

T.C.
KASTAMONU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MALZEME BİLİMİ VE MÜHENDİSLİĞİ ANABİLİM DALI



**DERİN ÖĞRENME İLE GERÇEK ZAMANLI TÜRKÇE İŞARET
DİLİ İFADE TANIMA**

CUMHUR TORUN

MALZEME BİLİMİ VE MÜHENDİSLİĞİ

DOÇ. DR. ABDULKADİR KARACI

2025
KASTAMONU

TEZ ONAYI

Cumhur TORUN tarafından hazırlanan “**DERİN ÖĞRENME İLE GERÇEK ZAMANLI TÜRKÇE İŞARET DİLİ İFADE TANIMA**” adlı tez çalışmasının savunma sınavı **07.07.2025** tarihinde yapılmış olup aşağıda verilen jüri tarafından oy birliği / oy çokluğu ile Kastamonu Üniversitesi Fen Bilimleri Enstitüsü **Malzeme Bilimi ve Mühendisliği Anabilim Dalı Malzeme Bilimi ve Mühendisliği** olarak kabul edilmiştir.

Danışman	Doç Dr. Abdulkadir KARACI Samsun Üniversitesi	
Jüri Üyesi	Doç. Dr. Salih GÖRGÜNOĞLU Kastamonu Üniversitesi	
Jüri Üyesi	Doç. Dr. Halil İbrahim AKYÜZ Kastamonu Üniversitesi	
Jüri Üyesi	Doç. Dr. Selçuk ÇAKMAK Samsun Üniversitesi	
Jüri Üyesi	Dr. Öğretim Üyesi Emel SOYLU Samsun Üniversitesi	

Jüri üyeleri tarafından kabul edilmiş olan bu tez Kastamonu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulunca onanmıştır.

Enstitü Müdürü Doç. Dr. Selçuk MEMİŞ

TAAHHÜTNAME

Bu tezin tasarımı, hazırlanması, yürütülmesi, araştırmalarının yapılması ve bulgularının analizlerinde bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu; ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını, bilimsel etiğe uygun olarak kaynak gösterildiğini bildirir ve taahhüt ederim.

Cumhur TORUN

ÖZET

MALZEME BİLİMİ VE MÜHENDİSLİĞİ

DERİN ÖĞRENME İLE GERÇEK ZAMANLI TÜRKÇE İŞARET DİLİ İFADE TANIMA

CUMHUR TORUN

KASTAMONU ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
MALZEME BİLİMİ VE MÜHENDİSLİĞİ ANABİLİM DALI
DANIŞMAN:DOÇ. DR. ABDULKADİR KARACI

Bu çalışma, Türk İşaret Dili (TİD) ifadelerinin gerçek zamanlı tanınmasını sağlamak amacıyla MediaPipe tabanlı iskelet verisi çıkarımı ve derin öğrenme modellerinin kullanımını kapsamaktadır. Çalışmanın temel amacı, işaret dili bilmeyen bireylerin TİD kullanıcılarıyla daha etkili iletişim kurmasını sağlamak ve işitme engelli bireylerin toplumsal entegrasyonunu desteklemektir. Bu doğrultuda, Milli Eğitim Bakanlığı'nın TİD Dersi Öğretim Materyali esas alınarak 446 kelimeden oluşan özgün bir veri seti geliştirilmiş ve TurkSign446 olarak adlandırılmıştır. Hem statik hem de dinamik işaretleri içeren bu veri seti, cihazdan bağımsız ve kapsamlı bir yapıda oluşturulmuş; MediaPipe Holistic kütüphanesi aracılığıyla vücut, el ve yüz anahtar noktalarından elde edilen iskelet görüntüleriyle desteklenmiştir. Veri seti %70 eğitim, %10 doğrulama ve %20 test olmak üzere dışarda tutma (hold-out) yöntemiyle ayrılmıştır. Bu veriler üzerinde LSTM, Bi-LSTM, GRU ve CNN gibi temel modellerin yanı sıra CNN+GRU, CNN+LSTM, CNN+BiLSTM, GRU+LSTM ve GRU+BiLSTM gibi hibrit mimariler de eğitilmiş; en yüksek test doğruluğu %97,72 ile CNN+GRU modeli tarafından elde edilmiştir. Bu model, eğitim sürecine dâhil edilmiş ve edilmemiş iki farklı katılımcı üzerinden gerçek zamanlı olarak test edilmiş; kelimeler için 10 tekrar içeren senaryolar sonucunda sırasıyla %85,50 ve %80,26 genel başarı oranlarına ulaşılmıştır. Ayrıca, her kelimeye ait 30 çerçevelik iskelet görüntüleri; ızgara, optik akış ve yatay birleştirme yöntemleriyle farklı temsillere dönüştürülmüş ve bu temsiller üzerinden ResNet18 ile dikkat mekanizması içeren CBAM+ResNet18 mimarileri eğitilmiştir. En yüksek doğruluk değeri, %97,30 ile CBAM+ResNet18 modeli ve optik akış temsili kullanılarak elde edilmiştir. Sonuç olarak, TurkSign446 veri seti ve bu çalışma kapsamında uygulanan derin öğrenme tabanlı yaklaşımlar, Türk İşaret Dili'nin gerçek zamanlı tanınmasına yönelik önemli bir temel oluşturmaktadır. Çalışma, hem akademik araştırmalarda hem de uygulamalı sistem geliştirmelerinde kullanılabilir niteliktedir. Gelecekte, veri setinin genişletilmesi ve hibrit modellerin daha karmaşık ifadeleri tanıyacak biçimde optimize edilmesi planlanmaktadır.

ANAHTAR KELİMELER:Türk İşaret Dili, Derin Öğrenme, MediaPipe, Gerçek Zamanlı Tanıma

Temmuz 2025, 118

ABSTRACT

MATERIALS SCIENCE AND ENGINEERING

REAL-TIME RECOGNITION OF TURKISH SIGN LANGUAGE EXPRESSION USING DEEP LEARNING

CUMHUR TORUN

**KASTAMONU UNIVERSITY INSTITUTE OF SCIENCE
DEPARTMENT OF MATERIALS SCIENCE AND ENGINEERING
SUPERVISOR: ASSOC. PROF. DR. ABDULKADİR KARACI**

This study focuses on the use of MediaPipe-based skeleton data extraction and deep learning models to enable the real-time recognition of Turkish Sign Language (TİD) expressions. The primary objective is to facilitate more effective communication between individuals who do not know sign language and TİD users, thereby promoting the social inclusion of hearing-impaired individuals. In this context, an original dataset consisting of 446 words was developed based on the Turkish Ministry of National Education's TİD Course Teaching Material and named TurkSign446. This dataset includes both static and dynamic signs and was constructed to be comprehensive and device-independent. Skeleton images were obtained using the MediaPipe Holistic library by capturing keypoints from the body, hands, and face. The dataset was split using the hold-out method into 70% for training, 10% for validation, and 20% for testing. On this dataset, various baseline models such as LSTM, Bi-LSTM, GRU, and CNN were trained, along with hybrid architectures including CNN+GRU, CNN+LSTM, CNN+BiLSTM, GRU+LSTM and, GRU+BiLSTM. The highest test accuracy of 97.72% was achieved with the CNN+GRU model. This model was also tested in real time with two participants—one included in the training data and one excluded. In test scenarios involving 10 repetitions, overall success rates of 85.50% and 80.26% were recorded for the seen and unseen participants, respectively. Additionally, for each word, 30-frame skeleton sequences were transformed into different visual representations using grid, optical flow, and horizontal concatenation methods. Based on these representations, ResNet18 and CBAM+ResNet18 architectures (the latter incorporating an attention mechanism) were trained. The highest recognition accuracy 97.30% was obtained using the CBAM+ResNet18 model with optical flow data. In conclusion, the TurkSign446 dataset and the deep learning-based approaches implemented in this study provide a significant foundation for real-time recognition of Turkish Sign Language. The results indicate that the system is applicable for both academic research and practical deployments. In future work, it is planned to expand the dataset and further optimize hybrid models to recognize more complex sign language expressions.

KEYWORDS: Turkish Sign Language, Deep Learning, MediaPipe, Real-Time Recognition

July 2025, 118 Page

TEŞEKKÜR

Bu çalışma, ailemin sarsılmaz desteği olmadan mümkün olamazdı. En büyük destekçim eşim Gülşen'e, kızım Ayşe Duru'ya, oğlum Ali Erdem'e, bugünlere gelmemi sağlayan anne ve babama, üzerimde emeği olan tüm öğretmenlerime sevgi ve teşekkürü borç bilirim. Onların sevgisi, inancı ve cesaretlendirmesi bu çalışmanın gerçekleşmesinde temel bir rol oynamıştır.

Bu tez çalışması boyunca önemli çabalar sarf ettim; ancak, danışmanım Doç. Dr. Abdulkadir KARACI'nın nazik desteği ve değerli yardımları olmadan bu çalışmayı tamamlamam mümkün olmazdı. Rehberliği ve katkıları için kendisine derin minnettarlığımı sunuyorum, doktora tezim için daha iyi bir danışman ve akıl hocasına sahip olmayı hayal edemeyeceğimi belirtmek isterim. Ayrıca, değerli önerileri ve destekleri için kıymetli jüri üyelerim Doç. Dr. Salih GÖRGÜNOĞLU ve Doç. Dr. Halil İbrahim AKYÜZ'e en içten teşekkürlerimi iletmek isterim.

Cumhur TORUN

Kastamonu, 2025

İÇİNDEKİLER

Sayfa

TEZ ONAYI	ii
TAAHHÜTNAME	iii
ÖZET.....	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ŞEKİLLER DİZİNİ	ix
TABLolar DİZİNİ	xi
SİMGELER VE KISALTMALAR DİZİNİ	xii
1. GİRİŞ.....	13
2. ÖNCEKİ ÇALIŞMALAR	19
3. İŞARET DİLİ VE TANIMA SİSTEMLERİ	32
3.1 İşaret dili ve İletişim	32
3.2 Türkçe İşaret Dili (TİD) Alfabesi	33
3.3 Türkçe İşaret Dili	34
3.4 Uluslararası İşaret Dili (International Sign Language – ISL) Alfabesi	35
3.5 Statik ve Dinamik İşaret Dili Kelimeleri	36
3.6 İşaret Dili Tanıma Sistemleri	37
3.6.1 Kinect Sensör V2	37
3.6.2 Leap Motion Controller	39
3.6.3 Veri Eldivenleri.....	40
3.6.4 Görüntü Tabanlı Sistemler.....	41
3.7 Görüntü Tabanlı Sistemlerde Kullanılan Anahtar Nokta Tespit Algoritmaları.....	42
3.7.1 MediaPipe	44
4. MATERYAL VE YÖNTEM	48
4.1 Veri Seti	48
4.2 Veri Seti Oluşturma	49
4.3 Veri Seti Klasör Yapısı	52
4.4 MediaPipe Holistic Anahtar Nokta Tespiti.....	54
4.5 Yöntem.....	55
4.5.1 LSTM (Long Short-Term Memory)	56
4.5.2 GRU (Gated Recurrent Unit).....	59
4.5.3 Bi-LSTM.....	61
4.5.4 CNN (Convolutional Neural Networks)	63
4.5.5 CNN + LSTM	65
4.5.6 CNN + BiLSTM	66
4.5.7 CNN+GRU	68
4.5.8 GRU+LSTM	69
4.5.9 GRU+Bi-LSTM.....	71
4.5.10 CBAM (Convolutional Block Attention Module).....	73
4.5.11 ResNet-18 (Residual Network – 18).....	74
4.6 Performans Metrikleri.....	75
5. DENEYSEL SONUÇLAR VE TARTIŞMA.....	76

5.1	Anahtar Nokta Koordinatlarını Ön İşleme.....	76
5.1.1	Kelimeleri Etiketleme	76
5.1.2	Dizileri Yükleme.....	77
5.2	İskelet Görüntüsü Dosyalarını Ön İşleme.....	78
5.3	Modellerin Eğitilmesi ve Test Edilmesi	80
5.3.1	Hiper Parametreler	80
5.3.2	Veri Setinin Bölünmesi.....	82
5.4	Dışarda Tutma Doğrulama ile Elde Edilen Sınıflandırma Sonuçları.....	82
5.5	İskelet Görüntüleri Üzerinden Elde Edilen Doğrulama Sonuçları	88
5.6	Tartışma	92
5.6.1	Dışarıda Tutma (Hold-Out) Doğrulama Yöntemi ile Doğruluk Değerlendirmesi.....	92
5.6.2	İskelet Görüntüleri Üzerinden Doğruluk Değerlendirmesi	93
5.6.3	Gerçek Zamanlı Değerlendirme.....	94
6.	SONUÇ VE ÖNERİLER	103
	KAYNAKLAR	106
	ÖZGEÇMİŞ.....	Hata! Yer işareti tanımlanmamış.

ŞEKİLLER DİZİNİ

Sayfa

Şekil 3.1 (a) Statik ve (b) Dinamik Kelime Örnekleri	36
Şekil 3.2 Microsoft Kinect V2	38
Şekil 3.3 Leap Motion.....	39
Şekil 3.4 Data Gloves.....	40
Şekil 3.5 MediaPipe Hand ile Tespit Edilen Yer İşaretleri	45
Şekil 3.6 MediaPipe Pose ile Tespit Edilen Yer İşaretleri	45
Şekil 3.7 MediaPipe Holistic Çalışma Yöntemi	46
Şekil 4.1 Türk İşaret Dili Dersi Öğretim Materyali Kitabı Birinci Ünite Kelimeleri	48
Şekil 4.2 Her Çerçeve İçin Anahtar Noktaların Tespiti	49
Şekil 4.3 Kırpma İşlemi	50
Şekil 4.4 Kırpma İşlemi	51
Şekil 4.5 Birim Değere Çevrilmiş Koordinat Değerleri.....	51
Şekil 4.6 Numpy Dosyaları ve İskelet Görünümü Görselleri	51
Şekil 4.7 Veri Seti Klasör Yapısı	53
Şekil 4.8 Veri Seti Oluşturma Aşaması	53
Şekil 4.9 Veri Seti Oluşturma Ortamı	54
Şekil 4.10 TİD Kelime Tanıma Sistemin Genel Tasarımı	55
Şekil 4.11 LSTM Yapısı	57
Şekil 4.12 LSTM Model Yapısı	57
Şekil 4.13 GRU ve LSTM Modelleri	59
Şekil 4.14 GRU Model Yapısı	60
Şekil 4.15 Bi-LSTM Modeli.	61
Şekil 4.16 Bi-LSTM Model Yapısı	63
Şekil 4.17 CNN Sınıflayıcı Genel Yapısı	64
Şekil 4.18 CNN Model Yapısı	65
Şekil 4.19 CNN + LSTM Hibrit Model Yapısı.....	65
Şekil 4.20 CNN + BiLSTM Hibrit Model Yapısı	67
Şekil 4.21 CNN+GRU Model Yapısı	69
Şekil 4.22 GRU+LSTM Model Yapısı	70
Şekil 4.23 GRU+Bi-LSTM Model Yapısı	72
Şekil 4.24 CBAM Modül Yapısı.....	74
Şekil 5.1 Kelimelerin Sayısal Karşılığı.....	76
Şekil 5.2 Izgara Birleştirme.....	78
Şekil 5.3 Optik Akış Temsili Birleştirme.....	79
Şekil 5.4 Yan Yana Birleştirme	79
Şekil 5.5 LSTM Modeli Kayıp ve Doğruluk Eğrisi	83
Şekil 5.6 GRU Modeli Kayıp ve Doğruluk Eğrisi	84
Şekil 5.7 BiLSTM Modeli Kayıp ve Doğruluk Eğrisi	84
Şekil 5.8 CNN Modeli Kayıp ve Doğruluk Eğrisi	85
Şekil 5.9 CNN+LSTM Modeli Kayıp ve Doğruluk Eğrisi	85
Şekil 5.10 CNN+BiLSTM Modeli Kayıp ve Doğruluk Eğrisi	85
Şekil 5.11 CNN+GRU Modeli Kayıp ve Doğruluk Eğrisi	86
Şekil 5.12 GRU+LSTM Modeli Kayıp ve Doğruluk Eğrisi	86
Şekil 5.13 GRU+BiLSTM Modeli Kayıp ve Doğruluk Eğrisi	86

Şekil 5.14 Izgara Birleştirme Yöntemi / ResNet-18 Modeli.....	89
Şekil 5.15 Optik Akış Birleştirme Yöntemi / ResNet-18 Modeli	89
Şekil 5.16 Yan Yana Birleştirme Yöntemi / ResNet-18 Modeli.....	90
Şekil 5.17 Izgara Birleştirme Yöntemi / CBAM+ResNet-18 Modeli.....	91
Şekil 5.18 Optik Akış Birleştirme Yöntemi / CBAM+ResNet-18 Modeli	91
Şekil 5.19 Yan Yana Birleştirme Yöntemi / CBAM+ResNet-18 Modeli.....	91
Şekil 5.20 Gerçek Zamanlı Test.....	95
Şekil 5.21 Modelin Tahmin Süresi	95
Şekil 5.22 Tahmin Edilmek İstenen Kelimelerin Optik Akış Görselleri	99



TABLÖLAR DİZİNİ

	<u>Sayfa</u>
Tablo 1.1 İşaret dili veri seti sınıf sayıları.....	16
Tablo 5.1 Modellerin hiper parametre ayarları	81
Tablo 5.2 Modellerin sınıflandırma sonuçları.....	88
Tablo 5.3 Modellerin iskelet görüntüleri üzerinde sınıflandırma sonuçları.....	92
Tablo 5.4 Modellerin karşılaştırılması	93
Tablo 5.5 Veri Setinde Verisi Olmayan Katılımcı ile Gerçek Zamanlı Değerlendirme Sonucu.....	96
Tablo 5.6 Tahmin Edilmesi İstenen Kelimeler ve Yanlış Tahmin Sonuçları	97
Tablo 5.7 Veri Setinde Verisi Olan Katılımcı ile Gerçek Zamanlı Değerlendirme Sonucu.....	100
Tablo 5.8 Gerçek Zamanlı Test Sonuçları	101

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar

CNN	: Convolutional Neural Network
RNN	: Recurrent Neural Network
LSTM	: Long Short Term Memory
Bi-LSTM	: Bidirectional Long Short Term Memory
GRU	: Gated Recurrent Unit
TİD	: Türkçe İşaret Dili
DÖ	: Derin Öğrenme
İDT	: İşaret Dili Tanıma
AUTSL	: Ankara University Turkish Sign Language
ERUSLR	: Erciyes University Sign Language Recognition
DTW	: Dynamic Time Warping
YAKA	: Yapay Arı Kolonisi Algoritması
HMM	: Hidden Markow Model
CAMSHIFT	: Continuously Adaptive Mean Shift
ESA	: Evrişimli Sinir Ağı
TSA	: Tekrarlayan Sinir Ağı
MEMP	: Multiple Extraction and Multiple Prediction
ISL	: International Sign Language
DNN	: Deep Neural Network
SVM	: Support Vector Machines

1. GİRİŞ

İletişim, bireylerin düşüncelerini, duygularını ve bilgilerini paylaşarak toplumda anlamlı bağlar kurmalarını sağlayan temel bir insani ihtiyaçtır. Bu ihtiyaç, insanların sosyal, ekonomik ve kültürel yaşamlarında etkin bir rol oynamalarını mümkün kılar. Ancak, işitme engelli bireyler için bu iletişim süreci, genellikle çeşitli engellerle sınırlanır. Dünya Sağlık Örgütü'ne göre dünya genelinde yaklaşık 466 milyon kişi işitme kaybı yaşamakta olup, bu rakamın 2050 yılına kadar 700 milyona ulaşması beklenmektedir. Türkiye'de ise işitme engelli bireylerin sayısının 3 milyonu aştığı tahmin edilmektedir (WHO, 2024). Bu bireylerin yaşamlarında karşılaştıkları iletişim zorlukları, eğitimden istihdama, sağlık hizmetlerine erişimden toplumsal katılıma kadar birçok alanda belirgin hale gelmektedir.

İşitme engelli bireyler, iletişim eksikliğinden dolayı toplumsal izolasyon, eğitim fırsatlarına erişimde kısıtlamalar ve iş bulma zorlukları gibi çok yönlü problemlere maruz kalmaktadır. Özellikle, toplumsal farkındalık eksikliği nedeniyle, işitme engelliler genellikle iletişim kurmada destekleyici araçlardan yoksun kalmaktadır. İşaret dili, işitme engelli bireyler için temel bir iletişim aracıdır; ancak bu dili öğrenen işitme engeli olmayan bireylerin sayısının düşük olması, iletişim engellerini daha da artırmaktadır. Ayrıca, işaret dilinin ülkeden ülkeye ve hatta bölgeden bölgeye değişiklik göstermesi, standart bir iletişim modelinin geliştirilmesini zorlaştırmaktadır (Yanikkerem vd., 2017).

Son yıllarda yapay zeka, derin öğrenme ve bilgisayarla görü teknolojilerindeki ilerlemeler, işaret dili tanıma sistemleri için önemli fırsatlar yaratmıştır. Derin öğrenme, insan davranışlarının ve hareketlerinin daha hassas ve doğru bir şekilde analiz edilmesini sağlayarak, bu sistemlerin etkinliğini artırmıştır. Evrişimsel Sinir Ağları (CNN), Tekrarlayan Sinir Ağları (RNN) ve Uzun Kısa Süreli Bellek (LSTM) gibi mimariler, özellikle görsel verilerin işlenmesi ve sınıflandırılması konusunda yüksek performans göstermektedir (Arun Prasath vd., 2023; Farooq vd., 2023; Sundar vd., 2022). Poz tanıma sistemleri ise, insan hareketlerini algılamak için temel bir teknoloji olarak öne çıkmaktadır. MediaPipe gibi kütüphaneler, eller, yüz ve vücut

pozlarını tespit ederek, bu verilerin işlenmesi için derin öğrenme modellerine hazır hale getirilmesini sağlamaktadır (Kuznetsova vd., 2024). Bu tür sistemler, gerçek zamanlı işaret dili tanıma uygulamalarının geliştirilmesine yönelik büyük bir potansiyel sunmaktadır.

Literatürde, işaret dili tanıma sistemleri üzerine yapılan çalışmalar, genellikle farklı veri setleri ve algoritmalar kullanılarak gerçekleştirilmiştir (Alaftekin vd., 2024; Colak vd., 2023; Gündüz vd., 2021; Katılmış vd., 2021). Hareket geçmişi görüntüleri (Motion History Images, MHI) ve iskelet anahtar noktaları gibi özellik çıkarım teknikleri, uzamsal ve zamansal verilerin bütünleşik bir şekilde analiz edilmesine olanak sağlamaktadır. Türk İşaret Dili (TİD) üzerine yapılan çalışmalar ise, genellikle sınırlı veri setleriyle gerçekleştirilmiş olup, sürekli işaret dili tanıma ve çeviri sistemlerinde önemli zorluklarla karşılaşmaktadır. Bu durum, daha büyük ve çeşitli veri setlerinin gerekliliğini ortaya koymaktadır.

Her dilin kendine özgü bir yapısı ve kelime hazinesi olduğu gibi, işaret dilleri de bu benzersiz özelliklere sahiptir. İşaret dilleri, dillerin doğal bir uzantısı olarak her toplumun kültürel, sosyal ve dilbilimsel bağlamına göre şekillenmiştir. Bu durum, farklı dillerin işaret diline yansıyan kelimelerinin ve dilbilgisel yapıların da farklılık göstermesine neden olmaktadır. Örneğin, TİD ile Amerikan İşaret Dili (ASL) arasında sadece işaretlerin görsel formları değil, aynı zamanda sözdizimsel yapılar ve ifade biçimleri açısından da belirgin farklılıklar bulunmaktadır. Bu yüzden yapılan işaret dili tanıma çalışmalarında farklı dillere ait veri setleri kullanılmaktadır.

TİD tanıma sistemlerini geliştirmek için az da olsa birkaç farklı veri seti bulunmaktadır. Sincan vd. (2020), Microsoft Kinect v2 ile kaydedilen ve 43 katılımcı tarafından gerçekleştirilen 226 işaret dili kelimesinden oluşan bir veri seti oluşturmuşlardır. Bu veri seti iç ve dış ortamlarda farklı arka planlara sahip renkli görüntüler, derinlik ve iskelet görüntülerini içermektedir. Güney vd. (2022), gerçek zamanlı TİD tanıma için evrişimli sinir ağlarını kullanarak bir model geliştirmişlerdir. Bu çalışmada 32 statik kelimelik bir veri seti 3 katılımcı ile oluşturulmuştur. Bankur vd. (2022), TİD harflerini tespit etmek için 29 harften oluşan bir veri seti oluşturmuşlardır. Veri setinde 29 harf için parlaklık, kontrast ve geometrik

değişikliklerde farklılıklar gösteren 10.000 etiketli görüntü bulunmaktadır. Kasapbaşı vd. (2023), Derin Öğrenme metotlarını kullanarak TİD alfabesini tahmin eden bir çalışma yapmışlardır. Bu çalışmada TİD alfabesinde yer alan 29 harf için görsellerden oluşan yeni bir veri seti oluşturulmuştur. Sonugür vd. (2023), TİD kelimelerinin tespiti için sensörlerden oluşan bir veri eldiveni geliştirmişlerdir. Çalışmada, 25 farklı katılımcıdan toplanan ve 120 kelimeyi kapsayan 3000 örnekten oluşan bir veri kümesini kullanmışlardır. Veri seti; 12 esneklik sensörü, iki atalet sensörü ve 10 Hall sensöründen gelen veriler kullanılarak oluşturulmuştur. Kindiroğlu vd. (2024), denetimli transfer öğrenme metodlarını karşılaştıran bir çalışma yapmıştır. Bu çalışmada Ankara University Turkish Sign Language Dataset (AUTSL) ve BosphorusSign22k (BSign22k) veri setleri kullanılmıştır. Özdemir vd. (2020), Microsoft Kinect v2 kullanarak BSign22k TİD veri setini geliştirmişlerdir. Bu veri setinde 6 katılımcıdan 744 kelime için alınan 22.542 görüntü bulunmaktadır. Veri setindeki kelimelerin 428 tanesi sağlık, 163 tanesi finans ve 174 tanesi de genel kelimelerden oluşmaktadır. Mercanoglu vd. (2020), TİD için Microsoft Kinect v2 kullanarak AUTSL veri seti geliştirmişlerdir. Veri setinde 43 farklı katılımcıdan 226 kelime için 38.336 örnek video bulunmaktadır. Özcan vd. (2021), hastane ortamında kullanılan ve genel kullanıma açık bir TİD veri seti oluşturmuşlardır. 25 sınıf üzerinden 49 katılımcıyla oluşturulan Erciyes University Sign Language Recognition (ERUSLR) isimli veri seti, toplamda 13186 örnek içermektedir. BSign22k ve AUTSL veri setlerine açık kaynak olarak erişilebilmektedir. Bu iki veri seti, TİD tanıma sistemlerinin geliştirilmesinde literatürde sıkça kullanılmaktadır.

Kelime düzeyinde hazırlanmış, literatürde sıkça kullanılan veri setleri ve veri setlerine ait özellikler Tablo 1.1’de sunulmuştur. Bu veri setleri incelediğinde TİD için oluşturulan veri setlerinin sadece üç tane olduğu görülmektedir. Bunlardan iki tanesi ise çok dar kapsamlı veri setleridir. Sadece bir tanesi (BSign22k) geniş kapsamlıdır. Ancak bu veri setinde bulunan 744 kelimenin 428 tanesi sağlık alanı, 163 tanesi finans alanı ve 174 tanesi de genel ifadelerdir. Bu dezavantajlardan dolayı günlük hayatta sık kullanılan TİD kelimelerinin bulunduğu bir veri setine ihtiyaç bulunmaktadır.

Tablo 1.1 İşaret dili veri seti sınıf sayıları

Veri seti	İşaret Dili	Sınıf	Katılımcı	Örnek
DGS Kinect 40	Almanca	40	15	3000
GSL 20	Yunanca	20	6	840
Boston ASL	İngilizce (ABD)	33	6	9800
PSL Kinect 30	Lehçe	30	1	300
PSL TOF 84	Lehçe	84	1	1680
LSA 64	İspanyolca	64	10	3200
DEVISIGN-D	Çince	50	8	6000
AUTSL	Türkçe	226	43	38336
BSign22k	Türkçe	744	6	22542
ERUSLR	Türkçe	25	49	13186

Bu çalışmanın temel amacı, işaret dili bilmeyen bireylerin işaret dili bilen bireylerle daha kolay iletişim kurmasını sağlamak için Derin Öğrenme modelleri geliştirmek ve bunu sağlayabilmek için de geniş kapsamlı orijinal bir veri seti oluşturmaktır. Bu ana amaç çerçevesinde belirlenen alt amaçlar ise aşağıdaki şekildedir:

- TİD alanında kullanılacak bir veri seti oluşturmak: Millî Eğitim Bakanlığı'nın yayınladığı TİD dersi öğretim materyalinden yararlanılarak 446 TİD kelimesini içeren geniş kapsamlı bir veri seti oluşturulması.
- Bağımsızlık ve çeşitlilik: Oluşturulan veri setinin cihazdan, çözünürlükten ve arka plandan bağımsız olması sağlanarak daha geniş bir kullanım alanı elde edilmesi.
- Yüksek çözünürlüklü veri temsili: Veri setindeki vücut anahtar noktalarının diğer veri setlerine göre çok daha fazla olması, daha zengin ve çeşitli veri temsili sağlanması.
- Halka açık veri seti: Oluşturulan veri setinin halka açık bir şekilde yayınlanması, TİD alanındaki çalışmaların nitelik ve nicelik olarak artmasını sağlayacak ve araştırmacıların daha geniş bir veri seti üzerinde çalışmalarına olanak tanıyacaktır.
- Gerçek zamanlı TİD tanıma uygulaması geliştirmek: Elde edilen veri seti üzerinde son teknoloji ürünü derin öğrenme mimarilerinin ve hibrit mimarilerin uygulanarak gerçek zamanlı TİD tanıma uygulamasının geliştirilmesi.

- Çeşitli derin öğrenme modellerinin değerlendirilmesi: Farklı derin öğrenme modelleri ve hibrit yöntemlerin kullanılmasıyla genişletilmiş veri seti üzerinde başarı durumlarının değerlendirilmesi hedeflenmektedir.

Bu amaçlar doğrultusunda, Millî Eğitim Bakanlığı Özel Eğitim ve Rehberlik Genel Müdürlüğünün yayınladığı TİD Dersi Öğretim Materyali kitabındaki temel iletişim kalıplarını içeren 446 adet kelime için MediaPipe kullanılarak özgün bir veri seti oluşturulmuştur. Ayrıca bu veri seti üzerinde hem gerçek zamanlı hem de dışarda tutma yöntemi ile derin öğrenme modelleri eğitilmiş ve test edilmiştir. Çalışma kapsamında kullanılan derin öğrenme modelleri, CNN, GRU, Bi-LSTM modelleri ve bu modellerin birbirine bağlanması ile oluşturulan CNN+GRU, CNN+BiLSTM, CNN+LSTM, GRU+BiLSTM, GRU+LSTM hibrit modellerdir. Bunun yanı sıra iskelet görüntüleri üzerinde Resnet18 ve dikkat mekanizması temelli CBAM+Resnet18 modeli de kullanılmıştır.

Bu çalışmanın ana katkıları aşağıdaki gibi özetlenebilir:

1. Daha önceki çalışmalarda kullanılmamış olan TİD ifadelerine ait özgün ve geniş kapsamlı (133 statik, 313 dinamik kelime) bir veri seti geliştirilmiştir. Bu veri setinin, TİD tanıma alanındaki gelecekteki araştırmalar için faydalı olması beklenmektedir. Ayrıca, veri setinin halka açık bir veri seti olarak yayınlanması ile TİD alanındaki çalışmaların nitelik ve nicelik olarak artması sağlanacaktır.
2. Veri seti cihazdan, çözünürlükten ve arka plandan bağımsız ve çok örnekli bir veri setidir. Bu, mevcut literatürdeki çalışmalardan farklılık gösterir ve TİD tanıma alanında önemli bir adımdır. Oluşturulan veri seti, diğer veri setlerine göre çok daha fazla sayıda vücut anahtar noktasını içerdiğinden dolayı daha zengin ve çeşitli veri temsili sağlamıştır.
3. TİD alanında daha önce yapılmayan gerçek zamanlı TİD uygulaması geliştirilmiştir.
4. Çalışmada, CNN, GRU, LSTM ve Bi-LSTM gibi temel derin öğrenme mimarileri kullanılmış; ayrıca bu mimarilerin kombinasyonu ile oluşan CNN+GRU,

CNN+BiLSTM, CNN+LSTM, GRU+BiLSTM, GRU+LSTM gibi hibrit modeller de tasarlanarak performans karşılaştırmaları yapılmıştır.

5. Hibrit model olan CNN+GRU, %97,72 gibi etkileyici bir doğruluk değerine ulaşarak literatürdeki benzer çalışmaları geride bırakmış ve hibrit modellerin işaret dili tanıma için etkinliğini göstermiştir.

6. En yüksek doğruluk değerine ulaşan CNN+GRU modeli hem eğitim sürecine dâhil edilmiş hem de edilmemiş iki farklı katılımcı üzerinde gerçek zamanlı olarak test edilmiştir ve gerçekleştirilen test senaryoları sonucunda, model sırasıyla; %85,50 ve %80,26 genel doğruluk değeri elde etmiştir.

7. İskelet görüntüleri üzerinde farklı birleştirme metotları kullanarak dikkat mekanizması temelli CBAM+ResNet-18 modeli ile de %97,30 doğruluk değeri elde edilmiştir.

Çalışmanın devamı şu şekilde yapılandırılmıştır: Bölüm 2, işaret dili tanıma sistemleri, işaret dili tanıma sistemlerinde kullanılan veri setleri ve vücut anahtar noktalarının tespiti için kullanılan algoritmalara ilişkin literatürde yer alan çalışmaların kapsamlı bir derlemesini sunmaktadır. Bölüm 3, işaret dilleri ve işaret dili tanıma sistemleri hakkında temel bilgileri ele almaktadır. Bu kapsamda, işaret dillerinin temel özellikleri, işaret dili tanıma sistemlerinde kullanılan yöntemler ve çalışma kapsamında kullanılan Media Pipe Holistic algoritması ayrıntılı bir şekilde açıklanmaktadır. Bölüm 4, TİD için oluşturulmuş veri setinin hazırlanma sürecine odaklanmaktadır. Bu bölümde, veri setinde yer alan kelime seçim kriterleri, veri seti oluşturma adımları ve bu süreçte kullanılan yöntemler detaylandırılmaktadır. Ayrıca, oluşturulan veri seti ile eğitilecek modeller hakkında bilgiler sunulmaktadır. Bölüm 5, deneysel tasarımı, derin öğrenme modellerinden elde edilen bulgu ve sonuçları ile bu sonuçların literatürdeki önceki çalışmalarla karşılaştırılmasını kapsamaktadır. Bölüm 6, çalışmanın genel sonuçlarını özetlemekte; genel değerlendirmeler ve öneriler sunmakta, gelecekte yapılması planlanan çalışmalara ilişkin öngörülerde bulunmaktadır.

2. ÖNCEKİ ÇALIŞMALAR

İşaret Dili Tanıma (İDT) teknolojileri, başlangıçta el yapımı özellikler ve klasik makine öğrenimi tekniklerine dayalı yaklaşımlarla başlamış, derin öğrenme yöntemlerinin benimsenmesiyle daha ileri seviyelere ulaşmıştır. Özellikle, özellik çıkarma ve sınıflandırma alanlarında kaydedilen gelişmeler, İDT çalışmalarının doğruluk oranlarını ve genellenebilirliğini önemli ölçüde artırmıştır. TİD ve diğer işaret dillerini konu alan çalışmalarda BosphorusSign22k (Camgoz vd., 2016; Özdemir vd., 2020), Arjantin İşaret Dili için LSA64 (Ronchetti vd., 2016b) ve Yunan İşaret Dili için GSL (Adaloglou vd., 2020) gibi veri kümeleri sıklıkla kullanılmaktadır. Bu veri kümeleri, farklı dil yapılarını temsil etmeleri ve önerilen yöntemlerin çeşitliliğini ve uyumluluğunu değerlendirmeye olanak tanımaları nedeniyle dikkat çekmektedir. İDT alanında izole işaretlerin tanınmasından sürekli çevrilere kadar geniş bir uygulama yelpazesinde kullanılan metodolojilerdeki ilerlemeler, bu teknolojilerin kullanıcılar açısından daha işlevsel hale gelmesini sağlamaktadır. Bu bağlamda, veri kümeleri ve derin öğrenme modellerinin etkinliği, İDT yöntemlerinin performansını belirlemede kilit bir rol oynamaktadır.

Görme tabanlı İDT sistemlerinin ilk sürümleri, el yapımı özellik çıkarma yöntemlerine ve makine öğrenimi tekniklerine dayalı bir yaklaşımı benimsemiştir. Bu yöntemlerden bazıları, Ölçekle-Değişmeyen Özellik Dönüşümü (Scale-Invariant Feature Transform - SIFT) (Yasir vd., 2016), Hızlandırılmış Sağlam Özellikler (Speeded-Up Robust Features - SURF) (Kour ve Mathew, 2017; Yang ve Peng, 2014), Yönlendirilmiş Gradyanların Histogramı (Histogram of Oriented Gradients - HOG) (Raj ve Jasuja, 2018) ve Temel Bileşen Analizi (Principal Component Analysis - PCA) (Gweth vd., 2012; Mohana ve Reddy, 2014) gibi yöntemlerdir. Bu teknikler, işaret dili harflerinden, sayılarından veya kelimelerinden elde edilen özellikleri çıkararak sınıflandırma amacıyla farklı makine öğrenimi algoritmalarına girdi sağlamaktadır. Bu tür yöntemler, özellikle harfler ve sayılar gibi statik işaretlerin tanınmasında yoğun şekilde kullanılmıştır. Bununla birlikte, izole işaret dili tanıma görevine yönelik, bu temel yaklaşımların üzerine inşa edilmiş ve daha gelişmiş teknikleri benimseyen

alıřmalar da literatürde yer almaktadır. Bu yeni yöntemler, statik işaretlerin ötesine geçerek daha karmařık ve dinamik işaretlerin tanınmasını hedeflemektedir.

