

**T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
YÜKSEK LİSANS TEZİ**

**GÖRÜNTÜ İŞLEME TEKNİKLERİ İLE
BALIKLARDA DURUŞ TAHMİNİ VE HAREKET
TANIMA**

Mehmet Furkan AKÇA

Danışman: Doç. Dr. Mehmet Fatih DEMİRAL

II. Danışman: Prof. Dr. Ali Hakan IŞIK

BURDUR, 2024

ETİK KURALLARA UYGUNLUK BEYANI

Burdur Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin ilgili hükümleri uyarınca **Yüksek Lisans** Tezi olarak sunduğum “**Görüntü İşleme Teknikleri ile Balıklarda Duruş Tahmini ve Hareket Tanıma**” başlıklı bu tezin;

- Kendi çalışmam olduğunu,
- Sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi,
- Bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi,
- Kullandığım verilerde değişiklik yapmadığımı,
- Tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını,
- Bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı,
- Bu tezin planlanmasından yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı,

bildirir, aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

12/08/2024

(İmza)

Mehmet Furkan AKÇA

ÖNSÖZ

“Görüntü İşleme Teknikleri ile Balıklarda Duruş Tahmini ve Hareket Tanıma” başlıklı bu araştırma için beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan değerli Danışman Hocam **Doç. Dr. Mehmet Fatih DEMİRAL** ve 2. Danışman Hocam **Prof. Dr. Ali Hakan IŞIK**’a teşekkürlerimi sunarım.

Araştırmalarım sırasında yardımlarını gördüğüm çalışma arkadaşlarım Ömer Can ESKİCİOĞLU ve Ali Tezcan SARIZEYBEK’e teşekkür ederim.

Eğitim hayatımın her aşamasında beni her anlamda destekleyen aileme sonsuz sevgi ve saygılarımı sunarım.

Ağustos, 2024

Mehmet Furkan AKÇA

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	i
İÇİNDEKİLER	ii
ŞEKİLLER DİZİNİ.....	iii
ÇİZELGELER DİZİNİ	v
SİMGELER VE KISALTMALAR DİZİNİ	vi
ÖZET.....	vii
ABSTRACT.....	viii
1. GİRİŞ.....	1
2. GENEL BİLGİLER.....	3
2.1. Literatür Araştırması.....	3
2.2. Yapay Zekâ	7
2.3. Yapay Sinir Ağları	7
2.4. Evrişimsel Sinir Ağları	8
2.5. Duruş Tahmini	9
2.6. Hareket Tanıma.....	9
3. MATERYAL VE YÖNTEM	10
3.1. Materyal	10
3.1.1. Veri Seti	10
3.1.2. Performans Metrikleri	10
3.2. Yöntem.....	12
3.2.1. Görüntü Ön İşleme Teknikleri ve Veri Setinin Hazırlanması.....	12
3.2.2. Veri Artırma	14
3.2.3. Nesne Tespiti ve Anahtar Nokta Tespiti	16
3.2.4. Transfer Öğrenme Yöntemleri	16
3.2.5. Hiper-Parametre Optimizasyonu.....	23
4. ARAŞTIRMA BULGULARI VE TARTIŞMA.....	24
4.1. Araştırma Bulguları	24
4.2. Tartışma	37
5. SONUÇ VE ÖNERİLER	38
KAYNAKLAR	40

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 2.1. Basitleştirilmiş yapay sinir ağları yapısı.....	8
Şekil 2.2. Evrişimsel sinir ağları mimarisi.....	8
Şekil 3.1. Örnek karmaşıklık matrisi	11
Şekil 3.2. Görüntü eşikleme uygulananarak siyah beyaz yapılmış görüntüler	12
Şekil 3.3. Anahtar nokta tespiti.....	13
Şekil 3.4. Hareketi oluşturan balık baş noktalarının düzleme aktarılması.....	13
Şekil 3.5. Model eğitimi için kırılıp yeniden boyutlandırılmış görüntü	14
Şekil 3.6. VGG-16 mimarisi	17
Şekil 3.7. “Bottleneck” bloğu	17
Şekil 3.8. ResNet50 ve ResNet101’in mimari yapıları.....	18
Şekil 3.9. Normal hücre ve indirgeme hücresinin yapısı	19
Şekil 3.10. Basitleştirilmiş NasNet mimarisi	19
Şekil 3.11. Inception modülleri	20
Şekil 3.12. Inception v3 model mimarisi	20
Şekil 3.13. DenseNet169 mimarisi	21
Şekil 3.14. EfficientNet-B4 model mimarisi	21
Şekil 3.15. Xception model mimarisi	22
Şekil 4.1. YOLOv8 sonuçları	24
Şekil 4.2. NasNetLarge 9 sınıf değerlendirme grafikleri.....	24
Şekil 4.3. NasNetLarge 9 sınıf karmaşıklık matrisi.....	25
Şekil 4.4. NasNetLarge 7 sınıf değerlendirme grafikleri.....	26
Şekil 4.5. NasNetLarge 7 sınıf karmaşıklık matrisi.....	26
Şekil 4.6. ResNet50 değerlendirme grafikleri.....	27
Şekil 4.7. ResNet50 karmaşıklık matrisi.....	28
Şekil 4.8. ResNet101 değerlendirme grafikleri.....	29
Şekil 4.9. ResNet101 karmaşıklık matrisi.....	29
Şekil 4.10. EfficientNet-B4 değerlendirme grafikleri.....	30
Şekil 4.11. EfficientNet-B4 karmaşıklık matrisi	30
Şekil 4.12. Inception v3 değerlendirme grafikleri	31
Şekil 4.13. Inception v3 karmaşıklık matrisi	32
Şekil 4.14. Xception değerlendirme grafikleri.....	33

