

## Basketbol Hakemi El İşaretlerinin Nesne Tanıma Yöntemleri Kullanılarak Sınıflandırılması

Ozan Biçer<sup>1\*</sup>, Cemil Zalluhoğlu<sup>2</sup>

<sup>1</sup>Veri ve Bilgi Mühendisliği, Bilişim Enstitü, Hacettepe Üniversitesi, Türkiye

<sup>2</sup>Bilgisayar Mühendisliği Bölümü, Hacettepe Üniversitesi, Türkiye

\*ozanbicer@hacettepe.edu.tr

**Özet** – Teknolojinin gelişmesiyle birlikte, yapay zekâ ve bilgisayar görüşü teknikleri spor alanında da insan hayatına girmeye başladı. Günümüzde basketbol hakemleri, maç içinde verdikleri kararları masa hakemlerine el işaretleri ile gösterir ve masa hakemleri bu doğrultuda bazı verileri skor tabelasına işler. Bu işlem insan kaynaklı hatalar sebebiyle bazen aksamaktadır ve maçın duraklamasına sebep olmaktadır. Bu çalışmada, bilgisayar görüşü teknolojisi ile bu probleme bir çözüm önerilmektedir. Bu amaçla basketbol hakemlerinin el işaretleri için bir veri seti elde edilmiştir. Bu veri seti üzerinden iki adet derin öğrenme modeli eğitilmiştir ve başarı sonuçları paylaşılmıştır. Gerçekleştirilen çalışma, bahsedilen sürecin yapay zekâ desteği ile geliştirilebileceği ve sadece basketbol değil, tüm spor dallarına sağlayabileceği potansiyel katkıları göstermektedir.

**Anahtar Kelimeler** – Derin Öğrenme, Bilgisayar Görüşü, Evrimsel Sinir Ağı, YOLO, Tensorflow

### I. GİRİŞ

Nesne tanıma, görüntü işleme alanında kullanılan bir yöntemdir. Bu yöntemde görüntüdeki nesnelerin pozisyonları belirlenir ve nesne tanımlanır. Nesne tanıma sistemleri genellikle derin öğrenme ve evrimsel sinir ağları tekniklerinden yararlanır ve haar-cascade, R-CNN, YOLO ve SSD gibi algoritmalar kullanır. Nesne tanıma teknolojisi günümüzde birçok alanda kullanılmaktadır. Öncelikli olarak güvenlik kameraları ve otomatik kamera sistemleri gibi alanlarda kullanılmaktadır. Bu sistemler insan ve araba trafiğini takip etmek, radar sistemlerinde araçların plakasını belirlemek için tercih edilebilir. Ayrıca nesne tanıma sistemleri, robotik, drone ve otonom sürüslü araç alanında da kendine yer edinmeye başlamıştır. Bu gelişmelerle beraber bazı ülkelerdeki kargo şirketleri drone ile servis hizmeti vermeye başlamıştır. Otonom sürüslü araçlar ise son yıllarda insanlar tarafından oldukça büyük bir ilgi ile takip edilmektedir. Bahsedilen bu teknolojik ilerlemeler, sadece insan hayatını kolaylaştırmakla kalmayıp ülkelerin orduları için de büyük bir potansiyel taşımaktadır.

Her alanda olduğu gibi, spor dünyasında da nesne tanıma sistemleri ile karşılaşmaya başladık. Güncel bir örnek olarak, Katar'da düzenlenen 2022 Futbol Dünya Kupası'nda oyuncuların pozisyonlarını maç sırasında canlı bir şekilde takip eden bir nesne tanıma teknolojisi kullanıldı. Bu sayede ofsayt pozisyonlarında hakemler saha içinde hata yapsa da bu teknoloji ile doğru karar alınması sağlandı. Nesne tanıma teknolojisinin kullanıldığı spor dalları sadece futbol ile sınırlı değildir. İstatistiklere göre dünyanın en çok izlenen sporlarından birisi olan basketbolda da kullanılmaktadır. Son yıllarda hem oyuncuların antrenman tekniklerinde hem de takımların strateji belirleme aşamasında çeşitli yapay zekâ uygulamalarından yararlandığı görülmektedir. Örnek olarak Amerikan Basketbol Ligi (NBA) takımlarından Toronto Raptors, "Noah Basketball - Shooting System" adında bir bilgisayar görüşü sistemini, takımdaki oyuncuların şut tekniklerini analiz etmek için kullanmakta ve bu analizler doğrultusunda oyuncuların şut antrenmanlarını dizayn etmektedir. Fakat bir basketbol karşılaşması sadece takımlar ve oyuncularından ibaret değildir. Hakemler de bir