Madani ve Nahvi (2013), izleme işlemleri için CAMSHIFT (Continuously Adaptive Mean Shift) algoritmasını, özellik çıkarımı için ise Radon dönüşümünü kullanarak geliřtirdikleri yöntemle, 20 izole Pakistan İşaret Dili kelimesini %95,56 doğruluk değeriyle sınıflandırmayı başarmışlardır. Benzer şekilde, Li vd. (2016), Tayvan İşaret Dili'nden 11 kelimenin tanınması amacıyla, sınıflandırıcı olarak Gizli Markov Modeli'ni (HMM) kullanmış ve entropi tabanlı bir K-Ortalamalar Algoritması ile TBA'yı (Temel Bileşen Analizi) birleřtirerek gereksiz boyutları ortadan kaldırmıştır. Bu alıřmada optimizasyon için Yapay Arı Kolonisi Algoritması (YAKA) ve Baum-Welch algoritmalarını da uygulayarak %91,30 doğruluk değeri elde etmişlerdir. Ahmed vd. (2017), Dinamik Zaman Bükme (Dynamic Time Warping - DTW) tabanlı bir el hareketi tanıma sistemi geliřtirmiş, bu sistemle hem ellerin hem de yüzün özelliklerini analiz ederek Hint İşaret Dili'nden 24 işareti tanımada %90 doğruluk sağlamıştır. Ronchetti vd. (2016a), konum, hareket ve el řekli gibi çeřitli özelliklere dayalı olasılıksal bir model önererek, LSA64 veri kümesinde (64 sınıflı) %97 doğruluk değeri ulařmış ve işaretçiden bağımsız testlerde %91,7 doğruluk sağlamıştır. Rodríguez ve Martínez (2018) ise kümülatif řekil farkı ve Destek Vektör Makinesi (DVM) tabanlı bir model tasarlamış ve LSA64 veri kümesinde %85 doğruluk değeri elde etmiştir. Özdemir vd. (2020), işaret dilini tanıma sürecinde Geliřtirilmiş Yoğun Yörüngeler (Improved Dense Trajectory - IDT) yöntemini kullanarak Türk İşaret Dili'nde 744 farklı işareti içeren BosphorusSign22k veri setinde %88,53 doğruluk değeri ulařmıştır. Bu yöntemde HOG (Histogram of Oriented Gradients), HOF (Histogram of Optical Flow) ve MBH (Motion Boundary Histogram) gibi özellikler, işaret diline özgü karmařık el hareketleri ve vücut dili gibi unsurları analiz etmek için kullanılmıştır. Geleneksel makine öğrenimi tekniklerine dayanan bu alıřmalar, İDT alanında önemli ilerlemeler sağlamış olsa da, sınıflandırılan kelime sayısı arttıkça doğruluk oranlarının düşmesi, işaret dilinin karmařık ve dinamik doğasını tam anlamıyla modelleyememe sorununa işaret etmektedir. Ayrıca, bu yöntemler genellikle el yapımı özellik mühendisliğı gerektirir ve bu süreç uzmanlık gerektirerek, algoritmalar için uygun özelliklerin seçilmesi amacıyla ayrıntılı veri analizi yapılmasını zorunlu kılar. Veri kümesinin çeřitliliğı arttıkça, yeni ve daha gelişmiş

yöntemlerin geliştirilmesi gerekliliği ortaya çıkmaktadır. Derin öğrenme tekniklerinin gelişmesiyle birlikte, özellik çıkarma sürecinin otomatikleşmesi, bu tür manuel mühendislik çabalarını önemli ölçüde azaltmıştır.

Benzer şekilde, Damaneh vd. (2023), statik el hareketlerini tanımak için ESA, Gabor Filtresi ve ORB (Oriented FAST and Rotated BRIEF) özellik tanımlayıcısını birleştiren hibrit bir yöntem önermiştir. Çalışmaları, Massey, Amerikan İşaret Dili Alfabeti ve Amerikan İşaret Dili veri kümelerinde sırasıyla %99,92; %99,80 ve %99,80 doğruluk oranlarına ulaşmıştır. Das vd. (2023), Bangladeş İşaret Dilindeki rakamları ve karakterleri tanımak için ESA ve Rastgele Orman sınıflandırıcılarını birleştiren bir yöntem kullanarak sırasıyla %97,33 ve %91,67 doğruluk oranları elde etmiştir. Aldhahri vd. (2023) ise, Arap İşaret Dilindeki harfleri tanımlayan bir ESA mimarisi geliştirerek %94,46 doğruluk sağlamıştır. Dinamik hareketlerin tanınmasına yönelik bir diğer çalışma, Ma vd. (2022) tarafından gerçekleştirilmiştir. J ve Z harfleri gibi zaman içindeki değişimleri içeren işaretleri tanımak için, İki Akışlı Karma Evrimsel Sinir Ağını kullanmış ve ResNet50 modeliyle Amerikan İşaret Dili Alfabeti veri kümesinde %97,57 doğruluk değerine ulaşmıştır. Alsharif vd. (2023), Amerikan İşaret Dili alfabetini sınıflandırmak için farklı derin öğrenme modellerini karşılaştırmıştır. AlexNet, ConvNeXt, EfficientNet, ResNet-50 ve Vision Transformer modellerini kullanarak sırasıyla %99,50; %99,51; %99,95; %99,98 ve %88,59 doğruluk oranları elde etmiştir. Bu sonuçlar, derin öğrenme tabanlı modellerin hem statik hem de dinamik işaretleri tanımda oldukça başarılı olduğunu göstermekte, aynı zamanda daha karmaşık ve çeşitli işaretlerin tanınması için yeni modellerin geliştirilmesine olanak tanımaktadır.

Ayrıca, 2 boyutlu Evrimsel Sinir Ağları (2D-ESA), harf ve rakam gibi statik işaretlerin tanınmasında önemli başarılar elde etmiştir, ancak zamansal bilgi içermemesi nedeniyle kelimeler ve cümleler gibi dinamik işaretlerin tanınmasında yetersiz kalmaktadır. İşaretlerin doğası gereği hem uzamsal hem de zamansal özellikler taşıması, araştırmacıları daha karmaşık ve entegre yöntemler geliştirmeye yöneltmiştir. Bu çerçevede hem uzamsal hem de zamansal boyutları birleştiren DÖ tabanlı modeller, İDT alanında önemli bir dönüşümü temsil etmektedir. Masood vd. (2018), ESA ile elde edilen uzamsal özellikleri, zamansal özellikleri öğrenmek için

Tekrarlayan Sinir Ağına (TSA) girdi olarak kullanmıştır. Bu model, LSA64 veri kümesinin bir alt kümesinde (46 sınıf) %95,2 doğruluk değeri sağlamıştır. Zhang ve Li (2019), dinamik videoların zamansal ve uzamsal özelliklerini değerlendirmek için 3 boyutlu ESA ile Evrişimsel Uzun Kısa Süreli Belleği birleştirerek MEMP (Multiple Extraction and Multiple Prediction) ağı adını verdikleri bir model geliştirmiş ve LSA64 veri kümesi üzerinde %99,063 doğruluk değerine ulaşmıştır. Imran ve Raman (2020), Hareket Tarihçesi Görüntüleri (HTG), RGB hareket görüntü şablonları ve dinamik görüntüler gibi farklı veri türlerini kullanarak önceden eğitilmiş ESA modellerine ince ayar yapmış ve bu modellerin sonuçlarını bir çekirdek tabanlı aşırı öğrenme makinesiyle birleştirmiştir. Bu yenilikçi füzyon metodolojisi, LSA64 veri kümesinde %97,81 doğruluk sağlamıştır. Özdemir vd. (2020), hem 2D hem de 3D evrişim katmanlarını birleştiren Karma Evrişimsel 3D Kalıntı Ağı (MC3) yöntemiyle, BosphorusSign22k veri kümesinde %78,85 doğruluk elde etmiştir. Benzer şekilde, Adaloglou vd. (2022), GSL veri kümesindeki 310 sınıfı sınıflandırmak için GoogLeNet+TConvs ve 3D-ResNet modellerini kullanmış ve işaretçiden bağımsız testlerde sırasıyla %86,03; %86,23 ve %89,74 doğruluk oranlarına ulaşmıştır. Daha sofistike modellerden biri olan Wang vd. (2021) tarafından önerilen (2+1)D-SLR ağı, 2D ve 3D evrişimlerin avantajlarını birleştirerek LSA64 veri kümesinde %98,7 doğruluk değerine ulaşmıştır. Sincan ve Keleş (2021) ise, I3D modelini RGB videolar üzerinde, ResNet-50 modelini ise RGB Hareket Tarihçesi Görüntüleri üzerinde eğitmiştir. Bu iki modelden elde edilen özellikleri geç füzyon yöntemiyle birleştirerek, BosphorusSign22k veri kümesinde %94,83 doğruluk değeri sağlamışlardır. Bu çalışmalar, zamansal ve uzamsal özellikleri birleştiren yöntemlerin, dinamik işaretlerin tanınmasında güçlü performans sergileyerek İDT sistemlerinin doğruluğunu ve genellenebilirliğini artırmada kritik bir rol oynadığını göstermektedir.

İDT alanındaki mevcut çalışmalar önemli başarılar sağlamış olsa da işaret dili videolarının karmaşıklığını ve çeşitliliğini tam anlamıyla modelleyememiştir. Özellikle, ellerin detaylı hareketleri ve pozisyonlarının yanı sıra yüz ifadeleri ve jestler gibi anlamın büyük bir kısmını taşıyan bileşenlere yeterince odaklanılmaması, sistem performansını sınırlamıştır. Bu eksiklikler, İDT sistemlerinin daha fazla geliştirilmesi için yeni yaklaşımların araştırılmasına yol açmıştır. Gökçe vd. (2020), el, vücut ve yüz görüntülerini kullanarak 18 katmanlı Karma Evrişimsel 3D Kalıntı Ağı (MC3-18)

modelini eğitmiş ve bu ağlardan elde edilen özellikleri skor seviyesinde birleştirmiştir. Bu yöntem, BosphorusSign22k test verisi üzerinde %94,94 doğruluk sağlamıştır. Gündüz ve Polat (2021) ise, yüz, eller ve vücut bölgelerinden elde edilen görüntüleri ve optik akış verilerini Inception3D modelinde kullanırken, UKSB ağı eğitiminde vücut ve el pozu verilerinden yararlanmıştır. Modellerden elde edilen özellik akışlarını iki katmanlı bir sinir ağına entegre ederek, 174 sınıflı BosphorusSign22k-genel veri kümesinde %89,3 doğruluk elde etmişlerdir. Jebali vd. (2024), ESA ve UKSB gibi derin öğrenme modellerini kullanarak manuel ve manuel olmayan bileşenlerden gelen bilgileri eş zamanlı işleyebilen bir sistem geliştirmiştir. Yüz ifadeleri gibi manuel olmayan özelliklerin dahil edilmesi, sistemin doğruluğunu önemli ölçüde artırmıştır. Önerilen sistem, 450 sınıflı bir veri kümesinde %90,12; 26 sınıflı bir veri kümesinde ise %94,87 doğruluk oranlarına ulaşmıştır. Bu çalışmalar, işaret dilindeki anlamın tüm yönlerini yakalayabilmek için hem manuel hem de manuel olmayan bileşenlere eş zamanlı olarak odaklanan yaklaşımların, İDT sistemlerinin performansını artırmada etkili olduğunu göstermektedir. Bu kapsamda, el hareketleri, yüz ifadeleri ve vücut pozisyonlarını bütüncül bir şekilde işleyen modellerin geliştirilmesi, İDT alanında gelecekteki ilerlemelere önemli katkılar sağlayacaktır.

İDT sistemlerinin performansını artırmak ve dayanıklılığını güçlendirmek amacıyla poz verilerine dayalı yaklaşımlar, son yıllarda yoğun bir şekilde araştırılmıştır. Poz verileri, arka plan varyasyonlarının ve işaretçi farklılıklarının etkisini azaltarak sistemlerin güvenilirliğini artırma potansiyeline sahiptir. Ellerin, yüz ifadelerinin ve vücut hareketlerinin ayrıntılı analizi, işaret dilinin anlamını daha doğru bir şekilde yakalamaya olanak sağlar. Konstantinidis vd. (2018b), RGB videolarından el ve vücut poz bilgilerini çıkararak bu verilerle eğitilen UKSB ağlarından elde edilen özellikleri geç füzyon yöntemiyle birleştiren bir model önermiştir. Bu model, LSA64 veri kümesinde %98,09 doğruluk elde etmiştir. Aynı yazarlar başka bir çalışmada (Konstantinidis vd., 2018a), RGB video ve optik akış bilgilerini analiz eden ek akışlar ekleyerek modelin doğruluğunu %99,84'e yükseltmişlerdir. Kindiroglu vd. (2019), BosphorusSign22k-genel veri kümesi üzerinde Zamansal Birikimli Özellikler (TAF) yöntemini kullanarak %81,58 doğruluk sağlamıştır. Daha sonra, değişken uzunluktaki videoları sabit uzunluktaki özelliklerle temsil etmek için Hizalanmış Zamansal Birikimli Özellikler (ATAF) yöntemini önermiş ve bu yöntemi MC3-18 modelleriyle

birleştirerek %94,9 doğruluk değerine ulaşmıştır. Selvaraj vd. (2021), GSL veri kümesinde UKSB, Transformer, ST-GCN ve SL-GCN modellerini kullanarak sırasıyla %86,6, %89,5, %93,5 ve %95,4 doğruluk oranlarına ulaşmıştır. Bu modeller, poz verilerini işleyerek işaret dilinin zamansal ve uzamsal boyutlarını etkili bir şekilde değerlendirmiştir. Samaan vd. (2022), MediaPipe tarafından elde edilen poz verilerini UKSB, çift yönlü UKSB ve GRU modelleriyle işleyerek on sınıflı bir veri kümesi üzerinde sırasıyla %99,6, %99,3 ve %100 doğruluk elde etmiştir. Ancak bu başarılar, daha geniş ve çeşitli veri kümelerine genelleştirme açısından sınırlı kalmıştır. Alyami vd. (2023), el ve yüz poz verilerini Dönüştürücü tabanlı bir modelle sınıflandırmış, LSA64 veri kümesinde işaretçiye bağlı ve işaretçiden bağımsız testlerde sırasıyla %98,25 ve %91,01 doğruluk oranlarına ulaşmıştır. Özdemir vd. (2023), ST-GCN ve MC-LSTM modellerini birleştirerek vücut poz verileri ile yüz ve ellerin jestsel bilgilerini entegre eden bir yöntem geliştirmiştir. Bu model, BosphorusSign22k veri kümesinde %92,58 doğruluk sağlamıştır. de Castro vd. (2023), ellerin ve yüzün bölümlere ayrılmış bölgelerini, eklem mesafelerini ve yapay derinlik haritalarını 3D-ESA ile işleyen birçok akışlı yöntem sunmuş ve 20 sınıflı bir veri kümesinde %91 doğruluk elde etmiştir. Hamza ve Wali (2023), veri artırma tekniklerini kullanarak sınırlı bir veri kümesi üzerinde 3D-ESA ile %93,33 doğruluk sağlamıştır. Laines vd. (2023), poz verilerini RGB görüntüsüne dönüştüren ve iskelet yapısını Ağaç Yapısı İskelet Görüntüsü (TSSI) ile temsil eden bir yöntem geliştirmiştir. DenseNet-121 modeliyle yapılan sınıflandırmalarda, 100, 226 ve 30 sınıflı veri kümelerinde sırasıyla %81,47, %93,13 ve %98 doğruluk oranlarına ulaşılmıştır. Bu yöntem, yüz ifadeleri ve el hareketleri gibi kritik bileşenleri dikkate alarak modelin doğruluğunu artırmıştır. Bu çalışmalar, poz verilerinin kullanımıyla İDT sistemlerinin doğruluğunu ve dayanıklılığını artırmanın mümkün olduğunu göstermektedir. Ellerin ve yüzün detaylı analiziyle işaret dilinin ince nüanslarını yakalamak, sistemlerin daha karmaşık ve çeşitli işaretleri tanıma yeteneğini geliştirmektedir.

Kim vd. (2023), evde yalnız yaşayan yaşlı bireyler için mobil robotlarla poz tahmini üzerine bir çalışma gerçekleştirmiştir. MediaPipe'in hafif ve hızlı 2D poz tahmin yeteneklerini kullanarak eklem açılarını doğru bir şekilde tahmin eden bir model geliştirmişlerdir. Model, hareket verilerini analiz ederek 3D poz tahmininde karşılaşılan zorlukları çözmeyi hedeflemiştir. Pratik senaryolarda test edilen sistem,

video görüntüleri üzerinden eklem açılarını başarılı bir şekilde tahmin etmiştir. Dill vd. (2023), MediaPipe Pose'un fiziksel egzersiz analizinde doğruluğunu inceleyen bir çalışma gerçekleştirmiştir. Dört katılımcıyla yapılan deneylerde, iki kamera ve bir hareket yakalama sistemi kullanılarak elde edilen veriler analiz edilmiştir. Sonuçlar, MediaPipe Pose'un doğruluğunun kameranın bakış açısına ve egzersiz türüne göre değişiklik gösterdiğini ortaya koymuştur. Uygun bakış açılarında yüksek doğruluk elde edilirken, öz-örtüşme ve elverişsiz açılarda doğruluk kayıpları gözlemlenmiştir. Çalışmada, çoklu kamera entegrasyonu ve biyomekanik modellerin doğruluğu artırabileceği belirtilmiştir. Garg vd. (2023), yoga duruşlarının sınıflandırılması için MediaPipe kütüphanesi kullanarak vücut anahtar noktalarını tespit etmiş ve iskeletleştirme işlemlerini gerçekleştirmiştir. Çalışma, iskeletleştirilmiş görüntülerle eğitilen derin öğrenme modellerinin, iskeletleştirilmemiş görüntülerden daha yüksek doğruluk sağladığını göstermiştir. Bu bulgu, iskelet temelli yaklaşımların yoga duruşlarının analizinde etkili olduğunu ortaya koymaktadır. Bora vd. (2023), Assamese İşaret Dili tanıma üzerine bir sistem geliştirmiştir. Çalışma kapsamında, Assamese el hareketlerini içeren iki ve üç boyutlu görüntülerden oluşan 2094 veri noktasına sahip bir veri seti hazırlanmıştır. MediaPipe, görüntülerden işaret noktalarını tespit etmek için kullanılmış ve ileri beslemeli yapay sinir ağı modeliyle dokuz statik kelimeyi sınıflandırmada %99 doğruluk elde edilmiştir. Singh vd. (2021), MediaPipe Holistic çözümünü insan-bilgisayar etkileşimi ve güvenlik uygulamaları için gerçek zamanlı yüz, el ve vücut pozlarını tahmin edebilen bir sistem geliştirmede kullanmıştır. Gerçek zamanlı video analiziyle insan hareketleri yakalanmış ve analiz edilmiştir. Çalışma, yüksek doğruluk değeriyle hızlı ve etkili bir çözüm sunarak insan-hareket analizinde önemli bir katkı sağlamıştır. Bu çalışmalar, MediaPipe ve benzeri araçların poz tahmini, insan hareket analizi ve işaret dili gibi farklı alanlardaki potansiyelini göstermekte ve bu sistemlerin, hızlı ve doğru sonuçlar elde etmek için değerli araçlar olduğunu ortaya koymaktadır.

Parashar vd. (2023), MediaPipe ve MoveNet modellerini entegre ederek yoga pozlarının tespiti ve sınıflandırılmasını gerçekleştirmiştir. Bu modellerin kombinasyonu, yoga pozlarının doğru şekilde iskeletleştirilmesi ve anahtar noktaların tespitinde önemli bir iyileşme sağlamıştır. Çalışmada, sağlık ve fitness uygulamaları için geniş bir potansiyele sahip olan sistem, çeşitli ve geniş yoga veri setleriyle

eđitilmiřtir. Bu yaklařım, yoga duruřlarının analizinde hem dođruluk hem de genelleme kapasitesini artırmıřtır. Jo vd. (2022) ise MoveNet, OpenPose ve PoseNet gibi poz tespiti modellerini karřılařtırmıřtır. alıřma sonuları, MoveNet'in zellikle mobil cihazlarda hız ve hesaplama verimliliđi aısından stn performans sergilediđini gstermiřtir. Buna karřın, OpenPose'un daha ayrıntılı anahtar nokta tespiti yapabilme avantajına sahip olduđu, ancak daha fazla kaynak tkettiđi belirtilmiřtir. MoveNet, hızlı ve kaynak aısından verimli bir zm sunarken, OpenPose'un yksek dođruluk gerektiren ayrıntılı uygulamalarda tercih edilebileceđi vurgulanmıřtır. Bu alıřmalar, poz tespiti alanında farklı modellerin gl ve zayıf ynlerini deđerlendirerek, uygulama gereksinimlerine uygun zmler geliřtirilmesine katkı sađlamaktadır. zellikle MediaPipe ve MoveNet gibi modellerin entegrasyonu, yoga gibi alanlarda kullanıcı dostu ve etkili zmler sunma potansiyelini gstermektedir.

Lin vd. (2024), yařlı bireylerin sađlık durumlarının takibi ve acil durum mdahaleleri iin gerek zamanlı dřme tespitine odaklanan bir alıřma gerekleřtirmiřtir. alıřmada, mevcut poz tanıma yntemlerinin byk model boyutları, zayıf zamanlılık ve oklu hedef pozları dođru řekilde tanıma gibi sorunlarını zmek amacıyla, hafif ve geliřtirilmiř bir OpenPose tabanlı algoritma nerilmiřtir. Bu algoritma, VGG-19 yerine daha hafif bir MobileNet ađı kullanarak model boyutunu ve hesaplama ykn nemli lde azaltmıřtır. Deneysel sonular, bu algoritmanın gerek zamanlı oklu kiři poz tespitinde bařarılı olduđunu ve uygulama gereksinimlerini karřıladıđını gstermektedir. Arkushin vd. (2023), HamNoSys iřaret dili notasyonunu iřaret hareket dizilerine dnřtren ilk yntem olan Ham2Pose'u geliřtirmiřtir. HamNoSys, dil bađımsızlıđı sađlayan evrensel bir iřaret dili transkripsiyon sistemidir. Bu alıřma, farklı iřaret dillerine uyumlu bir zm sunmayı hedeflemiřtir. Arařtırmada, HamNoSys metni ve bir referans poz erevesi alınarak dođru iřaret hareketlerini retecek bir model tasarlanmıřtır. Model, OpenPose tarafından ıkarılan anahtar noktalar zerinde eđitilmiřtir. Cao vd. (2017), OpenPose'u kullanarak insan vcut pozlarının oklu kiřiyle eř zamanlı tespiti zerine odaklanmıřtır. alıřma, Part Affinity Fields (PAFs) adlı bir teknik kullanarak insan vcudundaki paralar arasındaki iliřkileri đrenmiř ve bu paraları dođru bireylere atamayı bařarmıřtır. Bu yntem, zellikle oklu kiři takip ve poz analizi gerektiren uygulamalarda etkili olmuřtur. Badiola ve Mendez (2021), spor ve egzersizlerde insan vcut pozlarının

izlenmesi amacıyla OpenPose'un kullanıldığı çalışmaların sistematik bir incelemesini gerçekleştirmiştir. Çalışma, spor ve egzersiz uygulamalarında İnsan Poz Tahmini (Human Pose Estimation - HPE) yöntemlerinin nasıl geliştirildiğini ve OpenPose'un bu bağlamda nasıl optimize edilebileceğini incelemiştir. Özellikle spor performansını izleme ve analiz etme bağlamında OpenPose'un potansiyeli vurgulanmıştır. Bu çalışmalar, OpenPose'un insan poz tespiti ve hareket analizi konusundaki çok yönlü uygulama potansiyelini göstermektedir. Özellikle sağlık, spor, egzersiz ve işaret dili gibi farklı alanlarda OpenPose'un optimize edilerek kullanılması, ilgili sistemlerin doğruluk, hız ve kullanım kolaylığı açısından gelişimine önemli katkılar sağlamaktadır.

Song vd. (2024), masa tenisi teknik hareketlerini tanımak için iki model geliştirmiştir. Çalışma, dinamik ve statik karmaşık ağ analizlerini topluluk algılama algoritmalarıyla birleştirerek oyuncuların teknik ve taktiksel davranışlarını değerlendirmeyi amaçlamaktadır. 37 elit oyuncunun 8015 yüksek çözünürlüklü videosunu içeren iki veri seti oluşturulmuş ve bu videolar AlphaPose kullanılarak işlenmiştir. YOLOv8-AlphaPose tabanlı yapay zeka algoritmasının, diğer algoritmalara (LSTM, biGRU, CNN) kıyasla tanıma ve tahmin performansında üstün olduğu görülmüştür. Areerob vd. (2024), kişisel koruyucu ekipman (PPE) tespitini iyileştirmek için dikkat tabanlı YOLOv7 ve insan duruş tahmini için AlphaPose'u entegre ederek bir sistem geliştirmiştir. Bu sistem, kalabalık sahnelerde PPE'nin doğru bir şekilde konumlandırılıp konumlandırılmadığını insan poz tahmini yoluyla doğrulamış ve yanlış tespitleri azaltmıştır. AlphaPose'un kalabalık sahnelerde yüksek doğruluk sağladığı ve mevcut yöntemlerin sınırlamalarını aştığı belirtilmiştir. Fang vd. (2021), AlphaPose kullanarak vücut, yüz, el ve ayak anahtar noktalarının eşzamanlı tespiti ve izlenmesi üzerine odaklanmıştır. Önerilen yöntem, COCO-WholeBody ve Halpe-FullBody gibi geniş veri setlerinde test edilmiş ve mevcut yöntemlerden daha hızlı ve daha doğru sonuçlar elde etmiştir. Zhu vd. (2020), insan vücudunun 133 anahtar noktasını (vücut, yüz, el, ayak) içeren tam vücut poz tahmini üzerinde çalışmıştır. AlphaPose tabanlı bu çalışma, eller ve yüz gibi daha ince ayrıntıları algılamayı hedeflemiştir. YOLO insan algılayıcısı kullanılarak geliştirilen sistem, MSCOCO2017 veri setinde yüksek verimlilik ve doğruluk sağlamıştır. Bu çalışmalar, AlphaPose'un geniş bir uygulama yelpazesinde (spor analitiği, kişisel koruyucu ekipman tespiti,

insan hareket analizi) kullanılabilirliğini ve etkinliğini göstermektedir. AlphaPose'un sağladığı yüksek doğruluk, hızlı işlem kapasitesi ve detaylı anahtar nokta tespiti, ilgili sistemlerin performansını artırarak gerçek dünya uygulamalarına önemli katkılar sunmaktadır.

Bibin (2024), insan hareketlerinin tanınması için Detectron2'yi LSTM modeliyle entegre eden bir yaklaşım sunmuştur. Çalışmada, videolardaki her karede tespit edilen anahtar noktalar Detectron2 kullanılarak çıkarılmış ve bu veriler LSTM modeline aktarılmıştır. Model, analiz edilen hareketlerde %91'in üzerinde bir doğruluk değeri elde etmiş ve gerçek zamanlı hareket tanıma sistemlerinin geliştirilmesine önemli bir katkı sağlamıştır. Önerilen sistem, özellikle spor analitiği ve güvenlik uygulamaları gibi alanlarda potansiyel kullanım imkânı sunmaktadır. Lee vd. (2024), yanlış duruş alışkanlıklarından biri olan baş öne duruşunu (FHP- Forward Head Position) tespit etmek için bir sistem geliştirmiştir. Mevcut 2D insan duruş tahmini modellerinin, FHP tanısında kritik önem taşıyan C7 omuru ve kraniyovertebral açığı ölçme yetersizliği bu alandaki temel zorluklardan biri olarak tanımlanmıştır. Detectron2 tabanlı olarak geliştirilen sistem, bu eksikliği gidererek gerçek zamanlı ve etkili bir FHP tespiti sağlamaktadır. Davoudi Khaskoli vd. (2024), süper düşük çözünürlüklü görüntülerde nesne tespiti için Detectron2 tabanlı bir yöntem geliştirmiştir. Çalışmada, kişilerin algılanmasını sağlamak amacıyla sentetik veri setleri oluşturulmuş ve Faster R-CNN ile RetinaNet modelleri eğitilmiştir. COCO değerlendirme metrikleriyle test edilen modeller arasında RetinaNet'in daha yüksek doğruluk değeri sağladığı tespit edilmiştir. Önerilen yöntem, düşük çözünürlüklü senaryolar için optimize edilmiş bir çözüm sunarak farklı çözünürlüklerdeki verilerde başarılı sonuçlar elde etmiştir. Hernandez vd. (2021), insan-robot etkileşimi bağlamında, özellikle rehabilitasyon süreçlerinde Detectron2'yi uygulamışlardır. Bu çalışmada, omuz ve dirsek açılarının tespiti hedeflenmiş ve sistem, üst ekstremité rehabilitasyon hareketlerini değerlendirmede başarılı bulunmuştur. Ancak, Detectron2'nin doğruluk değerinin OpenPose gibi alternatif modellere kıyasla daha düşük olduğu belirtilmiştir. Bu çalışmalar, Detectron2'nin insan hareket analizi ve nesne tespiti gibi farklı uygulamalarda kullanılabilirliğini göstermekte olup, sistemlerin çözünürlük, hız ve doğruluk açısından optimize edilebileceğini ortaya koymaktadır.

Li vd. (2024), HRNet'in karmaşık yapısını hafifleterek sınırlı kaynaklara sahip cihazlarda daha verimli kullanılmasını amaçlayan bir model geliştirmiştir. Model, tıbbi rehabilitasyon ve insan-bilgisayar etkileşimi gibi alanlarda uygulanmıştır. Convolutional Block Attention Module (CBAM) kullanılarak ağırlık hafifliği korunmuş ve aynı zamanda yüksek doğruluk oranları elde edilmiştir. COCO 2017 ve MPII veri setlerinde test edilen model, yüksek performans sergilemiştir. Wang vd. (2024), HRNet'in bir varyasyonu olan HR-xNet adlı yüksek çözünürlüklü bir poz tahmin algoritması sunmuştur. HR-xNet, HRNet'in özellik çıkarım süreçlerini optimize ederek modelin kaynak tüketimini azaltmıştır. Düşük seviyeli detaylı özelliklerle yüksek seviyeli anlamsal özelliklerin birleşimini sağlayan model, küçük anahtar noktalar ve değişken insan pozları gibi zorlukların üstesinden başarıyla gelmiştir. Bu özellik, kaybolan anahtar noktaların doğru şekilde tespit edilmesine olanak tanımıştır. Bao vd. (2023), skyjump spor dalında hızlı hareket eden sporcuların analizini yapmak için HRNet'i Efficient Channel Attention (ECA) modülüyle birleştirmiştir. Bu birleşim, sporcuların poz tespitinde daha doğru anahtar noktaların belirlenmesini sağlamış ve HRNet'in doğruluğunu artırmıştır. Zhou vd. (2022), HRNet'in ağırlıklı hesaplama maliyetini azaltmak amacıyla X-HRNet modelini geliştirmiştir. X-HRNet, özellikle küçük boyutlu kişilerin yer aldığı karmaşık sahnelerde ölçek farklılıklarını daha iyi yönetmek için Spatially Unidimensional Self-Attention (SUSA) modülünü kullanmıştır. Model, COCO val2017 veri setinde yapılan testlerde, diğer hafif ağlara kıyasla rekabetçi bir performans sergilemiştir. Bu çalışmalar, HRNet ve türevlerinin farklı alanlardaki etkinliğini artırmak için optimizasyon ve modül entegrasyonu gibi tekniklerin önemini vurgulamaktadır. Hafif ağ yapıları ve dikkat modüllerinin kullanımı, sınırlı kaynaklara sahip cihazlarda yüksek doğrulukla çalışan sistemlerin geliştirilmesine katkı sağlamaktadır.

Chen ve Fisher (2024), yaşlı bireylerin evde hareketsizlik durumlarını izlemek amacıyla RGB-D kamera tabanlı bir sistem geliştirmiştir. Çalışma, düşük ışık koşulları ve TV ışığı değişimleri gibi zorlu çevresel faktörlere rağmen hareketsizlik algılamada yüksek doğruluk ve hassasiyet sağlamayı hedeflemiştir. ViTPose ve RAFT yöntemlerini kullanan sistem, çevresel değişimlere dayanıklılık göstererek çeşitli senaryolarda güçlü bir performans sergilemiştir. Chen ve Tan (2024), ViTPose'i temel olarak geliştirdikleri FanPose isimli bir insan poz tahmin modeli tanıtmışlardır.

Modelin performansını artırmak için iki temel iyileştirme yapılmıştır: İlk olarak, giriş verileri arasındaki bağlantıları güçlendiren gelişmiş bir dikkat mekanizması entegre edilmiştir. İkinci olarak, anahtar nokta yerelleştirme doğruluğunu artırmak için geleneksel Gauss çekirdekleri yerine Laplacian çekirdekleri kullanılmıştır. Bu yenilikler, modelin tanıma hassasiyetini belirgin şekilde artırmıştır. Xu vd. (2023), Vision Transformer mimarisine dayalı ViTPose++ modelini geliştirmiştir. Bu model, hem Top-Down hem de Bottom-Up tarama yöntemlerine uyarlanabilir şekilde tasarlanmıştır ve 20M ila 1B parametre arasında ölçeklenebilir bir yapı sunmaktadır. İnsan ve hayvan eklem noktalarını tespit etme kapasitesine sahip olan model, MAE pre-training yöntemiyle transfer öğrenmesini geliştirmiş ve çok görevli öğrenme uygulamalarında etkili sonuçlar elde etmiştir. Üstek vd. (2023), akıllı havalimanlarında gerçek zamanlı şiddet tespiti için iki aşamalı bir çerçeve önermiştir. İlk aşamada, ViTPose kullanılarak insan eklem noktaları tespit edilmiştir. İkinci aşamada ise, CNN-BiLSTM tabanlı bir sınıflandırıcı, bu eklem noktalarından elde edilen mekânsal ve zamansal bilgileri analiz ederek şiddet içeren davranışları sınıflandırmıştır. Sistem, şiddetli ve şiddetsiz davranışları yüksek doğrulukla ayırt edebilmiştir. Bu çalışmalar, ViTPose ve türevlerinin insan poz tahmini ve davranış analizi gibi çeşitli uygulamalarda etkili sonuçlar elde edebileceğini göstermektedir. Geliştirilen modeller, hem karmaşık çevresel faktörlere dayanıklılık göstermekte hem de daha doğru ve ölçeklenebilir tahminler yaparak farklı alanlarda geniş bir kullanım potansiyeline sahiptir.

Yukarıda özetlenen önceki çalışmalarda da görüldüğü üzere, insan poz tahmini ve hareket analizleri için farklı kütüphaneler ve modeller kullanılmıştır. MediaPipe, MoveNet, OpenPose, AlphaPose, Detectron2 ve HRNet gibi kütüphaneler, insan vücudunun anahtar noktalarının doğru ve hızlı bir şekilde tespit edilmesi için yaygın olarak kullanılmaktadır. Bu kütüphaneler, özellikle sağlık, güvenlik ve spor alanlarında önemli avantajlar sunmakla birlikte, her birinin kendine özgü avantaj ve dezavantajları bulunmaktadır. İşaret dili tanıma uygulamalarında ise her modelin farklı avantajlar sunduğu görülmektedir. MediaPipe Holistic, el ve parmak eklemlerini yüksek doğrulukla tespit etme yeteneği ve düşük donanım gereksinimi ile mobil ve gerçek zamanlı işaret dili uygulamaları için idealdir. MediaPipe Pose'un bilek noktasını tespit ederek elin konumunu belirlemesi ve ardından el üzerindeki anahtar

noktaları MediaPipe Hand çözümü ile yüksek doğrulukla analiz etmesi, bu kütüphaneyi işaret dili tanıma sistemleri için öne çıkarır (Torun ve Karacı, 2024).



3. İŞARET DİLİ VE TANIMA SİSTEMLERİ

3.1 İşaret dili ve İletişim

İşaret dili, el, kol, yüz hareketleri ve vücut jestlerinin kullanılması yoluyla iletişim kurulmasını sağlayan bir dildir. Bu dil, işitme engelli bireyler arasında ve işitme engelli bireylerle işitme yeteneği olan kişiler arasında etkin bir iletişim aracıdır. İşaret dili, işitme engellilerin yaşamında hem kişisel hem de toplumsal boyutlarda büyük bir öneme sahiptir. İşitme duyusu aracılığıyla bilgi alışverişi yapamayan bireyler için, işaret dili dünya ile bağlarını kurmalarına ve kendilerini ifade etmelerine olanak tanır. (Georgiou, 2023)

Günlük yaşamda işaret dili, işitme engellilerin eğitim, iş ve sosyal çevrelerde etkin bir şekilde yer alabilmesini sağlar. İşitme engelli bireyler, işaret dili aracılığıyla düşüncelerini ve duygularını daha rahat bir şekilde ifade edebilir. Ancak, işitme engelli olmayan bireylerle iletişim kurarken karşılaştıkları zorluklar sıkça gözlemlenmektedir. İşaret dili kullanıcıları sağlık hizmetlerine erişimde çeşitli zorluklarla karşılaşmaktadır. Örneğin, diş hekimliği ortamlarında maske kullanımı, arka plan gürültüsü ve bekleme odasından çağrılma gibi iletişim engelleri bulunmaktadır. Ayrıca, işitme engelli bireylerin genel sağlık durumlarının daha kötü olduğu ve kronik hastalıkların teşhis ve tedavisinde eksiklikler yaşandığı belirtilmektedir (Rezende vd., 2024). Bu nedenle, işaret dilinin yaygınlaştırılması ve daha fazla kişinin bu dili öğrenmesi, toplumdaki iletişim bariyerlerini azaltabilir.

İşaret dili, her ne kadar öğrenilmesi kolay olmayan bir dil olarak görülse de bu dili öğrenen bireyler hem kendi topluluklarında hem de daha geniş bir sosyal ağda etkili bir şekilde iletişim kurabilir. Bu durum, işitme engelli bireylerin topluma daha aktif katılımlarına ve özgün kimliklerini ortaya koymalarına katkı sağlar (Sömen, 2023).

İşaret dili, sözlü dillerden farklı olarak, kendine özgü gramer yapısı, sözcük sırası ve jest kullanımlarına sahiptir. Bu farklılıklar, işitme engelli bireylerin işitme yeteneği olan bireylerle iletişim kurması sırasında zorluklar oluşturabilir. Aynı zamanda, işaret dili, zaman içinde değişen ve gelişen bir yapıya sahiptir. Bu nedenle, işaret dili

kullanan bireyler arasında tutarlılık sağlamak için sürekli eğitim ve öğrenme süreci gereklidir (Erten ve Arıcı, 2022).

İşaret dilinin toplumsal farkındalığının artması, işitme engelli bireylerin yaşam kalitesini doğrudan etkiler. Özellikle eğitim kurumlarında işaret dili eğitiminin yaygınlaştırılması ve kamu hizmetlerinde işaret dili tercümanlarının bulunması, bu bireylerin toplumla olan bağlarını güçlendirebilir. Aynı zamanda, işaret dili öğrenen işitme engeli olmayan bireyler, hem işitme engelli topluluğuyla daha iyi bir iletişim kurabilir hem de kapsayıcı bir toplum oluşturulmasına katkı sağlayabilir (Akmeşe, 2016).

İşaret dili, işitme engelli bireylerin iletişim becerilerini geliştirerek, eğitimden sosyal yaşama kadar pek çok alanda aktif rol almalarını sağlar. Aynı zamanda, bu dilin toplumda daha fazla tanınması ve desteklenmesi, işitme engellilerin yaşam kalitesini artıracak önemli bir adımdır. Bu nedenle, işaret diline yönelik eğitim programları ve farkındalık kampanyalarının artması hem bireysel hem de toplumsal düzeyde büyük faydalar sağlayacaktır.

3.2 Türkçe İşaret Dili (TİD) Alfabeti

TİD alfabesi, işitme engelli bireylerin iletişim ihtiyaçlarını karşılamak amacıyla geliştirilmiş bir görsel-işitsel iletişim sistemidir. Bu alfabe, el hareketleri, parmak pozisyonları ve mimiklerin bir araya getirilmesiyle oluşturulur. Türk alfabesindeki 29 harf ile uyumlu olan TİD alfabesi, harf ve rakamların her biri için ayrı bir işaret tanımlar ve bu işaretler birbirleriyle karıştırılmayacak şekilde tasarlanmıştır. Bu sistem, işitme engelli bireylerin dil becerilerini geliştirmelerinde ve çevreleriyle etkili bir iletişim kurmalarında kritik bir rol oynar (Dikyuva vd., 2023 s. 57).