Şekil 4.15. Xception karmaşıklık matrisi.....	33
Şekil 4.16. VGG-16 değerlendirme grafikleri	34
Şekil 4.17. VGG 16 karmaşıklık matrisi.....	34
Şekil 4.18. DenseNet169 değerlendirme grafikleri	35
Şekil 4.19. DenseNet169 karmaşıklık matrisi	36



ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 3.1. Yedi sınıflı veri setinin görüntü sayıları	15
Çizelge 3.2. Dokuz sınıflı veri setinin görüntü sayıları	15
Çizelge 4.1. NasNetLarge 9 sınıflı sınıflandırma raporu	25
Çizelge 4.2. NasNetLarge 7 sınıflı sınıflandırma raporu	27
Çizelge 4.3. ResNet50 sınıflandırma raporu	28
Çizelge 4.4. ResNet101 sınıflandırma raporu	29
Çizelge 4.5. EfficientNet-B4 sınıflandırma raporu	31
Çizelge 4.6. Inception v3 sınıflandırma raporu	32
Çizelge 4.7. Xception sınıflandırma raporu	33
Çizelge 4.8. VGG 16 sınıflandırma raporu	35
Çizelge 4.9. DenseNet169 sınıflandırma raporu	36
Çizelge 4.10. Modellerin performans karşılaştırması	37
Çizelge 4.11. Literatür karşılaştırması	37

SİMGELER VE KISALTMALAR DİZİNİ

DSRN	: Dual-Stream Recurrent Network
DVM	: Destek Vektör Makinesi
ESA	: Evrişimsel Sinir Ağları
FPS	: Frame Per Second
GN	: Gerçek Negatif
GP	: Gerçek Pozitif
LR	: Lojistik Regresyon
MBConv	: Mobile Inverted Bottleneck Convolution
NACA	: National Advisory Committee for Aeronautics
ReLU	: Rectified Linear Unit
RO	: Rastgele Orman
YN	: Yanlış Negatif
YOLO	: You Only Look Once
YP	: Yanlış Pozitif
YSA	: Yapay Sinir Ağları

ÖZET

Yüksek Lisans Tezi

Görüntü İşleme Teknikleri ile Balıklarda Duruş Tahmini ve Hareket Tanıma

Mehmet Furkan Akça

**Burdur Mehmet Akif Ersoy Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı**

**Danışman: Doç. Dr. Mehmet Fatih DEMİRAL
II. Danışman: Prof. Dr. Ali Hakan IŞIK**

Temmuz, 2024

Balıkların besin değerinin yüksek olması, nüfus artışına bağlı olarak balık etine olan talebin artması ve gelecekte daha da artacak olması balıkların insan yaşamındaki önemini ortaya koymaktadır. Ancak bu öneme karşın yeterli veri setinin olmaması gibi etkenler balıklar üzerinde çalışma yapmayı zorlaştırmaktadır. Bir başka ifadeyle literatürde balıklar üzerine yapılmış çalışma sayısı yeterli düzeyde değildir. Bu alanda yapılacak çalışmalar hem literatüre katkı yapacağı hem de bu konudaki bilinçlenme düzeyini artıracığı söylenebilir. Bu çalışmada veri olarak Xu vd. tarafından 2017 yılında yapılan çalışmada açık olarak paylaşılan veri seti kullanılmıştır. 16528 görüntüden oluşan veri setinde 9 farklı sınıfta balık hareketi mevcuttur. Bu çalışmada literatürden farklı olarak direkt balık görüntülerinin üzerinden işlem yapılmasının aksine balığın hareketini duruş tahmini yöntemi ile iki boyutlu bir düzleme aktarıp hareketinin görüntüsü üzerinden transfer öğrenme yöntemlerini kullanarak sınıflandırılması önerilmiştir. Çalışmada hareketin düzleme aktarılma aşamalarında görüntü işleme yöntemleri kullanılmıştır. Anahtar nokta tespiti için YOLOv8 modeli ve sınıflandırma için evrimsel sinir ağıları modellerinden olan Xception, ResNet50, ResNet101, Inception v3, NasNetLarge, DensNet169, VGG 16, EfficientNet-B4 modelleri kullanılmıştır. Çalışmada detaylı olarak sınıflandırma raporları verilen modellerden en yüksek model performansına %89,43 doğruluk, %91,85 kesinlik, %88,80 duyarlılık ve %89,17 f1-skor değerleri ile EfficientNet-B4 ile ulaşılmıştır.

