmek amacıyla farklı derin öğrenme tabanlı sınıflandırma yöntemleri kullanılmıştır. Sequential (ardışık) verilerle çalışabilen güçlü bir RNN yapısı olan LSTM, gradyan kaybı ve gradyan patlaması sorunlarını çözmek üzere özel hücre yapısı ile seçilmiş ve videolardaki zaman akışını etkili bir şekilde analiz etme yeteneği sağlamıştır. Ayrıca, üç boyutlu veri üzerinde hem uzaysal hem de zamansal özellik çıkarımı yapabilen ResNet-3D-18 modeli, zaman içerisindeki değişimleri takip ederek videolardaki hareketleri detaylı bir şekilde anlamlandırmıştır.

Tablo 3. 7. Kullanılan veri setine ait bazı istatistiksel bilgiler

	Duygusal Durum						
	Öfke	Tiksinti	Korku	Mutluluk	Nötr	Üzüntü	Toplam
Eğitim	315	318	354	327	237	306	1857
Validasyon	69	69	75	69	51	66	399
Test	69	69	78	72	51	69	408
Toplam	453	456	507	468	339	441	2664

Hareket tanıma görevlerinde yüksek performans sunan SlowFast R50 modeli de kullanılmış; verideki yavaş ve hızlı hareketleri iki paralel yol üzerinden ayrıştırarak daha hassas analizler gerçekleştirilmiştir. Düşük hesaplama maliyetiyle doğruluk sağlayan X3D Medium modeli ise, farklı genişletme eksenlerinde kademeli olarak genişletilerek video tanıma görevlerinde optimal bir performans sunmuştur. Bu yöntemler, DEMOS

veri setine uygulanarak geniş bir analiz çerçevesi sunmakta ve video temelli görevler için uygun model seçimleri yapılmasını sağlamaktadır.

Tablo 3.7'de görüldüğü üzere DEMOS veri setindeki videolar, eğitim, doğrulama ve test setleri olmak üzere üçe ayrılmıştır. Eğitim seti, toplamda 1857 videodan oluşmakta olup, en çok video Fear (354 video) ve Happiness (327 video) duygularını içermektedir. Doğrulama setinde toplam 399 video bulunmaktadır ve en fazla video Fear (75 video) duygusuna aittir. Test seti ise toplam 408 videodan oluşmakta ve burada da yine en fazla Fear (78 video) ile Happiness (72 video) duyguları yer almaktadır. Genel olarak veri seti, toplamda 2664 video içermekte olup, bu videolar altı temel duyguya (Anger, Disgust, Fear, Happiness, Neutral, Sadness) eşit şekilde dağılmıştır.

Tablo 3. 8. Ön test işlemleri için yapılan denemeler

Model	**Görsel Boyutu**	**Örnekleme Çerçevesi**
Attentive3D-CNN_LSTM	14, 28, 56, 112, 224, 448, 540	4, 8, 16, 24, 32, 50
R3d_18	14, 28, 56, 112, 224, 448, 540	4, 8, 16, 24, 32, 50
Slowfast_r50	224, 448, 540	32
X3d_medium	224, 448, 540	16, 24, 32, 50

Tablo 3.8'de görüldüğü gibi, çalışmada kullanılan modellerin her biri farklı görüntü boyutları ve örnekleme çerçeveleri ile denenmiş ve nihai parametrelere daha sonra karar verilmiştir.

Tablo 3. 9. Parametre boyutuna ait girdi bilgileri

Parametre Adı	**Detay**
Görsel Boyutu	224*224
Çerçeve Boyutu	32
Batch Boyutu	8
Epoch Sayısı	60
Etiket Yumuşatma (label smoothing)	0.1
Öğrenme Oranı	0,001
Öğrenme Oranı Azalması (learning_rate_decay)	Adım sayısı = 2, gamma= 0.9
Optimizer	AdamW

Örneğin, attentive3D-CNN_LSTM ve r3d_18 modelleri geniş bir görüntü boyutu aralığında (14'ten 540 piksele kadar) çalışırken, slowfast_r50 ve x3d_medium modelleri daha dar bir görüntü boyutu yelpazesi (224, 448, 540 piksel) kullanmaktadır. Örnekleme çerçeveleri açısından, attentive3D-CNN_LSTM ve r3d_18 modelleri çeşitli çerçeve aralıklarını kapsarken (4'ten 50'ye kadar), slowfast_r50 modeli sadece 32 çerçeve ile sınırlandırılmıştır. Bu ön test işlemlerinden sonra hepsinde ortak olan 32 çerçeve seçilmiştir.

Tablo 3.9'de görüldüğü gibi, modelin eğitimi sırasında 224 piksel görüntü boyutu ve 32 çerçeve kullanılmıştır. Eğitim aşamasında her adımda 8 veriden oluşan bir batch ile model, 60 epoch boyunca eğitilmiştir. Doğru sınıflandırmayı artırmak için 0,1 oranında label smoothing uygulanmış, öğrenme oranı ise 0,001 olarak belirlenmiştir. Ayrıca, öğrenme oranı her iki adımda bir %10 oranında azaltılacak şekilde ayarlanmıştır. Optimizasyon için AdamW algoritması kullanılmıştır, bu algoritma L2 düzenleme (regularization) yöntemi ile ağırlıkların daha etkili güncellenmesini sağlamaktadır.

3.4. Performans Ölçütleri

Model performansını değerlendirmek için kullanılan ölçütler, sınıflandırma algoritmasının doğruluğunu ve çeşitli hata türlerini detaylı bir şekilde analiz etmemizi sağlar. Sınıflandırma problemlerinde modelin hangi sınıflarda başarılı olduğunu veya hata yaptığını belirlemek, modelin güvenilirliğini ve genel performansını anlamak için kritik öneme sahiptir. Bu başlık altında, modelin sınıflandırma performansını detaylı bir şekilde analiz etmeye olanak tanıyan temel metrikler ele alınacaktır: karmaşıklık matrisi (confusion matrix), duyarlılık (recall), kesinlik (precision), F1 ve balanced accuracy değeri (Japkowicz, 2013; Hossin vd., 2015). Bu ölçütler, doğru ve yanlış sınıflandırmaların sayısal dağılımını ortaya koyarken, modelin her sınıf için başarısını değerlendirmeye yardımcı olur.

Tablo 3. 10. Karmaşıklık matrisi

		Gerçek Etiket	
		Doğru	**Yanlış**
Tahmin Edilen Etiket	**Doğru**	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
	Yanlış	Yanlış Negatif (FN)	Doğru Negatif (TN)

Tablo 3.10'da görülen karmaşıklık matrisi, sınıflandırma problemlerinde model performansını değerlendirmek için sıkça kullanılan bir ölçüt olup modelin doğru ve yanlış sınıflandırmalarını detaylı bir şekilde sunar.

İki veya daha fazla sınıfa sahip veri setleri için oldukça yararlı bir analiz aracı olan karmaşıklık matrisi, modelin tahmin ettiği sonuçları gerçek değerlerle karşılaştırarak her bir sınıf için ayrıntılı bir performans değerlendirmesi yapılmasını sağlar. Bu matrisin dört temel bileşeni bulunur: Doğru Pozitif (TP), modelin bir örneği doğru şekilde pozitif olarak sınıflandırdığı durumu temsil eder; model bir örneği doğru sınıfa atadığında bu durum oluşur. Doğru Negatif (TN), modelin bir örneği doğru şekilde negatif olarak sınıflandırdığı durumu ifade eder; burada model, olması gerektiği gibi bir örneği negatif sınıfa yerleştirmiştir. Yanlış Pozitif (FP), modelin negatif olması gereken bir örneği yanlışlıkla pozitif sınıfa atadığı durumu gösterir. Son olarak, Yanlış Negatif (FN), modelin pozitif olması gereken bir örneği yanlışlıkla negatif sınıfa atadığı durumu ifade eder.

Önerilen yaklaşımı doğrulamak adına Denklem 3.5'te sunulan doğruluk metriği ile hesaplamalar yapılmıştır.

$$Doğruluk\ (\%) = 100 * \frac{\sum Doğru\ Sınıflandırılan\ Örnek\ Sayısı}{Toplam\ Kullanılan\ Örnek\ Sayısı} \tag{3.5}$$

Doğruluk metrikleri ile doğru sınıflandırılmış örneklerin toplam kullanılan örnek sayısına yüzdesi elde edilir. Ayrıca hedef niteliğe ait tahminlerin ve gerçek değerlerin karşılaştırıldığı karmaşıklık matrislerden de ilgili her bir duygu durumunun doğrulukları elde edilerek tahmin sonuçlarının özeti sunulmuştur.

Denklem 3.6'da gösterilen kesinlik (precision), modelin pozitif olarak tahmin ettiği örnekler arasından, gerçekten pozitif olanların oranını ifade eder. Yüksek bir kesinlik, modelin doğru pozitif tahmin oranının yüksek olduğunu gösterir.

$$Kesinlik\ (Precision) = \frac{Gerçek\ Pozitif\ (TP)}{Gerçek\ Pozitif\ (TP) + Yanlış\ Pozitif\ (FP)} \quad (3.6)$$

Denklem 3.7'de duyarlılık (recall), gerçek pozitif örnekler arasından, doğru tahmin edilenlerin oranını ifade eder. Yüksek bir duyarlılık, modelin pozitif örnekleri kaçırmadan tahmin edebilme başarısını gösterir.

$$Duyarlılık\ (Recall) = \frac{Gerçek\ Pozitif\ (TP)}{Gerçek\ Pozitif\ (TP) + Yanlış\ Negatif\ (FN)} \quad (3.7)$$

Ancak, özellikle dengesiz veri setlerinde model performansını değerlendirmek için duyarlılık gibi sınıf bazlı metriklere ek olarak balanced accuracy (dengeli doğruluk) metriği kullanılmaktadır. Balanced accuracy, her bir sınıfın duyarlılık değerlerinin ortalamasını alarak hesaplanır ve modelin tüm sınıflar üzerindeki genel başarısını ölçer. Denklem 3.8'de balanced accuracy metriğinin matematiksel formülü sunulmuştur.

$$Balanced\ accuracy\ (\%) = \frac{TPR + TNR}{2} \quad (3.8)$$

Balanced accuracy metriği, modelin hem pozitif hem de negatif sınıflardaki performansını denge içinde değerlendirmeyi amaçlar. Bu metrik, TPR (True Positive Rate) ve TNR (True Negative Rate) değerlerinin aritmetik ortalamasını alarak hesaplanır. TPR (Duyarlılık/Sensitivity), modelin gerçek pozitifleri doğru bir şekilde tanımlama oranını ifade ederken, TNR (Özgüllük/Specificity) ise modelin negatif örnekleri doğru bir şekilde sınıflandırma oranını belirtir. Başka bir deyişle, TPR, modelin pozitif örnekleri yakalama kabiliyetini ölçerken, TNR ise negatif sınıfları doğru tanımlama kapasitesini yansıtır. Bu iki oran, modelin genel performansını sınıflar arasında dengeleyerek daha adil bir değerlendirme sunar.

Balanced accuracy, özellikle sınıf dengesizliklerinin yüksek olduğu veri setlerinde klasik doğruluk (accuracy) metriğine kıyasla daha güvenilir bir ölçüm sağlar. Sınıf dengesizliği durumlarında, model nadir görülen sınıfları göz ardı edebilir ve bu durumda

doğruluk metriği yanıltıcı olabilir. Balanced accuracy, her sınıfın performansını eşit ağırlıkta ele alarak modelin tüm sınıflardaki başarısını daha doğru bir şekilde yansıtır. Bu yaklaşım, sınıflar arası dengesizliklerin fazla olduğu veri setlerinde, modelin hem duyarlılık hem de özgüllük açısından performansını objektif bir şekilde ortaya koymayı amaçlar.

Ayrıca modellerin performansı değerlendirmek adına Denklem 3.9'da sunulan F1 skor metriği de hesaplanmıştır.

$$F1\ skoru = 2 * \frac{Kesinlik\ (Precision)\ x\ Duyarlılık\ (Recall)}{Kesinlik\ (Precision) + Duyarlılık\ (Recall)} \tag{3.9}$$

F1 skoru, kesinlik (Precision) ve duyarlılık (Recall) değerlerinin harmonik ortalamasını alarak hesaplanır ve özellikle dengesiz veri setlerinde daha güvenilir bir performans ölçütü sağlar. F1 skoru, modelin doğru pozitif tahmin oranını (precision) ve gerçek pozitifler arasından doğru tahmin edilenlerin oranını (recall) dengeleyerek, modelin genel tahmin performansını özetler.

4. BULGULAR VE TARTIŞMA

Bu bölümde, ham kinematik veri seti ve DEMOS veri seti üzerinde gerçekleştirilen duygu sınıflandırma çalışmalarından elde edilen sonuçlar detaylı bir şekilde incelenmiştir. Her iki veri seti için kullanılan derin öğrenme ve makine öğrenimi tabanlı modellerin doğruluk, dengeli doğruluk gibi performans metrikleri üzerinden karşılaştırmaları yapılmış, modellerin genelleme kapasiteleri ve duygu sınıflarına özgü başarıları değerlendirilmiştir. Kinematik veri setinde makine öğrenimi ve öznitelik çıkarımı yöntemlerinin etkisi analiz edilirken, DEMOS veri setinde ise spatio-temporal özelliklerin duygu sınıflandırma üzerindeki etkisi ve farklı modellerin bu bağlamdaki performansları incelenmiştir. Her iki veri setinden elde edilen bulgular, duygu tanıma görevlerinde kullanılan modellerin etkinliğini ve bu görevler için en uygun yaklaşımların belirlenmesine yönelik önemli çıkarımlar sunmaktadır.

