



ÖN-EĞİTİLMİŞ DERİN SİNİR AĞLARI KULLANILARAK ULTRA GENİŞ BANT RADAR TABANLI EL HAREKETİ TANIMA

Nurullah ÇALIK*

İstanbul Medeniyet Üniversitesi, Mühendislik ve Doğa Fakültesi, Biyomedikal Mühendisliği Bölümü, İstanbul, Türkiye

Anahtar Kelimeler	Öz
<i>El Hareketi Tanıma, Ultra Geniş-Bant, Radar, Derin Öğrenme, Ön-eğitilmiş Model.</i>	Ultra geniş-bant (UGB) radar sensörleri insan hareketlerinin tanınmasında kritik bir öneme sahiptir. Sahip oldukları yüksek frekans özelliği ile hassas hareketlerin ayırt edilmesinde kullanılabilirler. Ayrıca bu özellik, sensör boyutlarının portatif olmasına ve birçok alanda kolay kullanımına imkân tanır. Her ne kadar RGB kameralar üzerinden bu işlem yapılabilse de özellikle mahremiyet gizliliği konusunda problemler ile karşılaşmaktadır. UGB sensörlerden elde edilen verilerde kişiler anonimleştiği için sadece hareket örüntüsü ele alınabilmektedir. Dolayısıyla, ev içi izleme ve yaşlı takibi gibi konularda önemli bir potansiyeli bulunmaktadır. Bu çalışmada, üç farklı sensör üzerinden toplanan radar verilerinden el hareketlerinin sınıflandırılması yapılmaktadır. Radar görselleri sağ, sol ve üst olacak şekilde arka arkaya bir araya getirilerek 3 kanallı bir görüntü elde edilmiştir. Ardından bu görüntüler üzerinden ön-eğitilmiş modeller eğitilmiş ve test edilmiştir. Eğitim ve test oranı 50:50 olduğu durumda %97.93 (DenseNet201) 75:25 oranında ise %97.65 (EfficientNetB0) oranında bir başarımla elde edilmiştir. Tek sensör kullanımı yerine öne sürülen strateji ile modellerin eğitilmesinin önemli bir iyileştirme yaptığı ortaya koyulmuştur.

ULTRA WIDEBAND RADAR-BASED HAND GESTURE RECOGNITION USING PRE-TRAINED DEEP NEURAL NETWORKS

Keywords	Abstract
<i>Hand Gesture Recognition, Ultra Wide-Band, RADAR, Deep Learning, Pre-trained Model.</i>	Ultra-wideband (UWB) radar sensors play a pivotal role in recognizing human movements. They can be used to distinguish sensitive movements with their high frequency feature. In addition, this feature allows the sensor dimensions to be portable and easy to use in many areas. Although this process can be done via RGB cameras, problems are encountered, especially regarding privacy. Since people are anonymized in the data obtained from UWB sensors, only their movement patterns can be considered. Therefore, it has significant potential in issues such as in-home and elderly monitoring. In this study, hand movements are classified from radar data collected through three different sensors. A 3-channel image was obtained by combining radar images one after the other, right, left, and top. Then, pre-trained models were trained and tested on these images. When the training and testing ratio was 50:50, a performance rate of 97.93% (DenseNet201) was achieved, and when the ratio was 75:25, a performance rate of 97.65% (EfficientNetB0) was achieved. It has been shown that training models with the proposed strategy instead of using a single sensor makes a significant improvement.

Alıntı / Cite

Çalık, N., (2024). Ön-Eğitilmiş Derin Sinir Ağları Kullanılarak Ultra Geniş Bant Radar Tabanlı El Hareketi Tanıma, Mühendislik Bilimleri ve Tasarım Dergisi, 12(1), 205-216.

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
N. Çalık, 0000-0002-7351-4980	Başvuru Tarihi / Submission Date	11.01.2024
	Revizyon Tarihi / Revision Date	10.02.2024
	Kabul Tarihi / Accepted Date	26.02.2024
	Yayın Tarihi / Published Date	25.03.2024

* İlgili yazar / Corresponding author: nurullah.calik@medeniyet.edu.tr, +90-216-280-3333

ULTRA WIDEBAND RADAR-BASED HAND GESTURE RECOGNITION USING PRE-TRAINED DEEP NEURAL NETWORKS

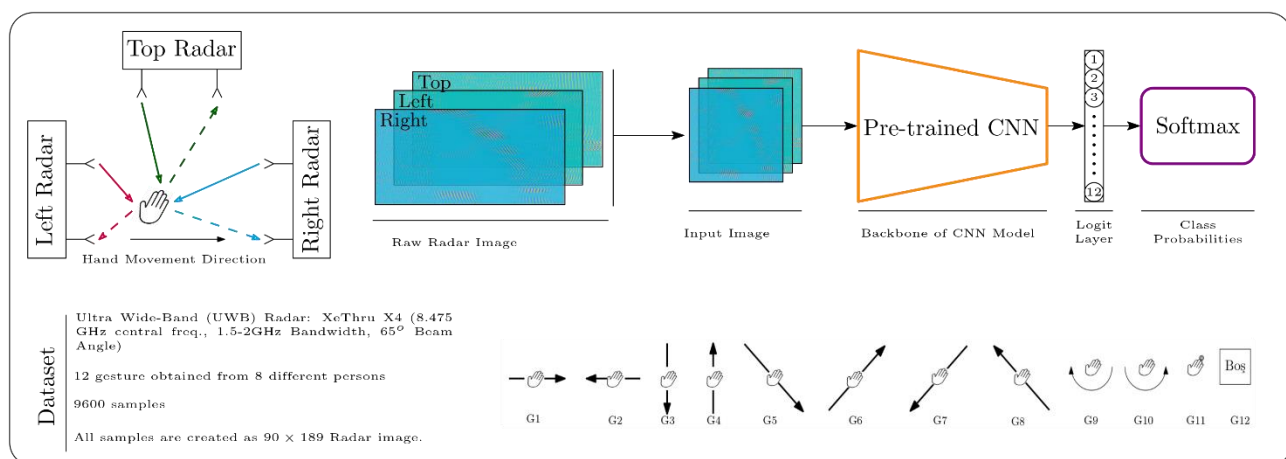
Nurullah CALIK[†]

Istanbul Medeniyet University, School of Engineering and Natural Sciences, Department of Biomedical Engineering, Istanbul, Turkiye.

Highlights

- Recognition of hand gesture with ultra-wideband radar sensors
- Converting data from 3 different sensors into a 3-channel image
- Conducting 5 random experiments for each models and training-test ratios of 50:50 and 75:25
- Achieving superior accuracy of 97.93% by training pre-trained models with fine tuning.

Graphical Abstract



Purpose and Scope

The aim of the paper is that the 3D images produced by combining images collected from different UWB sensors contribute to the recognition of hand gesture.

Design/methodology/approach

For this purpose, using the UWB hand gesture data set (Ahmed et al., 2021), radar images of 90×189 dimensions were gathered channel-wise and resized to 224×224 to feed the input of pre-trained models. Then the models were trained using the fine tuning methodology.

Findings

In the experimental setup section, each model was tested by separating random training and test samples 5 times for training and test ratios of 50:50 and 25:75. This process was also performed for single-channel images. It is seen that the proposed approach significantly increases the recognition performance achieved through a single channel.

Originality

The novel contribution of the article is the classification of UWB images as multi-channel instead of single channel and using pre-trained models for this purpose.