TİD alfabesi, görsel bir iletişim sistemi olarak, yalnızca el hareketlerini değil, aynı zamanda parmakların konumları ve yüz mimiklerini de içerir. Örneğin, "A" harfi, avuç içi yukarı dönük bir el pozisyonuyla ifade edilirken, "B" harfi işaret parmağı ve orta parmağın yan yana tutulduğu, diğer parmakların ise kapalı olduğu bir el şekliyle gösterilir. Rakamlar için de benzer şekilde belirgin işaretler mevcuttur. Bu işaretlerin

öğrenilmesi, dilin kavranması için kritik önem taşır ve özellikle işaret dili tercümanları için temel bir araçtır.

TİD alfabesi yalnızca işitme engelli bireylerin iletişim kurmalarını sağlamakla kalmaz, aynı zamanda bu dili öğrenmek isteyen bireyler için de etkili bir öğretim aracıdır. İşaret dili tercümanları, TİD alfabesini kullanarak işaretleri yazılı veya sözlü metne dönüştürebilir, böylece işitme engelli bireylerin çevreleriyle daha etkin bir iletişim kurmalarına yardımcı olabilirler. Bu, toplumsal entegrasyonun artmasını sağlarken, aynı zamanda dil bariyerlerini de ortadan kaldırır.

3.3 Türkçe İşaret Dili

TİD, Osmanlı İmparatorluğu dönemine kadar uzanan köklü bir geçmişe sahiptir. Ancak TİD'nin resmi olarak tanınması ve sistematik bir şekilde öğretilmesi, Türkiye Cumhuriyeti'nin eğitim ve sosyal politikalarındaki değişimlerle mümkün olmuştur. 2005 yılında Millî Eğitim Bakanlığı, işaret dilinin eğitim sisteminde yer almasını sağlayan bir kararname yayımlamış ve böylece TİD'nin resmi statüsü güçlenmiştir. Bu adım, işitme engelli bireylerin iletişim ve toplumsal entegrasyon süreçlerini desteklemek için önemli bir dönüm noktasıdır.

TİD, başlangıçta doğal işaretlerden oluşan ve herhangi bir standartlaşmaya sahip olmayan bir sistemdi. Ancak işitme engelli topluluğunun büyümesi ve eğitim ihtiyaçlarının artması, dilin belirli bir standart kazanmasını zorunlu hale getirmiştir. 1955 yılında işitme engelli bireyler tarafından hazırlanan ilk Türk İşaret Dili Sözlüğü, bu standartlaşma sürecinin önemli bir adımıdır. 1960'lı yıllarda İstanbul Üniversitesi'nde açılan İşaret Dili Bölümü, TİD'nin akademik incelemeleri ve gelişimi açısından bir dönüm noktası olmuş, dilin sistematik bir şekilde araştırılmasını sağlamıştır (Zeshan, 2006).

Günümüzde TİD, işitme engelli bireylerin günlük yaşamlarında iletişim kurmalarını sağlayan temel bir araç olmanın ötesinde, onların toplumsal hayata katılımını destekleyen güçlü bir araçtır. TİD, el hareketleri, yüz ifadeleri ve vücut duruşunu bir araya getiren görsel bir dil olarak, kullanıcılarına duygu ve düşüncelerini etkili bir şekilde ifade etme imkânı sunar. Eğitim alanında ise, işitme engelli bireyler için

sağlanan TİD eğitimi, bireylerin akademik başarılarını artırmakta ve sosyal hayata entegrasyonlarını kolaylaştırmaktadır. Ayrıca, TİD'nin yaygınlaştırılması ve toplumda bilinirliğinin artırılması, işitme engelli bireylerin daha kapsayıcı bir toplumda yaşamalarını mümkün kılmaktadır.

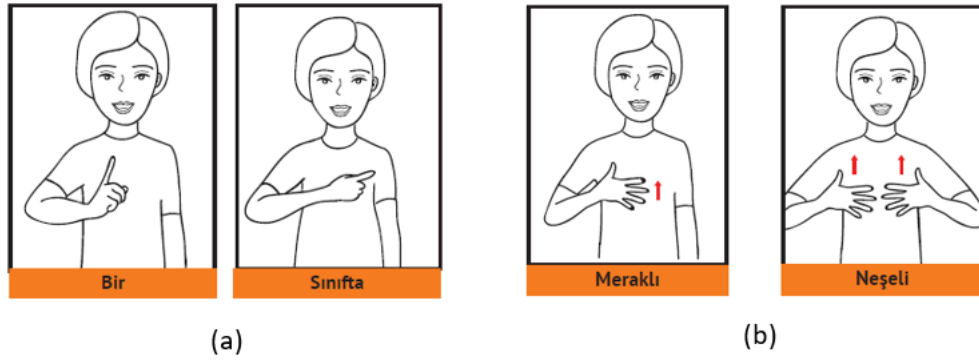
3.4 Uluslararası İşaret Dili (International Sign Language – ISL) Alfabetesi

Uluslararası İşaret Dili (ISL) alfabetesi, çeşitli dilsel çerçevelerde sohbet eden işitme engelli bireyler arasındaki iletişimi geliştirmeyi amaçlayan titizlikle geliştirilmiş bir sistem oluşturur. Bu alfabede, her harf için ayrı bir işaret belirlenir ve bu işaretler belirli el hareketleri, mekansal konfigürasyonlar ve yüz ifadeleri yoluyla ifade edilir. ISL alfabetesi, dünya çapında işitme engelli bireyler arasında etkili iletişimi teşvik etmek için standartlaştırılmış bir çerçeve görevi görür.

ISL alfabetesi, temelini Fransız İşaret Dili (Langue des Signes Française - LSF) alfabetesinden alan uluslararası bir dil sistemi olarak tanınır. LSF alfabetesi Fransa'daki işitme engelli topluluklar tarafından kullanılmaktadır ve ISL alfabetesi sistematik olarak bu dilsel modaliteden türetilmiştir. Toplam 26 harften oluşan ISL alfabetesi, her harf için benzersiz bir işaret sağlar. Bu işaretler hassas el hareketleri, mekansal yönelimler ve yüz ifadeleri ile üretilir. Örneğin, “A” harfine karşılık gelen işaret, işaret parmaklarını ve orta parmakları yan yana dik bir yönde konumlandırılarak yürütülür. “B” harfi durumunda, kalan üç parmak uzatılırken işaret parmakları ve orta parmaklar birleştirilir. ISL alfabetesi, işitme engelli bireyleri küresel iletişim kurma konusunda güçlendirir. Ayrıca, işaret dili tercümanları ve eğitimciler, ISL alfabetesini kullanarak diyalogları yazılı biçimlere aktarabilir ve işaret dili öğrenmekle ilgilenen bireyler için değerli bir kaynak haline getirebilir. Uluslararası İşaret Dili (ISL) alfabetesi, çeşitli dilsel geçmişlere sahip işitme engelli bireyler arasında iletişimi kolaylaştırmak için geliştirilmiş önemli bir sistemi temsil eder ve yaklaşık 15.000.000 kişi tarafından kullanıldığı belgelenmiştir. Alfabe çerçevesinde, her harf, belirlenmiş el hareketleri, mekansal düzenlemeler ve yüz ifadeleri ile ifade edilen belirli bir işaretle ilişkilendirilir. ISL alfabetesi, dünya çapında işitme engelli bireyler arasında etkili iletişim sağlamayı amaçlayan küresel bir standart olarak işlev görür (Çelik, 2019, s. 18-19).

3.5 Statik ve Dinamik İşaret Dili Kelimeleri

Son yıllarda, işaret dilinin tanınmasında, özellikle hesaplama yeteneklerinin geliştirilmesi ve üstün algoritmaların tanıtılmasıyla kolaylaştırılan önemli ilerlemeler kaydedilmiştir. İşaret dili kelimeleri, kelimeyi gösterirken yapılan statik ve dinamik hareketlere göre statik kelimeler ve dinamik kelimeler olarak ayrılır. Şekil 2.1’de örnek olarak gösterilen statik kelimelerde, hareket eksikliği ile karakterize edilen ellerin şeklini ve konfigürasyonunu vurgular. El ve kollar sabit konumda tutulur. Dinamik kelimeler ise, Şekil 3.1’de gösterildiği gibi ellerin tüm geçici hareketlerine odaklanır. El hareketlerindeki geçici varyasyonları analiz etme gerekliliği göz önüne alındığında, dinamik hareketlerin tespiti, statik hareketlerinkine kıyasla daha büyük bir zorluk sunar (Khotimah, 2018).



Şekil 3.1 (a) Statik ve (b) Dinamik Kelime Örnekleri (MEB TİD Öğretim Materyali, 2018)

Özellikle evrişimli sinir ağlarının ortaya çıkmasıyla, statik işaretlerin tanınması sağlanmış ve modellenmesi kolaylaşmıştır. Bununla birlikte, dinamik işaretler anlamlarını yüz, el ve vücut hareketlerindeki zamana bağlı varyasyonlardan aldığından, bu işaretleri etkili bir şekilde temsil etmek için yeni modeller geliştirmek zorunludur. Son zamanlarda, dinamik hareket tanıma için evrişimli sinir ağlarının kullanımının yanı sıra dinamik hareketleri modellemek için evrişimli sinir ağlarını uzun kısa süreli bellek modelleriyle entegre eden çok modlu yaklaşımlar kullanılmıştır. Ayrıca, çevresel ve bireye bağımlı değişkenlerin yarattığı zorluklara yanıt olarak, iskelet tabanlı işaret dili tanıma modelleri son yıllarda ortaya çıkmıştır. Bu metodoloji, başlangıçta bir insan pozu tahmin modeli aracılığıyla işlenen videolardan iskelet kilit noktalarının çıkarılmasını içerir. Tahmin daha sonra edinilen iskelet anahtar noktaları kullanılarak gerçekleştirilir.

3.6 İşaret Dili Tanıma Sistemleri

İşaret dili tanıma sistemleri, işitme engelli bireyler ile genel toplum arasındaki iletişim engelini aşmak için önemli bir araç haline gelmiştir. Bu sistemler, çeşitli sensör, kamera ve yapay zeka teknolojilerini kullanarak işaret dilini otomatik olarak algılayıp yorumlamayı amaçlamaktadır.

İşaret dili tanıma sistemleri genellikle görüntü işleme, elektromiyografi (EMG) sensörleri ve atalet ölçüm birimleri (IMU) gibi farklı veri toplama yöntemlerini kullanmaktadır (Amor vd., 2023; Khomami ve Shamekhi, 2020). Yüzey EMG ve IMU sensörlerinin birleştirilmesi, işaretleri yakalamada sistemin doğruluğunu artırabilmektedir (Khomami & Shamekhi, 2020). Ayrıca, iskelet tabanlı yaklaşımlar da son zamanlarda popülerlik kazanmıştır (Miah vd., 2023). Özellikle derin öğrenme teknikleri, CNN ve LSTM gibi modeller kullanılarak yüksek doğruluk oranları elde edilmiştir (Amor vd., 2023).

Bununla birlikte, SLR sistemleri hala bazı zorluklarla karşı karşıyadır. Karmaşık arka planlar, ışık koşulları ve kişiler arası farklılıklar gibi faktörler sistemlerin performansını etkileyebilmektedir (Miah vd., 2023). Ayrıca, işaret dillerinin ülkeden ülkeye değişmesi ve bazı dillerin (örneğin Hint İşaret Dili) daha karmaşık olması da zorluklar arasındadır (Das vd., 2022).

3.6.1 Kinect Sensör V2

Kinect, Xbox oyun konsolu için Microsoft tarafından üretilmiş ve insan doğal vücut hareketlerini bir girdi olarak alabilen hareket algılama aparatıdır. Bir dizi mikrofon ile birlikte renk ve kızılötesi (IR) gibi çeşitli sensörlerden oluşur. IR ışıklarını nesnelerin üzerine dökerek ve her bir ışının, sensörün IR alıcısı tarafından alınma zamanını hesaplayarak, 3D ortamında hareket algılama teknolojisini mümkün kılan bir derinlik haritası çizmektedir.

Kinect v2, Microsoft tarafından 2014 yılında piyasaya sürülen Kinect sensörünün ikinci versiyonudur (Şekil 3.2). Bir önceki modele göre birtakım iyileştirmeler yapılmıştır; Sensör, verileri saniyede iki gigabit ile daha hassas bir şekilde işleyebilir;

derinliđi ve kızılötesi sensör çözünürlüğü 512x424'e yükseltildi ve renk sensörü saniyede 30 kare (fps) hızında çalışan 1080p çözünürlüklü bir alıcı içermektedir. Sensörün algılayabildiđi iskelet eklemlerinin sayısı da 20'den 25'e çıkarılmıştır. Ayrıca, eşzamanlı kullanıcı saptama sayısı da ilk ikiden altıya çıkarılarak kameranın görüş alanı artırılmıştır. Bu durum kullanıcıların daha küçük bir alanda ve sensöre daha yakın çalışabilmelerini sağlamıştır (Tölgyessy vd., (2021).



Şekil 3.2 Microsoft Kinect V2

Bir işaret dili tanıma sistemi genellikle veri toplama, özellik çıkarımı, hareket tanıma ve çeviri gibi adımlardan oluşur. Kinect v2'nin RGB ve derinlik sensörleri, bu süreçlerde çeşitli veri türlerini sağlamaktadır. Hareketlerin görsel ve derinlik verileri kullanılarak kaydedilmesi, özellik çıkarım aşamasında kritik bir başlangıç noktası oluşturur. El, parmak ve vücut pozisyonlarının izlenmesi hem poz hem de hareket tabanlı özelliklerin çıkarılmasına olanak sağlar. Bu veriler, makine öğrenimi ve derin öğrenme algoritmaları kullanılarak analiz edilir. Örneğin, Destek Vektör Makineleri (SVM), K-En Yakın Komşu (KNN) gibi geleneksel yöntemlerin yanı sıra Long Short-Term Memory (LSTM) gibi zaman serisi analizine uygun derin öğrenme modelleri, işaret dili ifadelerindeki karmaşık hareket sekanslarını tanımlamak için kullanılmaktadır (Kaur vd., 2021).

Kinect v2 tabanlı işaret dili tanıma sistemleri, erişilebilirlik açısından önemli avantajlar sunmaktadır. Cihazın temassız algılama özelliđi, kullanıcı konforunu artırmakta ve geniş bir uygulama yelpazesi sunmaktadır. Eğitim, çeviri hizmetleri ve kamu erişimi gibi alanlarda işaret dili tanıma teknolojilerinin kullanımı, işitme engelli bireylerin iletişim ve bilgiye erişim süreçlerinde önemli bir kolaylık sağlamaktadır. Ancak, sistemlerin doğruluğunu artırmak için çeşitli zorlukların ele alınması gerekmektedir. Özellikle, parmak hareketlerinin algılanmasındaki sınırlamalar, hızlı veya karmaşık hareketlerin tanımlanmasındaki zorluklar ve çevresel faktörlerin

etkileri, bu teknolojilerin geliştirilmesi gereken başlıca alanlar arasında yer almaktadır (Bora vd., 2023).

3.6.2 Leap Motion Controller

Leap Motion Controller, masaüstüne yukarı bakacak şekilde yerleştirilmek üzere tasarlanmış küçük bir USB aygıtıdır (Şekil 3.3). Cihaz iki adet kızılötesi IR kamera ve üç adet kızılötesi LED kullanarak, yaklaşık 1 metrelik bir mesafeye yarım küre şeklindeki bir alanı gözlemler. Bu sensör sadece el hareketlerini değil, aynı zamanda parmakların 598 eklemlerini ayırt etme ve hareketlerini izleme yeteneğine de sahiptir (Bhiri vd., 2023).



Şekil 3.3 Leap Motion

İşaret dili tanıma sistemlerinde, Leap Motion'dan elde edilen ham veriler, genellikle makine öğrenimi ve derin öğrenme algoritmalarıyla işlenmektedir. Hareket ve pozisyon bilgilerinin çıkarılması aşamasında, Leap Motion tarafından sağlanan el koordinatları, hız, yönelim ve parmak eklemleri gibi detaylı veriler kullanılır. Bu tür veriler, özellikle işaret dilinin parmak alfabesi gibi hassas ve ayrıntılı ifadelerinde yüksek bir doğruluk seviyesi sağlar. Daha sonra bu özellikler, Destek Vektör Makineleri (SVM), Karar Ağaçları veya Derin Sinir Ağları (DNN) gibi algoritmalarla sınıflandırılarak tanıma işlemi gerçekleştirilir. Özellikle zaman tabanlı işaretlerin analizi için Long Short-Term Memory (LSTM) gibi algoritmalar tercih edilmektedir (Hisham vd., 2021).

Leap Motion'un sağladığı hassasiyet ve düşük maliyet avantajı, bu cihazı erişilebilir işaret dili tanıma çözümleri için cazip bir seçenek haline getirmektedir. Ancak, cihazın

sadece el hareketlerini algılayabilmesi, işaret dili ifadelerinde beden hareketlerinin de önem taşıdığı durumlarda sınırlayıcı bir faktör olabilir. Ayrıca, cihazın görüş alanı dışındaki hareketleri algılayamaması ve çevresel ışık koşullarına olan duyarlılığı, işaret dili tanıma sistemlerinin performansını etkileyebilmektedir. Bu nedenle, Leap Motion Controller tabanlı sistemlerin, genişletilmiş sensörler veya diğer teknolojilerle entegrasyonu, daha kapsamlı ve güvenilir bir işaret dili tanıma altyapısı sunabilir. Tüm bu faktörler göz önüne alındığında, Leap Motion Controller, işaret dili tanıma sistemleri için güçlü bir bileşen olmakla birlikte, daha geniş bir kapsam ve doğruluk elde etmek için destekleyici teknolojilerle bir arada kullanılmalıdır.

3.6.3 Veri Eldivenleri

Veri eldiveni, el duruşunu ayırt eden sensörlerle donatılmış benzersiz bir elektronik eldivendir (Şekil 3.4). Çoğu ticari işaret dili çeviri sistemi, eldivenlerin kullanılmasıyla elin eğilmesi ve 3D yönelimi hakkında veri elde etmek için veri eldiven yöntemini kullanırlar. Veri eldiveni, her parmağın üzerinde on adet esnek sensörle konumlandırılmıştır. Bu sensörler, parmakların her bir ekleminde bulunan bükülme noktasını tanıyabilir ve bilgiyi mikrodenetleyiciye gönderebilir. Veri eldiveninin dış tabakasına, parmak ve avuç içi bağlantılarının parmak uçlarına kadar monte edilir. Ayrıca, el pozunu tanımanın kesinliğini arttırmak, el hareketinin hızlanma değişimini yakalayabilmek için 3 eksenli bir ivmeölçer kullanılır. İvmeölçer, veri eldivenin arkasında bulunur. Veri eldiveni, statik ve hareket işaretleri içeren hem parmak izi hem de işaret hareketlerini algılamakta oldukça uygundur. Bununla birlikte, bu veri eldiveni maliyetli olabilmektedir. Daha az pahalı bir veri eldiveni üretilebilse de kullanılan sensörlerin sayısı azalırsa, el duruşu ile ilgili önemli verilerin kaybolmasına neden olur. Bu da işaret dil yorumunda doğruluk kaybına yol açmaktadır (Dong vd., 2021).



Şekil 3.4 Data Gloves

İşaret dili tanıma sistemlerinde, veri eldivenleri ile birlikte kameraların kullanılması, doğruluk oranlarını artırmaya yönelik etkili bir yaklaşım olarak literatürde yer bulmaktadır. Lu vd. (2023) yaptığı çalışmada, iki eksenli bükülme sensörleri ve bilgisayarla görme yöntemlerinin veri füzyonuyla birleştirilmesi yoluyla işaret dili tanımadaki doğruluk değerinin artırılması hedeflenmiştir.

Bahsedilen bu sistemlerde işaret dili tanıma için özel kameralara ve donanımlara ihtiyaç bulunmaktadır. Gelişen görüntü işleme teknikleri sayesinde eklem yerlerinin, vücuttaki anahtar noktaların tespiti daha kolay bir hale gelmiştir. Yüksek doğruluk ile çalışan bu teknikler sayesinde özel bir donanıma gerek kalmadan, normal bir kamera ile vücut anahtar noktalarının tespiti yapılabilmektedir.

3.6.4 Görüntü Tabanlı Sistemler

İşaret dili tanıma sistemlerinde normal web kameralarının kullanımı, ekonomik ve erişilebilir çözümler sunması nedeniyle önemli bir alan haline gelmiştir. Bu tür sistemler, görüntü toplama, ön işleme, özellik çıkarımı ve sınıflandırma gibi temel adımları içeren bir işleme hattı üzerinden çalışır. Web kameraları, kullanıcıdan elde edilen el hareketleri, yüz ifadeleri ve vücut duruşlarını analiz ederek işaret dilindeki anlamları tanımlamak için temel veri kaynağı olarak işlev görür (More vd., 2018).

Web kameraları genellikle RGB formatında video akışı sağlarken, 30 fps gibi bir kare hızında çalışarak hareketli işaretlerin tespiti için yeterli bir çözünürlük sunar. İşleme hattının ilk adımı olan görüntü toplama aşamasında, kameradan alınan veriler çeşitli ön işleme tekniklerine tabi tutulur. Arka plan çıkarma, aydınlatma koşullarını dengeleme ve gürültü azaltma gibi yöntemler, işaretlerin daha net bir şekilde algılanmasını sağlar. Özellikle hareketli arka planlar veya düşük ışık gibi zorlu çevresel koşullarda, bu tür işlemler sistemin performansını artırmada kritik rol oynar.

Ön işleme sonrası, işaret dilindeki jestlerin analizine yönelik özellik çıkarımı yapılır. Bu süreçte, elin pozisyonu, şekli ve hareket yönü gibi temel özellikler çıkarılarak sınıflandırma modellerine aktarılır. Normal web kameralarının sınırlı derinlik algısına rağmen, el ve parmak noktalarını doğru bir şekilde algılamak için renk tabanlı segmentasyon veya derin öğrenme tabanlı yaklaşımlar kullanılabilir. OpenPose ve

MediaPipe gibi araçlar, eklem noktalarını gerçek zamanlı olarak belirleyerek el hareketlerinin yanı sıra yüz ifadelerinin ve vücut pozisyonlarının da tanınmasını mümkün kılar. Bu araçlar, segmentasyon aşamasında yüksek doğruluk sağlamak için derin sinir ağı tabanlı algoritmalarla desteklenmiştir (Halder ve Tayade, 2021).

Özellik çıkarımından sonra, tanıma aşamasında makine öğrenimi ve derin öğrenme modelleri devreye girer. Evrişimsel Sinir Ağları (CNN), görüntülerden anlamlı özellikler öğrenmek için sıklıkla kullanılırken, hareketlerin zamansal sürekliliğini analiz etmek için LSTM gibi sıralı modeller tercih edilir. İşaret dili hareketlerinin zaman serisi verileriyle modellenmesi, jestlerin doğru şekilde tanımlanmasını sağlar. Eğitim aşamasında geniş bir veri kümesinin kullanılması, tanıma performansını artırmak için kritik öneme sahiptir (Arooj vd., 2024).

Web kameraları ile işaret dili tanıma sistemleri geliştirirken, sınırlı çözünürlük, değişken aydınlatma ve arka plan gürültüsü gibi teknik zorluklarla karşılaşılabilir. Bu kısıtlamalar, veri işleme ve algoritma tasarımında alınacak önlemlerle minimize edilebilir. Örneğin, ışık değişimlerine dayanıklı özellik çıkarma yöntemleri veya gerçek zamanlı çalışmayı hızlandıran optimizasyon teknikleri, sistemlerin daha dayanıklı hale gelmesini sağlar. Sonuç olarak, web kameralarının yaygın kullanılabilirliği ve düşük maliyeti, işaret dili tanıma teknolojilerinin geniş bir kitleye ulaşmasını ve bu alanda yenilikçi çözümler geliştirilmesini mümkün kılmaktadır.

3.7 Görüntü Tabanlı Sistemlerde Kullanılan Anahtar Nokta Tespit Algoritmaları

İnsan hareketinin analizi üzerine geçmişten günümüze kadar çeşitli çalışmalar yapılmıştır. Kinetik ve kinematik hareket analizi için birçok deney, yöntem ve araç geliştirilmiştir. Bilgisayarlı görü alanında eklem noktalarının tespiti, insan hareketlerinin ve duruşlarının dijital olarak anlaşılması ve modellenmesi açısından kritik bir öneme sahiptir. Bu tespit işlemi, insan vücudunun anatomik yapısına uygun olarak belirli anahtar noktaların (örneğin, eklemler ve uzuvlar) konumlarının hassas bir şekilde belirlenmesini içerir. Eklem noktalarının doğru ve etkin bir şekilde saptanması, çeşitli uygulama alanlarında derinlemesine analiz ve yenilikçi çözümler

sunar. Özellikle tıp ve rehabilitasyon alanlarında, eklem noktalarının analizi sayesinde hastaların hareket bozuklukları teşhis edilebilir ve tedavi süreçleri izlenebilir (Wu vd., 2020). Spor bilimlerinde atletlerin performans değerlendirmeleri ve hareket optimizasyonu için kullanılırken, insan-bilgisayar etkileşimi kapsamında doğal ve sezgisel arayüzlerin geliştirilmesine katkı sağlar (Jafarzadeh vd., 2021). Ayrıca, güvenlik ve gözetim sistemlerinde anormal davranışların tespiti, artırılmış ve sanal gerçeklik uygulamalarında gerçekçi avatarların oluşturulması ve robotik alanında insan hareketlerinin taklit edilmesi gibi çeşitli disiplinlerde de yaygın olarak kullanılmaktadır (Gao vd., 2019; Wu vd., 2020).

Bilgisayarlı görü alanında eklem noktalarının tespiti, işaret dili tanıma ve analizinde temel bir rol oynamaktadır. İşaret dili, el hareketleri, yüz ifadeleri ve beden duruşlarının kombinasyonu ile zengin bir iletişim biçimi sunar. Bu karmaşık hareketlerin doğru anlaşılması ve dijital ortama aktarılması için el ve parmak eklemlerinin hassas tespiti kritik öneme sahiptir. Eklem noktalarının tespiti, işaret dilinin bileşenlerinin ayrıştırılmasını ve hareketlerin zamansal ve mekânsal özelliklerinin analiz edilmesini sağlar. Bu sayede, derin öğrenme tabanlı modeller ve hareket yakalama teknikleri kullanılarak işaretlerin otomatik olarak tanınması ve sınıflandırılması mümkün hale gelir (Karacı vd., 2021). Ayrıca, bu yaklaşım işaret dili eğitimi, tercümesi ve işitme engelli bireylerle etkileşimde yeni teknolojik çözümlerin geliştirilmesine katkı sunar. Dolayısıyla, bilgisayarlı görü ile eklem noktalarının tespiti, işaret dili araştırmalarında ve uygulamalarında hem teorik hem de pratik açıdan önemli bir araştırma alanıdır.

Bilgisayarlı görü alanında eklem noktalarının tespiti için çeşitli yazılım kütüphaneleri geliştirilmiştir. Bu kütüphaneler, insan vücudu üzerindeki eklem noktalarını belirlemek için farklı algoritmalar ve mimariler kullanarak, hareket analizi ve tanıma görevlerinde önemli rol oynamaktadır. Her bir kütüphane, işlem hızı, doğruluk, platform uyumluluğu ve kullanım kolaylığı gibi farklı avantaj ve dezavantajlara sahiptir. Özellikle işaret dili kelimelerinin tespitinde, el ve parmak hareketlerinin hassas bir şekilde analiz edilmesi gerekmektedir. Bu nedenle, kullanılacak yazılım kütüphanesinin bu spesifik gereksinimleri karşılayabilmesi kritik öneme sahiptir. Bu

çalışmada, işaret dili kelimelerinin tespiti için uygun olabilecek yazılım kütüphaneleri karşılaştırılmış ve performansları detaylı bir şekilde değerlendirilmiştir.

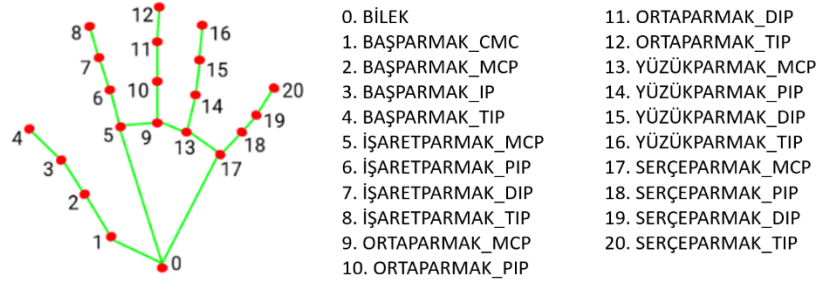
MediaPipe, MoveNet, OpenPose, AlphaPose, Detectron2 ve HRNet gibi kütüphaneler, insan vücudunun anahtar noktalarının doğru ve hızlı bir şekilde tespit edilmesi için yaygın olarak kullanılmaktadır. Bu kütüphaneler, özellikle sağlık, güvenlik ve spor alanlarında önemli avantajlar sunmakla birlikte, her birinin kendine özgü avantaj ve dezavantajları bulunmaktadır. MediaPipe Holistic, el ve parmak eklemlerini yüksek doğrulukla tespit etme yeteneği ve düşük donanım gereksinimi ile mobil ve gerçek zamanlı işaret dili uygulamaları için idealdir. MediaPipe Pose'un bilek noktasını tespit ederek elin konumunu belirlemesi ve ardından el üzerindeki anahtar noktaları MediaPipe Hand çözümü ile yüksek doğrulukla analiz etmesi, bu kütüphaneyi işaret dili tanıma sistemleri için öne çıkarır.

3.7.1 MediaPipe

MediaPipe, Google tarafından geliştirilen ve çapraz platformlarda çalışabilen, makine öğrenmesi çözümleri sunan açık kaynaklı bir çerçeve olarak tanımlanabilir. Video ve görüntü işleme gibi alanlarda daha önceden gerçekleştirilebilen nesne algılama gibi işlemleri, MediaPipe kullanıcı dostu bir hale getirerek, bu tür uygulamaların geliştirilmesini kolaylaştırmıştır. MediaPipe, esnek ve genişletilebilir bir altyapı sunarak farklı uygulama alanlarında kullanılabilirliği artırmıştır. MediaPipe çerçevesi ile yüz tanıma, yüz ağı (mesh) algılama, iris algılama, el algılama, vücut algılama (eklemleriyle birlikte), saç algılama ve nesne takibi gibi çok çeşitli görevler yerine getirilebilmektedir. Bu geniş uygulama yelpazesi, MediaPipe'in farklı platformlarda (Android, iOS vb.) ve farklı programlama dillerinde (C, Python, JavaScript) kullanılabilir olmasını sağlayan çapraz platform desteği ile güçlendirilmiştir. Bu özellikler, MediaPipe'in araştırma projelerinden gerçek zamanlı uygulamalara kadar geniş bir yelpazede kullanılmasını mümkün kılmaktadır (Dill vd., (2023).

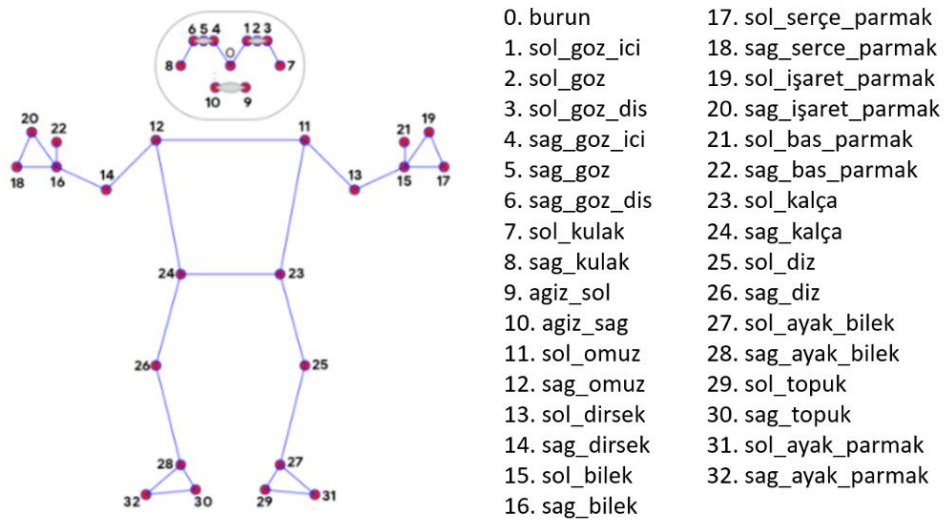
MediaPipe'in el algılama modülü olan "Hands," bir görüntü üzerindeki ellerin ve el eklemlerinin algılanmasını sağlamaktadır. Bu modül, bir elde 21 anahtar noktanın koordinatlarını yüksek doğrulukla tespit edebilmektedir (Şekil 3.5). Elde edilen bu

koordinatlar, işaret dili tanıma gibi uygulamalarda, el hareketlerinin detaylı analizi ve anlamlandırılması için kullanılabilir. MediaPipe Hands, işaret dili tanıma sistemlerinin geliştirilmesi için hızlı, doğru ve esnek bir çözüm sunmakta, aynı zamanda bu alanda standart bir altyapı sağlayarak araştırmaların ilerlemesine katkıda bulunmaktadır (Google AI, 2024).



Şekil 3.5 MediaPipe Hand ile Tespit Edilen Yer İşaretleri

İşaret dilinde yalnızca ellerin değil, kolların duruşunun ve hareketlerinin de önemli bir rolü bulunmaktadır. İşaret dili kelimeleri, kollar ve ellerin birlikte oluşturduğu hareketlerle ifade edilmektedir (Naz vd., 2023). Bu nedenle, yalnızca ellerin anahtar noktalarını tespit eden MediaPipe Hands çözümünü kullanmak yeterli olmamaktadır. Vücut üzerindeki temel eklem noktalarının tespiti için MediaPipe'ın "Pose" çözümü kullanılabilir. Şekil 3.6.'da gösterildiği gibi MediaPipe Pose, bir bireyin vücudundaki 33 farklı yer işaretini tespit ederek, vücut hareketlerinin kapsamlı bir şekilde analiz edilmesine olanak tanımaktadır (Arya ve Maji, 2024).



Şekil 3.6 MediaPipe Pose ile Tespit Edilen Yer İşaretleri

MediaPipe Hands ve Pose çözümleri, sırasıyla el ve vücut üzerindeki anahtar noktaları tespit etmek amacıyla geliştirilmiştir. Ancak, gerçek zamanlı uygulamalarda, bu iki çözümün bağımsız bir şekilde çalışması gecikmelere yol açabilmektedir. Bu gecikmeler, özellikle el anahtar noktalarının görüntü üzerinde aranması sırasında ortaya çıkmaktadır. MediaPipe’ın Holistic çözümü (Şekil 3.7), bu tür gecikmeleri önlemek ve daha hızlı bir tespit süreci sağlamak için geliştirilmiştir. Holistic çözümünde, vücut anahtar noktalarının belirlenmesinde kullanılan Pose modülü, bilek noktalarını tespit etmek için bir referans sağlamaktadır. Bilek noktalarının tespiti sonrası, el anahtar noktaları yalnızca yeniden boyutlandırılan bir alt görüntü üzerinde aranır. Bu yaklaşım, işlem süresini önemli ölçüde azaltarak tespit sürecini hızlandırır (Kuznetsova ve Kimmelman, 2024).



Şekil 3.7 MediaPipe Holistic Çalışma Yöntemi (Ivan vd., 2020)

MediaPipe Holistic, yeniden boyutlandırma işleminde çerçevedeki nesnenin önemli ölçüde hareket etmediğini varsayar. Ayrıca, mevcut çerçevedeki nesne bölgesine rehberlik etmek için önceki çerçeveden elde edilen tahmini kullanır. Bu yöntem, çerçeveler arasındaki tahmin doğruluğunu artırırken, işlem süresini kısaltır ve gerçek zamanlı uygulamaların performansını yükseltir. Holistic çözümü, aynı zamanda bir çerçeve içerisinde bulunan sol ve sağ ellerin veya vücut bölümlerinin karışmasını önler. Bu özellik, modelin gövde ve parçaları arasında anlamsal tutarlılığı korumasını sağlayarak tespit işlemlerinde daha yüksek doğruluk sunar. Bu nedenle, MediaPipe Holistic, işaret dili tanıma sistemleri gibi karmaşık uygulamalar için uygun ve etkili bir çözüm olarak değerlendirilebilir (Mallik vd., 2024).

MediaPipe Holistic, insan vücudunun farklı bölgelerindeki detaylı anahtar noktaların tespitini sağlayarak işaret dili tanıma gibi uygulamalar için kapsamlı bir altyapı

sunmaktadır. Holistic çözümü, yüz, eller ve vücut gibi farklı bölgelerdeki yer işaretlerini tespit eder ve bu işaretlerin üç boyutlu (x, y, z) koordinatlarını çıkarır. Bu kapsamlı veri, hareketlerin ve duruşların zamansal ve uzamsal olarak detaylı bir şekilde analiz edilmesine olanak tanır.

Yüz üzerinde 468 yer işareti tespit edilebilmekte ve her bir işaret için x, y ve z koordinatları hesaplanmaktadır. Bu işlem sonucunda yüz için toplamda 1404 anahtar nokta elde edilmektedir. Eller üzerinde, her biri 21 yer işaretinden oluşan sol ve sağ elin tespiti yapılabilmektedir. Her iki el için ayrı ayrı 63 anahtar nokta (21 yer işareti x 3 koordinat) hesaplanarak toplamda 126 anahtar nokta çıkarılmaktadır.

Vücut için ise, MediaPipe Holistic 33 yer işaretini tespit etmektedir. Bu işaretlerin üç boyutlu koordinatlarının hesaplanmasıyla toplamda 132 anahtar nokta elde edilir. Böylece yüz, eller ve vücut işaretlerinin birleştirilmesiyle toplamda 1662 anahtar noktanın tespiti gerçekleştirilebilir. Bu geniş kapsamlı veri seti, işaret dili tanıma gibi uygulamalarda hareketlerin detaylı bir şekilde analiz edilmesine olanak tanımakta ve modelin doğruluğunu artırmaktadır (Goyal, 2023). MediaPipe Holistic'in sağladığı bu detaylı veri yapısı, işaret dili hareketlerinin farklı bölgelerdeki koordinasyonunu anlamak ve modellemek için güçlü bir araçtır.

4. MATERYAL VE YÖNTEM

Bu bölümde, çalışma kapsamında oluşturulan özgün veri setinin ayrıntıları, kullanılan derin öğrenme modelleri ve bu modellerin değerlendirilmesinde kullanılan performans metrikleri açıklanmaktadır.