Anahtar Kelimeler: görüntü sınıflandırma, duruş tahmini, hareket tanıma, transfer öğrenme, görüntü işleme

ABSTRACT

M.Sc. Thesis

Pose Estimation and Action Recognition in Fish with Image Processing Techniques

Mehmet Furkan Akça

**Burdur Mehmet Akif Ersoy University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering**

**Supervisor: Assoc. Prof. Dr. Mehmet Fatih Demiral
Co-Supervisor: Prof. Dr. Ali Hakan Işık**

July, 2024

The high nutritional value of fish, raise in the demand for fish meat due to population growth and the fact that the increase will continue even more in the future reveal the place of fish in human life. However, despite this importance, factors such as the lack of sufficient data sets make it difficult to conduct studies on fish. In other words, the number of studies on fish in the literature is not adequate. It can be said that studies to be done in this field will, on the one hand, contribute to the literature, on the other hand, increase the level of awareness on this issue. In this study, the data set shared openly in the study conducted by Xu et al in 2017 was used as data. There are 9 different classes of fish movements in the dataset consisting of 16528 images. In this study, as different from the literature, it is proposed to classify the motion of the fish by transferring the motion of the fish to a two-dimensional plane with the pose estimation method and using transfer learning methods over the image of the motion. In the study, image processing methods were used to transfer the motion to the plane. In this study, following models were used: YOLOv8 model for key point detection and Xception, ResNet50, ResNet101, Inception v3, NasNetLarge, DenseNet169, VGG 16, EfficientNet-B4 models for classification. As a result, among the models, EfficientNet-B4 had the highest model performance with 89.43% accuracy, 91.85% precision, 88.80% recall and 89.17% f1-score values.

Keywords: image classification, pose estimation, action recognition, transfer learning, image processing

1. GİRİŞ

Balık, insan kültürünün, endüstrinin ve deniz ekosistemlerinin önemli bir bileşenidir. Dünya genelinde üç milyardan fazla bireyin düzenli olarak balık tükettiği düşünülürse balığın insanoğlunun yemek kültüründeki yeri yadsınamaz. Ayrıca, balık, endüstrinin ve deniz ekosisteminin önemli bir bileşenidir (Vianna vd., 2020). Bunun yanında artan dünya nüfusu ile birlikte gelecekte insanların gıdaya olan ihtiyacının artacak olması ve günümüzde kara yüzeyinin neredeyse yarısının gıda üretimi için kullanılması (Duda vd., 2015) insan nüfusunun artışına paralel olarak balık çiftliklerine ve deniz altı çalışmalarına yönelimine sebep olacağından bu alanlarda algoritmalar geliştirmenin gerekliliği ortaya çıkmıştır (Schellewald vd., 2021). Bu çalışmaların yapılabilmesi için gerekli olan yöntemlerden birisi uzaktan kumandalı araçlarda dahil olmak üzere insansız su altı araçlarıdır ancak bu araçların güvenli bir biçimde çalışabilmeleri için çalıştıkları ortama göre balık ağlarından ve balık sürülerinden zarar görmeden hareket etmeleri gerekmektedir (Schellewald vd., 2021). Balık ağlarının hareket kabiliyetleri sebebiyle tehlike dereceli çok yüksek olmamalarına karşın balıklar sürü halinde ve hızlı hareket edebilmeleri sebebiyle insansız araçların balıkların hareketlerini tahmin ederek hareketlerine devam etmeleri gerekmektedir. Bu noktada poz tahmini ve hareket tanıma çalışmalarının önemi artmaktadır.

Balık bilimcilere göre dünya genelinde 30.000'den fazla balık türünün olduğu belirtilmektedir. Balıkların çoğunun birbirine benzer şekillere sahip olması nedeniyle sadece fiziksel görünümlerine bakarak balıkları birbirinden ayırt etmek oldukça zordur (Rum ve Nawawi, 2021). Günümüzde balığa olan talebin artması bunun yanında balıkların da insanların yaşamlarını sürdürebilmeleri sağlayacak yüksek besin değerlerine sahip olması (Kılıçarslan vd., 2024; Yasin vd., 2023), balık çiftliklerindeki üreticilerin balıkları doğru bir şekilde yetiştirebilmesi için balıkların izlenmesi, büyüme durumu ve davranışsal özelliklerini algılayabilmesi gerekliliğini beraberinde getirmiştir (B. Lin vd., 2021). Bu zorluk ve gereklilik bir araya geldiğinde su ürünleri yetiştiriciliği açısından balık görüntü tanıma çalışmaları büyük önem taşımaktadır (Khalifa vd., 2018; Kononov vd., 2019; Rum ve Nawawi, 2021).

Su altındaki nesne hakkında birçok farklı yöntem kullanarak bilgi edinmek mümkündür. Ancak, araştırmacılara göre nesnenin bulunduğu su kütlesinin derinliği, ışık alıp-almama durumu, nesnenin arka ve ön planının hareketli olup-olmaması, suyun bulanıklığı ve düşük görüntü kalitesi gibi faktörler su altı görüntülerini kullanarak nesneyi

