



**OTİZM SPEKTRUM BOZUKLUKLARINDA KENDİNİ
UYARICI HAREKETLERİN GÖRSEL KELİME
ÇANTASI YAKLAŞIMI VE DERİN ÖĞRENME
YÖNTEMLERİ İLE TANINMASI**

Sibel KAÇDIOĞLU

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Ana Bilim Dalı

2020

(Her hakkı saklıdır.)

T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

**OTİZM SPEKTRUM BOZUKLUKLARINDA KENDİNİ UYARICI HAREKETLERİN
GÖRSEL KELİME ÇANTASI YAKLAŞIMI VE DERİN ÖĞRENME YÖNTEMLERİ
İLE TANINMASI**

(Recognition of Self-Stimulatory Behaviours with Bag-of-Visual-Words Approach and Deep Learning Methods in Autism Spectrum Disorders)

YÜKSEK LİSANS TEZİ

Sibel KAÇDIOĞLU

Danışman: Dr. Öğr. Üyesi Barış ÖZYER
İkinci Tez Danışmanı: Dr. Öğr. Üyesi Saeid AGAHIAN

ERZURUM

Haziran, 2020

KABUL VE ONAY TUTANAĞI

OTİZM SPEKTRUM BOZUKLUKLARINDA KENDİNİ UYARICI HAREKETLERİN GÖRSEL KELİME ÇANTASI YAKLAŞIMI VE DERİN ÖĞRENME YÖNTEMLERİ İLE TANINMASI

Dr. Öğr. Üyesi Barış ÖZYER danışmanlığında, Sibel KAÇDIOĞLU tarafından hazırlanan bu çalışma, 29/06/2020 tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı'nda yüksek lisans tezi olarak **oybirliği / oy çokluğu** (.../...) ile kabul edilmiştir.

Jüri Başkanı:	Doç Dr. Tansu FİLİK <i>Eskişehir Teknik Üniversitesi</i>	
Danışman:	Dr. Öğr. Üyesi Barış ÖZYER <i>Atatürk Üniversitesi</i>	
Jüri Üyesi:	Dr. Öğr. Üyesi Bilal USANMAZ <i>Atatürk Üniversitesi</i>	
Jüri Üyesi:	Dr. Öğr. Üyesi Ahmet COŞKUNÇAY <i>Atatürk Üniversitesi</i>	
İkinci Tez Danışmanı	Dr. Öğr. Üyesi Saeid AGAHIAN <i>Erzurum Teknik Üniversitesi</i>	

Enstitü Yönetim
Kurulunun .../.../.... tarih
ve sayılı kararı.

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

Prof. Dr. Mehmet KARAKAN
Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildiriş, çizelge, şekil ve fotoğrafların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Yüksek Lisans Tezi olarak *Dr. Öğr. Üyesi Barış ÖZYER* danışmanlığında sunulan “Otizm Spektrum Bozukluklarında Kendini Uyarıcı Hareketlerin Görsel Kelime Çantası Yaklaşımı ve Derin Öğrenme Yöntemleri ile Tanınması” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

Tez Bölümleri	Tezin Benzerlik Oranı (%)	Maksimum Oran (%)
Giriş	5	30
Kuramsal Temeller	5	30
Materyal ve Yöntem	6	35
Bulgular	5	20
Tartışma	0	20
Tezin Geneli	6	25

Not: Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'den büyük olmaması gerekir.

Beyan edilen bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ve beyan ederiz.

Tez Yazarı (Öğrenci)	Tez Danışmanı
Sibel KAÇDIOĞLU	Dr. Öğr. Üyesi Barış ÖZYER
6.7.2020	6.7.2020
İmza:	İmza:

* Tez ile ilgili YÖKTEZ’de yayınlamasına ilişkin bir engelleme var ise aşağıdaki alanı doldurunuz.

☒ Tezle ilgili patent başvurusu yapılması / patent alma sürecinin devam etmesi sebebiyle Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 2 (iki) yıl süreyle engellenmiştir.

☐ Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 6 (altı) ay süreyle engellenmiştir.

TEŞEKKÜR

Tezimin konusunun belirlenmesinde, araştırma aşamasında ve tamamlanma sürecindeki destek ve katkılarından dolayı danışmanım Sayın Dr. Öğr. Üyesi Barış ÖZYER'e, kısa sürede tezimin kapsamını genişletmem için katkı sağlayıp yol gösteren eş danışmanım Sayın Dr. Öğr. Üyesi Saeid AGAHIAN'a ve bilgi ve fikirleriyle destek olan Sayın Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER'e teşekkür ederim.

Tez çalışmamda bilgi ve tecrübelerini paylaşarak sağladığı katkılardan dolayı Çocuk ve Ergen Ruh Sağlığı Hastalıkları Anabilim Dalı Başkanı Dr. Öğr. Üyesi İbrahim Selçuk ESİN'e teşekkür ederim.

Tezimin başlangıcından bitimine kadar bana inanan, her zaman yanımda olan ve desteklerini esirgemeyen eşim Tayfur KAÇDIOĞLU'na ve mativasyonumu her daim canlı tutan başta babam olmak üzere KOÇ ve KAÇDIOĞLU ailelerine teşekkür ederim.

Sibel KAÇDIOĞLU

ÖZET

YÜKSEK LİSANS TEZİ

OTİZM SPEKTRUM BOZUKLUKLARINDA KENDİNİ UYARICI HAREKETLERİN GÖRSEL KELİME ÇANTASI YAKLAŞIMI VE DERİN ÖĞRENME YÖNTEMLERİ İLE TANINMASI

Sibel KAÇDIOĞLU

Danışman: Dr. Öğr. Üyesi Barış ÖZYER

İkinci Tez Danışmanı: Dr. Öğr. Üyesi Saeid AGAHIAN

Amaç: Otizm spektrum bozukluğu (OSB), sosyal etkileşim ve iletişim eksiklikleri, sınırlı ve tekrarlanan hareketler gibi belirtileri olan nörobiyolojik bir bozukluktur. Tekrarlanan hareketler, otizm tanısında kullanılan önemli ipuçlarıdır. Bu hareketler, kolları kanat gibi çırpma, ileri-geri kafa sallama, kendi etrafında dönme gibi kendini uyarıcı hareketlerdir. Kendini uyarıcı davranışlar, vücut parçalarının veya nesnelerinin kalıplaşmış, tekrarlayan hareketlerini ifade eder. Otizm tanısı için hekimlerin çocuklarla uzun süre etkileşimde bulunarak bu hareketleri gözlemlemesi gerekmektedir. Bu durum uzun süren seanslarla gerçekleştirildiği için çocuklarda otizmin erken tanısını geciktirmektedir. Bu tez çalışmasında OSB şüphesi bulunan çocukların klinik ortamdan bağımsız günlük yaşamlarında kaydedilmiş video görüntülerinden kendini uyarıcı hareketlerin klasik makine öğrenme yöntemleri ve derin öğrenme yöntemleri ile tanınması amaçlanmıştır.

Yöntem: Bu çalışmada 4 (dört) farklı tekrarlanan hareketten meydana gelen yeni bir veri seti oluşturulmuştur. Hareketlerin algılanması ve tanınması amacı ile Görsel Kelime Çantası yaklaşımı çerçevesinde yaygın olarak kullanılan HOG, HOF, HOG-HOF, SIFT VE SURF tanımlayıcıları ile öznitelik vektörleri çıkarılmıştır. Hareketleri sınıflandırmak için MLP, GNB ve SVM sınıflandırıcıları kullanılmıştır. Ayrıca, LSTM ve 3DCNN derin öğrenme algoritmaları kullanılarak hareketler tanınmaya çalışılmıştır.

Bulgular: HOF öznitelik tanımlayıcısı ve MLP sınıflandırıcısı genel olarak deneylerde iyi sonuç vermiştir. Hareketlerin tekrarlı olmasından dolayı görsel kelime çantası yaklaşımında yeterli ve az sayıda anahtar görsel kelime kullanılması sonucu iyileştirmiştir.

Sonuç: Deneysel sonuçları incelediğimizde, HOF yöntemi 200 anahtar görsel kelime sayısında MLP sınıflandırıcısı ile %80 doğruluk gözlemlenmiştir.

Anahtar Kelimeler: Otizm spektrum bozukluğu; hareket tanıma; kendini uyarıcı davranış; Görsel Kelime Çantası; 3DCNN; ConvLstm.

Haziran 2020, 82 sayfa

ABSTRACT

MASTER'S THESIS

RECOGNITION OF SELF- STIMULATORY BEHAVIOURS WITH BAG-OF-VISUAL-WORDS APPROACH AND DEEP LEARNING METHODS IN AUTISM SPECTRUM DISORDERS

Sibel KAÇDIOĞLU

Danışman: Dr. Öğr. Üyesi Barış ÖZYER

İkinci Tez Danışmanı: Dr. Öğr. Üyesi Saeid AGAHIAN

Purpose: Autism spectrum disorder (ASD) is a neurobiological disorder that some symptoms such as deficit of social interaction and communication, limited and repeated behavior are observed in patients. Repetitive behaviors are significant clues for diagnosis of ASD. These behaviors are self-stimulatory behaviors such as flapping arms like wings, shaking head back and forth, and spinning around itself. Self-stimulatory behaviors refers to the stereotyped, repetitive movements of body parts or objects. These repetitive behaviors, which is called self-stimulating behaviors, are described Physicians should observe and examine these self-stimulating behaviors by interacting with children for a long time that makes it difficult in early diagnosis of ASD. In this paper, it is aimed to recognize the self-stimulatory behaviors from the video recorded in the daily lives of the children with suspected ASD independent of the clinical environment with classical machine learning methods and deep learning methods.

Method: In this study, a new dataset consisting of 4 (four) different repetitive behaviours was created. In order to detect and recognize the behaviours, feature vectors are extracted with the HOG, HOF, HOG-HOF, SIFT and SURF descriptors, which are commonly used in the Bag-of-Visual-Word approach. MLP, GNB and SVM classifiers are used to classify behaviours. Additionally, LSTM and 3DCNN deep learning algorithms are used to recognize the behaviours.

Findings: The HOF features descriptor and MLP classifier generally worked well in experiments. Due to the repetition of the behaviors, using sufficient and few key visual words in the bag-of-visual-word approach improved accuracy result.

Results: When we examine the experimental results, 80% accuracy was observed in the HOF descriptor, MLP classifier and 200 key visual word.

Keywords: Autism spectrum disorder; action recognition; selfstimulatory behaviours; Bag-of-Visual-Words; 3DCNN; ConvLstm.

June 2020, 82 pages

İÇİNDEKİLER

KABUL VE ONAY TUTANAĞI.....	i
ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER.....	vi
TABLolar DİZİNİ.....	viii
ŞEKİLLER DİZİNİ	ix
KISALTMALAR ve SİMGELER DİZİNİ	x
GİRİŞ.....	1
Motivasyon.....	4
Tezin Amacı ve Katkısı.....	5
KURAMSAL TEMELLER.....	8
Otizm Spektrum Bozukluğu ve Tanı Yöntemleri	8
Otizm spektrum bozukluğu	8
Tanı yöntemleri	9
Otizmli Hastalarda Bulunan Davranış Bozuklukları.....	11
Görüntü Tabanlı İnsan Hareketi Tanıma Yaklaşımları	12
Tek katmanlı yaklaşımlar	13
Hiyerarşik yaklaşımlar	15
Davranış Bozukluklarının Video Görüntüleri Üzerinden Analizi	16
Klasik yöntemler	16
Derin öğrenme yöntemleri.	18
MATERYAL VE METOT.....	20
Veri Seti	20
Metot	22
Görsel kelime çantası yaklaşımı ile hareket analizi	22
Derin Öğrenme Yöntemleri.....	38
DENEYSEL SONUÇLAR VE TARTIŞMA	43
Değerlendirme Yöntemleri.....	43
Çapraz doğrulama	43
Performans ölçütleri	44
Önişlem aşamasında insan tespiti yöntemlerinin analizi	45

Görsel Kelime Çantası Yaklaşımı Sonuçları.....	46
Deneylerde kullanılan yöntemlerin parametre değerleri.....	46
Önişlem uygulanmamış veri seti ile alınan sonuçlar	49
Önişlem uygulanmış veri seti ile alınan sonuçlar	51
Önişlem uygulanmış verilerin üçlü histogram temsili sonrası sonuçları	52
Yöntemlerdeki işlem sürelerinin analizi	57
Derin Öğrenme Yöntemlerinin Sonuçları	58
SONUÇLAR VE ÖNERİLER	61
KAYNAKLAR.....	63
ÖZGEÇMİŞ.....	69



TABLÖLAR DİZİNİ

Tablo 1. Veri Setlerindeki Hareket Sınıflarına Ait Video Sayısı.....	21
Tablo 2. Veri Setimizdeki Hareket Sınıflarına Ait Video Kare Sayıları	21
Tablo 3. Videolardaki Kız-Erkek Dağılımı	22
Tablo 4. Karmaşıklık Matrisi.....	44
Tablo 5. Öznitelik Çıkarma Yöntemlerinin Parametre Değerleri	47
Tablo 6. K-means Parametre Değerleri	48
Tablo 7. Kullanılan Sınıflandırma Algoritmalarının Parametre Değerleri	48
Tablo 8. Orijinal Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre Doğruluk (%) Değerleri	49
Tablo 9. Orijinal Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre UAR Doğruluk (%) Değerleri	50
Tablo 10. Önişlem Uygulanmış Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre Doğruluk (%) Değerleri	51
Tablo 11. Önişlem Uygulanmış Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre UAR Doğruluk (%) Değerleri.....	51
Tablo 12. Üçlü Histogram Temsili Sonrası Doğruluk (%) Değerleri.....	53
Tablo 13. Üçlü Histogram Temsili Sonrası UAR Doğruluk (%) Değerleri	53
Tablo 14. En İyi Sonuç Alınan Deneyin 5 Fold İçin Doğruluk Değerleri.....	54
Tablo 15. Uygulamada Kullanılan Bilgisayarların Özellikleri	57
Tablo 16. Yöntemlerin İşlem Süreleri (saniye)	58
Tablo 17. Derin öğrenme yöntemlerinin parametre değerleri	59
Tablo 18. Orijinal Veri Seti İle Doğruluk (%) Değerleri.....	59
Tablo 19. Önişlem Uygulanmış Veri Seti İle Doğruluk (%) Değerleri	59
Tablo 20. Derin Öğrenme Yöntemlerinin İşlem Süreleri	59

ŞEKİLLER DİZİNİ

Şekil 1. Otizm spektrum bozukluğu tanı yöntemleri.....	10
Şekil 2. Görüntü tabanlı insan hareketi tanıma yaklaşımlarının sınıflandırması	12
Şekil 3. Veri setinden örnek görüntüler.....	20
Şekil 4. Yöntemin akış diyagramı (a) Anahtar görsel kelime oluşturma (b) Videoların histogramlarla temsil edilip sınıflandırılması	24
Şekil 5. YOLOv3 ağ mimarisi.....	26
Şekil 6. Önilem adımları.....	28
Şekil 7. Anahtar görsel kelime dizisi oluşturma örneği	33
Şekil 8. Üçlü histogram gösterimi.....	35
Şekil 9. Çok katmanlı algılayıcı yapısı.....	36
Şekil 10. Optimal hiper düzlem.....	37
Şekil 11. 2B ve 3B evrişim katmanları arasında karşılaştırma, (a) 2D evrişim çekirdekleri, (b) 3D evrişim çekirdekleri	39
Şekil 12. 3DCNN bileşenleri.....	40
Şekil 13. Evrişimli LSTM (CovnLSTM) yönteminin şeması	41
Şekil 14. İnsan tespiti yöntemlerinin çıktıları.	45
Şekil 15. Haar-Cascade yönteminde hatalı tespit.	46
Şekil 16. Orijinal veri setinde anahtar görsel kelime sayılarına göre sınıflandırıcıların değerlendirme grafiği.	50
Şekil 17. Önilemli veri setinde anahtar görsel kelime sayılarına göre sınıflandırıcıların değerlendirme grafiği	52
Şekil 18. En iyi sonuç alınan deneyin 5 fold için karmaşıklık matrisleri.....	55
Şekil 19. Önilemli veri setinde üçlü histogram temsili ile anahtar görsel kelime sayılarına göre sınıflandırıcıların değerlendirme grafiği.	56
Şekil 20. Anahtar görsel kelime sayılarına göre aşamaların doğruluk değerleri.....	56
Şekil 21. Aşamalara göre doğruluk analizi (HOF-200 Anahtar Görsel Kelime-MLP).....	57

KISALTMALAR ve SİMGELER DİZİNİ

2DCNN	: 2 Dimensions Convolutional Neural Network
3DCNN	: 3 Dimensions Convolutional Neural Network
ABD	: Amerika Birleşik Devletleri
ADI-R	: Autism Diagnostic Interview - Revised
ADOS	: Autism Diagnosis Observation Schedule
BoVW	: Bag-of-Visual-Words
C3D	: Convolutional 3 Dimensions
CDC	: Centers for Disease Control and Prevention
ConvLSTM	: Convolutional Long-Short Term Memory
CSBS	: Communication and Symbolic Behavior Scales
CT	: Computed Tomography
DBNs	: Dynamic Bayesian Networks
DSM	: Diagnostic and Statistical Manual of Mental Disorders
ECG	: Electrocardiogram
EEG	: Elektroensefalografi
fMRI	: Functional Magnetic Resonance Imaging
GKUDV	: Genişletilmiş Kendini Uyarıcı Davranış Veri Seti
GNB	: Gaussian Naive Bayes
GoD	: Gaussian of Distances
HDM	: Histogram of Dominant Motions
HMM	: Hidden Markov Model
HOF	: Histogram of Optical Flow
HOG	: Histogram of Oriented Gradient
LSTM	: Long-Short Term Memory
MAP	: Maximum a Posteriori Probability
MEG	: Magnetoencephalography
MLE	: Maximum Likelihood Estimation
MLP	: Multi Layer Perceptron
O-GAD	: One Glimpse Early ASD Detection
OSB	: Otizm Spektrum Bozukluğu
RBS-R	: The Repetitive Behavior Scale-Revised
RNN	: Recurrent Neural Network
ROI	: Region of Interest

SCQ	: Social Communication Questionnaire
SIFT	: Scale Invariant Feature Transform
SSBD	: Self-Stimulatory Behaviour Dataset
STAT	: The Screening Tool for Autism in Toddlers and Young Children
STIP	: Space Time Interest Points
SURF	: Speed Up Robust Features
SVM	: Support Vector Machine
TBoW	: Temporal Bag of Words
UAR	: Unweighted Average Recall
YOLOv3	: You Only Look Once Version 3



GİRİŞ

Otizm spektrum bozukluğu (OSB), sosyal etkileşim ve iletişim eksiklikleri, sınırlı ve tekrarlı davranışlar gibi belirtileri olan nörobiyolojik bir bozukluktur. OSB, gelişimsel dönemin gerektirdiği sosyal iletişim ve etkileşimde, dil gelişiminde ve zihinsel faaliyetlerde bireyin becerilerini sınırlamaktadır. Otizmin genetik, biyolojik ve çevresel nedenlerle ortaya çıkıp çıkmadığı tam olarak bilinmemektedir. Otizm spektrum bozukluğunun yaygınlığı ise her geçen yıl artmaktadır (Elsabbagh *et al.* 2012; Zwaigenbaum *et al.* 2015). ABD Sağlık Bakanlığı'nın Hastalık Kontrolü ve Önleme Merkezinde (CDC: Centers for Disease Control and Prevention) yapılan araştırmaya göre 1992 yılında doğan 150 çocuktan 1'ine OSB tanısı konurken, 2000'li yıllarda doğanlarda ise 68 çocuktan 1'i ne OSB tanısı konulmuştur. Bu yıl (2020) yayınladıkları raporda ise otizmin yaygınlığının arttığı ve 54 çocuktan 1'inde görülmeye başlandığı bildirilmiştir (Anonymous 2020).

OSB belirtileri ilk olarak 1943 yılında Leo Kanner tarafından tanımlanmıştır. Kanner, bu tanımlamayı iletişim problemleri, sosyal etkileşim eksikliği ve ileri-geri kafa sallama gibi tekrarlayan davranışları sergileyen 11 çocuğu gözlemleyerek yapmıştır. Ancak Otizm Spektrum Bozukluğu tanısı koymak hekimler için hala zor ve karmaşıklığını koruyan bir süreçtir (Anonymous 2013). Otizmle bağdaştırılan iletişimsel ve davranışsal bazı belirtiler otizme özgü olmayan başka birçok etkenlere bağlı olabilmektedir (Sigman and McGovern 2005; Bishop *et al.* 2006; Richler *et al.* 2007). Bu nedenle hekimler tanı konulurken tek bir ölçüt ya da araç kullanımını önermemektedir ve doğrudan uzun süren klinik gözlemlerle hastalarını (çocukları) takip etmektedirler (Huerta and Lord 2012). Ancak bu yöntemde, çocuk, hekim ile arasındaki etkileşimden etkilenebilmekte ve hekimi otorite olarak görüp farklı davranabilmektedir. Ayrıca hekim bir çocuk üzerindeki davranış değerlendirmesini bir başka çocukta doğru bir şekilde tekrarlayıp birebir aynı değerlendirmeyi sağlayamamaktadır (Lord *et al.* 2012). Her yönüyle bir hekim tarafından ayrıntılı bir gözlem uzun süren bir süreçtir. Araştırmalar, kesin tanı için tek bir gözlemin yeterli olmadığı çocuğun birçok farklı ortamda birden fazla hekim tarafından gözlemlenmesinin gerekli olduğunu göstermektedir. Bu gereklilik değerlendirmeyi uzun ve meşakkatli yaparak erken tanıyı ertelemektedir. OSB tanısı için kullanılan değerlendirme yöntemlerinin yukarıda bahsedildiği gibi çeşitli kısıtlamalar ve zorlukları vardır. Çocuklar gözlemlenirken sosyal iletişim, etkileşim ve davranışsal değerlendirmelerin nicel ölçümlerine ihtiyaç duyulmaktadır (Lam and Aman 2007). Bunun yanında OSB yaygınlığının artması ile günümüzde hekimlerin tanı koyarken yardım aldıkları

standartlaştırılmış klinik değerlendirme araçları oluşturulmuştur. Bu araçlar, ebeveyn görüş ve gözlemlerini, çocukla görüşme raporlarını, çocukların doğrudan gözlemlerini ayrıntılı olarak değerlendirmektedir (Sharma *et al.* 2018). Bir tanı aracının en iyi olması için sosyal işlevselliği ve ortamlardaki davranış farklılıklarını ayrıntılı bir şekilde ele alması gerekmektedir. Hazırlanan tanı araçları, çeşitli yaş gruplarındaki farklı belirtileri değerlendirmektedir. Otizm Teşhis Gözlem Programı (Autism Diagnosis Observation Schedule (ADOS)) (Lord *et al.* 1999), Yeni yürümeye başlayan ve küçük çocuklarda otizm tarama aracı (The Screening Tool for Autism in Toddlers and Young Children (STAT)) (Stone *et al.* 2000), İletişim ve Sembolik Davranış Ölçekleri (Communication and Symbolic Behavior Scales (CSBS)) (Wetherby *et al.* 2002), Otizm Tanı Görüşmesi (Autism Diagnostic Interview – Revised (ADI-R)) (Kim and Lord 2010), Sosyal İletişim Anketi (Social Communication Questionnaire (SCQ)) (Rutter *et al.* 2003) ve Tekrarlanan Davranış Ölçeği Revize (The Repetitive Behavior Scale-Revised (RBS-R)) (Lam and Aman 2007) gibi ölçekler bu araçlara örnek verilebilir. Yukarıda bahsedilen tanı araçlarının kullanımında kısıtlama ve zorluklar mevcuttur. Mesela, ADI-R ölçütü tanı koymada başarılı olsa da, kullanımında ciddi eğitim alınmasını gerektiren oldukça uzun bir araçtır. Karmaşık olmasından dolayı nihai raporun tamamlanması için ekstra araçlara ihtiyaç duyulmaktadır. Ayrıca, son zamanlarda birçok dile tercümesi yapılmış OSB tanı araçlarının geçerliliğinin onaylanması problemini içeren araştırmalar ortaya çıkmıştır. Fakat araçların tercüme versiyonunun kültürel karşılığı henüz belirlenememiştir (Huerta and Lord 2012). Yaygın olarak kullanılan bir diğer tanı ölçeği ise ABD’de yayınlanan Diagnostic and Statistical Manual of Mental Disorders (DSM–5)’de bildirilen kriterlerdir.

Otizm Spektrum Bozukluğunun kesin tanısı konulurken doğru araç ve ölçeklerin kullanılması önemli bir konudur. Geçmişte tanı konulurken özel günlerde kaydedilmiş ev video kayıtları ve ebeveynlerin geçmişe yönelik izlenimleri kullanılmaktaydı (Barbaro and Dissamayake 2009). Bu özel günlerde kaydedilmiş ev video kayıtlarını ilk kullanan araştırmacı Adrien *et al.* (1991)’dir. Çalışmalarında iletişimsel ve davranışsal farklılıkları değerlendirmektedir (Adrien *et al.* 1992,1993). Geçmişe yönelik özel günlerde kaydedilen ev videolarının kullanılması iyi bir fikir olmasına rağmen sınırlıdır. İlk olarak, çocukların sergilediği davranışlar herhangi bir ortam ve zamanda değil de, doğum günü gibi özel hazırlanmış ortamlar ve planlı zamanlarda kaydedildikleri için doğal değildir. Ayrıca, videolarda birinin gülümsemesine karşı verilen tepki gibi ince davranışları yakalamak neredeyse imkânsızdır (Baranek 1999). Bu videoların değerlendirilmesi ile birlikte ebeveynlerin geçmişe yönelik izlenimlerini öğrenerek tanı koymak da yaygın bir yöntemdir. Vostanis *et al.* (1998) ve Yaung *et al.* (2003) ebeveynlere yönelttikleri sorular ile erken dönem OSB gelişimi durumunda çocukların sosyal iletişim eksiklikleri ve problemli davranışları

hakkında bilgi alınan çalışmalar yapmışlardır. Ebeveynlerin yanlış hatırlama, çocuklarına karşı duygusal hassaslığa bağlı objektif bilgi vermeme ve olayları çarpıtarak aktarma tutumları alınan bilgilerin güvenilirliğini azaltmıştır (Zwaigenbaum *et al.* 2005).

Son yıllarda teknolojinin de ilerlemesi ile birlikte OSB'li bireylerin beyin sinyalleri (EEG, MEG) ve beyin görüntüleri (fMRI ve CT) araştırılmaya başlanmıştır. Normal ve otistik bireylerde beyin sinyallerinde ve görüntülerinde farklılık olduğunun saptandığı çalışmalar olsa bile, çalışmaların yeni ve deneysel olması otizmde işlevsel beyin görüntüleme çalışmalarının kesin tanı için kullanımını engellemektedir. Araştırmalarda, deney grubundaki otistik bireylerin istenen görevi yerine getiremediği için deney dışında bırakılması, örneklerin azalmasına sebep olmaktadır. Ayrıca, inceleme yapılacak beyin görüntülerinin genellikle ergen ya da erişkin olan otistik bireylere ait olması ve sağlıklı bireylerle yapılan deneylerin örneklem çeşitliliğinin az olması sonuçların güvenilirliğini azaltmaktadır. Normal gelişen bir bireyin beyninde gerçekleşebilecek olağan değişimlerin tam olarak bilinmemesi sonuçların yorumlanmasını zorlaştırmaktadır (Ulay ve Ertuğrul 2009). Ayrıca bazı beyin sinyallerini algılayan cihazların dokunularak yerleştirilmesi yakın temastan hoşlanmayan otistik bireyler için rahatsız ve tedirgin edici olabilmektedir.

Diğer sağlık problemlerinde olduğu gibi OSB'nin erken tanısı oldukça önemlidir. Erken tanı, çocukluğun erken döneminde gelişmekte olan OSB'nin bütün belirtilerinin ortaya çıkmasını önleyip, sosyal etkileşim ve davranış bozukluklarını en aza indirebilecek erken müdahale için önemli bir fırsattır (Barbaro and Dissamayake 2009). Son zamanlarda, bebeklik döneminde tanı konulan çocuklara uygulanan özel eğitim ve terapilerin olumlu sonuçları araştırmacıları erken tanı üzerine araştırmalar yapmaya yönlendirmiştir (Matson *et al.* 2008). Araştırmacıların çoğu, tedavi için uygulanan özel eğitim ve terapilerin en erken zamanda başlaması gerektiğini savunmaktadır (Landa 2007; Reichow 2012). Erken tanı konusundaki araştırma ve fikirlere rağmen hala tanıların çoğu ortalama 4 veya 5 yaşlarında konulmaktadır (Camarata 2014). Otizm spektrum bozukluğu olan çocukların nöronları arasındaki bağlar az sayıda ve sağlıklı olduğu için nörobiyolojik gelişiminin özel eğitim ve terapilerle desteklenmesi erken çocukluk döneminde olmalıdır. Yaşamın ilk 5 yılı erken çocukluk dönemi olarak adlandırılmaktadır. Bu dönem, çocuklarda beyin gelişiminin en hızlı olduğu dönem olduğu için çocuğun öğrenme yetisi her türlü uyarana karşı duyarlıdır. Araştırmalara göre tam olarak iyileşme mümkün olmasa bile, erken tanı ile başlayan erken müdahale OSB belirtilerinde azalma sağlayabilmektedir (Anonymous 2004).