4.1 Veri Seti

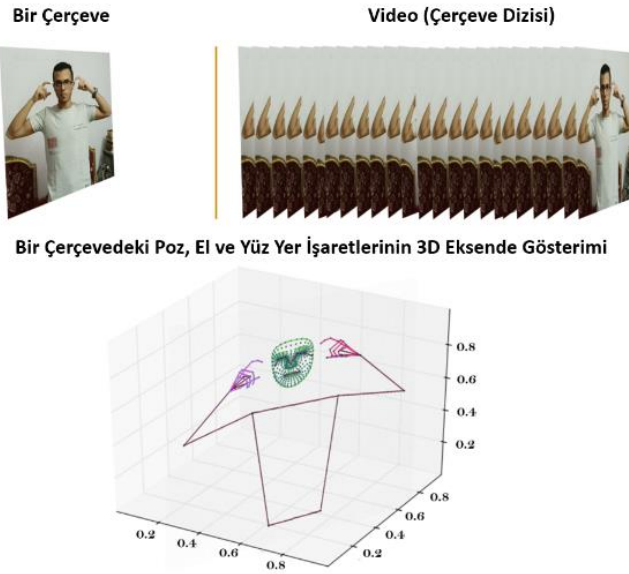
Bu çalışmada temel iletişim kalıplarından ve Millî Eğitim Bakanlığı Özel Eğitim ve Rehberlik Genel Müdürlüğünün yayınladığı Türk İşaret Dili Dersi Öğretim Materyali kitabında geçen 446 tane kelime seçilmiştir. Bu kelimeler günlük hayatta sıklıkla kullanılan temel işaret dili kelimeleridir. Bu, oluşturulan veri setinin pratik kullanıma uygun ve geniş bir yelpazede anlamlı bir kapsama sahip olmasını sağlamaktadır. Şekil 4.1’de öğretim materyalindeki ilk ünitedeki TİD kelimeleri örnek olarak gösterilmektedir. Buradaki dinamik işaret kelimelerini yaparken ellerin hareketleri kırmızı oklarla gösterilmiştir. Oluşturulan veri setindeki kelimelerin 313 tanesi dinamik, 133 tanesi statik kelimelerden oluşmaktadır. İşaret dilinde dinamik kelimelerde, kelimeyi anlatan birkaç farklı hareket yapılması gerekirken, statik kelimelerde ise kelimeyi göstermek için el, kol ve parmakların sabit bir hareketi yapması yeterlidir. Bu yüzden işaret dilindeki dinamik kelimelerin tespiti daha zordur.



Şekil 4.1 Türk İşaret Dili Dersi Öğretim Materyali Kitabı Birinci Ünite Kelimeleri

4.2 Veri Seti Oluřturma

Veri setinin oluřturulmasında OpenCV ve MediaPipe Holistic araları kullanılarak Python programlama dili ile bir veri toplama scripti geliřtirilmiřtir. Video akıř verileri kameradan alınarak her bir kelime iin 10 farklı kiři tarafından 10’ar rnek video kaydedilmiřtir. Bu yntem, veri setinin eřitliliğini ve genelleřtirilebilirliğini artırmayı amalamıřtır. Her video 30 erveden oluřmakta ve her erve iin 1662 anahtar noktası hesaplanmaktadır. Bu anahtar noktalar, hareketlerin ve ifadelerin zamansal ve uzamsal zelliklerini detaylı bir řekilde temsil eden numpy dizilerine kaydedilmiřtir. Bylece, veri seti iřaretlerin grsel ve hareket bazlı analizine olanak tanıyacak řekilde zengin bir yapıya sahiptir. řekil 4.2’de videodaki erve dizisinden alınan erve zerinden anahtar noktaların tespiti gsterilmektedir.

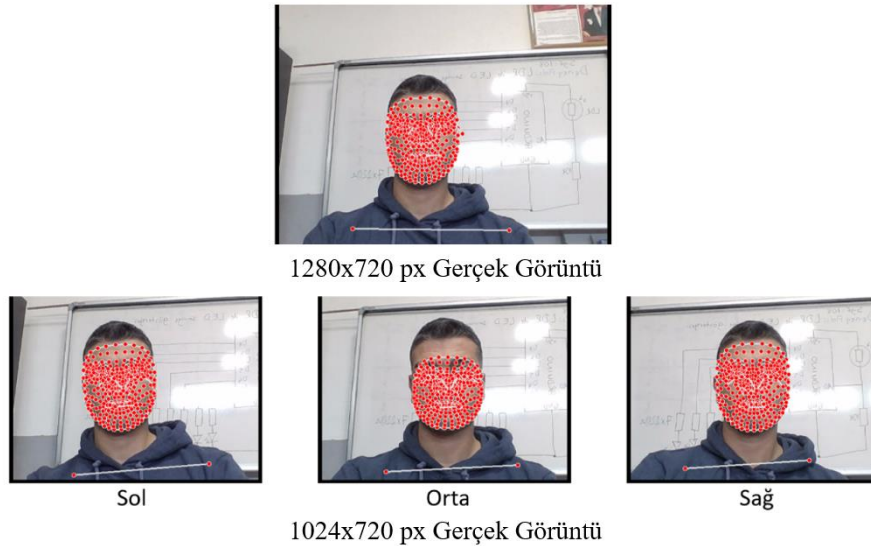


řekil 4.2 Her erve iin Anahtar Noktaların Tespiti

Veri seti oluřturma srecinde, standartlařtırılmıř bir yapı elde etmek amacıyla kamera grntleri, kiřiyi ervede merkezleyecek řekilde kırpılmıřtır. Bu yaklařım, farklı bireylerden elde edilen grntlerde eklem noktalarının konumlarının doėru bir řekilde tespit edilmesini ve tanıma başarısının artırılmasını hedeflemektedir. Kamera grntleri, 1280x720 piksel znrlkte elde edilmiřtir. İlk olarak, MediaPipe Holistic algoritması kullanılarak yzn orta noktası tespit edilmiřtir. Bu iřlem, grntnn kiřinin merkeziyle hizalanmasını saėlamak iin bir referans noktası oluřturmuřtur.

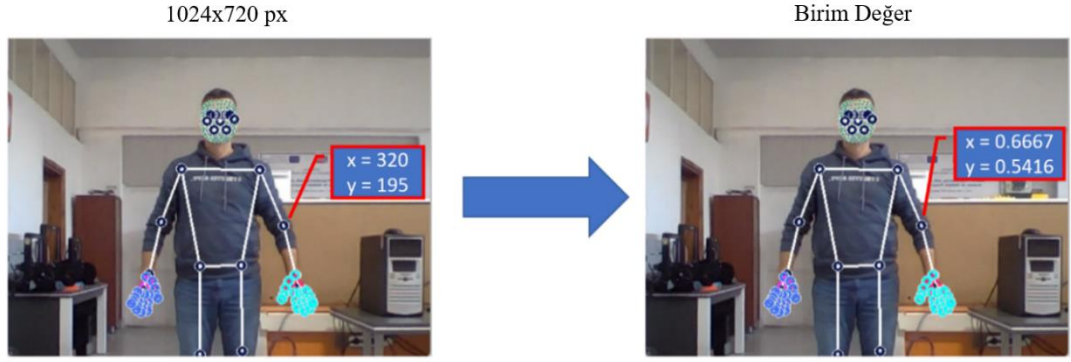
Yüz orta noktası referans alınarak, sahne çözünürlüğü 1024x780 piksele yeniden boyutlandırılmış ve kırpma işlemi bu çözünürlük üzerinde gerçekleştirilmiştir. Böylece, her bireyin sahnede aynı görsel standardı sağlaması hedeflenmiş ve işaret dili kelimelerinin tanınması için tutarlı bir veri seti yapısı oluşturulmuştur. Bu süreç, hem görüntüdeki görsel bileşenlerin tutarlılığını sağlamak hem de MediaPipe Holistic tarafından hesaplanan eklem noktalarının koordinatlarının farklı kişiler arasında karşılaştırılabilir olmasını temin etmek için kritik bir adım olmuştur.

Kırpma işlemi, kişinin sahneye ortalanmasını sağlayarak, işaret dilindeki el ve vücut hareketlerinin daha doğru bir şekilde analiz edilmesine olanak tanımaktadır. Bu işlem, özellikle yüz ve eller gibi işaret dilinde kritik öneme sahip olan bölgelerin daha hassas bir şekilde izlenmesini kolaylaştırmıştır. Standartlaştırılmış görsel ve koordinat sistemi, veri setinin hem kalite hem de uyarlanabilirlik açısından yüksek bir düzeye ulaşmasını sağlamış ve Türk İşaret Dili tanıma çalışmalarında kullanılacak sağlam bir altyapı sunmuştur. Şekil 4.3'te, bu işlem adımlarının görsel bir temsili ve sonuçları ayrıntılı olarak gösterilmiştir.



Şekil 4.3 Kırpma İşlemi

Mediapipe ile bulunan anahtar noktaların x ve y koordinatları sahne çözünürlüğüne göre değişmektedir. Bu yüzden tespit edilen anahtar noktaların x ve y koordinatları kamera çözünürlüğünden bağımsız olması için birim değere çevrilmiştir (Şekil 4.4).



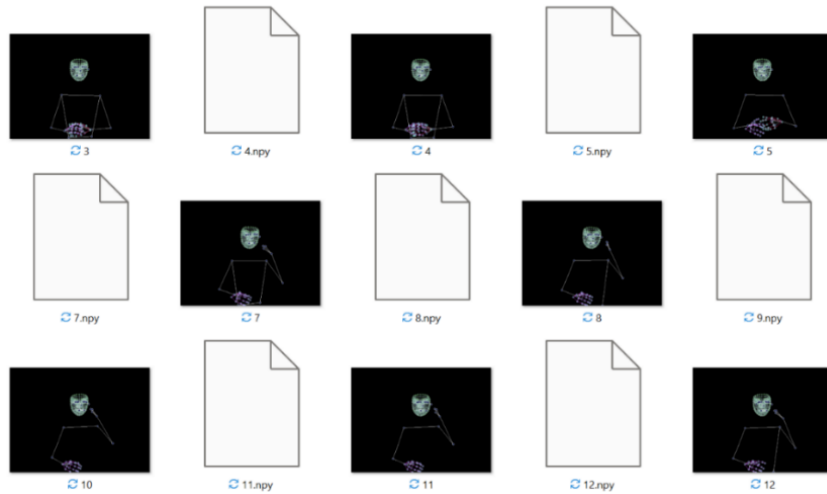
Şekil 4.4 Kırpma İşlemi

Birim değere çevrilen x, y ve z koordinatları numpy dizisi içerisinde saklanmaktadır (Şekil 4.5).

```
array([ 0.50711566,  0.59728622, -1.63998222, ...,  0.21593845,
        0.28581554, -0.0511836 ])
```

Şekil 4.5 Birim Değere Çevrilmiş Koordinat Değerleri

Veri setini genişletip görüntü tabanlı algoritmaları da test edebilme adına çerçeveler için tespit edilmiş anahtar nokta koordinatları ile birlikte siyah arkaplan üzerinde anahtar nokta ve anahtar nokta bağlantılarının olduğu (İskelet Görünümü) görsellerde oluşturulmuştur. Videolardaki 30 çerçeve için anahtar nokta ve anahtar nokta bağlantılarının olduğu görseller de script yardımı ile otomatik bir şekilde oluşturulmuştur. Oluşturulan numpy ve iskelet görüntüsü dosyaları kelime için oluşturulan klasöre kaydedilmiştir (Şekil 4.6).



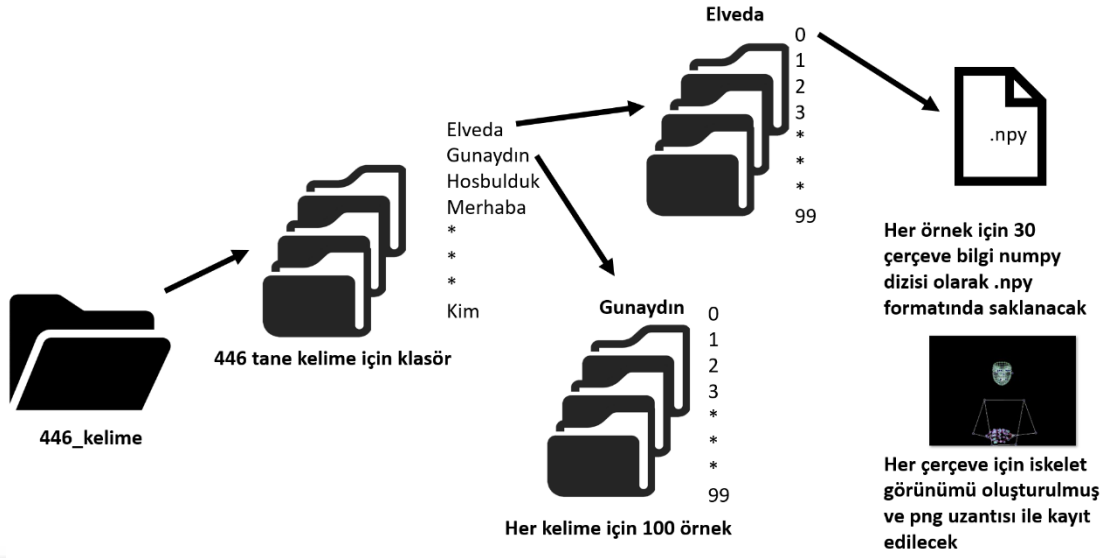
Şekil 4.6 Numpy Dosyaları ve İskelet Görünümü Görselleri

4.3 Veri Seti Klasör Yapısı

Çalışmada, 446 farklı kelimenin her biri için 10 katılımcı tarafından 10 video kaydedilmiştir. Veri setinin tutarlılığını sağlamak amacıyla, tüm kelimelerin ortalama işaretleme süreleri analiz edilmiş ve en uzun süreye sahip kelime referans alınarak sabit uzunlukta örnekleme stratejisi benimsenmiştir. Video kayıtları, OpenCV kütüphanesi kullanılarak 24 FPS (saniyede kare sayısı) hızında gerçekleştirilmiştir. Her bir video, en uzun işaret süresini kapsayacak şekilde 30 çerçeveye eşit aralıklarla bölünmüş ve bu çerçevelerden elde edilen anahtar nokta verileri kullanılarak NumPy dizileri oluşturulmuştur. Bu sayede, tüm kelimeler 1,2 saniyelik (30 kare) sabit uzunlukta temsil edilmiş, böylece zamanlama farklılıklarının model performansı üzerindeki etkisi minimize edilmiştir. Bu yaklaşım, modelin öğrenme sürecinde tutarlı bir veri yapısı sağlamış ve işaret dilindeki zamanlama varyasyonlarının neden olabileceği sapmaları ortadan kaldırmıştır. Ayrıca, her kelimenin farklı katılımcılar tarafından tekrarlanması, veri setinin çeşitliliğini ve genellenebilirliğini artırmıştır.

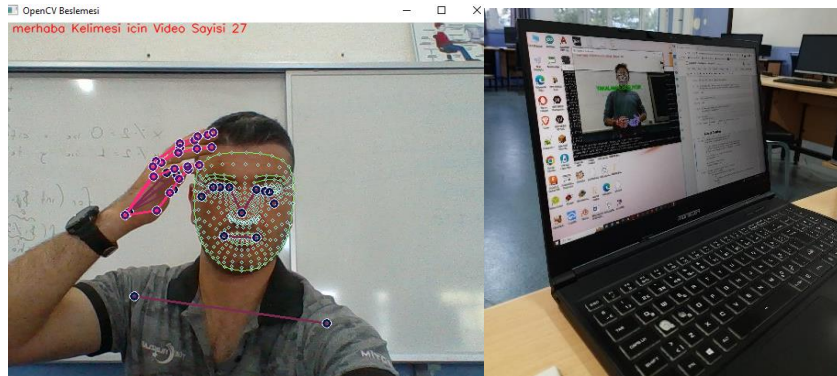
Veri setinin oluşturulmasında görev alan bireyler, yaşları 17 olan ve tamamı erkek olan 10 gönüllüden oluşmaktadır. Katılımcıların daha önce işaret dili tecrübeleri bulunmamaktadır. Ancak veri toplama sürecinden önce, Millî Eğitim Bakanlığı'nın Türk İşaret Dili Eğitimi Kitabı ve çeşitli çevrim içi kaynaklar kullanılarak katılımcılara hedef kelimelerin doğru gösterimi öğretilmiş, böylece kayıtların standardizasyonu sağlanmıştır.

Bu kapsamda geliştirilen Python betiği ile, her kelimeye ait örnekler, ilgili klasör yapısına hızlı ve sistematik biçimde yerleştirilmiş; veri seti düzenli bir şekilde yapılandırılmıştır. Şekil 4.7'de bu klasör yapısının mimarisi görsel olarak sunulmaktadır.



Şekil 4.7 Veri Seti Klasör Yapısı

Klasör yapısını oluşturduktan sonra veri oluşturma için çalıştırılan program ile 446 kelime için 100 örnek video alma işlemi sırası ile yapılmaktadır. Toplamda 44.600 (446x10x10) video için 1.338.000 çerçeve bilgisi oluşturulmuştur. Her çerçeve için 1662 anahtar nokta değeri numpy dizilerinde saklanmaktadır. Toplamda 2.223.756.000 anahtar nokta değeri oluşturulmuştur. Şekil 4.8’de örnek oluşturma için Python programlama dili oluşturulan programın çalışması gösterilmektedir.



Şekil 4.8 Veri Seti Oluşturma Aşaması

Video çekimleri, aydınlık ve gölgelenmeyen bir ortamda, sade arka planlı bir alanda gerçekleştirilmiştir. Kamera ile katılımcı arasında ortalama 2 metre mesafe bırakılarak, MediaPipe’in anahtar noktaları yüksek doğrulukla tespit edebilmesi sağlanmıştır (Şekil 4.9).



Şekil 4.9 Veri Seti Oluşturma Ortamı

4.4 MediaPipe Holistic Anahtar Nokta Tespiti

MediaPipe ile her çerçevede tespit edilen yer işaretlerinin x, y ve z eksenlerindeki koordinatları bulunmuş ve bunlardan 3 boyutlu bir dizi oluşturulmuştur. Yüz, eller ve vücut için oluşturulan diziler flatten() metodu kullanılarak 1 boyutlu dizi haline getirilmiştir. Bazı çerçevelerde tespit edilemeyen yer işaretlerinden dolayı dizi elemanlarının eksik kalmaktadır. Bundan dolayı veriler işlenirken hata alınmaktadır. Bunun önüne geçmek için tespit edilemeyen yer işaretleri yerine diziye 0 değeri atanmıştır. Son olarak 4 dizi birleştirilmiştir ve her çerçevedeki anahtar noktaları içeren bir dizi oluşturularak numpy dizisi olarak kaydedilmiştir.

MediaPipe holistic modelinde varsayılan ayarlar olarak aşağıdaki değerler bulunmaktadır.

Varsayılan Ayarlar:

```
Holistic (self, static_image_mode=False, model_complexity=1,  
smooth_landmarks=True, enable_segmentation=False,  
smooth_segmentation=True, refine_face_landmarks=False,  
min_detection_confidence=0.5, min_tracking_confidence=0.5)
```

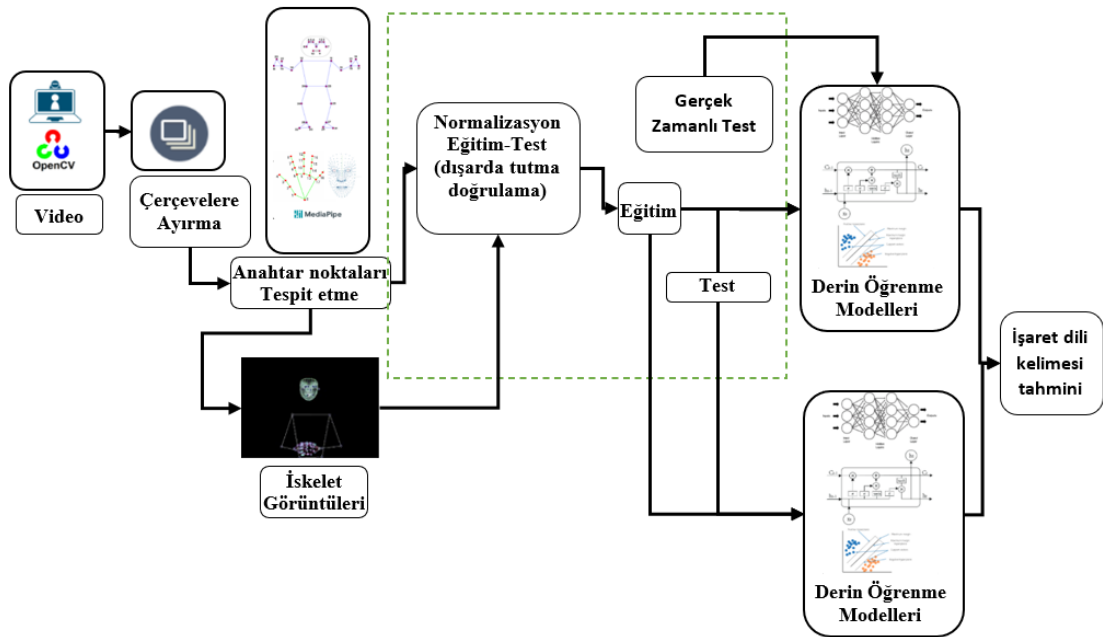
Bu varsayılan değerler içerisinde olan min_detection_confidence ve min_trackin_confidence değerleri 0,7 olarak değiştirilmiştir.

min_detection_confidence değeri, algılamanın başarılı sayılması için gerekli olan değerdir ve [0.0, 1.0] arasında değerler alır.

min_tracking_confidence değeri, yer işareti izleme modelinin başarılı sayılması için gerekli olan değerdir ve [0.0, 1.0] arasında değerler alır. Yüksek değer başarıyı artırır fakat bundan dolayı gecikme de oluşabilir. 0,7 değeri göz ardı edilebilecek gecikme sağlanmıştır.

4.5 Yöntem

Şekil 4.10’da TİD kelimelerini sınıflandırmak için kullanılacak yöntemlerin genel mimarisini göstermektedir. Elde edilen TİD veri seti üzerinde birçok derin öğrenme modellerinin eğitim ve test işlemi gerçekleştirilerek sonuçlar karşılaştırılmıştır. Derin öğrenme modelleri ilk olarak veri集中的 anahtar noktalar kullanılarak eğitilmiştir. Daha sonra iskelet görüntüleri ile eğitim işlemi gerçekleştirilmiştir.



Şekil 4.10 TİD Kelime Tanıma Sistemin Genel Tasarımı

Çalışma kapsamında, Türk İşaret Dili (TİD) sınıflandırması amacıyla CNN, GRU, LSTM ve Bi-LSTM gibi ardışıl modellerin yanı sıra, CNN+GRU, CNN+BiLSTM, CNN+LSTM, GRU+BiLSTM, GRU+LSTM hibrit mimarilere sahip modeller eğitilmiş ve test edilmiştir. Modellerin eğitimi ve değerlendirilmesinde dışarıda tutma (hold-out) yöntemi kullanılmıştır. Bu kapsamda veri seti üç alt kümeye ayrılmıştır: eğitim, doğrulama ve test. İlk aşamada verinin %80'i eğitim ve %20'si test için ayrılmış; ardından eğitim verisinin %12,5'i doğrulama kümesi olarak belirlenmiştir.

Sonuç olarak, veri seti %70 eğitim, %10 doğrulama ve %20 test şeklinde yapılandırılmıştır. Gerçek zamanlı performans değerlendirmesi amacıyla, biri veri setinde yer alan, diğeri ise veri setinde bulunmayan iki katılımcı ile deneysel bir test senaryosu gerçekleştirilmiştir. Bu senaryoda, gerçek zamanlı uygulamalarda kullanılacak nihai modelin belirlenmesinde doğrulama ve test başarımları temel alınmış; en yüksek performansı gösteren model yapısı tercih edilmiştir.

Veri bölme işlemleri sırasında, sınıf dengesinin korunmasına özel önem verilmiş; her küme içinde işaret sınıflarının temsili dengeli olacak biçimde stratified sampling yöntemi uygulanmıştır. Bu yaklaşım sayesinde, modelin azınlık sınıflarda da etkin bir şekilde genelleme yapması hedeflenmiştir. Veriler, NumPy dizileri biçiminde yapılandırılmış, sınıf etiketleri ise one-hot encoding yöntemi ile dönüştürülerek modelin çok sınıflı çıktı üretme yeteneği desteklenmiştir.

Ayrıca veri seti oluşturulurken her çerçeve için anahtar noktaları gösteren iskelet görüntüleri de oluşturulmuştur. Bu iskelet görüntüleri ızgara, optik akış ve yan yana birleştirme yöntemleri ile birleştirilerek ResNet-18 ve dikkat mekanizması temelli CBAM+ResNet-18 modelleri ile eğitilmiş ve test edilmiştir.

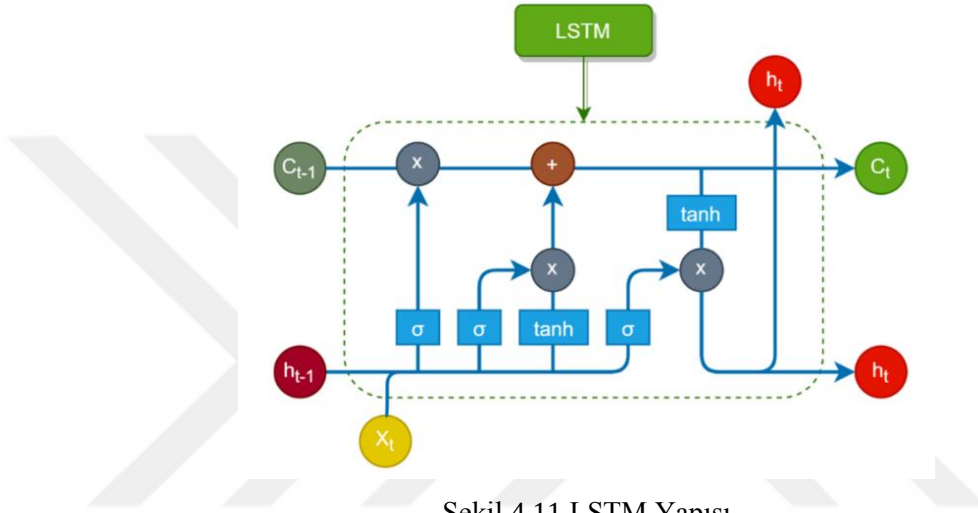
Çalışmada kullanılan derin öğrenme modelleri aşağıda ayrıntılı olarak açıklanmaktadır.

4.5.1 LSTM (Long Short-Term Memory)

LSTM (Long Short-Term Memory- Uzun Kısa Süreli Bellek), temel olarak klasik RNN yapısının daha gelişmiş bir biçimidir. Geleneksel RNN'lerdeki sınırlı hafıza sorununu aşmak için LSTM hücreleri, her adımda yalnızca bir çıktı üretmek yerine çeşitli kapı mekanizmaları aracılığıyla bilgiyi seçmeli olarak aktarır veya saklar. Bu yapı sayesinde, önceki zaman adımlarındaki önemli bilgiler gerektiğinde tekrar kullanılabilir hâle gelir (Sundar vd., 2022).

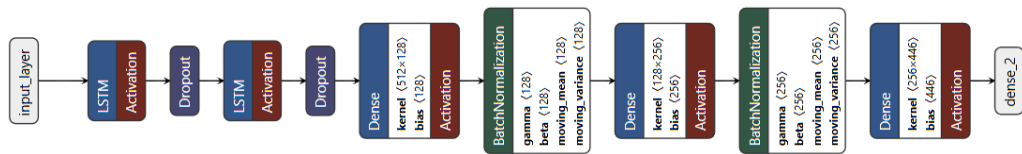
LSTM mimarisi, ardışık verileri işlemek ve hücreye giren ve çıkan bilgiyi kontrol ederek doğrusal olmayan bağımlılıkları yakalamak için tasarlanmıştır. Şekil 4.11'de gösterildiği gibi, LSTM yapısı birkaç temel bileşenden oluşur: X_t , giriş vektörünü; C_t ,

mevcut bloğun belleğini; h_t ise mevcut bloğun çıktısını temsil eder. σ ve $\tanh$ gibi doğrusal olmayan aktivasyon fonksiyonları, mimarinin içindeki dönüşümleri düzenler. Eleman bazlı çarpma ($\times$) ve toplama ($+$) işlemleri, kapı mekanizmalarının çalışmasını sağlar. Ayrıca, C_{t-1} ve h_{t-1} , bir önceki bloktan gelen bellek ve çıktı durumlarıdır; bu sayede ardışık veri işleme sürekliliği sağlanır. Bu tasarım, LSTM'lerin uzun vadeli bağımlılıkları öğrenmesini ve zaman serisi verilerindeki karmaşık desenleri etkili bir şekilde yakalamasını mümkün kılar (Crisóstomo de Castro Filho vd., 2020).



Şekil 4.11 LSTM Yapısı

Veri ön işleme adımının ardından model, ardışık olarak tanımlanmış ve LSTM tabanlı bir mimariyle yapılandırılmıştır. Söz konusu mimari, zamana bağlı hareket verilerindeki örüntüleri öğrenmeye yönelik olarak tasarlanmış ve işaret dili sınıflandırması gibi bağlamsal bilgi gerektiren görevler için uygulanmıştır. Modelin genel yapısı Şekil 4.12’de görsel olarak sunulmuştur.



Şekil 4.12 LSTM Model Yapısı

Modelin giriş katmanı, zaman adımları boyunca toplanan el hareketi verilerini alacak şekilde tasarlanmıştır. Ardışık bilgiyi işlemek üzere sırasıyla 512 birimden oluşan iki adet LSTM katmanı kullanılmıştır. İlk LSTM katmanı, her zaman adımına ait çıktıları üretirken (return_sequences=True), ikinci katman bu çıktıları özetleyerek yalnızca son

zaman adımını temel alan bir temsile dönüştürmektedir. Her iki katmanın ardından uygulanan Dropout işlemleri (%30 oranında), modelin aşırı öğrenmesini önlemek adına rastgele bağlantı koparmaları ile çalışır.

LSTM bloklarını takip eden Dense (yoğun bağlantılı) katmanlar sırasıyla 128 ve 256 birimden oluşmaktadır. Bu katmanlarda ReLU aktivasyon fonksiyonu kullanılmış, ardından Batch Normalization uygulanarak öğrenme sürecinde istikrar sağlanmıştır. Son katman, tahmin edilmesi hedeflenen işaret sayısına karşılık gelen nöronları içermekte ve çok sınıflı sınıflandırma için uygun olan softmax aktivasyon fonksiyonu ile sonlandırılmıştır.

Bu yapı, zamana yayılmış hareket desenlerini öğrenmenin yanı sıra soyut temsillerin sınıflandırılmasını da mümkün kılmaktadır. Eğitim sırasında elde edilen doğruluk oranları, modelin bu görev için etkili bir biçimde genelleme yapabildiğini göstermektedir. Özellikle LSTM katmanlarının sıralı bilgiye duyarlı yapısı, işaret dili gibi zamansal bağlam içeren sistemlerde önemli avantajlar sunmaktadır.

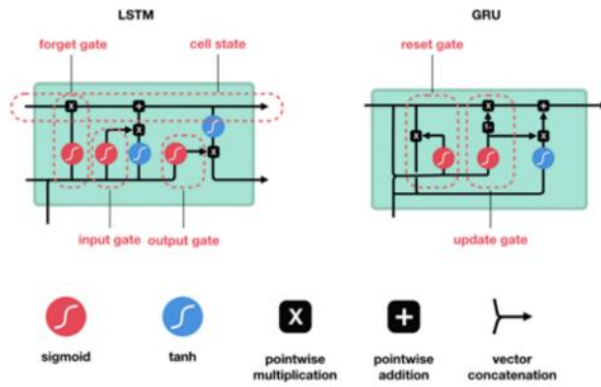
Modelin eğitimi sürecinde, optimizasyon algoritması olarak Adam ailesine ait Adamax yöntemi tercih edilmiştir. Bu seçimde, yüksek boyutlu parametre uzaylarında istikrarlı öğrenme sağlayabilmesi belirleyici olmuştur. Öğrenme oranı olarak 0.0001 değeri kullanılmış; bu sayede parametre güncellemeleri daha küçük adımlarla gerçekleştirilmiştir. Kayıp fonksiyonu olarak ise, çok sınıflı sınıflandırma problemlerinde yaygın kullanılan categorical crossentropy seçilmiştir. Model, 32 örnek içeren mini-batch'ler ile eğitilmiş ve eğitim süreci 1000 epoch ile sınırlandırılmıştır.

Aşırı öğrenmeyi önleyici stratejiler kapsamında, eğitim sırasında EarlyStopping mekanizması devreye alınmıştır. Doğrulama kaybında belirli bir süre boyunca iyileşme gözlenmediğinde eğitim otomatik olarak sonlandırılarak modelin genel performansı korunmuştur. Buna ek olarak, ReduceLROnPlateau yöntemi ile belirli epoch'larda doğrulama kaybı sabit kaldığında öğrenme oranı azaltılarak daha hassas bir optimizasyon süreci yürütülmüştür. Eğitim boyunca en iyi doğrulama başarımını

elde eden model ağırlıkları, ModelCheckpoint yöntemiyle kaydedilmiş ve test aşamasında bu ağırlıklar kullanılarak en yüksek performans elde edilmiştir.

4.5.2 GRU (Gated Recurrent Unit)

GRU (Gated Recurrent Unit – Kapılı Tekrarlayan Ünite) modelleri, tekrarlayan sinir ağlarının (Recurrent Neural Networks- RNNs) bir türüdür ve özellikle zaman serisi verileri, sıralı veriler veya metin gibi sıralı yapıya sahip veriler için kullanılır. GRU’lar kaybolan gradyan problemi (vanishing gradient problem)’ni çözmek amacıyla geliştirilmiştir. GRU, LSTM’in bir varyasyonu olarak da düşünülebilir çünkü her ikisi de benzer şekilde tasarlanmış ve bazı durumlarda eşit derecede başarılı sonuçlar verebilen ağ yapılarıdır. İkisinde de temel amaç aynıdır. LSTM’ler ve GRU’lar kısa süreli belleğe çözüm olarak oluşturulmuşlardır. Bilgi akışını düzenleyebilen “kapı (gate)” adı verilen iç mekanizmalara sahiptirler. Ancak GRU’lar da bir çıkış kapısı mevcut olmadığı için LSTM’den daha az parametresi bulunmaktadır (Kothadiya vd., 2022). Şekil 4.13’te GRU ve LSTM’a ait kapılar (gates) gösterilmektedir.

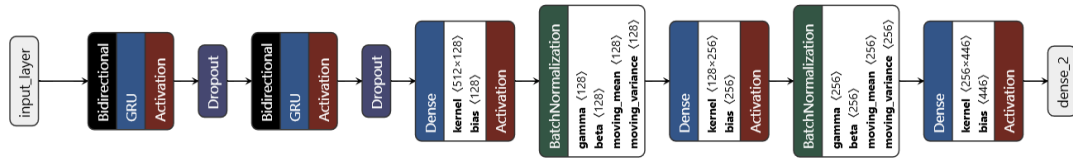


Şekil 4.13 GRU ve LSTM Modelleri

Gated Recurrent Unit (GRU), LSTM'lere kıyasla daha basitleştirilmiş bir yapıya sahiptir ve yalnızca iki kapı (sıfırlama kapısı ve güncelleme kapısı) kullanır. Bu kapılar, zaman serisi verilerinde hangi bilgilerin korunacağını ve hangi bilgilerin atlanacağını düzenler. GRU, hem kısa süreli hem de uzun süreli bağımlılıkları öğrenme kapasitesine sahiptir. Sıfırlama kapısı, geçmiş bilgilerin ne kadarının unutulacağını belirlerken, güncelleme kapısı, yeni bilgilere ne kadarının dahil edileceğini kontrol eder. Ayrıca, GRU’nun sahip olduğu daha az parametre, modelin eğitim süreçlerini hızlandırır ve bellek kullanımını azaltır; bu da özellikle küçük veri

setleriyle çalışıldığında veya sınırlı hesaplama kaynakları bulunduğunda önemli bir avantaj sağlar. GRU, metin üretimi, zaman serisi analizi ve konuşma tanıma gibi birçok alanda yaygın olarak kullanılmaktadır.

GRU ile hazırlanmış modelin yapısı Şekil 4.14'te gösterilmiştir. Modelin uygulanmasında, veri ön işleme adımları tamamlandıktan sonra GRU tabanlı derin öğrenme yapısı ardışık biçimde tanımlanmıştır. Giriş verisi olarak zaman serisi şeklinde yapılandırılmış olan el hareketi verileri kullanılmaktadır. Modelin ilk iki katmanında, sırasıyla 512 birimden oluşan iki adet GRU bloğu yer almakta olup, ilk katmanda `return_sequences=True` parametresiyle her zaman adımına ilişkin çıktı üretilmektedir. Bu sayede ikinci GRU katmanına tam bir ardışık bilgi aktarımı sağlanır. İkinci katman ise dizinin sonundaki zaman adımına ait özet çıktıyı üretir. Her iki katmanın ardından Dropout uygulanmış ve modelin öğrenme sürecinde aşırı uyum göstermesi engellenmiştir.



Şekil 4.14 GRU Model Yapısı

GRU bloklarını takiben, özellikleri ayrıştıran iki yoğun bağlantılı (Dense) katman yer almaktadır. Bunlardan ilki 128, ikincisi ise 256 nörona sahiptir. Her iki katmanda da ReLU aktivasyon fonksiyonu kullanılmıştır. Bu katmanların ardından Batch Normalization işlemleri gerçekleştirilmiş ve böylece modelin öğrenme sürecindeki içsel dağılım kaymalarının azaltılması hedeflenmiştir. Sonuç katmanında ise, tahmin edilmesi gereken sınıf sayısına denk gelen nöronlar yer almakta ve çok sınıflı sınıflandırma için yaygın olarak tercih edilen softmax aktivasyonu ile model sonlandırılmaktadır.

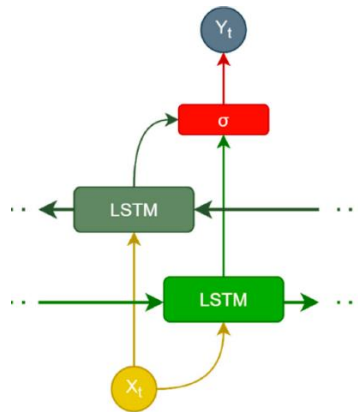
Bu mimari sayesinde, model hem zamana bağlı örüntüleri etkili biçimde öğrenebilmekte hem de GRU'nun parametre verimliliği sayesinde daha kısa sürede eğitilebilmektedir. Eğitim sürecinde elde edilen başarı oranları, modelin Türkçe işaret dili tanıma görevinde güçlü bir sınıflandırma performansı sergilediğini

göstermektedir. GRU'nun daha az sayıda kapı mekanizması içermesi, özellikle eğitim süresinin önemli olduğu gerçek zamanlı uygulamalarda büyük bir avantaj sağlamaktadır.

Modelin eğitimi sırasında, optimizasyon algoritması olarak yine Adamax yöntemi tercih edilmiş; düşük öğrenme oranı (0.0001) ile modelin parametre güncellemeleri daha kararlı hâle getirilmiştir. Kayıp fonksiyonu olarak categorical crossentropy kullanılmış ve eğitim işlemleri 32'lik mini-batch'lerle gerçekleştirilmiştir. Aşırı öğrenmeyi önlemek için EarlyStopping ve öğrenme oranını gerektiğinde azaltmak için ReduceLROnPlateau mekanizmaları etkinleştirilmiştir. En iyi doğrulama başarımını gösteren model ağırlıkları, ModelCheckpoint yöntemiyle kaydedilerek test aşamasında kullanılmıştır.