tespit etmeyi zorlaştırmaktadır (Pagire vd., 2024; L. Yang vd., 2021). Kullanılan yöntemlerden başlıcası balıkların hareketlerini daha iyi algılayabilmek için balıkların üzerinde derin öğrenmeye bağlı bilgisayarlı görme teknikleri kullanılmasıdır (Xin Wu vd., 2024). Bu tekniklerden nesne tespiti ve duruş tahmini, balık dağılımı, balığın gözlenmesi ve ölçülmesinde kullanılan en önemli iki destek teknoloji olarak ifade edilmektedir (Walther vd., 2004). Hem nesne tespiti hem de duruş tahmini makine görüşünün temel görevleri arasındadır. Birincisi, belirli bir görüntüdeki belirli bir kategorinin hedef nesnelerin olup olmadığını belirlemek için kullanılırken, ikincisi giriş görüntüsündeki hedef nesnenin duruşunu tahmin etmek için kullanılır (B. Lin vd., 2021; Rashid vd., 2017). Ancak bahsedildiği gibi su altı çekimleri daha zorlu olduğu ve bu durum maliyetli olabileceği için bu alanda çalışmalar yaparken balık benzeri robotlar da kullanılmaktadır. Bu robotların hareketleri üzerinde duruş tahmini çalışmaları yapılarak daha doğru tespit edilebilmektedir (T. Zhang vd., 2021). Bu durum da su altı çalışmalarına farklı bir bakış olarak değerlendirilebilir.

Bu durumlar dikkate alındığında balıklar üzerinde yapılmış ya da yapılacak olan çalışmalar hem literatür hem de gelecek nesiller açısından önem arz etmektedir. Bu tez çalışması ile üç boyutlu bir ortamı iki boyuta indirerek istenilen çalışmanın yapılması hem literatüre farklı bir bakış açısı kazandıracak hem de gelecek çalışmalara öncü olacaktır.

2. GENEL BİLGİLER

2.1. Literatür Araştırması

Literatürde, balık türlerinin sınıflandırılmasında uzmanlar tarafından farklı makine öğrenmesi ve derin öğrenme yöntemleri kullandığı ifade edilmektedir (Alaba vd., 2022). Örneğin, Huang vd. (P. X. Huang vd., 2013), ‘fish4knowledge’ veri setinde balık sınıflandırması için ‘Destek Vektör Makinesi (DVM)’ ve balıkların hiyerarşik özelliklerini önermiştir (Fisher vd., 2016). Jäger vd. (Jäger vd., 2016) özellik çıkarma için AlexNet mimarisini ve LifeCLEF 2015 balık veri seti için çok sınıflı DVM (Joly vd., 2015) kullanmıştır. Zhuang vd. (Zhuang vd., 2017) ise LifeCLEF 2015 balık verileriyle modern derin öğrenme tabanlı modeller üzerine ön işleme (pre-processing) ve işlem sonrası (post-processing) yöntemleri kullanmıştır.

Xu vd. çalışmalarında (Xu vd., 2017) derinlik görüntülerinden normal şartlarda ayrı ayrı çalışmalar olarak ele alınan eklemli nesnelerin poz tahmini, takibi ve hareket tanınması gibi problemleri, Lie grup teorisine dayalı bir paradigma ile toplu bir şekilde ele almaya çalışmışlardır. Yaptıkları çalışmada bu yaklaşımı fare, balık ve insan eli üzerinde değerlendirmişlerdir. 9 benzersiz kategoride yaptıkları balık hareket tanıma çalışmasında, teget vektör özelliklerine dayalı yaklaşımları sonucunda %91,33’lük ortalama doğruluğa ulaşmıştır. Bu sebeple çalışma sonuç olarak benzer çalışmalarla kıyaslandığında rekabetçi bir performans ortaya koymuştur.

Zheng vd. yaptıkları çalışmada (Zheng vd., 2018) balıkçı teknelerinden alınan balık görüntülerin sınıflandırılmasını amaçlamışlardır. Çalışmalarında balık türlerini tespit etmeyi ve sınıflandırmayı amaçlayan içerisinde 8 kategori olan bir Kaggle veri seti kullanmışlardır. Çalışmada sınıflandırmanın doğruluğu için iki yöntem ortaya konulmuş ve dengesiz veri setinden kaynaklanabilecek sorunları önlemek için veri çoğaltma yöntemi uygulanmıştır. Sınıflandırma için AlexNet, Caffenet gibi sinir ağlarının kullanıldığı çalışmada sonuçlar sınıflandırmanın başarılı bir sonuç verdiğini ortaya koymaktadır.

Måløy vd. yaptıkları çalışmada (Måløy vd., 2019) derin öğrenme yaklaşımlarının insanlarda hareket tanıma için uygun olduğunu fakat hayvanlarda ve hayvancılıkta uygulamasının az olmasına değinmiştir. Balıklarda poz değişikliklerinin az olmasına rağmen beslenme gibi su ürünleri yetiştiriciliğinde maliyetli bir işlemin optimizasyonu için hareket tanıma çalışmasını yapmışlardır. Bu doğrultuda somon balığının davranışlarını yakalamak için Dual-Stream Recurrent Network (DSRN) kullanmışlar ve

bu yöntem ile %80'lik bir başarı oranı yakalamaları sonucunda DSRN'nin balıklarda davranış tahmini alanında endüstriyel kullanıma uygun olduğunu ortaya koymuşlardır.