basketbol maçının önemli bir etkenidir. Hakemler, maç içinde verdikleri kararları belirli el işaretleri ile gösterirler. Bu kararlar, masa hakemlerinin de yardımıyla skor tabelasına elle giriş yapılarak, oyuncuların kaç faul yaptığı veya takımların kaç mola aldığı gibi belirli veriler kayıt altına alınır. Bu sürecin maç esnasında hızlı ve doğru bir şekilde gerçekleştirilmesi gerekir, ancak ara sıra insan kaynaklı hatalardan aksaklık yaşanmaktadır ve bu aksaklıklar maçın duraklamasına sebep olmaktadır.

Bu nedenle, bu çalışmada basketbol maçlarında hakemlerin el işaretlerini tespit eden ve sınıflandıran bir nesne tanıma modeli geliştirilmiştir. Bu model, basketbol maçlarında hakemler tarafından yapılan el işaretlerini tanımlayarak, masa hakemlerinin oyuncu faulleri gibi bazı bilgileri skor tabelasına elle girmesine gerek kalmamasını amaçlamaktadır. Bu çalışma, hakem masasına bir kamera sistemi yerleştirilerek, maç sırasında yaptıkları el işaretlerini gerçek zamanlı görüntü üzerinden tespit ederek, faul yapan oyuncuların ve mola alan takımların bilgilerini doğrudan skor tabelasına kaydetmeyi mümkün kılacaktır. Bu sayede masa hakemlerinin bu verileri skor tabelasına elle girmesine gerek kalmaması, bu işlemin daha hızlı ve doğru olarak gerçekleştirilmesi amaçlanmıştır.

Asıl amacının haricinde de bu çalışma biraz geliştirilerek televizyon yayıncılığı için de kullanışlı bir araç olabilir. Basketbol hakkında bilgisi kısıtlı olan izleyiciler için maç sırasında eğitici bir özellik olarak kullanılarak, hakemlerin el işaretlerini tanıtan bir ara yüz geliştirilebilir. Bu ara yüz, basketbol bilgisi kısıtlı seyirci kitlesini spora daha hızlı alıştırmaya ve sporun seyirci sayısını artırmaya yönelik bir pazarlama stratejisi olarak kullanılabilir.

Bahsedilen tüm bu sebepler doğrultusunda, bu çalışmada PyTorch YOLOv7 ve TensorFlow Faster R-CNN Inception v2 modelleri kullanılarak iki ayrı nesne tanıma modeli geliştirilmiş ve sonuçları karşılaştırılmıştır.

## II. İLGİLİ ÇALIŞMALAR

Oldukça sık kullanılan bir yöntem olması sebebiyle, günümüzde nesne tanıma ile ilgili birçok makaleye ulaşmak mümkündür. Zhong-Qiu Zhao vd.[1] çalışmasında, makinelerle tam bir görsel anlayış kazandırabilmek için görüntüleri sadece sınıflandırmakla kalmayıp, aynı zamanda görüntüdeki nesnelere ve konumlarını belirlemenin

önemini vurgulamıştır. Bu amaç ile nesne tanıma metodunun teknik detaylarını açıklamıştır. Zaidi