Objektif ve nicel sonuçlar elde ederek tanının en kısa zamanda doğru bir şekilde konulması için son yıllarda makine öğrenimi ve bilgisayarlı görme yöntemleri kullanılmaya

başlanmıştır. Makine öğrenmesi temelde, geçmiş OSB vakalarının ölçek değerlendirmelerini göz önüne alarak yeni vakaların sınıflandırma işlemini yapmaktadır (Wall *et al.* 2012). Bilgisayarlı görme yöntemleri ile uzun yıllardan beri tanı için kullanılan ev videoları ve görüntüler üzerinden duygu analizi, göz takibi ve tekrarlı hareket tanınması gibi çalışmalar yapılmıştır. Analizlerde, ebeveynler tarafından özel günlerde, özenle hazırlanmış ortamlarda ve planlanan bir zamanda kaydettikleri videolar yerine, herhangi bir zamanda çocukların doğal davranışlar sergiledikleri ev ortamında kaydedilen videolar kullanılmaktadır. Bu, değerlendirmelerin güvenilirliğini artırmaktadır. Davranışların analizinde bilgisayarlı görme yöntemlerinin kullanılması değerlendirme sürecini kısaltıp erken tanıyı mümkün kılmaktadır. Ayrıca, objektif ve nicel sonuçların elde edilmesini sağlamaktadır. Bu tez kapsamında OSB şüphesi olan çocukların belirgin hareketlerini tanımak amacı ile ev ortamında kaydedilen video görüntülerini güncel derin öğrenme algoritmaları ve görsel kelime çantası yaklaşımı ile incelenmiştir.

Motivasyon

Yaygınlığı gün geçtikçe artmakta olan otizm spektrum bozukluğunun tanı değerlendirme aşamaları hala uzun ve karmaşık olduğundan dolayı yeni tanı ölçeklerine ve araçlarına ihtiyaç duyulmaktadır. Erken tanı ile gelen erken müdahalenin otistik bireyler üzerindeki olumlu etkilerine rağmen hala erken ve kesin tanı konulmasında zorluklar mevcuttur. Örneğin; hekimlerin doğrudan videolar üzerinden hareketlerin incelenmesi ve analizi uzun zaman gerektirir. Aynı zamanda analizler sadece hekimin gözlemlerine ve tecrübesine bağlıdır. Tanı sürecindeki davranış değerlendirmelerine yardımcı olmak ve daha kısa sürede değerlendirme sonuçları alabilmek için makine öğrenmesi ve bilgisayarlı görme yöntemleri ile otizm şüphesi olan çocuklarda video tabanlı davranış tanıma gerçekleştirmek temel motivasyondur. Bu motivasyon ile tez çalışmasında OSB şüphesi olan çocukların ***ev ortamında çekilmiş kontrolsüz video görüntülerinden*** kendini uyarıcı hareketleri incelenmiştir.

Ev ortamında, kontrolsüz ve sınırlamalar olmadan, görüntü kalitesi ve çözünürlüğü belli olmayan bir kamera ile kaydedilen videolar üzerinde hareket tanıma işlemi, kamera hareketi, nesne görünümü ve pozunu, nesne ölçeği, kamera bakış açısı, dağınık arka plan, aydınlatma gibi koşulları göz önüne alındığında bilgisayarlı görme alanında birçok zorluk içermektedir. Bu zorluklara ek olarak, tez çalışmamızın temel motivasyonu olan, otizmli çocuklar tarafından normal hareketlerden farklı olarak aniden yapılmaya başlanan, hızlı değişen ve tekrarlanan hareketlere bağlı olarak aşağıda tanımlanmış zorluklar bulunmaktadır.

- 1. Güçl kle algılanan davranışlar:** Çocuklar t m video s resi boyunca hemen g ze  arpacak şekilde davranış sergilememektedirler. Davranışlar atak anlarında tam belirgin olarak g r lmektedir. Bu nedenle, se ilen tanıma algoritmasının belli belirsiz davranışları kesin olarak tespit edip davranış sınıflandırmasında kullanması gerekmektedir.
- 2. Yo unluk ve S rekli lik:** Bu  zel davranışlar bir seferde s rekli yapılmamakta ve bunun yerine de i en yo unluk seviyelerinde bir ok kez ger ekle tirilmektedir. Bu analizi zorla tırmaktadır. Belirli bir eylemin baskın, s rekli ve benzer yo unlukta oldu u standart eylem tanıma veri k melerinde g zlemlenen eylemlerin aksine bu t r yo unlu u ve s reklili i belli olmayan davranışları tanımak i in zaman serisi ge mi inin   renilmesi gerekmektedir.
- 3. Mek n De i imi:** Çocuk davranış belli bir me anda sergilemeye ba lamakta ve davranış devam ederken me an de i imi yapabilmektedir. De i ime ba lı olarak g r nt  karesine giren bir ok yeni nesne ya da insan olabilmektedir. Buna ba lı olarak, ilgili olmayan nesnelerin hareketi analiz yapılırken  n plana  ıkmaya ba lamaktadır.
- 4. Karma Davranışlar:**  ocu un belli bir davranış sergilerken aniden ba ka bir davranışa ge mesi normal bir durumdur. Hatta  ocuk d nme hareketi yaparken aynı anda kol  ırpma hareketini de yapabilmektedir. Bu t r karma davranışların analiz yaparken dikkate alınması gerekmektedir.
- 5. İli kili Nesne Etkisi:**  ocu un bu davranışları yaptı ı sırada, etkile imli oldu u nesneler de aynı davranış izleyebilir ve  ocuk durmasına ra men nesne davranış devam ettirebilir. Mesela,  ocuk d nebilen sandalyeye oturup d nme hareketini yaptıktan sonra d nmeyi bırakmasına ra men sandalye d nmeye devam edebilmektedir.  ocu un etkile imde oldu u nesneyi analiz sırasında iyice ayrıştırmak gerekmektedir.
- 6. Ebeveyn Kaygısı:** Bu davranışlar  ocu un atak anında  iddetini artırarak yapılmaya ba landığında g r nt y  kaydeden ebeveyni endi elendirebilmektedir. Buna ba lı olarak, video kalitesini d   rerek analizi zorla tırıcı kamera hareketleri olabilmektedir (Rajagopalan *et al.* 2013).

Tezin Amacı ve Katkısı

Bu tezin amacı OSB tanı s recini azaltarak, erken ve do ru tanı de erlendirmesine yardımcı olmak i in OSB   phesi olan  ocukların ev ortamlarında  ekilmi  video g r nt lerinden kendini uyarıcı hareketlerin tanınması ve incelenmesidir. Bu ama  do rultusunda tezin katkıları a a ıda sunulmu tur.

- Otizm spektrum bozukluğunun tanı ölçütlerinden biri olan kendini uyarıcı hareketleri tanıma işlemi için klinik ortamda oluşturulmuş herkese açık video verisi bulunmamaktadır. Ancak teknolojinin ilerlemesi ile ebeveynler günlük hayatta cep telefonlarından videolar çekerek çocukların davranışlarını kaydedip çeşitli sosyal medya platformlarında paylaşmaktadırlar (Zunino 2018). Bu çalışmada, herkese açık sosyal medya kanallarından toplanılan video görüntüleri kullanılarak, dönme, kol çırpma ve kafa sallama ve el-parmak hareketi videoları ile dört farklı davranışa ait ***Geniştirilmiş Kendini Uyarıcı Davranış Veri seti (GKUDV)*** adında yeni bir veri seti oluşturulmuştur.
- Kendini uyarıcı hareketlerin tanımlanması için çalışmamızda derin öğrenme yöntemlerinden Evrişimsel Uzun-Kısa Dönem Hafıza (ConvLSTM) ve 3 Boyutlu Evrişimsel Sinir Ağı (3DCNN) ve ayrıca Görsel Kelime Çantası (Bag-of-Visual-Word (BoVW)) yaklaşımı kullanılarak başarımlar analiz edilmiştir. Görsel Kelime Çantası yaklaşımı ile hareket tanımda yaygın olarak kullanılan (Histogram of Oriented Gradient), HOF (Histogram of Optical Flow), HOG-HOF, SIFT (Scale Invariant Feature Transform) ve SURF (Speed Up Robust Features) olmak üzere 5 (beş) öznitelik tanımlayıcısı kullanılmıştır. Sınıflandırmak için MLP (Multi Layer Perceptron), SVM (Support Vector Machine) ve GNB (Gaussian Naive Bayes) olmak üzere 3 (üç) sınıflandırıcı ile hareketler sınıflandırılmıştır. Ayrıca farklı anahtar görsel kelime sayıları ile deneyler yapılmıştır.
- Veri setimizde ki videolar ev ortamlarında ebeveynler tarafından çözünürlüğü belli olmayan bir kamera ile kaydedilmiş olmasından dolayı arka planda alakasız kişi ve nesneler olabilmektedir. Görüntülerden ilgili alan (region of interest (ROI)) bulunup, arka plandan ayrıştırılmıştır. Görüntülerde ki ilgi alanımız, asıl hareketi yapan çocuktur. Görüntülerde bu çocuğun tespiti ve takibi yapıp arka plandan ayrıştırılmıştır. Çocuğun tespiti yapılırken üç popüler yöntem olan HOG, Haar-Cascade ve YOLOv3 uygulanmıştır. En hızlı ve en doğru tespiti sağlayan yöntemler analiz edilmiştir.
- Görsel Kelime Çantası (Bag-of-Visual-Words) yaklaşımlarında videonun zaman bilgileri dikkate alınmamaktadır. Uzun süre zaman bilgisinin değerlendirilmemesi tanıma işleminin performansını düşürebilmektedir. Bu sorunun etkisini en aza indirmek ve yaklaşımın genel performansını iyileştirmek için videodaki zaman bilgisini kısmen dâhil etmek adına piramit histogram yaklaşımı önerilmiştir (Shukla *et al.* 2013). Bu yaklaşıma benzer olarak, video iki parçaya bölünüp her iki parçasından da histogram üretilmektedir. Daha sonra, tüm videoyu temsil eden histogram ve videonun her bir

yarısından üretilen histogramlar ard arda eklenmektedir. Böylece, video üç histogramla temsil edilerek zaman bilgileri hareket tanıma sürecine kısmen dâhil edilmektedir.



KURAMSAL TEMELLER

Bu tez çalışması ile literatür dört ana başlıkta incelenmiştir. İlk olarak OSB ve OSB tanısı için kullanılan ölçüt ve yöntemler hakkında bilgi verilmiştir. İkinci başlıkta Otistik bireylerde bulunan davranış bozukluklarından bahsedilmiştir. Üçüncü başlıkta insan hareketi tanıma yaklaşımları ele alınmıştır. Dördüncü ve son başlıkta ise otizmlili bireylerin davranış bozukluklarının video görüntüleri üzerinden analizi ile ilgili çalışmalar anlatılmıştır.

Otizm Spektrum Bozukluğu ve Tanı Yöntemleri

Otizm spektrum bozukluğu

Otizm, sosyal etkileşim ve iletişim eksiklikleri ile beraber sınırlı ve tekrarlayıcı davranışlarla kendini gösteren nöro-gelişimsel bir bozukluktur (Anonymous 2013). Erken çocukluk döneminde ortaya çıkan ve yaşam boyu süren bu belirtiler çocukların günlük işlerini yerine getirirken sorunlar yaşamasına sebep olmaktadır.

Dünyada ve Türkiye’de yaygınlık oranının artmasıyla dikkat çekmektedir ve her toplumda görülmektedir. Amerika Birleşik Devletlerinde (ABD) ilk tanının yapıldığı zamanlarda on bin çocuktan birinde görülürken, Hastalıkları Kontrol Etme ve Önleme Merkezi (CDC)’nin 2012 yılında yaptığı çalışmalarda 88 çocuktan birinde, 2014 yılında 68 çocuktan birinde görüldüğü saptanmıştır. Bu yıl (2020) yayınladıkları rapora göre otizmin 54 çocuktan 1’inde görülmeye başlandığı bildirilmiştir (Anonymous 2020). Türkiye’de ise yaygınlığın belirlenmesi için yapılan kapsamlı bir araştırma yoktur. Önceden yapılan araştırmalara göre tahmini olarak ülkemizde 500 bin civarında otizmlili çocuk bulunmaktadır (Anonymous 2012). Otizm kız çocuklarına göre erkek çocuklarda daha çok rastlanmaktadır (Loomes *et al.* 2017). Ancak otizmin sebepleri üzerinde çalışmalar hala devam etmekte ve şu ana kadar kesin bir neden bulunamamıştır (Anonymous 2015)

Otizm spektrum bozukluğu (OSB), uzun yıllar boyu zihinsel engel ve/veya ruhsal-duygusal bozukluk başlığı altında değerlendirilmiştir. Daha sonra ABD’de Ruhsal Bozuklukların Tanısal ve İstatistiksel El Kitabı (DSM-5) yayınlanarak tanı ölçütleri tanımlanmıştır (Anonymous 2013).

DSM ölçütleri ayrıntıları ile şunlardır;

- Sözel ve sözel olmayan iletişimdeki kısıtlılık,
- Akran ilişkisini başlatma veya sürdürmedeki zorluklar,

- Sosyal-duygusal kısıtlılık,
- Basmakalıp ve tekrarlayıcı motor hareketler,
- Rutinlere veya ritüellere sıkı bağlılık,
- Nesnelerin amacı dışında kullanımı ve bir bölümüne olan ilgi,
- Sınırlı ilgi alanları,
- Aşırı ya da olması gerekenden az duyuşal ilgi ya da tepkiler

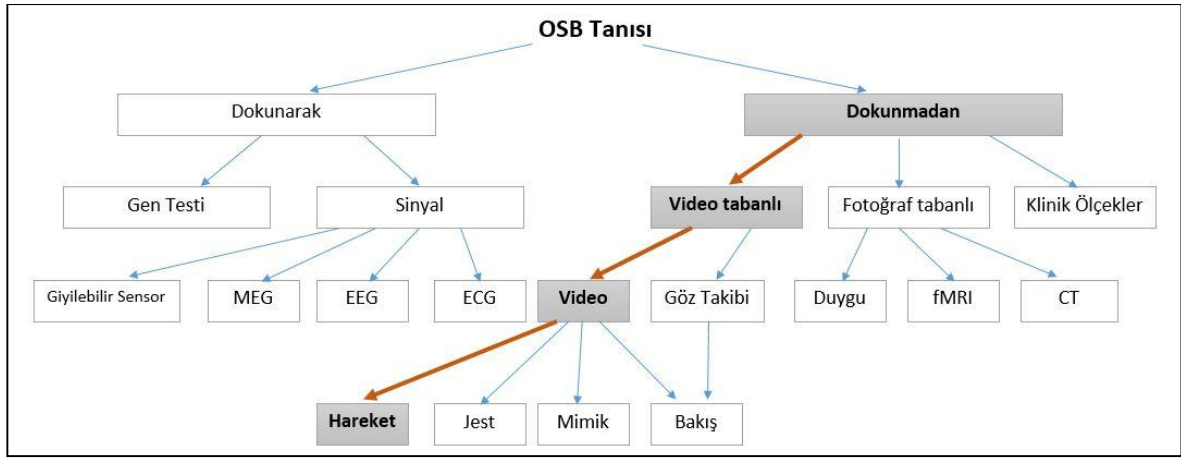
OSB'nin yaygın belirtileri temelde 3 başlıkta ifade edilebilir (Dur ve Mutlu 2018);

- **Sosyal etkileşimde bozukluk:** Yaşıtları ile ilişki kurma yürütmede zorlanırlar. Göz teması kurama, duygu ifadesi gösterememe, ismiyle çağrıldığında tepki vermeme görülmektedir.
- **Sözel ve sözel olmayan iletişimde yetersizlik ve konuşmada güçlük:** Çevresiyle bir konuşma başlatıp ve devam ettirmezler, duyuşal jest ve mimiklerde yetersiz olurlar.
- **Tekrarlı ve basmakalıp davranışlar:** Bu davranışların birçok türü bulunmaktadır. Bir nesnenin belli bir parçasına ve hareketine takılırlar, günlük rutinlerdeki değişiklikten hoşlanmazlar. Kollarını kuş gibi çırpma, sürekli etrafında dönme, kafa sallama, anlamsız parmak hareketleri gibi basmakalıp davranışları vardır.

Araştırmalara göre OSB gelişmesinde genetik faktörlerin etkili olabileceği düşünülse bile (O'Roak and State, 2008) tanı, genelde davranışsal ipuçları ile konulmaktadır. Pediatri Akademisi, 18 ve 24. ayda erken OSB taramasını önermesine rağmen, hala ortalama tanı yaşı için üzerindedir (Mandell *et al.* 2005). Bu kadar erken dönemlerde OSB tanı araştırmaları çok yeni ve bulgular kesinlik kazanmadığı için hekimler küçük çocuklar için tanı araçları ve ölçekleri uygulamakta önemli zorluklarla karşılaşmaktadır. Belirtileri her yaş grubu için ayrıntılı şekilde değerlendiren araçlara ihtiyaç duyulmaktadır (Guthrie *et al.* 2012).

Tanı yöntemleri

Literatürde yapmış olduğumuz ayrıntılı incelemeler sonucunda, OSB tanısında yardımcı olacak yöntemler sınıflandırılmış ve Şekil 1'de gösterilmiştir. Şekil 1'de görüldüğü gibi OSB tanısı yöntemin uygulanma şekline göre bireyler ile fiziksel teması gerektiren (EEG sinyallerinin alınması, gen testi vb.) yöntemler ve fiziksel temas gerektirmeyen video ve resim üzerinden uygulanan yöntemler olarak iki grupta incelenmiştir.



Şekil 1. Otizm spektrum bozukluğu tanı yöntemleri.

Otizm spektrum bozukluğu olan çocukların fiziksel temastan tedirgin ve huzursuz oldukları bilinmektedir. Bu bağlamda değerlendirme aşamasının fiziksel temas gerektirmeyen yöntemler ile olması daha uygundur. Otizmin genetik nedenleri olduğu ile ilgili bilimsel araştırmalar vardır. Fakat hala hangi genin ya da genlerin sebep olduğu kesin olarak bilinmemektedir. Bu sebeple Türkiye’de hala tanı konurken gen testi uygulanmamaktadır (Çöp *et al.* 2015).

Son zamanlarda normal gelişen çocuk ile otistik çocukların beyin sinyallerindeki gecikmeler ve anormallikler tanıda kullanılabileceği fikrini oluşturmuştur. Araştırmalarda otizmlili çocuklarda Elektroensefalografi (EEG) anomalilerinin oldukça sık olduğu gözlenmiştir. Fakat henüz EEG rutin tanı parametresi olarak kabul edilmemektedir (Yavral vd. 2015; Bosl *et al.* 2018). EEG, electrocardiogram (ECG)(Kushki *et al.* 2014), magnetoencephalography (MEG) (Roberts *et al.* 2010) gibi yöntemler sinyal verilerinin elde edilme aşamasında çocuğun vücuduna cihaz yerleştirmeyi gerektirmektedir.

Otizm spektrum bozukluğunda görülen davranışları tanımak için giyilebilir sensörler kullanılmaktadır. Westeyn *et al.* (2015), otizmdeki yedi kendini uyarıcı davranışın ve normal günlük davranışların simüle edilmesinden üretilen hızlanma verilerini içeren bir veri seti oluşturmuşlardır. Stereotip davranışlar; bir yüzeye arka arkaya vurmak, kol çıpmak, kapalı bir alanda amaçsız yürümek, ileri-geri sallanmak, dönmek, ayak parmak uçlarında yürümek ve el hareketleridir. Normal günlük davranışlar; oturmak, ayakta durmak, yazmak, konuşmak ve zıplamak gibi otistik davranışların dışında bulunan davranışlardır. Çalışmada, bireyin sağ bileğine, belinin arkasına ve sol bileğine üç kablosuz Bluetooth ivmeölçer sensörü yerleştirilmiştir. Her iki hareket grubu için de 15 denemede 105 stereotip davranış örneği ve 105 normal davranış örneği toplanmıştır. Hareket tanıma için Gizli Markov Model (HMM) önerilmiş ve %68.57 genel sistem doğruluğu elde edilmiştir. Bu çalışma gibi giyilebilir sensörler kullanılarak otistik davranışların tanındığı birçok çalışma mevcuttur (Rad *et al.* 2018; Gilchrist *et al.* 2018).

Fiziksel teması gerektiren yöntemler dışında fiziksel temasa gerek duyulmadan tanı

değerlendirmesine yardımcı olan video tabanlı, fotoğraf tabanlı ve ebeveynlerin görüşlerinin kâğıt üzerinde doldurulduğu klinik ölçekler vardır. Video ve fotoğraflardan davranış, duygu, jest, mimik, göz takibi gibi veriler incelenmektedir.

OSB tanısı kadar, tedavi yöntemi olan psikoeğitim terapi aşaması da önemlidir. İnsan-robot etkileşim teknolojisindeki son gelişmeler, insansı robotları kullanarak otizmlili insanlara duygu tanıma ve ifade öğretme girişimlerine yol açmıştır. Bu bağlamda, Avrupa Birliği tarafından finanse edilen proje koordinatörünün Hollanda olduğu ve 7 ortak ülke (İngiltere, Almanya, Romanya, Birleşik Krallık, Sırbistan, Belçika ve Portekiz.) ile yürütülen DE-ENIGMA projesi, Robokind şirketi tarafından geliştirilen Zeno'yu otistik çocukları duyguları öğrenmede desteklemek ve öngörülebilirlik tercihlerinden faydalanmak için kullanmaktadır. Teknoloji, çocuğun seslendirmelerini, davranışlarını ve yüz ifadelerini algılamak ve yorumlamak için derin öğrenme algoritmaları, çocuğun eylemlerine ve öğrenme aşamasına uyum sağlayan etkileşimli robot davranışları ve çocukların öğrenme adımlarında ilerlemelerine yardımcı olacak bir oyun motoru içermektedir (Anonymous 2018).

Yukarıda örnekleri verildiği gibi otistik bireylerin sergilediği çeşitli davranış bozuklukları vardır. Bu davranışlar genel olarak sınırlı ve tekrarlı davranışlar olarak adlandırılmaktadır.

Otizmlili Hastalarda Bulunan Davranış Bozuklukları

Otizm spektrum bozukluğu (OSB) tanısı almış bireyler sözel ve sözel olmayan iletişimde yetersizlik, bir takım sosyal sınırlılıklar ile sınırlı ve tekrarlayan davranışlar gibi belli başlı bir takım özelliklere sahiptirler (Anonymous 2013). OSB, Amerikalı psikiyatrist Leo Kanner tarafından tanımlandığından beri tekrarlanan diğer bir adıyla stereotip davranışlar bu bozukluğun değişmeyen temel tanı ölçütlerinden biri olmuştur.

Aman and Lam (2007) klinik araştırmalarına dayanarak, sınırlı tekrarlayıcı davranışları kavramsal olarak alt ölçeklere ayırmışlardır. Bunlar:

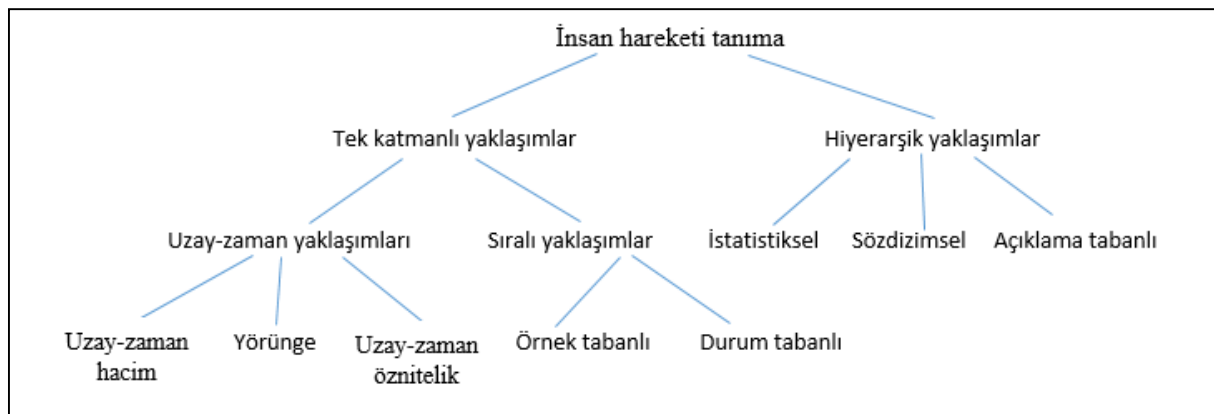
- **Belirgin bir amacı olmayan ve sürekli tekrarlanan davranışlar:** Basmakalıp davranışlardır. Kendini uyarıcı stereotip davranışları da kapsamaktadır. Bu davranışlara örnek olarak, kol hareketleri, kafa hareketleri, el-parmak hareketleri, kendi etrafında ya da bir daire etrafında dönme hareketi, nesnelerin kullanım hareketleri, duyuşal hareketler gösterilebilir.
- **Kızarıklık, morarma veya vücutta başka yaralanmalara sebep olabilecek davranışlar:** bir yüzeye çarpma, nesne ile vurma, kendini ısırma, saçlarını çekme, cildini çizmek ve koparmak gibi hareketlerdir.

- **Bir kurala göre tekrarlanan davranışlar:** Sıralama, yıkama, sayma, istifleme, tekrarlama, koklama gibi davranışlardır.
- **Günlük yaşam aktivitelerini her gün aynı şekilde gerçekleştirme davranışları:** Belli yemekler ve aynı yemek saatleri, aynı saatlerde uyuyup uyanma, kişisel bakım rutini gibi ritüel davranışlardır.
- **Değişime direnç gösterilen veya herşeyin aynı kalması için ısrar edilen davranışlar:** Nesnelerin yerlerini sabit tutması, belli şekilde yürümesi, belli yere oturması gibi davranışlardır.
- **Kısıtlı odak, ilgi alanı ve etkinlik davranışları:** Bir nesneye olan bağlılık, nesnenin sadece bir kısmıyla meşguliyet, nesnelerin dönme hareketiyle meşgul olma gibi davranışlardır.

Otizm spektrum bozukluklarında görülen bu davranışların ortak noktası tekrarı, değişmezliği ve tuhaflığıdır. Bu davranışlar çocuğun günlük hayatını zorlaştırdığı gibi sosyal çevresi ile iletişim ve etkileşim içerisine girmesine engel olmaktadır. Ayrıca ebeveynlerin endişelenmelerine sebep olan davranışlardır. Bu tez çalışmasında otistik bireylerde görülen kendini uyarıcı hareketler ev ortamında kaydedilmiş videolar üzerinden incelenmiştir.

Görüntü Tabanlı İnsan Hareketi Tanıma Yaklaşımları

İnsan hareketi tanıma başta güvenlik olmak üzere çeşitli alanlarda kullanılmaktadır. Güvenlik sistemleri, halka açık yerlerde, normal davranışların dışındaki anormal ve şüpheli davranışları tespit etmektedir. İnsan hareketleri, hareketlerin karmaşıklığına bağlı olarak dört gruptur. Bunlar; jestler (bacağın kaldırılması vb.), hareketler (yürüme vb.), bir nesne veya insanla etkileşimde bulunulan hareketler (kavga vb.) ve grup etkinlikleridir (grup toplantısı vb.). Şekil 2’de Aggarwal and Ryoo (2011) tarafından öngörülen görüntü tabanlı hareket tanıma yaklaşımlarının sınıflandırma şeması verilmiştir.



Şekil 2. Görüntü tabanlı insan hareketi tanıma yaklaşımlarının sınıflandırması (Aggarwal and Ryoo 2011).

Tek katmanlı yaklaşımlar

Bir hareketi bir video karesi dizisi olarak temsil ederek yeni bir videoyu sınıflandırmaktadır. Bu yaklaşımı kullanan yöntemlerin çoğu, video üzerinden otomatik hareket tanıma için tüm alt sınıf olasılıklarını sınıflandıran kayan pencere tekniğini kullanmaktadır. Tek katmanlı yaklaşımlar, yürüme ve koşma gibi basit hareketleri tanımak için daha uygundur. Şekil 2'ye göre tek katmanlı yaklaşımlar uzay-zaman yaklaşımları ve sıralı yaklaşımlar olmak üzere iki sınıfa ayrılmaktadır.

Uzay-zaman yaklaşımları: Uzay-zaman hacim, yörünge (trajectories) tabanlı ve yerel özniteliklere dayalı türleri vardır (Aggarwal and Ryoo 2011).