Lin vd. tarafından yapılan çalışmada (B. Lin vd., 2021) hayvanların hareketlerini algılayan video tabanlı bir yöntemin hayvanların davranışlarını analiz etmenin güçlü bir yolu olduğu belirtilmiştir. İnsanlar ve balıkların fizyolojik durumlarının kontrolü işleminin ise iskelet durumlarının kontrolü ile sağlanılabileceğinden bahsedilmiştir. Ancak çalışmada yapay zekâ için veri setlerinin azlığından bahsedilmiş ve çalışma kapsamında dünyanın 10 noktası işaretlenmiş 1000 adet tekil balık görüntüsünden oluşan dünyanın ilk balık duruşu veri seti oluşturulmuştur. Balık duruş tahmini çalışması için iki aşamalı bir tahmin modeli kullanılmıştır. Sonuç olarak modelin hassasiyeti %90,61, F1 skoru ise %90 olarak bulunmuştur.

Wang vd. yaptıkları çalışmada (G. Wang vd., 2021) balık davranışlarının hızlı ve kesin olarak tanınmasının su ürünleri sistemleri konusunda çiftçilere sağlayacağı kolaylığa değinmiş, geleneksel yöntemlerin balığın davranışlarını ve refahını etkileyebilme ihtimalinden dolayı derin öğrenmeye dayalı uzay-zaman ve hareket bilgilerine sahip yeni bir yöntem sunmuşlardır. Sundukları yöntemde 5 farklı davranışın (beslenme, hipotermi vb.) tanınması için FlowNet2 ve 3B CNN kullanan çift akışlı bir 3B evrişimsel sinir ağı kullanmışlar ve sonuç olarak ortalama %95,79 doğruluk oranına ulaşmışlardır. Bu sonuç, yöntemlerin balık sağlık durumlarının algılanmasında etkili bir araç olduğunu ortaya koymuştur.

Zhang vd. yaptıkları çalışmada (T. Zhang vd., 2021) biyomimetik su altı cihazlarının bir örneği olan balık benzeri robotların, balıkların hareketlerini iyi derecede taklit edebilmeleri sebebiyle balıkların hareketlerinin araştırılmasında önemli olduklarını belirtmiştir. Konunun zorluğu sebebiyle alanda çok fazla çalışma olmamasından dolayı mevcut olan metotlardan daha hızlı bir metot üretmek için uğraşmışlardır. Yaptıkları çalışma kapsamında TAB-IOL adını verdikleri yeni bir derin sinir ağı tabanlı yöntemi ortaya atmışlardır. Mevcut algoritmalar ile yaptıkları karşılaştırmalarda hem genel doğruluk hem de FPS (Frame Per Second) değerleri karşılaştırıldığında TAB-IOL yönteminin daha başarılı sonuçlar verdiğini ortaya koymuşlardır.

Habe vd. tarafından yapılan çalışmada (Habe vd., 2021) balık sürülerinin denetimi için National Advisory Committee for Aeronautics (NACA) kanat profili modelini kullanan bir poz tahmini yöntemi önerilmiştir. Bu yöntem ile balığın duruşuna göre yüzme durumunun kontrolü sağlanmaktadır. Model aşamasında balığın baş, gövde ve kuyruk olacak şekilde vücut parçalarının tespitinde DeepLabCut kullanılmıştır. Elde

edilen parçalar balık üzerinde 3 nokta olarak durum vektörünün bileşenlerini ayarlama kullanılmıştır. Çalışma sonucunda oluşturulan NACA modeli hem simülasyon hem de gerçek video verileri üzerinde yüzme yönünde hızlı değişimler olduğu durumlar dahil olmak üzere başarılı sonuçlar sağlamıştır.

Rum ve Nawawi (Rum ve Nawawi, 2021), VGG16 kullanarak geliştirdikleri FishDeTec modelini Malezya’da yetişen sekiz farklı tatlı su türüne ait 178 görüntü içeren veri kümesine uygulamış ve 15 yinleme sonucunda balık türlerini %60-80 oranında doğru tahmine ulaşmışlardır. FishDeTec’in özelliği büyük ölçekli bir görüntü sınıflandırma işlemine yönelik derin bir evrişimsel sinir ağı (ESA) modeli olmasıdır. FishDeTec’i geliştirmek için Python, Java, TensorFlow Mobile, Keras, VGG16, Google Colab, Android Studio ve Firebase Database teknolojilerini kullanmışlardır.

Er (Er, 2021) araştırmasında, önceden eğitilmiş ESA yardımıyla göğüs röntgen görüntülerini pnömoni ve pnömoni olmayan şekilde sınıflandırmada özellik çıkarıcı olarak kullanmıştır. Araştırmacı, önceden eğitilmiş derin öğrenme modelleri olarak ise AlexNet, VGG16, ResNet (ResNet18, ResNet50, ResNet101) modellerini tercih etmiştir. Bu modellerden elde edilen özellikler birleştirilerek hibrit özellik vektörü elde edilmiştir. Sınıflandırıcı olarak DVM ve derin öğrenme modellerinin son katmanında bulunan Softmax kullanılmıştır. Araştırmada en yüksek sınıflandırma başarısı %98,32 ile hibrit özellik vektöründen elde edilmiştir.