4.5.3 Bi-LSTM

Bi-LSTM (Bidirectional Long Short-Term Memory - Çift Yönlü Uzun Kısa Süreli Bellek) ağları hem ileri hem de geri yönlerde bağımlılıkları yakalayarak sıralı verileri işlemek için tasarlanmış gelişmiş sinir ağı mimarileridir. Bu ikili işleme yeteneği, modelin çeşitli uygulamalarda, özellikle geçmiş ve gelecekteki bağlamları anlamının çok önemli olduğu zaman serisi tahmininde performansını artırır (Şekil 4.15).



Şekil 4.15 Bi-LSTM Modeli (Crisóstomo de Castro Filho vd., 2020).

Bi-LSTM, biri geçmişten geleceğe, diğeri gelecekteki geçmişe veri işleyen iki LSTM katmanını kullanır ve bu da zamansal bağımlılıkların kapsamlı bir şekilde anlaşılmasına olanak tanır (Xu vd., 2024)). Çalışmalar, Bi-LSTM modellerinin, yağış tahmini ve

yaprak alanı indeksi tahmini gibi uygulamalarda RMSE ve R^2 gibi doğruluk ölçümlerinde geleneksel LSTM modellerinden daha iyi performans gösterdiğini göstermektedir (Liu vd., 2022).

Bu çalışmada, işaret dili tanıma görevine yönelik olarak ardışık verilerdeki bağlamsal ilişkileri hem ileri hem de geri yönde analiz edebilen çift yönlü Uzun Kısa Süreli Bellek (Bidirectional LSTM – BiLSTM) mimarisi kullanılmıştır. Bu yapı, özellikle hareketlerin zaman içindeki sırasına duyarlı olan uygulamalarda geçmiş ve gelecek bağlamı aynı anda modelleyerek sınıflandırma başarımını artırmaktadır.

Modelin giriş katmanı, zaman serisi biçiminde yapılandırılmış el hareketi verilerini alacak şekilde tanımlanmıştır. Ardışık bilgiyi işlemek üzere ardışık olarak yerleştirilmiş iki adet Bidirectional LSTM katmanı kullanılmıştır. Her biri 256 birimden oluşan bu katmanlardan ilki, tüm zaman adımlarına ait çıktıları üretirken (return_sequences=True), ikincisi yalnızca dizinin sonundaki bilgiye karşılık gelen özet temsili çıkarmaktadır. Bu katmanların ardından %30 oranında Dropout uygulanarak modelin aşırı öğrenmeye karşı dayanıklılığı artırılmıştır.

BiLSTM katmanlarının ardından gelen yoğun bağlantılı (Dense) yapılar, öğrenilen temsilleri soyut ve ayrıştırıcı hale getirmek üzere yapılandırılmıştır. İlk dense katman 128, ikinci katman ise 256 nörondan oluşmakta olup her iki katmanda da ReLU aktivasyon fonksiyonu kullanılmıştır. Bu katmanları takip eden Batch Normalization işlemleri, eğitim sırasında öğrenme sürecinin dengelenmesini ve hızlandırılmasını sağlamaktadır. Modelin son katmanında ise, sınıflandırılması hedeflenen işaret sayısına karşılık gelen 446 nöron bulunmaktadır ve çok sınıflı sınıflandırma için uygun olan softmax aktivasyon fonksiyonu ile tamamlanmaktadır.

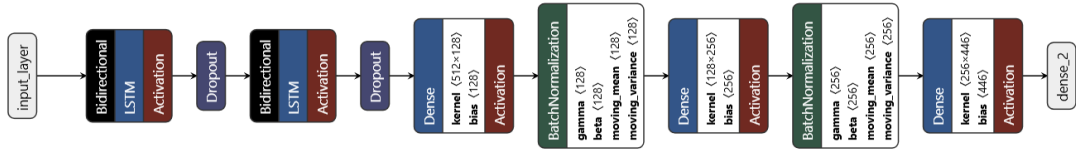
Modelin genel mimari yapısı, Şekil 4.16'de görselleştirilmiştir. Görselden de anlaşılacağı üzere, BiLSTM katmanlarıyla elde edilen çift yönlü bağlamsal temsil ve ardından gelen tam bağlantılı katmanlar sayesinde, model karmaşık hareket örüntülerini başarılı bir şekilde sınıflandırabilmektedir.

Modelin eğitimi sırasında optimizasyon algoritması olarak Adamax tercih edilmiştir. Bu algoritma, Adam ailesine ait olup özellikle yüksek boyutlu parametre uzaylarında

kararlı öğrenme sağlayabilmektedir. Öğrenme oranı olarak 0.0001 değeri kullanılmış, kayıp fonksiyonu olarak ise çok sınıflı sınıflandırma problemlerinde yaygın şekilde tercih edilen categorical crossentropy uygulanmıştır. Eğitim işlemleri 32 örnek içeren mini-batch'lerle gerçekleştirilmiş ve maksimum epoch sayısı 1000 olarak belirlenmiştir.

Modelin doğrulama başarımını izlemek ve aşırı öğrenmeyi önlemek amacıyla çeşitli geri çağırma (callback) mekanizmaları kullanılmıştır. EarlyStopping, doğrulama kaybında belirli bir süre boyunca iyileşme gözlenmediğinde eğitim sürecini erken sonlandırmak üzere etkinleştirilmiştir. Bununla birlikte, öğrenme oranının adaptif biçimde güncellenebilmesi için ReduceLROnPlateau mekanizması kullanılmış; doğrulama kaybında gelişme görülmeyen durumlarda öğrenme oranı azaltılarak daha hassas optimizasyon sağlanmıştır. Ayrıca, doğrulama kümesindeki en iyi performansa sahip modelin ağırlıkları ModelCheckpoint yöntemiyle kaydedilmiş ve test aşamasında bu ağırlıklar kullanılmıştır.

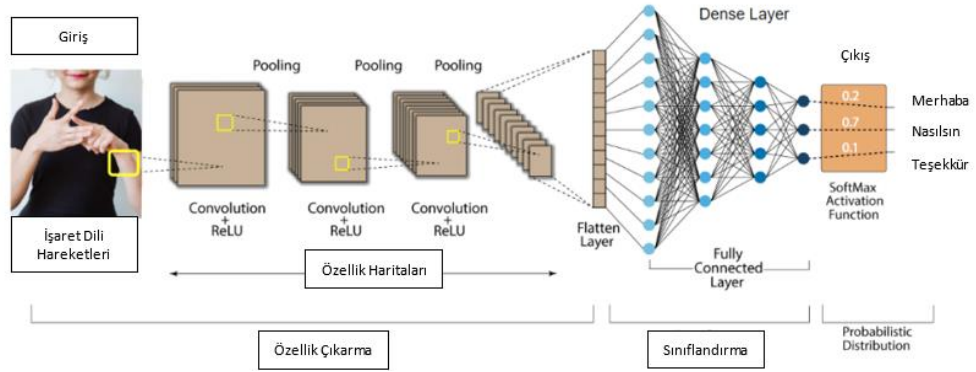
Bu BiLSTM temelli yapı, işaret dili gibi zamansal yapıya sahip, karmaşık ve bağlama duyarlı sınıflandırma problemleri için güçlü ve genellenebilir bir çözüm sunmaktadır.



Şekil 4.16 Bi-LSTM Model Yapısı

4.5.4 CNN (Convolutional Neural Networks)

CNN (Convolutional Neural Networks – Evrimsel Sinir Ağları), görüntü ve video işleme için kullanılır ve özellikle görüntü sınıflandırma, nesne algılama ve yüz tanıma gibi görevlerde etkilidir (Kasapbaşı vd., 2023). CNN, görüntü ve video işleme görevleri için kullanılan derin öğrenme algoritmalarıdır. Model yapısında, özellikleri çıkarmak için 3 katmanlı bir yapı oluşturulmuştur. Burada aktivasyon fonksiyonu olarak ReLU kullanılmıştır. Ardından sınıflandırma için Full Connected Layer kullanılmıştır. Burada da aktivasyon fonksiyonu olarak SoftMax kullanılmıştır. Oluşturulan modelin genel yapısı Şekil 4.17’de gösterilmektedir.



Şekil 4.17 CNN Sınıflayıcı Genel Yapısı

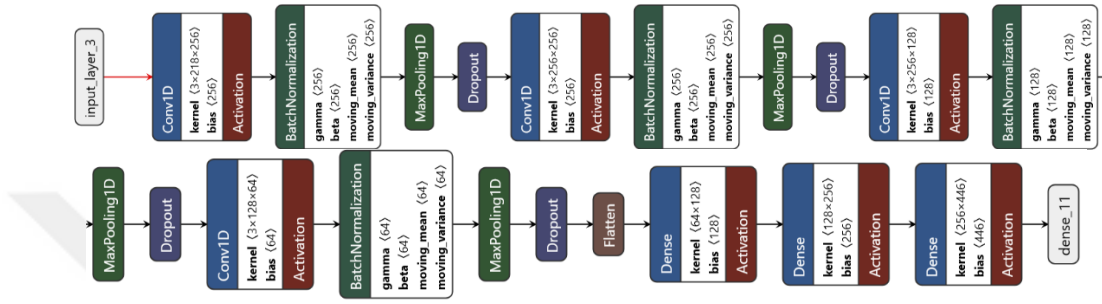
Bu çalışmada, zamana bağlı işaret dili hareketlerinden anlamlı örüntüler çıkarmak ve sınıflandırmak amacıyla 1-boyutlu Evrişimsel Sinir Ağı (1D CNN) mimarisi kullanılmıştır. CNN mimarileri, özellikle lokal örüntü çıkarımı konusunda güçlü yapılar olduğundan, zaman serisi formatındaki el hareketi verilerinde uzamsal özelliklerin öğrenilmesinde tercih edilmiştir.

Model, giriş olarak sabit uzunluktaki zaman serisi verisini kabul etmektedir. İlk dört katman bloğu, sırasıyla 256, 256, 128 ve 64 filtreden oluşan dört Conv1D katmanıdır. Her bir konvolüsyon katmanı, 3 boyutunda filtrelerle çalışmakta olup ReLU aktivasyon fonksiyonu ile etkinleştirilmiştir. Bu katmanların ardından sırasıyla Batch Normalization, MaxPooling1D ve Dropout (%30) işlemleri uygulanmıştır. Bu yapı, evrişim sonrası elde edilen özellik haritalarının standartlaştırılmasını, boyutsal azaltma yoluyla öznetelik yoğunluğunun dengelenmesini ve modelin aşırı öğrenmeye karşı dayanıklılığını sağlar.

Son evrişim bloğunun ardından, çok boyutlu yapı düzleştirilmek üzere bir Flatten katmanı uygulanır. Bu katman, özellik haritalarını tek boyutlu bir vektöre dönüştürerek tam bağlantılı katmanlara geçiş sağlar.

Tam bağlantılı (dense) bölümde, sırasıyla 128 ve 256 birimden oluşan iki adet Dense katman yer almaktadır. Her biri ReLU aktivasyonu ile çalışmakta ve modelin karar mekanizmasını derinleştirmektedir. Bu katmanların ardından gelen çıktı katmanı, sınıflandırılması hedeflenen işaret sayısına denk gelen 446 birim içermektedir ve çok sınıflı sınıflandırma için softmax aktivasyonu ile yapılandırılmıştır.

Modelin mimari düzeni, Şekil 4.18’de detaylı şekilde sunulmuştur. Katmanlar arası geçişlerin istikrarlı olması için her evrişim bloğunda batch normalization uygulanmış, böylece öğrenme süreci boyunca içsel dağılım kaymalarının önüne geçilmiştir. Modelin sonunda yer alan üç yoğun katman, CNN katmanlarında öğrenilen lokal örüntüleri sınıflandırmaya yönelik global temsillere dönüştürerek son kararı vermektedir.

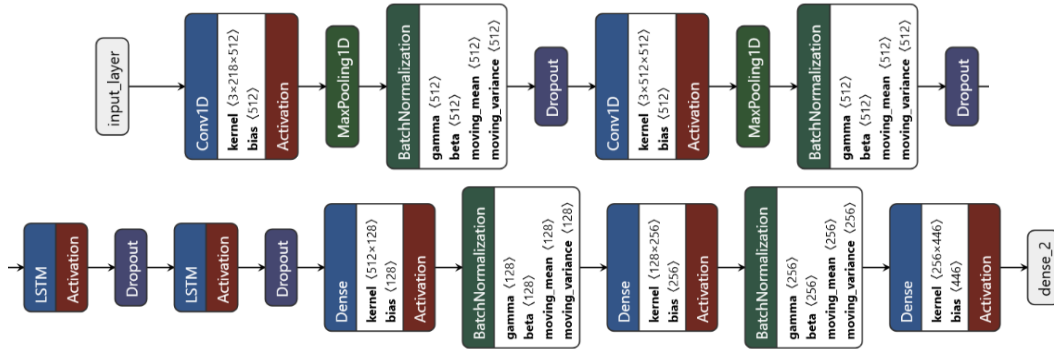


Şekil 4.18 CNN Model Yapısı

Bu yapı, hızlı eğitilebilmesi ve düşük hesaplama maliyeti ile özellikle gerçek zamanlı işaret dili tanıma uygulamaları için uygun bir çözüm sunmaktadır.

4.5.5 CNN + LSTM

Yerel hareket örüntülerini ve zamana bağlı ilişkileri aynı yapıda birleştirebilen CNN + LSTM hibrit mimarisi, işaret dili tanıma problemine yönelik güçlü bir çözüm sunar. Şekil 4.19’da gösterilen bu yapı, evrimsel katmanlar aracılığıyla el hareketlerinden yerel öznitelikler çıkartırken, LSTM katmanları sayesinde ardışık verilerdeki bağlamsal ilişkileri yakalamayı hedefler.



Şekil 4.19 CNN + LSTM Hibrit Model Yapısı

Modelin başlangıç katmanları, sırasıyla 512 filtrelili iki Conv1D katmanından oluşur. Her evrişim katmanını, sırasıyla MaxPooling1D, Batch Normalization ve Dropout (%10) işlemleri izler. Bu üçlü yapı, modelin öğrenme sürecini daha kararlı ve genellenebilir hâle getirir. Konvolüsyon bloğu, hareket verilerindeki lokal örüntülerin etkili biçimde çıkarılmasını sağlar.

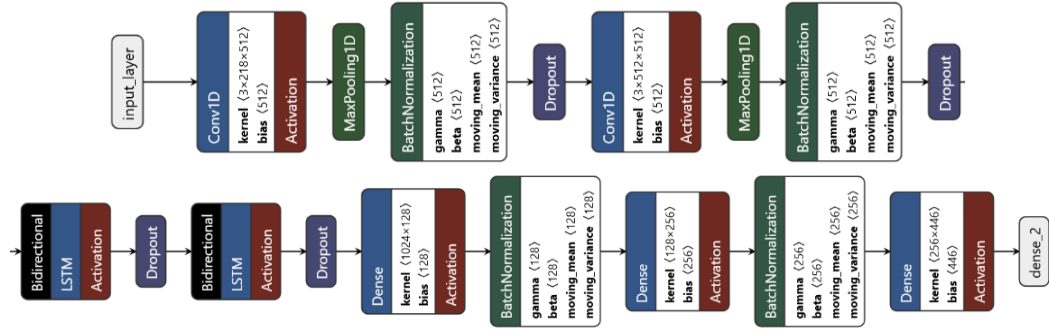
Evrişimsel çıkış daha sonra iki ardışık LSTM katmanına aktarılır. İlk LSTM katmanı her zaman adımındaki çıktıyı aktaracak şekilde `return_sequences=True` ile yapılandırılmıştır. İkinci LSTM katmanı ise bu çıktıyı tek bir vektöre özetler. LSTM sonrası yerleştirilen Dropout katmanları, modelin fazla öğrenmesini engellemek amacıyla aktif edilmiştir.

Zamansal bilgiyi barındıran bu temsil, sırasıyla 128 ve 256 nöron içeren iki Dense katmana gönderilir. Bu katmanlarda ReLU aktivasyonu tercih edilmiş; ardından gelen Batch Normalization katmanları ile öğrenme süreci dengelenmiştir. Modelin son katmanında, sınıflandırılacak toplam işaret sayısına (446) karşılık gelen nöronlar bulunur ve softmax aktivasyonu yardımıyla çok sınıflı bir çıktı elde edilir.

CNN ve LSTM yapılarını bir araya getiren bu mimari hem uzamsal hem de zamansal bilgiye aynı anda odaklanarak, işaret dili gibi karmaşık ardışık örüntülere sahip görevlerde yüksek sınıflandırma başarımı sağlamaktadır.

4.5.6 CNN + BiLSTM

Yerel öznitelik çıkarımı ile zaman içindeki bağlamsal bilgiyi birlikte değerlendirebilen bir yapı oluşturmak amacıyla, evrişimsel sinir ağları (CNN) ve çift yönlü uzun kısa süreli bellek (BiLSTM) katmanları bir araya getirilmiştir. Şekil 4.20’de görselleştirilen bu hibrit mimari, işaret dili gibi zamana yayılan karmaşık hareket dizilerinin sınıflandırılmasında güçlü bir çözüm sunmaktadır.



Şekil 4.20 CNN + BiLSTM Hibrit Model Yapısı

Modelin ilk aşamasında, giriş verileri iki adet 1-boyutlu Conv1D katmanından geçirilir. Her iki katman da 512 filtre içermekte olup, ReLU aktivasyonu ile çalışmaktadır. Bu katmanların ardından sırasıyla MaxPooling1D, Batch Normalization ve %10 oranında Dropout uygulanır. Bu işlemler modelin öğrenme sürecini dengelerken aşırı uyumlamayı önlemekte ve yerel hareket örüntülerinin öne çıkarılmasını sağlamaktadır.

Evrişimsel yapıdan elde edilen özellik haritaları, zaman sıralı bilgiyle çalışmaya uygun olan iki katmanlı Bidirectional LSTM yapısına aktarılır. Her iki BiLSTM katmanı 512 birimden oluşur. İlk katman tüm zaman adımlarına ilişkin çıktıları üretirken (return_sequences=True), ikinci katman bu çıktıları özetleyerek bir vektöre indirger. Her iki BiLSTM bloğunun ardından %20 oranında Dropout işlemi uygulanmıştır.

LSTM çıktısı sırasıyla 128 ve 256 nörondan oluşan iki adet Dense katmana aktarılır. Her biri ReLU aktivasyon fonksiyonu ile etkinleştirilmiş bu katmanlar, sırasıyla Batch Normalization işlemleri ile desteklenmiştir. Modelin son katmanı, sınıflandırılması hedeflenen toplam işaret sayısı kadar nöron içerir (446) ve softmax aktivasyonu ile çok sınıflı çıktı üretmektedir.

Optimizasyon sürecinde Adam algoritması tercih edilmiştir. 0.00001'lik düşük bir öğrenme oranı ile eğitilen model, categorical crossentropy kayıp fonksiyonu ile derlenmiştir. Eğitim işlemleri 32'lik mini-batch'lerle gerçekleştirilmiş, maksimum 150 epoch ile sınırlandırılmıştır. Eğitim sırasında doğrulama başarımı gözlemlenmiş ve EarlyStopping, ReduceLROnPlateau ile öğrenme dinamikleri optimize edilmiştir. En iyi model ağırlıkları ModelCheckpoint ile kaydedilmiştir.

Mimari düzeyde hem evrişimsel hem de ardışık öğrenme bloklarını içeren bu hibrit yapı, işaret dilinde zamana yayılan hareketlerin etkili biçimde modellenmesini sağlamış; test sonuçları, doğruluk değeri ve F1-Macro skorları açısından yüksek performans sergilemiştir.

4.5.7 CNN+GRU

Convolutional Neural Network (CNN) ve Gated Recurrent Unit (GRU) kombinasyonu, her iki modelin güçlü yönlerini bir araya getirerek karmaşık veri türleri için etkili bir mimari sunar. CNN'ler, evrişim ve havuzlama işlemleri aracılığıyla verilerden mekânsal ve yerel özellikler çıkarmada oldukça başarılıdır ve bu nedenle görüntü işleme ve doğal dil anlama gibi görevlerde sıkça kullanılır. Diğer yandan, GRU'lar, güncelleme ve sıfırlama kapıları gibi mekanizmalarla bilgi akışını düzenleyerek zamansal bağımlılıkları ve ardışık ilişkileri modellemede etkilidir. GRU'ların, LSTM'lere kıyasla daha az parametre içermesi, uzun süreli bağımlılıkları daha hızlı ve verimli bir şekilde öğrenmesini sağlar (Wang vd., 2023).

Bu çalışmada, hem uzamsal (spatial) hem de zamansal (temporal) özellikleri aynı model içerisinde çıkararak işaret dili tanıma başarımını artırmak amacıyla CNN ve GRU yapılarının birleştirildiği hibrit bir derin öğrenme mimarisi tasarlanmıştır. CNN katmanları ile giriş verisindeki yerel örüntüler yakalanırken, GRU katmanları zaman içindeki bağımlılıkları öğrenerek sınıflandırma sürecine bağlamsal derinlik kazandırmaktadır.

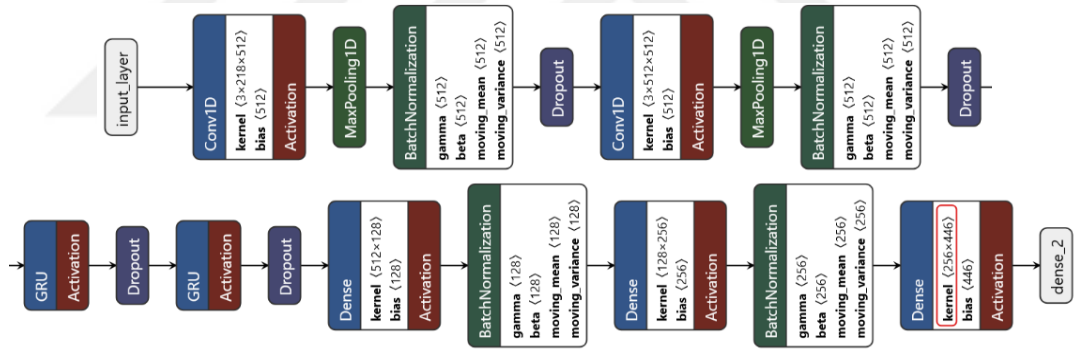
Modelin giriş katmanı, zaman serisi biçiminde yapılandırılmış el hareketi verilerini kabul eder. İlk iki katman bloğunda yer alan Conv1D katmanları sırasıyla 512 filtre ve 3'lük çekirdek boyutuyla yapılandırılmıştır. Her bir konvolüsyon işlemini takiben MaxPooling1D ile boyut azaltma, Batch Normalization ile içsel dağılım dengeleme ve Dropout ile aşırı öğrenmenin önüne geçme mekanizmaları devreye alınmıştır. Bu yapı, modelin istikrarlı bir şekilde öğrenmesini ve genellenebilirliğini artırmaktadır.

CNN katmanlarının ardından, zamansal ilişkilerin öğrenilebilmesi için ardışık olarak iki adet GRU katmanı yerleştirilmiştir. Her iki GRU katmanında da, zaman içindeki örüntüler sırasıyla işlenerek hareket dizilerinin ardışık bağlamı modellenmiştir. GRU

katmanlarının her biri 512 nöron içerir ve ardından gelen Dropout katmanları ile aşırı öğrenme riski azaltılmıştır.

GRU yapısının çıktısı, sırasıyla 128 ve 256 birimden oluşan iki adet Dense (yoğun bağlantılı) katmana aktarılmıştır. Bu katmanlarda ReLU aktivasyon fonksiyonu kullanılmış, her biri sonrasında Batch Normalization uygulanarak öğrenmenin dengelenmesi sağlanmıştır. Son olarak, sınıflandırılması hedeflenen işaret sayısı kadar nöron içeren ve softmax aktivasyon fonksiyonu ile yapılandırılmış bir çıkış katmanı modelin karar mekanizmasını tamamlamaktadır.

Modelin mimarisi, Şekil 4.21’de görsel olarak sunulmuştur. CNN bloklarının lokal özellik çıkarımı yeteneği ile GRU’nun zaman içindeki örüntüleri işleme gücü birleştirilerek, modelin hem doğruluk değeri hem de bağlamsal duyarlılığı artırılmıştır. Bu hibrit yaklaşım, özellikle ardışık ve karmaşık hareket serilerini içeren işaret dili tanıma görevlerinde güçlü bir performans sergilemiştir.



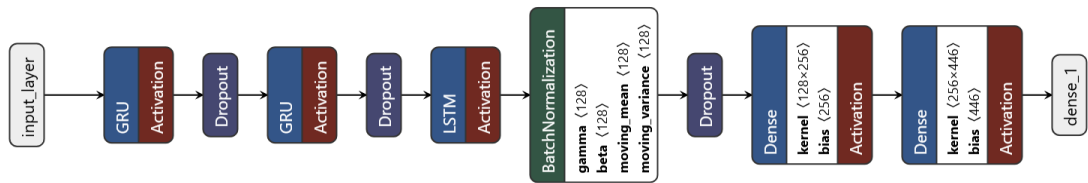
Şekil 4.21 CNN+GRU Model Yapısı

4.5.8 GRU+LSTM

Gated Recurrent Unit (GRU) ve Long Short-Term Memory (LSTM) modellerinin bir arada kullanılması, zaman serisi ve ardışık veri işlemede farklı bağımlılık türlerini modellemek için etkili bir yaklaşımdır. GRU, daha az parametre ile hızlı bir şekilde eğitilir ve kısa vadeli bağımlılıkları öğrenmede avantajlıdır. LSTM ise unutma, giriş ve çıkış kapıları sayesinde daha karmaşık uzun vadeli bağımlılıkları modellemeye odaklanır. Bu iki modelin birlikte kullanılması, bir yandan GRU’nun hız ve hesaplama avantajlarını sunarken, diğer yandan LSTM’nin derin bağlam öğrenme kapasitesinden faydalanmayı mümkün kılar (Pirani vd., 2022).

Bu yaklaşım, özellikle hem kısa hem de uzun vadeli ilişkilerin kritik olduğu uygulamalarda tercih edilmektedir. Örneğin, konuşma tanıma, doğal dil işleme ve biyomedikal sinyal işleme gibi alanlarda GRU ve LSTM'nin farklı görevlerde işlevsel hale getirilmesi, modelin genel performansını artırır. Genellikle GRU, verilerden hızlı özellik çıkarımı yaparken, LSTM uzun süreli bağlamların işlenmesinde devreye girer. Böyle bir yapı, zaman serilerindeki eğilimleri ve kalıpları öğrenirken aynı zamanda hesaplama yükünü dengeler. Ayrıca, bu kombinasyonun farklı veri türlerine uyarlanabilir olması, akademik araştırmalardan endüstriyel uygulamalara kadar geniş bir yelpazede kullanılmasını sağlamaktadır (Chakraborty vd., 2023).

Farklı türdeki ardışık katmanları bir araya getirerek daha zengin bir zaman serisi modellemesi elde etmek amacıyla, GRU ve LSTM yapılarının birleşiminden oluşan hibrit bir mimari tasarlanmıştır. Şekil 4.22'de gösterilen bu yapı, özellikle uzun ardışık bağımlılıklar içeren işaret dili hareketlerini tanımada güçlü bir performans sergilemektedir.



Şekil 4.22 GRU+LSTM Model Yapısı

Modelin ilk aşamasında iki ardışık GRU (Gated Recurrent Unit) katmanı yer almaktadır. İlk GRU katmanı, zaman adımlarına göre tüm ardışık çıktıları üretmek üzere return_sequences=True parametresiyle yapılandırılmıştır. İkinci GRU katmanı ise 256 nöronla daha kompakt bir temsil oluşturur. Her iki katmanın ardından Dropout uygulanarak modelin aşırı öğrenme riskine karşı dayanıklılığı artırılmıştır.

GRU bloklarının çıktısı, daha derin zamansal ilişkileri öğrenmek amacıyla bir adet LSTM (Long Short-Term Memory) katmanına aktarılır. LSTM katmanı, geçmiş zaman adımlarından gelen bilgiyi daha uzun süre saklayabilme özelliği sayesinde GRU'dan elde edilen özellikleri daha güçlü şekilde özetlemektedir. LSTM'in ardından Batch Normalization ve %20 oranında Dropout uygulanmıştır. Bu katmanlar, modelin öğrenme sürecini hem dengeli hem de genellenebilir kılar.

Modelin karar yapısı, sırasıyla 256 ve 446 nörondan oluşan iki adet Dense (yoğun bağlantılı) katmandan oluşmaktadır. İlk dense katmanda ReLU aktivasyon fonksiyonu kullanılırken, son katman softmax aktivasyonu ile çok sınıflı çıktı üretmektedir. Böylece model, işaretlerin olasılık dağılımına göre en uygun sınıf tahminini yapabilmektedir.

Modelin optimizasyonunda Adam algoritması tercih edilmiş ve 0.0001'lik düşük öğrenme oranı ile parametre güncellemeleri kademeli olarak yapılmıştır. Eğitim süreci boyunca categorical crossentropy kayıp fonksiyonu kullanılmış, eğitim işlemi 32'lik mini-batch'lerle yürütülmüştür. Eğitim sürecinde EarlyStopping, ReduceLROnPlateau ve ModelCheckpoint gibi callback mekanizmaları etkin biçimde kullanılarak en iyi model ağırlıkları kaydedilmiştir.

Bu hibrit yapı, hem GRU'nun hesaplama açısından verimli yapısından hem de LSTM'in uzun vadeli bağımlılıkları öğrenme yeteneğinden faydalanarak, işaret dili tanıma görevinde yüksek doğruluk ve F1-Macro başarımı sunmuştur.

4.5.9 GRU+Bi-LSTM

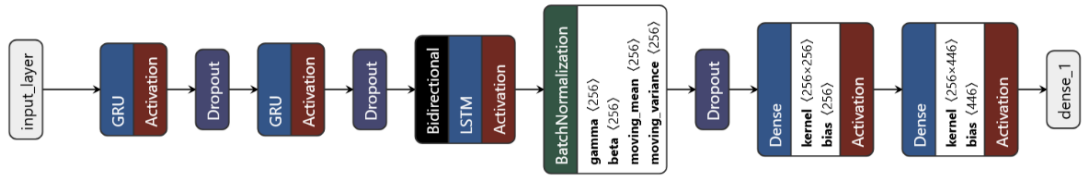
Gated Recurrent Unit (GRU) ve Bidirectional Long Short-Term Memory (Bi-LSTM) modellerinin birleştirilmesi, ardışık ve zaman serisi verilerle çalışmak için güçlü bir mimari oluşturur. GRU, daha az parametre ve hızlı eğitimi sayesinde hesaplama açısından verimli bir modeldir. Özellikle kısa vadeli bağımlılıkları öğrenmede etkili olan GRU, LSTM'ye göre daha sade bir yapıya sahiptir. Bi-LSTM ise geleneksel LSTM'nin iki yönlü bir versiyonudur; veriyi hem ileri hem de geri yönde işler. Bu özellik, sekans verilerinde hem geçmiş hem de gelecekteki bağlamı aynı anda öğrenmeyi mümkün kılar ve özellikle bağlamın tamamının kritik olduğu uygulamalarda avantaj sağlar (Karim vd., 2022).

Bu kombinasyon genellikle ardışık özelliklerin çıkarılması ve çift yönlü bağlamın modellenmesi gereken görevlerde kullanılır. GRU, veri girişinde hızlı ve etkin bir özellik çıkarımı yaparak modelin hesaplama yükünü azaltırken, Bi-LSTM derinlemesine bağlam analizi yapar ve ileri-geri bağlam ilişkilerini modelleyerek daha güçlü bir anlam çıkarımı sağlar. Bu nedenle, GRU ve Bi-LSTM'nin bir arada

kullanımı, dil modelleme, konuşma tanıma ve biyomedikal sinyal işleme gibi alanlarda etkili sonuçlar verir.

Örneğin, doğal dil işleme uygulamalarında GRU, metindeki kısa vadeli ilişkileri öğrenirken, Bi-LSTM aynı metni çift yönlü olarak analiz ederek anlamı bütüncül bir şekilde modelleyebilir. Bu yaklaşım, özellikle metin, konuşma veya biyolojik sinyal gibi ardışık veri türlerinde bağlamdan bağımsız anlam kaybını en aza indirmede yararlıdır. Ek olarak, bu mimari, görev gereksinimlerine bağlı olarak hiperparametre optimizasyonu ve katman düzenlemesiyle esnek bir şekilde özelleştirilebilir. Bu sayede, GRU'nun verimliliği ile Bi-LSTM'nin bağlam öğrenme gücü birleştirilerek hem performans hem de hesaplama açısından dengeli bir model elde edilebilir (Chaikaew vd., 2021).

Zamansal bağımlılıkları farklı yönlerden analiz edebilmek amacıyla Gated Recurrent Unit (GRU) ve çift yönlü LSTM (BiLSTM) yapılarını bir araya getiren hibrit bir mimari tasarlanmıştır. Bu yapı, hem kısa vadeli ilişkileri düşük hesaplama maliyetiyle öğrenebilen GRU bloklarını hem de uzun vadeli bağımlılıkları çift yönlü biçimde modelleyebilen LSTM katmanlarını aynı ağda birleştirerek performans avantajı sağlamaktadır. Katman yapısı Şekil 4.23'te görselleştirilmiştir.



Şekil 4.23 GRU+Bi-LSTM Model Yapısı

Model, giriş olarak zaman sıralı işaret dili verilerini almakta ve ilk olarak iki ardışık GRU katmanından geçirilmektedir. İlk GRU katmanı return_sequences=True ayarıyla zaman eksenindeki tüm adımlara ait çıktıları üretir; ikinci GRU katmanı ise bu çıktıyı daha kompakt hâle getirerek modeli derinleştirir. Her GRU katmanının ardından %30 oranında Dropout işlemi uygulanarak öğrenmenin genelleştirilebilirliği artırılmıştır.

GRU bloklarının ardından gelen Bidirectional LSTM katmanı, ardışık hareket dizilerini hem geçmişten geleceğe hem de gelecekte geçmişe okuyarak daha zengin

bağlamsal temsiller oluşturur. Bu katman, özellikle işaret dilindeki simetrik veya ters yönlü hareketlerin etkili şekilde öğrenilmesini sağlar. LSTM bloğu sonrasında Batch Normalization uygulanmış ve ardından tekrar %30 oranında Dropout ile modelin aşırı öğrenmeye eğilimi azaltılmıştır.

Ağın karar mekanizmasını oluşturan Dense katmanları, sırasıyla 256 ve 446 birimden oluşur. İlk dense katmanda ReLU aktivasyon fonksiyonu kullanılmış, ikinci dense katman ise çok sınıflı sınıflandırma görevine uygun olarak softmax aktivasyonu ile yapılandırılmıştır. Bu sayede model, her örnek için olasılık dağılımı üreterek sınıf tahmininde bulunmaktadır.

Eğitim sürecinde optimizasyon algoritması olarak Adamax tercih edilmiş ve öğrenme oranı 0.0001 olarak ayarlanmıştır. Modelin kaybı, çok sınıflı sınıflandırma problemlerinde standart olarak kullanılan categorical crossentropy ile hesaplanmıştır. Eğitim 32'lik mini-batch'lerle 150 epoch'a kadar gerçekleştirilmiş; doğrulama kaybı izlenerek EarlyStopping, ReduceLROnPlateau ve ModelCheckpoint geri çağırma mekanizmaları devreye alınmıştır. Böylece doğrulama başarımı en yüksek olan model ağırlıkları korunmuştur.

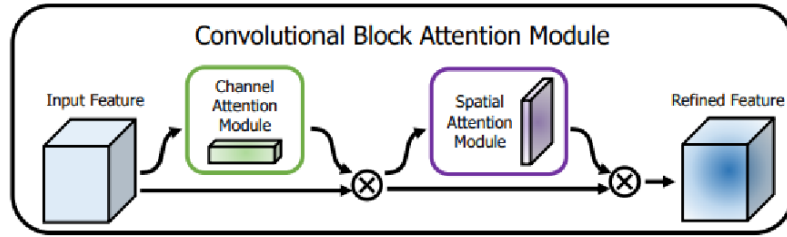
Sonuçlar değerlendirildiğinde, GRU ve BiLSTM yapılarını birleştiren bu hibrit yaklaşımın hem işlem verimliliği hem de sınıflandırma başarımı açısından tatmin edici sonuçlar verdiği görülmektedir. Model, işaret dili gibi yüksek boyutlu ve zaman bağımlı verilerde bağlamsal bütünlüğü etkili bir şekilde öğrenebilmektedir.

4.5.10 CBAM (Convolutional Block Attention Module)

CBAM (Convolutional Block Attention Module - Evrimsel Blok Dikkat Modülü), geniş ölçekli sınıflandırma problemlerinde derin sinir ağlarının (özellikle evrimsel sinir ağlarının) performansını artırmak amacıyla Woo ve arkadaşları tarafından önerilen bir dikkat (attention) mekanizmasıdır (Woo vd, 2018). Bu modül, ardışık olarak uygulanan kanal dikkat modülü (channel attention module) ve uzamsal dikkat modülü (spatial attention module) olmak üzere iki bileşenden oluşmaktadır. CBAM, giriş görüntülerden elde edilen özellik haritalarını önemli bölgelere odaklanarak rafine eder ve böylece daha anlamlı temsil öznetelikleri oluşturur.

Kanal dikkat modülü, girişten elde edilen özniteliklerin kanallar arası önem derecelerini değerlendirerek, daha anlamlı kanalların öne çıkarılmasını sağlar. Bu işlemin ardından gelen uzamsal dikkat modülünde ise, ortalama havuzlama (average pooling) ve maksimum havuzlama (max pooling) işlemleri birlikte kullanılarak öznitelik haritasının uzamsal boyuttaki dikkat bilgisi çıkarılır. Böylece modelin, hangi bölgelere odaklanması gerektiği ve hangi bölgelerin etkisinin azaltılması gerektiği konusunda yönlendirilmesi sağlanır.

CBAM modülü sayesinde, ağıın içinde daha etkin ve odaklanmış bir bilgi akışı gerçekleştirilerek, sınıflandırma doğruluğu ve genel öğrenme performansı önemli ölçüde iyileştirilebilmektedir. İlgili modülün yapısı Şekil 4.24'te sunulmaktadır.



Şekil 4.24 CBAM Modül Yapısı (Woo vd, 2018)

4.5.11 ResNet-18 (Residual Network – 18)

Derin öğrenme modelleri, özellikle görüntü sınıflandırma görevlerinde önemli başarılar elde etmiştir. Ancak, model derinliği arttıkça öğrenme sürecinde karşılaşılan gradyan sönmesi gibi problemler, model performansını olumsuz etkileyebilmektedir. Bu sorunu çözmek amacıyla He vd., (2016) tarafından geliştirilen Residual Network (ResNet) mimarisi, öğrenme sürecini kolaylaştırmak için rezidüel öğrenme yaklaşımını benimsemiştir. Bu mimari, derin ağların daha kolay eğitilebilmesini sağlamak amacıyla atlama bağlantıları kullanarak giriş bilgisinin doğrudan daha ileri katmanlara iletilmesini mümkün kılar.

ResNet mimarisinin 18 katmanlı versiyonu olan ResNet-18, bu yapının en temel ve yaygın kullanılan örneklerinden biridir. ResNet-18, 8 adet rezidüel bloktan oluşmakta ve her blok, iki adet 3×3 evrişim (convolution) katmanı içermektedir. Bu yapı, modelin parametre sayısını sınırlı tutarken, derinlikten gelen avantajları korumasını sağlar.