- **Uzay-zaman hacim yaklaşımları,** hareket videolarının uzay-zaman hacimlerini değerlendirerek hareket tanıma yapan yaklaşımlardır. Bir video karesi, gerçek hayattaki 3 boyutlu sahnelerin gösterimi olan 2 boyutlu verilerdir. Bu veriler insanların veya nesnelerin şekilleri ve görünüşleri gibi yerel özniteliklerini barındırmaktadır. Bir hareketin gerçekleştirilmesini gösteren bir video, video karelerinin zamansal (T) bilgilerinin kronolojik sırasına göre birleştirilmesiyle oluşturulan 3 boyutlu uzay-zaman hacmi olarak temsil edilebilmektedir. Bu yaklaşımla insan hareketi tanıma aşamaları şunlardır; eğitim için kullanılacak video verilerinin 3 boyutlu XYT uzay-zaman hacmi modelleri oluşturulur. Etiketlenmemiş bir video sağlandığında, sistem yeni videoya karşılık gelen 3 boyutlu bir uzay-zaman hacim modeli oluşturulur. Sınıflandırma için iki model arasındaki görünüm benzerliği değerlendirilir. Uzay-zaman hacmi yaklaşımları görüntüde birden fazla kişi olduğunda hareket tanımda zorlanmaktadır. Bu zorluğun üstesinden gelmek için yaklaşımların çoğu klasik kayan pencere algoritmasını kullanmaktadır. Bununla birlikte, bu, hareketlerin doğru konumu için birçok hesaplama yapmayı gerektirir (Aggarwal and Ryoo 2011).
- **Yörünge tabanlı yaklaşımlar,** bir insanın eklem pozisyonlarını 2 boyutlu (XY) veya 3 boyutlu (XYZ) noktalardan oluşan bir set olarak temsil edildiği yaklaşımlardır. Çubuk modellemesi insanların vücut parçalarının eklem pozisyonlarını çıkarma işleminde yaygın olarak kullanılmaktadır. Bir hareket yapıldığında, eklem pozisyonundaki değişimler, 3 (XYZ) ve 4 (XYZT) boyutlu temsillerini oluşturan uzay zaman yörüngeleri olarak kaydedilmektedir. Bu yaklaşım eklemleri iyi temsil ettiği için hareketin detaylarını analiz etmekte iyidirler ve görüş açısından bağımsızdır. Yöntemin iyi sonuç vermesi için videoda görünen kişilerin 3 boyutlu (XYZ) eklem yerlerinin doğru tahmin edilmesi gerekmektedir. 3 boyutlu vücut

parçalarının tespiti ve takibi hala üzerinde çalışılan bir problemdir (Aggarwal and Ryoo 2011).

- **Yerel özniteliklere dayalı yaklaşımlar**, 3 boyutlu uzay-zaman hacimlerinden yerel öznitelikleri çıkarmaktadır. Bu öznitelikleri hareket temsili ve tanınması için kullanmaktadır. Bu yaklaşımlarda arkaplan çıkarımına gerek yoktur. Ayrıca ölçek, döndürme ve ötelemeden bağımsızdır. Diğer tek katmanlı yaklaşımlar gibi basit hareketler için uygun olup, karmaşık hareketleri tanımak için yetersizdir (Aggarwal and Ryoo 2011).

Sıralı yaklaşımlar: Uzay-zaman yaklaşımlarından farklı olarak bir insan hareketini belirli gözlemlerin bir sırası olarak ele almaktadır. Yani, bir insan hareketini görüntülerden çıkarılan bir dizi öznitelik vektörü olarak temsil etmekte ve böyle bir sekans arayarak hareket türlerini tanımaktadırlar. Sıralı yaklaşımlar iki kategoriye ayrılır: örnek tabanlı tanıma yaklaşımları ve durum modeline dayalı tanıma yaklaşımları (Aggarwal and Ryoo 2011).

- **Örnek tabanlı yaklaşımlar**, direk eğitim verilerinin öznitelik vektörlerinin bir örneğini kullanarak hareketleri temsil eder. Her hareket sınıfı için bir dizi öznitelik vektör örneği tutarlar ve yeni video girişinin öznitelik vektörleri ile bu örneği karşılaştırırlar. Benzerliğin oranına göre yeni videoda hareketin gerçekleştirip gerçekleştirmediği sonucuna varırlar. İnsanlar aynı hareketi farklı şekillerde ve hızlarda yapabildikleri için benzerlik ölçümü bu değişkenler dikkate alınarak yapılmalıdır. Farklı boyutlardaki öznitelik vektörlerinin eşleştirilmesi için çoğunlukla dinamik zaman atlama (DTW) algoritması kullanılmaktadır (Aggarwal and Ryoo 2011).
- **Durum modeline dayalı yaklaşımlar**, bir grup durumdan oluşan bir model olarak bir insan hareketini temsil etmektedir. Videodaki bir kişinin her bir zaman aralığında başka bir durumda olduğu varsayılır. Bu durumların herbiri için bir özellik vektörü üretilir. Genel olarak, her hareket için bir model oluşturulur. Modeller için durumlar arası geçiş ve öznitelik vektör olasılıkları eğitildikten sonra hareketler uygun bir değerlendirme yöntemi seçilerek tanınır. Değerlendirme işlemi, belirli bir durum modeli tarafından üretilen belirli bir dizinin (girdi) olasılığını hesaplama işlemidir. Hesaplanan olasılık yüksekse, modele karşılık gelen hareketin verilen girdi dizisinde olduğu sonucuna varılır. Gizli Markov modelleri (HMMs) ve dinamik Bayes ağları (DBNs), durum modeli tabanlı yaklaşımlar için yaygın olarak kullanılmaktadır (Aggarwal and Ryoo 2011).

Hiyerarşik yaklaşımlar

Karmaşık insan hareketlerinin daha basit alt hareketlerin tanınmasına dayanarak tanınmasını sağlayan yaklaşımlardır. Tanıma sistemlerini birden fazla katmandan oluşacak şekilde tasarlayıp karmaşık hareketlerin tanınması için uygun hale getirmektedir. Grup aktivitelerinin tanınmasında kullanışlıdır. Mesela, dövüş eylemi “tekme” ve “yumruk” gibi kolay tanınabilecek alt eylemlere ayrıştırılarak temsil edilebilmektedir. Temel olarak, karmaşık insan hareketlerinde sık görülen hareketler basit hareketler olarak modellenir ve karmaşık hareketler hiyerarşik olarak birleştirilerek temsil edilip tanınır. Bu basit hareketler, tek katmanlı yöntemler uygulanarak tanınırlar. Hiyerarşik yaklaşımlar üç alt gruba ayrılır: istatistiksel, sözdizimsel ve açıklama tabanlı.

İstatistiksel yaklaşımlar, hareketleri tanımak için istatistiksel durum tabanlı modelleri kullanmaktadır. Tanınması amaçlanan üst düzey hareketleri oluşturan basit hareketler tek katmanlı yaklaşımlar ile tanınmaktadır. Tanıma işlemi tek katmanlı sıralı yaklaşımlarda olduğu gibi öznitelik vektörlerinin sekansları ile yapılır. Sonuç olarak, bir dizi özellik vektörü bir dizi basit harekete dönüştürülür. Üst düzey hareketler, bu basit hareket dizisini üst düzey modeller tarafından üretilen gözlemler olarak ele alır. Her model için, modelin, bir dizi basit hareket içermesi olasılığı, hareket ile giriş görüntü dizisi arasındaki olasılığı ölçmek için hesaplanır (Aggarwal and Ryoo 2011).

Sözdizimsel yaklaşım, insan hareketlerini bir sembol dizisi olarak modellemektedir. Her bir sembol üst düzey hareketlerin içerdiği bir basit harekete karşılık gelmektedir. Bu yaklaşım da hiyerarşik istatistiksel yaklaşımlar gibi, üst düzey hareketleri tanımadaki ilk aşama olan basit hareketleri tanımak için tek katmanlı yaklaşımları kullanmaktadır. Sözdizimsel yaklaşımlar, tüm özniteliklerin kendi üretim kurallarını uygulayarak ayrıştırıldığını varsayar. Üst düzey bir hareketi, onları oluşturan bir dizi basit hareketler olarak modellediğinden, basit hareketlerin ard arda olması bu yaklaşımlar için önemlidir. Yaklaşımın sınırlamalarından biri, eşzamanlı hareketlerin tanınmasıdır (Aggarwal and Ryoo 2011).

Açıklamaya dayalı yaklaşım, insan hareketlerinin mekânsal-zamansal özniteliklerini koruyan bir hareket tanıma yaklaşımıdır. Zamansal, mekânsal ve mantıksal ilişkilerini tanımlayan basit hareketler üst düzey bir insan hareketini temsil eder. Bu yaklaşım, bir insan hareketini, belirli ilişkisi sağlayan kendi alt basit hareketlerinin bir nüshası olarak modellemektedir. Açıklamaya dayalı yaklaşımlar üst düzey hareketleri tanımlamak için hareketin içerdiği basit hareketleri kullandığı için hiyerarşıktır (Aggarwal and Ryoo 2011).

Davranış Bozukluklarının Video Görüntüleri Üzerinden Analizi

Otizm spektrum bozukluğu tanısı hala karmaşık bir konudur. Tanının tek bir ölçüt veya araç üzerinden konulması uygun değildir. Davranış bozukluklarını saptamanın temel yöntemleri kâğıt üzerinde derecelendirme ölçekleri, doğrudan davranışsal gözlemler ve video tabanlı incelemelerdir (Rad and Furlanetto 2016). Derecelendirme açısından, bireylerin içsel değerlendirme ölçekleri farklıdır ve son raporun şüpheli bir doğruluğa sahip olma ihtimali yüksektir. Ayrıca, zamanla davranış bozukluğu diye adlandırılan hareketlerin şekli, miktarı, yapılma sıklığı ve süresi bireysel değişkenlikleri ölçmede başarısız olabilir (Pyles *et al.* 1997). Çocukların davranışlarının doğrudan izlenmesi, birçok tekrarlı davranışın aynı anda hem gözlenip hem de kaydedilmesini gerektirir. Tekrarlayıcı hareketler hızlı yapıldığı için tam olarak izleme ve kaydetme doğruluğunun az olması, hareketin başlama ve bitiş zamanlarını belirlenememesi, gözlemleyen kişinin hafıza hataları, bir zaman diliminde ne kadar sıklıkla yapılacağını tahmin edememe, eş zamanlı birden fazla tekrarlayıcı hareketin gözlemlenme güçlüğü ve çevresel etkenleri kaydetme gibi faktörler bu yöntemi güvenilir yapabilmektedir (Sprague 1996).

Video tabanlı analizler, hareketler video kaydına alındığı için tekrar tekrar izlenebilme ve arzu edilen yerde oynatma hızının düşürülebilmesi seçeneğinden dolayı diğer yöntemlere göre daha doğru ve güvenilirdir. Matson *et al.* (2007)'ye göre videolardan hareket analizi yöntemi sıkıcı ve uzun zaman gerektirdiğinden hekimler için pratik değildir. Bu problemi çözmek için makine öğrenmesi ve bilgisayarlı görme yöntemleri ile video tabanlı analizler yapılarak zaman kaybı en aza indirilmektedir (Goodwin *et al.* 2011).

Otizmin en belirgin hareket türü olan basmakalıp ve tekrarlayıcı hareketlerden olan kendini uyarıcı hareketleri bilgisayarlı görme yöntemleri ile video üzerinden tanıma çalışmaları yapılmıştır. Tanıma yapılırken klasik yöntemler veya derin öğrenme yöntemleri kullanılmıştır.

Klasik yöntemler

Araştırmalarımıza göre otizmlili hastalarda kendini uyarıcı davranış tanıma işlemi için klinik ortamda oluşturulmuş herkese açık video verisi bulunamamıştır. Ancak teknolojinin ilerlemesi ile ebeveynler günlük hayatta cep telefonlarından videolar çekerek çocukların davranışlarını kaydedip çeşitli sosyal medya platformlarında paylaşmaktadırlar. Rajagopalan *et al.* (2013), ebeveynler / bakıcılar tarafından yayınlanan çocukların kendini uyarıcı davranış sergiledikleri bu videoları toplayarak Kendini Uyarıcı Davranış Veri Seti (SSBD) adında yeni bir halka açık veri seti sunmuşlardır. Veri seti, kol çırpma, kafa sallama ve dönme olmak üzere üç davranış sınıfından oluşmaktadır. Veri seti, ortalama 90 saniye süren 75 video içermektedir.

Bu videolar, gerçek zamanlı davranış analizi için çeşitli zorluklar içermektedir. Çünkü videolar ebeveynler/bakıcılar tarafından doğal ev ortamında hiçbir kontrol ve kısıtlama olmadan herhangi bir kamera ile kaydedilmiştir. Bu nedenle için otomatik davranış analizi için son derece zordur. Çalışmalarında sınıflandırıcıyı eğitmek için Kelime çantası (BOW) yaklaşımında Harris3D detektörleri ve öznitelik çıkarmak için Uzay Zamanı İlgi Noktaları (STIP) (Laptev 2005) kullanılmıştır. "Bir Grubu Dışarıda Bırak (Leave-One-Group-Out)" yönteminde k-kat doğrulama (k-fold validation) ile 500, 1000, 1500 kod çizelgesi (codebook) boyutlarında deneyler yapılmıştır. Her sınıftan rastgele 5 video test için ayrılırken, diğer kalan videolar eğitim verisi olarak kullanılmıştır. Bu eğitim yöntemi 5 kez tekrarlanmıştır. Kod çizelgesi boyutu 500 iken %44, 1000 iken %50,7, 1500 iken %46,6 doğruluk sonuçları elde edilmiştir. Ortalama %47,1 doğruluk sonucu bildirilmiştir. Çocukların kendini uyarıcı davranışları sergilediği ev ortamlarındaki videolarından halka açık oluşturulan ilk veri seti bu çalışmada sunulmuştur. SSBD veri setindeki videolar yaklaşık 9 yıl önce sosyal medya platformlarına yüklenen videolardan oluşmaktadır. Veri seti herkese açık şekilde sunulurken videolar linklerle gösterildiği için günümüzde platformdan kaldırılmış erişilemeyen birçok video içermektedir.

Bu veri seti kullanılarak Rajagopalan *and* Goecke (2014)'in yaptığı başka bir çalışmada kendini uyarıcı davranışları tespit etmek için çocuğun pozlarına odaklanmayı sağlayacak bir algoritma önermişlerdir. Çocuğun pozlarını içeren sınırlayıcı kutu tahminleri elde etmek için pozelet belirleyicileri kullanılmıştır. Poselet, bir görüntüdeki kişinin konumlarını tanımlamak için kullanılan insan pozları ile önceden eğitilmiş vücut parçası belirleyicisidir. Tüm tahmin edilen sınırlayıcı kutulara güven puanı verilmekte ve en yüksek puana sahip olan sınırlayıcı kutunun takibi yapılmaktadır. Belirlenen poz sınırlayıcı kutuları izlenmektedir. Bu çalışmanın katkılarından biri bir çocuğun mümkün olan en iyi poselet sınırlayıcı kutu tahminini elde etmesidir. Tespit edilip sınırlayıcı kutu ile gösterilen vücut bölümlerindeki baskın hareketi global hareket tanımlayıcısı olan Baskın Hareketlerin Histogramı (HDM) kullanılarak hesaplanmaktadır. Bu tanımlayıcı kullanılarak önerilen model, OSB belirtilerinden olan kendini uyarıcı davranışların tespiti için kullanılmaktadır. Ayrıca, önerilen yöntem UCF101 (Soomro 2012), Weizmann (Gorelick 2007) gibi hareket tanıma çalışmalarında yaygın olarak kullanılan veri setleri ile denenmiş ve performansı doğrulanmıştır. SSBD veri seti üzerinde iki sınıflı (kafa sallama ve dönme) ve üç sınıflı (kafa sallama, dönme ve kol çırpma) olarak ayrı ayrı k (k=5 ve k=10) doğrulama yöntemi ile deneyler yapılmıştır. İki sınıflı yapıda k = 5 iken standart yöntemde %58, önerilen yöntemde %83,4; k = 10 iken standart yöntemde %69,5, önerilen yöntemde ise %86,6 doğruluk sonucu elde edilmiştir. Üç sınıflı yapıda ise k =5 iken standart

yöntemde %54,5, önerilen yöntemde %73,6; k = 10 iken standart yöntemde %45,8, önerilen yöntemde ise %76,3 doğruluk sonucu elde edilmiştir.

Derin öğrenme yöntemlerinin geliştirip yaygınlaşması ile hareket tanıma çalışmalarında sıklıkla kullanılmaktadır. Bu tür otizmde görülen davranışları tanımada derin öğrenme yöntemlerinin kullanıldığı çalışmalar da vardır.

Derin öğrenme yöntemleri.

Tian *et al.* (2019), OSB atipik eylem tipleriyle, özellikle tekrarlayan davranış etiketiyle açıklanmış video veri seti azlığı ve bu veri setlerinin erişime açık olmamaları sebebiyle OSB hareketlerini tanımak için yeni bir video veri seti oluşturmuşlardır. ASD40h adını verdikleri bu veri seti OSB'nin en yaygın beş atipik ve tekrarlayan davranışlarından olan kol çırpma, kafa sallama, dönme, parmak ucunda yürüme ve parmak hareketlerinin sergilendiği 30 video içermektedir (yaklaşık toplam video süresi 40 saat). Bu çalışma ile video tabanlı erken OSB tesbiti için hareket tanıma yapan etkili ve verimli derin mimari olan Bir Bakışta Erken OSB Algılama (O-GAD) ağını tanıtmışlardır. Tasarlanan bu ağ dört bileşenden oluşur. Bu bileşenler bir 3DCNN zamansal öznitelik çıkarıcısı, bir zamansal piramit ağı, bir OSB davranış detektörü ve tekrarlayan davranış ayırıcısıdır. Modelin performansının iyi olması için C3D (Tran *et al.* 2015) geçici öznitelik haritalarını kullanır. Zamansal piramit ağı, gölge katmanının üst düzey anlamını korurken, sonraki algılama görevleri için özellik haritasını ve çıktı özellik haritalarını sürekli olarak kısaltmak için geçici evrişim kullanır. OSB davranış detektörü, bağlantı parçalarını birden çok kategoriye ayırır ve bölüm sınırlarını bir aşamada rafine eder. Tekrarlayan davranış ayırıcı, tüm videonun tekrarlayan davranış içerip içermediğini ayırt eder. Önerilen ASD40h veri seti ile state-of-art yöntemlerden olan İDT (Wang and Schmid 2013), çift akışlı ağını (two-stream network) (Simonyan and Zisserman 2014) ve zaman mekânsal C3D ağını denemişlerdir. Deneylerde İDT %88,2, çift akışlı ağ (two-stream network) %89 ve zaman mekânsal C3D ağı %93,6 doğruluk verirken, önerilen yöntem %95,2 doğruluk değerine ulaşmıştır.

Zunino (2018), Otizm Spektrum Bozukluğun erken tanısı için sadece video dizileri üzerinden değerlendirmeler yaparak hekimler için otomatik ve objektif destek sağlamayı amaçlayan bir çalışma yapmıştır. Bu çalışmayı yaparken, psikolojik ve nöro bilimcilerin araştırmalarına göre temel motor hareketleri gerçekleştirilirken sağlıklı ve otistik bireylerdeki farklılıktan yola çıkmışlardır. Bu çalışmada, temel motor hareketlerinin video dizisi üzerinden analiz eden bir makine öğrenmesi modeli önererek normal gelişen ve otistik gelişen çocukları ayırt etmenin mümkün olduğunu göstermişlerdir. Bu amaçla, iki grup için bir şişeye ulaşma, kavrama, dökme, yerine koyma davranışlarının sergilendiği video kliplerden oluşturulan yeni

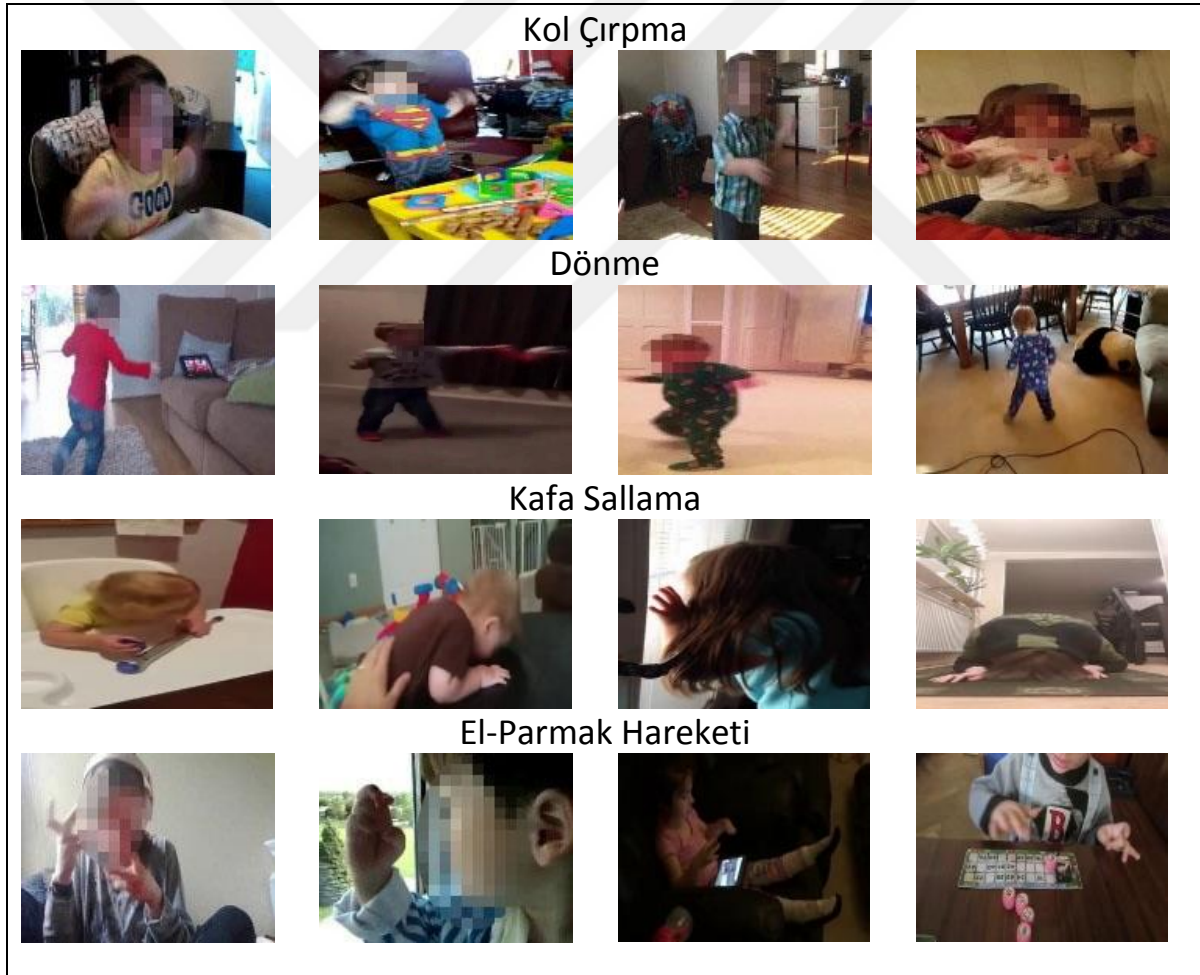
bir veri seti önermişlerdir. Otistik ve normal gelişen çocukların aynı basit hareketleri nasıl gerçekleştirdiğini değerlendirerek otizm tanısı sorununa video tabanlı bir çözüm üretmişlerdir. Bu çalışmayı yapan araştırmacılar, duyu cihazlarını (örn. MRI, mikrofonlar, göz izleyici, hareket yakalama) kullanmayıp, sadece video gözlemlerine güvenmişlerdir. Sınıflandırma için performansı kanıtlanmış bir derin öğrenme ağı olan Uzun-Kısa Dönem Hafıza (LSTM) (Sharma *et al.* 2015) kullanmışlardır. Deneyler sonucunda otistik çocukların otomatik tanısının iyi bir doğruluk derecesi ile mümkün olduğunu göstermişlerdir.

Marinoiu *et al.* (2018) De-Enigma projesindeki (Anonymous 2018), otizmli çocukların robot yardımlı terapi seansları sırasında kaydedilen videolarda tanımlanan hassas eylem ve duygu tanıma ile ilgili çalışma yapmışlardır. Bu çalışmada uzun videolara sahip büyük bir veri kümesi, çok sayıda değişken eylem, tüm vücut parçalarının görünmediği görüntüler, farklı yaşlardaki çocuklar ve öngörülemeyen davranışların yanı sıra standart olmayan kamera bakış açıları gibi sorunların üstünden gelmek için çeşitli yaklaşımlar önermişlerdir. Ayrıca, 3 boyutlu insan poz verilerinden eylem (kolları açma, elleri kaldırma, elleri birleştirme, yüze dokunma vb.) ve duygu tanımanın çoklu yaklaşımları analiz edilmiştir. Tahmin ve tanıma yapılırken kişinin hem 2d hem de 3d eklem pozisyonlarını ve semantik insan vücudu parçası etiketlemesini tahmin eden çok görevli bir derin sinir ağı olan DMHS (Popa *et al.* 2017) kullanılmıştır. DMHS, tamamen görünür insanlar üzerinde eğitildiği için kısmen görünen vücut parçaları için poz tahmini yöntemi önererek tanıma başarılarını arttırmışlardır.

MATERYAL VE METOT

Veri Seti

Bu tez kapsamında, herkese açık sosyal medya platformlarından (YouTube) toplanılan OSB şüphesi olan çocukların video görüntüleri kullanılarak dört farklı harekete (dönme, kol çırpma ve kafa sallama ve el-parmak hareketi) ait Genişletilmiş Kendini Uyarıcı Davranış Veri seti (GKUDV) adında yeni bir veri seti oluşturulmuştur. Arama yapılırken ‘autism stimming’, ‘autism self-stimulatory behaviour’, ‘autism arm flapping’, ‘autism hand action’, ‘autism head banging’, ‘autism spinning’ anahtar kelimeleri kullanılmıştır. Veri setimizin tamamı ebeveynlerin ev ortamlarında cep telefonları ile çocuklarının günlük yaşamını çektikleri video görüntülerinden oluşmaktadır. Şekil 3’de veri setine ait örnek görüntüler bulunmaktadır.



Şekil 3. Veri setinden örnek görüntüler.

Her bir videonun süresi birbirinden farklı olmakla beraber videolarda farklı kişilerin uzun süreli görüntüleri olduğu için hareketlerin olduğu bölümler manuel olarak belirlenip kesme işlemi gerçekleştirilmiştir. Tablo 1’de önceden oluşturulmuş SSBD (Rajagopalan *et al.* (2013)) ve GKUDV veri seti hakkında video sayı bilgileri mevcuttur. SSBD veri setine göre

yeni oluşturduğumuz GKUDV veri setine bir hareket sınıfı daha eklenmiştir. GKUDV veri setinin içeriği SSBD veri setinden tamamen farklı videolardan oluşturulmuştur. Gösterildiği üzere GKUDV veri setinde kol çırpma, el-parmak hareketi, kafa sallama ve dönme olmak üzere 4 adet hareket sınıfı vardır. Kol çırpma sınıfında 43 video, el-parmak hareketi sınıfında 31 video, kafa sallama sınıfında 27 video ve dönme sınıfında 40 video olmak üzere toplamda 141 video bulunmaktadır. İki veri setinde kol çırpma, dönme ve kafa sallama olmak üzere ortak üç hareket sınıfı bulunmaktadır.

Tablo 1. Veri Setlerindeki Hareket Sınıflarına Ait Video Sayısı

VERİ SETİ	Kol Çırpma	Dönme	Kafa Sallama	El-Parmak Hareketi	Toplam
GKUDV	43	40	27	31	141
SSBD (erişilebilir)	24	21	23	YOK	68

Veri setimizi oluşturan videolar ebeveynlerin kendi kameraları ile kaydedildiği için kamera tipi, modeli ve teknik özellikleri hakkında bilginiz bulunmamaktadır. Bu nedenle video karelerinden hareket algısını yakalamak için günümüzde televizyonlarda da kullanılan kare oranı (saniye başına 25 resim karesi) dikkate alınarak standartlaştırılmıştır. Yani, her bir video görüntüsü saniye başına ortalama 25 video karesine dönüştürülmüştür. Bu işleminden sonra veri setimizde kol çırpma hareket sınıfında 5.914, el-parmak hareketi sınıfında 11.331, kafa sallama hareket sınıfında 6.962 ve dönme sınıfında 21.908 video karesi bulunmaktadır. Tablo 2’de her hareket sınıfı için en fazla, en az ve ortalama video karesi sayıları ayrıntılı olarak gösterilmiştir.

Tablo 2. Veri Setimizdeki Hareket Sınıflarına Ait Video Kare Sayıları

Orijinal Veri Seti	Kol Çırpma	El-Parmak Hareketi	Kafa Sallama	Dönme	Toplam
Toplam Video Kare Sayısı	5 938	11 342	6 982	21 951	46 213
Maksimum Video Kare Sayısı	313	3 828	1 679	2 545	
Minimum Video Kare Sayısı	45	30	30	90	
Ortalama Video Kare Sayısı	138	365	258	548	

Veri setindeki görüntüler kız ve erkek çocuklardan oluşan 108 farklı kişinin hareketlerinden oluşmaktadır. Kol çırpma hareket sınıfında 24 erkek, 10 kız çocuk toplamda 34 farklı kişi, el-parmak hareket sınıfında 6 erkek, 7 kız çocuk toplamda 13 farklı kişi, kafa sallama hareket sınıfında 17 erkek, 7 kız çocuk toplamda 24 farklı kişi ve dönme hareket sınıfında 29 erkek, 8 kız çocuğu toplamda 37 farklı kişinin hareket görüntüleri bulunmaktadır. Ayrıca, bir

kişinin sadece bir hareket sınıfında görüntüsü bulunmaktadır. İlgili veriler Tablo 3’de ayrıntılı olarak gösterilmiştir.

Tablo 3. Videolardaki Kız-Erkek Dağılımı

	Kol Çırpma	El-Parmak Hareketi	Kafa Sallama	Dönme
Erkek	24	6	17	29
Kız	10	7	7	8
Toplam	34	13	24	37

Metot

Bu tez kapsamında hareket analizi için iki farklı yaklaşım kullanılarak başarımlar oranları analiz edilmiştir: 1) Görsel Kelime Çantası çerçevesinde klasik makine öğrenme yöntemleri 2) Derin öğrenme algoritmaları.