Kandimalla vd. (Kandimalla vd., 2022) çalışmalarında ilk olarak Ocqueoc Nehri’nden (Michigan-ABD) elde ettikleri yüksek çözünürlüklü DIDSON görüntüleme sonar veri kümesinden sekiz balık türünü YOLO (you only look once) makine öğrenimi modelinin bir uyarlamasını kullanarak ve doğru bir şekilde tespit ederek sınıflandırmaya çalışmışlardır. İkinci olarak, Columbia Nehri (Washington-ABD) üzerindeki Wells Barajı balık merdiveninden optik kameralar tarafından çekilen bir video veri kümesini kullanarak balıkları izlemek ve saymak için Norfair nesne izleme çerçevesini (framework) kullanmışlardır. Sonuçta hem yüksek çözünürlüklü görüntüleme sonarını hem de bir balık merdiveninden alınan su altı videosunu kullanarak derin öğrenme modelleri yardımıyla balıkların türlerinin tespit edilebileceğini göstermişlerdir. Çalışmada elde edilen bir başka sonuç ise gerçek zamanlı balık tespiti ve sınıflandırması için akustik/sesli verileri kullanırken YOLOv3’ün Mask-RCNN’ye kıyasla daha iyi bir model olabileceğidir. Ayrıca, görüntü karelerini işleme açısından saniyede 24 FPS hızla YOLOv3, 8 FPS hıza sahip Mask-RCNN’e kıyasla 3 kat daha hızlı olduğu belirlenmiştir.

Hindarto (Hindarto, 2023) çalışmasında iki sinir ağı mimarisinin (Inception ResNet ve Xception) deniz balıklarının nesne tespitinde karşılaştırmalı bir analizini yapmıştır. Araştırma sonucunda her iki yapının da çeşitli su altı çevresel bağlamlarında nesneleri tanımlamada güçlü bir yeteneğe sahip olduğunu ve dikkate değer bir hassasiyet derecesi gösterdiğini ifade etmektedir. ‘Derinlemesine ayrılabilir evrişim’ konseptini kullanan Xception, parametre sayısı ve hesaplama işlemleri açısından daha verimli bir seçenek olarak karşımıza çıkmakta ve böylece eğitim ve çıkarım prosedürlerinin hızını artırmaktadır. Inception ResNet, özellikle aydınlatma ve arka plandaki değişimlerin varlığında güçlü bir performans gösterirken, Xception su altı nesnelerinin hızlı ve verimli bir şekilde tespit edilmesi için umut verici bir alternatif olarak ortaya çıkmaktadır. Bu çalışmadan elde edilen bulgular, karmaşık deniz ortamlarında nesne algılama çözümleri yapanlar için önemli sayılacak bilgiler sunmaktadır.

Akgül vd. (Akgül vd., 2023) balığın taze olup olmadığını belirlemek için balık tazelik algılama modeli geliştirmişlerdir. Veri Kümesi1 (Hamsi) ve Veri Kümesi2 (İstavrit) üzerinde balık gözü ve solungaç bölgelerini kullanarak balığın taze olup olmadığını tespit etmek için yeni bir hibrit model yapısı önermişlerdir. Ele alınan veri kümelerinden iyi bir özellik çıkarabilmek için transfer öğrenme ile eğitilmiş olan Inception-ResNet-v2 ve Xception model yapıları Yolo-v5 model yapısı ile hibrit bir şekilde kullanılmıştır. Sonuç olarak, ‘Yolo-v5 + Inception-ResNet-v2’ ve ‘Yolo-v5 + Xception hibrit model yapılarının balıkların (hamsi ve istavrit) tazeliğini belirlemede üstün başarı gösterdiğini bulmuşlardır.

Zhou vd. (Zhou vd., 2023) çalışmalarında su altındaki bulanıklık ve düşük kontrasttan kaynaklanan düşük sınıflandırma sorununu çözmek için özellik çıkarıcı olarak önceden eğitilmiş ‘Resnet50 sinir ağını’ kullanmıştır. Ayrıca, çalışmada orijinal softmax sınıflandırıcı yerine ESA modelinde sınıflandırıcı olarak kullanılan hata minimize edilmiş rastgele vektör fonksiyonel bağlantı (EEMRVFL) sinir ağının geliştirilmesi önerilmektedir. ESA modeli, hâlihazırda kullanılan yöntemlere göre su altı görüntü sınıflandırması için daha ileri sınıflandırma doğruluğu sergilemektedir. Bu durum URPC, LifeCLEF 2015 ve Fish4Knowledge gibi mevcut veri kümeleri üzerinde sırasıyla %99,68, %97,34 ve %99,77’ye varan doğruluk oranları sergilemiştir.

Jia vd. (Jia vd., 2023)’in belirttiğine göre; deniz organizmaları için çok sayıda küçük ve üst üste gelen hedeflerin bulunduğu su altı ortamları karmaşık bir yapıya sahiptir. Bu sorunların üstesinden gelmek için YOLOv8 dedektörünü geliştirmişlerdir.

Deneyisel sonuçlar, geliştirilmiş YOLOv8'in %84,5'lik mAP değerine ulaştığını ve bunun orijinal YOLOv8'e göre yaklaşık %6,2'lik bir iyileşme olduğunu göstermiştir.