S.S.A. vd.[2], farklı çalışma prensiplerine sahip en çok kabul gören nesne tanıma modellerini hem teorik hem uygulamalı olarak karşılaştırmıştır. X. Zhou vd.[3], futbol maçlarından görüntüler elde edip, farklı R-CNN modellerinden yararlanarak, oyuncular haricinde kale ve korner direklerini de dahil eden kapsamlı bir nesne tanıma çalışması gerçekleştirmiştir. M. Buric vd.[4] ve Y.Zhang vd.[5], farklı spor dalları için, YOLO ve Deep Sort modelleri ile oyuncu takip çalışmalarını gerçekleştirmişlerdir. Basketbol hakemlerinin el işaretlerini algılama üzerine gerçekleştirilmiş ilgili çalışmalarda, bilgisayarlı görü tekniği haricinde giyilebilir sensör yardımı ile veri toplama tekniği de kullanılmıştır. Naik, B.T. vd[6], bilgisayar görüşü ve nesne tanıma methodları ile farklı spor dallarında gerçekleştirilen çalışmalara dair detaylı bir literatür taraması yapmıştır. Žemgulys J. vd.[7] maçların televizyon yayınlarından elde edilmiş görüntülerin, histogram of oriented gradients (HOG) ve local binary pattern (LBP) özneliklerini çıkarmıştır. Support vector machines ve random forest algoritmalarından yararlanarak %95 doğruluk ile sınıflandırma gerçekleştirmiştir. Yeh C.W. vd.[8] yüzey elektromiyografisi (sEMG) ve üç eksenli ivmeölçer (ACC) kullanarak %97.9'luk bir doğruluğa ulaşmıştır. Hakem haricinde, oyuncu hareketlerinin takibi için Y. Yoon vd.[9] ve Acuna D.[10] YOLO algoritmasını kullanarak çalışmalar gerçekleştirmişlerdir.

## III. BULGULAR

### A. Veri Seti

Veri setinde toplam 144 adet görüntü bulunmaktadır. Görüntülerin büyük bir kısmı Macbook Air video kamerası ile elde edilmiştir. Haricinde Youtube üzerinden basketbol hakemi eğitim videolarından bazı görüntüler de çalışmada kullanılmıştır. Bu 144 adet görüntüye karşılık gelen 8 adet sınıf bulunmaktadır. Her sınıfa ait 18 adet görüntü mevcuttur. Bu sınıflar aşağıdaki gibidir:

- Dört adet oyuncu numaralarını temsil eden sınıflar: No4, No5, No7, No13,
- Üç adet faul çeşidini temsil eden sınıflar: PersonalFoul, FlagrantFoul, TechnicalFoul ,
- Mola sınıfı: Timeout

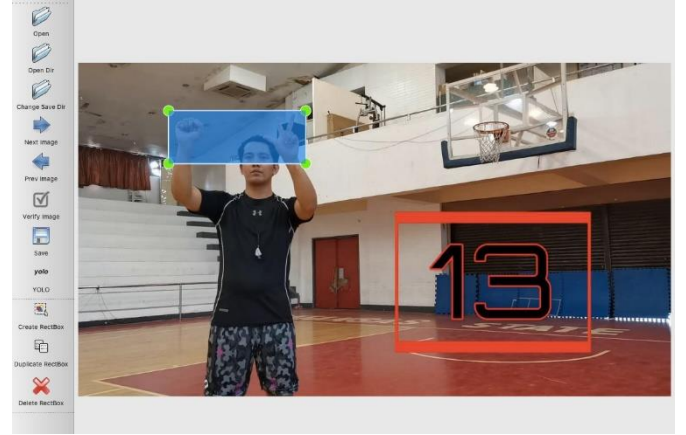
Şekil 1’de veri setinde bulunan görsellerden her sınıfa ait birer adet örnek paylaşılmıştır.



Şekil 1 Veri setinden örnek görseller

## B. Görsel Etiketleme

Eğitim işlemi başlamadan önce veri setindeki tüm görseller Labellmg programı kullanılarak etiketlenmiştir. Bu etiketleme işlemi, Şekil 2’de gösterildiği üzere, el işaretlerinin görsel içindeki konumunu belirlemek için gerçekleştirilir.