Modelin ilk katmanında geniş çekirdek boyutuna sahip (7×7) bir evrişim katmanı yer alır ve bu katmanı bir maksimum havuzlama (max pooling) katmanı takip eder. Ardından gelen dört farklı evrişim bloğu, sırasıyla 64, 128, 256 ve 512 adet filtre içeren yapıdadır. Her artık blokta, giriş ile blok çıktısı arasında yapılan toplama işlemi sayesinde, modelin yalnızca yeni bilgi öğrenmesine olanak tanınır; bu sayede öğrenilmesi zor olan fonksiyonlar yerine, daha kolay optimize edilebilen artık fonksiyonlar öğrenilmiş olur. Son bölümde global ortalama havuzlama uygulanır ve sınıflandırma işlemi için tam bağlantılı bir katman kullanılır (He vd., 2016).

ResNet-18, özellikle hesaplama kaynaklarının sınırlı olduğu durumlarda derin öğrenme uygulamaları için uygun bir mimaridir. Daha derin versiyonları (örneğin ResNet-50, ResNet-101) daha yüksek doğruluk sunabilse de ResNet-18 modeli daha az eğitim süresi ve daha düşük donanım gereksinimleri ile etkili sonuçlar üretebilmektedir. Bu nedenle, transfer öğrenme uygulamalarında da sıklıkla tercih edilmektedir.

4.6 Performans Metrikleri

Sınıflandırma sonuçlarında, doğruluk (accuracy), sınıflandırıcının doğruluğunu ölçmek için en yaygın kullanılan değerlendirme göstergelerinden biridir. Doğruluk, doğru sınıflandırılmış örneklerin toplam örneklerle oranını ifade eder. Bu çalışmada, doğruluğu bir ölçüm göstergesi olarak kullandık; ancak eğitim süreci boyunca, doğru pozitifler (TP – True Positive), doğru negatifler (TN – True Negative), yanlış pozitifler (FP – False Positive) ve yanlış negatifler (FN – False Negative) dahil olmak üzere dört sonuca dayalı olarak hesaplanan F1 skoru, kesinlik (precision) ve duyarlılık (recall) değerleri elde edildi.

$$\text{Doğruluk (Accuracy)} = \frac{TP}{(TN+FN+TP+FP)} \quad (4.1)$$

$$\text{Kesinlik (Precision)} = \frac{TP}{(TP+FP)} \quad (4.2)$$

$$\text{Duyarlılık (Recall)} = \frac{TP}{(TP+FN)} \quad (4.3)$$

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4.4)$$

5. DENEYSEL SONUÇLAR VE TARTIŞMA

Bu bölüm, derin öğrenme modellerine ait eğitim ve test süreçlerinden elde edilen deneysel sonuçların detaylı analizini içermekte; bu sonuçlar, gerçek zamanlı ve çevrimdışı değerlendirmeler ışığında önceki çalışmalarla karşılaştırılmaktadır.

5.1 Anahtar Nokta Koordinatlarını Ön İşleme

Veri setindeki numpy dizilerinde tutulan anahtar nokta koordinatları, modellerin eğitiminde kullanılmadan önce belirli ön işlemlerden geçirilmiştir. Öncelikle kelimeler etiketlenmiş daha sonra da numpy dizileri yüklenip, normalizasyon işlemi yapılmıştır. Bu işlemlerin ayrıntıları aşağıda sunulmuştur.

5.1.1 Kelimeleri Etiketleme

Her videoda veya çerçeve dizisinde gerçekleştirilen belirli eylemi belirtmek için etiketler kullanılır. Her kelimeye bir etiket atamak, kelimeleri görsel temsillerine göre sınıflandırmak için makine öğrenimi modellerini eğitmemize olanak tanır. Kelimeler listesindeki her etikete, sıfırdan başlayarak artan sırayla benzersiz bir numara atanmış ve bu eşlemeden oluşan yapı `label_map` adlı bir liste olarak oluşturulmuştur.

```
actions=np.array(['Gulegule','Gunaydın','Hosbulduk','Hosgeldiniz',  
'Merhaba','Ozurdilerim','Tesekkurederim','Neden','Yapmak','Kim',])  
label_map = {label:num for num, label in enumerate(actions)}
```

Bu listeyi kullanarak, etiket dizilerini Şekil 5.1'de gösterilen karşılık gelen sayısal değerlerine kolayca dönüştürebiliriz. Örneğin, Merhaba kelimesinin sayısal değeri 4'tür.

```
{'Gulegule': 0, 'Gunaydın': 1, 'Hosbulduk': 2, 'Hosgeldiniz': 3,  
'Merhaba': 4, 'Ozurdilerim': 5, 'Tesekkurederim': 6, 'Neden': 7,  
'Yapmak': 8, 'Kim': 9}
```

Şekil 5.1 Kelimelerin Sayısal Karşılığı

5.1.2 Dizileri Yükleme

Veri seti oluşturulduktan sonra bunların derin öğrenme modellerinde kullanılabilmesi için yüklenmesi gerekir. Veri setinde 446 kelime klasörü içerisinde bulunan 100 örnek video için 30 kare bulunmaktadır. Toplam örnek sayısı 44600'tür. Her video için 30 çerçeve vardır ve her çerçeve içinde 1662 anahtar nokta bulunmaktadır. Bunlar birleştirilerek 44600,30,1662 boyutlu bir dizi oluşturulmaktadır.

Veri setinde bulunan veri noktalarını belirli bir aralık veya ölçeğe göre yeniden düzenleme işlemine normalizasyon denir. Bu, genellikle makine öğrenimi ve veri analizi uygulamalarında verilerin daha etkili bir şekilde işlenmesini ve modellerin daha iyi performans göstermesini sağlamak için yapılır. Normalizasyonun temel amacı, farklı özelliklerin (features) birbiri üzerindeki etkisini dengelenmektir.

Veri setindeki değerleri normalize etmek için scikit-learn kütüphanesinde bulunan "StandardScaler" sınıfı kullanılmıştır. Bu sınıf ile her bir özelliğin ortalama değerini 0 ve standart sapmasını 1 olacak şekilde ayarlar. Bu işlem Z-skoru normalizasyonu olarak adlandırılır. Bu sayede verilerin özellikler arasındaki ölçek farklılıklarından etkilenmeden işlenmesi sağlanır. Aşağıda hem test hem de eğitim veri setinin normalizasyon kodları gösterilmektedir.

```
from sklearn.preprocessing import StandardScaler
scalers = {}
for i in range(X_train.shape[2]): #eğitim seti
    scalers[i] = StandardScaler()
    X_train[:, :, i] = scalers[i].fit_transform(X_train[:, :, i])
for i in range(X_test.shape[2]): #test seti
    X_test[:, :, i] = scalers[i].transform(X_test[:, :, i])
```

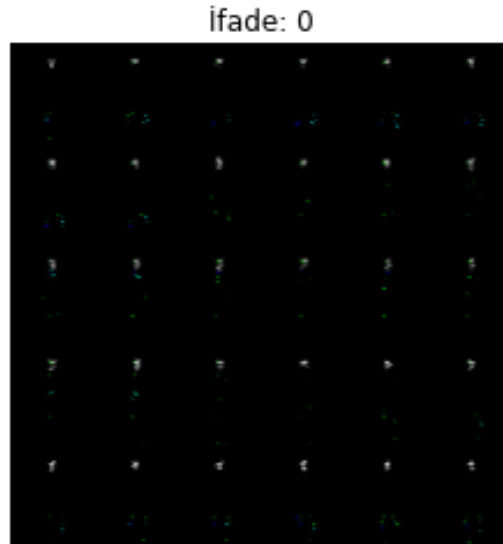
Normalizasyon işlemi sonrasında modellerin herhangi bir özelliğin ölçeğine karşı önyargılı olmamasını ve her bir özelliğin eğitim sürecine eşit katkıda bulunması sağlanmış olur. Normalizasyon, özellikle farklı birimlere veya ölçeklere sahip özelliklerle çalışıldığında önemlidir, ki bu da birçok makine öğrenimi görevinde

yaygındır. Özellikle sinir ağıları gibi girdi verilerinin ölçeğine hassas olan modellere veri beslemeden önce verilerin normalizasyonu modellerin başarısını arttırmaktadır.

5.2 İskelet Görüntüsü Dosyalarını Ön İşleme

Daha önce de belirtildiği üzere veri setinde MediaPipe tarafından verilen özelliklerin yanı sıra iskelet görüntüleri de elde edilmiştir. İşaret dili tanıma sistemlerinde, yalnızca tekil karelerin kullanılması yerine hareketin bütünsel yapısını koruyacak şekilde ardışık görsellerin birleştirilmesi, derin öğrenme modellerinin zamansal örüntüleri daha etkin biçimde öğrenmesine olanak tanımaktadır. Bu amaçla, her biri 30 kareden oluşan iskelet temsili görüntüler; ızgara, optik akış ve yatay birleştirme olarak üç farklı yöntemle görsel olarak birleştirilmiştir.

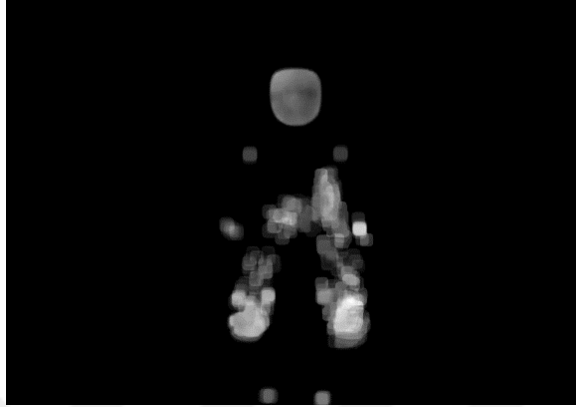
Izgara birleştirme yönteminde, 30 iskelet görüntüsü 5 satır ve 6 sütundan oluşan bir yapı içerisinde sıralanarak tek bir karede birleştirilmiştir. Bu yaklaşım hem zamansal geçişleri hem de mekânsal yapıyı koruyarak modelin hareket desenlerini daha kompakt bir biçimde gözlemlemesini sağlamaktadır. Şekil 5.2’de ızgara şeklinde birleştirilen örnek bir kelime gösterilmektedir.



Şekil 5.2 Izgara Birleştirme

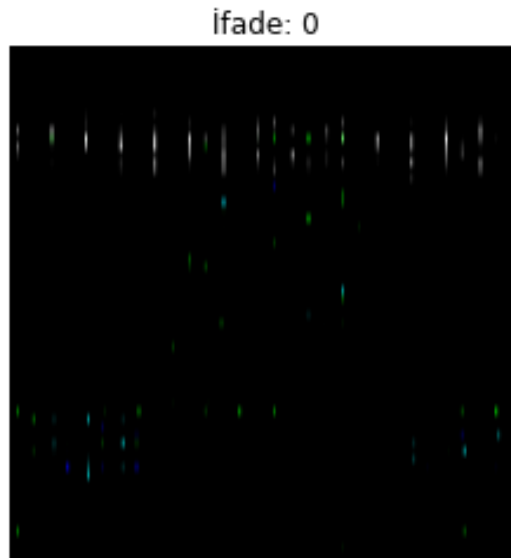
Optik akış temsili (Optical Flow Representation Fusion), kareler arası hareketin yoğunluğunu ve yönünü göstermek amacıyla uygulanmıştır. Farneback algoritması ile

elde edilen hareket vektörlerinin büyüklük bileşenleri (magnitude), yoğunluğa dayalı bir harita halinde görselleştirilmiş ve işaretin yapıldığı bölgelerdeki hareketin daha net analiz edilebilmesi sağlanmıştır. Şekil 5.3’de optik akış temsili şeklinde birleştirilen örnek bir kelime gösterilmektedir.



Şekil 5.3 Optik Akış Temsili Birleştirme

Yatay birleştirme yönteminde ise, 30 ardışık kare yatay ekseninde sıralanarak geniş bir görüntü oluşturulmuş, ardından evrişimli sinir ağı giriş boyutlarına uygun olarak yeniden ölçeklendirilmiştir. Bu yöntem, hareketin zaman içindeki akışını doğrudan yatay düzlemde temsil ederek modelin ardışıklığı daha açık biçimde öğrenmesini desteklemektedir. Şekil 5.4’te yan yana birleştirilen örnek bir kelime gösterilmektedir.



Şekil 5.4 Yatay Birleştirme

Izgara düzenleme, optik akış temsili ve yatay birleştirme teknikleri kullanılarak, her kelimeye ait 30 kareden oluşan tüm görseller tek bir görüntüde birleştirilmiştir. Bu işlemler sonucunda, 446 farklı kelimenin her biri için her yönteme özel olarak 100 görsel oluşturulmuştur. Böylece, her bir birleştirme yöntemi için toplam 44.600 örnekten oluşan üç ayrı görsel veri kümesi elde edilmiştir. Bu veri setleri, derin öğrenme tabanlı sınıflandırma modellerinin eğitimi ve karşılaştırmalı değerlendirmesi için kullanılmıştır.

5.3 Modellerin Eğitilmesi ve Test Edilmesi

5.3.1 Hiper Parametreler

LSTM, BiLSTM, CNN, GRU, CNN+LSTM, CNN+BiLSTM, CNN+GRU, GRU+LSTM ve GRU+BiLSTM modelleri, hold-out doğrulama yöntemi kullanılarak eğitilmiş ve test edilmiştir.

Modellerin eğitiminde kullanılan ve en yüksek sınıflandırma performansını sağlayan bazı parametreler Tablo 5.1'de sunulmuştur; bunun dışındaki parametreler varsayılan değerlerle kullanılmıştır. Optimizasyon, milyonlarca hatta milyarlarca parametrenin eğitim sürecinde nasıl ayarlanacağını belirleyen öğrenme algoritmasını ifade eder (Mahardika vd., 2023). Modellerin eğitiminde Adam, AdamW, Adadelata, SGD, RMSprop, Adamax ve Nadam gibi çeşitli optimizasyon algoritmaları test edilmiştir. Derin öğrenme modellerinin eğitiminde kullanılan optimizasyon algoritmaları, model parametrelerini düzenleyerek kayıp fonksiyonunu minimize etmeyi amaçlar. Bu bağlamda, Adam (Adaptive Moment Estimation), Adadelata, SGD (Stochastic Gradient Descent), RMSprop, Adamax ve Nadam gibi çeşitli optimizasyon algoritmaları sıklıkla tercih edilmektedir. Adam algoritması, momentum tabanlı optimizasyon ile adaptif öğrenme oranı yöntemlerini birleştirerek hızlı ve kararlı bir yakınsama sağlar ve genellikle varsayılan hiperparametrelerle bile etkili sonuçlar verir (Reyad vd., 2023). Adadelata, Adagrad algoritmasının bir türevi olup öğrenme oranının zamanla azalmasını engelleyerek daha kararlı bir eğitim süreci sunar. SGD, klasik bir optimizasyon yöntemi olarak, veri setinden alınan küçük mini-batch'lerle çalışarak basit ve etkili bir güncelleme sağlar; ancak yakınsama süresi daha uzun olabilir.

RMSprop algoritması, adaptif öğrenme oranlarını kullanarak gradyanların hareketli ortalamalarına dayalı güncellemeler yapar ve özellikle sıralı veriler üzerinde çalışan modellerde başarılıdır. Adamax, Adam algoritmasının sonsuz norm (infinity norm) kullanılarak daha stabil hale getirilmiş bir versiyonudur ve özellikle büyük veri setlerinde etkili sonuçlar verir. Nadam ise, Adam algoritmasını Nesterov momentum ile birleştirerek daha hassas güncellemeler ve hızlı yakınsama sağlamayı hedefler (Mandasari vd., 2023). Bu süreçte, LSTM modeli için en yüksek sınıflandırma doğruluğu Adadelat optimizasyon algoritması ile elde edilirken, diğer tüm modellerde Adam optimizasyon algoritması ile en iyi sonuçlar elde edilmiştir.

Eğitim sürecinde kritik bir diğer faktör ise öğrenme oranıdır. Öğrenme oranının çok yüksek olması durumunda modelin tahminleri kararsız hale gelebilir ve dalgalanabilirken, çok düşük bir öğrenme oranı eğitim süresini ciddi ölçüde uzatabilir. Bu nedenle, eğitim aşamasında çeşitli öğrenme oranları test edilmiş ve en iyi performansı sağlayan oran seçilmiştir. Batch size, model eğitimi sırasında, veri setinin küçük gruplar halinde işlenmesini sağlayan bir parametredir. Eğitim sürecinde veri seti, modelin işlem gücüne ve bellek kapasitesine uygun şekilde parçalara (mini-batches) ayrılır ve batch size, her bir adımda modelin bu mini-batch'lerden kaç örnek üzerinde işlem yapacağını belirler (Paharia vd., 2023). Bu işlem, modelin kayıp fonksiyonunu minimize etmek için ağırlıklarını optimize ederken gradyan hesaplamalarını etkiler. Bunun yanı sıra, tüm modellerde erken durdurma (early stopping) stratejisi uygulanmıştır. Bu durum, modellerin eğitiminde kullanılan epoch sayılarının birbirinden farklı olmasına neden olmuştur.

Tablo 5.1 Modellerin hiper parametre ayarları

Model	Optimizasyon Algoritması	Öğrenme Oranı	Küme Boyutu	Epoch	Activation Func.	Loss Func.
LSTM	Adamax	0,0001		172		
GRU	Adamax	0,0001		231		
BiLSTM	Adamax	0,00001		237		
CNN	Adamax	0,0001		128		
CNN+LSTM	Adadelat	0,0001	32	214	Relu Softmax	Categorical Crossentropy
CNN+BiLSTM	Adam	0,00001		161		
CNN+GRU	Adamax	0,0001		186		
GRU+LSTM	Adam	0.0001		71		
GRU+BiLSTM	Adamax	0.001		90		

5.3.2 Veri Setinin Bölünmesi

Veri setinin eğitim ve test alt kümelerine ayrılmasında, scikit-learn kütüphanesinin model_selection modülünde yer alan train_test_split fonksiyonu kullanılmıştır. Bu işlem sırasında verilerin %80'i modelin eğitimi için, kalan %20'si ise genel performansın bağımsız olarak değerlendirilmesi amacıyla test kümesi olarak ayrılmıştır. Eğitim verisinin bir bölümü de doğrulama (validation) amacıyla %12.5 oranında ayrılarak, eğitim sürecinde modelin aşırı öğrenme eğilimi izlenmiştir. Bu kapsamda veri seti yaklaşık olarak %70 eğitim, %10 doğrulama ve %20 test olacak şekilde üç alt kümeye bölünmüştür.

Modelin nihai performansını gerçek zamanlı koşullarda test edebilmek amacıyla, test süreci sonrasında en yüksek sınıflandırma başarımına ulaşan model seçilmiştir. Gerçek zamanlı performans değerlendirmesi amacıyla, biri veri setinde yer alan, diğeri ise veri setinde bulunmayan iki katılımcı ile deneysel bir test senaryosu gerçekleştirilmiştir. Gerçek zamanlı testlerde, Türk İşaret Dili (TİD) işaretleri, kamera karşısında 10 kez tekrarlanarak gerçekleştirilmiş; elde edilen her örnek doğrudan eğitilmiş modele girdi olarak verilmiş ve modelin çıktıları kayıt altına alınmıştır. Bu sayede, sistemin pratikteki doğruluğu güvenilir biçimde ölçülmüş ve elde edilen doğruluk değeri, modelin sahadaki başarımını yansıtır nitelikte olmuştur.

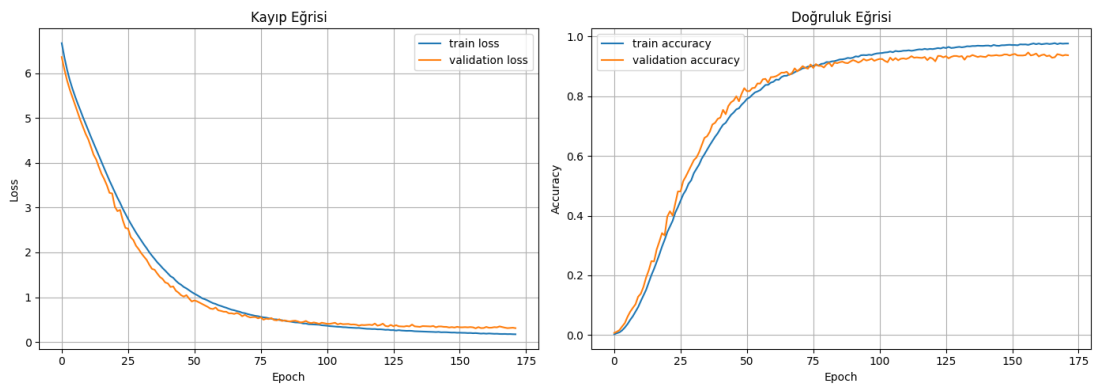
5.4 Dışarda Tutma Doğrulama ile Elde Edilen Sınıflandırma Sonuçları

Makine öğrenimi modellerinin başarı düzeylerini değerlendirmek amacıyla kullanılan temel yöntemlerden biri olan dışarda tutma doğrulama (hold-out validation) yöntemi, bu çalışmada da uygulanmıştır. Her modelin eğitim sürecinde erken durdurma (early stopping) ve öğrenme oranı azaltma (ReduceLROnPlateau) gibi geri çağırma (callback) mekanizmaları kullanılarak, modelin doğrulama verisi üzerindeki başarımı izlenmiş ve gereksiz yere uzun süren, aşırı öğrenmeye neden olabilecek eğitim süreçlerinin önüne geçilmiştir. Böylece modellerin aşırı uyum (overfitting) yapmadan daha iyi genellenebilirlik göstermesi amaçlanmıştır.

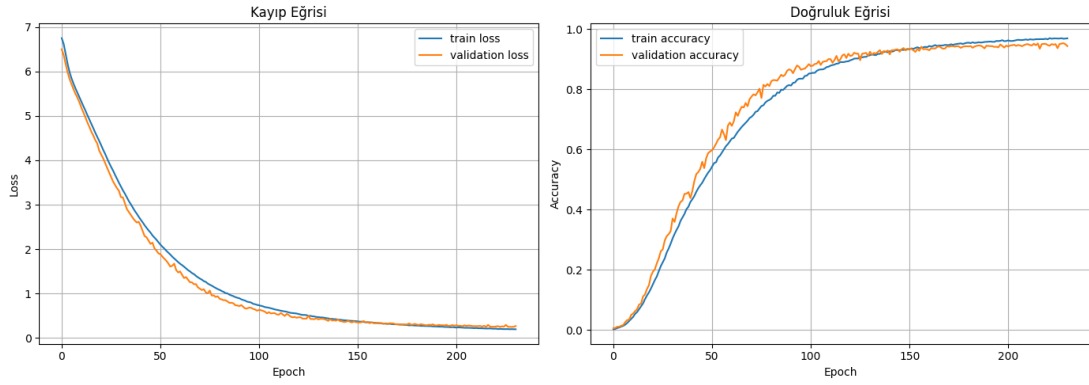
Modellerin eğitimleri sonunda Kayıp ve Doğruluk Eğrisi grafikleri oluşturulmuştur. Derin öğrenme modellerinin eğitim sürecinde performans takibi için kullanılan en

temel ölçütler kayıp ve doğruluk değerleridir. Bu ölçütlerin zamana bağlı değişimini gösteren grafikler, modelin öğrenme eğilimini anlamak açısından büyük önem taşır. Kayıp eğrisi, modelin tahminleri ile gerçek değerler arasındaki farkı temsil eder. Eğitim sırasında bu değerlerin düzenli şekilde azalması, modelin öğrenme sürecinin sağlıklı ilerlediğini gösterir. Eğriler genellikle eğitim ve doğrulama verileri için ayrı ayrı çizilir. Eğitimin devam ederken doğrulama kaybının artması, modelin ezberlemeye yöneldiğine ve genelleme yeteneğinin zayıfladığına işaret eder. Doğruluk eğrisi ise modelin doğru sınıflandırma oranını ifade eder. Bu eğrinin yükselmesi, modelin veriyle daha doğru şekilde örtüştüğünü gösterir. Eğitim ve doğrulama doğruluğunun benzer düzeylerde seyretmesi, modelin hem öğrenme hem de genelleme açısından dengeli çalıştığını ortaya koyar (Terven vd., 2025).

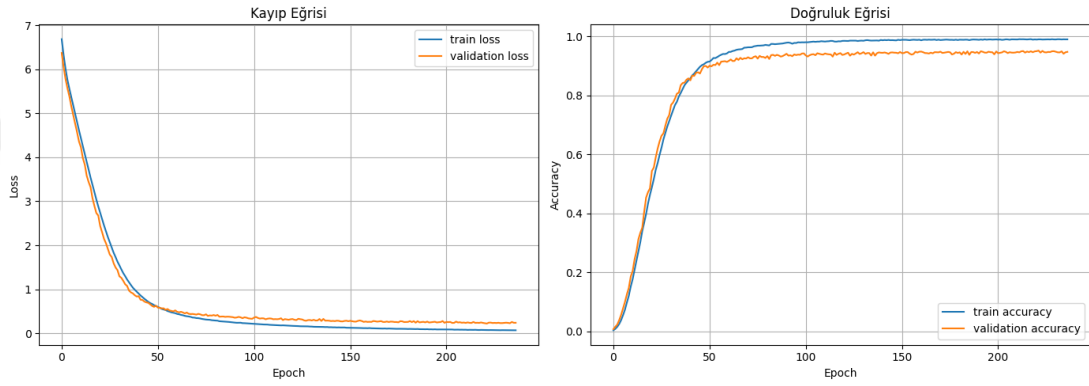
Şekil 5.5, Şekil 5.6 ve Şekil 5.7; LSTM, GRU ve BiLSTM modellerinin eğitim ve doğrulama eğrilerini göstermektedir. Bu eğriler incelendiğinde benzer bir performans sergilediği görülmektedir. Bu modellerde eğitim süresi boyunca kayıp değerleri düzenli şekilde azalırken, doğruluk oranları da istikrarlı bir şekilde artış göstermiştir. LSTM modeli 100 epoch, GRU modeli 150 epoch ve BiLSTM modeli 50 epoch sonrasında doğrulama doğruluğu açısından durağanlaşma eğilimine girmiştir. LSTM, GRU ve BiLSTM modellerinin doğrulama doğruluklarının sırasıyla %93,70, %94,28 ve %94,73 olduğu tespit edilmiştir.



Şekil 5.5 LSTM Modeli Kayıp ve Doğruluk Eğrisi



Şekil 5.6 GRU Modeli Kayıp ve Doğruluk Eğrisi

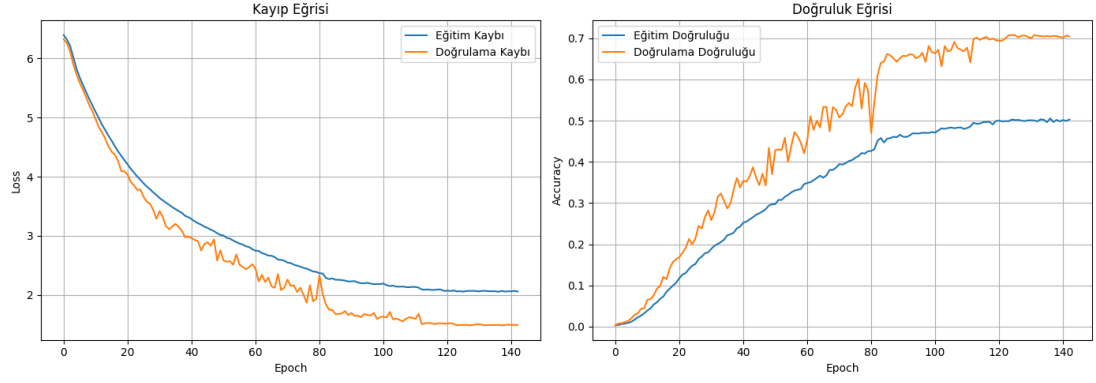


Şekil 5.7 BiLSTM Modeli Kayıp ve Doğruluk Eğrisi

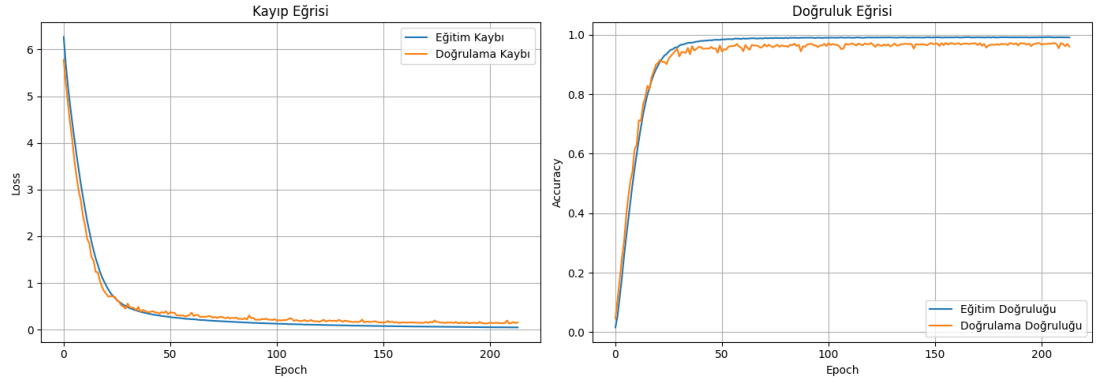
Şekil 5.8, Şekil 5.9, Şekil 5.10 ve Şekil 5.11 ise CNN tabanlı modellerin (CNN, CNN+LSTM, CNN+BiLSTM ve CNN+GRU) eğitim sürecindeki performansını göstermektedir. Özellikle Şekil 5.8’de yer alan CNN modelinde eğitim ve doğrulama kayıplarının düzenli şekilde azaldığı ve doğrulama kaybının eğitim kaybından düşük seyrettiği gözlemlenmektedir. Doğruluk eğrileri ise doğrulama doğruluğunun %70 seviyelerine ulaşırken, eğitim doğruluğunun %50 civarında kaldığını ortaya koymaktadır. Bu durum, modelin doğrulama verisinde daha başarılı olduğunu ancak eğitim verisinde yeterince genelleme yapamadığını düşündürmektedir.

Hibrit yapılar olan CNN+LSTM, CNN+BiLSTM ve CNN+GRU modelleri (Şekil 5.9–5.11), klasik RNN temelli modellere göre daha hızlı öğrenme eğrileri sergilemiştir. CNN+LSTM, CNN+BiLSTM ve CNN+GRU modellerinin doğrulama doğruluklarının sırasıyla %96,01; %97,35 ve %96,88 olduğu tespit edilmiştir. Bu modellerde hem eğitim hem doğrulama kayıplarının hızlı şekilde düştüğü ve doğruluk eğrilerinin yaklaşık %96 seviyesinden yüksek olduğu görülmektedir. Özellikle

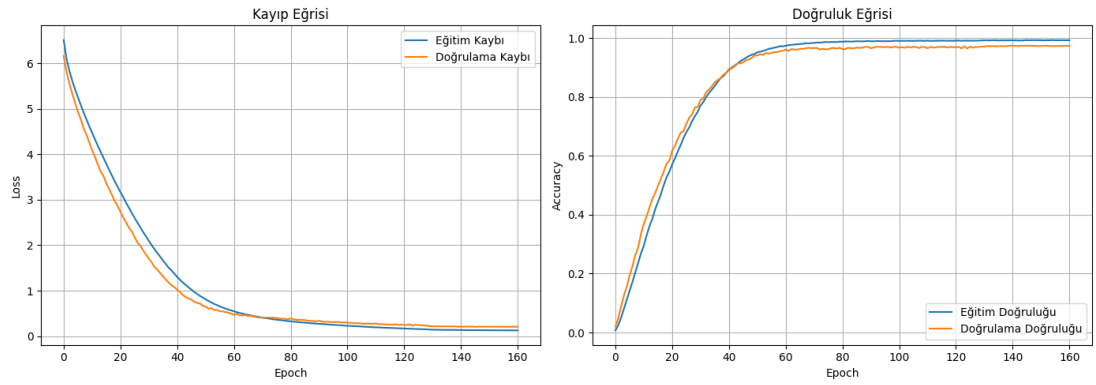
CNN+GRU modeli, en düşük kayıp ve en yüksek doğruluk değeriyle genel performans açısından dikkat çekici bir başarı göstermiştir.



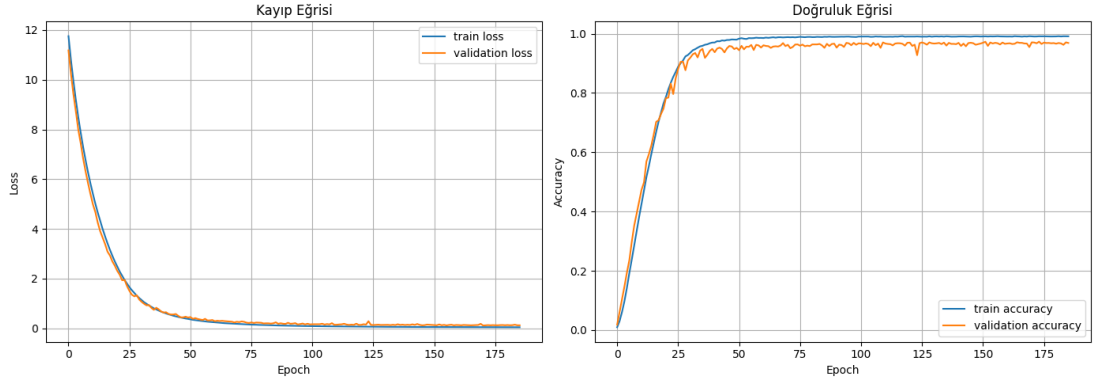
Şekil 5.8 CNN Modeli Kayıp ve Doğruluk Eğrisi



Şekil 5.9 CNN+LSTM Modeli Kayıp ve Doğruluk Eğrisi

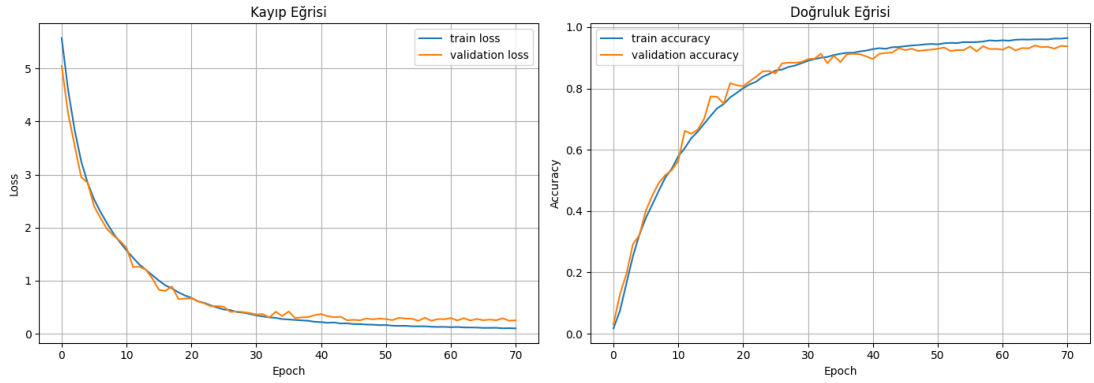


Şekil 5.10 CNN+BiLSTM Modeli Kayıp ve Doğruluk Eğrisi

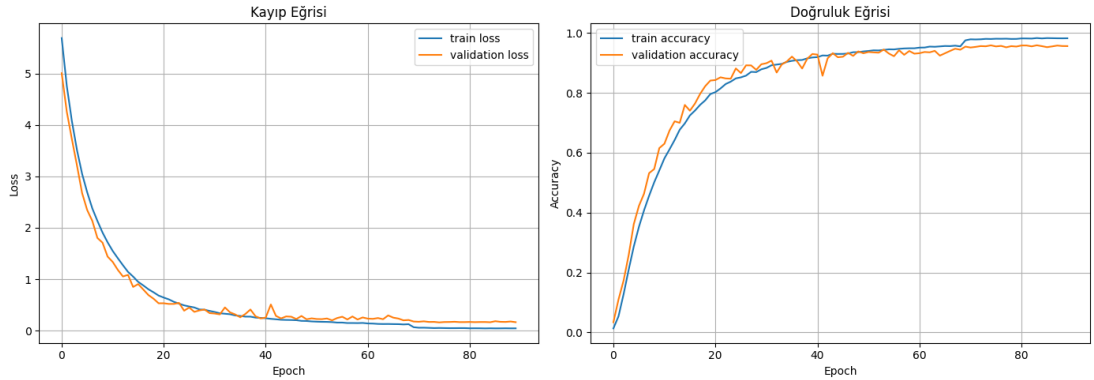


Şekil 5.11 CNN+GRU Modeli Kayıp ve Doğruluk Eğrisi

Son olarak, Şekil 5.12 ve Şekil 5.13'te yer alan GRU+LSTM ve GRU+BiLSTM hibrit modellerinin, eğitim süreci boyunca istikrarlı şekilde azalan kayıp eğrileriyle birlikte yüksek doğruluk oranlarını korudukları görülmektedir. Bu modellerin eğitim sürecinde aşırı öğrenme göstermemesi ve doğrulama performansının sabit kalması, öğrenme sürecinin kararlılığını ortaya koymaktadır. GRU+LSTM ve GRU+BiLSTM modellerinin doğrulama doğruluklarının sırasıyla %93,77 ve %95,5 olduğu tespit edilmiştir.



Şekil 5.12 GRU+LSTM Modeli Kayıp ve Doğruluk Eğrisi



Şekil 5.13 GRU+BiLSTM Modeli Kayıp ve Doğruluk Eğrisi

Tablo 5.2’de, dışarıda tutma doğrulama yöntemiyle eğitilen modellerin elde ettiği performans değerleri karşılaştırmalı olarak sunulmaktadır. Model çıktıları, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-macro skoru gibi metriklerle değerlendirilmiş; sınıflandırma performansları kapsamlı şekilde analiz edilmiştir. Sonuçlara göre, yalnızca temel CNN modeli dışında kalan tüm mimariler yüksek sınıflandırma performansı elde etmiştir. En yüksek test doğruluğu, CNN+GRU modelinde elde edilmiştir (%97,72). Bu model aynı zamanda %97,88 kesinlik, %97,72 duyarlılık ve %97,71 F1-Macro skoruna sahiptir. Bu sonuçlar, CNN tabanlı uzamsal özellik çıkarımı ile GRU’nun zamansal bağımlılıkları etkin bir şekilde modelleme yeteneğinin birleşiminin yüksek bir performans sağladığını ortaya koymaktadır. Eğitim süresi diğer modellere kıyasla kısa olup (1148,07 sn), verimlilik açısından da avantajlıdır.