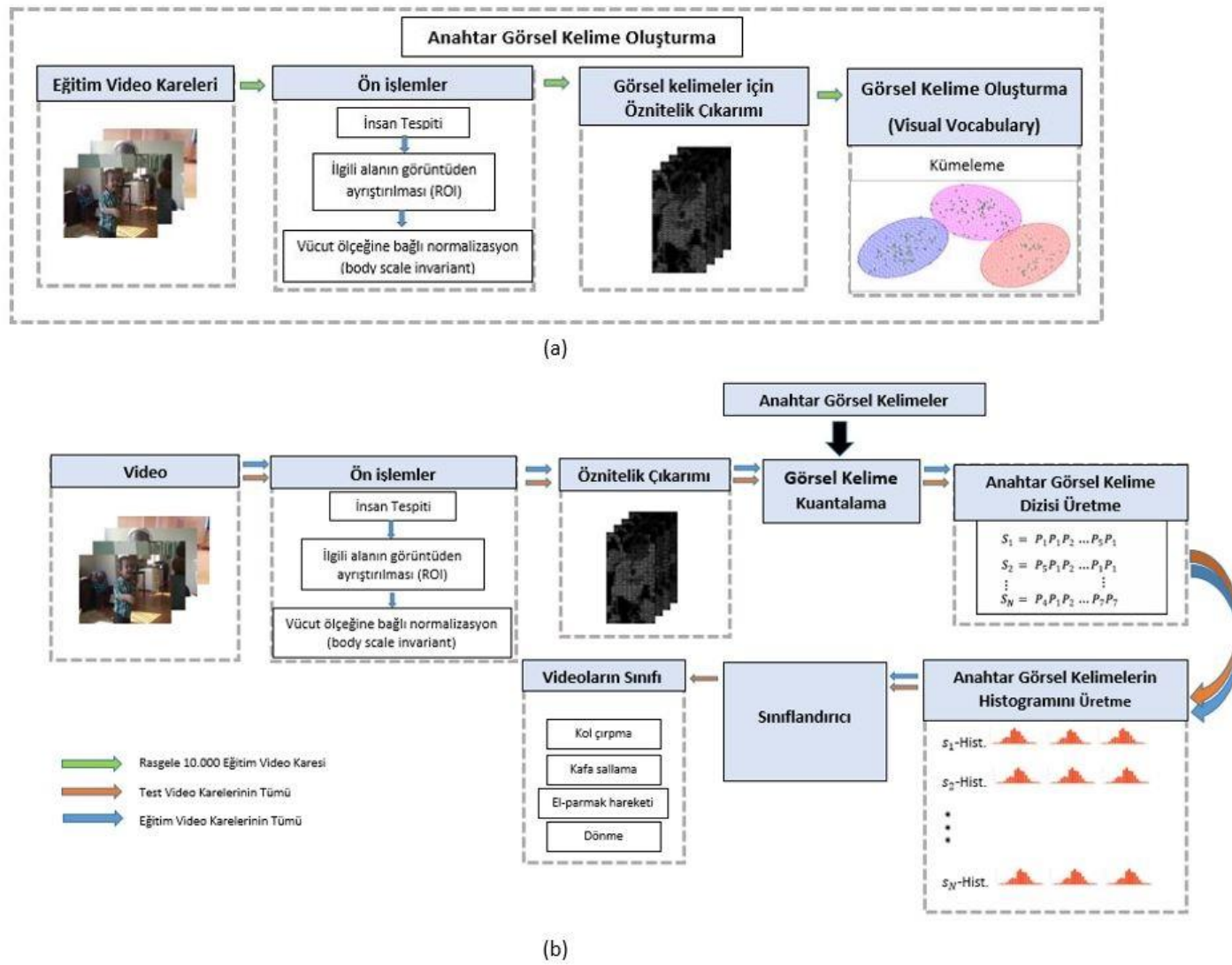
Görsel Kelime Çantası çerçevesinde kullandığımız tanımlayıcılar HOG, HOF, HOG-HOF, SIFT VE SURF yöntemleridir. Bu tanımlayıcılar videolar üzerinden insan hareketi tanımada iyi performans göstermektedirler (Bilinski and Bremond 2011). Bu yöntemler değişiklikleri görme ve ölçeklendirme konusunda sağlam, uygulanmaları kolay ve hesaplama hızları yüksektir. Ayrıca Görsel Kelime Çantası yaklaşımında sıkça kullanılmaktadırlar.

Son zamanlarda Derin Öğrenme ve Evrimsel Sinir Ağları (CNN) yaygın olarak kullanılmaya başlanmıştır. Bu yöntemlerin yaygınlaşmasıyla hareket tespitinde (Yeung 2016) ve hareket tanınmasında (Simonyan and Zisserman 2014) önemli ve etkileyici gelişmeler olmuştur. Bu gelişmeler, çalışmamızda hareket tanıma için Derin öğrenme tabanlı yaklaşımları kullanmaya yönlendirmiştir. Hem mekânsal hem de zamansal bilgileri doğrudan ham videodan yakalamak için Evrimsel uzun-kısa dönem hafıza ağı (ConvLSTM) ve 3 boyutlu evrim kullanan 3 boyutlu evrimsel sinir ağı (3DCNN) kullanılmıştır.

Görsel kelime çantası yaklaşımı ile hareket analizi

Şekil 4’de önerilen yöntemin akış diyagramı verilmiştir. Yöntemimiz iki aşamadan oluşmaktadır. Birinci aşamada anahtar görsel kelimeler üretilmektedir. Veri setimizdeki her bir videoda çok fazla görsel kelime bulunmaktadır. Literatürde bahsedildiği gibi hareket tahmini için tüm görsel kelimeler gerekli değildir. Hareketleri daha çok temsil eden görsel kelimeler yeterlidir (Schindler and Gool 2008). Bu aşamada amaç, bir hareketi diğer bir hareketten ayırmada daha çok bilgilendirici olan anahtar görsel kelimeleri çıkarmaktır.

İkinci aşamada eğitim ve test olmak üzere tüm videolardan çıkarılan görsel kelimeleri birinci aşamada oluşturulan anahtar görsel kelime temsilleri ile görsel kelime dizileri üretilmektedir. Video verilerimizin süreleri farklılık gösterdiği için anahtar görsel kelime dizi boyutları aynı değildir. Bunun üstesinden gelmek için, bu anahtar görsel kelime dizileri ile sabit boyutlu (anahtar görsel kelime sayısı boyutunda) histogramlar üretilmektedir. Bu histogramlar, bir videoyu temsil eden anahtar görsel kelime dizisindeki her görsel kelimenin dizide kaç kez tekrar ettiğini sayarak oluşturulmaktadır. Anahtar görsel kelime dizileri ile temsil edilen videolar artık her bir video için ayrı üretilen sabit boyutlu histogramlarla temsil edilmektedir. Anahtar görsel kelime dizilerini histogramla temsil etmek zamansal özelliklerin kaybolmasına neden olmaktadır. Bunun için, çalışmamızda tanıma yapılırken zamansal özellikleri kısmen dâhil etmek için videoyu belli sayıda parçaya bölüp her parçanın Anahtar görsel kelime dizisinden histogram oluşturulmuştur. Videonun tamamının Anahtar görsel kelime dizisinden oluşturulan histogram ve her iki parçadan oluşturulan histogramlar olmak üzere bir video üç histogramla temsil edilmektedir. BoVW yaklaşımının avantajlarından biri hareket sınıflarına ait videoları sabit uzunluktaki histogramlar ile temsil ederek videonun uzun ya da kısa olmasının üstesinden gelebilmesidir (Foggia 2013).



Şekil 4. Yöntemin akış diyagramı (a) Anahtar görsel kelime oluşturma (b) Videoların histogramlarla temsil edilip sınıflandırılması.

Veri önışleme

Veri setimizdeki videolar klinik ortam ya da özel tasarlanmış ortamlarda kaydedilmeyip ev gibi doğal ortamlarda kaydedildiğı için arka plan karmaşıklığı bulunmaktadır. Bu karmaşıklıklar ilgili olmayan nesne veya başka bir insan hareketinden kaynaklanabilmektedir. Görüntülerdeki arka plan karmaşıklığını azaltmak, hareket tanıma işlemini kolaylaştırıp tanıma doğruluğunu arttırabilmek için verilerin ön işlemden geçirilmesi gerekmektedir. Şekil 4’de görüleceğı üzere verilerin ön işlemleri üç aşamadan oluşmaktadır. i) Görüntülerden insan tespiti ii) İlgili alanın bulunması iii) Vücut ölçeğine bağı nomalizasyon

i) İnsan Tespiti

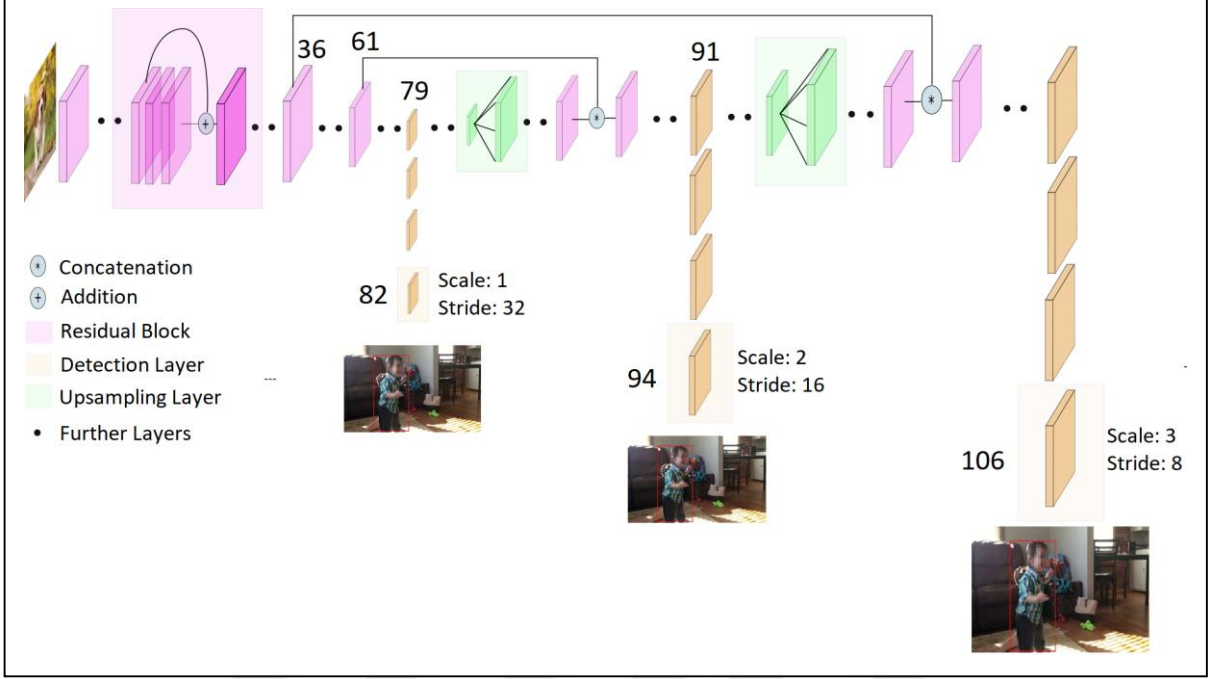
Görüntülerdeki insan hareketleri bir nesne veya başka bir insanla etkileşimde bulunulmayan hareket türlerindendir. Bu yüzden bizim çalışmamızda ilgili alanımız (region of interest (ROI)) (Foggia 2013) hareketi yapan çocuktur. İlgili alanı bulmak için görüntü üzerinde insan tespiti ve takibi yapılmaktadır. İnsan tespiti için YOLOv3 (Redmon and Farhadi 2018), Haar-Cascade (Viola and Jones 2001), HOG Descriptor (Dalal and Triggs 2005) yöntemleri kullanılmıştır.

Haar-Cascade: Makine öğrenimi tabanlı bir yaklaşımdır. Yöntem eğitim verileri ile eğitilmekte ve daha sonra diğer görüntülerdeki nesneleri tespit etmek için bu eğitilen model kullanılmaktadır. Yüz tespiti için yaygın olarak kullanılan bir metoddur. Haar-Cascade özneteliğı piksellerin yoğunluk değeri yerine bitişik dikdörtgen piksel grupları arasındaki kontrast değeriindeki değışimi kullanmaktadır. OpenCV kütüphanesinde, bu yöntem ile nesne tespiti sağlamak için önceden eğitilmiş modüller bulunmaktadır. Çalışmamızda OpenCV kütüphanesinin “Tam Vücut Algılama” modeli kullanılmıştır.

HOG Descriptor: Yöntem, hücelere bölünmüş görüntünün gradyan yönelimlerinin iyi normalize edilmiş yerel histogramlarının değerlendirilmesine dayanmaktadır. Görüntü 8 x 8 boyutunda hücelere bölünmektedir ve hücre içindeki pikseller için eğim ve kenar yönlerinin histogramını hesaplamaktadır. Eğim yönüne göre her hücreyi açısız kutulara bölmektedir. Normalleştirmek için 16 x 16 bloklara ayrılmaktadır. Bu normalleştirilmiş histogram grubu blok diagramı temsil etmektedir. Bu blok histogramları kümesi ise görüntüdeki nesneleri tespit edip tanımlamaktadır.

YOLOv3: Bu yöntem görüntüyü belli boyuttaki hücelere ayırıp her hücre için sınırlayıcı kutuları ve bu kutuların sınıf olasılıklarını tahmin etmektedir. Tahmin edilen olasılıklar sınırlayıcı kutuları ağırlıklandırmaktadır. Tüm olasılıkları tahmin ederken tek bir evrişimsel ağı kullanması ise diğer metodlara göre hızlı olmasını sağlar. Her hücrenin iki

sınırlayıcı kutuyu tahmin etmesi ve sadece bir sınıfa sahip olması yönteme uzamsal kısıtlama getirir. Bu kısıtlama ise yoğun ve küçük nesnelerin tespitinde zorlanması anlamına gelmektedir (Redmon *et al.* 2016). YOLOv3 birçok yönteme göre hızlı ve doğru tespit yapmaktadır (Redmon and Farhadi 2018)(Mohana 2019). Şekil 5’de YOLOv3 ağ mimarisi gösterilmiştir.



Şekil 5. YOLOv3 ağ mimarisi (Kathuria 2018).

YOLOv3 atlama bağlantılarına ve yukarı örneklemeye sahiptir. Giriş görüntüsünün boyutlarını sırasıyla 32, 16 ve 8 oranında küçülterek verilen üç ölçekte tahmin yapmaktadır. YOLO tamamen kıvrımlı bir ağıdır ve son çıktısı, ağıdaki üç farklı yerde üç farklı boyuttaki özellik haritalarına 1 x 1 algılama çekirdeği uygulanarak yapılmaktadır.

Algılama çekirdeğinin şekli 1 x 1 x (B x (5 + C)). Burada B, özellik haritasındaki bir hücrenin tahmin edebileceği sınırlayıcı kutu sayısıdır, “5” 4 sınırlayıcı kutu özelliği ve bir nesne güveni içindir ve C sınıf sayısıdır. Çekirdek boyutu 1 x 1 x 255’tir. Bu çekirdek tarafından üretilen özellik haritası, önceki özellik haritasının aynı yüksekliğine ve genişliğine sahiptir.

Girdi olarak verilen görüntünün boyutu 416 x 416 olarak belirlenmiştir. İlk tespit 82. katman tarafından yapılmaktadır. İlk 81 katman için, görüntü ağ tarafından alt örneklere ayrılmaktadır, böylece 81. katman 32’lik bir adım atmaktadır. Ortaya çıkan özellik haritası 13 x 13 boyutunda olur ve 13 x 13 x 255 algılama özelliği haritası vermektedir.

Daha sonra, katman 79’daki özellik haritası, 26 x 26 boyutlarına 2 katı oranında alt örneklere ayrılmadan önce kıvrımlı katmana tabi tutulmaktadır. Bu özellik haritası, katman 61’deki özellik haritasına derinlikle birleştirilir. Daha sonra, ikinci tespit 94. tabaka tarafından yapılır ve 26 x 26 x 255’lik bir tespit özelliği haritası vermektedir.

Katman 91'deki özellik haritasının, katman 36'daki özellik haritasına derinlik birleştirilmeden önce birkaç kıvrımlı katmana tabi tutulduğu benzer bir prosedür izlenir. Daha önce olduğu gibi, önceki aşamalardan gelen bilgileri birleştirmek için birkaç 1 x 1 kıvrımlı katman izlenir. 52 x 52 x 255 büyüklüğündeki özellik haritasından yararlanarak 106. katmanda çıktının son hali alınır.

Veri setimizdeki videolar hareketi yapan çocuğa odaklı kaydedilmeye başlanmış fakat çocuk harekete devam ederken ev ortamında yer değiştirdiği için farklı kişiler görüntü karesine dâhil olmuştur. Bu durum, videodaki alakasız insanların tespitinden dolayı asıl hareketi yapan çocuğun pozlarının gözden kaçırılmasına sebep olmaktadır. Bu durumu çözmek için hareketi yapan asıl kişinin tespiti ile beraber takip işleminin de yapılması gerekmektedir. Sıralı video karelerindeki kişiyi YOLOv3 algılayıcısı ile tespit edip takip için SORT (Bewley *et al.* 2016) izleyicisi kullanılmaktadır. SORT, nesneleri izlemek için Kalman filtresi kullanmaktadır (Bathija and Sharma 2019).

YOLO, tespit edilen nesnenin sınıf etiketine sahip sınırlayıcı kutular üretmektedir. Bu tespitler daha sonra SORT takipçisine iletmektedir. SORT, tespiti gösteren her sınırlayıcı kutuya numara atayarak takip işlemini yapmaktadır. Tespit yapılarak takip işleminin uygulanması atanan numaraların her video karesinde başka sınırlayıcı kutuya atanmasını engellemektedir. Böylelikle sadece ilgili kişinin tespiti ve takibi sağlanmaktadır.

ii) İlgili alanının (ROI) görüntüden ayrıştırılması

İncelenen hareketler bir kişi tarafından gerçekleştirilen, bir nesne veya başka bir kişiyle etkileşime girilmeyi gerektirmeyen türdendir. Bu yüzden görüntülerdeki hareketi yapan asıl kişinin görüntüsü ve pozu yeterlidir. Asıl kişinin tespit ve takibinde elde edilen sınırlayıcı kutular görüntü karelerinden kırılarak çıkarılmıştır. Bu işlem sonucunda arka plandaki farklı kişi ve nesneler, analiz penceresinden uzaklaştırılmıştır.

iii) Vücut ölçeğine bağlı normalizasyon

Veri setinde neredeyse her yaşta çocukların görüntüleri bulunmaktadır. Üç yaşındaki çocuk ile altı yaşındaki çocuğun vücut ölçüleri aynı değildir. Analizin vücut ölçeğinden bağımsız olmasını sağlamak için kırılan görüntü en-boy oranı sabit tutularak yeniden boyutlandırma yapıp her kare için sabit bir boyut ayarlanmıştır.

Yeniden boyutlandırılırken öncelikle görüntüden kırarak çıkartığımız ilgi alanımızın yükseklik (h) ve genişlik (w) bilgileri alınmaktadır. Çalışmamızda hareket eden insanlar analiz edildiği için yükseklik değerinin makul bir değer olması gerekmektedir. Bu sebeple yükseklik için 320 piksel belirlenmiştir. Değişim oranı (r);

$$r = 320 / h$$

formülü ile hesaplanır. Hesaplanan r (değişim oranı) değerine göre yeni genişlik değeri (W);

$$W = w * r$$

şeklinde hesaplanır. Yeni boyutlar ($(w*r)$, 320) olarak hesaplanmıştır. Bu boyutlara göre kırılan alan yeniden boyutlandırılır.

Hareket tanımadaki üstesinden gelinmesi gereken başka konu ise hareketin kameranın görüş açısından bağımsız yapılmasıdır. Bir hareket 3 boyutlu olarak gerçekleştiğinden ve kameranın bakış açısına bağlı olarak 2 boyutlu görüntüye yansıtıldığından, yansıtılan 2 boyutlu yörünge değişebilir. Bu nedenle, aynı eylemin çok farklı yörüngeleri olabilir ve farklı eylemlerin yörüngeleri aynı görünebilir. Bu, yörüngelerin daha yüksek seviyede yorumlanmasında bir sorun yaratabilir. Bununla birlikte, eylemlerin temsili yalnızca görünümle değişmeyen özellikleri yakalarsa, daha yüksek düzeyli yorum herhangi bir belirsizlik olmadan devam edebilir (Rao 2001). Videoalar ebeveynler tarafından tek bir kamera ile geçmişte kaydedildiği için kamera görüş açısından bağımsız olması sağlanamamaktadır. Bu çalışmada uygulanan ön işlem adımları Şekil 6’da gösterilmiştir.



Şekil 6. Ön işlem adımları.

Öznitelik çıkarma

Ön işlem kısmında tespit edilen ilgi alanımızın (ROI) görünümü ve hareketin tanımlanması için özniteliklerinin çıkarılması gerekmektedir. Çıkarılan öznitelikler görüntüyü temsil etmektedir.

Çalışmamızda HOG (Histogram of Oriented Gradient) (Dalal and Triggs 2005), HOF (Histogram of Optical Flow) (Laptev *et al.* 2008), HOG-HOF, SIFT (Scale Invariant Feature Transform) (Lowe 2004) ve SURF (Speed Up Robust Features) (Bay *et al.* 2008) yöntemleri görüntülerden öznitelik çıkarmak için kullanılmıştır. Bu tanımlayıcılar videolar üzerinden insan

hareketi tanımda iyi performans gösterdiklerini kanıtlamışlardır (Bilinski and Bremond 2011). Tanımlayıcıların uygulanmaları kolay ve hesaplama hızları yüksektir.

HOG, izleme ve hareket tahminine dayanmayan, nesne görünüm bilgilerini yerel yoğunluk gradyanlarının veya kenar yönelimlerinin dağılımıyla tespit eden bir yöntemdir. Yoğun bir hücre ızgarasıyla döşenmiş görüntünün gradyan yönelimlerinin iyi normalize edilmiş yerel histogramlarını değerlendirmektedir. Gradyan büyüklüğü, ani yoğunluk değişimi bölgeleri olan kenarlar ve köşeler etrafında büyük olduğu için düz bölgelere göre nesne şekli hakkında daha fazla bilgi vermektedir. HOG tanımlayıcısı, her bir video karesinden **[1, 3780]** boyutunda bir öznitelik vektörü çıkarmaktadır.

HOF, optik akış bilgilerini kullanarak görüntü dizisinin yerel hareket bilgilerini tanımlamaktadır. Optik akış, nesnenin veya kameranin hareketinden kaynaklanan iki ardışık video karesi arasındaki nesnelerin görünür hareket modelidir.

Yöntem temel olarak görüntünün alt bölgelerindeki baskın hareketi hesaplamaktadır. Her görüntü pikselinin birinci görüntü karesi ile ikinci görüntü karesindeki yer değiştirme vektörleri bulunur. Hem yer değiştirme büyüklüğü hem hareket yönü, 2 boyutlu optik akış histogramları kullanılarak nicelendirilmekte ve bu histogram bölmelerinin her birine bir sembol atanarak baskın hareket kodlanmaktadır. Bu şekilde görüntüdeki tüm hareketlerin kompakt bir temsili oluşturulmaktadır. İki varsayım üzerine çalışır;

1. Bir nesnenin piksel yoğunluğu ardışık kareler arasında değişmez,
2. Komşu pikseller benzer harekete sahiptir.

En belirgin avantajı, değişken ışıklandırma ve karmaşık arka planda hareketi doğru tahmin etme sorumluluğunun tamamen optik akış hesaplaması ile sınırlı olmasıdır (Perš *et al.* 2010). HOF tanımlayıcısı, her bir video karesinden **[1, 18954]** boyutunda bir öznitelik vektörü çıkarmaktadır.

HOG-HOF, görünümü kodlamak için HOG, hareketi kodlamak için HOF öznitelik vektörlerinin arda arda eklenerek kombine edilen öznitelik çıkarıcısıdır. HOG-HOF tanımlayıcısı her video karesinden **[1, 22734]** boyutunda bir öznitelik vektörü çıkarmaktadır.

SIFT, bir görüntünün, aydınlatma, döndürme ve ölçeklendirmeye karşı değişmeyen bölgesel özniteliklerini belirleyip tanımlayan bir yöntemdir. Özniteliklerin çıkarılması için, öncelikle görüntüye farklı ölçeklerde Gaussian filtresi uygulanıp resimler arasındaki farklar alınır. Alınan bu Gaussian farklarının (GoD) ekstremum noktaları bulunur. Bu noktalar lokasyonu belli olmayan birçok aday anahtar nokta sunar. Anahtar nokta adaylarının her birine interpolasyon uygulanarak lokasyonlar doğru bir şekilde saptanır. Her lokasyona orijinal

görüntü eğim yönlerine göre bir veya daha fazla yön atanır. Bu aşamadan sonraki işlemler, her bir özellik için atanan yönlendirmeye, ölçeğe ve konuma göre dönüştürülmüş görüntü verileri üzerinde gerçekleştirilir ve böylece bu dönüşümlere değişmezlik sağlar. Her biri 8 kutu içeren 4x4 piksel büyüklüğünde yön histogramları oluşturulur. Anahtar nokta etrafındaki 16x16 lık seçilen bir bölgede yer alan, yön ve büyüklük ile ilgili bilgi veren histogramlar hesaplanır. Her bir histogram gerçek komşuluk bölgesinin 4x4 lük bir alt bölgesini içerir. Her biri 8 kutu içeren, 16(4x4) histogramlar, toplamda 128 adet vektör belirtmiş olurlar. Bu vektörler, anahtar nokta tanımlayıcılarıdır. Bunlar, önemli düzeyde yerel şekil bozulması ve aydınlatma değişikliğine izin veren bir temsile dönüştürülür. SIFT tanımlayıcısı, her bir video karesinden [1, 128] boyutunda bir öznitelik vektörü çıkarmaktadır.

SURF, Bay *et al.* (2008) tarafından, ölçeklendirme ve döndürmeye karşı değişmezlik için SIFT ile aynı yaklaşımları kullanıp, hesaplamalardaki yavaşlığı gidermede etkili yöntemler benimsediği için SIFT'e göre daha hızlı bir yöntem olarak önerilmiştir. SIFT yönteminin anahtar nokta tahmini için kullandığı Gaussian filtreleri, hesaplama olarak daha pahalıdır. SURF tanımlayıcı, Gaussian filtreleri yerine kutu filtrelerine verilen yanıtlardan türetilmiştir (Gauglitz *et al.* 2011). SURF tanımlayıcısı, her bir video karesinden [1, 64] boyutunda bir öznitelik vektörü çıkarmaktadır.

Anahtar görsel kelime üretme

Videolardan çıkarılan her bir öznitelik vektörü bir görsel kelimedir. Bir video, video kare sayısı kadar görsel kelime içerir. Öznitelik çıkarma sürecinin, video karesinde görünen gölgeler, arka plan dokuları alakasız nesneler gibi durumlardan etkilenebilmektedir. Bu, çıkarılan her öznitek vektörünün hareket sınıflandırması için bilgilendirici olmadığı anlamına gelmektedir (Schindler and Gool 2008). Bilgilendirici olmayan görsel kelimeleri değerlendirmek modelimizin performansını düşürebilir. Bu aşama ile hareketler için bilgilendirici görsel kelimelerin üretilmesi amaçlanmıştır. Bu bilgilendirici görsel kelimelere anahtar görsel kelimeler denilmektedir.

Anahtar görsel kelimeler, eğitim video karelerinden çıkarılan öznitelik vektörlerini kümeleyerek oluşturulur. Kümelendirken elde edilen küme merkezlerinin her biri bir anahtar görsel kelimedir.

Çalışmamızda, anahtar görsel kelimeler oluşturulurken en yaygın kümeleme algoritmalarından biri olan K-means (MacQueen *et al.* 1967) algoritması kullanılmıştır.

Uygulamamızda K-means kümeleme algoritmasının hesaplama karmaşıklığını azaltmak için Wang *et al.* (2009)'de önerildiği gibi eğitim video karelerinden rastgele örnek seçilmiştir. Veri setindeki videolar,

$$x_i = [x_{i1}, x_{i2} \dots, x_{im}]$$

$i \in \mathbf{R}^n$ n eğitim video sayısı ve m her bir videodaki video kare sayısıdır. Tüm eğitim video kareleri dizisi $X = [x_1, x_2, x_3 \dots, x_n] \in R^{n \times m}$ ($\sum_{i=1}^n x_i$) şeklindedir. Bu diziden rastgele 10.000 video karesi seçilir. Rastgele seçilen video karelerinin öznitelik vektörleri;

$$f = [f_1, f_2 \dots, f_N] \in R^{D \times N}$$

şeklindedir. D çıkarılan öznitelik vektör boyutu, N rastgele seçilen video kare sayısıdır. Çıkarılan öznitelik vektörleri K-means yöntemi ile kümelendirilir. Bu yöntemde ilk başta çıkarılacak anahtar görsel kelime sayısı kadar küme merkezleri ($v = \{v_1, v_2, \dots, v_c\}$, c küme merkezi sayısı) rastgele atanır. Daha sonra küme merkezleri dışında kalan her bir öznitelik vektörü ile küme merkezleri arasındaki uzaklık,

$$Dist_{ab} = \sqrt{\sum_{k=1}^m (a_{ik} - b_{ik})^2} \quad (1)$$

Öklid mesafe ölçütü formülü ile hesaplanır. Daha sonra bir öznitelik vektörünün hangi küme merkezine olan uzaklığı en az ise o kümeye atanır ve yeni küme merkezleri,

$$v_i = \left(\frac{1}{c_i}\right) \sum_{i=1}^{c_i} f_i \quad (2)$$

şeklinde hesaplanır. c_i , i'ninci kümedeki öznitelik vektör sayısıdır. Bir öznitelik vektörünün ait olduğu küme sabit kalana kadar kümeye atama ve küme merkezi güncelleme işlemleri tekrarlanır. Kümeleme tamamlandıktan sonra elde edilen küme merkezleri anahtar görsel kelimelerimizdir. Anahtar görsel kelimeler,

$$S = [P_1, P_2, P_3 \dots, P_k] \in R^k$$

şeklindedir ve k küme merkezi (anahtar görsel kelime) sayısıdır. Üretilcek anahtar görsel kelimelerin diğer görsel kelimeleri temsil etme gücü, kümeleme yapılırken girdi olarak verilen yeterli öznitelik örneği ve sonucunda oluşacak toplam anahtar görsel kelime sayısı ile ilişkilidir.