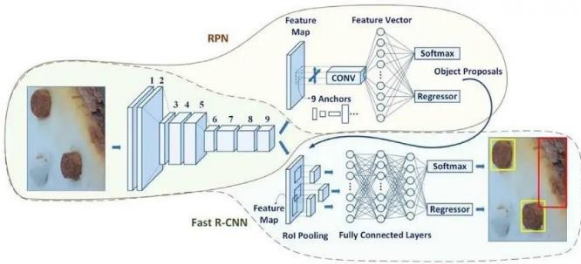
Şekil 2 Labellmg ile görsel etiketleme

Faster R-CNN modeli için PascalVOC, YOLOv7 modeli için YOLO formatında tüm görseller etiketlenmiştir. Bu iki format da prensip olarak aynı işlemi gerçekleştirmektedir. PascalVOC(.xml) formatı etiketlenen nesnenin konumunu (x\_min, y\_min, x\_max, y\_max) değerleri ile belirler. YOLO(.txt) formatı ise bu bilgiyi (x\_center, y\_center, w, h) değerlerini 0 ile 1 arasında normalize edilmiş bir şekilde tutar. Nesne konumlarını barındıran dosyaları PascalVOC formatından YOLO formatına dönüştürmek için formül (1) kullanılabilir.

## C. Kullanılan Yöntemler

### 1. Faster R-CNN Inception v2

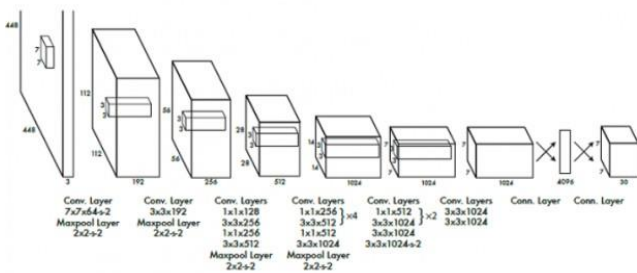
Faster R-CNN [11,12] algoritması, 2015 yılında Microsoft tarafından geliştirilmiş, derin öğrenme tabanlı bir nesne tanıma algoritmasıdır. Faster R-CNN Inception V2 modeli, Faster R-CNN algoritmasının bir versiyonudur ve Inception V2 mimarisini temel ağ olarak kullanır. Inception V2 mimarisi, Google tarafından geliştirilmiş bir evrimsel sinir ağı (CNN)dır ve görüntü analizi, robotik ve otomatik araçlar gibi birçok uygulamada kullanılmaktadır. Genel olarak, Faster R-CNN, bir giriş görüntüsünü aldıktan sonra bu görüntüyü bir CNN aracılığıyla geçirir ve bir dizi özellik haritası çıkarır. Bu özellik haritaları, bir bölge öneri ağı (RPN) tarafından ele alınır ve potansiyel nesne kutuları önerir. Son olarak, sınıflandırıcı her kutu için uygulanır, nesnenin sınıfını tahmin eder ve kutunun konumunu düzgün hale getirir.



Şekil 3 Faster R-CNN mimarisini [6]

## 2. Yolov7

YOLOv7 [13,14] (You Only Look Once), bir başka derin öğrenme tabanlı nesne tanıma modelidir. Bu model bir tek aşamalı tanımlayıcıdır, yani diğer nesne tanıma modellerine (örneğin Faster R-CNN) göre birden fazla aşamaya sahip olmak yerine tüm giriş görüntüsünü tek bir geçişte işler. Genel olarak, YOLOv7 görüntüyü hücreler grubuna bölerek her hücrede nesnelerin varlığını ve konumunu tahmin etmek için bir sinir ağı kullanır. Sinir ağı, her hücrenin bir nesne içerdiği olasılığını, nesnenin sınıfını ve nesnenin kutusunun koordinatlarını tahmin etmek üzere eğitilir. YOLOv7-Tiny, YOLOv7 nesne tanıma modelinin daha küçük ölçekli versiyonudur ve YOLOv7 modelinden az doğruluğa sahip olsa bile daha hızlı olması amaçlanmıştır. Bu sebeple gerçek zamanlı nesne tanıma amacını içeren çalışmalarda sıklıkla tercih edilmektedir.