İkinci sırada yer alan CNN+BiLSTM modeli %97,57 test doğruluğu ve %97,55 F1-Macro skoruyla dikkat çekmektedir. Bu model, çift yönlü LSTM yapısı sayesinde geçmiş ve gelecek bilgiye aynı anda erişerek sınıflandırma başarımını artırmıştır. Eğitim süresi ise CNN+GRU’ya göre daha yüksektir. CNN+LSTM modeli de benzer şekilde %97,51 test doğruluğu ve %97,50 F1-Macro skoru sunarak üst düzey bir performans sergilemiştir. CNN tabanlı modellerin, ardıl tekrarlayan katmanlarla birlikte kullanılması genelde yüksek doğruluk seviyelerine ulaşılmasını sağlamıştır. GRU+BiLSTM modeli %96,05 doğruluk ve %96,04 F1-Macro skoru ile önemli bir başarı göstermektedir. Eğitim süresi açısından ise oldukça avantajlıdır (1321,33 sn). Bu yapı, GRU’nun daha hızlı öğrenme kapasitesini BiLSTM’nin iki yönlü bağlam anlayışıyla birleştirmektedir. GRU ve BiLSTM modelleri sırasıyla %95,48 ve %95,26 doğruluk oranlarıyla hibrit modellere göre daha düşük performans göstermiştir. Yine de bu modellerin F1-Macro skorları %95’in üzerindedir ve dengeli sınıflandırma yaptıkları gözlemlenmektedir. LSTM modeli %93,99 doğruluk ile daha sınırlı bir performans sunmuş, özellikle test kaybının yüksek olması (%0,3216) bu modelin öğrenme sürecinde daha fazla dalgalanma yaşadığını göstermiştir. Tek başına kullanılan CNN modeli, %70,89 test doğruluğu ile en düşük sınıflandırma başarımına sahip model olmuştur. Bu durum, uzamsal özellikleri çıkarmada etkili olan CNN’in, zamansal örüntüleri tek başına yeterince yakalayamadığını göstermektedir.

Tablo 5.2 Modellerin sınıflandırma sonuçları

Model	En İyi Epoch	Test Doğruluğu (%)	Kesinlik (%)	Duyarlılık (%)	F1-Macro Skoru (%)	Test Kaybı	Eğitim Süresi (sn)
CNN+GRU	171	97,72	97,88	97,72	97,71	0,0929	1148,07
CNN+BiLSTM	146	97,57	97,66	97,57	97,55	0,1918	2685,37
CNN+LSTM	199	97,51	97,60	97,51	97,50	0,1225	2413,05
GRU+BiLSTM	75	96,05	96,24	96,05	96,04	0,1685	1321,33
GRU	216	95,48	95,73	95,48	95,45	0,2432	2145,08
BiLSTM	222	95,26	95,48	95,26	95,23	0,2319	3173,77
LSTM	157	93,99	94,29	93,99	93,97	0,3216	1599,36
GRU+LSTM	56	93,91	94,38	93,91	93,91	0,2464	879,98
CNN	128	70,89	74,30	70,89	70,18	1,4687	552,54

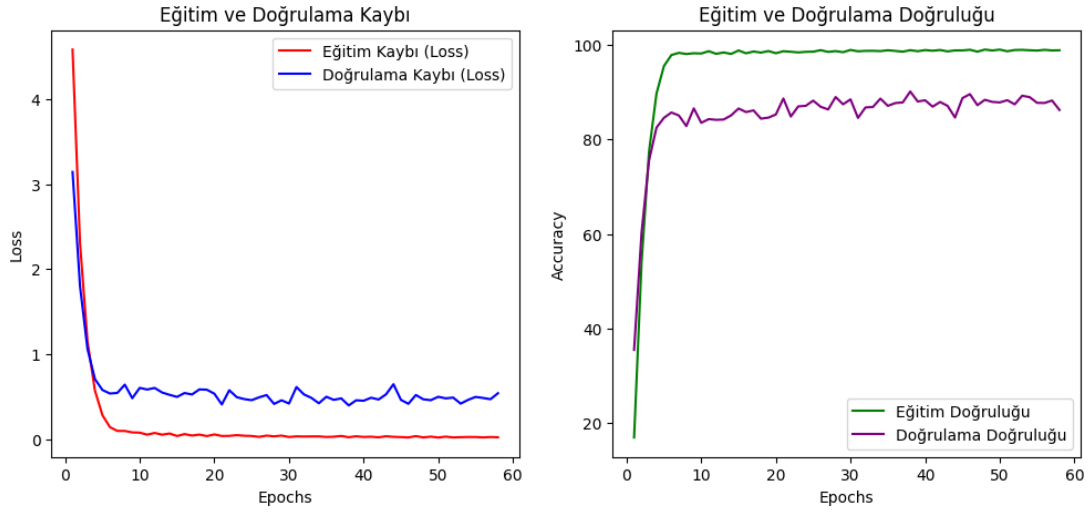
5.5 İskelet Görüntüleri Üzerinden Elde Edilen Doğrulama Sonuçları

Veri setinde bulunan iskelet görüntüleri üzerinde de deneysel çalışmalar yürütülmüştür. Model eğitimleri için veri seti önceki modellerin eğitimindekine benzer şekilde %70 eğitim, %10 doğrulama ve %20 test olarak ayrılmıştır. Bu bölme işlemi, sınıflar arası dengenin korunmasına dikkat edilerek rastgele ancak kontrollü bir şekilde gerçekleştirilmiştir. Eğitim kümesi, modelin parametrelerini öğrenmesi amacıyla; doğrulama kümesi, hiperparametre ayarlarının yönlendirilmesi ve modelin aşırı öğrenme eğiliminin izlenmesi için; test kümesi ise modelin daha önce hiç görmediği veriler üzerindeki genelleme performansını değerlendirmek amacıyla kullanılmıştır.

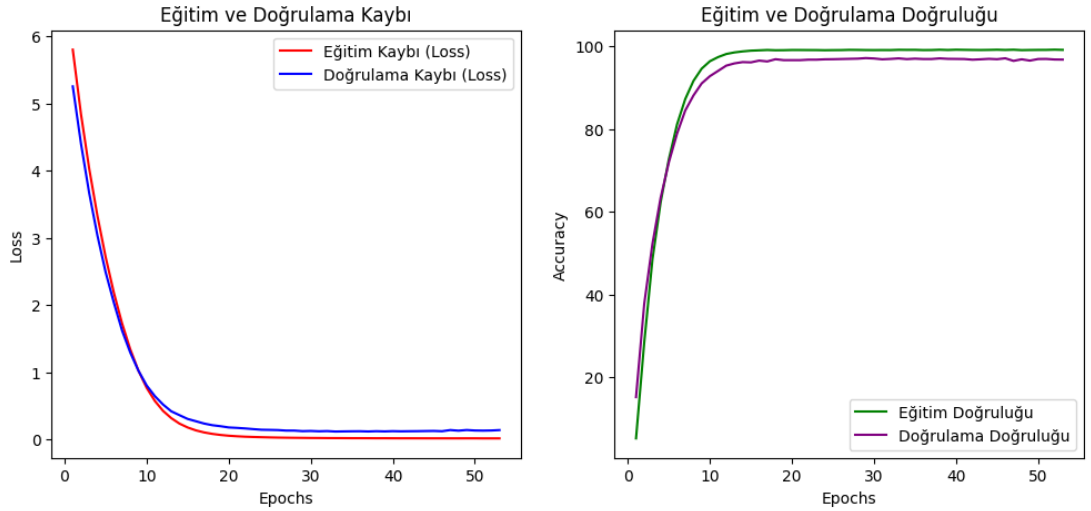
Farklı birleştirme yöntemleriyle oluşturulan görsel temsiller üzerinden eğitilen ResNet-18 ve CBAM+ResNet-18 modellerinin eğitim süreçleri, eğitim ve doğrulama doğruluğu ile kayıp eğrileri aracılığıyla ayrıntılı olarak incelenmiştir. Grafikler, modellerin öğrenme davranışlarını karşılaştırmalı olarak değerlendirme olanağı sunmakta ve temsil yöntemlerinin model başarısı üzerindeki etkisini ortaya koymaktadır.

ResNet-18 modeli için ızgara birleştirme yöntemiyle elde edilen sonuçlar (Şekil 5.14), eğitim ve doğrulama doğruluğunun yaklaşık 20. epoch'ta dengelendiğini ve kayıpların zamanla istikrarlı biçimde azaldığını göstermektedir. Optik akış temsiline ait grafikler (Şekil 5.15), hem eğitim hem de doğrulama doğruluklarının erken epoch'larda yüksek

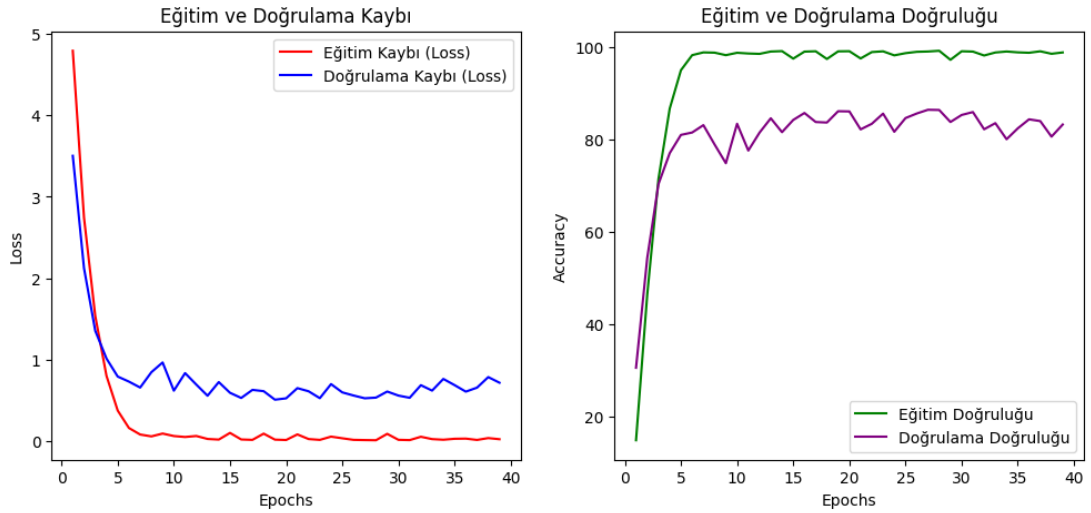
değerlere ulaştığını ve doğrulama kaybının düşük seviyelerde sabitlendiğini göstermektedir. Bu durum, modelin zamansal hareket bilgisini içeren bu temsilden etkin şekilde yararlandığını ortaya koymaktadır. Buna karşın, yan yana birleştirme yöntemi (Şekil 5.16) ile eğitilen modelin doğrulama doğruluğunda dalgalanmalar gözlenmiş; doğrulama kaybının zaman zaman arttığı izlenmiştir. Bu bulgular, modelin bu temsilde genelleme kabiliyetinin sınırlı olduğunu ve bazı sınıflarda düşük performans sergilemiş olabileceğini düşündürmektedir.



Şekil 5.14 Izgara Birleştirme Yöntemi / ResNet-18 Modeli



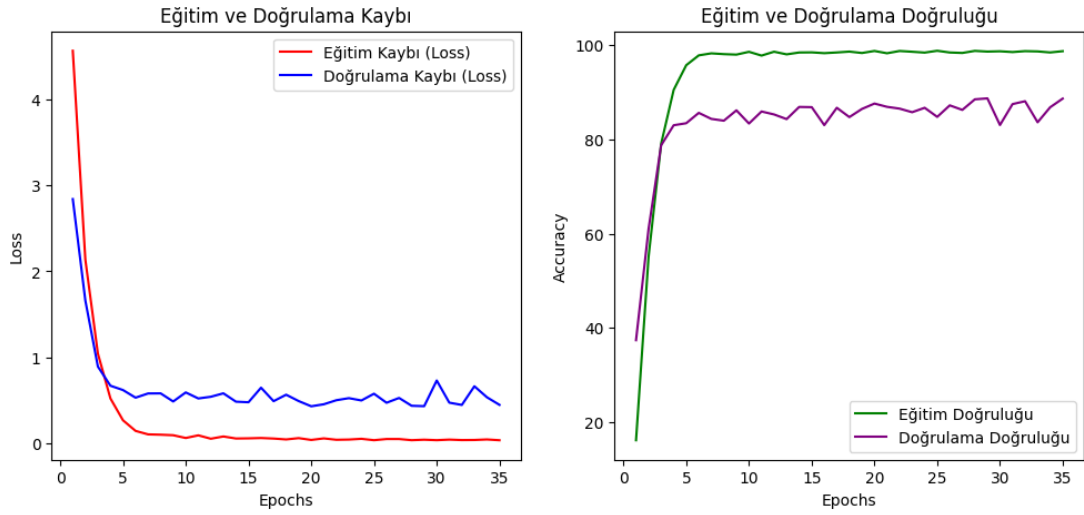
Şekil 5.15 Optik Akış Birleştirme Yöntemi / ResNet-18 Modeli



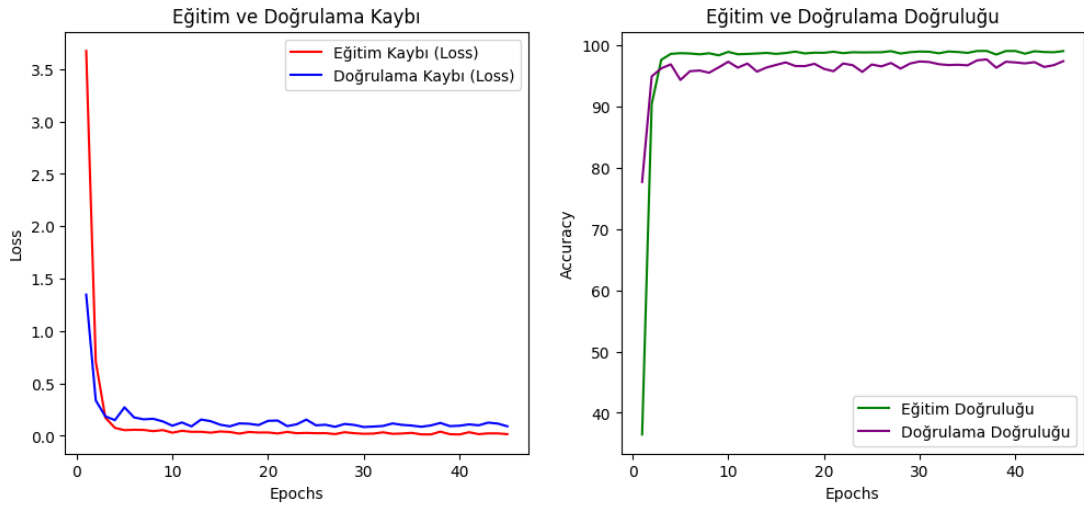
Şekil 5.16 Yan Yana Birleştirme Yöntemi / ResNet-18 Modeli

CBAM modülünün entegre edildiği ResNet-18 modellerinde ise eğitim süreçlerinin daha kararlı ve hızlı seyrettiği gözlemlenmiştir. CBAM+ResNet18 modeline ait ızgara temsilli öğrenme eğrileri (Şekil 5.17), doğruluk seviyelerinin daha erken epoch’larda yüksek değerlere ulaştığını ve doğrulama eğrisinin eğitim eğrisiyle yüksek uyum gösterdiğini ortaya koymaktadır. Optik akış temsili ile desteklenen CBAM+ResNet18 modeli (Şekil 5.18), doğrulama doğruluğunun yüksekliği ve kaybın minimum düzeyde seyretmesi bakımından en dengeli ve başarılı yapı olarak öne çıkmaktadır. Bu sonuçlar, CBAM modülünün özellikle hareket bilgisinin öne çıkarıldığı temsillerde anlamlı katkılar sunduğunu göstermektedir. Yan yana birleştirme yöntemiyle oluşturulan veriyle eğitilen CBAM+ResNet18 modeli (Şekil 5.19), yüksek doğruluk değerlerine ulaşmakla birlikte doğrulama doğruluğunda düzensiz dalgalanmalar sergilemiştir. Bu durum, temsilin sınırlı yapısal bilgisi nedeniyle modelin dikkat mekanizmasının potansiyelinden tam anlamıyla faydalanamadığını düşündürmektedir.

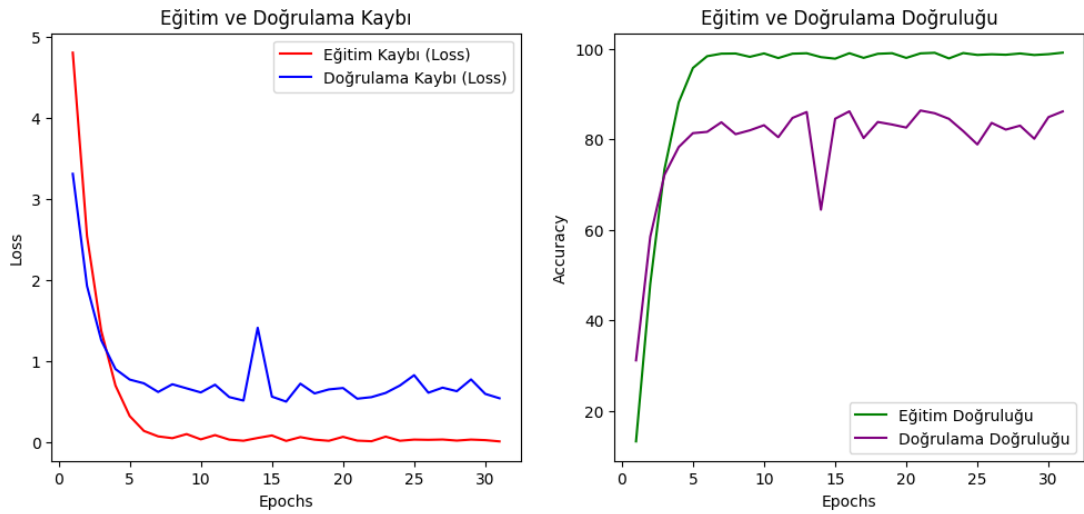
Görsel temsillerin yapısı ve içeriği, derin öğrenme modellerinin öğrenme eğrileri ve genel doğruluk düzeyleri üzerinde belirleyici bir rol oynamaktadır. Özellikle optik akış temsili hem temel hem de dikkat destekli mimarilerde daha dengeli, hızlı ve istikrarlı bir öğrenme süreci sunarak öne çıkmıştır. CBAM modülünün kullanımı ise doğrulama başarısının daha erken epochlarda dengelenmesini sağlamakta ve genel model kararlılığını güçlendirmektedir.



Şekil 5.17 Izgara Birleştirme Yöntemi / CBAM+ResNet-18 Modeli



Şekil 5.18 Optik Akış Birleştirme Yöntemi / CBAM+ResNet-18 Modeli



Şekil 5.19 Yan Yana Birleştirme Yöntemi / CBAM+ResNet-18 Modeli

Tablo 5.3 incelendiğinde, hem ResNet-18 hem de CBAM ile zenginleştirilmiş ResNet-18 (CBAM+ResNet18) mimarilerinin, TurkSign446 veri setindeki iskelet görüntüleri üzerinde yüksek sınıflandırma başarıları elde ettiği görülmektedir. Ancak bu başarı düzeyleri, kullanılan görsel temsil yöntemine bağlı olarak farklılık göstermektedir. Bu durum, veri ön işleme sürecinde tercih edilen temsillerin, modelin öğrenme sürecini ve çıktılarını doğrudan etkileyen önemli bir faktör olduğunu ortaya koymaktadır.

Tablo 5.3 Modellerin iskelet görüntüleri üzerinde sınıflandırma sonuçları

Model	Birleştirme	Test Doğruluğu (%)	Kesinlik (%)	Duyarlılık (%)	F1-Macro Skoru (%)	Test Kaybı
ResNet-18	Izgara	86,20	88,15	86,67	86,71	0,0260
ResNet-18	Optik Akış	96,77	96,78	96,64	96,65	0,0128
ResNet-18	Yan yana	83,26	83,81	82,04	82,19	0,0265
CBAM + ResNet18	Izgara	88,64	89,04	88,07	88,09	0,0333
CBAM + ResNet18	Optik Akış	97,44	97,49	97,30	97,26	0,0162
CBAM + ResNet18	Yan yana	86,18	85,99	85,12	85,08	0,0151

5.6 Tartışma

5.6.1 Dışarıda Tutma (Hold-Out) Doğrulama Yöntemi ile Doğruluk Değerlendirmesi

Çalışma kapsamında oluşturulan veri seti kullanılarak LSTM, Bi-LSTM, GRU, CNN, CNN+LSTM, CNN+BiLSTM, CNN+GRU, GRU+LSTM ve GRU+BiLSTM olmak üzere toplam dokuz farklı model eğitilmiştir. Bu modellerden CNN+GRU modeli %97,72 test doğruluğu ile en yüksek sınıflandırma performansını elde etmiştir. Literatürdeki diğer TİD veri setleri ve modelleriyle karşılaştırıldığında, bu çalışmada geliştirilen TurkSign446 veri seti ve önerilen tanıma yönteminin, sınıflandırma başarımları açısından daha yüksek doğruluk oranlarına ulaştığı görülmektedir. Literatürdeki çeşitli Türk İşaret Dili (TİD) veri setleri ile gerçekleştirilen çalışmalar, sınıflandırma doğruluğu bakımından farklı sonuçlar ortaya koymuştur. Tablo 5.4'te, yaygın kullanılan veri setleri (BSign22k, ERUSLR, AUTSL) ile elde edilen doğruluk

oranları, bu çalışmada geliştirilen TurkSign446 veri seti ile ulaşılan sonuçlarla karşılaştırmalı olarak sunulmuştur.

Gündüz ve Polat (2021), BSign22k Genel veri seti ile %89,35 doğruluk elde ederken, Gökçe vd. (2020) BSign22k üzerinde %94,94 doğruluk sağlamıştır. Akdağ (2024) çalışmasında ise aynı veri seti için %96,94'e kadar çıkan sonuçlar raporlanmıştır. Bunun yanı sıra, Özcan vd. (2021) ERUSLR veri seti ile %93,93 doğruluk bildirirken, Mercanoğlu vd. (2020) AUTSL veri setinde %95,95 doğruluğa ulaşmıştır.

Bu çalışma kapsamında oluşturulan TurkSign446 veri seti ve CNN+GRU tabanlı sınıflandırma modeli ile elde edilen %97,72'lik doğruluk değeri, mevcut literatürde raporlanan pek çok çalışmanın üzerinde bir başarı göstermektedir. Tüm çalışmaların ortak noktası olarak "dışarıda tutma" (hold-out validation) yöntemi ile değerlendirme yapılmış olması, bu karşılaştırmayı anlamlı kılmaktadır. Sonuçlar, önerilen yaklaşımın TİD kelime tanıma alanında güçlü bir alternatif sunduğunu ortaya koymaktadır.

Tablo 5.4 Modellerin karşılaştırılması

Çalışmalar	Doğruluk (%)					Doğrulama Yöntemi
	BSign22k Genel	BSign22k	ERUSLR	AUTSL	TurkSign446	
Gündüz ve Polat (2021)	89,35	-	-	-	-	Dışarıda Tutma
Gökçe vd. (2020)	-	94,94	-	-	-	Dışarıda Tutma
Akdağ (2024)	96,94	95,13	-	-	-	Dışarıda Tutma
Özcan vd. (2021)	-	-	93,93	-	-	Dışarıda Tutma
Mercanoğlu vd. (2020)	-	-	-	95,95	-	Dışarıda Tutma
Bu çalışma	-	-	-	-	97,72	Dışarıda Tutma

5.6.2 İskelet Görüntüleri Üzerinden Doğruluk Değerlendirmesi

ResNet-18 mimarisi özelinde, en yüksek doğruluk değeri %96,77 ile optik akış temsili kullanılarak elde edilmiştir. Bu yöntemde ayrıca %96,78 kesinlik, %96,64 duyarlılık ve %96,65 F1-Macro skoru hesaplanmıştır. Bu değerler, optik akış temsiliyle hareket bilgisinin etkin biçimde modellendiğini ve modelin sınıflar arasında daha net ayrım

yapabildiğini göstermektedir. Optik akış yöntemini, %86,20 doğruluk değeri ile ızgara temsili takip ederken, yatay birleştirme yöntemiyle %83,26 doğruluk elde edilmiştir. Ancak yan yana (yatay) birleştirme tekniğiyle hesaplanan %82,19'luk F1-Macro skoru, bu temsilde modelin bazı sınıflarda ayırım yapmada zorluk yaşadığını düşündürmektedir.

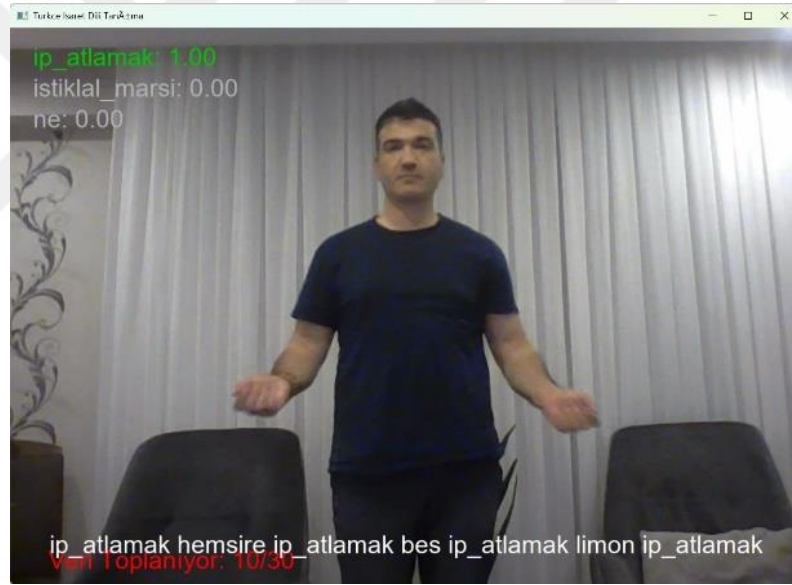
CBAM entegrasyonunun model performansına etkisi incelendiğinde, tüm temsil yöntemlerinde genel bir iyileşme sağladığı gözlemlenmektedir. CBAM+ResNet18 mimarisi ile en yüksek doğruluk %97,44 ile optik akış temsili kullanılarak elde edilmiş; bu temsilde ayrıca %97,49 kesinlik, %97,30 duyarlılık ve %97,26 F1-Macro skoru raporlanmıştır. Bu durum, CBAM modülünün dikkat mekanizması sayesinde hareketin yoğunlaştığı bölgeleri vurgulayarak sınıflandırma başarısını anlamlı şekilde artırdığını göstermektedir. Literatürdeki önceki çalışmalarla karşılaştırıldığında bu model de CNN+GRU modelinde olduğu gibi daha yüksek bir doğruluk değeri elde etmiştir. Öte yandan, yatay birleştirme yönteminde %86,18 doğruluk ve %85,08 F1-Macro skoru elde edilmiştir; bu temsil yöntemi, CBAM modülü eklenmiş olmasına rağmen modelin bazı sınıflarda kararsızlık yaşadığına işaret etmektedir. Izgara temsili ile elde edilen %88,64 doğruluk ve %88,09 F1-Macro skoru ise optik akış kadar olmasa da CBAM etkisinin diğer temsillere kıyasla daha dengeli bir iyileşme sağladığını göstermektedir.

Sonuç olarak, her iki model için de en başarılı sonuçlar optik akış temsili ile elde edilmiştir. CBAM modülü, özellikle F1-Macro skoru açısından anlamlı iyileştirmeler sunmuş ve modelin sınıflar arası denge kurma yeteneğini artırmıştır. Tüm temsillerde düşük test kaybı değerleri gözlemlenmiş olup bu durum, modellerin öğrenme sürecinde aşırı öğrenmeden kaçındığını ve yüksek genelleme kapasitesine sahip olduğunu göstermektedir. Bulgular hem görsel temsil seçiminin hem de dikkat modüllerinin işaret dili tanıma sistemlerinin başarımında kritik rol oynadığını açıkça ortaya koymaktadır.

5.6.3 Gerçek Zamanlı Değerlendirme

Gerçek zamanlı testlerde model doğruluk değeri en yüksek olan CNN+GRU mimarisi

kullanılmıştır. Gerçek zamanlı değerlendirme, aydınlık ve gölgelenmeyen bir ortamda, sade arka planlı bir alanda gerçekleştirilmiştir. Kamera ile katılımcı arasında ortalama 2 metre mesafe bırakılarak, görüntü işleme algoritmasının anahtar noktaları yüksek doğrulukla tespit edebilmesi sağlanmıştır. Gerçek zamanlı performans değerlendirmesi amacıyla, biri veri setinde yer alan, diğeri ise veri setinde bulunmayan iki katılımcı ile deneysel bir test senaryosu gerçekleştirilmiştir. Her bir kelime için toplam 10 tekrar gerçekleştirilmiştir. Bu testler sonucunda, modelin her bir kelimeyi doğru veya yanlış sınıflandırma durumu dikkatle kaydedilmiş; yanlış sınıflandırılan örneklerde modelin hangi kelimeyi tahmin ettiği de ayrıca not edilmiştir. Oluşturulan Python script'i ile katılımcıdan 30 çerçeve alınmakta, bu çerçevelerdeki anahtar nokta koordinatları birleştirilerek CNN+GRU modelinden tahminde bulunması istenmektedir (Şekil 5.20).



Şekil 5.20 Gerçek Zamanlı Test

Modelin tahmin süresi 35-45 ms arasında değişmektedir (Şekil 5.21). Bu da modelin hızlı bir şekilde tahminde bulunduğunu göstermektedir.

1/1	0s	39ms/step
1/1	0s	38ms/step
1/1	0s	42ms/step
1/1	0s	39ms/step
1/1	0s	44ms/step
1/1	0s	38ms/step
1/1	0s	52ms/step
1/1	0s	41ms/step
1/1	0s	61ms/step
1/1	0s	39ms/step
1/1	0s	35ms/step
1/1	0s	36ms/step
1/1	0s	39ms/step

Şekil 5.21 Modelin Tahmin Süresi

Oluşturulan script sayesinde, modelin genelleme yeteneği ile birlikte hatalı tahmin eğilimleri de analiz edilmiştir. Gerçek zamanlı performansın değerlendirilmesine yönelik olarak hazırlanan sonuçlar Tablo 5.5'te ayrıntılı şekilde sunulmaktadır. Türkçe işaret dili kelimelerinden oluşan 446 örnek için her kelimenin 10 tekrar üzerinden doğru tahmin edilme sayıları analiz edilmiştir. Elde edilen sonuçlara göre, modelin en yüksek başarıyı gösterdiği %100 doğru sınıflandırılan kelime sayısı 96 olup, bu oran tüm kelimeler içerisinde %21,52'ye karşılık gelmektedir. Bunun yanı sıra, %90 doğru tahminle sınıflandırılan kelime sayısı 102'dir ve bu oran %22,87 olarak hesaplanmıştır. Modelin orta üstü düzeyde başarı gösterdiği tahminler incelendiğinde, %80 ve %70 doğruluk değerine sahip kelimelerin sayıları sırasıyla 72 (%16,14) ve 92 (%20,63) olarak tespit edilmiştir. Sonuç olarak %70 ve üzeri doğruluk değeri ile tahmin edilen kelimelerin sayısı 362'dir. Bu tüm kelimelerin %81,17'sini oluşturmaktadır. Bu da yüksek bir orandır. Modelin orta düzeyde başarı gösterdiği tahminler incelendiğinde, %60 ve %50 doğruluk değerine sahip kelimelerin sayıları sırasıyla 62 (%13,9) ve 22 (%4,93) olarak tespit edilmiştir. Bu durum, modelin bazı kelimeleri sınıflandırmakta zaman zaman kararsızlık yaşadığını ve varyasyonlardan etkilenebildiğini göstermektedir. Tüm kelimeler için gerçek zamanlı doğruluk değeri ise %80,27'dir. Yani yapılan denemelerde kelimelerin %80,27'si doğru sınıflandırılmıştır.

Tablo 5.5 Veri Setinde Verisi Olmayan Katılımcı ile Gerçek Zamanlı Değerlendirme Sonucu

Doğruluk Değeri (%)	Kelime Sayısı	Oran (%)
100	96	21,52
90	102	22,87
80	72	16,14
70	92	20,63
60	62	13,9
50	22	4,93

Tablo 5.6'da doğru sayısı 5 ve 6 olan kelimeler gösterilmektedir. Tablodaki Kelime1, Kelime2, Kelime3 ve Kelime 4 sütunlarında yanlış tahminler sonucunda hangi kelimelerin tespit edildiğini göstermektedir.

Tablo 5.6 Tahmin Edilmesi İstenen Kelimeler ve Yanlış Tahmin Sonuçları

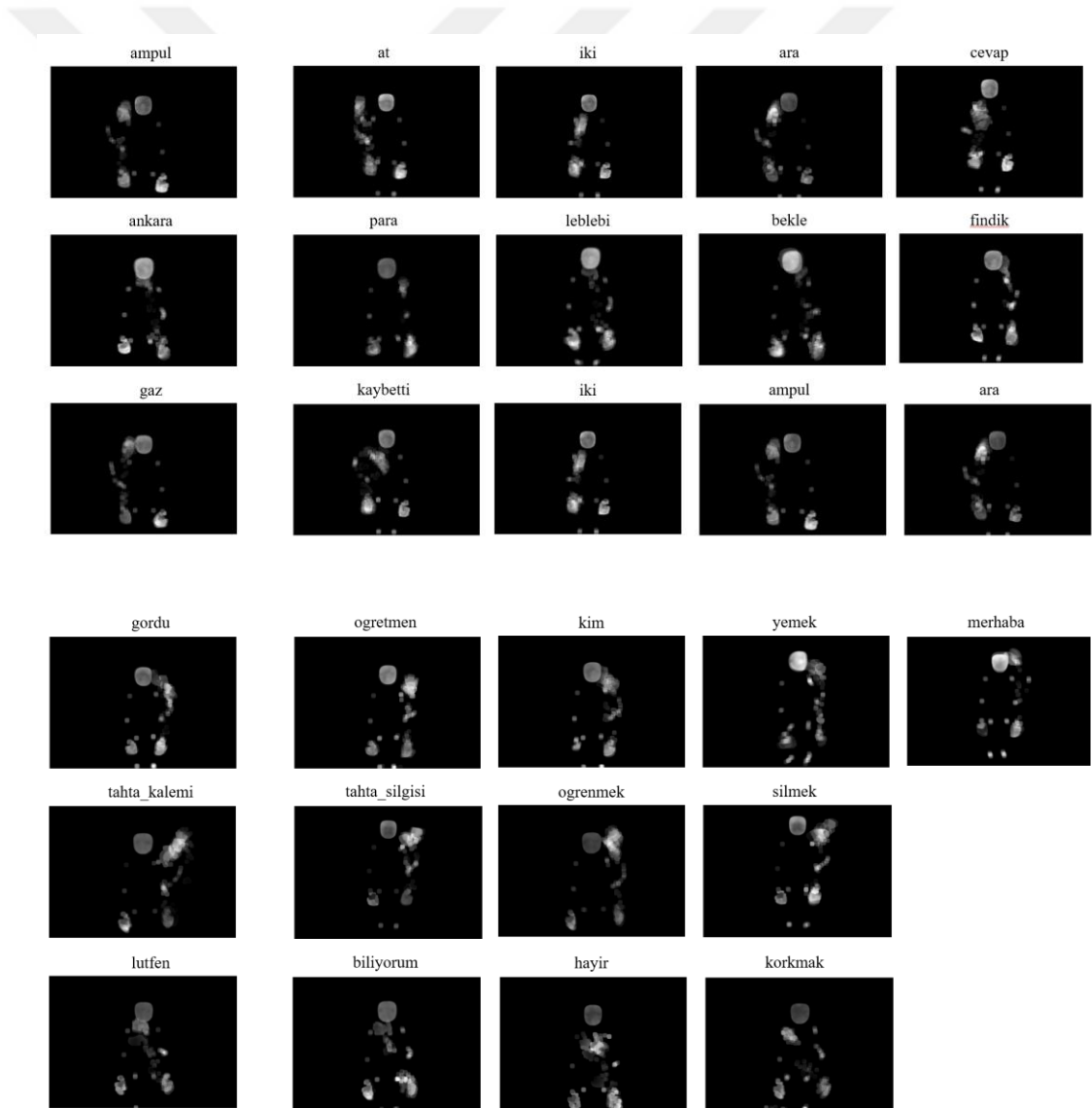
Yanlış tahmin edilen dört kelime						
Kelime	D	Y	Kelime1	Kelime2	Kelime3	Kelime4
ampul	5	5	at	iki	ara	cevap
ankara	5	5	para	leblebi	bekle	findik
aralik	5	5	nisan	haziran	mayis	mart
at	5	5	cevap	iki	ampul	ara
gaz	5	5	kaybetti	sivi	ekim	ayran
gordu	5	5	kim	ogretmen	yemek	merhaba
gosterebilir_misin	5	5	vuruyor	pil	yukaridaki	gram
kaybetti	5	5	sivi	gaz	vuruyor	yukaridaki
musluk	5	5	pasta	soru	secilir	yagmak
mutfak	5	5	prizde	yapiyor	tas	ve
sivi	5	5	kaybetti	gaz	kizilay	etti
haziran	5	5	temmuz	mayis	aralik	
iki	5	5	cevap	at	ampul	
kutlu	5	5	dost	olsun	doner	
lutfen	5	5	biliyorum	hayir	korkmak	
pasta	5	5	musluk	yagmak	soru	
tahta_kalemi	5	5	tahta_silgisi	ogrenmek	silmek	
rapor	5	5	kazandi	oldu		
sinif	5	5	merakli	koyum		
yatak_odasi	5	5	yemekhane	oyun_hamuru		
koyun	5	5	halay			
topa_vurmak	5	5	yakan_top			
cevap	6	4	at	iki	ampul	ara
olsun	6	4	bulustur	kutlu	dost	komsu
sinifta	6	4	bes	sari	heyecanli	cocuk
domates	6	4	ogretmen	anne	gordu	
durak	6	4	zurafa	kamyon	ust_gecit	
evli	6	4	karınca	odev	akraba	
guzel	6	4	benim	pazartesi	buldu	
koft	6	4	utu	sut	calisir	
leblebi	6	4	bekle	ankara	para	
litre	6	4	tahmin	lamba	bogaz	
mikrop	6	4	limon	dayi	tuvalet	
nisan	6	4	aralik	mart	lamba	
no	6	4	elma	koyum	bakmak	
ogretmen	6	4	domates	anne	gordu	
para	6	4	ankara	findik	leblebi	
salca	6	4	yok	sinek	korkmak	
seker	6	4	adi_ne	krem	neresi	
tahta	6	4	gecmek	taze	gunu	
taze	6	4	gunu	tahta	gecmek	
yok	6	4	salca	koyune	cevrende	
adam	6	4	dede	hediye		

Tablo 5.7 devamı

adi_ne	6	4	seker	adi
ayva	6	4	ilkbahar	hamam
bekle	6	4	leblebi	ankara
dost	6	4	kutlu	olsun
gecmek	6	4	tahta	taze
gunu	6	4	taze	tahta
gunun	6	4	baba	verdi
hayir	6	4	cevrende	lutfen
hayvanlar	6	4	aslan	evcil
istiklal_marsi	6	4	masa	bulustur
kasim	6	4	tahmin	gunes
kazandi	6	4	rapor	oldu
kes	6	4	anlat	zor
kirmizi	6	4	kiz	bebek
koyune	6	4	yok	cevrende
kurban	6	4	kasik	soylemek
merhaba	6	4	yemek	gordu
mevsim	6	4	hava	oldu
muz	6	4	adres	fotograf_cekiyor
ogrenmek	6	4	tahta_kalemi	silmek
okula	6	4	kurt	cikolata
pastel_boya	6	4	kantin	suluboya
tahmin	6	4	kasim	litre
tatil	6	4	ders	cizmek
yemek	6	4	merhaba	gordu
agirlik	6	4	tirnak	
ates	6	4	itfaiye	
banyo	6	4	yikanmak	
halay	6	4	koyun	
hava	6	4	mevsim	
ilkbahar	6	4	ayva	
itfaiye	6	4	ates	
karınca	6	4	evli	
kim	6	4	gordu	
kis	6	4	yaramaz	
kuzen	6	4	akraba	
limon	6	4	mikrop	
masa	6	4	istiklal_marsi	
meyve	6	4	ilcesinde	
nereye	6	4	trafik_kazasi	
simdi	6	4	usumek	

Tablodaki kelimeler incelendiğinde doğru sayısı 5 ve 6 olan kelimelerin Türkçe işaret dilindeki karşılıklarını yaparken benzer hareketleri içerdiği görülmüştür. Örneğin

tahta_kalemi kelimesi test edilirken hatalı tahmin kelimeleri içerisinde tahta_silgisi, ogrenmek ve silmek kelimeleri bulunmaktadır. tahta_kalemi kelimesi yapılırken elde bir kalem tutar gibi yapılıp, el omuz üzerine geçecek gibi kaldırıp yazı yazılmaktadır. tahta_silgisi kelimesini de yaparken aynı hareket yapılmakta fakat parmaklar kalem tutar gibi yapılmamaktadır. Yine silmek kelimesinde benzer hareket yapılıp parmaklar açık tutulmaktadır. Şekil 5.22’de ampul, ankara, gaz, gordu, tahta_kalemi, lutfen kelimeleri için optik akış yöntemi ile birleştirilen görselleri ve yanlış tahmin edilen kelimelerin optik akış görselleri bulunmaktadır. Kelimelerin optik akış yöntemi ile birleştirilmiş görselleri incelendiğinde tahmin edilmek istenen kelimeler ile yanlış tahmin edilen kelimelerin birbirlerine benzedikleri görülmektedir.