Pagire vd. (Pagire vd., 2024) çalışmalarında, belirli bir veri kümesi (Fish4knowledge) kullanarak su altı balıklarının tespiti için üç gelişmiş nesne algılama modeli YOLOv3, SSD MobileNetV2 ve Faster R-CNN ResNet50'nin kullanılıp-kullanılmayacağını araştırmıştır. Çalışmada, hassasiyet değerleri Faster R-CNN ResNet50, YOLOv3 ve SSD MobileNetV2 için sırasıyla %45,06, %79 ve %98,21 olarak bulunmuştur. Bir başka ifadeyle, SSD MobileNetV2 modeli YOLOv3 ve Faster R-CNN ResNet50 modellerine kıyasla daha yüksek hassasiyet değerine sahiptir.

Kılıçarslan vd. (Kılıçarslan vd., 2024) çalışmalarında 4476 balık görüntüsünden oluşan bir veri kümesi üzerinde üç transfer öğrenme modeli (MobileNetV2, Xception, VGG16) kullanarak özellik çıkarma işlemini gerçekleştirmiştir. Sınıflandırma işlemi için dört makine öğrenmesi algoritması (DVM, Lojistik Regresyon (LR), Yapay Sinir Ağları (YSA), Rastgele Orman (RO)) uygulanmıştır. DVM ve LR algoritmalarıyla Xception ve MobileNetV2'nin sinerjisi %100 başarı oranına ulaşmıştır.

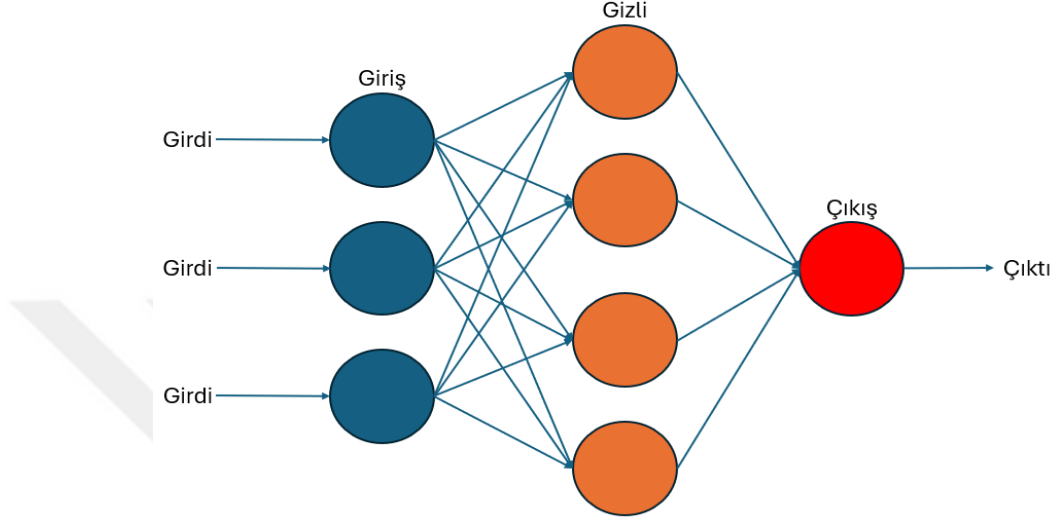
2.2. Yapay Zekâ

Ortaya ilk defa 1956 yılında atılan yapay zekâ (Salehi ve Burgueño, 2018) temelde insan zekasını anlayabilmek için bilgisayarların kullanılması, bu doğrultuda akıllı makineler ve akıllı bilgisayar programları yapma bilimidir (McCarthy, 2007). Makine öğrenimi, doğal dil işleme, görüntü işleme ve veri madenciliği gibi alanlarda sıklıkla kullanılmaktadır (Gupta, 2017; Salehi ve Burgueño, 2018). Bu özellikleri sayesinde günlük hayatta çok fazla karşımıza çıkan yapay zekâ teknolojiye sağladığı fayda ile yapılan işlerdeki insan gücünün kullanımını azaltmaktadır. Nesne tespiti, görüntü sınıflandırma, doğal dil işleme gibi alanlarda kullanılan yapay zekâ bu tez çalışmasında görüntü üzerinden balıkların tespit edilip anahtar noktalarının tespit edilmesinde ve bu noktalar ile oluşturulan hareket görüntülerinin sınıflandırılmasında kullanılmıştır.

2.3. Yapay Sinir Ağları

YSA makine öğrenmesinin bir alt dalı olarak ortaya çıkmış (LeCun vd., 2015) ve günümüzde doğal dil işleme (Devlin vd., 2018), bilgisayarlı görme (Krizhevsky vd., 2012) ve stratejik oyunlar (Silver vd., 2017) da dahil olmak üzere birçok alanda kullanılmaktadır. İnsan beyninde birbirine bağlı on milyardan fazla nöron vardır. YSA bu biyolojik sinir sisteminin matematiksel modellenmesi ile ortaya çıkmıştır (Abraham,