[†]Corresponding author: nurullah.calik@medeniyet.edu.tr, +90-216-280-3333

1. Giriş (Introduction)

Son yıllarda, insan hareketlerinin takibi alanında yaşanan teknolojik gelişmeler, sağlık, güvenlik ve bireysel yaşam kalitesi gibi önemli konularda çeşitli çözümler sunmaktadır. Özellikle görüntü işleme ve yapay zekâ teknolojilerinin entegrasyonu, bu alandaki potansiyeli büyük ölçüde artırmış ve insan hareketlerinin analizi konusunda yeni ufuklar açmıştır. Günümüzde, görüntü işleme teknikleri, kameralar aracılığıyla elde edilen video verilerini analiz ederek insan hareketlerini anlamak ve izlemek amacıyla yaygın bir şekilde kullanılmaktadır. Bu teknikler, gelişmiş algoritmalar sayesinde karmaşık ve dinamik hareketleri hassas bir şekilde tespit edebilmekte, bu durum çeşitli sektörlerde önemli uygulama alanları yaratmaktadır. Özellikle sağlık, güvenlik ve otomasyon alanlarında, insan hareketlerinin detaylı bir şekilde takip edilmesi, önceden belirlenmiş hedeflere ulaşma süreçlerini optimize etmekte ve toplumsal fayda sağlamaktadır. Sağlık sektöründe, yaşlı veya engelli bireylerin hareketlerini izlemek, onların günlük aktivitelerini değerlendirmek ve olası sağlık sorunlarını önceden belirlemek için görüntü işleme teknikleri oldukça etkili bir araçtır. Bu teknikler, kameralar aracılığıyla elde edilen video verilerini işleyerek, yaşlıların normal hareket örüntülerini belirlemekte ve anormal durumları tespit etmektedir. Örneğin, bir yaşlı bireyin düşmesi durumunda, görüntü işleme algoritmaları bu durumu tespit edebilir ve acil durum ekiplerine otomatik bir uyarı gönderebilir. Fakat görüntü işleme sistemlerinin mahremiyet konusunu egale etme potansiyeline sahip olmalarından dolayı, kişilerin evlerine yerleştirmekte tereddüt ettiği sistemlerdir. Bunun yerine ultra geniş-bant radar sistemleri yüksek frekans çözünürlüğü sayesinde insan hareketlerini hassas bir şekilde ayırt edebilmektedir. Dolayısıyla, kamera sistemleri yerine böylesi yenilikçi çözümlerin kullanılması daha elverişlidir.

Ultra geniş-bant (UGB) radar sensörleri, çok geniş bir frekans bandında çalışabilme özelliği olan radar sistemleridir. Bu sensörler, milimetrik ve santimetrik dalga boylarında yüksek çözünürlük ve hassasiyet sağlayarak, çeşitli nesnelerin, özellikle de insanların hareketlerini tanıma yeteneğine sahiptir (Ahmed ve Kallu vd., 2021). Gelişmiş işaret işleme algoritmaları ve veri analizi yöntemleri ile birleştirildiğinde, UGB radardan gelen çok sayıda veri aynı anda işlenebilmektedir. Bu sayede, insan hareketi tanıma konusunda daha doğru ve güvenilir sonuçlar elde edilebilmektedir. Yüksek çözünürlük ve düşük güç tüketimi avantajları, UGB radarlarının endüstriyel uygulamalarda da büyük ilgi görmesini sağlamaktadır (Jiang vd., 2011). Bu avantajlar, endüstriyel sistemlerin hassas ve etkili bir şekilde çalışmasını destekleyerek, üretim süreçlerindeki verimliliği artırmaktadır. Yüksek çözünürlük, radar sensörlerinin nesnelere arasındaki küçük farklılıkları dahi ayırt edebilme yeteneğini sağlar. Bu özellik, endüstriyel ortamlarda çok sayıda nesnenin ve bileşenin aynı anda takip edilmesinde, hataların hızlı bir şekilde tespit edilmesinde ve üretim süreçlerinin optimize edilmesinde kritik bir öneme taşır. Özellikle radar sinyallerinin toplanmasıyla elde edilen görüntülerin anonim bir yapıda olması nedeniyle kişilerin izlenmeden sadece hareketlerinin belirlenmesi UGB radar uygulamasının önemli bir avantajını oluşturmaktadır.

UGB sensörlerden elde edilen radar imgeleri RGB görüntüler kadar net bir bilgi sunmazlar. Sensörlerin hesaplayabildiği mesafe ve hız bilgileri üzerinden oluşturulan görüntüler ayrıca birçok arkaplan gürültüsünü de içermektedir. Bu gibi problemler ile başa çıkmak için son zamanlarda derin öğrenme yöntemlerinin kullanılması revaçtadır. Evrimsel sinir ağlarının (Convolutional Neural Network - CNN) kullanılmasıyla ham olan radar görüntülerinden sınıflandırma için etkili öznitelikler çıkarılabilmektedir. CNN'lerin görüntü çıkarmadaki kabiliyetleri sayesinde UGB görüntüleri de işlenebilmektedir. Bu görüntüler tek sensör kullanılarak elde edilebileceği gibi birden fazla sensör tarafından toplanan imgeler de olabilmektedir. Çoklu görüntülerin bir arada değerlendirilip sınıflandırma için kullanılması incelenmesi gereken kayda değer bir alandır.

Bu çalışma kapsamında, UGB radar görsellerinin sınıflandırılması için yeni bir yaklaşım öne sürülmektedir. Literatürde tek sensör veya üç kanal için ayrı model tasarımı gibi maliyetli çözümlere alternatif olarak sensörlerden gelen görüntüleri arka arkaya birleştirerek 3-kanallı bir görüntü elde edilmektedir. Ardından bu görüntü ön-eğitilmiş bir model üzerinden sınıflandırılmaktadır. 3-sensör kullanılarak elde edilen görüntü el hareketinin örüntüsünü içermektedir. RGB katmanlar yerine sağ, sol ve üst radardan elde edilen ölçüm değerleri eklenen bu yöntem dört farklı ön-eğitilmiş ağ üzerinde test edilmiş ve hepsinde diğer sensörlerden gelen sınıflandırmalara nazaran üstün bir performans sergilemiştir.

1.1. Kaynak Araştırması (Literature Survey)

UGB radar sensörleri özellikle kişi mahremiyetini koruyarak gözlem yapabileme imkânı vermesi nedeniyle üzerine yoğun çalışılan konulardan biridir. Saçılma verileri üzerinden elde edilen görseller, hareketin kendisini betimlemek için net bir bilgi sunmazlar. Arkaplan gürültülerinin de işin içine girmesiyle birlikte klasik öznitelik çıkartma ve sınıflandırma algoritmalarının baş edemeyeceği karmaşıklıkta bir problem ortaya çıkar. Bu nedenle, UGB radarlar için literatürde CNN tabanlı birçok çalışma yürütülmüştür.

Ahmed vd., 2021 çalışmasında UGB darbe radarı kullanılarak toplanan yeni bir el hareketleri veri kümesinin oluşturulmasını ve tanımlanmasını içermektedir. Veri seti, sekiz farklı insan gönüllüden toplanan 9.600 örnekten oluşmakta ve bu veriler hem .mat hem de .csv dosya formatlarında saklanmaktadır. Yalnızca sol radar sensöründen gelen girdiyi kullanan 4 katmanlı CNN modeliyle sınıflandırma doğruluğu, %94 seviyelerinde elde edilmiştir.

Mekruksavanich vd., 2023 çalışmasında (Ahmed vd., 2021)'de sunulan 12 el hareketini içeren veri seti kullanılarak bir model geliştirilmiştir. Modelin amacı, her 3 sensörden gelen veriyi kullanarak sınıflandırma yapan bir CNN modeli tasarlamaktır. Ortaya koyulan CNN mimarisi dört katmanlıdır. Makalede yazarlar 5-kat çapraz doğrulama (eğitim test oranı 80:20) %97.19 F1 skor değerine ulaşmışlardır.

Hendy vd., 2022'de ise yazarlar tek UGB radar sensörü kullanarak havaya yazılan sayıların sınıflandırmasına odaklanmışlardır. [0,9] arasında rakamların kullanıldığı veri seti 1800 örnekten oluşmaktadır. 1200 eğitim 300 validasyon ve 300 test örneği olarak ayrılmıştır. Makalede UGB sensörden gelen veriler görsel hale getirilmiş ve 3D CNN yapısına verilmiştir. Ayrıca çalışmada LSTM modelleri de kullanılarak başarı %98.5 seviyelerine çıkarılmıştır.

Park vd., 2023'te UGB sensör kullanılarak işaret dili tanıma uygulaması geliştirilmiştir. 50 cm uzaklıktan 5 farklı işaret için her işaretten 500 tane örnek toplanarak özgün bir veri seti oluşturulmuştur. Çalışmada yazarlar çift kanallı bir CNN yapısı geliştirmişlerdir. Ayrıca, doğruluğun artırılması için hem radar görüntülerinden hem de bu görüntülerin 2D Fourier dönüşümü olan görsellerden de faydalanmışlardır. Geliştirilen modellerin yanı sıra AlexNet (Krizhevsky vd., 2012), VGG-19 (Simonyan vd., 2014) ve ResNet-50 (He vd., 2016) gibi ön-eğitilmiş modeller de kullanılarak %98.98 oranında bir başarı elde etmişlerdir. Modellerin eğitimi için 70:30 oranı tercih edilmiştir.