Şekil 5.22 Tahmin Edilmek İstenen Kelimelerin Optik Akış Görselleri

Modelin kişiye bağımlı olarak gerçek zamanlı performansını ortaya koymak için veri setinde verisi bulunan bir katılımcı ile model doğruluk değeri en yüksek olan CNN+GRU mimarisi test edilmiştir. Bu testte, verisi olmayan katılımcı ile yapılan değerlendirme sonucundaki doğru tahmin edilen kelime gruplarından 10'ar örnek alınmıştır. Bu kelimeler veri setinde verisi bulunan katılımcı ile test edilmiştir. Değerlendirme ortamı aydınlık gölge oluşturmayan, MediaPipe kütüphanesinin anahtar noktaları tespit edebileceği bir ortamda ve katılımcı ile kamera arasında 2 metre bırakılmıştır.

Bu testler sonucunda, modelin her bir kelimeyi doğru veya yanlış sınıflandırma durumu dikkatle kaydedilmiş; yanlış sınıflandırılan örneklerde modelin hangi kelimeyi tahmin ettiği de ayrıca not edilmiştir. Gerçek zamanlı performansın değerlendirilmesine yönelik olarak hazırlanan sonuçlar Tablo 5.7'de ayrıntılı şekilde sunulmaktadır. Veri setinde verisi bulunan katılımcı ile yapılan testlerde, modelin sınıflandırma doğruluğunun oldukça yüksek olduğu gözlemlenmiştir. Özellikle %100 ve 90 doğrulukla sınıflandırılan kelimelerin oranları %26,67 olup toplamda %53,34'lük yüksek başarı düzeyine işaret etmektedir. Bu sonuç, modelin daha önce gördüğü katılımcı örnekleri üzerinde güçlü bir genelleme kapasitesi geliştirdiğini ortaya koymaktadır. Ayrıca, %80 doğru tahmin edilen kelimelerin oranı %25,00 iken %70 doğru tahmin edilen kelimelerin oranı %18,33'tür. %70 ve daha fazla doğruluk değeri ile tahmin edilen kelimelerin oranı ise %96,67'dir. Bu durum, modelin tanıdığı örnekler üzerinden oldukça kararlı ve etkili bir performans sergilediğini göstermektedir. Modelin ortalama doğruluk değeri ise %85,5'dir. Bu gerçek zamanlı test açısından yüksek bir doğruluk değeridir.

Tablo 5.8 Veri Setinde Verisi Olan Katılımcı ile Gerçek Zamanlı Değerlendirme Sonucu

Doğruluk Değeri (%)	Kelime Sayısı	Oran (%)
100	16	26,67
90	16	26,67
80	15	25,00
70	11	18,33
60	2	3,33

Gerçek zamanlı performans değerlendirmesi kapsamında, modelin eğitim sürecinde verileri kullanılan bir katılımcı ile verileri daha önce model tarafından görülmemiş bir katılımcı üzerinde gerçekleştirilen test senaryoları sonucunda Tablo 5.8’de yer alan veriler elde edilmiştir. Eğitim verisinde yer almayan katılımcıyla yapılan testlerde doğruluk değeri %50 olan kelimelerin oranı %4,93 olarak kaydedilirken, eğitim verisinde yer alan katılımcıyla yapılan testlerde bu doğruluk seviyesine sahip herhangi bir kelime tespit edilmemiştir. Benzer şekilde, %60 doğruluk düzeyine sahip kelimelerin oranı eğitimde yer almayan katılımcı için %13,9 iken, eğitimde yer alan katılımcı için bu oran %3,33’e düşmüştür. Elde edilen sonuçlar, modelin daha önce karşılaştığı kullanıcı verileriyle yapılan gerçek zamanlı testlerde daha yüksek doğruluk oranlarına ulaştığını ve bu bağlamda öğrenilmiş veriler üzerinde daha başarılı performans sergilediğini ortaya koymaktadır.

Tablo 5.9 Gerçek Zamanlı Test Sonuçları

Doğruluk Değeri (%)	Veri Setinde Yer Almayan Katılımcılarda Bu Doğrulukta Tahmin Edilen Kelime Oranı (%)	Veri Setinde Yer Alan Katılımcılarda Bu Doğrulukta Tahmin Edilen Kelime Oranı (%)
100	21,52	26,67
90	22,87	26,67
80	16,14	25,00
70	20,63	18,33
60	13,9	3,33
50	4,93	-

Genel başarı oranı hesaplamak için her doğruluk seviyesi değeri o seviyedeki kelime oranı ile çarpılmış ve 100’e bölünmüştür. Elde edilen bu değerler toplanarak genel başarı oranı hesaplanmıştır.

$$Genel\ Başarı\ Oranı = \sum \left(\frac{Doğruluk\ Değeri\ (\%) \times (Kelime\ Oranı\ \%) }{100} \right)$$

Tablodaki oranlara göre hesaplanan genel başarı oranları aşağıdaki gibidir

- Veri setinde yer almayan katılımcı için genel başarı oranı: %80,26
- Veri setinde yer alan katılımcı için genel başarı oranı: %85,50

Bu fark, modelin daha önce gördüğü kullanıcılar üzerinde daha yüksek doğrulukla çalıştığını; ancak yeni kullanıcılar karşısında genel doğrulukta bir düşüş yaşadığını göstermektedir.



6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, TİD kelimelerinin gerçek zamanlı olarak tanınmasını sağlayan bir sistem geliştirilmiştir. MediaPipe Holistic kütüphanesi aracılığıyla elde edilen iskelet temsilleri ile çeşitli derin öğrenme mimarileri kullanılarak model eğitimi gerçekleştirilmiş; hem eğitim verisinde yer alan katılımcılarla hem de daha önce modele tanıtılmamış kullanıcılarla gerçek zamanlı testler uygulanmıştır. Kullanıcıya bağımlı gerçek zamanlı test doğruluğu %85,5 iken kullanıcıdan bağımsız gerçek zamanlı test doğruluğu %80,27'dir. Bu oranlar gerçek zamanlı test için son derece yüksektir. Bunun yanı sıra çevrim dışı test doğruluğu en yüksek %97,72'dir. Bu oran literatürdeki TİD üzerine yapılan önceki çalışmalardan daha yüksektir.

Geliştirilen TurkSign446 veri seti hem statik hem de dinamik işaretleri içeren 446 kelimelik kapsamıyla TİD alanında bugüne kadar oluşturulmuş en geniş ve ayrıntılı veri setlerinden biridir. Bu veri setinin cihazdan, çözünürlükten ve arka plan koşullarından bağımsız olması, platformlar arası uyumluluğu artırarak hem eğitim hem de uygulama aşamalarında esneklik sağlamaktadır. Ayrıca, açık erişime sunulması, Türk İşaret Dili üzerine çalışan araştırmacılar için önemli bir kaynak oluşturmakta ve benzer çalışmalarda karşılaştırılabilirlik ile yeniden üretilebilirlik açısından katkı sunmaktadır.

Dünyada işaret dili tanıma sistemleri büyük oranda Amerikan İşaret Dili (ASL) üzerine yoğunlaşmış olup, çok sayıda açık veri seti ve sistem bu alanda geliştirilmiştir. Buna karşın, Türk İşaret Dili üzerine yapılan çalışmalar sınırlı sayıdadır. Literatürde yalnızca birkaç veri seti (AUTSL, BSign22k, ERUSLR) bulunmaktadır. Bu bağlamda, TurkSign446 veri seti hem içerik zenginliği hem de güncel öğretim programlarına dayalı yapısıyla bu alandaki önemli bir boşluğu doldurmaktadır.

Modelleme sürecinde kullanılan CNN, LSTM, Bi-LSTM ve GRU gibi farklı mimarilerin yanı sıra, bu modellerin hibrit kombinasyonlarının da test edilmesi, derin öğrenme mimarilerinin işaret dili tanımadaki performanslarını karşılaştırmalı biçimde değerlendirme imkânı sunmuştur. Özellikle CNN+GRU mimarisi hem doğruluk hem de işlem süresi açısından öne çıkmış ve gerçek zamanlı uygulamalarda yüksek

kullanılabilirlik göstermiştir. Ayrıca, CBAM+ResNet18 gibi dikkat mekanizması içeren mimarilerin iskelet temsilleri üzerinde eğitilerek yüksek F1-Macro skoru elde etmesi, dikkat tabanlı yapıların işaret dili tanıma alanındaki potansiyelini gözler önüne sermektedir.

Geliştirilen sistemin platformdan bağımsız çalışması ve düşük donanım gereksinimiyle mobil ya da gömülü sistemlere entegre edilebilmesi, gelecekte işaret dili tanıma teknolojilerinin daha geniş bir kullanıcı kitlesine ulaşmasını mümkün kılmaktadır. Gerçek zamanlı işaret dili tanıma sistemlerinin sağlık hizmetleri, eğitim ortamları, kamu hizmetleri ve mobil uygulamalar gibi alanlarda yaygınlaştırılması, işitme engelli bireylerin toplumsal yaşama daha etkin katılımlarını destekleyecektir.

Çalışmada geliştirilen sistem, temel olarak kelime düzeyinde sınıflandırmaya odaklanmaktadır. Ancak işaret dili yalnızca kelime tanımaya indirgenemez; dilin yapısal bütünlüğü, bağlam, cümle yapısı ve yüz ifadeleri gibi çok boyutlu bileşenlerle tamamlanmaktadır. Bu nedenle, gelecekteki çalışmaların cümle düzeyinde tanıma ve çeviri yeteneklerine yönelmesi, sistemin ifade gücünü ve gerçek hayattaki işlevselliğini artıracaktır.

Veri setinin çeşitliliği ve kapsayıcılığı, derin öğrenme temelli sistemlerin başarısını doğrudan etkilemektedir. Bu doğrultuda, mevcut veri setinin farklı yaş gruplarından, cinsiyetlerden ve bölgesel işaret dili varyasyonlarını temsil eden katılımcılarla genişletilmesi, modelin genelleme yeteneğini artırarak kullanıcı bağımsızlığını güçlendirecektir.

Son olarak, sistemin mobil cihazlar, web tabanlı platformlar ve gömülü sistemlerde test edilmesi, gerçek hayat senaryolarına yönelik daha nitelikli geri bildirimler elde edilmesine olanak sağlayacaktır. Uygulama koşullarına göre optimizasyon yapılması, sistemin erişilebilirliğini ve pratikliğini artıracaktır.

Bu tez çalışmasıyla Türk İşaret Dili'ne yönelik geliştirilen sistem, yalnızca teknik bir başarı ortaya koymakla kalmamış; aynı zamanda erişilebilirlik, kapsayıcılık ve sosyal katkı açısından da değerli bir zemin oluşturmuştur. Gelecekteki araştırmaların bu

temeller üzerinde ilerleyerek işaret dili tanıma alanını daha da ileriye taşıması beklenmektedir.



KAYNAKLAR

- Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., Xydopoulos, G. J., Atzakas, K. ve Daras, P., 2020, A comprehensive study on sign language recognition methods, arXiv.
- Ahmed, W., Chanda, K. ve Mitra, S., 2017, Vision based hand gesture recognition using dynamic time warping for indian sign language, Proceedings - 2016 International Conference on Information Science, ICIS 2016.
- Akdağ, A. (2024). Evrişimsel Sinir Ağları Kullanılarak Video Tabanlı İzole İşaret Dili Tanıma. Doktora Tezi, *Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü*. Konya.
- Akmeşe, P. (2016). Türk İşaret Dili TİD Eğitimi ve Yüksek Öğretim Programlarında İşaret Dili Dersi. Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi, 17(3), 341-361.
- Alaftekin, M., Pacal, I., & Cicek, K. (2024). Real-time sign language recognition based on YOLO algorithm. *Neural Computing and Applications*, 36(14), 7609-7624.
- Aldahhri, E., Aljuhani, R., Alfaidi, A., Alshehri, B., Alwadei, H., Aljojo, N., Alshutayri, A. ve Almazroi, A., 2023, Arabic sign language recognition using convolutional neural network and mobilenet, *Arabian Journal for Science and Engineering*, 48 (2).
- Alyami, S., Luqman, H. ve Hammoudeh, M., 2023, Isolated arabic sign language recognition using a transformer-based model and landmark keypoints, *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*
- Areerob, P., Matangkasombut, T., Monnikhof, K. O., & Kumwilaisak, W. (2024, May). Crowded Scene PPE Detection Using Attention Based YOLOv7 and Alpha Pose. In 2024 21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 1-6). IEEE.
- Arkushin, R. S., Moryossef, A., & Fried, O. (2023). Ham2pose: Animating sign language notation into pose sequences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21046-21056).
- Arooj, S., Altaf, S., Ahmad, S., Mahmoud, H., & Mohamed, A. S. N. (2024). Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language. *Journal of King Saud University-Computer and Information Sciences*, 36(2), 101934.
- Arun Prasath, G., & Annapurani, K. (2023). Prediction of sign language recognition based on multi layered CNN. *Multimedia Tools and Applications*, 82(19), 29649-29669.

- Arya, V., & Maji, S. (2024, August). Enhancing Human Pose Estimation: A Data-Driven Approach with MediaPipe BlazePose and Feature Engineering Analysing. In 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT) (pp. 1-6). IEEE.
- Badiola-Bengoia, A., & Mendez-Zorrilla, A. (2021). A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. *Sensors*, 21(18), 5996.
- Bankur, F., & Kaya, M. (2022). Deep Learning Based Recognition of Turkish Sign Language Letters with Unique Data Set. *Turkish Journal of Science and Technology*, 17(2), 251-260.
- Bao, W., Niu, T., Wang, N., & Yang, X. (2023). Pose estimation and motion analysis of ski jumpers based on ECA-HRNet. *Scientific Reports*, 13(1), 6132.
- Barczak, A. L. C., Reyes, N. H., Abastillas, M., Piccio, A. ve Susnjak, T., 2011, A new 2d static hand gesture colour image dataset for asl gestures.
- Bhiri, N. M., Ameer, S., Alouani, I., Mahjoub, M. A., & Khalifa, A. B. (2023). Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions. *Expert Systems with Applications*, 226, 120125.
- Bibin Sebastian. (2024, Eylül 10). Human Action Recognition using Detectron2 and LSTM. <https://learnopencv.com/human-action-recognition-using-detectron2-and-lstm/>
- Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time assamese sign language recognition using mediapipe and deep learning. *Procedia Computer Science*, 218, 1384-1393.
- Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time assamese sign language recognition using mediapipe and deep learning. *Procedia Computer Science*, 218, 1384-1393.
- Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291-7299).
- Chaikaew, A., Somkuan, K., & Yuyen, T. (2021, March). Thai sign language recognition: an application of deep neural network. In 2021 joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunication engineering (pp. 128-131). IEEE.
- Chakraborty, S., Paul, P., Bhattacharjee, S., Sarkar, S., & Chakraborty, A. (2023). Sign Language Recognition Using Landmark Detection, GRU and LSTM. *American Journal of Electronics & Communication*, 3(3), 20-26.

- Chen, L., & Fisher, R. B. (2024). Miso: Monitoring inactivity of single older adults at home using rgb-d technology. *ACM Transactions on Computing for Healthcare*, 5(3), 1-19.
- Chen, M., & Tan, G. (2024, July). FANpose: 2D human pose estimation with fully attentional networks under vision transformer baselines. In *Third International Symposium on Computer Applications and Information Systems (ISCAIS 2024)* (Vol. 13210, pp. 879-884). SPIE.
- Colak, S., Sadeghzadeh, A., & Islam, M. B. (2023, March). T-SignSys: An Efficient CNN-Based Turkish Sign Language Recognition System. In *International Conference on Advanced Engineering, Technology and Applications* (pp. 226-241). Cham: Springer Nature Switzerland.
- Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O. L., Pozzobon de Bem, P., dos Santos de Moura, R., Olino de Albuquerque, A., ... & Trancoso Gomes, R. A. (2020). Rice crop detection using LSTM, Bi-LSTM, and machine learning models from Sentinel-1 time series. *Remote Sensing*, 12(16), 2655.
- Cui, R., Liu, H. ve Zhang, C., 2019, A deep neural framework for continuous sign language recognition by iterative training, *IEEE Transactions on Multimedia*, 21 (7).
- Çelik, Ö. (2019). *Yapay Zeka Tekniği İşitme Engelliler İçin Uzaktan İletişim Sistemi*. Eskişehir: Eskişehir Osmangazi Üniversitesi.
- Damaneh, M. M., Mohanna, F. ve Jafari, P., 2023, Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using orb descriptor and gabor filter, *Expert Systems with Applications*, 211, 118559-118559.
- Das, S., Imtiaz, M. S., Neom, N. H., Siddique, N. ve Wang, H., 2023, A hybrid approach for bangla sign language recognition using deep transfer learning model with random forest classifier, *Expert Systems with Applications*, 213.
- Davoudi Kashkoli, M., Javied, A., Barrera-Animas, A. Y., & Davila Delgado, J. M. (2024). A synthetic data approach for object detection in super low resolution images. *Proceedings of the 2024 International Conference on Innovation in Artificial Intelligence (ICIAI '24)*, 86–91.
- de Castro, G. Z., Guerra, R. R. ve Guimarães, F. G., 2023, Automatic translation of sign language with multi-stream 3d cnn and generation of artificial depth maps, *Expert Systems with Applications*, 215.
- Dikyuva, H., Makaroğlu, B. ve Arık, E. (2015). *Türk işaret dili dilbilgisi kitabı*. Ankara: Aile ve Sosyal Politikalar Bakanlığı Yayınları.
- Dill, S., Rösch, A., Rohr, M., Güney, G., De Witte, L., Schwartz, E., & Hoog Antink, C. (2023, September). Accuracy Evaluation of 3D Pose Estimation with

- MediaPipe Pose for Physical Exercises. In *Current Directions in Biomedical Engineering* (Vol. 9, No. 1, pp. 563-566). De Gruyter.
- Dill, S., Rösch, A., Rohr, M., Güney, G., De Witte, L., Schwartz, E., & Hoog Antink, C. (2023, September). Accuracy Evaluation of 3D Pose Estimation with MediaPipe Pose for Physical Exercises. In *Current Directions in Biomedical Engineering* (Vol. 9, No. 1, pp. 563-566). De Gruyter.
- Dong, Y., Liu, J., & Yan, W. (2021). Dynamic hand gesture recognition based on signals from specialized data glove and deep learning algorithms. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-14.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G. ve Gelly, S., 2020, An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929.
- Erten, H., & Arıcı, N. (2022). İşaret dilinin tarihi serüveni ve Türk işaret dili. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 24(1), 1-14.
- Fang, Y., Han, Z., Hu, Z., & Wang, Z. (2021, December). Human Posture Estimation. In *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)* (pp. 220-234). IEEE.
- Farooq, U., Mohd Rahim, M. S., & Abid, A. (2023). A multi-stack RNN-based neural machine translation model for English to Pakistan sign language translation. *Neural Computing and Applications*, 35(18), 13225-13238.
- Gao, Q., Liu, J., Ju, Z., & Zhang, X. (2019). Dual-hand detection for human–robot interaction by a parallel network based on hand detection and body pose estimation. *IEEE Transactions on Industrial Electronics*, 66(12), 9663-9672.
- Garg, S., Saxena, A., & Gupta, R. (2023). Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application. *Journal of Ambient Intelligence and Humanized Computing*, 14(12), 16551-16562.
- Georgiou, S. (2023). Sign Language as a Universal Language of Communication: An Alternative Approach to Nation Branding. In *Global Developments in Nation Branding and Promotion: Theoretical and Practical Approaches* (pp. 156-168). IGI Global.
- Google AI. (2024, Eylül 10). MediaPipe Solutions guide. <https://ai.google.dev/edge/mediapipe/solutions/guide>
- Goyal, K. (2023). Indian sign language recognition using mediapipe holistic. arXiv preprint arXiv:2304.10256.
- Gökçe, Ç., Özdemir, O., Kındıroğlu, A. A. ve Akarun, L., 2020, Score-level multi cue fusion for sign language recognition, *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*).

- Gökçe, Ç., Özdemir, O., Kındıroğlu, A. A. ve Akarun, L., 2020, Score-level multi cue fusion for sign language recognition, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).
- Gündüz, C. ve Polat, H., 2021, Turkish sign language recognition based on multistream data fusion, Turkish Journal of Electrical Engineering and Computer Sciences, 29 (2).
- Gündüz, C. ve Polat, H., 2021, Turkish sign language recognition based on multistream data fusion, Turkish Journal of Electrical Engineering and Computer Sciences, 29 (2).
- Gündüz, C., & POLAT, H. (2021). Turkish sign language recognition based on multistream data fusion. Turkish Journal of Electrical Engineering and Computer Sciences, 29(2), 1171-1186.
- Güney, S., & Erkuş, M. (2022). A real-time approach to recognition of Turkish sign language by using convolutional neural networks. Neural Computing and Applications, 34(5), 4069-4079.
- Gweth, Y. L., Plahl, C. ve Ney, H., 2012, Enhanced continuous sign language recognition using pca and neural network features, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.
- Halder, A., & Tayade, A. (2021). Real-time vernacular sign language recognition using mediapipe and machine learning. Journal homepage: www.ijrpr.com ISSN, 2582, 7421.
- Hamza, H. M. ve Wali, A., 2023, Pakistan sign language recognition: Leveraging deep learning models with limited dataset, Machine Vision and Applications, 34 (5).
- He, K., Zhang, X., Ren, S. ve Sun, J., 2016, Deep residual learning for image recognition, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- Hernández, Ó. G., Morell, V., Ramon, J. L., & Jara, C. A. (2021). Human pose detection for robotic-assisted and rehabilitation environments. Applied Sciences, 11(9), 4183.
- Hisham, B., & Hamouda, A. (2021). Arabic sign language recognition using Ada-Boosting based on a leap motion controller. International Journal of Information Technology, 13(3), 1221-1234.
- Imran, J. ve Raman, B., 2020, Deep motion templates and extreme learning machine for sign language recognition, Visual Computer, 36 (6).
- Ivan Grishchenko and Valentin Bazarevsky. 2020. <https://research.google/blog/mediapipe-holistic-simultaneous-face-hand-and-pose-prediction-on-device/>

- Jafarzadeh, P., Virjonen, P., Nevalainen, P., Farahnakian, F., & Heikkonen, J. (2021, October). Pose estimation of hurdles athletes using openpose. In 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-6). IEEE.
- Jebali, M., Dakhli, A. ve Bakari, W., 2024, Deep learning-based sign language recognition system using both manual and non-manual components fusion, AIMS Mathematics, 9 (1), 2105-2122.
- Jo, B., & Kim, S. (2022). Comparative analysis of OpenPose, PoseNet, and MoveNet models for pose estimation in mobile devices. *Traitement du Signal*, 39(1), 119.
- Karacı, A., Akyol, K., & Turut, M. U. (2021). Real-Time Turkish Sign Language Recognition Using Cascade Voting Approach with Handcrafted Features. *Applied Computer Systems*, 26(1), 12-21.
- Karim, M. E., Foysal, M., & Das, S. (2022, November). Stock price prediction using Bi-LSTM and GRU-based hybrid deep learning approach. In *Proceedings of Third Doctoral Symposium on Computational Intelligence: DoSCI 2022* (pp. 701-711). Singapore: Springer Nature Singapore.
- Kasapbaşı, A., & Canbolat, H. (2023, June). Prediction of Turkish Sign Language Alphabets Utilizing Deep Learning Method. In 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-7). IEEE.
- Kasapbaşı, A., Elbushra, A. E. A., Omar, A. H., & Yilmaz, A. (2022). DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals. *Computer methods and programs in biomedicine update*, 2, 100048.
- Katılmış, Z., & Karakuzu, C. (2021). ELM based two-handed dynamic turkish sign language (TSL) word recognition. *Expert Systems with Applications*, 182, 115213.
- Katoch, S., Singh, V., & Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. *Array*, 14, 100141.
- Kaur, S., Gupta, A., Aggarwal, A., Gupta, D., & Khanna, A. (2021). Sign language recognition using Microsoft Kinect. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 1* (pp. 637-645). Springer Singapore.
- Khotimah, W. N., Suciati, N., & Benedict, I. (2018, August). Indonesian Sign Language Recognition by Using the Static and Dynamic Features. In 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 293-298). IEEE.
- Kındıroğlu, A. A., Özdemir, O. ve Akarun, L., 2023, Aligning accumulative representations for sign language recognition, *Machine Vision and Applications*, 34 (1).

- Kim, J. W., Choi, J. Y., Ha, E. J., & Choi, J. H. (2023). Human pose estimation using mediapipe pose and optimization method based on a humanoid model. *Applied sciences*, 13(4), 2700.
- Kindiroglu, A. A., Kara, O., Ozdemir, O., & Akarun, L. (2024). Transfer Learning for Cross-dataset Isolated Sign Language Recognition in Under-Resourced Datasets. *arXiv preprint arXiv:2403.14534*.
- Kindiroglu, A. A., Ozdemir, O. ve Akarun, L., 2019, Temporal accumulative features for sign language recognition, *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*.
- Konstantinidis, D., Dimitropoulos, K. ve Daras, P., 2018a, A deep learning approach for analyzing video and skeletal features in sign language recognition, *IST 2018 - IEEE International Conference on Imaging Systems and Techniques, Proceedings*.
- Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Sign language detection and recognition using deep learning. *Electronics*, 11(11), 1780.
- Kour, K. P. ve Mathew, L., 2017, Sign language recognition using image processing, *International Journal of Advanced Research in Computer Science and Software Engineering*, 7 (8).
- Krizhevsky, A., Sutskever, I. ve Hinton, G. E., 2012, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*.
- Kuznetsova, A., & Kimmelman, V. (2024). Testing MediaPipe Holistic for Linguistic Analysis of Nonmanual Markers in Sign Languages. *arXiv preprint arXiv:2403.10367*.
- Kuznetsova, A., & Kimmelman, V. (2024). Testing MediaPipe Holistic for Linguistic Analysis of Nonmanual Markers in Sign Languages. *arXiv preprint arXiv:2403.10367*.
- Laines, D., Gonzalez-Mendoza, M., Ochoa-Ruiz, G. ve Bejarano, G., 2023, Isolated sign language recognition based on tree structure skeleton images, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 276-284.
- Lee, H., Oh, B., & Kim, S. C. (2024). Recognition of Forward Head Posture Through 3D Human Pose Estimation With a Graph Convolutional Network: Development and Feasibility Study. *JMIR Formative Research*, 8(1), e55476.
- Li, N., Wang, Y., Liu, F., & Huang, W. (2024, August). Real-time multitarget fall detection based on OpenPose. In *Seventh International Conference on Advanced Electronic Materials, Computers, and Software Engineering (AEMCSE 2024)* (Vol. 13229, pp. 751-755). SPIE

- Li, T. H. S., Kao, M. C. ve Kuo, P. H., 2016, Recognition system for home-service-related sign language using entropy-based k-means algorithm and abc-based hmm, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46 (1).
- Lin, H., Chen, H., & Lin, J. (2024). Deep neural network uncertainty estimation for early oral cancer diagnosis. *Journal of Oral Pathology & Medicine*, 53(5), 294-302.
- Liu, T., Jin, H., Xie, X., Fang, H., Wei, D., & Li, A. (2022). Bi-LSTM model for time series leaf area index estimation using multiple satellite products. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.
- Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T. ve Xie, S., 2022, A convnet for the 2020s, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11966-11976.
- Lu, C., Kozakai, M., & Jing, L. (2023). Sign Language Recognition with Multimodal Sensors and Deep Learning Methods. *Electronics*, 12(23), 4827.
- Madani, H. ve Nahvi, M., 2013, Isolated dynamic persian sign language recognition based on camshift algorithm and radon transform, 1st Iranian Conference on Pattern Recognition and Image Analysis, PRIA 2013.
- Mahardika T, N. Q., Fuadah, Y. N., Jeong, D. U., & Lim, K. M. (2023). PPG signals-based blood-pressure estimation using grid search in hyperparameter optimization of CNN–LSTM. *Diagnostics*, 13(15), 2566.
- Mallik, B., Rahim, M. A., Miah, A. S. M., Yun, K. S., & Shin, J. (2024). Virtual Keyboard: A Real-Time Hand Gesture Recognition-Based Character Input System Using LSTM and Mediapipe Holistic. *Comput. Syst. Sci. Eng.*, 48(2), 555-570.
- Mandasari, S., Irfan, D., Wanayumini, W., & Rosnelly, R. (2023). COMPARISON OF SGD, ADADELTA, ADAM OPTIMIZATION IN GENDER CLASSIFICATION USING CNN. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 9(3), 345-354.
- Masood, S., Srivastava, A., Thuwal Harish, C. ve Ahmad, M., 2018, Real-time sign language gesture (word) recognition from video sequences using cnn and rnn, *Intelligent Engineering Informatics*, Singapore, 623-632.
- Mercanoglu Sincan, O., & Yalim Keles, H. (2020). AUTSL: A Large Scale Multimodal Turkish Sign Language Dataset and Baseline Methods. *arXiv e-prints*, arXiv-2008.
- More, V., Sangamnerkar, S., Thakare, V., Mane, D., & Dolas, R. (2018). Sign language recognition using image processing. *SIGN*, 21, 22nd.
- Naz, N., Sajid, H., Ali, S., Hasan, O., & Ehsan, M. K. (2023). Signgraph: An efficient and accurate pose-based graph convolution approach toward sign language recognition. *IEEE Access*, 11, 19135-19147.

- Özcan, T. A. Y. Y. İ. P., & Baştürk, A. (2021). ERUSLR: A new Turkish sign language dataset and its recognition using hyperparameter optimization aided convolutional neural network. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36(1), 527-542.
- Özdemir, O., Baytaş, İ. M. ve Akarun, L., 2023, Multi-cue temporal modeling for skeleton-based sign language recognition, *Frontiers in Neuroscience*, 17.
- Özdemir, O., Baytaş, İ. M. ve Akarun, L., 2023, Multi-cue temporal modeling for skeleton-based sign language recognition, *Frontiers in Neuroscience*, 17.
- Özdemir, O., Kındıroğlu, A. A., Camgöz, N. C., & Akarun, L. (2020). Bosphorussign22k sign language recognition dataset. *arXiv preprint arXiv:2004.01283*.
- Öztürk, Ş., & Keles, H. Y. (2024, May). E-TSL: A Continuous Educational Turkish Sign Language Dataset with Baseline Methods. In *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1-7). IEEE.
- Paharia, N., Jadon, R. S., & Gupta, S. K. (2023). Optimization of convolutional neural network hyperparameters using improved competitive gray wolf optimizer for recognition of static signs of Indian Sign Language. *Journal of Electronic Imaging*, 32(2), 023042-023042.
- Parashar, D., Mishra, O., Sharma, K., & Kukker, A. (2023). Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Model. *Revue d'Intelligence Artificielle*, 37(5).
- Pirani, M., Thakkar, P., Jivrani, P., Bohara, M. H., & Garg, D. (2022, April). A comparative analysis of ARIMA, GRU, LSTM and BiLSTM on financial time series forecasting. In *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1-6). IEEE.
- Pu, J., Zhou, W. ve Li, H., 2019, Iterative alignment network for continuous sign language recognition, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Rahim, M. A., Shin, J. ve Yun, K. S., 2020, Hand gesture-based sign alphabet recognition and sentence interpretation using a convolutional neural network, *Annals of Emerging Technologies in Computing*, 4 (4).
- Raj, R. D. ve Jasuja, A., 2018, British sign language recognition using hog, 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2018.
- Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, 35(23), 17095-17112.

- Rezende, R. F., Guerra, L. B., & Carvalho, S. A. D. S. (2020). Satisfaction of deaf patients with the health care. *Revista CEFAC*, 22, e8119.
- Rodríguez, J. ve Martínez, F., 2018, Towards on-line sign language recognition using cumulative sd-vlad descriptors, *Communications in Computer and Information Science*.
- Ronchetti, F., Quiroga, F., Estrebou, C. A., and Lanzarini, L. C. (2016a). Handshape recognition for argentinian sign language using probsom. *Journal of Computer Science & Technology*, 16.
- Ronchetti, F., Quiroga, F., Estrebou, C. A., Lanzarini, L. C., and Rosete, A. (2016b). Lsa64: an argentinian sign language dataset. In *XXII Congreso Argentino de Ciencias de la Computación*.
- Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L. ve Rosete, A., 2016a, Sign language recognition without frame-sequencing constraints: A proof of concept on the argentinian sign language, *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*).
- Samaan, G. H., Wadie, A. R., Attia, A. K., Asaad, A. M., Kamel, A. E., Slim, S. O., Abdallah, M. S. ve Cho, Y. I., 2022, Mediapipe's landmarks with rnn for dynamic sign language recognition, *Electronics* (Switzerland), 11 (19).
- Selvaraj, P., C, G. N., Kumar, P. ve Khapra, M. M., 2021, Openhands: Making sign language recognition accessible with pose-based pretrained models across languages, *CoRR*, abs/2110.05877.
- Sharma, A., Mittal, A., Singh, S. ve Awatramani, V., 2020, Hand gesture recognition using image processing and feature extraction techniques, *Procedia Computer Science*.
- Sincan, O. M. ve Keles, H. Y., 2021, Using motion history images with 3d convolutional networks in isolated sign language recognition, *CoRR*, abs/2110.12396.
- Sincan, O. M., & Keles, H. Y. (2020). Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE access*, 8, 181340-181355.
- Singh, A. K., Kumbhare, V. A., & Arthi, K. (2021, June). Real-time human pose detection and recognition using mediapipe. In *International conference on soft computing and signal processing* (pp. 145-154). Singapore: Springer Nature Singapore
- Sonugür, G., & Çaylı, A. (2023). Turkish sign language recognition using fuzzy logic asisted ELM and CNN methods. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-13.
- SÖMEN, O. (2023). TÜRK SAĞIR VE İŞİTME ENGELLİLERDE İLETİŞİM OLGUSU. *Premium e-Journal of Social Science (PEJOSS)*, 7(27), 106-126.

- Sundar, B., & Bagyammal, T. (2022). American sign language recognition for alphabets using MediaPipe and LSTM. *Procedia Computer Science*, 215, 642-651.
- Tan, M. ve Le, Q. V., 2019, Efficientnet: Rethinking model scaling for convolutional neural networks, 36th International Conference on Machine Learning, ICML 2019.
- Terven, J., Cordova-Esparza, D. M., Romero-González, J. A., Ramírez-Pedraza, A., & Chávez-Urbiola, E. A. (2025). A comprehensive survey of loss functions and metrics in deep learning. *Artificial Intelligence Review*, 58(7), 195.
- Torun, C., & Karacı, A. (2024). Derin Öğrenmeye Dayalı 2 Boyutlu İnsan Poz Tahmin Modellerinin Karşılaştırılması. *Sürdürülebilir Mühendislik Uygulamaları Ve Teknolojik Gelişmeler Dergisi*, 7(2), 185-196. <https://doi.org/10.51764/smutgd.1573626>
- Tölgyessy, M., Dekan, M., Chovanec, L., & Hubinský, P. (2021). Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors*, 21(2), 413.
- Tran, D., Bourdev, L., Fergus, R., Torresani, L. ve Paluri, M., 2015, Learning spatiotemporal features with 3d convolutional networks, 2015 IEEE International Conference on Computer Vision (ICCV), 4489-4497.
- Üstek, İ., Desai, J., Torrecillas, I. L., Abadou, S., Wang, J., Fever, Q., ... & Tsourdos, A. (2023, August). Two-Stage Violence Detection Using ViTPose and Classification Models at Smart Airports. In 2023 IEEE Smart World Congress (SWC) (pp. 797-802). IEEE.
- Wang, F., Du, Y., Wang, G., Zeng, Z. ve Zhao, L., 2021a, (2+1)d-slr: An efficient network for video sign language recognition, *Neural Computing and Applications*.
- Wang, J., Wang, P., Tian, H., Tansey, K., Liu, J., & Quan, W. (2023). A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Computers and Electronics in Agriculture*, 206, 107705.
- Wang, Y., Wang, R., Shi, H., & Liu, D. (2024). MS-HRNet: multi-scale high-resolution network for human pose estimation. *The Journal of Supercomputing*, 1-23.
- WHO. 2024. Deafness and hearing loss. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>, Erişim Tarihi;10/06/2024
- Wu, Q., Xu, G., Zhang, S., Li, Y., & Wei, F. (2020, July). Human 3D pose estimation in a lying position by RGB-D images for medical diagnosis and rehabilitation. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 5802-5805). IEEE.

- Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).
- Xu, S., Sun, H., Li, B., Zhao, D., Li, H., Wang, K., ... & Wang, X. (2024, June). Bi-LSTM-based load forecasting for power grid. In Journal of Physics: Conference Series (Vol. 2781, No. 1, p. 012002). IOP Publishing.
- Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2023). Vitpose++: Vision transformer for generic body pose estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence.
- Yang, Q. ve Peng, J. Y., 2014, Chinese sign language recognition method based on depth image information and surf-bow, Moshu Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence, 27 (8).
- Yanikkerem, E., & Esmeray, N. (2017). İşitme ve konuşma engelli kadınların yaşadığı güçlükler. Annals of Health Sciences Research, 6(2), 38-46.
- Yasir, F., Prasad, P. W. C., Alsadoon, A. ve Elchouemi, A., 2016, Sift based approach on bangla sign language recognition, 2015 IEEE 8th International Workshop on Computational Intelligence and Applications, IWCIA 2015 - Proceedings.
- Zeshan, U. (2006). Interrogative and negative constructions in sign language (p. 375). Ishara Press.
- Zhang, X. ve Li, X., 2019, Dynamic gesture recognition based on memp network, Future Internet, 11, 91-91.
- Zhou, Y., Wang, X., Xu, X., Zhao, L., & Song, J. (2022, July). X-hrnet: Towards lightweight human pose estimation with spatially unidimensional self-attention. In 2022 IEEE international conference on multimedia and expo (ICME) (pp. 01-06). IEEE.
- Zhu, H., Jie, C., & Jiang, S. (2020). Multi-Person Full Body Pose Estimation. arXiv preprint arXiv:2008.10060