4.1. Ham Kinematik Veri Seti Sonuçları

Bu çalışmada, kinematik veri seti kullanılarak yedi duygu sınıfını sınıflandırmaya yönelik çeşitli derin öğrenme ve makine öğrenimi tabanlı modellerin doğruluk (accuracy) performansları karşılaştırılmıştır. Kullanılan modeller arasında GRU, LSTM, MobileNetV3-Large, RegNetY-800MF gibi derin öğrenme mimarilerinin yanı sıra, KNN, RF, XGBoost ve CatBoost gibi makine öğrenimi algoritmaları da yer almaktadır. Her modelin mimarisi ve eğitim süreçleri farklı olup, kinematik veri seti üzerinde duygu sınıflandırma doğruluğu açısından çeşitli sonuçlar elde edilmiştir. Bu çalışmanın amacı, her bir modelin doğruluk üzerinden karşılaştırılarak, hangi modelin duygu sınıflarında daha başarılı olduğunu belirlemektir. Çalışma kapsamında, modellerin genel doğruluk performansları detaylı bir şekilde incelenmiş ve aralarındaki karşılaştırmalar yapılmıştır.

Verilerin hem ham pozisyon hali hem de öznitelik çıkarılmış (FE) hali hem ML hem de DL algoritmaları ile sınıflandırılmıştır.

Tablo 4. 1. ML algoritmaları ile elde edilen test sonuçları

	Sınıflandırıcı adı	Veri Senaryosu	Pencere boyutu (milisaniye cinsinden)					
			40	80	160	240	320	400
Test Başarısı (%)	CatBoost	Ham	96,783	93,537	88,208	83,844	80,472	77,032
		FE	98,374	96,302	92,421	88,818	86,388	83,798
	kNN	Ham	94,450	88,414	78,105	70,374	64,770	60,400
		FE	98,706	95,750	88,165	82,235	77,296	73,914
	RF	Ham	99,067	96,331	90,256	84,540	80,049	75,802
		FE	**99,224**	96,927	92,032	87,334	83,164	79,725
	XGBoost	Ham	92,248	90,832	86,295	82,339	78,776	75,753
		FE	95,219	92,549	88,007	84,203	81,035	78,344
Ortalama dosya hesaplama maliyeti (fold başına) (milisaniye)	CatBoost	Ham	0,63	1,56	4,53	8,71	13,72	19,86
		FE	0,53	0,69	1,01	1,33	1,66	2,00
	kNN	Ham	0,26	0,36	0,53	0,52	0,53	0,53
		FE	0,27	0,13	0,07	0,04	0,04	0,03
	RF	Ham	3,89	4,87	5,95	6,58	7,30	8,12
		FE	3,68	3,11	2,59	2,35	2,19	2,15
	XGBoost	Ham	0,22	0,59	1,76	3,55	5,65	8,10
		FE	0,18	0,24	0,37	0,49	0,63	0,74

Tablo 4.1'te kullanılan dört adet ML tabanlı sınıflandırıcı için hem başarı oranları hem de fold başına her bir dosya için harcanan toplam süre (milisaniye) cinsinden verilmiştir. En başarılı sonuç RF algoritmasıyla 40 milisaniyelik pencere boyutu için %99,22 ile elde edilmiştir. Tüm pencere boyutları için ham pozisyon verisi sınıflandırmada genel ortalamada en başarılı algoritma RF algoritması olmuştur. Hem FE verinin tasnifinde hem de tüm (ham-FE) ortalamada en başarılı algoritma ise CatBoost olmuştur. CatBoost ile 400 ms'lik en uzun pencere süresi için bile FE verisi için %83,798 oranında başarı sağlanabilmektedir. Dosya başına ortalama süre hesabında özellikle dosyanın ham hali ile yapılan sınıflandırmanın boyut arttıkça maliyetinin arttığı görülmektedir. Özellikle CatBoost ve XGBoost algoritmasının hem kullandığı hafıza hem de zaman maliyetinin boyut arttıkça ham veri durumu için oldukça fazla olduğu görülmüştür. Tüm ML sınıflandırıcılarında öznitelik çıkarılmış veriden elde edilen başarı oranı, verinin ham halinden çok daha başarılı sonuçlar elde edilmesini sağlamıştır.

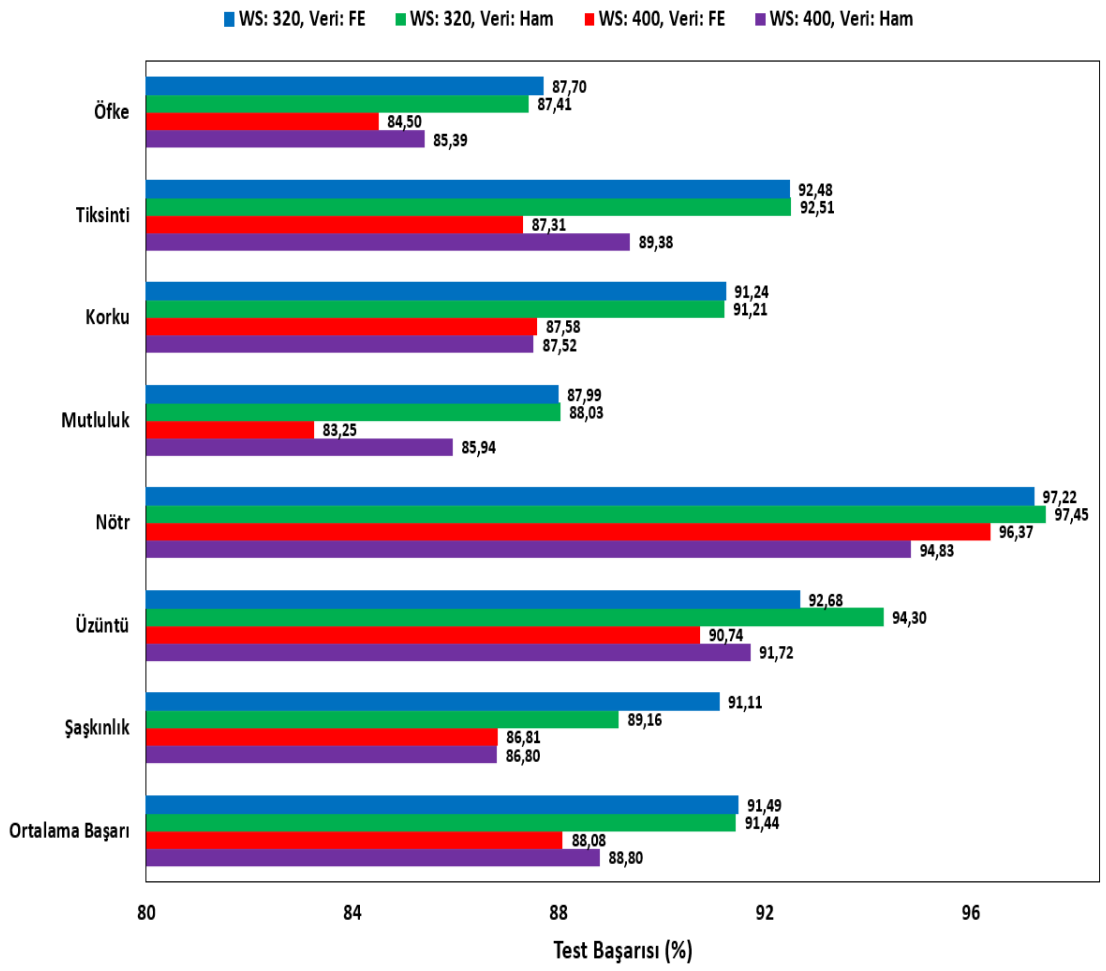
Tablo 4. 2. DL algoritmaları ile elde edilen test sonuçları

	Sınıflandırıcı adı	Veri Senaryosu	Pencere boyutu (milisaniye cinsinden)					
			40	80	160	240	320	400
Test Başarısı (%)	GRU	Ham	99.642	97.789	93.917	89.344	85.059	79.294
		FE	99.847	98.849	96.057	91.921	88.338	84.369
	LSTM	Ham	99.483	96.800	89.622	83.191	74.770	71.924
		FE	99.685	98.063	93.470	89.115	84.171	81.466
	MobileNetV3-Large	Ham	99.931	99.270	96.488	93.233	89.775	85.993
		FE	99.899	98.920	94.819	91.054	85.624	82.663
	RegNetY-800MF	Ham	99.970	99.623	98.196	94.955	91.441	88.799
		FE	**99.986**	99.524	97.262	94.461	91.488	88.081
Ortalama dosya hesaplama maliyeti (epoch başına) (milisaniye)	GRU	Ham	0.38	0.38	0.38	0.39	0.40	0.41
		FE	0.38	0.39	0.39	0.39	0.39	0.38
	LSTM	Ham	0.53	0.52	0.53	0.54	0.55	0.57
		FE	0.53	0.53	0.53	0.53	0.53	0.53
	MobileNetV3-Large	Ham	0.14	0.19	0.31	0.46	0.60	0.69
		FE	0.13	0.13	0.13	0.13	0.13	0.13
	RegNetY-800MF	Ham	0.23	0.36	0.68	1.12	1.18	2.00
		FE	0.23	0.23	0.23	0.23	0.23	0.23

Tablo 4.2'te kullanılan dört adet DL tabanlı sınıflandırıcı için hem başarı oranları hem de epoch başına her bir dosya için harcanan toplam süre (milisaniye) cinsinden verilmiştir. Elde edilen en başarılı sonuçlar hem ham hem FE veri durumları için RegNetY-800MF algoritması ile elde edilmiştir. Daha sonra en başarılı algoritma MobileNetV3-Large algoritması olup, onu sırasıyla GRU ve LSTM takip etmiştir. FE verisi özellikle GRU ve LSTM için oldukça işe yaramış ve başarı oranını etkileyici bir şekilde artırmıştır. RegNetY-800MF algoritmasında bazı pencere uzunlukları için ham hali bazen de FE durumu daha başarılı olmuştur. Genel olarak CNN kökenli DL sınıflandırıcılarında öznitelik çıkarımının ML sınıflandırıcıları ve RNN tabanlı DL algoritmaları kadar başarılı çıktılar sağlamadığı görülmektedir.

DL modelleri için harcanan süre karşılaştırması pencere boyutu ve verinin türüne göre değişiklikler göstermektedir. Genel olarak GRU LSTM'den, MobileNetV3-Large modeli de RegNetY-800MF modelinden az da olsa daha hızlıdır. Ham veri için penceredeki çerçeve boyutu küçükken CNN tabanlı modeller, penceredeki çerçeve boyutu büyükken RNN tabanlı modeller daha hızlı işlem kapasitesine sahiptir. FE veri durumu için ise CNN tabanlı modellerin RNN tabanlı modellere az da olsa üstünlüğü vardır. RegNetY-800MF modelinin parametre sayısı diğer üç DL modeline göre nispeten daha fazla olmasına rağmen test için harcanan sürede bu fazlalığın etkisiz olduğu görülmektedir.

Şekil 4.1'te yukarıdaki sınıflandırma sonuçlarına ek olarak RegNetY-800MF algoritması ile her bir duygunun doğru sınıflandırma oranı sunulmuştur. Şekil 3'te sadece 320 ve 400 milisaniyelik pencere boyutları ile ilgili detaylı sınıflandırma bilgisi verilmesinin nedeni diğer pencere boyutlarındaki sonuçlara göre daha düşük sonuçlar vermeleri ve sınıflandırma ayrımının daha belirgin gösterilmek istenmesidir. En belirgin ve başarılı şekilde ayırt edilebilen duygu veya hal Neutral durumudur. Daha sonra en belirgin duygular sırasıyla Sadness ve Disgust duygularıdır. En az duygulanım başarısı sırasıyla Anger ve Happiness duygularında elde edilmektedir.

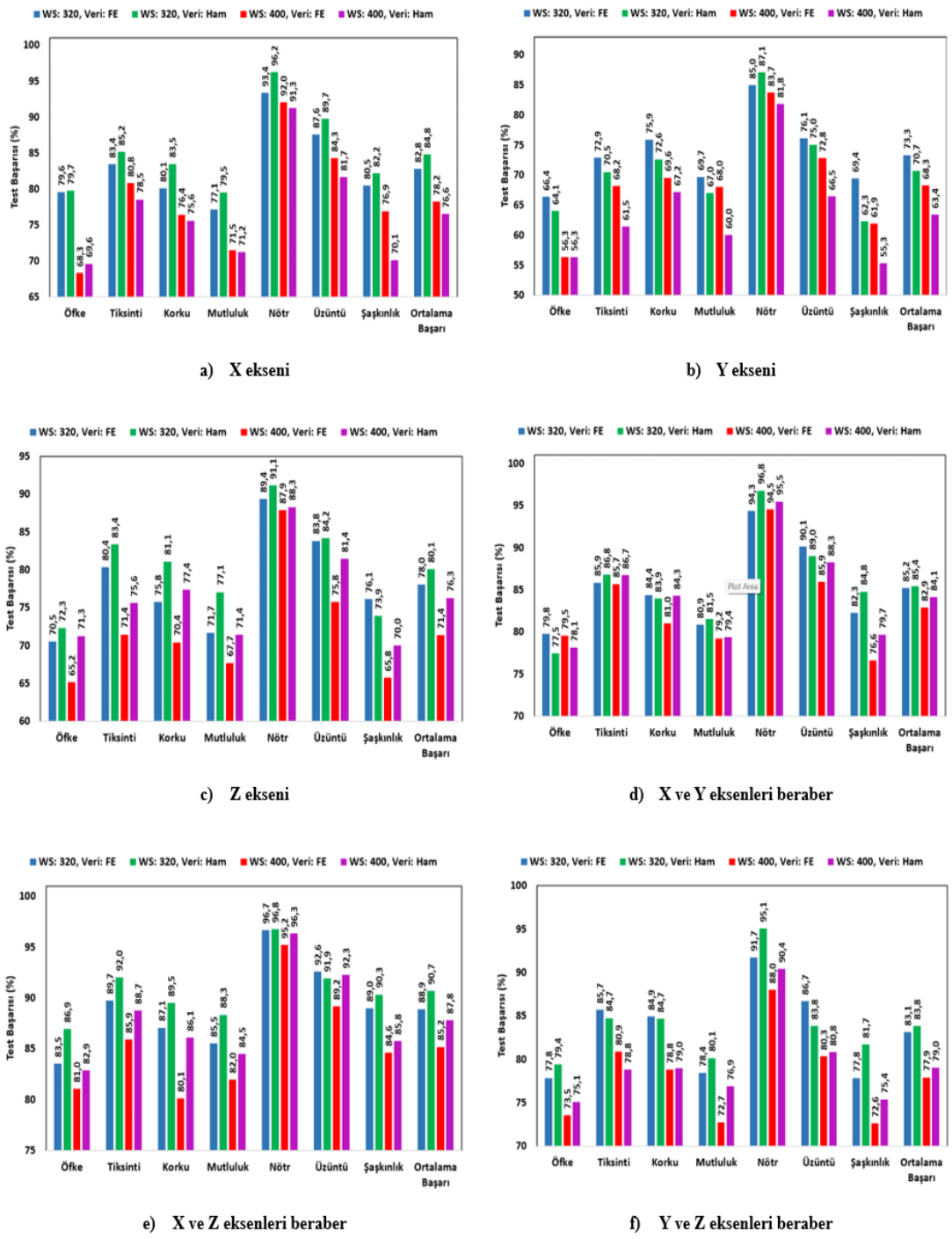


Şekil 4. 1. Duyguların hem ham hem FE hallerinin başarı oranları

Ayrıca eldeki veriler kapsamında sadece bir veya iki eksendeki koordinat bilgisi kullanılarak hesaplama yapılırsa başarı düzeyi ne olabilir sorusunun da cevapları aranmıştır. Şekil 4.2'de yine RegNetY-800MF algoritması ile her bir eksenin sınıflandırma başarısına katkısı hem farklı veri hem de duyguların üzerinden incelenerek sunulmuştur. Yine Şekil 4.2'de sadece 320 ve 400 milisaniyelik pencere boyutları için