Ahmed ve Cho, 2020 çalışmasında, Ahmed vd., 2021'da olduğu gibi 12 farklı hareketten üst ve sağ olmak üzere 2 radar kullanarak Xethru-X4 UGB sensörü ile veri toplamışlardır. Bu çalışmada yazarlar ön-eğitilmiş model kullanabilmek için iki görseli satır bazlı üst üste birleştirmişler, ardından 3 kanallı hale getirmek için gri seviye görüntülere MATLAB'ın "ind2rgb8" fonksiyonu uygulanmıştır. Ön-eğitilmiş model olarak GoogleNet (Szegedy vd., 2015) kullanılmıştır. Çalışmada iki sensörden alınan verinin birlikte işlenmesiyle %95 oranında bir başarının elde edildiği ortaya koyulmuştur.

Skaria tarafından kaleme alınan makalede (Skaria vd., 2020), derin öğrenme tekniklerinin radar sinyalleri üzerinden el jesti tanıma uygulamalarında önemli bir role sahip olduğu vurgulanmıştır. Çalışmada, ultra geniş bant (UGB) darbe radarından el jesti imzalarını sınıflandırmak için bir yöntem sunulmuştur. Bu çerçevede, 14 farklı el hareketinin işaretleri çıkarılmış ve her jestin bir menzil-Doppler çerçeve dizisi olarak temsil edildiği belirtilmiştir. Hareketlerden öznitelikleri çıkarmak için CNN modelleri kullanılmış ve ardından sınıflandırıcıya iletilmiştir. k-NN, SVM ve LSTM gibi farklı sınıflandırma mimarileri karşılaştırılmış ve sınıflandırma doğruluğunu maksimize etmek için tensör şekli ve sınıflandırıcı parametrelerinin optimize edildiği belirtilmiştir. Önerilen mimarilerin %96'nın üzerinde yüksek bir doğruluk seviyesine ulaşılmıştır.

Lai vd., 2023'te ise uyku pozisyonu izleme, obstrüktif uyku apne hastaları için önemli bir değerlendirme olarak belirtilmiştir. Bu çalışmanın amacı, çift ultra geniş bant radar sistemi kullanarak bir makine öğrenimi tabanlı uyku pozisyonu tanıma sistemi geliştirmektir. Makale kapsamında, dört uyku pozisyonunu (sırt üstü, sol ve sağ yan yatış ve eğimli) gerçekleştiren 16 hastadan gelen radyo frekansı verileri iki radar tarafından yatak üzerine ve yanına yerleştirilerek toplanmıştır. Önerilen ve değerlendirilen derin öğrenme yaklaşımları, öznitelik çıkarımı ve sınıflandırma süreçlerini optimize eden yöntemler içermekte olup, ayrıca farklı öznitelik çıkarıcılar ve sınıflandırıcıların kombinasyonlarını içeren geleneksel makine öğrenimi yaklaşımları da değerlendirilmiştir. Çalışmanın sonunda, çift radar sisteminin tek bir radardan daha iyi performans gösterdiğini ortaya koymuştur.

Ahmed vd., 2024'te yürütülen çalışmada, radar sensörlerinin temas ve invaziv olmayan doğası nedeniyle köpeklerde hayati belirtilerin izlenmesinde yaygın olarak kullanıldığı belirtilmektedir. Ancak, köpeklerden veri toplamanın özel eğitim ve onay gerektirmesi nedeniyle, köpeklerin hayati belirtilerine yönelik genel bir veri seti henüz bu zamana kadar önerilmemiştir. Ele alınan çalışmada, ilk kez ultra geniş bant radar tabanlı köpek hayati belirti (UWB-DVS) veri setini iki senaryo bağlamında sunulmuştur. İlk senaryoda, bayılmış köpeklere klinik referans sensörler takılmış, UGB radar ve referans sensörlerden veriler eşzamanlı olarak toplanmıştır. İkinci senaryoda ise köpekler serbestçe hareket edebilmektedir. Hareket tespiti ve solunum çıkarımı için referans olarak video kayıtları da toplanmıştır. Teknik doğrulama için, Senaryo 1'de hem kalp atış hızı hem de solunum hızı ölçümleri için radar ile klinik referans sensörler arasında yüksek bir korelasyon (0.9) bulunmuştur.

1.2. Makalenin Özgün Katkıları (Novel Contributions of the Paper)

Kaynak araştırmasında sunulduğu üzere, bu zamana kadar önerilen yöntemler ya bir yön üzerinden sınıflandırma yapmaya çalışmışlar (Ahmed vd., 2021) ya da 3 yönden çekilmiş radar görüntüleri için 3 girişi değerlendiren farklı model kurarak çalışmalarını yürütmüşlerdir (Mekruksavanich vd., 2023). Bu makalede önerilen çerçeve ise, her 3 harekete ait radar görüntüsü arka arkaya bir araya getirilerek 3-kanallı bir hareket görüntüsü elde edilmiş ve bu görsel doğrudan ön-eğitilmiş modellere verilmiştir. Çalışmada ele alınan diğer özellikler şu şekilde listelenebilir:

- Ön-eğitilmiş modeller kullanılarak az sayıda devir (epoch) ile model eğitimi yapılmıştır.
- Eğitim ve test oranında 50:50 ve 25:75 oranları kullanılarak zor durumlar altında bile önerilen yöntemin gürbüz (robust) çalıştığı sergilenmiştir.
- Öne sürülen yaklaşımın etkili çalıştığı için dört farklı ön-eğitilmiş model (GoogleNet, ResNet50, DenseNet201 (Huang vd., 2017) ve EfficientNetB0 (Tan vd. 2019)) kullanılmıştır.
- Bu modeller, rastgele ayrılmış 5 farklı eğitim ve test kümeleri üzerinden denenmiş ve bu denemelerin ortalamaları olarak sonuç tabloları oluşturulmuştur.

1.3. Çalışmanın Organizasyonu (Outline of the Paper)

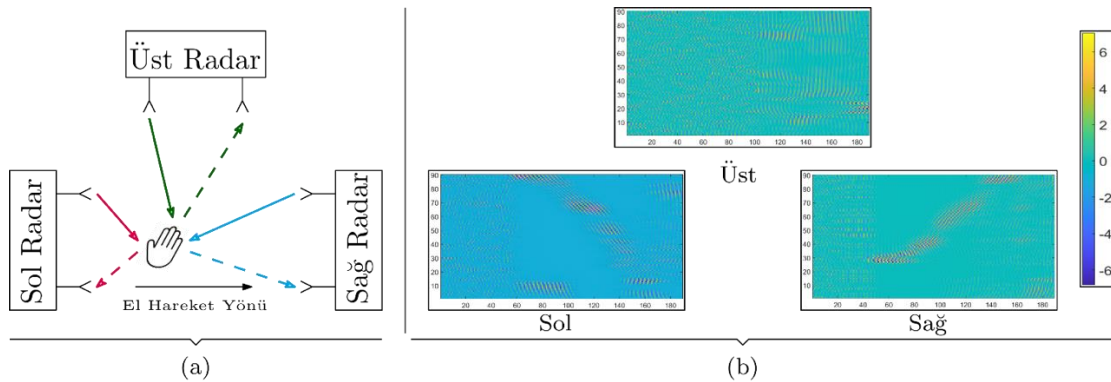
Makale toplam 4 bölümden oluşmaktadır. 2. bölüm (Materyal ve Yöntem) kapsamında çalışmada kullanılan materyal ve ortaya koyulan özgün yaklaşım sunulmuştur. 3 bölümde (DeneySEL Sonuçlar) ise modellerin karşılaştırmalı sonuçları sergilenmiş ve 4. bölümde burada ortaya çıkan bulguların tartışması yürütülmüştür.

2. Materyal ve Yöntem (Material and Method)

2.1. El Hareketi Tanıma İçin UGB Radar Veri Kümesi (UWB Radar Dataset for Hand Gesture Recognition)

El hareketlerinin tanınması için radar dalgaları ile toplanan veri seti genellikle tek sensör üzerindedir. (Ahmed vd., 2021)'deki çalışmada çoklu sensör kullanımı ile üretilen veri seti bu makalenin materyal tarafını oluşturmaktadır.