Anahtar görsel kelimelerin temsil gücü ne kadar çok ise hareketin sınıflandırma başarısı da o oranda iyi olabilmektedir (Cortés *et al.* 2018).

Görsel kelime kuantalama

Bu aşama, bir videodan çıkarılan öznitelik vektörlerinin (görsel kelimelerin) k-means algoritmasında bulunan küme merkezlerinden (anahtar görsel kelimelerden) birine atanarak nicelendirilmesidir.

Çalışmamızda, hangi öznitelik vektörünün hangi anahtar görsel kelime ile temsil edileceği mesafe ölçütlerinden olan Öklid mesafesi ile belirlenir. Öklid mesafesi kullanılan en yaygın kuantalama yöntemlerinden biridir. (1) ve (2) formülleri ile hesaplanır.

Sande *et al.* (2013) BOW yaklaşımının kuantalama aşamasında yüksek hesaplama maliyeti olduğunu göstermiştir.

Anahtar görsel kelime dizisi oluşturma

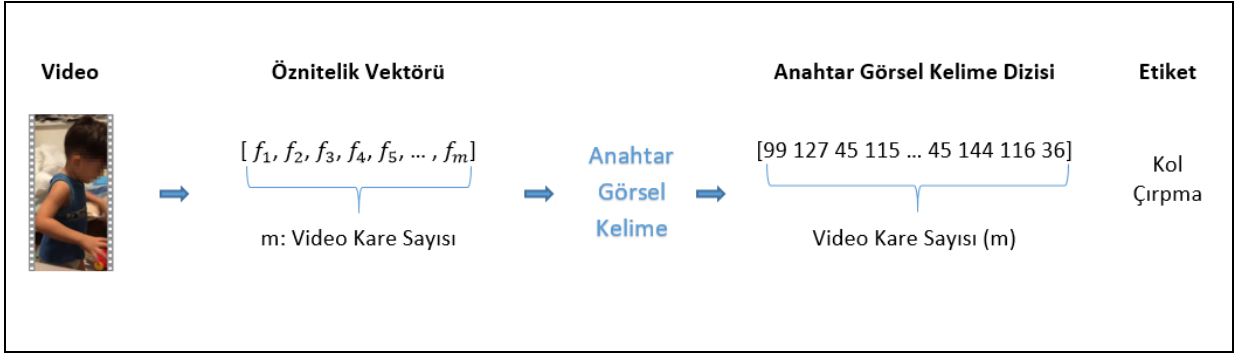
Bütün eğitim ve test videolarının öznitelik vektörleri K-means kümeleme modelinden elde edilen anahtar görsel kelimelerle temsil edilmektedir (Agahian 2018). Videoların öznitelik vektörlerine karşılık gelen anahtar görsel kelimelerden diziler oluşturulmaktadır. Aynı birer dizi olmak üzere eğitim ve test videolarının öznitelik vektörleri çıkarılır. Herhangi bir video için öznitelik vektör dizisi;

$$f_i = [f_{i1}, f_{i2} \dots, f_{im}] \in R^{D \times m}$$

şeklinde. f_i , i 'nci videonun öznitelik vektörü, D öznitelik vektör boyutu, m video kare sayısıdır. Her video için çıkarılan öznitelik vektör dizilerine karşılık gelen anahtar görsel kelime dizileri K-means modelinden alınır. Bir videonun anahtar görsel kelime dizisi S_i ;

$$s_i = [P_{i1}, P_{i2}, P_{i3}, P_{i4}, P_{i5} \dots, P_{im}] \in R^m$$

şeklinde gösterilmektedir. P_i , anahtar görsel kelime dizisindeki i 'nci görsel kelime; m ilgili videonun görsel kelime sayısıdır. Şekil 7'de bir videonun öznitelikleri temsil eden anahtar görsel kelime dizisi ve sınıf etiketi örneği gösterilmiştir. Şekildeki anahtar görsel kelime dizisindeki her bir eleman (99,127...) anahtar görsel kelimelerin etiket değeridir. Anahtar görsel kelime dizisi öznitelik vektörlerinin anahtar görsel kelime temsillerinin etiketleri ile oluşturulmuştur. Videolardaki video kare sayısı m ile gösterilmiştir.



Anahtar görsel kelime dizi boyutları farklılık göstermektedir. Bunun sebebi her videonun uzunluğu farklı olduğu için video kare sayılarının farklı olmasıdır.

Şekil 7. Anahtar görsel kelime dizisi oluşturma örneği.

Anahtar görsel kelimelerin histogramı ile hareket temsili

Bu aşama ile amaçlanan, bir hareketi, anahtar görsel kelimelerin kaç kez görüldüğünü sayan bir frekans histogramı aracılığıyla temsil etmektir. Her video için oluşturulan anahtar görsel kelime dizilerinin aynı boyutta olmasını sağlamak için iki yöntem kullanılmaktadır. İlki, videoların belirli sayıda video karesinden oluşmalarını sağlayacak ön işlemler yapıldıktan sonra öznitelik çıkarımı yapmaktır. Diğeri ise görsel kelime kuantalama ile çıkarılan değerleri temsil eden anahtar görsel kelime dizilerinden frekans histogramı oluşturmaktır (Ramanathan *et al.* 2014). D öznitelik vektör boyutu, K anahtar görsel kelime sayısı olmak üzere;

$$S = [P_1, P_2, \dots, P_k, \dots, P_K] \in R^{D \times K}$$

anahtar görsel kelime dizimizdir. N öznitelik vektör sayısı ve D öznitelik vektörünün boyutu olmak üzere;

$$F = [f_1, f_2, \dots, f_i, \dots, f_N] \in R^{D \times N}$$

bir videodan çıkarılan görsel kelime (öznitelik vektörü) dizisidir. Bu dizideki her görsel kelime en yakın anahtar görsel kelimeye atanmaktadır. Aynı anahtar görsel kelimeye atanan görsel kelimeler benzer görünümü paylaşmaktadırlar. Mesela, F dizisindeki f_1 , S dizindeki P_3 anahtar görsel kelimesine atanmış ise videoya ait anahtar görsel kelime dizinde f_1 , P_3 olarak gösterilir. Tüm videolar için oluşturulan anahtar görsel kelime dizileri frekans histogramı ile temsil edilir. Histogram $Hist_k(F)$ ve k'nıncı histogram bölmesi aşağıdaki gibi hesaplanır.

$$Hist_k(F) = |\{f_i | LB(f_i) = k\}|$$

$LB(f_i) = k$ (k f_i 'nin anahtar görsel kelime etiketi ve $|\cdot|$ anahtar görsel kelime sayma). Her histogram bölmesi için anahtar görsel kelimeler sayılır ve böylelikle histogram oluşturulur.

Hareket anahtar görsel kelime dizilerinden oluşturulan histogram ile temsil edilmektedir. Histogramın boyutu anahtar görsel kelime sayısı olan K 'ya eşittir (Sun *et al.* 2016). Algoritma-1'de histogram oluşturmak için kullanılan algoritma açıklanmıştır.

Algoritma-1 Histogram Oluştur

Girdi: Küme Merkezi Sayısı (N),

Kuantalanmış Anahtar Görsel Kelime Dizileri $S_{Tümü} = [S_1, S_2, \dots, S_k, \dots, S_K] \in R^K$

K video sayısı olmak üzere, S_k , k 'nıncı videonun kuantalanmış anahtar görsel kelime dizisidir.

$S_k = [P_1, P_2, \dots, P_i] \in R^i$

Çıktı: Histogram (Hist)

1: ilk değer atama: $Hist = [[0, \dots, 0], \dots, [0, \dots, 0]]_{K \times N}$

2: $sıra = 0$

3: **for** S_k to S **do**

4: **for** P_i to S_k **do**

5: $Hist[sıra, P_i] = Hist[sıra, P_i] + 1$

6: $sıra = sıra + 1$

7: **return** Hist

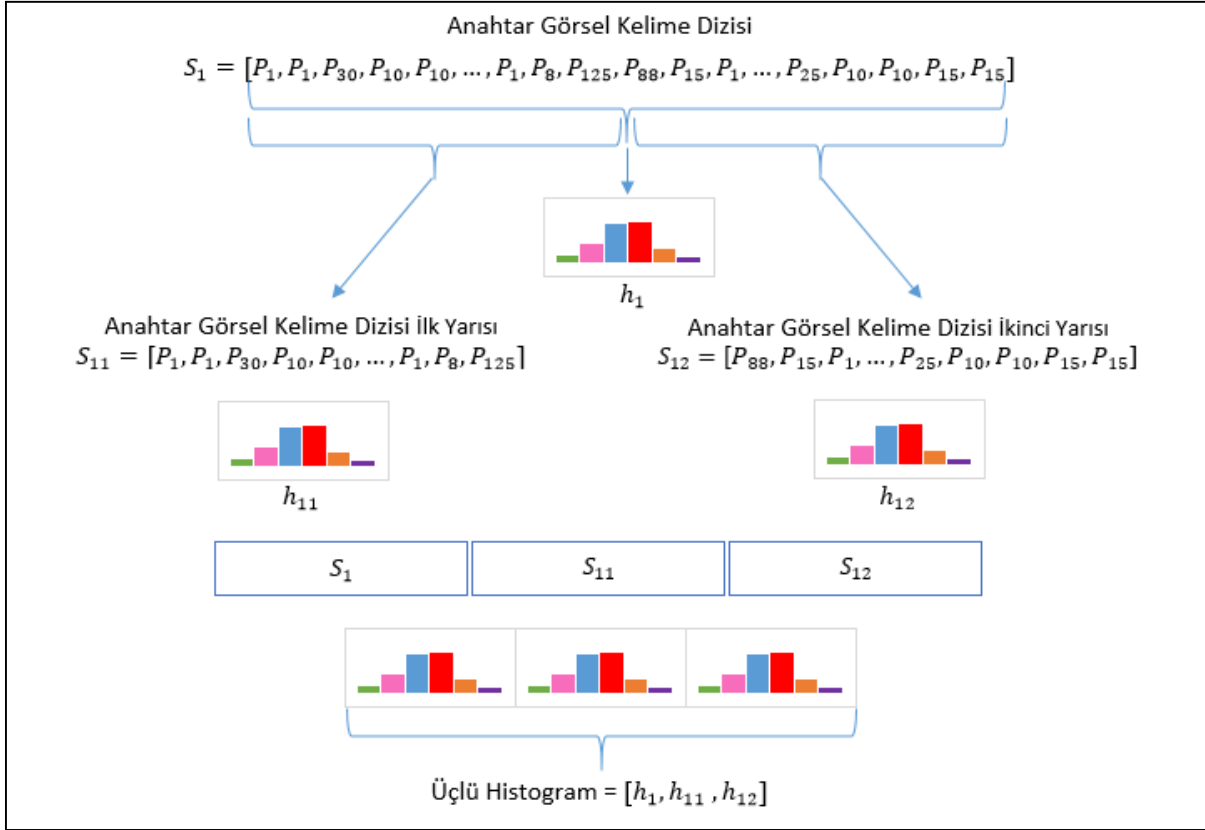
Üçlü histogram ile hareket temsili

Videoları anahtar görsel kelimelerin histogramı ile temsil etmek, öğrenmeyi etkili hale getiren mekânsal-zamansal özniteliklerin zamansal düzenlemelerini göz ardı etmektedir. Uzun süreli zamansal bilginin eksikliği hareketlerin modellenmesini engellemektedir (Wang *et al.* 2008).

Shukla *et al.* (2013)'de zamansal özellikleri dâhil etmek için her videoyu belli sayıda geçici bölümlere ayırıp ve her bölmeyle ait öznitelik vektörlerine karşılık gelen anahtar görsel kelimelerden histogram oluşturan Geçici Kelime Torbası (TBoW) modeli önerilmektedir. Bu yaklaşıma göre bir video ayrılan bölüm sayısı kadar histogramla temsil edilmektedir.

Uygulamamızda, video verilerimizin zamansal özelliklerinin kaybolmasını kısmen önlemek için bu yaklaşımdan yararlanılmıştır. Önceki aşamalarda oluşturulan anahtar görsel kelime dizilerinden üç histogram oluşturulmuştur. Her videoyu temsil eden anahtar görsel kelime dizileri ikiye ayrılmakta ve her iki bölmenin ayrı ayrı histogramı alınmaktadır. Üçüncü histogram ise dizinin tamamından oluşturulan histogramdır. Artık her bir video Şekil 8'deki gibi üç histogramla gösterilmektedir. Şekilde S bir videodaki her bir video kareden çıkarılan öznitelik vektörlerini temsil eden anahtar görsel kelime dizisidir ($(P_1, P_2, \dots, P_k)$, k belirlenen anahtar görsel kelime sayısı). Her bir öznitelik vektörü kuantalama ile atandığı anahtar görsel

kelime ile temsil edilir. h_1 bir videonun tamamından, h_{11} videonun ilk yarısından ve h_{12} videonun son yarısından çıkarılan frekans histogramlarıdır.



Şekil 8. Üçlü histogram gösterimi.

Sınıflandırma

Sınıflandırma iki aşamadan oluşur: bir öğrenme aşaması ve değerlendirme aşamasıdır. Öğrenme aşamasında, sınıflandırıcı modelini tüm eğitim videolarından oluşturulan histogramlarla eğitir ve değerlendirme aşamasında test videoları için sınıflandırıcıdan sınıf etiket tahmini alınır.

Hareketleri sınıflandırmak için Çok Katmanlı algılayıcı (MLP), Destek Vektör Makinesi (SVM) ve Gaussian Naive Bayes (GNB) sınıflandırıcıları kullanılmıştır.

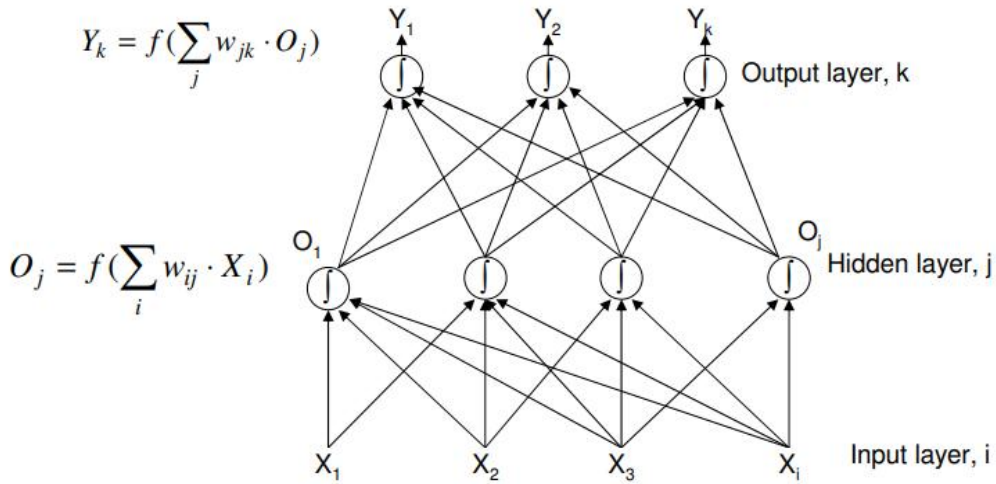
MLP, denetimli öğrenme ve sınıflandırma için ileri beslemeli yapay sinir ağına dayanan bir yöntemdir. Bu sınıflandırıcının önemi, giriş bilgilerinin önceki bilgilere göre istenen yanıt nasıl dönüştürüleceğini öğrenen denetimli bir ağ olmasıdır.

Birincil sınıflandırıcı olarak kullanılan MLP, birbirine bağlanan çoklu düğüm katmanlarından oluşur. Giriş katmanı, gizli katmanlar ve çıkış katmanı bulunmaktadır. Çalışmamızda, oluşturulan histogramlar ve sınıf etiketleri, sınıflandırma için MLP ağından geçirilir. Sınıflandırıcı, eğitim verilerinin histogramı ile eğitilirken, test verilerinin histogramını olası dört hareket sınıfından birine eşlemektedir.

Veriler, algılayıcının x_0 'dan x_{N-1} 'e numaralandırılmış giriş düğümleri ve algılayıcının her bir dalındaki girdi ağırlıkları ile beslenir. Yöntem, ağırlıklı girdileri toplar, sapma değeri ekler ve algılayıcının aktivasyonunu y olacak şekilde dönüştürür (3);

$$y = f_h \left[\left(\sum_{i=0}^{N-1} w_i x_i \right) + \theta \right] \quad (3)$$

Bu nedenle, eşiği (threshold) bir sonraki katmana bağlayan eşik değeri (threshold) ve sapma değeri (bias) çarpımı sabittir. MLP ağ yapısı Şekil 9'da verilmiştir.



Şekil 9. Çok katmanlı algılayıcı yapısı (Swingler 2013).

Toplam sistem doğruluğu, öznelitekt vektörlerinin sayısı, MLP'deki gizli düğümlerin sayısı ve toplam eğitim örnekleri optimize edilerek artırılır. Gizli katmanda yetersiz veya aşırı sayıda nöron bulunan MLP modeli kötü genelleme ve aşırı uyum sorunlarına neden olabilmektedir. Gizli katmandaki nöron sayısını belirlemek için analitik bir yöntem yoktur. Bu nedenle, sadece deneme yanılma yoluyla keşfedilmiştir (Subasi and Ercelebi 2005).

MLP'nin diğer geleneksel sınıflandırma tekniklerine göre iki önemli avantajı vardır. İlki, verilerin dağılımı veya sınıflandırılacak veri gruplarının kovaryans matrislerinin eşitliği ile ilgili varsayımlar gerektirmez. İkincisi, ayrık bölgeler de dâhil olmak üzere doğrusal olmayan karar bölgelerinin oluşmasına izin verir.

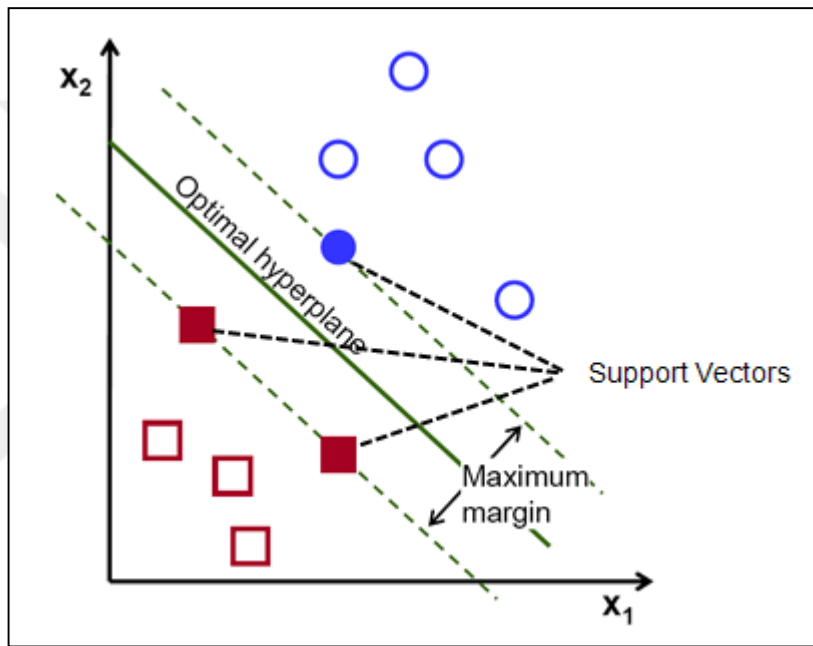
SVM, sınıflandırma ve regresyon analizi için kullanılan, veriyi analiz eden ilişkili öğrenme algoritmalarıyla denetimli öğrenme modelleridir. SVM'nin arkasındaki ana fikir, veri kümesinin en optimal hiper düzlem ayrımını bulmaktır. Bu hiper düzlemi istatistiksel öğrenme teorisini kullanarak hesaplar.

Şekil 10'da görüldüğü gibi iki sınıfı bir birinden ayıran hiper düzleme en yakın veri noktasına olan mesafesine margin denmektedir. SVM margin mesafesinin en büyük olduğu

hiper düzlemi arar. Hiper düzleme en yakın veri noktalarına destek vektörleri denmektedir. Denklem 4’de iki sınıfı birbirinden ayıran hiper düzlemin fonksiyonu verilmiştir. W ağırlık vektörü, b ise sabit bir sayı ve x sınıfı bilinmeyen bir noktayı ifade eder. Denklem (5)’de birbirinden doğrusal olarak ayrılabilen verilerin karar fonksiyonu verilmiştir. Bu fonksiyona göre $f(x) \geq 0$ (bir sınıf) veya $f(x) < 0$ (diğer sınıf) için sınıf tayini yapılmaktadır.

$$w \cdot x + b = 0 \quad (4)$$

$$f(x) = w \cdot x + b = \sum_i^N w_i \cdot x_i + b \quad (5)$$



Şekil 10. Optimal hiper düzlem (Bhalla 2017).

Pal (2008) tüm sınıflar arasında ikili sınıflandırmalar yapmıştır (one-against-one). Çalışmamızda bu yaklaşım kullanılmaktadır. Olası her sınıf çifti için bir sınıflandırıcı eğitir, yani $\frac{k(k-1)}{2}$ (k : hareket sınıfı sayısı) sınıflandırıcı. Test verileri her sınıflandırıcıda denenir ve sonuçta tahmin edilen sınıfa bir oy verilir. Sınıflandırıcıdan çıkan sınıf etiketi en fazla oya sahip olan sınıf etiketidir.

GNB, Naive bayes sınıflandırıcılarının bir uzantısı olan bir sınıflandırma modelidir. Naive Bayes sınıflandırıcıları, bir sınıftaki belirli bir özneliliğin etkisinin diğer özneliliklerden bağımsız olduğunu varsaymaktadır. En basit denetimli öğrenme algoritmalarından biridir. Hızlı, doğru ve güvenilir bir algoritmadır (John and Langley 2013).

Öznitelik olasılık dağılımlarına göre üç türü vardır. Gauss naive bayes, Multinomial naive bayes ve Bernoulli naive bayesdir. GNB aralarında en basiti ve en yaygın olarak kullanılan modeldir.

GNB’de her bir sınıfla ilişkili öznitelik olasılığının normal (veya Gauss) bir dağılıma göre dağıldığı varsayılmaktadır. Sınıflandırma yapmak için eğitim verilerinden öznitelik ortalaması ve standart sapmayı hesaplamaktadır. Bu değerleri kullanarak Olasılık Yoğunluk Fonksiyonu (Probability Density Function (PDF)) ile en yüksek olasılığa sahip giriş özniteliğini seçerek bir sınıf etiketini atamaktadır (Rafique *et al.* 2019).

$$X = [x_1, x_2, \dots, x_i, \dots, x_N] \in R^{D \times N}$$

y, sınıf değişkenidir ve X, N boyutunda bağımlı bir özellik vektörü olmak üzere koşullu olasılık;

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\pi\sigma_y^2}\right) \quad (6)$$

şeklinde hesaplanır (Anonymous 2017).

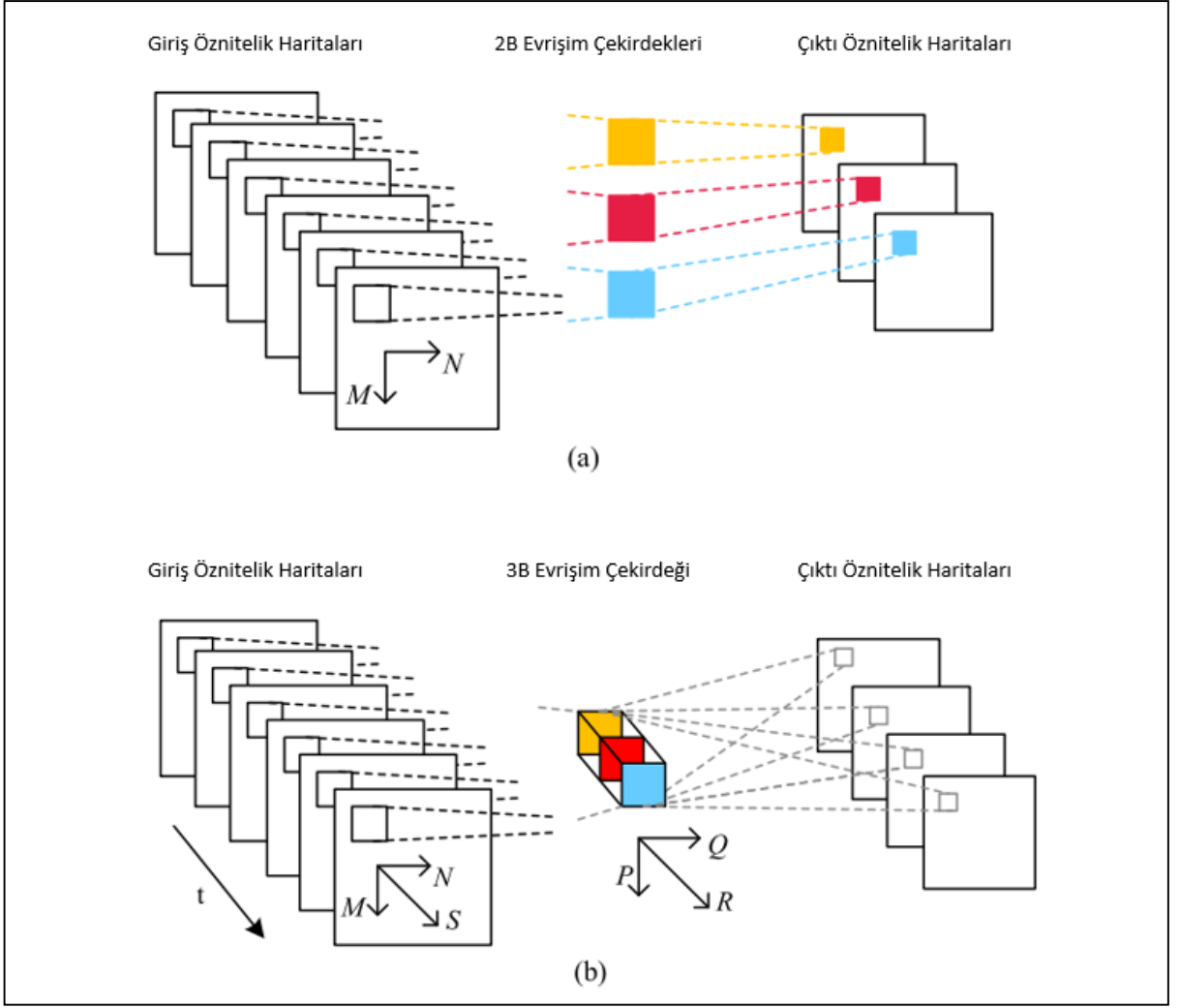
Derin Öğrenme Yöntemleri

Hem mekânsal hem de zamansal bilgileri doğrudan ham videodan yakalamak için ConvLSTM ve 3 boyutlu evrişim kullanan 3DCNN yöntemleri kullanılarak tanıma başarımları analiz edilmiştir.

3 boyutlu evrişimsel sinir ağı

Yöntemlerin çoğu, görüntü düzleminde kaydırma-değişmez (shift-invariant) gösterimlere olanak tanıyan 2 boyutlu (2B) evrişime (2DCNN) sahip mimariler kullanmaktadır. Fakat insan hareketi tanımada eylemin başlangıç ve bitişi genellikle bilinmediği için zaman bilgilerinin dönüşümlerinin değişmezliği önemlidir. Bu sorun 3 boyutlu (3B) zaman-mekânsal evrişimlere sahip olan 3DCNN mimarisi ile çözülür ve video görüntüleri için 2DCNN’nin genişletilmiş halini sunar.

Wang *et al.* (2017)’de açıklandığı gibi, Şekil 11 (a)’da gösterildiği üzere 2 boyutlu evrişim çekirdekleri M ve N ekseninde kayarak tüm görüntüler için görüntü dizisinden bağımsız öznitelik haritası çıkarır. 3 boyutlu evrişim çekirdekleri ise Şekil 11 (b) ‘deki gibi M, N ve S ekseninde kayar ve hem mekânsal (uzaysal) hem de zamansal bilgilerden çıkış öznitelik haritaları çıkarır.



Şekil 11. 2B ve 3B evrişim katmanları arasında karşılaştırma, (a) 2D evrişim çekirdekleri, (b) 3D evrişim çekirdekleri (Wang *et al.* 2017).