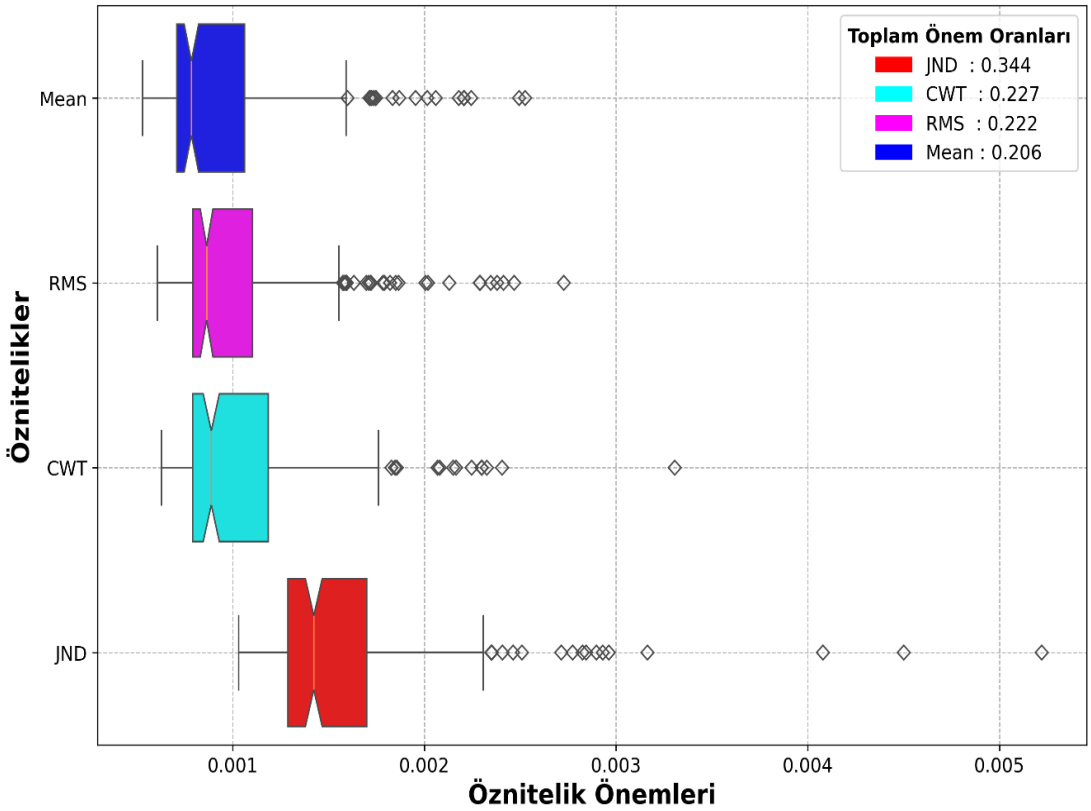
yapılan denemeler sunulmuştur. Tekil bazda eksenler (x, y, z) içinde en başarılı bilgileri sağlayan tek eksenin her durum ve pencere uzunluğu için X ekseni olduğu görülmüştür. Başarılı tekil eksenler sıralamasında X eksenini sırasıyla Z ve Y eksenleri takip etmektedir. İki eksenli kombinasyonlarda da en başarılı birleşim XZ eksenlerinin birleşiminden oluşmaktadır. Daha sonra başarı sıralamasında XY ve ZY birleşimi gelmektedir.

Bu çalışma bazında Y ekseninin hesaplamalarda başarı katkısını düşürdüğü gözlemlenmiştir. Bu durumun veri seti toplanırken deneklerin ayakta olmasından kaynaklı olabileceği düşünülmektedir. 3 eksenli verilerin işlenmesi için daha fazla hesaplama gücü gerekebileceğinden, bu sonuçlar ışığında bazı durumlarda 2 eksenli hesaplamaların verilerin daha hızlı ve daha kolay bir şekilde işlenmesi için uygulanabileceği düşünülmektedir.



Şekil 4. 2. Duyguların eksenlere göre başarı oranları

Veriye uygulanan özniteliklerin sonucunda tüm eklem ve tüm eksenler bazında özniteliklerin önem derecesi ağırlıklarının dağılımı Şekil 4.3'te sunulmuştur. Bu yaklaşımda önem derecesi için bahsedilen ağırlıkların belirlenmesinde RF algoritmasından faydalanılmıştır. Ayrıca şekilde toplamda özniteliklerin elde edilen oransal ağırlık dağılımları da görülebilmektedir.



Şekil 4. 3. Çıkarılan özniteliklerin sınıflandırmada önemi

Öznitelik çıkarım kısmında da belirtildiği üzere onlarca öznitelik kombinasyonu denemesinden sonra karar kılınan dört adet öznitelik kümesinin içinde en önemli ağırlığa sahip öznitelik JND yaklaşımımız olmuştur. JND parametresini sırasıyla CWT, RMS ve Ort. parametreleri takip etmiştir. JND dışındaki özniteliklerin katkısı birbirine yakın iken, JND diğerlerinden daha belirleyici olmuştur. Özniteliklerin önem derecesinin Tablo 3'te bahsedilen öznitelik çıkarımı zaman maliyeti ile göz önüne alınarak çözülmek istenen probleme entegre edilmesi daha uygun olacaktır.

Bu çalışmada kullanılan kodlar, veri setleri ve eğitilmiş model dosyaları, GitHub platformu üzerinden açık erişime sunulmuştur [2]. Bu sayede, diğer araştırmacılar CNN tabanlı derin öğrenme modellerimizi kullanarak örnek veri setlerini test edebilme imkânına sahip olacaktır. Çalışma kapsamında toplam 12 yeni veri seti oluşturulmuş ve her bir veri setine her duygudan 100 örnek eklenmiştir. Bu düzenleme sonucunda toplamda 700 dosya oluşturulmuştur. İlgili dizinde, MobileNetV3-Large ve RegNetY-

[2] https://github.com/ahalikoguz/Kinematic_Emotion_Recognition

800MF algoritmalarının eğitilmiş sürümlerini içeren 24 model dosyasına ek olarak, bu küçük veri setlerinin test edilmesine yönelik ayrıntılı kullanım talimatları yer almaktadır.

4.2. DEMOS Video Veri Seti Sonuçları

Bu çalışmada, DEMOS veri seti üzerinde spatio-temporal duygu tanıma amacıyla farklı derin öğrenme modellerinin performansı kapsamlı bir şekilde incelenmiştir. Altı farklı duygu sınıfını sınıflandırmaya yönelik çeşitli derin öğrenme tabanlı modellerin performansları karşılaştırılmıştır. Bu modeller arasında Attentive3D-CNN-LSTM, R3D_18, Slowfast_r50, X3D_Medium yer almaktadır. Her modelin mimarisi ve eğitim süreçleri farklı olup, duygu sınıflandırma başarısı açısından çeşitli sonuçlar elde edilmiştir. Çalışma, modellerin doğruluk ve dengeli doğruluk skoru gibi metrikler üzerinden genelleme kapasitelerini ve sınıflar arası duyarlılıklarını değerlendirmeyi hedeflemiştir. Bunun yanı sıra, eğitim süreleri ve epoch sayıları gibi hesaplama maliyetlerine ilişkin faktörler de göz önünde bulundurulmuştur. Elde edilen sonuçlar, farklı model mimarilerinin avantaj ve dezavantajlarını ortaya koymakta ve bu alandaki gelecek çalışmalar için önemli referans noktaları sunmaktadır. Çalışma kapsamında, modellerin genel performansları ile duygu sınıflarına özgü başarıları detaylı bir şekilde incelenmiş, bu sonuçlar üzerinden karşılaştırmalı bir analiz yapılmıştır.

Tablo 4. 3. DEMOS veri seti genel sonuçlar

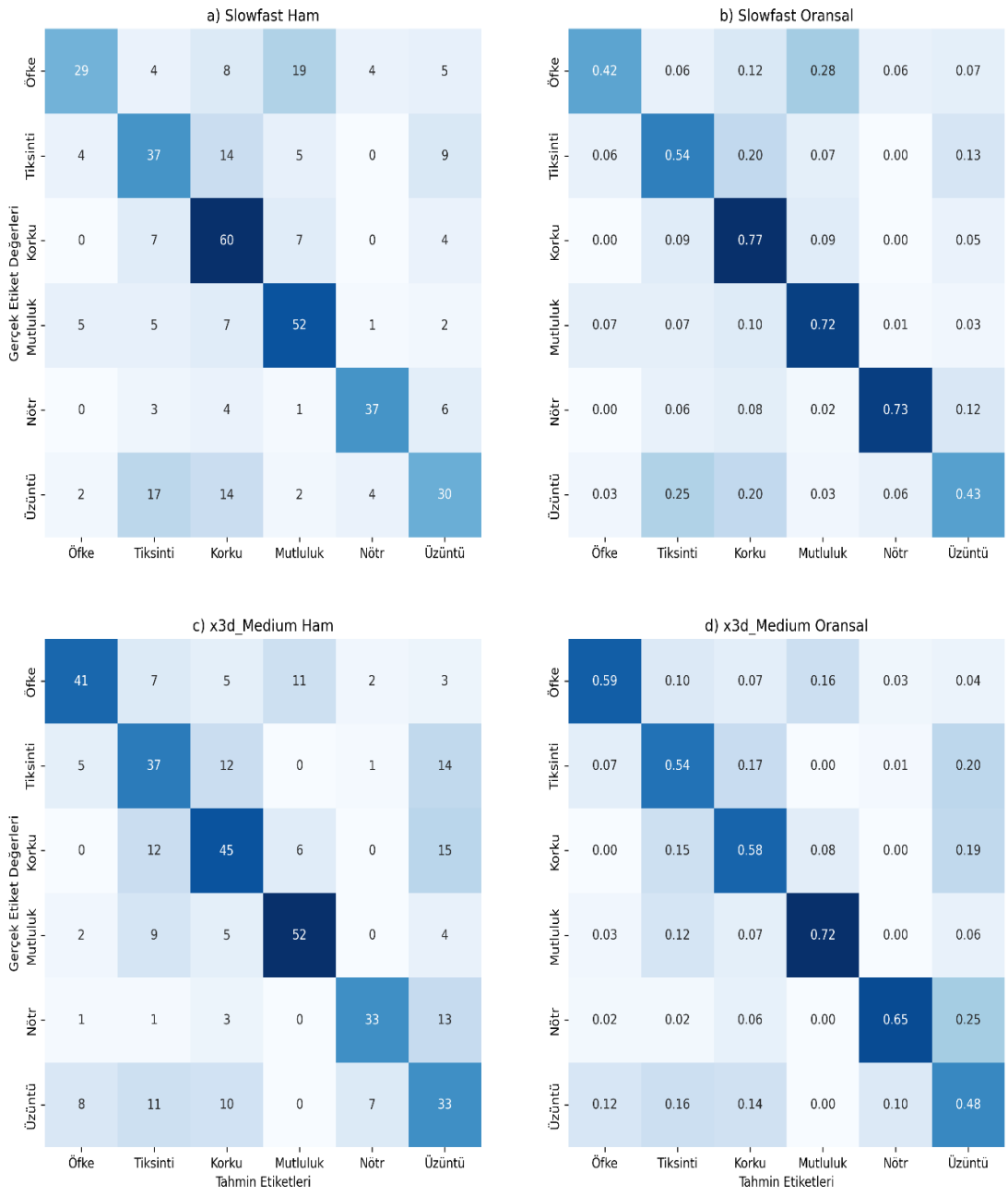
Model	Eğitim Acc	Validasyon Balanced Acc	En İyi Epoch Sayısı	Ortalama Epoch (saniye)	Test Acc	Test Balanced Acc	Test Weighted F1 Skoru
CNN3D_LSTM	40,23	45,61	8	601,85	47,55	47,36	47,46
R3D_18	99,57	51,19	43	536,28	49,02	49,70	49,41
Slowfast_r50	92,03	60,80	41	744,77	60,05	60,14	59,48
X3D_medium	67,85	57,71	57	619,05	59,07	59,25	59,65

Tablo 4.3'de yer alan sonuçlara göre, SlowFast_r50 modeli, %60,80 validasyon dengeli doğruluk ve %60,14 test dengeli doğruluk oranlarıyla en yüksek performansı sergilemiştir. Bu model, spatio-temporal özelliklerin öğreniminde güçlü bir potansiyel ortaya koymuş ve sınıf dengesizliği durumlarında dahi kararlı sonuçlar sağlamıştır. X3D_medium modeli ise %57,71 validasyon dengeli doğruluk ve %59,25 test dengeli doğruluk oranıyla benzer bir performans sergilemiş, %59,65 test F1 skoru ile sınıf

dengelerine duyarlılığını daha iyi bir şekilde yansıtmıştır. Bu bulgu, modelin özellikle sınıflar arası ağırlıklandırmayı daha verimli gerçekleştirdiğini göstermektedir. R3D_18 ve CNN3D_LSTM modelleri ise daha düşük test dengeli doğruluk oranları (%49,70 ve %47,36) ile genelleme kapasitelerinin sınırlı olduğunu göstermiştir. Özellikle R3D_18 modelinin eğitim doğruluğunun %99,57 gibi yüksek bir seviyede olmasına rağmen validasyon ve test doğruluğunda önemli düşüşler göstermesi, overfitting problemi yaşadığını işaret etmektedir.