Ahmed vd., 2021'deki çalışmada, 3 adet UGB radar kullanılarak Şekil 1(a)'da verilen düzenek oluşturulmuştur. Sensör olarak, 8.745 GHz merkez frekansına ve 1.5 GHz-2 GHz bant genişliğine sahip olan XeThru X4 tercih edilmiştir. Sensörün kendi üzerinde 23 GHz'lik bir örnekleme frekansı da bulunmaktadır. Saniyede 20 çerçeve (frame) gönderim kapasitesine ve 65° bir hüzme genişliğine sahiptir. Üzerinde bir alıcı ve bir verici anten bulunmaktadır. Üç radarın arasında 12 farklı en hareketi icra edilmiş ve buradan örnekler toplanmıştır. Her el hareketinin kendisine özgü karakteristiği, radar sensörleri tarafından algılanarak Şekil 1(b)'de verilen radar görüntülerini oluşturmaktadır. Bu görseller üzerinde oluşan dalgalanmalar hareketin örüntüsünü ortaya koyar. Her bir yön hareketin farklı bir bileşenini oluşturmaktadır.



Şekil 1. (a) El hareketleri veri setinin toplanması. (b) Radar saçılım görüntülerinin oluşturulması ((a) Collection of hand gestures dataset. (b) Creation of radar scatter images)

UGB radar sistemi geniş bir frekans spektrumuna sahip bir sinyali çok kısa bir süre boyunca iletebilmektedir. Gönderilen darbe benzeri işaretlere karşılık alınan sinyal Denklem (1)'de verilmiştir.

$$x[n, k] = \sum_{j=1}^{\Phi_{yol}} A_{nj} \times s[n, k - T_j] + G \quad (1)$$

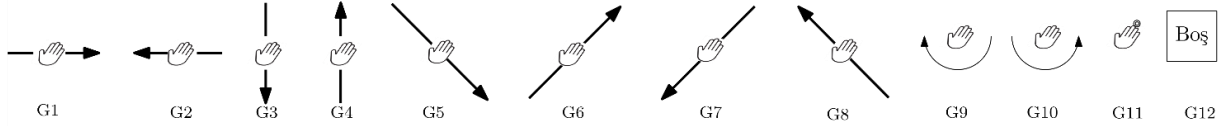
Burada $x[n, k]$ anten tarafından alınan Φ farklı yol üzerinden gelmiş ve üzerine G gürültüsü eklenmiş işareti temsil etmektedir. $x[n, k]$ işareti üzerinden ilgili indislere (n, k) ait bir sayısal değer elde edilmektedir. $n \in [1, N]$ ve $k \in [1, K]$ olan bir aralık için tarama yapılmaktadır ve Denklem (2)'de ifade edilen matris yapısı elde edilmektedir.

$$X = \begin{bmatrix} x[N, 1] & \dots & x[N, K] \\ \vdots & \ddots & \vdots \\ x[1, 1] & \dots & x[1, K] \end{bmatrix} \quad (2)$$

X matrisinde satırlar yavaş zaman (slow time), sütunlar ise hızlı zaman (fast time) durumlarını göstermektedir. İlgili veri kümesi oluşturulurken $N = 90$ ve $K = 189$ olarak seçilmiştir. Her hareketin sonunda üç radarda Şekil.1(b)'de verilen görseller oluşmaktadır. Radar işareti toplanırken çevresel saçılmalardan (clutter) kaynaklı bozucu arkaplan gürültüsü de meydana gelir. Gürültünün giderilmesi için geri-döngü (loopback) filtresi olarak adlandırılan Denklem (3)'te verilen filtre kullanılmıştır. Bu filtreden geçen görüntüler veri kümesinde "ClutterRemoved" klasörüne ham verinin haricinde ayrıca kaydedilmiştir. Bu çalışma kapsamında, temizlenmiş görseller üzerinden deneyler yapılmıştır.

$$z_n[k] = \alpha z_n[k-1] + (1 - \alpha)x_n[k] \quad (3)$$

12 farklı hareketin bulunduğu veri kümesinin oluşturulması için farklı yaş ve kilolarda toplam 8 farklı kişiden veriler toplanmıştır. Her gönüllüden bir hareket için toplam 100 örnek alınmıştır. Bu bağlamda, her sınıfta 800 örnek, toplamda 9600 örnekten oluşan bir veri kümesi ortaya çıkmıştır. 12 sınıfa ait görseller Şekil.2'de sunulmuştur. Hareket kümesinin içerisinde, sağ-sol ve yukarı-aşağı doğrusal hareketlerden 4 tane (G1-G4), çapraz (diagonal) hareketlerden 4 tane (G5-G8) ve saat yönünde ve tersinde olmak üzere 2 tane de dönme hareketi (G9-G10) bulunmaktadır. Ayrıca elin içe doğru gelmesi (G11) ile birlikte boşluğun tespit edilmesi için hareketsizlik durumu da (G12) ek birer sınıf olarak ele alınmıştır.



Şekil 2. Veri setinde bulunan 12 el hareketi (12 hand gestures included in the dataset)

Veri seti içinde her gönüllü için ayrı bir klasör bulunmaktadır. Bu klasörün altında da "ClutterRemovedData" ve "RawData" isimlerine alt klasörler mevcuttur. Veriler temizlenmiş ve ham olan klasörlerin altında ".mat" formatında tutulmaktadır. Her ".mat" dosyası kaydedilirken "HV_<Gönüllü ID>_<Hareket ID>_ClutterRemoved.mat" isimlendirilmektedir.

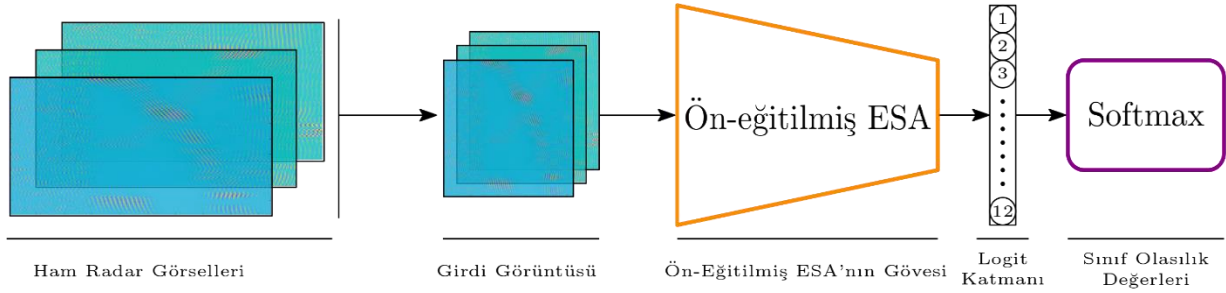
2.2. 3-Sensör Hareket Görüntüsünün Sınıflandırılması (Classification of 3-Sensor Gesture Images)

Derin öğrenme tekniklerini kullanarak eğitilen ön-eğitilmiş (pre-trained) Evrişimli Sinir Ağı (Convolutional Neural Network-CNN) modelleri, bilgisayarlı görü, nesne tanıma ve sınıflandırma gibi görevlerde geniş bir uygulama alanı bulmuştur (Chen vd., 2021; Yuan vd., 2021; Qiu vd., 2020). Son yıllarda yapılan çalışmalarda sadece bilgisayarlı görü alanı değil aynı zamanda farklı disiplinlerde de kendisini göstermektedir (Kozziel vd., 2022; Mahouti vd. 2022). Bu modeller, önceden büyük veri kümesi üzerinde eğitilmiş ve genel özellikleri öğrenmiş olduğu için, çeşitli görevlerde daha yüksek performans ve verimlilik sağlama potansiyeline sahiptir. Ön-eğitilmiş CNN modellerini, büyük veri kümesindeki eğitiminden öğrendiği genelleme kabiliyeti sayesinde başka veri kümelerine olan adaptasyon daha hızlı ve verimli bir şekilde gerçekleşir (Bouaafia vd., 2021; Han vd., 2021). Ayrıca, genel özelliklere odaklanan bu modeller, özel bir görev için daha az eğitim verisine ihtiyaç duyabilir ve böylece eğitim maliyeti daha az olabilmektedir. Modelin küçük veri setlerine uyarlanması için gerekli olan devir (epoch) sayısı büyük veri setlerine göre daha azdır.