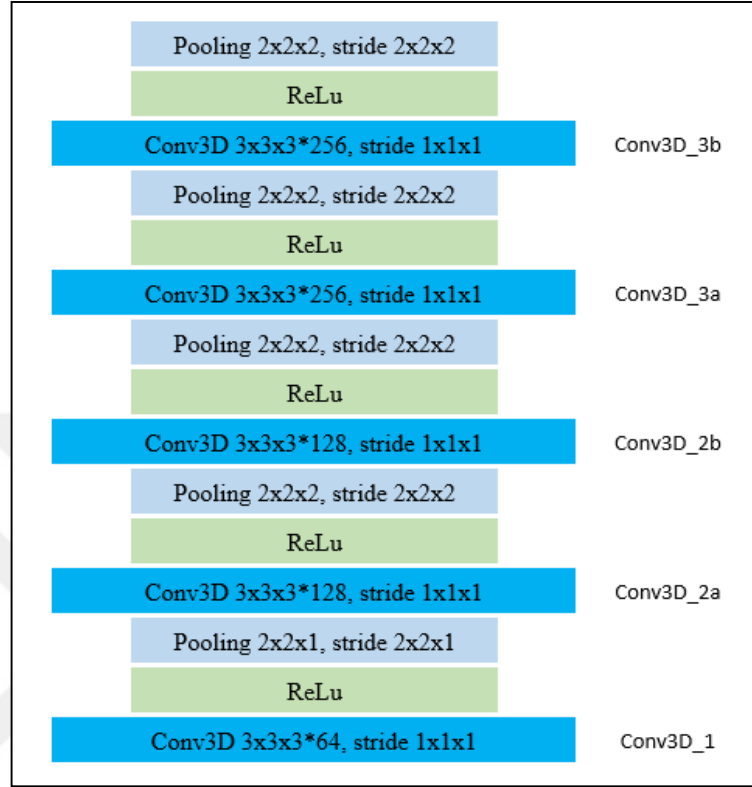
3DCNN, 3 boyutlu çekirdeği, ardarda gelen birden çok video karesini bir araya getirerek oluşturulan küpün içine doğru döndürerek elde edilir. Evrişim katmanındaki öznitelik haritaları bir önceki katmandaki birden çok komşu kareye bağlanarak hareket bilgisi yakalanır (Boualia 2018).

3DCNN yönteminde, i 'ninci katmandaki j 'ninci öznitelik haritasında (m,n,s) pikselinin birim değeri Ji *et al.* (2013) tarafından verilen denklem (7) ile hesaplanır.

$$v_{ij}^{mns} = \tanh(b_{ij} + \sum_t \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijt}^{pqr} v_{(i-1)t}^{(m+P)(n+q)(s+r)}) \quad (7)$$

$\tanh()$ hiperbolik tanjant fonksiyonu, b_{ij} j 'ninci öznitelik haritası için sapma (bias) değeri, R_i zamansal boyuttaki 3 boyutlu çekirdek boyutu, t mevcut öznitelik haritasına bağlı $(i-1)$ 'inci katmandaki öznitelik haritaları üzerindeki dizinler, P_i çekirdeğin yüksekliği, Q_i çekirdeğin genişliğidir. w_{ijt}^{pqr} ise önceki katmandaki t 'ninci öznitelik haritasına bağlanan çekirdeğin (p,q,r) 'inci değeridir.

Şekil 12’de yöntemin bileşenleri gösterilmiştir. Mimaride, 3B Evrişim Katmanı (conv3d), ReLU, BatchNorm, 3B Max Havuzlama Katmanı (MaxPooling3D), Softmax Sınıflandırıcısı vardır. Her 3B evrişim katmanı için kernel boyutu 3x3x3, adım (stride) boyutu 1x1x1’dir. Pooling katmanı için kernel boyutu 2x2x2 ve adım (stride) boyutu 2x2x2’dir.



Şekil 12. 3DCNN bileşenleri (Zhang *et al.* 2017).

3B evrişim katmanı 5 parametre almaktadır. Bunlar sırasıyla; yığın(batch) boyutu, girdi genişliği, girdi yüksekliği, girdi derinliği ve renk kanal sayısıdır. Görüntüler 32x32 olarak yeniden boyutlandırılmıştır. Çıktı parametreleri ise sırasıyla; batch boyutu, çıktı genişliği, çıktı yüksekliği çıktı derinliği ve evrişimdeki çıktı filtresi sayısıdır. Görüntüler RGB olarak işlendiği için renk kanal sayısı 3 tür.

Evrişimsel uzun-kısa dönem hafıza ağı

LSTM derin öğrenme alanında kullanılan tekrarlayan sinir ağı(RNN) mimarisidir (Hochreiter and Schmidhuber 1997). Standart ileri beslemeli sinir ağlarının aksine, LSTM’nin geri bildirim bağlantıları vardır. Görüntü, konuşma ve video gibi tek boyutlu olmayan verileri işleyemez. Girdi olarak öznitelik vektörlerini alan tam bağlantılı LSTM, zamansal özellikleri öğrenmek için kullanılır. Vektörleştirmenin sınırlandırılması, yinleme sırasında mekânsal bağlantı bilgisinin kaybına sebep olur. İnsan hareket tanıma için ise vücut bölümlerinin pozisyon değişimleri önemlidir. Bu nedenle, ConvLSTM (Shi *et al.* 2015), uzun dönemli zamansal-mekânsal öznitelikleri öğrenmek için kullanılmaktadır.

ConvLSTM mimarisinde, model önceki gizli durumu dizinin bir sonraki adımına taşır. Bu nedenle ağın daha önceki verilerini tutar ve bir sonraki adımda karar vermek için kullanır. Yani, ConvLSTM, belirli bir hücrenin gelecekteki durumunu, yakın komşularının girdileriyle ve geçmiş durumlarıyla belirler.

Yöntemde, evrişimli katmanda işlenen görüntü işlenir ve 1 boyutlu özellik vektörü oluşturulur. Bu işlem her görüntü için tekrarlanır ve oluşan özellik vektörleri LSTM katmanına girdi olarak verilir. Girdi vektörü $x_1, \dots, x_t$, hücre durumları (states) $c_1, \dots, c_t$, gizli durumla (hidden states) $h_1, \dots, h_t$ ve i_t, f_t, o_t ConvLSTM'nin kapıları olmak üzere ConvLSTM aşağıdaki gibi formüle edilebilir:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i) \quad (8)$$

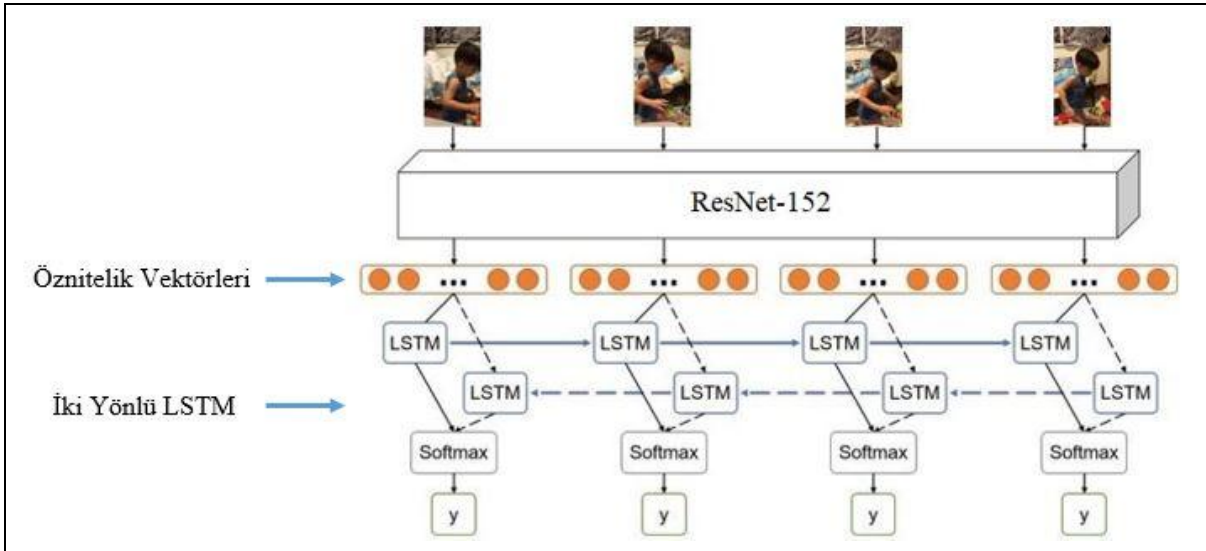
$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f) \quad (9)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o) \quad (10)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (11)$$

$$H_t = o_t \circ \tanh(C_t) \quad (12)$$

Burada "*" evrişim operatörünü, σ sigmoid fonksiyonunu ve $W_x \sim$ ve $W_h \sim$ 2 boyutlu evrişim çekirdeğini temsil etmektedir. Şekil 13'de yöntemin şemasında gösterildiği gibi evrişim katmanında mekânsal özniteliklerin çıkarımı için ResNet-152 kullanılmıştır.



Şekil 13. Evrişimli LSTM (ConvLSTM) yönteminin şeması (Jiang et al. 2018).

Sınıflandırma

Softmax Sınıflandırıcı: Softmax Regresyon, lojistik regresyonu çok sınıflı problemlere genelleştiren bir sınıflandırma yöntemidir (Wolfe *et al.* 2017). Lojistik regresyon bağımlı ikili değişken ile bir veya daha fazla nominal veya oran düzeyi bağımsız değişkenler arasındaki ilişkiyi açıklamak için kullanılır. Softmax algoritması ise, aynı anda birden çok sınıfa ikili lojistik regresyon uygulamaktadır. Softmax, birbirinden farklı olan birden fazla sınıfla kullanılmak üzere tasarlanmıştır. Sınıf üyeliğini diğer sınıflardan bağımsız olarak değerlendirmez. Her bir veri ögesinin, her çıkış sınıfı için etkinleştirme değerlerini hesaplar. Toplamı bir olan bir olasılık kümesi elde etmek için değerleri normalleştirir. En yüksek olasılığa sahip sınıf kazanır ve veri ögesi bu sınıfa atanır (Collobert *et al.* 2002). Denklem (13)'de softmax fonksiyonu verilmiştir.

$$P(y = j | \theta^{(i)}) = \frac{e^{\theta^{(i)}}}{\sum_{j=0}^k e^{\theta_k^{(i)}}} \quad (13)$$

$$\theta = W_0 X_0 + W_1 X_1 + \dots + W_k X_k = \sum_{i=0}^k W_i X_i = W^T X$$

θ girdi vektörü, X eğitilmiş öznitelikler, w ağırlık ve k girdi vektörü sayısı olmak üzere, girdi vektörümüzdeki her değer için Softmax değeri, tek tek girdilerin üstel değerinin tüm girdilerin üstel değerlerinin toplamına bölünmesi ile hesaplanmaktadır. Üstel fonksiyon sayesinde negatif girdiler negatif olmayan değerlere dönüştürülmektedir. Her girdi değeri (0,1) aralığında ve hesaplamada payda hep aynı olduğu için çıktılar birbirleri ile orantılı olacaktır (Mahmood 2018).

DENEYSEL SONUÇLAR VE TARTIŞMA

Önerilen yöntemi değerlendirmek için kendini uyarıcı hareketlerin bulunduğu veri seti oluşturulmuştur. Yöntem 3 aşama için test edilmiştir. İlk aşamada veriler üzerinde önışlem uygulanmadan orijinal verilerle hareket tanıma, ikinci aşamada verilere önışlem uygulandıktan sonra hareket tanıma ve son olarak üçlü histogram yönteminden sonra hareket tanıma yapılmıştır. Her 3 senaryoda, 5 farklı öznitelik çıkarma yöntemi ve 3 farklı sınıflandırıcının performansları değişen kod çizelgesi (codebook) boyutları ile değerlendirilmiştir. Veri setimiz üzerinde yöntem ve yöntemi oluşturan adımların sistemin doğruluğuna olan etkileri analiz edilmiştir.

Değerlendirme Yöntemleri

Çapraz doğrulama

Makine öğrenmesinin tanıma doğruluğunu doğrudan etkileyen önemli faktörlerden biri veri setinin eğitim ve test olarak bölünme kısmıdır. Eğitim verisi makine öğrenme yöntemlerini amaca yönelik eğitir. Test verileri ile eğitilen modelin performansı ölçülebilir. Çalışmalarda genelde %70 eğitim-%30 test verisi ya da %80 eğitim-%20 test verisi olarak ayrılmaktadır. Üzerinde çalışılan problemin türüne ve kapsamına göre bu oranlarda değişiklik yapılabilmektedir.

Çalışmalarda performans değerlendirmelerinin adil ve güvenilir olması için ayırımın doğru bir şekilde yapılması gerekmektedir. Ayrıca, makine öğrenmesi yöntemlerinin aşırı uyum (overfitting) ya da yetersiz öğrenme (underfitting) gibi problemlerin üstesinden gelmesi gerekmektedir. Aşırı uyum; bir modelin eğitim verisindeki detay ve gürültüyü, modelin yeni veriler üzerindeki performansını olumsuz yönde etkilediği ölçüde öğrenme durumudur. Bu, eğitim verilerindeki gürültü veya rastgele dalgalanmaların model tarafından öznitelik olarak algılandığı ve öğrenildiği anlamına gelir. Sonuç olarak modellerin genelleme yetenekleri olumsuz etkilendiği için performans düşer. Yetersiz öğrenme ise hem eğitim verileri ile model oluşturmama hem de test verisi üzerinde genelleme yapamama durumudur.

Hem aşırı uyum hem de yetersiz öğrenme kötü performansa neden olabilir. Ancak, uygulamalı makine öğreniminde açık ara en yaygın sorun aşırı uyum sağlamaktır. Eğitilen makine öğrenme algoritmasının aşırı uyum probleminden kaçınmak için k-kat çapraz doğrulama (k-fold cross validation) tekniği kullanılabilir. Çapraz doğrulama modelin bağımsız verilere nasıl genelleştirildiğini değerlendirmek için bir model doğrulama tekniğidir. Bu

tekniknin amacı, modelin pratikte ne kadar doğru performans gösterdiğini tahmin etmek ve aşırı uyum gibi sorunları sınırlamaktır. Çapraz doğrulamada tüm veri seti rastgele k tane eşit büyüklükteki alt kümeye bölünür. Bu kümelerden her defasında birisi test diğerleri eğitim için kullanılır. Her bir alt küme bir kez test verisi olarak kullanılacak şekilde eğitim k kere tekrarlanır. Sonunda, sistemin performansı elde edilen k tane doğruluk değerinin ortalaması alınarak bulunur.

Bu çalışmada, doğruluk değerinin güvenilir olması ve aşırı uyum probleminin üstesinden gelmek için “birini grubun dışında bırak (Leave-one-Group-out)” senaryosunda 5 kat çapraz doğrulama tekniği kullanılmıştır.

Performans ölçütleri

Makine öğrenmesinde en uygun yöntemin seçilmesi için yöntemin öğrenme performansının hesaplanması gerekmektedir. Eğitilen yöntemin test sonuçları karmaşıklık matrisi (confusion matrix) ile gösterilir. Bir karmaşıklık matrisi Tablo 4’de gösterildiği gibi dört kategoriden oluşur: Doğru pozitifler (TP), pozitif olarak doğru sınıflandırılan durumları ifade eder. Yanlış pozitifler (FP) yanlış pozitif olarak sınıflandırılan negatif örnekleri temsil eder. Doğru negatifler (TN), negatif olarak doğru bir şekilde sınıflandırılan negatif durumları ifade eder. Son olarak, yanlış negatifler (FN) yanlış negatif olarak sınıflandırılan pozitif örnekleri temsil eder.

Tablo 4. Karmaşıklık Matrisi

		<i>Tahmin Edilen Sınıf</i>	
		<i>Pozitif</i>	<i>Negatif</i>
<i>Gerçek Sınıf</i>	<i>Pozitif</i>	TP	FN
	<i>Negatif</i>	FP	TN

Karmaşıklık matrisindeki veriler ile sınıflandırıcıların performansı doğruluk hesaplaması ile belirlenmektedir. Doğruluk tüm test veri kümesi içerisinde doğru sınıflandırılmış pozitif ve negatif örneklerin toplamının tüm veri kümesindeki örnek sayısına bölümüdür. Doğru sınıflandırılmış verilerin yüzdesini verir. Denklem (14)’de karmaşıklık matrisinden elde edilen doğruluk formülü verilmiştir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Bir diğer değerlendirme ölçüsü ise 2009’dan bu yana kullanılan Ağırlıksız Ortalama Hatırlama (UAR)dır. UAR, doğru tahmin edilen örneklerin oranıdır (yani toplam doğru

tahminlerin, toplam örneklere bölümü). Her sınıf için hatırlama değeri Denklem (15)'deki gibi hesaplanıp ortalaması alınarak ağırlıksız doğruluk hesaplanır. Bu iki değerlendirme ölçüsü arasındaki ayırım, veri setinde örneklerin yetersiz ve dengesiz dağıldığı sınıflar varsa yararlıdır. Ağırlıksız doğruluk çok sınıflı problemler için daha uygundur. Klasik doğruluk her sınıfı veri setindeki o sınıfa ait örnek sayısına göre değerlendirirken, ağırlıksız doğruluk her sınıfa aynı ağırlığı verir (Yean 2015)(Schuller *et al.* 2017).

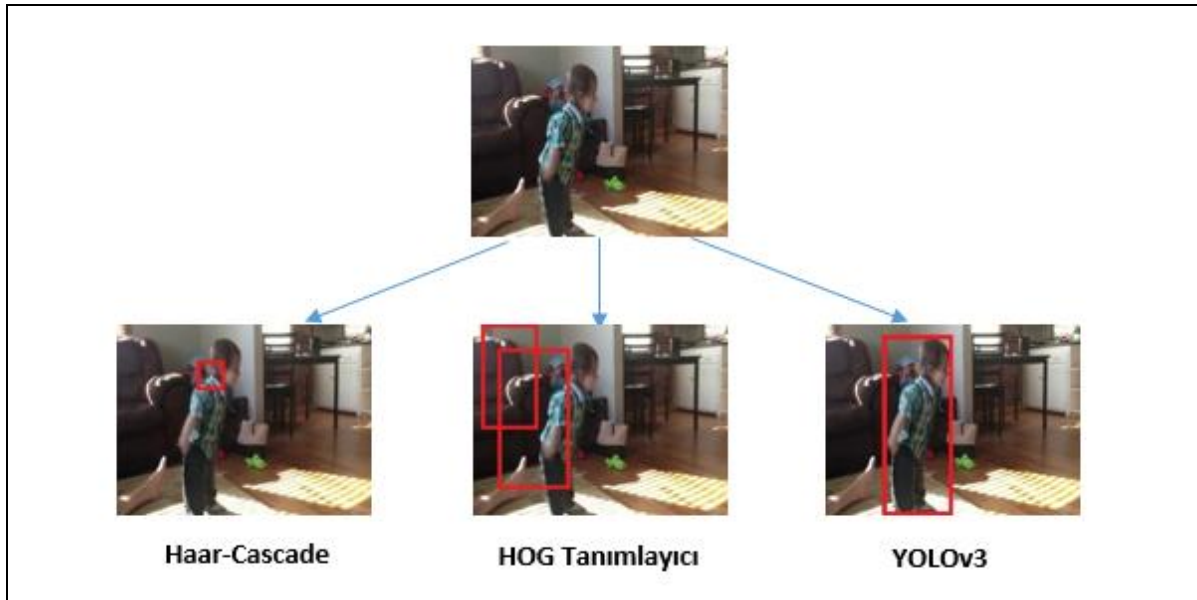
$$\text{Hatırlama (recall)} = \frac{TP}{TP+FN} \quad (15)$$

Sınıf 1 'in hatırlaması (recall) R1, sınıf 2'nin hatırlaması (recall) R2 ise UAR değeri R1 ve R2'nin ortalamasıdır.

Oluşturduğumuz veri setinde kafa sallama hareket sınıfı en az örnek bulunan sınıftır. Tüm hareket sınıflarının video sayısı farklı olduğu için değerlendirmelerde UAR doğruluğu da hesaplanmıştır. Bu tez çalışmasında, yöntemlerin tanıma doğruluğu klasik (ağırlıklı) doğruluk ve UAR ile hesaplanmıştır.

Önişlem aşamasında insan tespiti yöntemlerinin analizi

Çalışmamızda insan tespiti için HOG descriptor, Haar-Cascade ve YOLOv3 yöntemleri denenmiştir. Şekil 14'de uygulanan yöntemlerden elde edilen çıktılar verilmiştir.



Şekil 14. İnsan tespiti yöntemlerinin çıktıları.

Haar-Cascade ve HOG Descriptor yöntemlerinin OpenCV'de ki önceden eğitilmiş modelleri bir kişinin ayakta durma pozisyonunda tespit edilmesi için eğitilmiştir. Bu sebeple yöntemlerin ayakta duran kişinin ön ve arka görünümünden tespit edilebilmekte fakat yan görünümünde tespit gücü zayıflamaktadır. Şekil 15'de gösterilen örnek gibi görüntüde masa

ayağı gibi insan bacağına benzetilen nesneler bulunmakta ise masa ayağını insan olarak tespit etmeleri büyük olasılıktır.



Şekil 15. Haar-Cascade yönteminde hatalı tespit.

En çok karşılaşılan sorunlardan birisi de tespit sınırı hatasıdır. Şekil 14’de Haar-Cascade ve HOG Descriptor tarafından belirlenen tespit alanı, kişinin vücut bölümlerinin tamamını kapsamamaktadır.

Deneyler sonucunda en iyi sonucu YOLOv3 yönteminden elde ettiğimiz için, çalışmamızda insan tespit ve takibinde YOLOv3 ve SORT kullanılmıştır.

Görsel Kelime Çantası Yaklaşımı Sonuçları

Görsel Kelime Çantası yaklaşımı iki kısımda ele alınmıştır. İlk kısmında tüm eğitim video karelerinin %30’una karşılık gelen 10.000 video karesi rastgele seçilir. Daha sonra rastgele seçilen 10.000 eğitim video karesinin öznitelik vektörleri HOG, HOF, HOG-HOF, SIFT ve SURF tanımlayıcı yöntemleri kullanılarak çıkarılır. Hareket sınıflarını en iyi temsil edebilecek anahtar görsel kelimelerin bulunması için K-means kümeleme yöntemi kullanılmıştır. Çalışmamızda farklı sayıda küme merkezi sayıları (anahtar görsel kelime sayısı) ile deneyler yapılmıştır. İkinci kısımda tüm eğitim ve test videolarının öznitelikleri çıkarılıp birinci kısımda oluşturulan anahtar görsel kelimelerden görsel kelime dizileri oluşturulur. Oluşturulan dizilerden frekans histogramları üretilir ve videolar bu histogramlarla temsil edilir. Eğitim videolarının histogramları ile eğitilen GNB, SVM ve MLP sınıflandırıcı modellerinden test videolarının histogramları için sınıf etiket tahminleri üretilir.

Deneylerde kullanılan yöntemlerin parametre değerleri

Bu tez çalışmasında, en iyi sonucu alabilmek için farklı yöntem kombinasyonları denenmiştir. Kullanılan yöntemlerin en uygun parametre değerleri uygulamamızda kullanılmıştır. Öznitelik vektörü çıkarmak için kullanılan bu yöntemlerin parametre değerleri Tablo 5’de verilmiştir. Öznitelik tanımlayıcı yöntemler için OpenCV (Bradski and Kaehler 2008) kütüphanesi kullanılmıştır.

Tablo 5. Öznitelik Çıkarma Yöntemlerinin Parametre Değerleri

Öznitelik	Parametre	Açıklama	Değer
HOG	winSize	Algılama penceresi boyutu	(64,128)
	blockSize	Piksel cinsinden blok boyutu	(16,16)
	blockStride	Adım boyutu	(8,8)
	cellSize	Hücre boyutu	(8,8)
	nbins	Eğim histogramının hesaplanmasında kullanılan kutu sayısı	9
	winSigma	Gauss düzeltme penceresi parametresi	-1
	L2HysThreshold	L2-Hys normalleştirme yöntemi eşik değeri	2.0000000000000001e-01
	gammaCorrection	Gama düzeltme ön işlemenin gerekli olup olmadığı	0
HOF	nlevels	Maksimum algılama penceresi sayısı	64
	orientations	Yönlendirme kutularının sayısı.	9
	pixels_per_cell	Her hücrenin piksel boyutu	(8, 8)
	cells_per_block	Her bloktaki hücre sayısı	(3, 3)
	normalise	İşlemeden önce görüntüyü normalleştirmek	false
SIFT	motion_threshold	Hareketsizlik eşik değeri	1.
	nfeatures	Tutulacak en iyi özelliklerin sayısı	0
	nOctaveLayers	Her oktavdaki katman sayısı	3
	contrastThreshold	Kontrast eşiği	0,04
	edgeThreshold	Kenar benzeri özellikleri filtreleme eşiği	10
SURF	sigma	Giriş görüntüsüne uygulanan Gaussian sigması	1.6
	hessianThreshold	Hessian anahtar nokta dedektörü için eşik değeri	100
	nOctaves	Anahtar nokta dedektörünün kullanacağı piramit oktav sayısı	4
	nOctaveLayers	Her oktavdaki katman sayısı	3
	extended	Genişletilmiş tanımlayıcı bayrağı	false
SURF	upright	Yönlendirme hesaplama	false

Kümeleme işlemi için kullanılan K-means algoritmasının parametre değerleri Tablo 6'da verilmiştir. Uygulamada ScikitLearn (Pedregosa *et al.* 2011) kütüphanesi kullanılmıştır.

Tablo 6. K-means Parametre Değerleri

K-means	Açıklama	Değer
n_clusters	Oluşturulacak küme sayısı	200/400/600/800/1000
init	Başlatma metodu	'k-means++'
n_init	Her küme merkezi için çalıştırılma sayısı	10
max_iter	Tek bir çalışma için maksimum yineleme sayısı	300
tol	Birbirini izleyen iki yinelemenin küme merkezlerindeki farkın nispi toleransı	0.0001
precompute_distances	Kümeler arası mesafelerin önceden hesaplanması	'auto'
verbose	Ayrıntı modu	0
random_state	Küme merkezi tanımlamak için rastgele sayı üretimi	None
copy_x	Verilerin kopyalanması	True
n_jobs	Hesaplama için kullanılacak OpenMP iş parçacığı sayısı	None
Algorithm	Kullanılacak algoritma	'auto'

Deneylerde K-means küme sayısı parametresi 100, 200, 400, 600, 800 veya 1000 olarak ayarlanıp sonuçlar alınmıştır. Farklı küme sayılarının (görsel kelime sayısının) sistem doğruluğundaki etkisi değerlendirilmiştir.

Sınıflandırma aşamasında ayrı ayrı MLP, SVM ve GNB yöntemleri ScikitLearn (Pedregosa *et al.* 2011) kütüphanesi kullanılarak uygulanmıştır. Kullanılan sınıflandırma yöntemlerinin parametre değerleri Tablo 7'de gösterilmiştir.

Tablo 7. Kullanılan Sınıflandırma Algoritmalarının Parametre Değerleri

Sınıflandırıcı	Parametre	Açıklama	Değer
MLP	solver	Ağırlık optimizasyonu için çözücü	'lbfgs'
	alpha	L2 ceza (düzenleme süresi) parametresi	0.13
	hidden_layer_size	Gizli katmandaki nöron sayısı	(200,50)
	max_iter	Maksimum yineleme sayısı	10000
	learning_rate	Ağırlık güncellemeleri için öğrenme oranı programı	'adaptive'
GNB	var_smoothing	Hesaplama kararlılığı için varyanslara eklenen tüm özelliklerin en büyük varyansının miktarı	1e-9
SVM	decision_function_shape	Karar fonksiyonu	'ovo'
	C	Düzenleme parametresi	1,0
	kernel	Algoritmada kullanılacak çekirdek türü	rbf
	tol	Kriteri durdurma toleransı.	0,001
	gamma	Çekirdek katsayısı	scale

Gizli düğümünde yetersiz veya aşırı nöron bulunan MLP sınıflandırıcı modeli kötü genelleme ve aşırı uyum sorunlarına neden olabilmektedir. Gizli katmandaki nöron sayısını belirlemek için analitik bir yöntem yoktur (Subasi and Ercelebi 2005). Bu nedenle, çalışmada en uygun parametre değerleri deneme yanılma yolu ile bulunmuştur. Tüm deneylerde aynı parametre değerleri kullanılmıştır.

Önişlem uygulanmamış veri seti ile alınan sonuçlar

Veri setindeki video görüntüleri üzerine hiçbir önişlem uygulanmadan orijinal veriler ile hareket tanıma deneyleri yapılmıştır. Bunun amacı çalışmada önerilen önişlem aşamalarının tanıma doğruluğuna olan katkısını gözlemlemektir. Tablo 8’de BoVW yaklaşımı ile orijinal veri setindeki kendini uyarıcı hareketlerin tanıma doğruluğu verilmiştir.