Modellerin eğitim süreleri incelendiğinde, SlowFast_r50 modelinin ortalama 744,77 saniyelik epoch süresi ile en yüksek hesaplama maliyetine sahip olduğu görülmektedir. Bununla birlikte, gösterdiği yüksek performans, bu maliyetin belirli uygulama senaryolarında kabul edilebilir olduğunu göstermektedir. X3D_medium modeli, hesaplama süresi açısından daha verimli bir seçenek sunmakta (%17 daha düşük eğitim süresi), ancak performans olarak SlowFast_r50 modeline çok yakın bir sonuç elde etmektedir. Bu durum, hesaplama maliyetlerini optimize etmek isteyen uygulamalar için X3D_medium modelini daha uygun bir seçenek haline getirebilir. Daha kısa eğitim sürelerine sahip olan R3D_18 ve CNN3D_LSTM modelleri ise, bu avantajlarına rağmen sınırlı genel performansları nedeniyle tercih edilme olasılığı düşüktür.

Slowfast ve X3D_Medium modellerinin duygu sınıflandırma başarısı Şekil 4.4'te detaylı bir şekilde göstermektedir. Slowfast modeli, Korku (Fear) ve Mutluluk (Happiness) sınıflarında en yüksek doğruluk oranlarını elde etmiştir. Korku sınıfında 60 doğru tahmin ile %77 doğruluk oranı sağlarken, Mutluluk sınıfında ise 52 doğru tahmin ile %72 doğru sonuçlar elde edilmiştir. Tiksinme (Disgust) sınıfı da nispeten iyi performans göstermiş, 37 doğru tahmin yapılmış olsa da Korku ve Üzüntü sınıflarıyla bazı karışıklıklar yaşanmıştır. Nötr (Neutral) sınıfı, 37 doğru tahmin ile diğer sınıflarla sıkça karıştırılmış ve düşük doğruluk oranları ortaya çıkmıştır. Öfke (Anger) ve Üzüntü (Sadness) sınıflarında ise belirgin bir zayıflık gözlemlenmiştir; özellikle Öfke sınıfı, yalnızca %42 doğruluk oranı ile büyük oranda yanlış sınıflandırılmış ve sıklıkla Mutluluk ve Nötr sınıfları ile karıştırılmıştır. Üzüntü sınıfı da %43 oranla en düşük performansı sergileyen sınıf olmuştur.



Şekil 4. 4. Slowfast ve X3D_medium modellerine ait karmaşıklık matrisleri

X3D_Medium modelinin sonuçlarına bakıldığında, Öfke ve Korku sınıflarında Slowfast modeline kıyasla daha iyi performans gösterdiği söylenebilir. Öfke sınıfı, 41 doğru tahmin ile %59 doğruluk oranı sağlamış, bu oran Slowfast modelindeki %42 oranından belirgin şekilde yüksektir. Korku ve Mutluluk sınıfları, sırasıyla %58 ve %72 doğruluk oranları ile güçlü performanslar sergilemiştir. Ancak Tiksinme ve Üzüntü sınıflarında, Slowfast modeline kıyasla biraz daha düşük doğruluklar elde edilmiştir. Tiksinme sınıfı, 37 doğru tahmin ile %54 doğruluk sağlarken, Üzüntü sınıfı ise %48 doğrulukla zayıf performans sergileyen diğer bir sınıf olmuştur. Nötr sınıfı ise

X3D_Medium modelinde %65 doğruluk oranıyla, Slowfast modelinin %73 doğruluğunun gerisinde kalmıştır.

Genel karşılaştırmalarda, her iki modelin Korku ve Mutluluk sınıflarında benzer yüksek doğruluklar sergilediği gözlemlenmiştir. Ancak, Öfke sınıfında X3D_Medium modelinin daha başarılı olduğu, Tiksinme ve Üzüntü sınıflarında ise Slowfast modelinin daha iyi sonuçlar verdiği görülmektedir. Nötr sınıfında Slowfast modeli daha başarılı olurken, X3D_Medium modelinde doğruluk oranı daha düşük kalmıştır. Bu karşılaştırma, her iki modelin güçlü yönlerinin birleştirilmesiyle daha hassas ve doğru bir duygu sınıflandırma sistemi elde edilebileceğini göstermektedir.

Sonuçlar itibarıyla, SlowFast-R50 modeli, Anger, Neutral, Fear ve Happiness sınıflarında yüksek doğruluk oranları ile genel başarı açısından en iyi performansı göstermektedir. X3D-Medium ise özellikle Fear, Happiness ve Neutral sınıflarında güçlü performansı ve Ağırlıklı F1 skoru açısından (%59,65) en yüksek başarıyı elde etmiştir. Diğer taraftan, MC3-18 ve S3D modelleri, Disgust ve Anger sınıflarında iyileştirilebilecek potansiyel göstermektedir, ancak yine de bu sınıflarda katkı sağlamakta önemli bir rol oynamaktadırlar. R3D-18 modeli de Sadness sınıfında daha fazla geliştirme gerektiren bir alanda performans göstermektedir. SlowFast-R50, genel doğruluk ve sınıflandırma başarısı açısından öne çıkarken, X3D-Medium ise F1 skoru açısından daha yüksek başarı göstermektedir.

DEMOS veri setindeki 0°, 45° ve 90° açılardan alınan veriler, spatio-temporal duygu tanıma performansını görüş açısına bağlı olarak değerlendirmek amacıyla ayrı ayrı eğitilmiştir. Bu analiz, her bir açının sınıflandırma doğruluğu ve dengeli doğruluk gibi metriklere olan katkısını anlamaya yönelik önemli bir perspektif sunmaktadır. Modellerin farklı açılardan elde edilen verilerle eğitildiğinde sınıflandırma performanslarında nasıl bir değişim gösterdiği detaylı olarak incelenmiş ve bu bulgular üzerinden görüş açısının duygu tanıma başarı oranlarına etkisi değerlendirilmiştir.

Ancak CNN3D_LSTM modelinin genel sınıflandırma performansında diğer modellerin gerisinde kaldığı ve özellikle sınıflandırma doğruluğu, dengeli doğruluk ve F1 skoru gibi temel metriklerde yeterli başarı sağlayamadığı görülmüştür. Bu modelin hem genelleme kapasitesindeki sınırlılıklar hem de spatio-temporal ilişkileri öğrenmedeki yetersizliği nedeniyle, 0°, 45° ve 90° açılardan elde edilen verilerle yapılan analizlerde anlamlı bir katkı sunamayacağı değerlendirilmiştir. Bu nedenle, CNN3D_LSTM modeli analizden çıkarılmış ve diğer modellerin performanslarının daha net ve tutarlı bir şekilde karşılaştırılmasına olanak tanınmıştır.

DEMOS veri setindeki 0°, 45° ve 90° açılarında yapılan eğitimlerde, veri kümesinin toplam uzunluğu 752 örnekten oluşmaktadır. Her bir örnek, 224x224 piksel boyutlarında görüntülerden oluşmakta olup, spatio-temporal özelliklerin etkin bir şekilde işlenmesi için 32 çerçeve üzerinden örnekleme yapılmıştır. Girdi boyutu ise toplamda 1.605.632 özellik içermekte olup, modellerin karmaşık mekânsal ve zamansal ilişkileri öğrenmesine olanak sağlamaktadır. Bu yapılandırma, 0° açısındaki verilerin duygu tanıma performansına etkisini değerlendirmek amacıyla güçlü bir temel sunmuştur.

Tablo 4. 4. 0° açı eğitim ve test sonuçları

Model Adı	Eğitim Doğruluğu	Validasyon Dengeli Doğruluğu	Test Dengeli Doğruluğu	En İyi Epoch Sayısı	Ortalama Epoch (saniye)
Slowfast_r50	74,64	68,94	62,34	30	358,62
X3D_medium	47,50	47,44	47,23	44	349,07
R3D_18	27,30	35,80	33,50	9	218,58

0° açısına yönelik eğitim sonuçları Tablo 4.4'de gösterilmektedir. Buna göre; SlowFast_r50 modeli, eğitim doğruluğu %74,64, validasyon dengeli doğruluğu %68,94 ve test dengeli doğruluğu %62,34 ile en yüksek performansı sergileyen model olmuştur. Özellikle yüksek test dengeli doğruluk oranı, modelin genelleme kapasitesinin güçlü olduğunu göstermektedir. Ortalama epoch süresi 358,62 saniye ile diğer modellere kıyasla biraz daha yüksektir. Ancak en iyi sonuçlara 30 epoch sonrasında ulaşması, eğitim süresi bakımından da verimli bir model olduğunu ortaya koymuştur. SlowFast_r50, spatio-temporal özellikleri 0° açıdan etkin bir şekilde öğrenerek, duygu tanıma görevinde diğer modellere göre daha başarılı sonuçlar elde etmiştir.

X3D_Medium modeli, eğitim doğruluğu %47,50, validasyon dengeli doğruluğu %47,44 ve test dengeli doğruluğu %47,23 oranları ile orta düzeyde bir performans sergilemiştir. Modelin validasyon ve test dengeli doğruluğu değerlerinin birbirine yakın olması, genelleme kapasitesinde bir istikrar olduğunu göstermektedir. Ancak performans oranları, diğer modellere kıyasla sınırlı kalmıştır. Ortalama epoch süresi 349,07 saniye olup, bu süre SlowFast_r50'ye göre biraz daha düşük bir hesaplama maliyeti sunmaktadır. Ancak en iyi sonuçlara ulaşmak için 44 epoch gerektirmesi, eğitim süresinin toplamda uzamasına neden olmaktadır. X3D_Medium modeli, düşük hesaplama maliyetine sahip olmasına rağmen, performans açısından sınırlı kalmış ve 0° açı verileriyle spatio-temporal duygu tanıma başarısında yeterli düzeye ulaşamamıştır.

R3D_18 modeli, eğitim doğruluğu %27,30, validasyon dengeli doğruluğu %35,80 ve test dengeli doğruluğu %33,50 ile en düşük performansı sergileyen model olmuştur. Bu sonuçlar, modelin spatio-temporal özellikleri öğrenmede yetersiz kaldığını ve genelleme kapasitesinin oldukça sınırlı olduğunu göstermektedir. Ortalama epoch süresi 218,58 saniye ile en düşük hesaplama maliyetine sahip modeldir. Ayrıca en iyi sonuçlara 9 epoch gibi kısa bir sürede ulaşması, bu modelin eğitim açısından hızlı sonuçlar verdiğini göstermektedir. R3D_18 modeli, düşük hesaplama maliyeti sunmasına rağmen, sınıflandırma doğruluğu ve genelleme başarısı açısından diğer modellere kıyasla oldukça geride kalmıştır.

0° açısına ilişkin eğitim sonuçları, SlowFast_r50 modelinin hem doğruluk hem de genelleme kapasitesi açısından üstün olduğunu net bir şekilde ortaya koymuştur. X3D_Medium modeli, dengeli bir performans sergilese de sınırlı doğruluk oranları nedeniyle tatmin edici sonuçlar elde edememiştir. R3D_18 modeli ise düşük doğruluğu ve genelleme başarısı ile duygu tanıma görevinde yeterli bir performans sunamamıştır. Bununla birlikte, eğitim süresi açısından R3D_18 modeli daha kısa bir eğitim maliyeti ile avantaj sağlasa da bu avantaj sınıflandırma performansındaki düşüklüğü telafi etmemektedir. Bu bulgular, 0° açı verilerinde spatio-temporal duygu tanıma için SlowFast_r50 modelinin en uygun seçenek olduğunu göstermektedir. Yüksek performansı ve dengeli sonuçları, bu modeli bu açıdan en güçlü seçenek haline getirmektedir.

Tablo 4. 5. 45° açı eğitim ve test sonuçları

Model Adı	Eğitim Doğruluğu	Validasyon Dengeli Doğruluğu	Test Dengeli Doğruluğu	En İyi Epoch Sayısı	Ortalama Epoch (saniye)
Slowfast_r50	95,90	65,29	62,73	43	364,37
X3D_medium	33,12	35,39	37,40	39	373,73
R3D_18	40,39	41,92	32,98	20	218,29

45° açısına yönelik eğitim sonuçları Tablo 4.5'de gösterilmektedir. Buna göre; SlowFast_r50 modeli, eğitim doğruluğu %95,90 ile oldukça yüksek bir seviyededir. Validasyon dengeli doğruluğu %65,29 ve test dengeli doğruluğu %62,73 ile 45° açı sonuçlarında en yüksek performansı sergileyen model olmuştur. Bu değerler, modelin genelleme kapasitesinin güçlü olduğunu ve spatio-temporal ilişkileri 45° açıdan başarılı bir şekilde öğrenebildiğini göstermektedir. Ortalama epoch süresi 364,37 saniye ile diğer

modellere kıyasla biraz daha yüksektir. En iyi sonuçlara 43 epoch sonrasında ulaşması, modelin hesaplama açısından makul bir denge sunduğunu göstermektedir. SlowFast_r50 modeli, hem doğruluk oranları hem de genelleme kapasitesi açısından 45° açıda diğer modellere kıyasla belirgin bir üstünlük sağlamıştır.