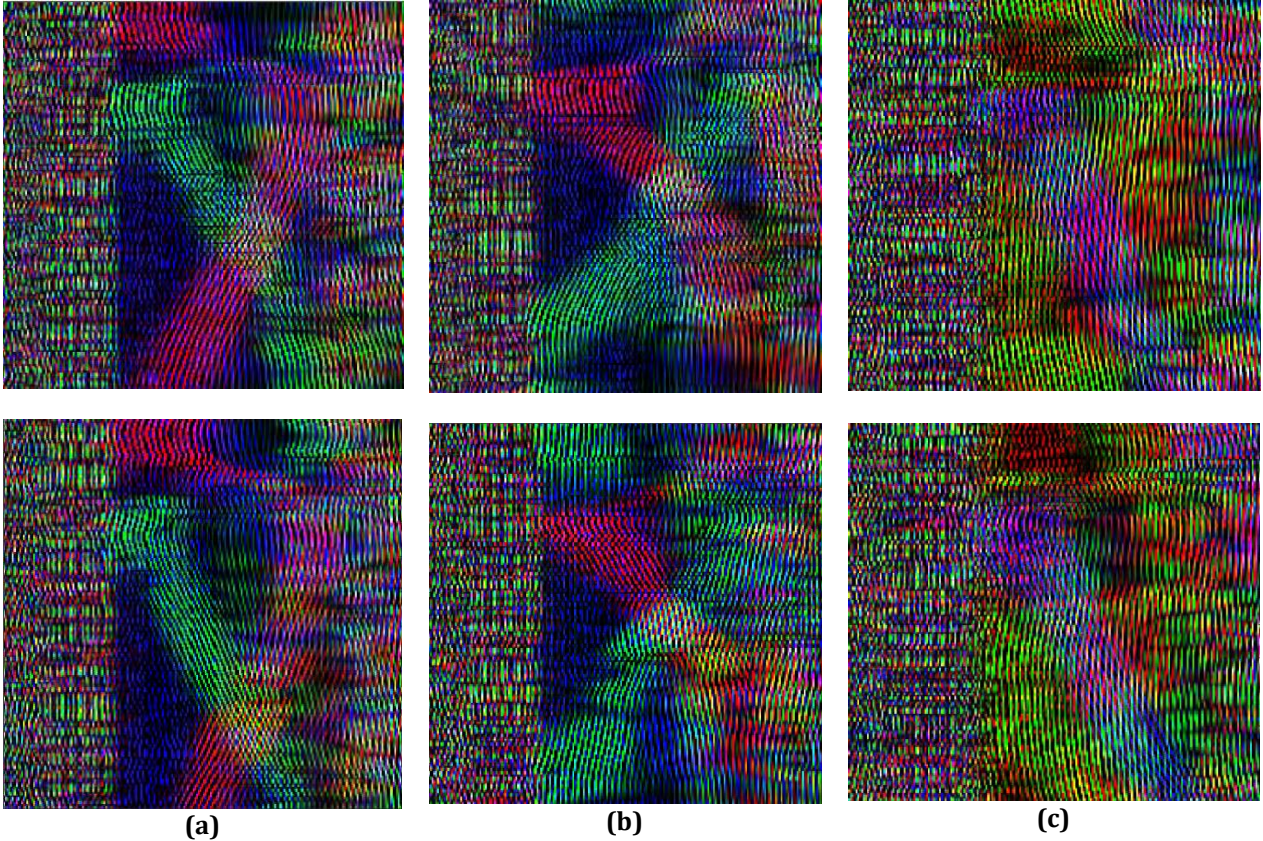
Ön-eğitilmiş modellerin az sayıda örnek içeren veri kümeleri ile eğitiminde transfer öğrenimi (transfer learning) ve ince ayar (fine tuning) (Ding vd., 2023) eğitim stratejileri vardır. Bu çalışmada ince ayar eğitim stratejisi kullanılmıştır. Bu yöntemde, modelin son katmanı olan ve sınıf skorlarının üretildiği nöronların bulunduğu logit katmanı temel modelden çıkarılarak yerine yeni veri kümesinin sınıf sayısı kadar bir nörona sahip logit katmanı eklenmektedir. Bu katman tam-bağlı (fully-connected) olarak modelin içinde bulunmaktadır. Sonrasında, yeni model düşük bir öğrenme oranı ile yeni görev için eğitilir. Öğrenme oranının düşük olmasının sebebi, temel modeldeki genelleme yeteneğinin ilgili veri kümesi için sadece adapte olmasının istenmesidir. Katsayılarına rastgele değerler atanmış bir CNN yapısı tercih edilseydi öğrenme oranının yüksek tutulması gerekirdi. Ön-eğitilmiş modeller incelendiğinde farklı mimariye sahip birçok CNN yapısı ile karşılaşılmaktadır. Her bir modeli özgün kılan taraf, katmanlar arasındaki bağlantı yapıları ve katmanlar için kullanılan farklı aktivasyon

fonksiyonları, konvolüsyon çeşitleri ve havuzlama (pooling) katmanındaki değişikliklerdir. Örnek vermek gerekirse, GoogleNet (Szegedy vd., 2015) ilk defa "inception" yapısını ortaya atmıştır. Bu katman kendisine gelen tensör verisine farklı boyutlardaki filtreleri paralel olarak uygular ve elde edilen aktivasyon haritalarını kanal bazında bir araya getirerek bir sonraki katmana aktarır. Buradaki amaç, hangi filtre çıkışının daha anlamlı olduğuna modelin kendisinin karar vermesidir. Bir diğer ön-egitilmiş model yapısı ise ResNet (He vd., 2016) mimarisidir. Bu model, diğerlerinden farklı olarak atlamalı bağlantı yapısı ile tensör akışını tek bir kanal üzerinden yapılmaya zorlamaz. t . katman kendi içinde $F(x^t)$ ile x^t tensör verisini işler ve bir sonraki katmana $x^{t+1} = F(x^t) + x^t$ olarak veriyi sunar. Eğer model içerisinde ilgili katman gereksiz ise $F(\cdot)$ içinde bulunan parametreler sıfıra doğru gitmektedir. Bu sayede gradyanların sönümlenmesi probleminin önüne geçilir. Diğer birçok ön-egitilmiş modellerde kendilerine has mimari yapıları ile çeşitli görevlerde farklı başarımlar sergilerler.

Ele alınan bu çalışmada, el hareketi tanıma için UGB radardan toplanan görüntülerin sınıflandırılması için Şekil 3'te gösterilen yeni bir eğitim modeli önerilmiştir. Ortaya koyulan bu yapıda, sağ, sol ve üst sensörden edinilen verilerin ($90 \times 189 \times 1$) ayrı ayrı sınıflandırılması yerine, ön-egitilmiş modelin girişine uygun olarak şekillendirildikten sonra kanal ekseninde bir araya getirilerek modele ($90 \times 189 \times 3$) boyutlarında verilmiştir. 3-sensör görüntüsü oluşturulmadan önce her görsel kendi içinde normalize edilmektedir. Aksi takdirde, farklı radarlardan elde edilen değerler kendi aralarında ölçekleme problemine sebebiyet verecektir. Bu durumun önüne geçmek için MATLAB'ın "normalize" komutu kullanılmıştır. Ardından, görüntüler ön-egitilmiş modelin girişine uygun olması için $224 \times 224 \times 3$ boyutlarına getirilmiştir.



Şekil 3. 3 sensör üzerinden alınan görüntülerin ön-egitilmiş modeller üzerinden işlenmesi (Processing of images taken from 3 sensors through pre-trained models)



Şekil 4. 3-sensör üzerinden kanal boyunca bir araya getirilen görüntüler (Images combined across the channel via 3-sensors)

Logit katmanı için veri setinde bulunan 12 sınıf sebebiyle 12 nöronlu bir yapı tercih edilmiştir. Buraya eklenen nöronlara atanan değerler rastgele olduğu için öğrenme çarpan oranı (learning rate) CNN içindeki katmanlar için 1 iken bu katman için çarpan 10 olarak seçilmiştir. Eğitim için toplamda 5 devir kullanılmıştır. Mini-yığın (mini-batch) miktarı 32 olarak tercih edilmiştir. Öğrenme oranı 0.001'den başlatılarak, modeller Adam (Kingma ve Ba, 2014) uyarlayıcısı ile eğitilmiştir. Her 2 devirde öğrenme oranı 0.65 katsayısı ile çarpılarak kademeli olarak azaltılmıştır.