Şekil 4 Örnek YOLO mimarisini [8]

## IV. TARTIŞMA

### A. Gerçekleştirilen Ortam

Bu çalışmadaki tüm deneyler, PyCharm ortamında Python(3.9) dili ile gerçekleştirilmiştir. Modellerin eğitim sürecinde Intel i7-6700 HQ CPU ve NVIDIA GeForce GTX 950M GPU kullanılmıştır. Tablo 1’de modellerin eğitiminde tercih edilen hiper parametreler paylaşılmıştır.

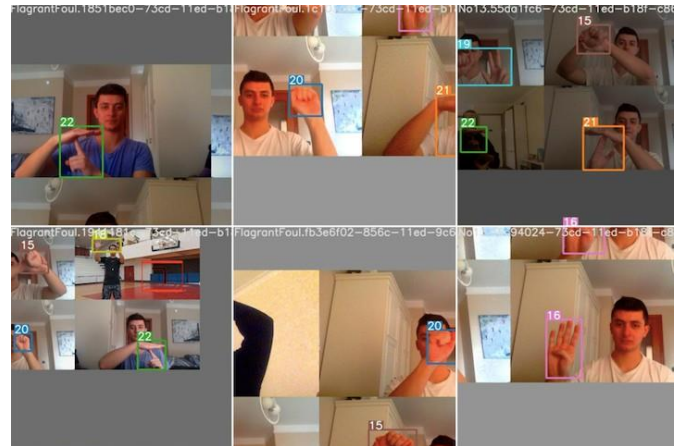
Tablo 1. Model hiper parametreleri

Başlık 1	Faster R-CNN	YOLOv7
Düzenleyici	L2	L2
Öğrenme Katsayısı (LR)	4e-3, 2e-4, 2e-5*	0.01
Optimizör	SGD	SGD
Veri Çoğaltma	Rastgele Yatay Döndürme (RHF)	Döndürme, HSV

\* Faster R-CNN modelinde ilk 5000 adıma kadar 4e-3, 5000 ile 10000 adım arasında 2e-4, 10000 adımdan sonra 2e-5 learning rate kullanılmıştır.

### B. Deneyler ve Sonuçları

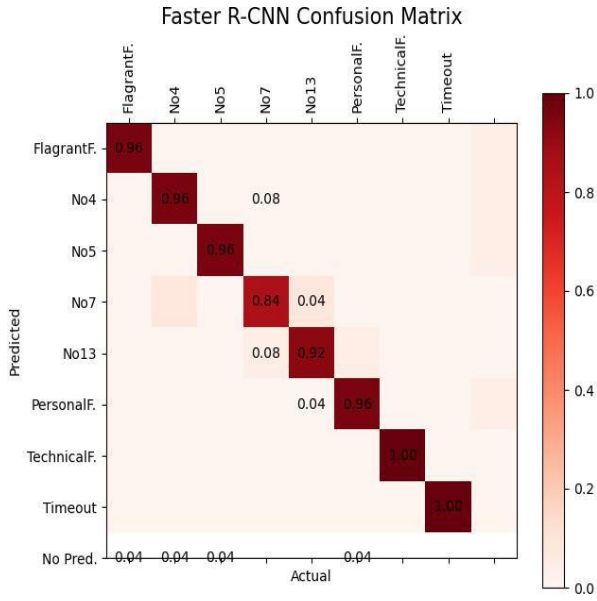
Faster R-CNN Inception v2 modelini eğitim süreci yaklaşık 17000 adım sürmüştür. İlk 5000 adımda learning rate 0.0004, 5000 ile 10000 adım arası learning rate 0.00002, 10000 adımdan sonra ise learning rate 0.000002 olarak belirlenmiştir. Eğitim veri setinde görsel sayısı nispeten az olduğu için veri artırımı teknikleri kullanılmıştır. Bu modelin eğitimi 198 dakikada tamamlanmıştır. YOLOv7 modelinde learning rate 0.01 olarak belirlenmiş ve veri artırımı teknikleri kullanılmıştır. YOLOv7 modelinin eğitimi 500 epoch boyunca devam etmiş ve 163 dakikada tamamlanmıştır. Her iki modelin eğitim aşamasında da, epoch veya adım sayısına paralel olarak loss değerlerinin azaldığı gözlemlenmiştir. Modellerin eğitim aşamasından örnek bir görsel grubu Şekil 5’te paylaşılmıştır.