X3D_Medium modeli, eğitim doğruluğu %33,12, validasyon dengeli doğruluğu %35,39 ve test dengeli doğruluğu %37,40 oranlarında kalmıştır. Test dengeli doğruluk oranının validasyon doğruluğundan daha yüksek olması, modelin test verileri üzerinde bir miktar daha iyi performans gösterdiğini işaret etse de genel olarak düşük seviyelerde seyretmektedir. Ortalama epoch süresi 373,73 saniye ile SlowFast_r50 modeline kıyasla biraz daha yüksektir. En iyi sonuçlara 39 epoch sonrasında ulaşması, eğitim süresi açısından daha verimli seçeneklerin olabileceğini göstermektedir. X3D_Medium modeli, 45° açı verileri ile spatio-temporal duygu tanıma performansında tatmin edici bir başarı elde edememiştir. Ancak test dengeli doğruluk oranının stabil olması, modelin daha fazla optimize edilmesi durumunda iyileşme gösterebileceğine dair bir ipucu sunmaktadır.

R3D_18 modeli, eğitim doğruluğu %40,39, validasyon dengeli doğruluğu %41,92 ve test dengeli doğruluğu %32,98 ile en düşük genel performansı sergileyen model olmuştur. Model, 45° açı verilerinde genelleme kapasitesinde ciddi sınırlılıklar göstermiş ve düşük test doğruluk oranı ile dikkat çekmiştir. Ortalama epoch süresi 218,29 saniye ile en kısa eğitim süresine sahip modeldir. Ayrıca, en iyi sonuçlara yalnızca 20 epoch sonrasında ulaşması, eğitim süresi açısından avantaj sağlamaktadır. R3D_18 modeli, düşük hesaplama maliyetine sahip olmasına rağmen, sınıflandırma doğruluğu ve genelleme başarısı açısından diğer modellere kıyasla oldukça geride kalmıştır.

45° açısına yönelik sonuçlar, SlowFast_r50 modelinin diğer modellere kıyasla belirgin bir üstünlük sağladığını göstermektedir. Yüksek doğruluk oranları ve dengeli genelleme kapasitesi, bu modeli 45° açıda en uygun seçenek haline getirmiştir. X3D_Medium modeli düşük doğruluk oranlarına rağmen stabil bir performans sunarken, R3D_18 modeli düşük doğruluğu ve sınırlı genelleme kapasitesiyle 45° açı verilerinde yeterli bir başarı elde edememiştir. Bu sonuçlar, 45° açı verileri ile spatio-temporal duygu tanıma görevlerinde SlowFast_r50 modelinin tercih edilmesi gerektiğini, diğer modellerin ise farklı senaryolarda optimize edilmesiyle daha iyi performans gösterebileceğini işaret etmektedir.

Tablo 4. 6. 90° açı eğitim sonuçları

Model Adı	Eğitim Doğruluğu	Validasyon Dengeli Doğruluğu	Test Dengeli Doğruluğu	En İyi Epoch Sayısı	Ortalama Epoch (saniye)
Slowfast_r50	62,52	51,60	49,04	40	373,97
X3D_medium	66,07	51,45	51,15	60	393,56
R3D_18	17,45	28,20	30,50	4	330,07

90° açısına yönelik eğitim sonuçları Tablo 4.6 da gösterilmektedir. Buna göre; SlowFast_r50 modeli, eğitim doğruluğu %62,52, validasyon dengeli doğruluğu %51,60 ve test dengeli doğruluğu %49,04 ile orta düzey bir performans göstermiştir. Bu modelin test dengeli doğruluğu, genelleme kapasitesinin 90° açı verilerinde biraz sınırlı olduğunu işaret etmektedir. Ortalama epoch süresi 373,97 saniyedir ve en iyi sonuçlara 40 epoch sonrasında ulaşmıştır. Eğitim süresi ve doğruluk açısından dengeli bir yapı sunmaktadır. SlowFast_r50, 90° açı verilerinde kararlı bir performans sergilemiş olsa da diğer açılardaki üstün başarısını burada tam anlamıyla yansıtamamıştır. Yine de genelleme kapasitesi açısından güçlü bir aday olmaya devam etmektedir.

X3D_Medium modeli, eğitim doğruluğu %66,07 ile en yüksek değerlerden birini göstermiştir. Validasyon dengeli doğruluğu %51,45 ve test dengeli doğruluğu %51,15 ile, 90° açı verilerinde SlowFast_r50 modeline yakın bir performans sergilemiştir. Bu sonuçlar, modelin genelleme kapasitesinin 90° açı verilerinde biraz daha etkili olduğunu göstermektedir. Ortalama epoch süresi 393,56 saniye ile en uzun eğitim süresine sahip modeldir. En iyi sonuçlara 60 epoch sonunda ulaşması, toplam eğitim süresini artırmış, ancak test doğruluk oranlarında bu süreye karşılık tatmin edici bir kazanç elde edilmiştir. X3D_Medium modeli, 90° açı verilerinde stabil bir performans sergileyerek hem genelleme kapasitesi hem de test doğruluğu açısından rekabetçi bir seçenek sunmuştur. Eğitim süresi uzun olsa da sonuçları dikkate alındığında bu maliyet kabul edilebilir düzeydedir.

R3D_18 modeli, eğitim doğruluğu %17,45, validasyon dengeli doğruluğu %28,20 ve test dengeli doğruluğu %30,50 ile en düşük genel performansı göstermiştir. Bu değerler, modelin 90° açı verilerinde spatio-temporal özellikleri yeterince öğrenemediğini ve genelleme kapasitesinin oldukça sınırlı olduğunu ortaya koymuştur. Ortalama epoch süresi 330,07 saniye ile hesaplama açısından avantaj sağlasa da yalnızca 4 epoch sonunda en iyi sonuçlara ulaşması, modelin öğrenme kapasitesinin sınırlı olduğunu ve daha fazla

eğitime ihtiyaç duymadığını göstermektedir. R3D_18 modeli, düşük eğitim doğruluğu ve genelleme kapasitesiyle, 90° açı verilerinde yeterli performansı sağlayamamıştır.

90° açı verilerine yönelik yapılan analizler, X3D_Medium ve SlowFast_r50 modellerinin birbirine oldukça yakın performans sergilediğini göstermiştir. X3D_Medium modeli, test dengeli doğruluk açısından en yüksek sonuçları elde ederek bu açıda bir miktar avantaj sağlamıştır. Ancak, eğitim süresinin uzunluğu dikkate alındığında SlowFast_r50 modeli daha dengeli bir alternatif olarak değerlendirilebilir. R3D_18 modeli ise düşük doğruluğu ve genelleme kapasitesi ile 90° açı verilerinde beklenen performansı sergileyememiştir. Bu sonuçlar, 90° açı verilerinde X3D_Medium ve SlowFast_r50 modellerinin spatio-temporal duygu tanıma görevlerinde güçlü alternatifler sunduğunu, ancak hesaplama maliyeti ve doğruluk oranları açısından tercihlerin uygulama senaryosuna bağlı olarak yapılması gerektiğini göstermektedir.

Bu çalışmada kullanılan test kodları, örneklemler ve eğitilmiş model dosyaları, GitHub platformu üzerinden açık erişime sunulmuştur [3]. Bu platformda tezimizde de kullanılan test verilerinden her bir sınıf için birkaç örnek rastgele seçilip verinin hazırlanması ve gerçek veriyle eğitilmiş modellerin test edilme senaryosu kodlarla gösterilmiştir.

4.3. Karşılaştırmalı Sonuçlar

Bu doktora tezi kapsamında, kinematik veri seti ve DEMOS veri seti kullanılarak duygu sınıflandırması yapılmış ve her iki veri seti için farklı modellerin başarıları detaylı bir şekilde incelenmiştir. Çalışma, derin öğrenme ve makine öğrenimi tabanlı modellerin, duygusal ifadelerin doğru şekilde sınıflandırılmasında nasıl performans gösterdiğini ve hangi yöntemlerin en iyi sonuçları verdiğini karşılaştırmak amacıyla yapılmıştır.

İlk olarak, kinematik veri setinde yapılan analizlere odaklanılmıştır. Bu veri setinde altı farklı duygu sınıfı üzerinde çeşitli makine öğrenimi (ML) ve derin öğrenme (DL) algoritmalarının doğruluk performansları karşılaştırılmıştır. Makine öğrenimi algoritmaları arasında CatBoost, kNN, RF ve XGBoost yer alırken, derin öğrenme modelleri arasında GRU, LSTM, MobileNetV3-Large ve RegNetY-800MF gibi gelişmiş modeller bulunmaktaydı. Her bir modelin başarısı, kullanılan ham veri ve öznitelik

[3] https://github.com/ahalikoguz/DEMOS_emotion-recognition

çıkarılmış veri (FE) üzerine yapılmış testler ile değerlendirilmiştir. Sonuçlar, ML modelleri için RF algoritmasının hem ham verilerde hem de FE verilerinde en yüksek başarıyı elde ettiğini göstermiştir. RF algoritması özellikle 40 milisaniyelik pencere boyutunda %99,22 doğrulukla dikkat çekerken, FE verisinde ise CatBoost algoritması daha yüksek doğruluk oranları sergilemiştir. CatBoost FE verisiyle %98,37 doğruluk elde etmiştir. DL algoritmalarında ise RegNetY-800MF en yüksek doğruluk oranlarını sağlayan model olmuştur. MobileNetV3-Large, GRU ve LSTM modelleri de sırasıyla iyi sonuçlar elde etmiştir. RegNetY-800MF özellikle FE verisi ile yüksek doğruluk oranlarına ulaşırken, CNN tabanlı DL sınıflandırıcılarında FE verisinin başarısının, RNN tabanlı modellere göre daha sınırlı kaldığı gözlemlenmiştir. Zaman maliyeti açısından ise, GRU ve LSTM gibi RNN modelleri, MobileNetV3-Large gibi CNN tabanlı modellere kıyasla daha hızlı işlem sürelerine sahip olmuşlardır. Bu bulgular, kinematik veri setinde öznitelik çıkarımının (FE) başarı oranlarını önemli ölçüde artırdığı ve daha verimli modellerin elde edilmesini sağladığını göstermiştir.

Bunun yanı sıra, DEMOS veri seti üzerinde yapılan analizde, altı duygu sınıfını sınıflandırmaya yönelik derin öğrenme tabanlı modellerin performansları karşılaştırılmıştır. Bu modeller arasında cnn3d_and_lstm, R3D_18, Slowfast_R50, X3D_Medium, mc3_18 ve s3d yer almıştır. R3D-18 Modeli eğitim doğruluğunda %99,57 ile en yüksek değeri elde ederken, test doğruluğunda ise %49.02 gibi bir düşüş göstermiştir, bu da modelin aşırı uyum (overfitting) yaptığına işaret etmektedir. Diğer modellerin test doğrulukları ise SlowFast-R50 ile %60.05'e kadar ulaşmıştır ve bu model, doğruluk dengesi açısından da güçlü bir performans sergilemiştir. X3D-Medium modeli ise, test doğruluğu ve F1 skoru açısından iyi bir performans göstermiştir. CNN3D+LSTM ise daha düşük doğruluk oranlarına sahip olmasına rağmen, duygu sınıflandırma görevinde önemli bir öğrenme kapasitesine sahip olduğunu kanıtlamıştır. Duygu sınıflarına göre yapılan değerlendirmelerde, Anger (öfke) sınıfında SlowFast-R50 modeli %71,7 doğrulukla en yüksek başarıyı elde ederken, Fear (korku) sınıfında R3D-18 modelinin yüksek doğruluğu dikkat çekmiştir. Disgust (tiksinme) sınıfında ise SlowFast-R50 ve X3D-Medium modelleri, ortalama doğruluklar sağlarken, mc3_18 ve s3d modelleri bu sınıfta daha düşük doğruluk oranları göstermiştir. Neutral (nötr) sınıfında SlowFast-R50 modeli %82,19 doğruluk oranı ile en başarılı model olmuşken, X3D-Medium modeli %61,64 doğrulukla ikinci sırada yer almıştır. Sadness (üzüntü) sınıfında ise, X3D-Medium modeli %58,12 doğrulukla en iyi sonucu elde etmiştir.

DEMOS veri setindeki 0°, 45° ve 90° açılardan elde edilen verilerle yapılan analizlerde, görüş açısının spatio-temporal duygu tanıma başarısı üzerindeki etkisi detaylı bir şekilde incelenmiştir. 0° açısında, SlowFast_R50 modeli, %62,34 test dengeli doğruluk oranı ile en yüksek performansı sergilerken, X3D_Medium modeli %47,23 doğruluk oranı ile orta düzeyde bir başarı göstermiştir. R3D_18 modeli ise düşük doğruluk oranı (%33,50) ve genelleme kapasitesi ile bu açıda sınırlı bir performans sergilemiştir. 45° açısında, SlowFast_R50 modeli %62,73 test dengeli doğruluk ile yine en yüksek performansı sağlamış, X3D_Medium modeli ise %37,40 ile sınırlı bir başarı elde etmiştir. 90° açısında ise X3D_Medium modeli %51,15 test dengeli doğruluk oranı ile sınırlı bir farkla en iyi sonuçları elde etmiş, SlowFast_R50 modeli ise %49,04 doğruluk oranı ile ikinci sırada yer almıştır. Bu sonuçlar, görüş açısının duygu tanıma başarı oranlarına etkisinin önemli olduğunu ve modellerin performanslarının farklı açılarda değişiklik gösterebileceğini ortaya koymuştur.

Her iki veri seti için yapılan karşılaştırmaların genelinde, kinematik veri setinde RegNetY-800MF ve CatBoost modelleri en yüksek doğrulukları elde ederken, DEMOS veri setinde ise SlowFast-R50 ve X3D-Medium modelleri daha iyi sonuçlar vermiştir. Kinematik veri setindeki başarının, öznitelik çıkarımının etkisiyle büyük ölçüde artırıldığı ve ham verilerle yapılan sınıflandırmanın performansını arttırmak için daha fazla hesaplama gücü gerektiği sonucuna varılmıştır. DEMOS veri setinde ise, her bir duygu sınıfının doğru sınıflandırılması için modellerin farklı başarılara sahip olduğu, bazı modellerin belirli sınıflarda daha güçlü performans gösterdiği gözlemlenmiştir. Örneğin, Fear ve Neutral gibi sınıflarda R3D-18 ve SlowFast-R50 modelleri öne çıkarken, Disgust ve Sadness gibi sınıflarda daha fazla iyileştirme yapılması gerektiği anlaşılmıştır.