Ön-eğitilmiş modellerin hareketleri algılaması Şekil.4'te gösterilmektedir. Farklı sensörlerden gelen görseller bir araya getirildiğinde her ne kadar karmaşık örüntüler oluştursalar da CNN'ler tarafından tanınabilecek desenlere sahiptirler. Şekil.4(a)(b) ve (c) G1, G2 ve G3 hareketlerinden verilen ikişer örnek içermektedir. Görüntülerin bir araya getirilme sırası sağ, sol ve üst olduğu için kırmızı (R), yeşil (G) ve mavi (B) renkleri ilgili sensörlere karşılık gelmektedir. CNN mimari ilk katmanlarda $3 \times 3 \times 3$ veya $5 \times 5 \times 3$ boyutlarında çekirdekler (kernel) kullanarak bu ham bilgiyi işlerler ve edindikleri örtüntü bilgisini damıtarak sınıflandırma bloguna aktarırlar. Bunları yaparken modeller ayrıca görsellerin sağ ve sol tarafında oluşan kargaşadan da etkilenmemeleri gerekmektedir. Ön-eğitilmiş modeller bu zorlukların üstesinden gelerek üç bilgi kullanımıyla üstün bir başarıyı sergilemektedir.

Her ne kadar Şekil 4'teki görsellerden göz ile belli bir örüntü çıkarılamasa bile CNN modelleri sahip olduğu karmaşık ilişki yakalama kabiliyetleri neticesinde sınıf içi benzerlikleri üç boyutlu uzayda renk uzayında elde edebilmektedirler. Buradan damıtılan düşük seviyeli bilgi, tüm-bağlı (fully-connected) katmana aktarılır. Saklı uzayda (latent space) oluşan öznitelik vektörü hareketlerin sınıflandırılması için gerekli yapılandırılmış değerleri barındırmaktadır.

3. Deneysel Sonuçlar (Experimental Results)

Önerilen yöntemin performansının gösterilmesi için 4 farklı ön-eğitilmiş model kullanılmıştır. Karşılaştırma için bu modeller her üç sensörden gelen veriler (tek kanallı) üzerinden ayrı ayrı eğitilerek test edilmiş. Ön-eğitilmiş modellere tek kanal görsellerinin verilmesi için bunların da 3 kanal haline getirilmesi gerekmektedir. Bunun için ele alınan görüntü çoğullanarak 3-kanal formatına dönüştürülmüştür.

Eğitim ve testler 50:50 ve 25:75 eğitim ve test oranları üzerinden 5 rastgele denemenin ortalaması olarak yapılmıştır. Her rastgele denemede modellerin aynı eğitim ve test örnekleri üzerinden çalışması için özen gösterilmiştir. Tablo.1 ve Tablo.2'de verilen değerler bu 5 denemenin ortalaması ve standart sapması üzerinden verilmiştir. 25:75 oranının denemesinin sebebi, zor şartlar altında modellerin gürbüzlüğünün test edilmesidir. Değerlendirmeler doğruluk (accuracy) ve makro F-Skor üzerinden yapılmıştır.

Tablo 1. Ön-eğitilmiş modellerin farklı sensörler üzerinden test edilmesi. Eğitim-test oranı: 50-50. Deneme sayısı: 5 (Testing pre-trained models on different sensors. Training-test ratio: 50-50. Number of trials: 5)

Sensör	Modeller							
	GoogleNet		ResNet50		DenseNet201		EfficientNetB0	
	Doğruluk	F-Skor	Doğruluk	F-Skor	Doğruluk	F-Skor	Doğruluk	F-Skor
Sağ	89.39±0.83	89.39±0.82	95.30±0.35	95.30±0.36	96.14±0.30	96.14±0.31	95.67±0.09	95.66±0.09
Sol	88.36±0.96	88.28±1.07	96.64±0.26	96.64±0.26	97.24±0.13	97.21±0.14	96.89±0.24	96.89±0.24
Üst	82.40±2.72	82.29±2.74	94.09±0.32	94.10±0.32	94.47±0.87	94.48±0.88	94.22±0.43	94.25±0.46
3-Sensör*	95.28±0.61	95.29±0.61	97.83±0.18	97.84±0.16	97.93±0.14	97.93±0.14	97.89±0.14	97.90±0.14

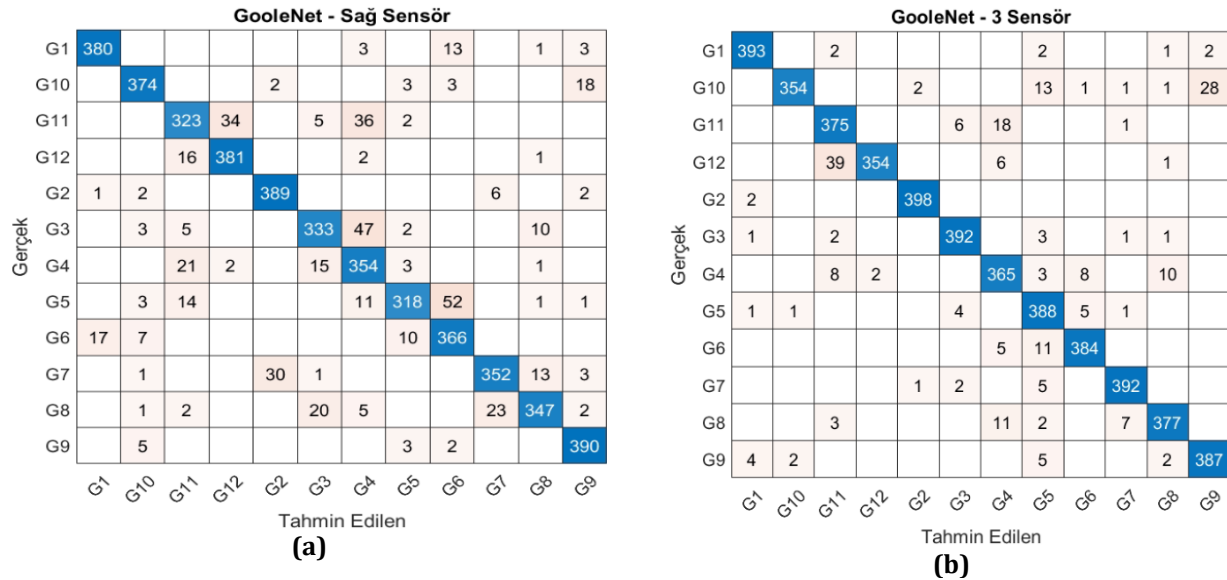
Tablo 1'de verilen sonuçlar 50:50 eğitim test oranı altında ortaya çıkan değerlerdir. Her dört modele önerilen yöntemin uygulanması diğer kanallardan yapılan sınıflandırmalara göre önemli bir iyileştirme sağlamıştır. GoogleNet modeller arasında en düşük performansa sahipken en yüksek değer DenseNet201'de ortaya çıkmıştır. Diğer kayda değer durum ise modellerin standart sapma oranlarıdır. Ortaya koyulan tanıma başarısı sadece bir denemenin değil, 5 ayrı eğitim-test süreci üzerinden olduğu için standart sapma, modellerin gürbüzlüğü için önemli bir göstergedir. 3 sensör bilgisi kullanılarak yapılan çalışmada, sapmanın diğer kanallar üzerinden yapılan sınıflandırmalara nazaran daha düşük olması diğer bir önemli bulgudur.

Tablo 2. Modellerin 25:75 eğitim ve test ayırım oranında gösterdikleri başarımlar. En iyi skor * ile gösterilmiştir. (Performances of the models at 25:75 training and test ratio. The best score is indicated by “*”.)