Şekil 5 Eğitim süreci- Örnek görsel grubu

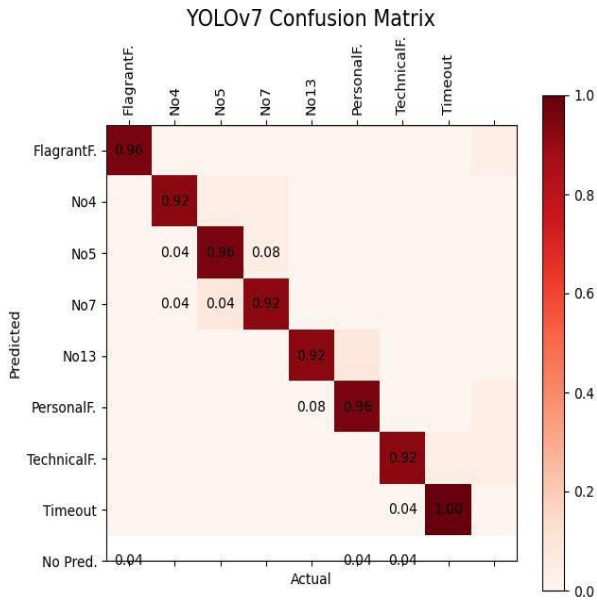
Eğitimi tamamlanan Faster R-CNN Inception v2 ve YOLOv7-Tiny modellerinin test aşamasında 5 katlı çapraz doğrulama yöntemi kullanılmıştır. Her doğrulama sırasında sınıf başına 5, toplamda 40 adet görüntü test edilmiştir. Şekil 6’da modellerin çapraz doğrulama yönteminden elde edilen karmaşıklık matrisleri (confusion matrix) paylaşılmıştır.





Tablo 2 Faster R-CNN, 5 katlı çapraz doğrulama ile test sonuçları

	Faster R-CNN Inception v2			
	Precision	Recall	F1-Score	Support
PersonalFoul	1,000	0,960	0,980	25
TechnicalFoul	0,923	0,960	0,941	25
FlagrantFoul	1,000	0,960	0,980	25
No4	0,955	0,840	0,894	25
No5	0,920	0,920	0,920	25
No7	0,960	0,960	0,960	25
No13	1,000	1,000	1,000	25
Timeout	1,000	1,000	1,000	25
accuracy			0,950	200
macro avg	0,862	0,844	0,853	200
weighted avg	0,970	0,950	0,959	200



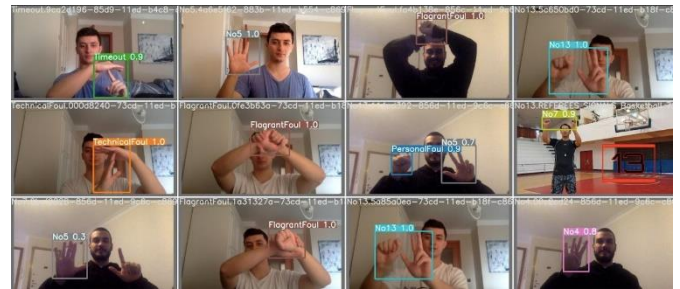
Tablo 3 YOLOv7, 5 katlı çapraz doğrulama ile test sonuçları