Bu iki veri seti arasındaki farklar, kullanılan algoritmaların duygu sınıflandırma yeteneklerini ve sınıflar arasındaki karışıklık oranlarını ortaya koymaktadır. Kinematik veri seti, daha fazla fiziksel hareket ve pozisyon verisi içerdiği için, doğru sınıflandırma başarılarını elde etmek için öznitelik çıkarımının çok önemli bir rol oynadığını ortaya koymuştur. DEMOS veri seti ise, daha çok videolar üzerinden duygu sınıflandırması yapılması gerektiği bir bağlamda çalışılmakta olup, burada CNN tabanlı modellerin performansı, özellikle belirli duygu sınıflarında daha iyi sonuçlar elde edilmesini sağlamıştır. Bu sonuçlar, her iki veri seti ile yapılan çalışmanın, duygu sınıflandırma modellerinin etkinliğini ve uygulama alanlarını değerlendirmede oldukça faydalı olduğunu göstermiştir.

4.4. Literatürle Karşılaştırma ve Değerlendirme

Bu çalışmada ham kinematik veriler kullanılarak gerçekleştirilen duygu tanıma literatürü detaylı bir şekilde analiz edilmiştir. Çalışmalar, kullanılan veri toplama yöntemleri, anatomik düğüm sayısı, katılımcı sayısı, örnek sayısı, duygu sınıfları ve uygulanan makine öğrenmesi (ML) ve derin öğrenme (DL) teknikleri açısından incelenmiştir. Bu analizlerin sonunda literatürdeki eğilimler, sınırlılıklar ve boşluklar belirlenmiştir. Sonuçlar, Tablo 4.7 ve 4.8'de detaylı bir şekilde özetlenmiştir.

Literatürde, Kinect ve Motion Capture (MocaP) sistemleri en yaygın kullanılan veri toplama yöntemleridir. Kinect sisteminin kullanımına ilişkin çalışmalar genellikle düşük anatomik düğüm sayılarıyla sınırlıdır (11–25 arasında), buna karşın MocaP daha geniş anatomik kapsama alanı sunarak 72 düğüme kadar ulaşmıştır. Bu durum, MocaP sistemlerinin özellikle hareket dinamiklerinin detaylı analizinde üstünlük sağladığını göstermiştir. Bununla birlikte, Kinect sisteminin daha hızlı ve düşük maliyetli bir seçenek olduğu görülmüştür.

Veri setlerinin içeriği pozisyon (Pos) ve rotasyon (Rot) bilgileriyle sınırlı kalırken, yalnızca birkaç çalışma (örn. Ahmed vd., 2020) RGB video verilerini de entegre etmiştir. Bu tür multimodal yaklaşımlar, duygu tanıma doğruluğunu artırabilme potansiyeli taşımaktadır.

Tablo 4.7'te görüldüğü gibi, katılımcı sayısı ve örnek büyüklüğü çalışmalardaki doğruluk ve genellenebilirlik açısından kritik bir faktördür. Çalışmalar arasında katılımcı sayısının 6 ile 72 arasında değiştiği, örnek büyüklüğünün ise 50 gibi düşük seviyelerden başlayarak 1835'e kadar ulaştığı gözlemlenmiştir. Özellikle Wu vd. (2021), Ghaleb vd. (2021) ve Farinelli (2022) gibi çalışmalarda daha yüksek örnek sayısı kullanılarak daha sağlam sonuçlar elde edilmiştir.

Duygu sınıflandırmasında öfke, korku, mutluluk, nötr, üzüntü ve şaşkınlık gibi temel duyguların çoğunlukla yer aldığı gözlenmiştir. Bununla birlikte, Saha vd. (2014) ve Avola vd. (2022) gibi bazı çalışmalar, daha az yaygın olan duygusal durumları (örneğin rahatlama, zafer) da dahil ederek daha geniş bir duygusal yelpazeyi hedeflemiştir. Ancak bu tür genişletilmiş sınıflandırmaların, genellikle daha az veri ve sınırlı katılımcı sayısı nedeniyle düşük doğruluk oranlarına sahip olduğu dikkat çekmiştir.

Tablo 4.7'te belirtildiği üzere, makine öğrenmesi ve derin öğrenme yöntemlerinin kullanımı yıllar içinde önemli bir evrim geçirmiştir. Erken dönem çalışmalarda (örn. Saha vd., 2014) SVM, kNN ve AdaBoost gibi geleneksel ML teknikleri tercih edilirken, son

dönem çalışmalarda (örn. Farinelli, 2022; Bhatia vd., 2022) LSTM, GRU ve MobileNetV3 gibi derin öğrenme yöntemleri öne çıkmıştır. Ayrıca, Farinelli (2022) gibi çalışmalar hem ML hem de DL yöntemlerini bir arada kullanarak karma bir model önerisi sunmuştur. Bu tür hibrit yaklaşımlar, duygu tanıma performansını artırmada umut vaat etmektedir.

Ham kinematik verilerle duygu tanıma çalışmaları, veri toplama yöntemlerindeki çeşitlilik, katılımcı ve örnek büyüklükleri, duygu sınıfları ve kullanılan modeller açısından geniş bir yelpaze sunmaktadır. MocaP sistemleri, yüksek anatomik doğruluk ve veri kapsamıyla öne çıkarken, Kinect sistemleri düşük maliyet avantajı sağlamaktadır. Ancak, sınırlı sayıda çalışmanın multimodal veri setlerini (örneğin, RGB video ile pozisyon/rotasyon birleşimi) kullandığı görülmektedir. Gelecekte, bu tür verilerin daha yaygın kullanımı, duygu tanıma modellerinin doğruluğunu artırabilir.

Özellikle derin öğrenme tabanlı modellerin kullanımıyla birlikte, duygu tanıma doğruluğu artmış olsa da katılımcı sayısındaki sınırlamalar ve veri setlerinin homojen olmayan yapısı, genel sonuçların genellenebilirliğini sınırlamaktadır. Bu nedenle, daha büyük, çeşitli ve standartlaştırılmış veri setlerinin oluşturulması, literatürdeki bu önemli boşluğu doldurabilir.

Sonuç olarak, Tablo 4.7 verileri ışığında, ham kinematik verilerle duygu tanıma alanında daha fazla multimodal yaklaşımın benimsenmesi, katılımcı çeşitliliğinin artırılması ve derin öğrenme yöntemlerinin daha fazla optimize edilmesi, duygu tanıma doğruluğunda ve genellenebilirlikte önemli ilerlemeler sağlayabilir.

Tablo 4. 7. İskelet tabanlı duygu tanıma üzerine yapılan çalışmalar

Referans bilgisi	Vücut pozisyonu temeli	Veri seti toplama yöntemi	Veri seti orijinal içeriği	Çalışmada kullanılan anatomik noktalar	Denek sayısı	Veri setinin orijinal örnek sayısı	İncelenen duygular	ML/DL tekniği	Kriter	Sonuç
Saha vd., 2014	Duruş	Kinect	Pos, Rot	11	10	50	[Öf, Ko, Mt, Ra, Üz]	AdaBoost, DT, kNN, MLP, SVM	Doğruluk	AdaBoost kullanılarak %90,83
Fourati ve Pelachaud, 2015	Hareket	MocaP	Pos, Rot	23	11	8206	[Öf, Ank, Ne, Nt, PF, Pr, Üz, Ut]	RF	Doğruluk	%84,8
Daoudi vd., 2017	Hareket	MocaP	Pos, Rot	43	8	156	[Öf, Ko, Ne, Nt, Üz]	kNN	Doğruluk	%71,12
Sapiński vd., 2019	Duruş	KinectV2	Pos, Rot	25	16	474	[Öf, Tk, Ko, Mt, Nt, Üz, Şa]	CNN, RNN, RNN+LSTM	Doğruluk	RNN+LSTM kullanılarak altı sınıf için %72, dört sınıf için %82,7
Ahmed vd., 2020	Hareket	KinectV2	Pos, Rot (RGB video)	15	30	300	[Öf, Ko, Mt, Nt, Üz]	DT, GNB, kNN, LDA, SVM	Doğruluk	Yürüme için %90, oturma için %96, bağımsız aksiyon için %86,66
Razzaq vd., 2020	Duruş	KinectV2	Pos, Rot	15	6	NI	[Öf, Ko, Mt, Nt, Üz, Şa]	SVM	Doğruluk	%96,73
H. Zhang vd., 2021	Duruş	KinectV2	Pos, Rot	25	16	474	[Öf, Tk, Ko, Mt, Nt, Üz, Şa]	AS-LSTM	Doğruluk	Yedi sınıf için %74,1, altı sınıf için %75, dört sınıf için %94,2 (AS-LSTM kullanılarak)

Referans bilgisi	**Vücut pozisyonu temeli**	**Veri seti toplama yöntemi**	**Veri seti orijinal içeriği**	**Çalışmada kullanılan anatomik noktalar**	**Denek sayısı**	**Veri setinin orijinal örnek sayısı**	**İncelenen duygular**	**ML/DL tekniği**	**Kriter**	**Sonuç**
Zacharatos vd., 2021	Duruş	MocaP	Pos, Rot	17	NI	402	[Mt, Üz]	Inception-v3	Doğruluk	%81
Ghaleb vd., 2021	Duruş	MocaP	Pos, Rot	21	22	1402	[Öf, Tk, Ko, Mt, Nt, Üz, Şa]	ST-GCN	Doğruluk	Veri ayrımı olmadan %65
Bhatia vd., 2022	Hareket	MocaP	Pos	16	1	1835	[Öf, Mt, Nt, Üz]	LSTM+MLP	Macro mAP, Micro mAP	Makro mAP için %86, makro mAP için %97
Avola vd., 2022	Duruş	MocaP	Pos, Rot	18	11	103	[Kn, Yn, Kos, Tr]	LSTM+MLP	Doğruluk	%71,21
Wu vd., 2022	Duruş	Kinect	Pos	25	15	664	[Öf, Ko, Mt, Üz, Şa]	Bi-LSTM+ (HPN and SAE)	Harmonic mean	Partition1 için %78,25, Partition2 için %66,48
Farinelli L, 2022	Duruş	MocaP	Pos, Rot	72	22	1402	[Öf, Tk, Ko, Mt, Nt, Üz, Şa]	LSTM, GC-LSTM	Doğruluk	LSTM kullanılarak 8 ms veri için %96, 24 ms veri için %93
Bizim çalışmamız	Duruş	MocaP	Pos, Rot	72	22	1402	[Öf, Tk, Ko, Mt, Nt, Üz, Şa]	(CatBoost, kNN, RF, XGBoost) and (GRU, LSTM, MobileNetV3, RegNetY)	Doğruluk	RegNetY kullanılarak ham veri için %99,97, FE ile 40 ms veri için %99,98

Ank: Anksiyete, Hk: Hayal Kırıklığı, Kn: Konsantrasyon, Ko: Korku, Mt: Mutluluk, Ne: Neşe, Nt: Nötr, Öf: Öfke, Ra: Rahatlama, Şa: Şaşkınlık, Tk: Tiksinti, Ut: Utanç, Üz: Üzüntü, Yn: Yenik, Za: Zafer.

Dinamik video verilerinde duygu analizi hem hareket hem de zaman-mekân (spatio-temporal) temelli yöntemlerin etkili bir şekilde uygulanmasını gerektiren kompleks bir problem olarak öne çıkmaktadır. Bu bağlamda, literatür analizine dayanarak, kullanılan veri toplama yöntemleri, duygu sınıfları, anatomik noktalar, veri seti içeriği ve makine öğrenmesi ile derin öğrenme yaklaşımları incelenmiştir. Tablo 4.8 bu literatür incelemesinin detaylı bir özetini sunmaktadır.

Dinamik video verilerinin analizi için kamera tabanlı veri toplama yöntemleri baskın olarak kullanılmaktadır. Çalışmaların büyük çoğunluğu, pozisyon (Pos) ve rotasyon (Rot) verilerini temel alırken, yalnızca birkaç çalışma metin (Txt) ve görüntü (Img) gibi ek veri türlerini entegre etmiştir (örn. Xu vd., 2021; Vaiani vd., 2024). Multimodal veri kaynaklarını kullanan çalışmalar (örn. Chen vd., 2023; Vaiani vd., 2024) daha zengin bilgi sunarken, duygu tanıma doğruluğunu artırmak için önemli bir fırsat yaratmaktadır.

Tablo 4.8 verileri incelendiğinde, katılımcı sayısının 15 ile 25 arasında değiştiği, örnek büyüklüğünün ise 300 ile 500 arasında değiştiği görülmektedir. Bu sayıların, genelleme açısından sınırlı olduğu, ancak duygu analizi modellerinin temel doğrulama ve geçerlilik testleri için yeterli olduğu görülmüştür. Özellikle Wang vd. (2023) ile Xu vd. (2021) çalışmalarında yüksek örnek sayılarıyla daha kapsamlı bir analiz gerçekleştirilmiştir.

Çalışmalar genel olarak mutluluk, üzüntü, öfke, korku ve şaşkınlık gibi temel duygulara odaklanmıştır. Bununla birlikte, Xu vd. (2021) ve Vaiani vd. (2024) çalışmalarında rahatlama (Ra) gibi daha ince duygu durumlarının da analize dahil edildiği gözlemlenmiştir. Geniş bir duygu spektrumunun ele alınması, duygu analizi modellerinin gerçek dünyadaki karmaşık duygusal ifadeleri daha iyi kavramasına olanak sağlamaktadır.