Sensör	Modeller							
	GooleNet		ResNet50		DenseNet201		EfficientNetB0	
	Doğruluk	F-Skor	Doğruluk	F-Skor	Doğruluk	F-Skor	Doğruluk	F-Skor
Sağ	81.20±3.99	81.33±3.94	93.19±0.36	93.21±0.37	93.24±0.14	93.25±0.15	93.48±0.14	93.49±0.14
Sol	78.97±0.96	78.83±0.76	94.98±0.42	94.96±0.46	95.38±0.32	95.33±0.30	95.71±0.23	95.72±0.23
Üst	76.03±3.96	75.58±4.17	90.18±0.51	90.19±0.53	90.35±0.59	90.36±0.59	90.49±0.22	90.50±0.22
3-Sensör*	90.08±2.06	90.11±2.08	97.14±0.25	97.12±0.26	97.55±0.27	97.55±0.22	97.65±0.21	97.66±0.22

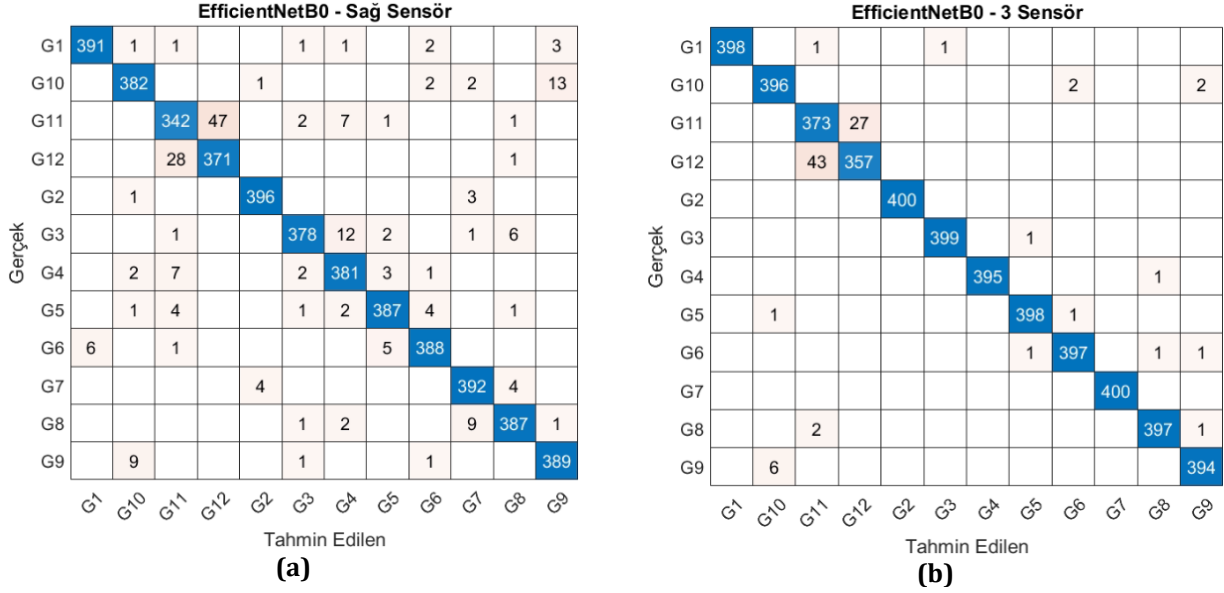
Tablo 2’de verilen değer ise 25:75 oranı çerçevesinde elde edilen başarımlardır. Eğitim verisi düştüğünde özellikle üst sensörün başarımlarında %4'lere diğer modellerde ise %2 civarında düşüşler yaşanmıştır. Özellikle GooleNet üzerinden yapılan denemelerde tüm yöntemlerde dramatik düşüşler görülmektedir. Fakat bu durum model mimarisinin kendisinden ileri gelmektedir. 3-sensör yaklaşımının diğer modeller altında elde ettiği değerler 50:50 durumuna nazaran kayıplar %0.5 civarlarındadır. Dolayısıyla, önerilen yaklaşımın düşük eğitim oranlarında bile üstün bir başarımlar sergilediği ortaya koyulmuştur.

Şekil 5 ve Şekil 6, 3-sensör yaklaşımının hangi sınıflarda iyileştirmeye katkı sağladığı görülmektedir. 50:50 oranlar için gösterilen karışım matrislerinde (confusion matrix) GooleNet ve EfficientNetB0 modelleri tercih edilmiştir. Şekil 5(a)'da GooleNet'in sağ sensörden gelen görseller üzerinden yapılan sınıflandırma başarımları görülmektedir. Model, özellikle G11-G12, G3-G4 ve G5-G6 sınıflarında karışım yaşamaktadır. 3-sensör görüntüsü kullanıldığında ilgili sınıflardaki başarımlar önemli ölçüde arttığı için doğruluk ve F-skor metriklerinde bir yükselme görülmektedir.



Şekil 5. GooleNet üzerine önerilen yöntemin ortaya çıkardığı katkı. (a) Sadece sağ sensörün kullanımı (b) Her üç sensörün kullanımı. (Contribution of the proposed method over GooleNet. (a) Use of right sensor only (b) Combination of all three sensors)

EfficientNetB0 modeline ait karışım matrisi ise Şekil 6'da verilmiştir. Şekil 6(a) ile Şekil 5(a) karşılaştırıldığında EfficientNetB0 mimarisinin GooleNet'e olan üstünlüğü görülmektedir. Şekil 6(b) incelendiğinde ise 3-sensör görüntü sınıflandırmanın edindiği iyileştirme gözler önüne serilmiştir. Şu not edilmelidir ki, özellikle modeller G11 ve G12 sınıflarını sınıflandırmada zorlanmaktadır. Şekil 6(b)'de neredeyse tüm sınıflarda önemli derecede bir iyileştirme söz konusu iken G11 ve G12 sınıflarında aynı oranda bir iyileştirme yakalanamamıştır.



Şekil 6. EfficientNetB0 modeli üzerinden 3 sensör yaklaşımının karışım matrisleri üzerinden değerlendirilmesi (Evaluation of 3 sensor approaches through confusion matrices over the EfficientNetB0 model)

4. Sonuç ve Tartışma (Result and Discussion)

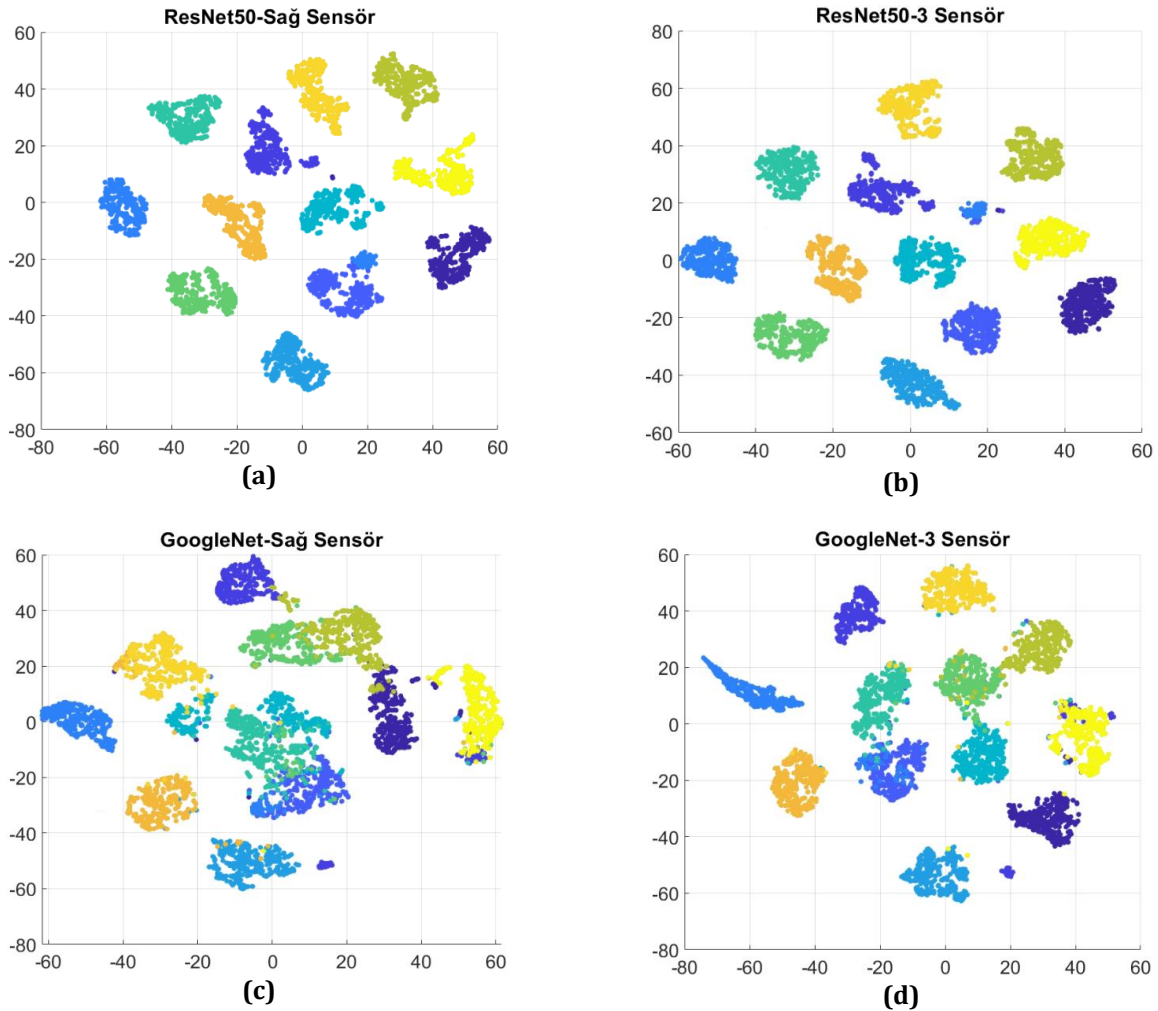
Ön-ēitilmiş modellere kanal bazlı birleştirme yöntemiyle verilen UGB radar görselleri üzerinden yüksek doğruluk oranıyla sınıflandırma yapılabildiği gösterilmiştir. Farklı sensörlerden gelen bilgiler el hareketinin tanınmasına olanak kılan örüntüyü barındırmaktadır. Sensörler tekil olarak ele alındığında belli bir başarımın üzerine çıkamazlarken bir arada kullanılması hem doğruluğu arttırmış hem de daha kararlı sonuçlar elde edilmesini sağlamıştır.