	YOLOv7-Tiny			
	Precision	Recall	F1-Score	Support
PersonalFoul	1,000	0,960	0,980	25
TechnicalFoul	1,000	0,920	0,958	25
FlagrantFoul	0,889	0,960	0,923	25
No4	0,920	0,920	0,920	25
No5	1,000	0,920	0,958	25
No7	0,923	0,960	0,941	25
No13	1,000	0,920	0,958	25
Timeout	0,962	1,000	0,980	25
accuracy			0,945	200
macro avg	0,855	0,840	0,847	200
weighted avg	0,962	0,945	0,952	200

Şekil 6 Modellerin karmaşıklık matrisleri

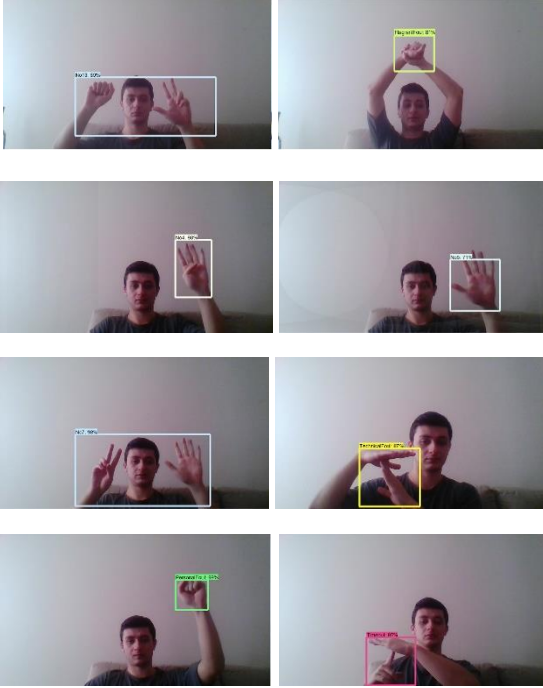
Tablo 2 ve Tablo 3'te modellerin kesinlik (precision), duyarlılık (recall), F1 ve doğruluk (accuracy) skorları paylaşılmıştır.

Modellerin, 5 katlı çapraz doğrulama ile test aşamasında gerçekleştirdiği tahminlerden örnek bir kesim Şekil 7'de gösterilmektedir.

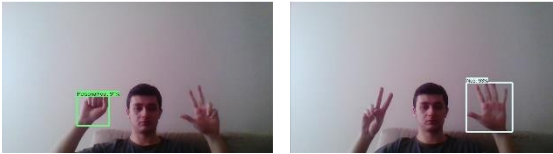


Şekil 7 Test seti tahmin sonuçları

Faster R-CNN modelinin, canlı görüntü üzerinden (bilgisayar video kamerası aracılığıyla) gerçekleştirdiği başarılı ve başarısız sınıflandırmalardan örnek görseller sırasıyla Şekil 8 ve Şekil 9'da paylaşılmıştır.



Şekil 8 Canlı görüntü üzerinden gerçekleştirilen doğru sınıflandırma örnekleri



Şekil 9 Canlı görüntü üzerinden gerçekleştirilen yanlış sınıflandırma örnekleri

## V. TARTIŞMA VE SONUÇ

Basketbol hakemlerinin el işaretlerini algılamaya yönelik bu çalışmada, önerilen metodlar kullanılarak iki adet nesne tanıma modeli oluşturulmuştur. Modellerin performanslarının karşılaştırılabilir olması amacıyla iki model de aynı veri seti ile eğitimi gerçekleştirilmiştir. Faster R-CNN Inception v2 modeli test seti üzerinde %95, YOLOv7- Tiny modeli %94.5 doğruluk oranında başarı göstermiştir. İki model de, çalışma boyunca birbirine oldukça yakın bir performans sergilemiştir. Modeller bilgisayar video kamerası aracılığıyla canlı görüntü üzerinden test edildiğinde ise bu doğruluk değerleri ile paralel bir başarı oranı göstermiştir. Modeller, el işaretleri belirli