Tablo 4. 8. Dinamik video verisi ile duygu tanıma üzerine yapılan çalışmalar

Araştırmacılar	Yıl	Vücut Pozisyonu Temeli	Veri Toplama Yöntemi	Veri Seti Orijinal İçeriği	Denek Sayısı	Veri Seti Orijinal Örnek Sayısı	Duygular	ML/DL Tekniği
Zhang vd.	2018	Davranış hareketi	Kamera	Pos, Vel	30	300	Öf, Mt, Üz	CNN-RNN
Kahou vd.	2019	Davranış hareketi	Kamera	Pos, Vel	30	300	Mt, Şa, Ko	Deep Neural Networks
Wang vd.	2021	Duruş	Kamera	Pos, Rot	50	500	Öf, Mt, Üz, Ko	SVM, CNN
Xu vd.	2019	Duruş	Kamera	Pos, Txt	50	500	Mt, Üz, Ko, Ra	CNN, Attention
Zhang vd.	2021	Keyframes	Kamera	Pos, Img	40	400	Öf, Mt, Üz	Keyframe Extraction, CNN
Zhang vd.	2022	Duruş	Kamera	Pos, Img	40	400	Öf, Ko, Mt	Reinforcement Learning
Abdurrahman vd.	2022	Davranış hareketi	Kamera	Pos, Txt	35	350	Mt, Üz, Ko, Şa	CNN, RNN
Chen vd.	2023	Multimodal	Kamera	Pos, Img	40	400	Öf, Ko, Mt, Ra, Üz	AdaBoost, DT, kNN, MLP, SVM
Wang vd.	2023	Duruş	Kamera	Pos, Rot	50	500	Öf, Mt, Üz	CNN
Vaiani vd.	2024	Multimodal	Çok Modlu Kaynaklar	Pos, Rot, Txt	35	350	Öf, Mt, Üz, Şa, Ra	Transformers, LLM

Ko: Korku, Mt: Mutluluk, Öf: Öfke, Ra: Rahatlama, Şa: Şaşkınlık, Üz: Üzüntü

Tablo 4.8'de belirtildiği üzere, literatürde CNN, RNN, SVM ve Transformers gibi çeşitli teknikler kullanılmıştır. Geleneksel makine öğrenmesi yöntemleri (örn. AdaBoost, kNN, MLP) daha önceki çalışmalarda öne çıkarken (örn. Chen vd., 2023), daha yeni çalışmalarda Transformers ve büyük dil modelleri (LLM) gibi modern yaklaşımlar benimsenmiştir (örn. Vaiani vd., 2024). Bu durum, duygu analizi modellerinde derin öğrenmenin giderek baskın hale geldiğini ve daha yüksek doğruluk oranları sağladığını göstermiştir.

Dinamik video verilerinde duygu analizi üzerine yapılan çalışmalar, kamera tabanlı veri toplama yöntemleri ve spatio-temporal özelliklerin kullanımı açısından güçlü bir temel sunmaktadır. Bununla birlikte, Tablo 4.8'de görüldüğü gibi, veri seti büyüklükleri genellikle sınırlıdır ve daha büyük, çeşitli veri setlerinin oluşturulması literatürde önemli bir gereklilik olarak ortaya çıkmıştır.

Son yıllarda multimodal veri kaynaklarının ve Transformers tabanlı modellerin kullanımı, duygu analizi performansında önemli bir iyileşme sağlamış olsa da duygu sınıflarının kapsamı hâlâ genişletilmelidir. Özellikle daha ince duygu durumlarının (örneğin rahatlama ve hayal kırıklığı) analizine odaklanmak, modellerin gerçek dünyadaki duygusal ifadeleri daha iyi anlamasına katkı sağlayacaktır.

Gelecekte, multimodal yöntemlerin daha yaygın olarak benimsenmesi, katılımcı ve örnek sayısındaki artışla birleşerek duygu analizi modellerinin doğruluğunu ve genellenebilirliğini artırabilir. Bu bağlamda, Tablo 4.8'de yer alan veriler, mevcut yaklaşımların güçlü ve zayıf yönlerini göstererek, ileride yapılacak çalışmalar için bir rehber niteliği taşımaktadır.

Bu tezde, gerçek zamanlı kullanım için uygun model mimarilerinin incelenmesi özellikle ilgi çekici bir odak noktası oluşturmaktadır. Bu modellerin geniş eklem hiyerarşik verileri üzerinde milisaniyeler içinde işlem yapabilmesi, gerçek zamanlı uygulamalar için büyük bir potansiyel sunmaktadır. Kinematik verilerden duygu tespiti için gerçek zamanlı modellerin kullanımı, çeşitli alanlarda anlamlı yorumlar sağlayabilir. Örneğin, oyun sektöründe, oyuncuların duygularını tespit etmek ve bu duygulara uygun bir yanıt veren bir rol belirlemek, dikkat çekici bir uygulama potansiyeline sahiptir. Ayrıca, benzer bir model yaklaşımına dayalı bir kimlik doğrulama yönteminin etik sınırlar içerisinde oldukça uygulanabilir olması da mümkündür. Bu ve benzeri olasılıklar, bu araştırmanın önemini daha da artırmıştır.

5. SONUÇLAR VE ÖNERİLER

Bu doktora çalışması, ham kinematik ve video tabanlı veri setlerini kullanarak beden hareketlerine dayalı duygu tanıma alanında kapsamlı bir analiz sunmuştur. Araştırma, beden hareketlerinin duygusal durumların doğru şekilde sınıflandırılmasında önemli bir bilgi kaynağı sunduğunu göstermiş ve bu alanda kullanılan makine öğrenimi ve derin öğrenme tabanlı yaklaşımların performansını değerlendirmiştir. Çalışmanın temel bulguları, kinematik veriler ve video tabanlı verilerin duygu tanımada etkili olduğunu ortaya koymuştur.

Ham kinematik verilerin detaylı analizleri, bu verilerin duygu tanıma süreçlerinde yüksek doğruluk oranlarına ulaştığını ve düşük gürültü seviyeleriyle etkili sonuçlar sunduğunu göstermiştir. Özellikle, Random Forest ve CatBoost gibi makine öğrenimi algoritmalarının, özellik çıkarımı yapılmış (FE) verilerle kullanıldığında yüksek doğruluk oranlarına ulaştığı tespit edilmiştir. Derin öğrenme modelleri arasında ise RegNetY-800MF, diğer modellerden daha üstün performans sergilemiştir. Bununla birlikte, RNN tabanlı yaklaşımların hem ham sinyal hem de video tabanlı analizlerde CNN tabanlı modellere kıyasla daha sınırlı kaldığı gözlemlenmiştir. Kinematik verilerin, yüz ifadelerinin yetersiz kaldığı durumlarda dahi beden hareketleriyle duygusal durumların güvenilir bir şekilde tanımlanabildiğini göstermesi, bu veri türünün önemini vurgulamaktadır.

DEMOS veri setiyle gerçekleştirilen çalışmalarda, mekânsal-zamansal özelliklerin duygu tanıma süreçlerine katkılar sunduğu görülmüştür. SlowFast-R50 ve X3D-Medium modelleri, farklı açılardan en iyi performansı sergileyen algoritmalar olarak öne çıkmış; ancak her bir modelin belirli duygusal sınıflarda farklı başarı oranlarına sahip olması, bağlama duyarlı model seçiminin önemini ortaya koymuştur. Örneğin, öfke ve nötr sınıflarında SlowFast-R50 daha başarılı sonuçlar verirken, korku ve üzüntü sınıflarında X3D-Medium üstün bir performans göstermiştir. Bu bulgular, duygu tanıma sistemlerinde model seçiminin bağlama göre optimize edilmesi gerektiğini işaret etmektedir.

Çalışmada elde edilen sonuçlar, kinematik verilerin yüksek doğruluğunu ve video verilerinin bağlamsal zenginliğini birleştirerek çok modaliteli bir duygu tanıma sisteminin oluşturulabileceğini göstermiştir. Böyle bir sistemin, sağlık, güvenlik ve insan-

makine etkileşimi gibi alanlarda geniş uygulama potansiyeline sahip olabileceği düşünülmektedir. Özellikle sağlık sektöründe, hastaların duygusal durumlarını izleyerek tedavi süreçlerini iyileştiren sistemlerin geliştirilmesi, bu alandaki teknolojik yeniliklerin etkinliğini artırabilir.

Bu araştırma, değerli katkılar sunmakla birlikte, bazı sınırlamalarının da olduğunu kabul etmek önemlidir. Çalışmada kullanılan veri seti, ham sinyal senaryosunda yalnızca yedi, video senaryosunda ise yalnızca altı duyguyu içermektedir. Bu durum, gerçek hayattaki duyguların karmaşıklığını tam olarak temsil edememe riskini taşımaktadır. Ayrıca, veri seti benzerlerinden daha kapsamlı olmasına rağmen, fiziksel engeli olmayan sınırlı sayıda katılımcıdan ve yalnızca ayakta durma pozisyonunda toplanmıştır. Daha geniş ve çeşitli bir örneklem grubu ile doğal ortamları yansıtan bir veri toplama süreci, bu alandaki araştırmalara daha güvenilir ve uygulanabilir sonuçlar sağlayabilir.

Bunun yanı sıra, kinematik ve video tabanlı veri türlerinin bazı teknik sınırlamaları da dikkat çekmektedir. Kinematik verilerin laboratuvar ortamında toplanması, gerçek dünya koşullarındaki uygulanabilirliği sınırlandırabilir. Video tabanlı veriler ise yüksek işlem maliyetleri ve aşırı öğrenme (overfitting) riski gibi zorluklar barındırmaktadır. Bu sınırlamaların üstesinden gelmek için veri artırma, transfer öğrenme ve model küçültme gibi optimizasyon tekniklerinin daha fazla araştırılması gerekmektedir.

Ayrıca Avrupa Birliği'nin 2023 yılında kabul ettiği Yapay Zekâ Yasası (Artificial Intelligence Act), biyometrik veriye dayalı duygu tanıma ve algılama sistemlerinin etik ve yasal çerçevede kullanımına dair güçlü düzenlemeler getirmiştir [4]. Bu düzenlemeler, bireylerin temel haklarını ve özgürlüklerini koruma amacıyla, özellikle iş yerleri ve eğitim kurumlarında duygu tanıma sistemlerinin kullanımını sınırlandırmakta veya yasaklamaktadır. Tez çalışmasında önerilen çok modaliteli duygu tanıma sistemlerinin etik çerçevede geliştirilmesi ve uygulanması, bu tür yasal düzenlemelerle uyumlu olacak şekilde dikkatle ele alınmalıdır. Gelecekte yapılacak çalışmaların, AB'nin belirlediği etik standartları göz önünde bulundurarak, duygu tanıma teknolojilerinin kullanım alanlarını ve sınırlarını daha detaylı şekilde araştırması önerilmektedir. Bu bağlamda, önerilen sistemlerin tasarımında şeffaflık, kullanıcı onayı ve veri güvenliği gibi prensiplerin öncelikli hale getirilmesi önem arz etmektedir.

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Sonuç olarak, bu tez çalışması, kinematik ve video tabanlı verileri birleştiren yenilikçi bir çerçeve sunmuş ve beden hareketlerine dayalı duygu tanıma alanında kıymetli katkılar sağlamıştır. Gelecekteki çalışmaların, çok modaliteli veri kullanımını destekleyen sistemlerin geliştirilmesine, model optimizasyonuna ve veri çeşitliliğini artırmaya odaklanması önerilmektedir. Bu tür sistemlerin, insan-makine etkileşimini geliştirme, sağlık ve güvenlik uygulamalarını iyileştirme gibi çeşitli alanlarda faydalı etkiler yaratması beklenmektedir.

KAYNAKLAR

Abdulkareem, N. M., & Abdulazeez, A. M. (2021). Machine learning classification based on random forest algorithm: A review. *International Journal of Science and Business, 5*(2), 128–142. https://ideas.repec.org/a/aif/journl/v5y2021i2p128-142.html

Abdurrahman, A. H., Hossain, M. S., & Muhammad, G. (2022). Video-based emotion estimation using deep neural networks: A review. *Lecture Notes in Computer Science, 1142*, 22–32.

Abro, A. A., Khan, A. A., Talpur, M. S. H., Kayijuka, I., & Yaşar, E. (2021). Machine learning classifiers: A brief primer. *University of Sindh Journal of Information and Communication Technology, 5*(2), 63–68. https://sujo.usindh.edu.pk/index.php/USJICT/article/view/2373

Ahmed, F., Bari, A. S. M. H., & Gavrilova, M. L. (2020). Emotion recognition from body movement. *IEEE Access, 8*, 11761–11781. https://doi.org/10.1109/ACCESS.2019.2963113

Al-Khater, W., & Al-Madeed, S. (2024). Using 3D-VGG-16 and 3D-ResNet-18 deep learning models and FABEMD techniques in the detection of malware. *Alexandria Engineering Journal, 89*, 39–52. https://doi.org/10.1016/j.aej.2023.12.061

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46*(3), 175–185. https://doi.org/10.2307/2685209

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*, Article 74. https://doi.org/10.1186/s40537-021-00444-8

Arığ, E. (2020). Video duygu analizi [Yüksek lisans tezi, İstanbul Ticaret Üniversitesi]. *İstanbul Ticaret Üniversitesi Tez Arşivi*. https://katalog.ticaret.edu.tr/e-kaynak/tez/88752.pdf

Arığ, E., & Turan, M. (2020). Video duygu analizi. *Avrupa Bilim ve Teknoloji Dergisi, 61*, 133–143. https://doi.org/10.31590/ejosat.779059

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Avola, D., Cinque, L., Fagioli, A., Foresti, G. L., & Massaroni, C. (2022). Deep temporal analysis for non-acted body affect recognition. *IEEE Transactions on Affective Computing, 13*(3), 1366–1377. https://doi.org/10.1109/TAFFC.2020.3003816

Bailenson, J. N. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do.* W.W. Norton & Company.

Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., & Zhou, J. (2017). Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 65–74). https://doi.org/10.1145/3097983.3097997

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Bhatia, Y., Bari, A. H., Hsu, G.-S. J., & Gavrilova, M. (2022). Motion capture sensor-based emotion recognition using a bi-modular sequential neural network. *Sensors, 22*(1), Article 403. https://doi.org/10.3390/s22010403

Cao, Z., Hidalgo, G., Simon, T., Wei, S. -E., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 161–168). https://doi.org/10.1145/1143844.1143865

Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ebrahim, N. A. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science, 9*(5), 18–26. https://doi.org/10.5539/ass.v9n5p18

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chen, J., Liu, Z., & Li, P. (2023). Video multimodal emotion recognition system for real-world applications. *arXiv preprint*, arXiv:2308.14320. https://doi.org/10.48550/arXiv.2308.14320

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical*

Methods in Natural Language Processing (EMNLP) (pp. 1724–1734). https://doi.org/10.3115/v1/D14-1179

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint*, arXiv:1412.3555. https://doi.org/10.48550/arXiv.1412.3555

Chun-Lin, L. (2010). A tutorial of the wavelet transform. *NTUEE*, 21, 22.

Daoudi, M., Berretti, S., Pala, P., Delevoye, Y., & Del Bimbo, A. (2017). Emotion recognition by body movement representation on the manifold of symmetric positive definite matrices. In *Lecture Notes in Computer Science* (Vol. 10484, pp. 550–560). Springer. https://doi.org/10.1007/978-3-319-68560-1_49

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets, and applications. *Multimedia Tools and Applications, 82*(6), 9243–9275. https://doi.org/10.1007/s11042-022-13644-y

Doğan, F., & Türkoğlu, İ. (2019). Derin öğrenme modelleri ve uygulama alanlarına ilişkin bir derleme. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 10*(2), 409–445.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition, 44*(3), 572–587. https://doi.org/10.1016/j.patcog.2010.09.020

Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., & Feichtenhofer, C. (2021). Multiscale vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6824–6835).

Farinelli, L. (2022). Design and implementation of a multi-modal framework for scenic actions classification in autonomous actor-robot theatre improvisations. Politecnico di Milano. https://www.politesi.polimi.it/handle/10589/186325

Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). SlowFast networks for video recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6202–6211).

Feichtenhofer, C. (2020). X3D: Expanding architectures for efficient video recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 203–213). https://doi.org/10.1109/CVPR42600.2020.00028

Fletcher, S., & Islam, Md. Z. (2020). Decision tree classification with differential privacy. *ACM Computing Surveys, 52*(1), Article 21. https://doi.org/10.1145/3337064

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Ghaleb, E., Mertens, A., Asteriadis, S., & Weiss, G. (2021). Skeleton-based explainable bodily expressed emotion recognition through graph convolutional networks. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)* (pp. 1–8). IEEE. https://doi.org/10.1109/FG52635.2021.9667052

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation, 101*(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems, 28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Hassan, A., Mahmood, A., & Khan, A. (2021). A review of machine learning algorithms for text-documents classification. *Journal of King Saud University-Computer and Information Sciences, 33*(4), 447–461. https://doi.org/10.1016/j.jksuci.2018.09.014

Hara, K., Kataoka, H., & Satoh, Y. (2018). Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6546–6555). https://doi.org/10.48550/arXiv.1711.09577

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP), 5*(2), 1–15. https://doi.org/10.5121/ijdkp.2015.5201

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. *arXiv*. http://arxiv.org/abs/1905.02244

Hume-Reaction, X., & MLLM Consortium. (2024). Emotion recognition from videos using multimodal large language models. *Future Internet, 16*(7), Article 247. https://doi.org/10.3390/fi16070247

Japkowicz, N. (2013). Assessment metrics for imbalanced learning. In *Imbalanced learning: Foundations, algorithms, and applications* (pp. 187–206). Wiley.

Jovic, A., Brkic, K., & Bogunovic, N. (2015). A review of feature selection methods with applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1200–1205). https://doi.org/10.1109/MIPRO.2015.7160458

Kahou, S. E., Bouthillier, X., Lamblin, P., Gülçehre, C., & Bengio, Y. (2019). EmoNets: Multimodal deep learning approaches for emotion recognition in video. *Journal of Multimodal User Interfaces, 13*(3), 367–380. https://doi.org/10.1007/s12193-015-0195-2

Kaynar, O., Yıldız, M., Görmez, Y., & Albayrak, A. (2016). Makine öğrenmesi yöntemleri ile duygu analizi - Sentiment analysis with machine learning techniques. In *Proceedings of the International Artificial Intelligence and Data Processing Symposium (IDAP'16)*. https://www.researchgate.net/publication/311136507

Khanna, P., & Sasikumar, M. (2015). Recognizing emotions from keyboard stroke pattern. *International Journal of Computer Applications, 111*(15), 12–16. http://dx.doi.org/10.5120/1614-2170

Kumar, A., & Talukdar, S. S. (2008). Semi-supervised clustering with metric learning using relative comparisons. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 616–624). https://doi.org/10.1109/TKDE.2007.190715

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems* (pp. 396–404). https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf

Lee, T.-H., Ullah, A., & Wang, R. (2019). Bootstrap aggregating and random forest. In *Macroeconomic Forecasting in the Era of Big Data* (pp. 389–429). https://doi.org/10.1007/978-3-030-02194-8_13

Lin, Y., Wang, R., & Chen, F. (2023). Affective video content analysis: Decade review and new perspectives. *arXiv preprint*, arXiv:2310.17212. https://doi.org/10.48550/arXiv.2310.17212

Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2016). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint*, arXiv:1511.03677. https://doi.org/10.48550/arXiv.1511.03677

Martinez, J., Black, M. J., & Romero, J. (2017). On human motion prediction using recurrent neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2891–2900). https://doi.org/10.1109/CVPR.2017.497

Mathew, S. D. A., Shree, N. D., & Chowdhary, C. L. (2023). An experimental study on the deviations in performance of FNNs and CNNs in the realm of grayscale adversarial images. *Journal of Engineering Science and Technology Review, 16*(3), 66–73. http://dx.doi.org/10.25103/jestr.163.09

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence, 29*(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

Moghe, B., Kachhara, M., & M, K. (2024). Multimodal emotion analysis for depression detection: Integrating facial expression and speech recognition. In *Proceedings of the 2024 Second International Conference on Inventive Computing and Informatics (ICICI)* (pp. 37–42). https://doi.org/10.1109/ICICI62254.2024.00015

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective.* MIT Press.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics, 7*, Article 21. https://doi.org/10.3389/fnbot.2013.00021

Ng, A. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. In *Proceedings of the Twenty-First International Conference on Machine Learning (ICML)*. https://doi.org/10.1145/1015330.1015435

Nguyen, T. H., Tran, D. K., & Vo, N. T. (2023). Multimodal group emotion recognition in-the-wild using privacy-compliant features. *arXiv preprint*, arXiv:2312.05265. https://doi.org/10.48550/arXiv.2312.05265

Oğuz, A., & Ertuğrul, Ö. F. (2022). Human identification based on accelerometer sensors obtained by mobile phone data. *Biomedical Signal Processing and Control, 77*, Article 103847. https://doi.org/10.1016/j.bspc.2022.103847

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning* (pp. 1310–1318). https://proceedings.mlr.press/v28/pascanu13.html

Penney, D. D., & Chen, L. (2019). A survey of machine learning applied to computer architecture design. *arXiv preprint*, arXiv:1909.12373. https://doi.org/10.48550/arXiv.1909.12373

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems* (pp. 6638–6648). https://arxiv.org/abs/1706.09516

Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollar, P. (2020). Designing network design spaces. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10425–10433. https://doi.org/10.1109/CVPR42600.2020.01044

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion, 37*, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Razzaq, M. A., Bang, J., Kang, S. S., & Lee, S. (2020). UnSkEm: Unobtrusive skeletal-based emotion recognition for user experience. In *Proceedings of the 2020 International Conference on Information Networking (ICOIN)* (pp. 92–96). https://doi.org/10.1109/ICOIN48656.2020.9016601

Robert-Lachaine, X., Mecheri, H., Muller, A., Larue, C., & Plamondon, A. (2020). Validation of a low-cost inertial motion capture system for whole-body motion analysis. *Journal of Biomechanics, 99*, Article 109520. https://doi.org/10.1016/j.jbiomech.2019.109520

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536. https://doi.org/10.1038/323533a0

Saha, S., Datta, S., Konar, A., & Janarthanan, R. (2014). A study on emotion recognition from body gestures using Kinect sensor. In *Proceedings of the 2014 International Conference on Communication and Signal Processing (ICCSP)* (pp. 056–060). https://doi.org/10.1109/ICCSP.2014.6949798

Saganowski, S., Perz, B., Polak, A., & Kazienko, P. (2022). Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2022.3176135

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4510–4520). https://doi.org/10.1109/CVPR.2018.00474

Sapiński, T., Kamińska, D., Pelikant, A., & Anbarjafari, G. (2019). Emotion recognition from skeletal movements. *Entropy, 21*(7), Article 646. https://doi.org/10.3390/e21070646

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus, and Dimensions: A comparative analysis. *Scientometrics, 126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access, 7*, 53040–53065. https://doi.org/10.1109/ACCESS.2019.2912200

Shi, Q., & Cui, Y. (2024). Research on motion capture system of motor skill based on computer vision technology. *Journal of Electrical Systems, 20*(2), 1181–1191.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems* (pp. 28–34).

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404*, Article 132306. https://doi.org/10.1016/j.physd.2019.132306

Srivastava, D., & Bhambhu, L. (2010). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology, 12*(1), 1–7.

Sun, J., Du, W., & Shi, N. (2018). A survey of kNN algorithm. *Information Engineering and Applied Computing, 1*(1). https://doi.org/10.18063/ieac.v1i1.770

Taddeo, M., & Floridi, L. (2019). How AI can be a force for good. *Minds and Machines, 29*(1), 119–129. https://doi.org/10.1007/s10676-019-09508-z

Vaiani, L., Cagliero, L., & Garza, P. (2024). Emotion recognition from videos using multimodal large language models. *Future Internet, 16*(7), Article 247. https://doi.org/10.3390/fi16070247

Wang, Y., Shen, J., Wang, W., & Xie, X. (2021). Video-based emotion recognition using aggregated features and spatio-temporal CNN. *IEEE Transactions on Multimedia*, *20*(6), 1153–1164. https://doi.org/10.1109/TMM.2021.8545441

Wang, Y., Shen, J., Wang, W., & Xie, X. (2023). A multimodal fusion-based deep learning framework combined with keyframe extraction and spatial and channel attention for group emotion recognition from videos. *Pattern Analysis and Applications, 24*, 214–230. https://doi.org/10.1007/s10044-023-01178-4

Wang, J., Li, J., Wang, X., Wang, J., & Huang, M. (2021). Air quality prediction using CT-LSTM. *Neural Computing and Applications, 33*, 4779–4792.

Wu, J., Zhang, Y., Sun, S., Li, Q., & Zhao, X. (2022). Generalized zero-shot emotion recognition from body gestures. *Applied Intelligence, 52*, 8616–8634. https://doi.org/10.1007/s10489-021-02927-w

Wu, X., Sun, H., Xue, J., Zhai, R., Kong, X., Nie, J., & He, L. (2023). eMotions: A large-scale dataset for emotion recognition in short videos. *arXiv preprint arXiv:2311.17335*. https://doi.org/10.48550/arXiv.2311.17335

Xu, B., Zheng, Y., Ye, H., Wu, C., Wang, H., & Sun, G. (2019). Video emotion recognition with concept selection. *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 406–411. https://doi.org/10.1109/ICME.2019.00077

Xue, F., Ji, H., & Zhang, W. (2020). Mutual information guided 3D ResNet for self-supervised video representation learning. *IET Image Processing*. https://doi.org/10.1049/iet-ipr.2020.0019

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation, 31*(7), 1235–1270. https://doi.org/10.1162/neco_a_01199

Zacharatos, H., Gatzoulis, C., Charalambous, P., & Chrysanthou, Y. (2021). Emotion recognition from 3D motion capture data using deep CNNs. In *Proceedings of the 2021 IEEE Conference on Games (CoG)* (pp. 1–5). IEEE. https://doi.org/10.1109/CoG52621.2021.9619065

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2018). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39–58. https://doi.org/10.1109/TPAMI.2008.52

Zhang, C.-B., Jiang, P.-T., Hou, Q., Wei, Y., Han, Q., Li, Z., & Cheng, M.-M. (2021). Delving deep into label smoothing. *IEEE Transactions on Image Processing, 30*, 5984–5996. https://doi.org/10.1109/TIP.2021.3089942

Zhang, H., Yi, P., Liu, R., & Zhou, D. (2021). Emotion recognition from body movements with AS-LSTM. In *Proceedings of the 2021 IEEE 7th International Conference on Virtual Reality (ICVR)* (pp. 26–32). IEEE. https://doi.org/10.1109/ICVR51878.2021.9483833

Zhang, M., Yu, L., Zhang, K., Du, B., Zhan, B., Chen, S., Jiang, X., Guo, S., Zhao, J., Wang, Y., Wang, B., Liu, S., & Luo, W. (2020). Kinematic dataset of actors expressing emotions. *Scientific Data, 7*, Article 292. https://doi.org/10.1038/s41597-020-00635-7

Zhang, M., Yu, L., Zhang, K., Du, B., Zhan, B., Jia, S., Chen, S., Han, F., Li, Y., Liu, S., Yi, X., & Luo, W. (2023). Construction and validation of the Dalian emotional movement open source set (DEMOS). *Behavior Research Methods, 55*, 2353–2366. https://doi.org/10.3758/s13428-022-01887-4

Zhang, S., Zhao, X., & Li, B. (2018). Video-based emotion recognition using CNN-RNN and C3D hybrid networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 158–167). https://openaccess.thecvf.com/content_cvprw_2018/papers/zhang_Video-Based_Emotion_Recognition_CVPRW_2018_paper.pdf

Zhang, S., Zhao, X., Li, B., & Wang, W. (2021). User-generated video emotion recognition based on key frames. *Multimedia Tools and Applications, 80*(1), 10203–10217. https://doi.org/10.1007/s11042-020-10203-1

Zhang, S., Zhao, X., & Li, B. (2022). Real-time video emotion recognition based on reinforcement learning and domain knowledge. *IEEE Transactions on Multimedia, 23*(4), 887–899. https://doi.org/10.1109/TMM.2021.3058589