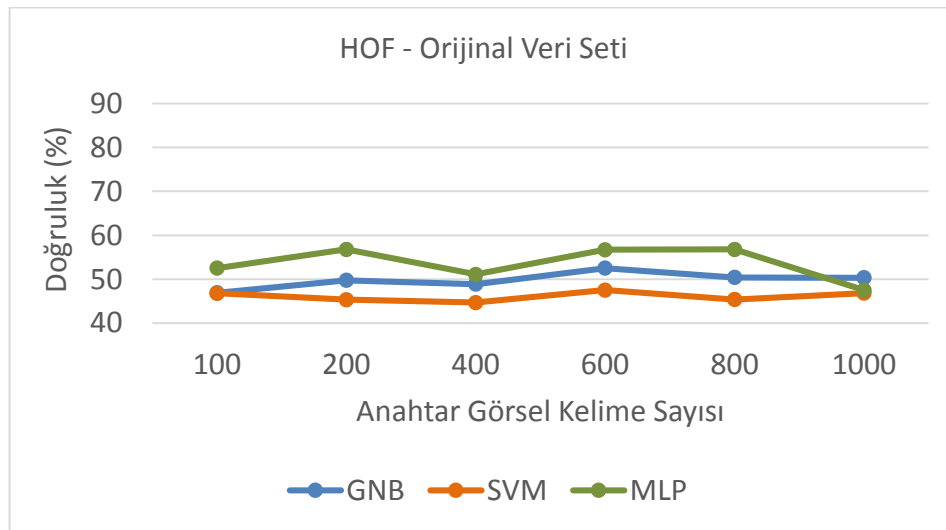
Tablo 8. Orijinal Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre Doğruluk (%) Değerleri

Orijinal Veri Seti		Anahtar Görsel Kelime Sayısı					
Yöntem	Sınıflandırıcı	100	200	400	600	800	1000
HOG	GNB	46,89	47,52	43,29	47,63	42,63	43,27
	SVM	41,81	40,41	40,41	39,01	38,98	41,86
	MLP	42,12	41,86	43,32	49,82	43,29	44,03
HOF	GNB	51,08	49,75	48,89	52,51	50,39	50,32
	SVM	45,34	45,34	44,67	47,51	45,36	46,84
	MLP	52,44	56,77	51,10	56,69	56,75	47,53
HOG/HOF	GNB	46,89	49,69	46,91	51,10	53,96	54,65
	SVM	46,79	46,07	44,65	46,76	46,05	45,41
	MLP	52,50	51,82	49,65	50,38	46,81	53,19
SIFT	GNB	51,08	48,24	38,29	33,32	37,60	43,29
	SVM	41,81	40,36	38,32	37,58	36,86	36,17
	MLP	46,12	47,50	46,87	43,34	44,11	41,25
SURF	GNB	33,24	41,12	39,69	39,01	38,27	41,12
	SVM	42,59	41,10	41,10	41,10	41,81	41,82
	MLP	38,27	40,39	45,36	40,43	43,98	41,84

Tablo 9. Orijinal Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre UAR Doğruluk (%) Değerleri

Orijinal Görüntü (UAR)		Anahtar Görsel Kelime Sayısı					
Yöntem	Sınıflandırıcı	100	200	400	600	800	1000
HOG	GNB	45,47	46,44	41,95	45,78	40,28	41,95
	SVM	35,39	34,26	34,40	32,78	33,11	35,49
	MLP	43,10	41,03	41,85	47,50	42,44	42,72
HOF	GNB	50,05	48,21	47,42	50,60	48,46	48,71
	SVM	38,50	38,49	37,96	40,76	38,75	40,16
	MLP	50,52	55,05	49,25	54,59	54,03	45,68
HOG/HOF	GNB	45,66	48,19	46,38	49,98	51,19	53,79
	SVM	39,79	39,50	37,91	40,01	39,39	38,49
	MLP	49,75	49,75	47,72	47,74	44,52	50,68
SIFT	GNB	48,94	45,48	35,03	30,95	35,82	41,10
	SVM	36,35	34,48	32,47	31,80	31,17	30,46
	MLP	43,84	46,57	45,01	41,12	41,92	39,22
SURF	GNB	31,86	40,44	37,90	36,98	37,70	38,76
	SVM	36,04	34,58	34,86	34,85	35,76	35,69
	MLP	36,81	39,00	44,32	40,65	42,62	40,61

Tabloyu incelediğimizde, önişlemsiz veri seti ile yöntemler arasında genel anlamda en düşük sonuçlar SIFT ve SURF yöntemlerindedir. Düşük sonuçlar vermeleri çıkardıkları özneteliklerin diğer yöntemlere göre az sayıda olmasıdır. Deney sonuçlarında görüldüğü gibi alınan en iyi sonuç %56,77'dir. Bu sonuç HOF yönteminin 200 anahtar görsel kelime sayısı ve MLP sınıflandırıcısı ile alınmıştır. HOF yönteminde en iyi sonucun alınması çıkarılan öznetelik sayısının fazla olması ve videolardaki hareketin yakalanmasında iyi olmasıdır. Orijinal veri seti üzerinde HOF tanımlayıcısı ile yapılan deneylerin ayrıntılı grafiği Şekil 16'da verilmiştir.



Şekil 16. Orijinal veri setinde anahtar görsel kelime sayılarına göre sınıflandırıcıların değerlendirme grafiği.

Önişlem uygulanmış veri seti ile alınan sonuçlar

Veri setinde ki videolar ebeveynler tarafından herhangi bir kamera ile ev ortamlarında kaydedilmişlerdir. Doğal ev ortamında kaydedilmeleri arka plan karmaşıklığı, farklı nesne ve kişilerin görüntüye dâhil olması gibi problemlere sebep olmaktadır. Ayrıca bir kişi ve bir nesneyle etkileşimi gerektirmeyen hareketlerin tanınması için asıl kişi ve pozu yeterlidir. Önişlemler ile görüntülerdeki asıl kişi ve poza odaklanılmıştır. Önişlem uygulandıktan sonra alınan doğruluk değerleri Tablo 10’da, UAR değerleri Tablo 11’de verilmiştir.

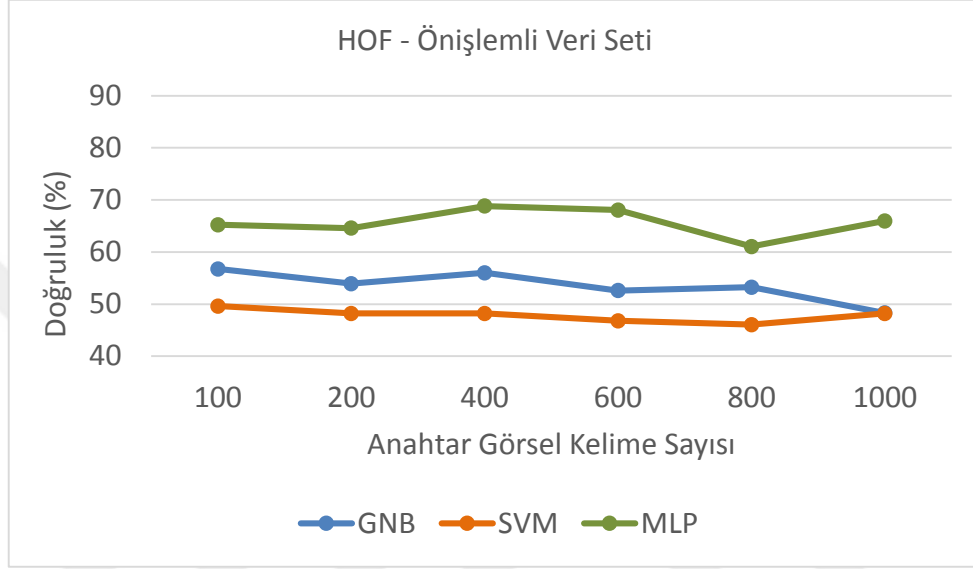
Tablo 10. Önişlem Uygulanmış Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre Doğruluk (%) Değerleri

Önişlemli Veri Seti		Anahtar Görsel Kelime Sayısı					
Yöntem	Sınıflandırıcı	100	200	400	600	800	1000
HOG	GNB	41,82	43,27	46,77	47,53	53,88	49,67
	SVM	41,13	44,67	41,15	43,91	41,84	43,96
	MLP	53,20	50,44	51,08	46,84	48,24	48,24
HOF	GNB	56,79	53,94	56,03	52,61	53,25	48,30
	SVM	48,22	48,22	48,22	46,79	46,08	48,22
	MLP	64,22	64,60	68,83	68,05	61,05	65,96
HOG/HOF	GNB	56,77	59,67	58,15	56,06	51,79	54,63
	SVM	49,63	48,24	48,22	48,22	48,22	49,65
	MLP	65,24	64,58	67,45	67,43	63,15	69,50
SIFT	GNB	43,97	44,01	37,56	35,49	36,25	39,77
	SVM	41,87	39,75	41,15	39,03	39,03	39,03
	MLP	46,94	41,25	47,58	45,46	50,42	54,00
SURF	GNB	37,51	36,94	34,75	31,94	30,49	35,51
	SVM	39,03	37,58	34,77	33,35	34,04	33,35
	MLP	38,27	42,53	38,37	41,18	42,58	42,58

Tablo 11. Önişlem Uygulanmış Veri Seti İle Yöntem, Sınıflandırıcı ve Anahtar Görsel Kelime Sayısına Göre UAR Doğruluk (%) Değerleri

Önişlemli Veri Seti (UAR)		Anahtar Görsel Kelime Sayısı					
Yöntem	Sınıflandırıcı	100	200	400	600	800	1000
HOG	GNB	40,78	41,35	46,65	47,64	53,68	49,39
	SVM	35,02	38,36	34,63	37,72	36,26	37,75
	MLP	53,05	48,70	49,48	49,45	47,48	48,67
HOF	GNB	52,80	51,06	52,70	50,25	51,00	45,55
	SVM	41,75	41,72	41,93	40,06	39,43	41,93
	MLP	63,17	64,82	67,62	68,46	60,05	65,63
HOG/HOF	GNB	52,49	57,18	55,06	54,30	48,94	52,66
	SVM	43,26	41,72	41,51	41,72	41,93	43,39
	MLP	64,23	64,85	67,53	66,37	61,81	68,88
SIFT	GNB	39,61	39,93	35,50	32,88	33,54	37,59
	SVM	35,50	33,70	35,50	33,42	33,42	33,42
	MLP	45,20	39,44	44,88	42,57	46,70	50,36
SURF	GNB	35,20	34,46	32,98	29,78	28,21	32,71
	SVM	32,71	31,56	29,06	27,88	28,58	27,88
	MLP	34,61	39,70	34,76	36,60	39,15	37,83

Önişlem öncesi değerler ile önişlem sonrası değerler karşılaştırıldığında ortalama %3 oranında artışlar olmuştur. Önişlemsiz veriler ile aldığımız en iyi sonuç olan %56,77, önişlemden sonra %64,60'a yükselmiştir. HOF öznitelik çıkarıcısı 200 anahtar görsel kelime sayısı ile MLP sınıflandırıcı yönteminde %7,83 oranında iyileşme olmuştur. Öznitelikler çıkarılırken görüntüdeki alakasız nesne ve kişi olmasının verdiği karmaşıklık ve gürültüyü en aza indirerek asıl kişiye ve pozuna odaklanmak sonucu iyileştirmiştir. Önişlem uygulanmış veri seti üzerinde HOF tanımlayıcısı ile yapılan deneylerin ayrıntılı grafiği Şekil 17'de verilmiştir.



Şekil 17. Önişlemli veri setinde anahtar görsel kelime sayılarına göre sınıflandırıcıların değerlendirme grafiği.

Önişlem uygulanmış verilerin üçlü histogram temsili sonrası sonuçları

İnsan hareket tanıma için vücut bölümlerinin ardışık zamanlardaki pozisyon değişimleri önemlidir. Pozisyon değişimleri zamansal bilgilerle elde edilir. Zaman bilgisinin uzun süre eksik olması hareketlerin modellenmesini engellemektedir. Zaman bilgisi kaybını kısmen önlemek için önişlem uygulanmış videonun anahtar görsel kelime dizisini iki geçici bölüme ayrılarak her dizi için histogram oluşturulmuştur. Video, oluşan iki geçici bölümün histogramları ve tüm dizinin histogramı olmak üzere üç histogramla temsil edilmiştir. Önişlem ve üçlü histogram temsilinden sonra elde edilen doğruluk sonuçları Tablo 12'de gösterilmiştir.

Tablo 12. Üçlü Histogram Temsili Sonrası Doğruluk (%) Değerleri

Üçlü Histogram		Anahtar Görsel Kelime Sayısı					
Yöntem	Sınıflandırıcı	100	200	400	600	800	1000
HOG	GNB	41,25	48,94	44,03	45,36	47,56	53,89
	SVM	41,15	40,46	41,82	41,18	40,44	39,72
	MLP	50,68	52,58	55,44	52,46	53,94	49,65
HOF	GNB	52,55	60,00	60,34	55,39	53,89	60,27
	SVM	48,91	50,71	47,51	46,79	47,51	47,51
	MLP	67,45	80,00	70,22	66,72	73,84	68,86
HOG/HOF	GNB	57,80	57,46	54,60	58,94	53,79	52,53
	SVM	47,73	47,44	47,51	45,39	48,94	47,51
	MLP	61,15	60,27	61,82	61,01	63,17	65,24
SIFT	GNB	41,20	43,30	33,27	34,75	33,37	37,58
	SVM	42,58	43,30	39,72	39,75	39,75	39,72
	MLP	40,53	48,37	48,96	49,09	43,37	48,30
SURF	GNB	28,82	29,09	31,25	30,51	31,92	31,97
	SVM	35,61	34,80	33,37	34,08	34,77	33,35
	MLP	39,84	43,25	41,89	39,08	35,56	39,03

Tablo 13. Üçlü Histogram Temsili Sonrası UAR Doğruluk (%) Değerleri

Üçlü Histogram (UAR)		Anahtar Görsel Kelime Sayısı					
Yöntem	Sınıflandırıcı	100	200	400	600	800	1000
HOG	GNB	39,41	48,19	44,10	44,79	47,85	52,79
	SVM	35,10	34,32	35,41	35,00	34,05	33,47
	MLP	50,15	51,34	55,23	52,78	54,86	50,64
HOF	GNB	49,59	57,16	58,03	52,98	51,90	57,58
	SVM	42,34	42,87	41,10	40,33	41,17	40,89
	MLP	66,19	79,36	68,45	65,80	74,98	68,28
HOG/HOF	GNB	54,00	54,49	51,87	57,23	51,30	51,00
	SVM	42,15	42,55	41,81	39,08	42,55	40,68
	MLP	60,17	60,99	61,61	60,23	62,28	63,29
SIFT	GNB	37,42	39,13	30,63	31,06	30,83	35,47
	SVM	36,33	37,38	34,04	34,25	34,25	34,13
	MLP	41,44	47,05	47,17	46,11	42,17	45,63
SURF	GNB	26,60	26,56	28,68	28,33	29,55	29,37
	SVM	34,04	28,98	27,80	28,35	29,06	27,88
	MLP	38,18	39,43	39,06	35,47	30,80	35,44

Tablo 12’de görüldüğü üzere 200 anahtar görsel kelime sayısında HOF tanımlayıcısı ve MLP sınıflandırıcısı %80 doğruluk değeri ile en iyi sonucu vermiştir. Sadece önışlem uygulanmış verilerle aynı yöntem ve parametrelerle yapılan deneyde doğruluk değeri %64,60 iken üçlü histogram temsilinden sonra %15,40 artmıştır. Artış oranının yüksek olması hareket tanıma işlemlerinde zamansal bilgilerin ne kadar önemli olduğunu göstermektedir.

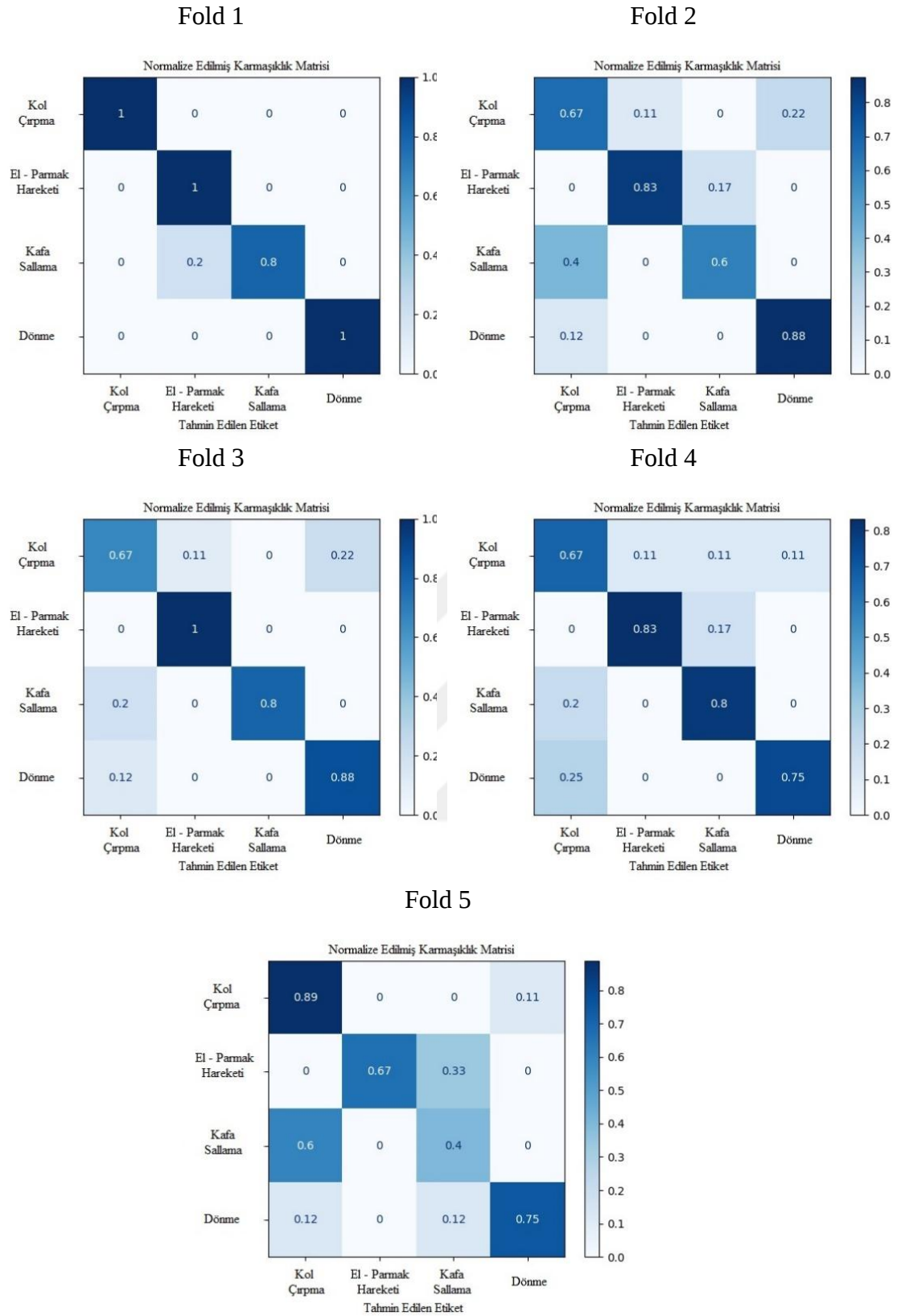
Görsel Kelime Çantası yaklaşımında sistemin doğruluğunu etkileyen parametrelerden biri anahtar görsel kelime sayısıdır. Anahtar görsel kelime sayısının incelenen harekete göre en uygun değerini bulmak hareketin yeterli temsil edilmesi için önemlidir. Gereğinden az seçilmesi hareketlerin yetersiz temsiline sebep olurken, fazla seçilmesi gereksiz karmaşıklığa sebep olmaktadır. Deneylerin birçoğunda görüldüğü gibi anahtar görsel kelime sayısının yüksek olması hareketin en iyi şekilde temsil edileceği anlamına gelmemektedir. Çünkü analiz edilen hareketler basmakalıp ve tekrarlı olduğu için aynı pozların birden fazla görünmesi olasıdır. Alınan sonuçlar incelendiğinde az sayıda anahtar görsel kelime, tekrarlı olan hareketleri temsil edebilmektedir. 100 görsel kelime, yeterli olmayan hareket temsiline sebep olduğu için başarısı düşüktür.

İki türlü tanımlayıcı vardır. Bazıları görünümü yakalarken bazıları da hareketi yakalamaktadır. HOF hareketi yakalayan tanımlayıcılardan biridir. Ardışık iki video karesindeki nesne hareketini tanımlamaktadır. En belirgin avantajı, değişken aydınlatma koşullarında ve dağınıklığıdaki hareketi doğru tahmin etme yükünün tamamen optik akış hesaplamasıyla ilgili olmasıdır. HOF’un hareketi yakalaması ve diğer yöntemlere göre çok sayıda öznitelik çıkarması hareket tanımada başarılı olmasını sağlamıştır.

Tablo 14’de 5 foldun ayrı ayrı doğruluk değerleri ve Şekil 18’de karmaşıklık matrisleri verilmiştir.

Tablo 14. En İyi Sonuç Alınan Deneyin 5 Fold İçin Doğruluk Değerleri

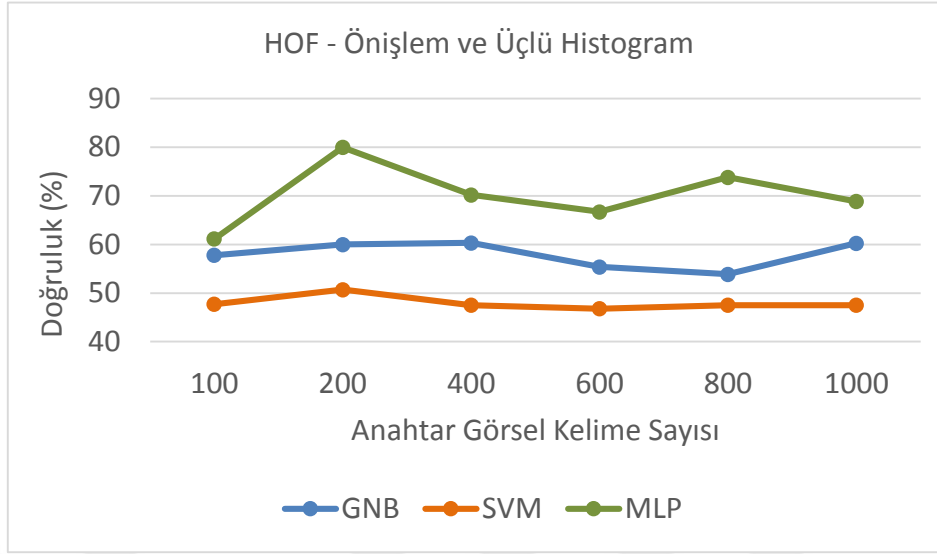
	Doğruluk (%)	UAR (%)
Fold 1	96,42	95,00
Fold 2	75,00	74,37
Fold 3	82,14	83,54
Fold 4	75,00	76,25
Fold 5	71,44	67,63
Ortalama	80,00	79,36



Şekil 18. En iyi sonuç alınan deneyin 5 fold için karmaşıklık matrisleri.

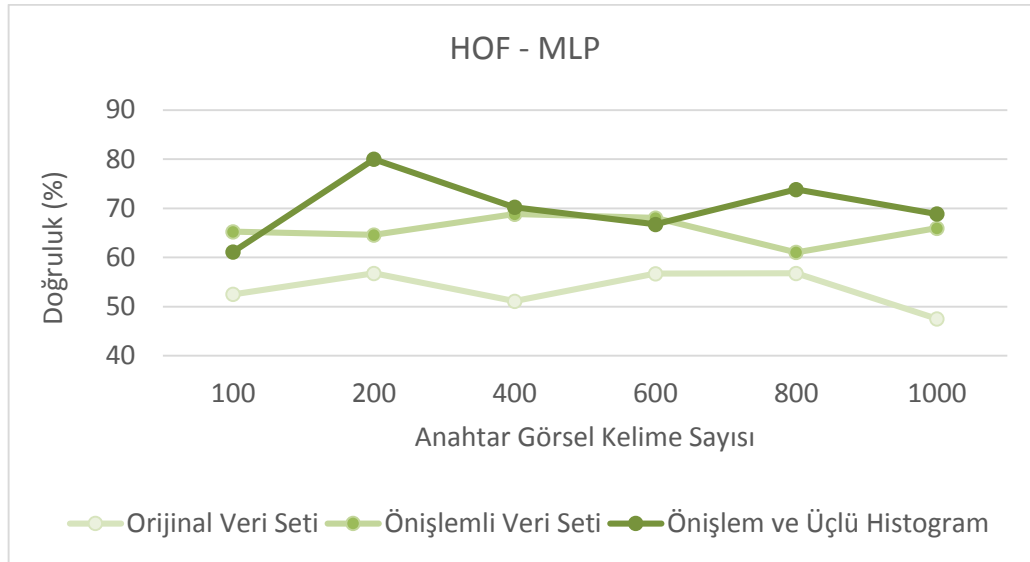
Karmaşıklık matrisleri incelendiğinde hatalı tahmin edilen hareket sınıfları vardır. İncelen hareketler vücudün farklı bölümleri ile gerçekleştirilen hareketlerdir. Bu yüzden çocuk “dönme” ve ya “kafa sallama” hareketini yaparken aynı zamanda en çok kullanılan vücut parçası olan kollarını da hareket ettirebildiği için sistem hatalı tahmin yapmıştır.

Önişlem uygulanmış videoların üçlü histogram temsili ile yapılan deneylerde elde edilen sonuçların ayrıntılı grafiği Şekil 19’da verilmiştir.



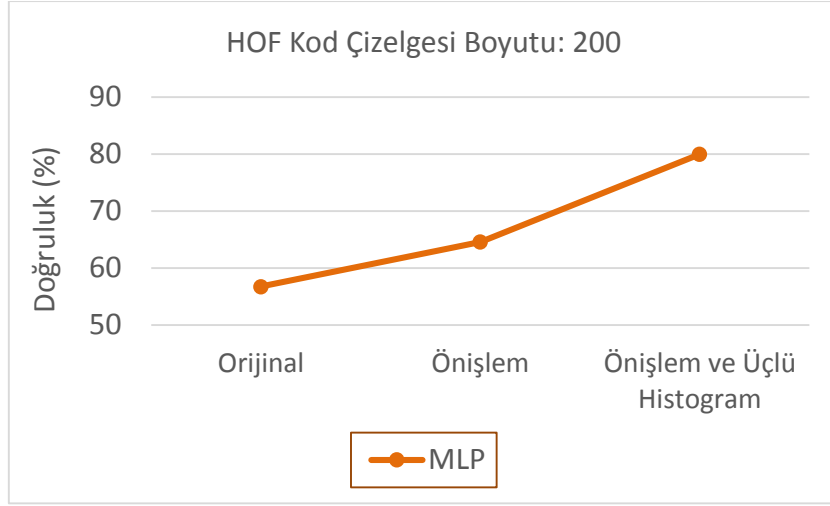
Şekil 19. Önişlemli veri setinde üçlü histogram temsili ile anahtar görsel kelime sayılarına göre sınıflandırıcıların değerlendirme grafiği.

Şekil 19’da görüldüğü gibi en yüksek doğruluk %80 değeri ile 200 görsel kelime sayısındadır. Grafiğe genel olarak bakıldığında her bir görsel kelime sayısı için MLP sınıflandırıcısı en iyi sonuçları vermiştir. Anahtar görsel kelime sayısına göre üç aşamanın doğruluk değerlerindeki değişim grafiği Şekil 20’de gösterilmiştir.



Şekil 20. Anahtar görsel kelime sayılarına göre aşamaların doğruluk değerleri.

Şekil 20’deki grafiğe bakıldığında önişlem ve üçlü histogram temsil yaklaşımının uygulandığı deneyler genel anlamda daha iyi sonuç vermiştir. 200 anahtar görsel kelime sayısında en yüksek sonuç alınmıştır.



Şekil 21. Aşamalara göre doğruluk analizi (HOF-200 Anahtar Görsel Kelime-MLP).

Sonuç olarak, yöntemlerin aşama aşama farklı kombinasyonlarla deneyleri yapıldıktan sonra HOF-200 Anahtar Görsel Kelime-MLP yöntem kombinasyonunun en iyi sonuç verdiği anlaşılmıştır. Üç aşama için bu kombinasyonunun başarısındaki artış Şekil 21'deki grafikte gösterilmiştir. Orijinal veri seti %56,77 sonuç verirken, önerilen önişlemler ve üçlü histogram yöntemleri sonucu %80'e çıkararak %23 iyileşme sağlamıştır.

Yöntemlerdeki işlem sürelerinin analizi

Deneylerimizde iki ana bilgisayar kullanılmıştır. Kullanılan bilgisayarların özellikleri Tablo 15'de verilmiştir. HOG, HOF, HOG-HOF yöntemleri Bilgisayar-1 ile SIFT ve SURF yöntemleri Bilgisayar-2 ile çalıştırılmıştır.

Tablo 15. Uygulamada Kullanılan Bilgisayarların Özellikleri

Özellik	Bilgisayar-1	Bilgisayar-2
İşlemci	Intel Core i7-8700T 2.40 GHz	AMD FX-8350 4.00 GHz
Bellek	16,0 GB	16,0 GB
Ekran Kartı	Intel UHD Graphics 630	NVIDIA GeForce GTX 970
İşletim Sistemi	Windows 10 Pro	Windows 10 Enterprise

Görsel Kelime Çantası yaklaşımındaki uygulama aşamalarının öznitelik çıkarma yöntemlerine göre işlem süreleri Tablo 16'da verilmiştir. Tablodaki veriler önişlem ve üçlü histogram temsili sonrasında en iyi sonuç alınan 200 anahtar görsel kelime sayısındaki bir fold için harcanan süreleri göstermektedir. Bu değerlere verilerin okunma süresi dâhil edilmemiştir.