mesafeden gösterildiğinde oldukça başarılıdır fakat el işareti kameradan uzaklaştıkça modellerin başarısının düşmeye başladığı görülmüştür. Haricinde ise, 'No5' ve 'No7' (iki sınıfta da beş işareti bulunur.) veya 'PersonalFoul' ve 'No13' (iki sınıfta da yumruk işareti bulunur.) gibi birbirlerine benzer işaretlerin sınıflandırılmasında modelin hata yapabildiği gözlemlenmiştir. Bu durum Şekil 8'de örneklendirilmiştir. Bahsedilen bu eksiklikleri gidermek için gelecek çalışmalarda, veri setinin daha farklı mesafe ve açılardan elde edilmiş görseller eklenerek zenginleştirilmesi ve yeni el işaretleri tanımlanarak da sınıf sayısının artırılması hedeflenmektedir. Daha kapsamlı veri seti üstünde eğitilecek bir algoritma ile basketbol maçlarında kullanılabilir bir nesne tanıma modeli geliştirmek amaçlanmaktadır.

## KAYNAKLAR

- [1] Z.-Q. Zhao, P. Zheng, S. -T. Xu and X. Wu, "Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865
- [2] Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, Brian Lee, A survey of modern deep learning-based object detection models, Digital Signal Processing, Volume 126, 2022, 103514, ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2022.103514
- [3] X. Zhou, W. Gong, W. Fu and F. Du, "Application of deep learning in object detection," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 2017, pp. 631-634, doi: 10.1109/ICIS.2017.7960069.
- [4] M. Buric, M. Ivasic-Kos and M. Pobar, "Player Tracking in Sports Videos," 2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Sydney, NSW, Australia, 2019, pp. 334-340, doi: 10.1109/CloudCom.2019.00058.
- [5] Y. Zhang, Z. Chen and B. Wei, "A Sport Athlete Object Tracking Based on Deep Sort and Yolo V4 in Case of Camera Movement," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 2020, pp. 1312-1316, doi: 10.1109/ICCC51575.2020.9345010.
- [6] Naik BT, Hashmi MF, Bokde ND. A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. Applied Sciences. 2022; 12(9):4429. <https://doi.org/10.3390/app12094429>
- [7] Žemgulys, J., Raudonis, V., Maskeliūnas, R. et al. Recognition of basketball referee signals from real-time videos. J Ambient Intell Human Comput 11, 979–991 (2020). [https:// doi.org/10.1007/s12652-019-01209-](https://doi.org/10.1007/s12652-019-01209-)

1FLEXChip Signal Processor (MC68175/D), Motorola, 1996.

- [8] Yeh, CW., Pan, TY., Hu, MC. (2017). A Sensor-Based Official Basketball Referee Signals Recognition System Using Deep Belief Networks. In: Amsaleg, L., Guðmundsson, G., Gurrin, C., Jónsson, B., Satoh, S. (eds) MultiMedia Modeling. MMM 2017. Lecture Notes in Computer Science(), vol 10132. Springer, Cham. [https://doi.org/10.1007/978-3-319-51811-4\\_46](https://doi.org/10.1007/978-3-319-51811-4_46)
- [9] Y. Yoon et al., "Analyzing Basketball Movements and Pass Relationships Using Realtime Object Tracking Techniques Based on Deep Learning," in IEEE Access, vol. 7, pp. 56564-56576, 2019, doi: 10.1109/ACCESS.2019.2913953
- [10] Acuna, D. (2017, December). Towards real-time detection and tracking of basketball players using deep neural networks. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA (pp. 4-9).
- [11] J. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. <https://doi.org/10.48550/arXiv.1506.01497>
- [12] Achraf KHAZRI, <https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4> (Ocak, 2023)
- [13] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. <https://doi.org/10.48550/arXiv.2207.02696>
- [14] ODSC - Open Data Science, <https://odsc.medium.com/overview-of-the-yolo-object-detection-algorithm-7b52a745d3e0> (Ocak, 2023)