Modellerin başarımının neden arttığına dair önemli bir araştırma olarak Şekil.7 verilmiştir. CNN modellerinin amacı girdi olarak verilen görüntülerin olabildiğince sınıflar arası varyansın yüksek sınıf içi varyansın düşük olduğu bir vektör uzayına haritalandırmaktır. Sınıf skorlarının oluştuğu logit katmanından önce modellerde sıklıkla Global Ortalama Havuzlama (Global Average Pooling) katmanı bulunmaktadır. Bunun amacı, kendisine gelen $N \times M \times C$ boyutlarında tensör yapısını, kanal bazında ortalama olarak $1 \times 1 \times C$ boyutlarında bir vektöre dönüştürmektir. Bu vektör C boyutlu uzayda girdi görüntüsünü temsil etmektedir. Bu bağlamda, önerilen yaklaşımın neden daha iyi çalıştığı, eğitim verileri üzerinden elde edilen bu öznelik vektörlerinin tSNE algoritması ile görselleştirilmesi ile tartışılmıştır.

Şekil 7(a)(b) ve Şekil.7(b) ResNet50 modeli için tSNE (Van der Maaten ve Hinton, 2008) grafikleridir. tSNE algoritması, veri setinin dağılım gösterdiği uzaydaki (C uzayı) örnekler arasındaki komşuluk metriklerini koruyabilecek bir D boyutlu uzay arar. Eğer iki örnek C boyutlu uzayda uzaksa, haritalandırılmış D boyutlu uzayda da uzaktır. Eğer D 2 veya 3 seçilirse, verinin kendi dağılımını sergilediği C boyutlu uzaydaki manifoldu gözler önüne serilmiş olur. Sınıfların C boyutlu öznelik uzayında nasıl bir dağılıma sahip olduklarını gösteren bu grafikler 3 sensör verisinin kullanılmasının ne gibi bir katkı sunduğunu ön plana çıkarmaktadır. Şekil 7(b) incelendiğinde önerilen yöntem ile eğitilen modellerin sınıfları daha bütünleşik olarak bir arada tuttuğu görülmektedir.

Özellikle bu durum sarı, mor ve turuncu sınıflar için net bir şekilde görülmektedir. Şekil.7(c)(d) ise 3 sensör yaklaşımının GoogleNet'e eklediği başarımın kaynağı incelenmiştir. Yine bu model için de Şekil.7(d) incelendiğinde birleştirilmiş görüntü üzerinden oluşturulan öznelik uzayında sınıfların kendi içinde daha bir arada olduğu incelenebilmektedir.

Sonuç olarak, bu çalışmada UGB sensörler üzerinden el hareketlerinin sınıflandırması yapılırken her sensörden gelen verilerin bir arada değerlendirilmesinin daha avantajlı olacağı ortaya koyulmuştur. Özellikle ön-ēitilmiş ağların sağladığı güçlü modelleme yeteneği sayesinde az sayıda eğitim adımı ile problemin üzerinden yüksek doğruluk oranlarında bir başarımın elde edileceği incelenmiştir.



Şekil 7. tSNE algoritması kullanılarak GoogleNet ve ResNet50 üzerinden elde edilen öz niteliklerin görselleştirilmesi (Visualization of features obtained via GoogleNet and ResNet50 using the tSNE algorithm)

Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the author.

Kaynaklar (References)

- Ahmed, S., & Cho, S. H. (2020). Hand gesture recognition using an IR-UWB radar with an inception module-based classifier. *Sensors*, 20(2), 564.
- Ahmed, S., Kallu, K. D., Ahmed, S., & Cho, S. H. (2021). Hand gestures recognition using radar sensors for human-computer-interaction: A review. *Remote Sensing*, 13(3), 527.
- Ahmed, S., Wang, D., Park, J., & Cho, S. H. (2021). UWB-gestures, a public dataset of dynamic hand gestures acquired using impulse radar sensors. *Scientific Data*, 8(1), 102.
- Ahmed, S., Yoon, S., & Cho, S. H. (2024). A public dataset of dogs vital signs recorded with ultra wideband radar and reference sensors. *Scientific Data*, 11(1), 107.
- Bouaafia, S., Messaoud, S., Maraoui, A., Ammari, A. C., Khriji, L., & Machhout, M. (2021, March). Deep pre-trained models for computer vision applications: traffic sign recognition. In *2021 18th International Multi-Conference on Systems, Signals & Devices (SSD)* (pp. 23-28). IEEE.
- Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., ... & Gao, W. (2021). Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12299-12310).
- Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., ... & Sun, M. (2023). Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 5(3), 220-235.
- Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., ... & Zhu, J. (2021). Pre-trained models: Past, present and future. *AI Open*, 2, 225-250.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hendy, N., Fayek, H. M., & Al-Hourani, A. (2022). Deep Learning Approaches for Air-Writing Using Single UWB Radar. *IEEE Sensors Journal*, 22(12), 11989-12001.

- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
- Jiang, S., Skibniewski, M. J., Yuan, Y., Sun, C., & Lu, Y. (2011). Ultra-wide band applications in industry: a critical review. *Journal of Civil Engineering and Management*, 17(3), 437-444.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Koziel, S., Çalık, N., Mahouti, P., & Belen, M. A. (2022). Reliable computationally efficient behavioral modeling of microwave passives using deep learning surrogates in confined domains. *IEEE Transactions on Microwave Theory and Techniques*, 71(3), 956-968.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lai, D.K., Zha, L., Leung, T.Y., Tam, A.Y., So, B.P., Lim, H., Cheung, D.S., Wong, D.W., & Cheung, J.S. (2023). Dual ultra-wideband (UWB) radar-based sleep posture recognition system: Towards ubiquitous sleep monitoring. *Engineered Regeneration*.
- Mahouti, P., Belen, M. A., Çalık, N., & Koziel, S. (2022). Computationally efficient surrogate-assisted design of pyramidal-shaped 3-D reflectarray antennas. *IEEE Transactions on Antennas and Propagation*, 70(11), 10777-10786.
- Mekruksavanich, S., Jantawong, P., Tanchaen, D., & Jitpattanakul, A. (2023, June). A Convolutional Neural Network for Ultra-Wideband Radar-Based Hand Gesture Recognition. In *2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC)* (pp. 1-4). IEEE.
- Park, G., Chandrasegar, V. K., & Koh, J. (2023). Accuracy Enhancement of Hand Gesture Recognition Using CNN. *IEEE Access*, 11, 26496-26501.
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10), 1872-1897.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Skaria, S., Al-Hourani, A., & Evans, R.J. (2020). Deep-Learning Methods for Hand-Gesture Recognition Using Ultra-Wideband Radar. *IEEE Access*, 8, 203580-203590.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- Yuan, L., Chen, D., Chen, Y. L., Codella, N., Dai, X., Gao, J., ... & Zhang, P. (2021). Florence: A new foundation model for computer vision. arXiv preprint arXiv:2111.11432.