Tablo 16. Yöntemlerin İşlem Süreleri (saniye)

İşlem (Saniye)	HOG	HOF	HOG-HOF	SIFT	SURF
Eğitim Öznitelik Çıkarma Süresi	79,940	442,684	475,475	1 050,240	684,772
Test Öznitelik Çıkarma Süresi	19,782	112,763	132,768	274,287	161,132
K-Means Uygulama Süresi	327,417	1 793,767	2 096,359	26,495	20,516
Eğitim Histogram Çıkarma Süresi	1,915	19,097	20,905	12,202	11,412
Test Histogram Çıkarma Süresi	0,290	1,425	1,549	0,154	0,204
GNB Uygulama Süresi	0,002	0,011	0,012	0,050	0,047
GNB Tahmin Süresi	0,000	0,000	0,004	0,386	0,317
SVM Uygulama Süresi	0,013	0,018	0,013	0,015	0,016
SVM Tahmin Süresi	0,001	0,001	0,001	0,002	0,002
MLP Uygulama Süresi	2,504	4,010	2,778	4,514	4,777
MLP Tahmin Süresi	0,000	0,000	0,003	0,001	0,000
Toplam Geçen Süre	431,864 (~7 dakika)	2 373,776 (~40 dakika)	2 729,867 (~45 dakika)	1 368,346 (~23 dakika)	883,195 (~15 dakika)

Öznitelik çıkarma aşamasının süreleri kullanılan yöntemin çıkardığı öznitelik vektörü boyutuna göre değişmektedir. Tablo 16’da görüldüğü gibi HOG yöntemiyle eğitim verilerinin özniteliklerinin çıkarma süresi yaklaşık 80 saniye iken HOF yönteminde yaklaşık 443 saniyedir. Aradaki yaklaşık 5,5 katlık fark çıkarılan öznitelik boyutları farkından kaynaklanmaktadır. SIFT ve SURF yöntemleriyle çıkarılan öznitelik vektör boyutlarının daha az olmasına karşın fazla zaman almaları kullanılan bilgisayarın işlemcisinin diğer bilgisayara göre düşük olmasından kaynaklanmaktadır. K-means yöntemi, hesaplama karmaşıklığından dolayı en çok zaman alan aşamadır. Toplam harcanan sürenin yaklaşık %75’i K-means kümeleme yönteminin uygulama süresidir. Sınıflandırma aşaması sistemin en az zaman alan kısmıdır. Sınıflandırıcılar arasında MLP çok katmanlı bir ağ yapısına sahip olduğu için diğer sınıflandırıcılara göre daha çok zaman almıştır. Sonuç olarak toplam süreler çıkarılan öznitelik vektörleri boyutuyla doğru orantılı olarak artmaktadır.

Derin Öğrenme Yöntemlerinin Sonuçları

Ağ parametrelerinin etkisini ölçmek için deneyler yapılarak derin öğrenme yöntemlerinin en uygun parametre değerleri bulunmuştur. Bu değerler Tablo 17’de verilmiştir.

Tablo 17. Derin öğrenme yöntemlerinin parametre değerleri

	ConvLSTM	3DCNN
Batch size (Yığın boyutu)	16	8
Epoch (Eğitim tekrarlarının sayısı)	100	100
Sequence length (her dizideki kare sayısı)	10	-
Loss function (Kayıp fonksiyonu)	CrossEntropyLoss	categorical_crossentropy
Optimizer (İyileştirici)	Adam	SGD
Evaluation metric (değerlendirme ölçütü)	accuracy	accuracy

Derin öğrenme yöntemlerinin orijinal veri seti ile yapılan deneylerdeki doğruluk değerleri Tablo 18’de, önışlem uygulanmış veri seti ile yapılan deneylerin sonuçları ise Tablo 19’da verilmiştir. Yöntemleri uygulamak için iki Python derin öğrenme kütüphanesi kullanılmıştır. 3DCNN için Keras (Gulli and Pal 2017) ve ConvLSTM için Torch (Collobert 2002) kütüphaneleri ile uygulamalar çalıştırılmıştır.

Tablo 18. Orijinal Veri Seti İle Doğruluk (%) Değerleri

Orijinal veri	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Ortalama
ConvLSTM	78,12	68,75	71,88	65,62	56,25	68,12
3DCNN	33,33	33,33	33,33	33,33	25,00	31,66

Tablo 19. Önışlem Uygulanmış Veri Seti İle Doğruluk (%) Değerleri

Önışlemli veri	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Ortalama
ConvLSTM	78,12	78,12	78,12	63,54	67,30	73,04
3DCNN	37,50	37,50	33,33	37,50	29,17	35,00

Sonuçlar incelendiğinde ConvLSTM’nin 3DCNN’e göre çok daha iyi sonuç verdiği görülmektedir. LSTM ağı önceki adımlarındaki çıktıyı hafızada tutup gerektiğinde kullanabilmesinden dolayı performansı daha iyidir. Ayrıca aşırı uyumu (overfitting) engellemek için dropout değerleri 0,5 alınmıştır.

Tablolar incelendiğinde önışlem uygulanmış verilerle yapılan deneylerde iki yöntemin de sonuçları iyileşmiştir. ConvLSTM doğruluğu yaklaşık %5 oranında artarken, 3DCNN %3 oranında yükselmiştir.

Tablo 20. Derin Öğrenme Yöntemlerinin İşlem Süreleri

Yöntem	Geçen Süre
ConvLSTM	40 dakika
3DCNN	325 dakika (~5,5 saat)

Derin Öğrenme yöntemlerinin işlem süreleri Tablo 20’de verilmiştir. ConvLSTM Bilgisayar-2 de, 3DCNN Bilgisayar-1 de uygulanmıştır. ConvLSTM Bilgisayar-2 üzerindeki ekran kartı grafik işlemci birimi (GPU) ile gerçekleştirildiği için çok daha hızlı bir şekilde tamamlanmıştır.



SONUÇLAR VE ÖNERİLER

Otizm Spektrum Bozukluğu (OSB) sosyal iletişim ve etkileşim eksikliğiyle beraber bir takım sınırlı ve tekrarlı davranışlarla karakterize edilmiştir. Yaygınlığı gün geçtikçe artmaktadır. OSB'nin tanı değerlendirme aşamaları hala uzun ve karmaşıktır. Tanı genelde davranışsal gözlemlere dayanarak konulduğu için uzun süreçler gerektirmektedir. Bu uzun süreçler ise erken tanı ve müdahalenin önemli olduğu OSB'nin tanısını geciktirmektedir. Bu yüzden davranışların değerlendirme aşamasında hızlı ve doğru sonuç veren tanı araçlarına ihtiyaç duyulmaktadır. Kısa zamanda analizlerin gerçekleştirilmesi için son zamanlarda makine öğrenmesi ve bilgisayarlı görü yöntemleri ile video tabanlı analizler yapılmaktadır.

Bu tez çalışması kapsamında, OSB'nin önemli tanı ölçütlerinden olan kendini uyarıcı hareketlerin tanınması için Görsel Kelime Çantası yaklaşımı önerilmiştir. Farklı anahtar görsel kelime sayısı (100, 200, 400, 600, 800, 1000) ile 5 (beş) öznitelik tanımlayıcı ve 3 (üç) sınıflandırıcı yöntemi kullanılarak deneyler yapılmış ve en iyi sonuç 200 görsel kelime, HOF tanımlayıcı ve MLP sınıflandırıcısında %80 doğruluk ile alınmıştır. Yani videolardaki hareketler %80 oranında doğru sınıflandırılmıştır. Bu çerçevede veri setindeki videoların arka plan karmaşıklığını giderip asıl çocuğa ve onun hareketine odaklanmayı sağlayacak önışlemler yapıp başarımlar yaklaşık %8 oranında artırılmıştır. Görsel Kelime Çantası yaklaşımındaki videoların histogramla temsili aşamasında Geçici Kelime Torbası (TBoW) modeli önerilerek zamansal bilginin kaybı kısmen önlenmiştir (Shukla *et al.* 2013). Önerilen Geçici Kelime Torbası (TBoW) modelinin tanıma doğruluğunu %15 oranında iyileştirdiği gözlemlenmiştir. Orijinal verilerle yapılan deney sonucuna göre (%56,77) önerilen önışlemler ve TBoW modeli sistemin doğruluğunu toplamda %23 oranında artırarak %80 doğruluk sonucuna ulaşılmıştır. Görsel kelime sayısının doğru seçilmesi önemlidir. Gereğinden az seçilmesi hareketlerin yetersiz temsiline sebep olurken, fazla seçilmesi gereksiz karmaşıklığa sebep olmaktadır. Çalışmamızda genel anlamda az sayıda görsel kelime ile yapılan deneylerin iyi sonuç vermesinin sebebi incelenen hareketlerin basmakalıp ve tekrarlı olmasıdır.

Bu çalışmada, hareket tespiti ve tanımda kullanımı yaygınlaşan Derin Öğrenme yöntemlerinden bazıları uygulanarak sonuçlar değerlendirilmiştir. Uygulanan derin öğrenme yöntemleri videodaki hem zamansal hem mekânsal öznitelikleri çıkarabilen 3DCNN ve ConvLSTM yöntemleridir. ConvLSTM daha iyi sonuç vermiştir. Orijinal verilerle %68,12 doğruluk gözlenirken, önışlem uygulanmış verilerde yaklaşık %5 oranında artarak %73,04 doğruluk oranına ulaşmıştır.

Bu yöntemler bu çalışma kapsamında ebeveynler tarafından ev ortamında gelişigüzel çekilen videolardan oluşturulan GKUDV adı verilen yeni veri seti üzerinde uygulanmıştır. Elde edilen sonuçlar incelendiğinde önceden aynı amaçla ve aynı yolla oluşturulmuş SSBD veri seti ile yapılan çalışmalara göre daha iyi sonuçlar alınmıştır (Rajagopalan *et al.* 2013; Rajagopalan and Goecke 2014).

Gelecek çalışmalarda öncelikle oluşturulan veri setinde bulunan videolar üzerinde uzman görüşü alınarak değerlendirmeler yapılacaktır. Ayrıca veri setinin hareket sınıfları ve her hareket sınıfındaki video sayısı artırılarak genişletmeyi ve ileride ses, beyin görüntüleri gibi veriler ile multimodal çalışmalar yapmayı planlamaktayız.



KAYNAKLAR

- Adrien, J.L., Fauer M., Perrot A., Hameury L., Garreau B., Barthelemy C. and Sauvage D., 1991. Autism and family home movies: Preliminary findings. *Journal of Autism and Developmental Disorders*, 21, 43-49.
- Adrien, J.L., Barthelemy C., Perrot A., Roux S., Lenoir P., Hameury L. and Sauvage D., 1992. Validity and reliability of the infant Behavioral Summarized Evaluation (IBSE): A rating scale for the assessment of young children with autism and developmental disorders. *Journal of Autism and Developmental Disorders*, 22, 375-394.
- Adrien, J.L., Lenoir P., Matineau J., Perrot A., Hameury L., Larmande C. and Sauvage D., 1993. Blind Ratings of Early Symptoms of Autism Based upon Family Home Movies. *Journal of the American Academy of Child & Adolescent Psychiatry*, 32, 617-626.
- Agahian, S., 2018. Uzay-Zaman Poz Çantası ile 3B İnsan Hareketlerinden Eylem Tanıma. Doktora Tezi, Fen Bilimleri Enstitüsü, Karadeniz Teknik Üniversitesi, Trabzon.
- Aggarwal, J.K. and Ryoo M.S., 2011. Human activity analysis: A review. *ACM Computing Surveys*, 43 (3), article 16.
- Anonymous, 2004. Erken Teşhis ve Tedavi. <https://www.otizmvakfi.org.tr/erken-teshis-ve-tedavi/19> (27.03.2020).
- Anonymous, 2012. Otizm Platformu 2012 Basın Duyurusu. Türkiye Tohum Otizm Vakfı, <http://www.odfed.org/otizm-platformu-2012-basin-duyurusu/> (14.03.2020).
- Anonymous, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th. Edition (DSM-5). American Psychiatric Association, Washington, DC, USA.
- Anonymous, 2015. Causes. <https://www.autism-society.org/what-is/causes/> (29.03.2020).
- Anonymous, 2017. Naive Bayes Classifiers. <https://www.geeksforgeeks.org/naive-bayes-classifiers/> (26.03.2020).
- Anonymous, 2018. DE-ENIGMA: Otistik çocukların duygularını tanımasına ve ifade etmesine yardımcı olmak için oyunu kullanma. <https://ec.europa.eu/digital-single-market/en/news/de-enigma-using-play-help-autistic-children-recognise-and-express-emotions> (01.06.2020).
- Anonymous, 2020. CDC Estimates 1 in 54 Children has been Identified with Autism Spectrum Disorder. Centers for Disease Control (CDC), <https://www.cdc.gov/ncbddd/autism/addm-community-report/documents/addm-community-report-2020-h.pdf> (07.03.2020).
- Baranek, G.T., 1999. Autism During Infancy: A Retrospective Video Analysis of Sensory-Motor and Social Behaviors at 9–12 Months of Age. *Journal of Autism and Developmental Disorders*, 29, 213-224.
- Barbaro, J. and Dissanayake C., 2009. Autism spectrum disorders in infancy and toddlerhood: A review of the evidence on early signs, early identification tools, and early diagnosis. *Journal of Developmental and Behavioral Pediatrics*, 30 (5), 447-459.
- Bathija, A. and Sharma G., 2019. Visual Object Detection and Tracking using YOLO and SORT. *International Journal of Engineering Research & Technology (IJERT)*, 8 (11), 705-708.
- Bay, H., Ess A., Tuytelaars T. and Gool L.V., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110 (3), 346-359.
- Bewley, A., Ge Z., Ott L., Ramos F. and Upcroft B., 2016. Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA.
- Bhalla, D., 2017. Support Vector Machine Simplified Using R. <https://www.listendata.com/2017/01/support-vector-machine-in-r-tutorial.html> (04.04.2020).

- Bilinski, P. and Bremond F., 2011. Evaluation of Local Descriptors for Action Recognition in Videos. *International Conference on Computer Vision Systems ICVS*, 61-70.
- Bishop, S.L., Richler J. and Lord C., 2006. Association Between Restricted and Repetitive Behaviors and Nonverbal IQ in Children with Autism Spectrum Disorders. *Child Neuropsychology*, 12 (4-5), 247-267.
- Bosl, W.J., Tager-Flusberg H. and Nelson C.A., 2018. EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach. *Scientific Reports*, 8, 6828.
- Boualia, S.N. and Amara N.E.B., 2018. 3D CNN for Human Action Recognition. *International Conference on Sensors, Systems, Signals and advanced technologies, (SSS'18)*, Hammamet, Tunisia.
- Bradski, G. and Kaehler A., 2008. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, 557 p, USA.
- Camarata, S., 2014. Early identification and early intervention in autism spectrum disorders: Accurate and effective?. *International Journal of Speech-Language Pathology*, 16 (1), 1-10.
- Collobert, R., Bengio S. and Mariethoz J., 2002. Torch: A modular machine learning software library. Technical Report IDIAP-RR 02-46.
- Cortés, X., Conte D. and Cardot H., 2018. A new bag of visual words encoding method for human action recognition. *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, China.
- Dalal, N. and Triggs B., 2005. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA.
- Elsabbagh, M., Divan G., Koh Y.J., Kim Y.S., Kauchali S., Marcin C., Montiel-Nava C., Patel V., Paula C.S., Wang C., Yasamy M.T. and Fombonne E., 2012. Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, 5 (3), 160-179.
- Foggia, P., Percannella G., Saggese A. and Vento M., 2013. Recognizing Human Actions by a Bag of Visual Words. *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Manchester, UK.
- Gauglitz, S., Hollerer T. and Turk M., 2011. Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision*, 94, 1-26.
- Gilchrist, K.H., Hegarty-Craver M., Christian R.B., Grego S., Kies A.C. and Wheeler A.C., 2018. Automated Detection of Repetitive Motor Behaviors as an Outcome Measurement in Intellectual and Developmental Disabilities. *Journal of Autism and Developmental Disorders*, 48, 1458-1466.
- Goodwin, M.S., Intille S.S., Albinali F. and Velicer W.F., 2011. Automated Detection of Stereotypical Motor Movements. *Journal of Autism and Developmental Disorders*, 41, 770-782.
- Gulli, A., and Pal S., 2017. *Deep Learning with Keras*. Packt Publishing, 296, Birmingham, UK.
- Guthrie, W., Swineford L.B., Nottke C. and Wtherby A.M., 2013. Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *The Journal of Child Psychology and Psychiatry*, 54 (5), 582-590.
- Hochreiter, S. and Schmidhuber J., 1997. Long Short-Term Memory. *Neural Computation*, 9 (8), 1735-1780.
- Huerta, M. and Lord C., 2012. Diagnostic evaluation of autism spectrum disorders. *Pediatr Clin North Am*, 59 (1), 103-111.
- Ji, S., Xu W., Yang M. and Yu K., 2013. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (1), 221-231.

- Jiang, Y., Xu Q. and Cao X., 2018. Outfit Recommendation with Deep Sequence Learning. 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), Xi'an, China.
- John, G.H. and Langley P., 2013. Estimating Continuous Distributions in Bayesian Classifiers. arXiv:1302.4964.
- Kathuria, A., 2018. What's new in YOLO v3? <https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b> (01.07.2020).
- Kim, S.H. and Lord C., 2010. Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS). *Autism Research*, 3 (4), 162-173.
- Kushki, A., Brian J., Dupuis A. and Anagnostou E., 2014. Functional autonomic nervous system profile in children with autism spectrum disorder. *Molecular Autism*, 4 (5), article 39.
- Lam, K.S.L. and Aman M.G., 2007. The repetitive behavior scale-revised: independent validation in individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 37, 855-866.
- Landa, R., 2007. Early communication development and intervention for children with autism. *Mental Retardation and Developmental Disabilities Research Reviews*, 13 (1), 16-25.
- Laptev, I., 2005. On Space-Time Interest Points. *International Journal of Computer Vision*, 64, 107-123.
- Laptev, I., Marszalek M., Schmid C. and Rozenfeld B., 2008. Learning realistic human actions from movies. 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA.
- Loomes, R., Hull L. and Mandy W.P. L., 2017. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56 (6), 466-474.
- Lord, C., Rutter M., DiLavore P. and Risi S., 1999. Autism Diagnostic Observation Schedule (ADOS). Calif: Western Psychological Services, Los Angeles.
- Lord, C. and Jones R.M., 2012. Annual research review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 53(5), 490-509.
- Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 91-110.
- MacQueen, 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1nd ed., Ed: Cam L.M.L., Neyman J, California, USA, 281-296.
- Mahmood, H., 2018. The Softmax Function, Simplified. <https://towardsdatascience.com/softmax-function-simplified-714068bf8156> (13.12.2019).
- Mandell, D.S., Novak M.M. and Zubritsky C.D., 2005. Factors Associated With Age of Diagnosis Among Children With Autism Spectrum Disorders. *Pediatrics*, 116 (6), 1480-1486.
- Marinoiu, E., Zafir M., Olaru V. and Sminchisescu C., 2018. 3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children With Autism. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2158-2167.
- Matson, J.L., Boisjoli J.A., González M.L., Smith K.R. and Wilkins J., 2007. Norms and cut off scores for the autism spectrum disorders diagnosis for adults (ASD-DA) with intellectual disability. *Research in Autism Spectrum Disorders*, 1 (4), 330-338.
- Matson, J.L., Wilkins J. and González M., 2008. Early identification and diagnosis in autism spectrum disorders in young children and infants: How early is too early?. *Research in Autism Spectrum Disorders*, 2 (1), 75-84.
- Mohana, 2019. Object Detection and Classification Algorithms using Deep Learning for Video Surveillance Applications. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8 (8) 386-395.

- Niebles, J.C., Wang H. and Fei-Fei L., 2008. Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words. *International Journal of Computer Vision*, 79, 299-318.
- O’Roak, B.J. and State M.W., 2008. Autism genetics: strategies, challenges, and opportunities. *Autism Research*, 1 (1), 4-17.
- Pal, M., 2008. Multiclass Approaches for Support Vector Machine Based Land Cover Classification. *arXiv:0802.2411*.
- Pedregosa, F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M. and Duchesnay É., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Perš, J., Sulić V., Kristan M., Perše M., Polanec K. and Kovačič S., 2010. Histograms of optical flow for efficient representation of body motion. *Pattern Recognition Letters*, 31 (11), 1369-1376.
- Popa, A., Zafir M. and Sminchisescu C., 2017. Deep Multitask Architecture for Integrated 2D and 3D Human Sensing. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6289-6298.
- Pyles, D.A.M., Riordan M.M. and Bailey J.S., 1997. The stereotypy analysis: An instrument for examining environmental variables associated with differential rates of stereotypic behavior. *Research in Developmental Disabilities*, 18 (1), 11-38.
- Rad, N.M. and Furlanello C., 2016. Applying Deep Learning to Stereotypical Motor Movement Detection in Autism Spectrum Disorders. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, Barcelona, Spain.
- Rad, N.M., Kia S.M., Zarbo C., van Laarhoven T., Jurman G., Venuti P., Marchiori E. and Furlanello C., 2018. Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing*, 144, 180-191.
- Rafique, A.A., Jalal A. and Ahmed A., 2019. Scene Understanding and Recognition: Statistical Segmented Model using Geometrical Features and Gaussian Naïve Bayes. *IEEE conference on International Conference on Applied and Engineering Mathematics*.
- Rajagopalan, S.S., Dhall A. and Goecke R., 2013. Self-Stimulatory Behaviours in the Wild for Autism Diagnosis. *The IEEE International Conference on Computer Vision (ICCV) Workshops*, 755-761.
- Rajagopalan, S.S. and Goecke R., 2014. Detecting self-stimulatory behaviours for autism diagnosis. *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France.
- Ramanathan, M., Yau W.Y. and Teoh E.K., 2014. Human Action Recognition With Video Data: Research and Evaluation Challenges. *IEEE Transactions on Human-Machine Systems*, 44 (5), 650-663.
- Rao, C., Shah M., 2001. View-invariance in action recognition. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA.
- Redmon, J., Divvala S., Girshick R. and Farhadi A., 2016. You Only Look Once: Unified, Real-Time Object Detection. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.
- Redmon, J. and Farhadi A., 2018. YOLOv3: An Incremental Improvement. *arXiv:1804.02767*
- Reichow, B., 2012. Overview of Meta-Analyses on Early Intensive Behavioral Intervention for Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 42, 512-520.
- Richler, J., Bishop S.L., Kleinke J.R. and Lord C., 2007. Restricted and repetitive behaviors in young children autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37 (1), 73-85.

- Roberts, T.P.L., Khan S.Y., Rey M., Monroe J.F., Cannon K., Blaskey L., Woldoff S., Qasmieh S., Gandal M., Schmidt G.L., Zarnow D.M., Levy S.E. and Edgar J.C., 2010. MEG detection of delayed auditory evoked responses in autism spectrum disorders: towards an imaging biomarker for autism. *Autism Research*, 3 (1), 8-18.
- Rutter, M., Bailey A., Lord C. and Berument S.K., 2003. Social Communication Questionnaire (SCQ). Western Psychological Services, Los Angeles, CA, USA.
- Schindler, K. And Gool L.V., 2008. Action snippets: How many frames does human action recognition require?. 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA.
- Schuller, B., Steidl S., Batliner A., Bergelson E., Krajewski J., Janott C., Amatuni A., Casillas M., Seidl A., Soderstrom M., Warlaumont A., Hidalgo G., Schnieder S., Heiser C., Hohenhorst W., Herzog M., Schmitt M., Qian K., Zhang Y., Trigeorgis G., Tzirakis P., and Zafeiriou S., 2017. The INTERSPEECH 2017 computational paralinguistics challenge: Addressee, cold & snoring. In *Computational Paralinguistics Challenge (ComParE)*, Interspeech.
- Sharma, S., Kiros R. and Salakhutdinov R., 2015. Action Recognition using Visual Attention. arXiv:1511.04119
- Sharma, S.R., Gonda X. and Tarazi F.I., 2018. Autism Spectrum Disorder: Classification, diagnosis and therapy. *Pharmacology & Therapeutics*, 190, 91-104.
- Shi, X., Chen Z., Wang H., Yeung D.Y., Wong W.K. and Woo W.C., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, 802-810.
- Shukla, P., Biswas K.K. and Kalra P.K., 2013. Action Recognition using Temporal Bag-of-Words from Depth Maps. *MVA2013 IAPR International Conference on Machine Vision Applications*, Kyoto, Japan.
- Sigman, M. and McGovern C.W., 2005. Improvement in Cognitive and Language Skills from Preschool to Adolescence in Autism. *Journal of Autism and Developmental Disorders*, 35, 15–23.
- Sprague, R.L., 1996. Instruments for assessing stereotypic movements: A review of the literature. *American Psychological Association*, 85-113.
- Stone, W.L., Coonrod E.E. and Ousley O.Y., 2000. Brief report: screening tool for autism in two-year-olds (STAT): development and preliminary data. *Journal of Autism and Developmental Disorders*, 30 (6), 607-612.
- Subasi, A. and Erçelebi E., 2005. Classification of EEG signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine*, 78 (2), 87-99.
- Sun, Q., Liu H., Ma L. and Zhang T., 2016. A novel hierarchical Bag-of-Words model for compact action representation. *Neurocomputing*, 174, 722-732.
- Swingler, K., 2013. Multi-Layer Perceptrons. *Computing Science&Math*, University of Stirling, <http://www.cs.stir.ac.uk/courses/ITNP4B/lectures/kms/4-MLP.pdf> (01.05.2020).
- Tian, Y., Min X., Zhai G. and Gao Z., 2019. Video-Based Early ASD Detection via Temporal Pyramid Networks. 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China.
- Ulay, H.T. ve Ertuğrul A., 2009. Otizmde Beyin Görüntüleme Bulguları: Bir Gözden Geçirme. *Türk Psikiyatri Dergisi*, 20(2), 164-174.
- Wall, D.P., Kosmicki J., DeLuca T.F., Harstad E. and Fusaro V.A., 2012. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*, 2, e100.
- Wang, H., Shao M., Liu Y. and Zhao W., 2017. Enhanced Efficiency 3D Convolution Based on Optimal FPGA Accelerator. *IEEE Access*, 5, 6909-6916.

- Wang, H., Ullah M.M., Klaser A., Laptev I. and Schmid C., 2009. Evaluation of local spatio-temporal features for action recognition. British Machine Vision Conference, London, United Kingdom.
- Westeyn, T., Vadas K., Bian X., Starner T. and Abowd G.D., 2005. Recognizing mimicked autistic self-stimulatory behaviors using HMMs. Ninth IEEE International Symposium on Wearable Computers (ISWC'05), Osaka, Japan.
- Wetherby, A., Allen L., Cleary J., Kublin K. and Goldstein H., 2002. Validity and reliability of the Communication and Symbolic Behavior Scales Developmental Profile with very young children. *Journal of Speech, Language, and Hearing Research*, 45 (6), 1202-1218.
- Wolfe, J., Jin X., Bahr T. and Holzer N., 2017. Application Of Softmax Regression and Its Validation For Spectral-Based Land Cover Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42 (1), 455-459.
- Viola, P. and Jones M., 2001. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA.
- Vostanis, P., Smith B., Corbett J., Sungum-Paliwal R., Edwards A., Gingell K., Golding R., Moore A. and Wiilliams J., 1998. Parental Concerns of Early Development in Children with Autism and Related Disorders. *Autism*, 2 (3), 229-242.
- Yavral, F., Bebek N., Abalı O., Gürses C., Baykan B. ve Gökyiğit A., 2015. Dikkat Eksikliği ve Otizm Tanılı Çocuk Olgularda Elektroensefalografi Bulguları. *Epilepsi*, 21 (3), 133-138.
- Yean, C., 2015. Data preparation: The Balancing Act. <https://devblogs.microsoft.com/cse/2015/11/29/data-preparation-the-balancing-act/> (13.03.2020).
- Yeung, S., Russakovsky O., Mori G. and Fei-Fei L., 2016. End-to-end learning of action detection from frame glimpses in videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2678–2687.
- Young, R., Brewer N. and Pattison C., 2003. Parental Identification of Early Behavioural Abnormalities in Children with Autistic Disorder. *Autism*, 7 (2), 125-143.
- Zang, L., Zhu G., Shen P., Song J., Shah S.A. and Bennamoun M., 2017. Learning Spatiotemporal Features Using 3DCNN and Convolutional LSTM for Gesture Recognition. *The IEEE International Conference on Computer Vision (ICCV)*, 3120-3128.
- Zunino, A., Morerio P., Cavallo A., Ansuini C., Podda J., Battaglia F., Veneselli E., Becchio C. and Murino V., 2018. Video Gesture Analysis for Autism Spectrum Disorder Detection. 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China.
- Zwaigenbaum, L., Bryson S., Rogers T., Roberts W., Brian J. and Szatmari P., 2005. Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, 23 (2-3), 143-152.
- Zwaigenbaum, L., Bauman M.L., Choueiri R., Kasari C., Carter A., Granpeesheh D., Mailloux Z., Roley S.S., Wagner S., Fein D., Pierce K., Buie T., Davis P.A., Newschaffer C., Robins D., Wetherby A., Stone W.L., Yirmiya N., Estes A., Hansen R.L., McPartland J.C. and Natowicz M.R., 2015. Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research. *Pediatrics*, 136 (1), 60-81.

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı:	Sibel KAÇDIOĞLU
Doğum tarihi:	28 Temmuz 1991
Doğum Yeri:	Yakutiye/Erzurum
Uyruğu:	T.C.
Adres:	Erzurum Teknik Üniversitesi Mühendislik Mimarlık Fakültesi Kat:2
E-mail:	sibel.kacdioglu@erzurum.edu.tr
Eğitim	
Lise:	Erzurum Merkez Anadolu Lisesi
Lisans:	Anadolu Üniversitesi, Mühendislik Fakültesi
Yabancı Dil Bilgisi	
İngilizce:	İyi
Yayınlar	
Bildiri:	Kaçdioğlu, S., Özyer B., Tümöklü Özyer G., 2020. Otizm Spektrum Bozukluklarında Kendini Uyarıcı Davranışları Tanıma. 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Türkiye, 05 - 07 October 2020.