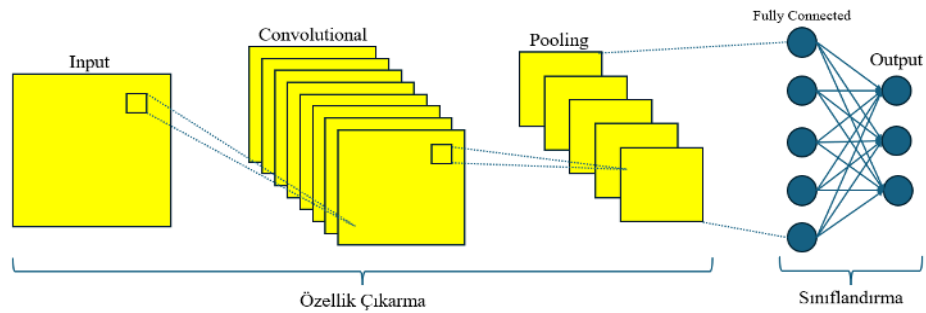
2005). Şekil 2.1’ de basitleştirilmiş şekli görülebilen yapay sinir ağları giriş katmanı gizli katman ve çıkış katmanı olarak üç katmandan oluşan (Lopez ve Kalita, 2017) YSA’nın temel işlem birimlerine nöron denir (Abraham, 2005; Lopez ve Kalita, 2017) ve nöronlar sonraki katmanlar ile bağlantılı olarak çalışır. Nöronların öğrenme yeteneği seçilen algoritmaya göre değişebilir (Abraham, 2005).



Şekil 2.1. Basitleştirilmiş yapay sinir ağları yapısı

2.4. Evrimsel Sinir Ağları

ESA canlıların görsel algılama mekanizmasından alınan ilham ile ortaya çıkmış yaygın kullanılan bir derin öğrenme mimarisidir (Gu vd., 2018), bu nedenle ESA’lar öncelikle girdinin görüntü olmasına dikkat eder (O’shea ve Nash, 2015). ESA’lar kullanılarak nesne tespiti (Dhillon ve Verma, 2020), görüntü tanıma (Q. Liu vd., 2017; Rawat ve Wang, 2017), görüntü segmentasyonu (Ajmal vd., 2018) çalışmaları yapılabilmektedir (Z. Li vd., 2021). ESA temelde 3 katmandan oluşur, bu katmanlar evrişim katmanı, havuzlama katmanı ve tam bağlantı katmanıdır (O’shea ve Nash, 2015). Şekil 2.2’de ESA’ların genel mimarisi verilmiştir.



Şekil 2.2. Evrimsel sinir ağları mimarisi

Evriřim ve havuzlama katmanları  zellik  ıkarıcı oluřturmak i in ađın ilk yarısını oluřturur. Tam bađlantı katmanları ile  ıkarılan  zellikler  zerinden sınıflandırma ve regresyon gibi g revler yerine getirilir (Zhao vd., 2024).

2.5. Duruř Tahmini

Duruř tahmini canluların iskeletlerini taklit eden anahtar noktaların tespit edilmesi ile anlık pozisyonlarının  ıkarımını yapmayı hedefleyen bir alandır. İskelet yapılarının daha anlaşılır olması ve g nl k hayatta insan bilgisayar etkileřimi, oyunlar, sanal ger eklik, spor analizleri, tıbbi yardım gibi alanlarda yaygın kullanılması sebebiyle insanlar (Dang vd., 2019; J. Wang vd., 2021)  zerinde sıklıkla kullanılmaktadır (Cao vd., 2017; Insafutdinov vd., 2016; Papandreou vd., 2017; Wei vd., 2016).

2.6. Hareket Tanıma

Temelde hareket tanıma g r nt ler  zerinden canlının ne yaptığının tespit edilmesini ama layan bir  alıřma alanıdır. Duruř tahmini ile benzer olarak  ođunlukla insanlar  zerinde  alıřmaların olduđu bu alanda tehlikeli olabilecek insan faaliyetlerini g zlemlemek i in g zetim sistemlerinde (W. Lin vd., 2008), otonom navigasyon sistemlerinde (Lu vd., 2020), eđitim, eđlence (Shotton vd., 2011) ve tıp alanlarında kullanılmaktadır (Pareek ve Thakkar, 2021). Bu  alıřmada kullanılan y ntem geređi hareket tanıma  alıřması duruř tahmininden yardım alınarak yapılmıřtır.

3. MATERYAL VE YÖNTEM

3.1. Materyal

3.1.1. Veri Seti

Veri kümesi Xu vd. tarafından yapılan çalışmada (Xu vd., 2017) kullanılmış ve makalede de belirtilmiş olduğu üzere açık bir şekilde paylaşılmış olan içerisinde Scoot (Kaçma), J-turn (J Dönüşü), C-Turn (C-Dönüşü), R-Turn (R-Dönüşü), Surface (Yüzeye Çıkma), Dive (Dibe Dalma), Zigzag (Zikzak), Thrash (Kıvrınma), Freeze (Donma) şeklinde ayrılmış olan 9 sınıfta hareket bulunmaktadır. Veri setindeki toplam görüntü sayısı 16528 olup bu görüntüler balık hareket halindeyken tepeden çekilmiş videolardan alınmış olan görüntülerdir.

Bu tez çalışmasında kullanılan yöntem gereği dibe dalma ve yüzeye çıkma hareketleri 2 boyut üzerinde anlaşılamayacağı ve görüntü sınıflandırma modelleri tarafından eğitimi negatif etkilemesi sebebiyle veri seti içerisinden çıkarılıp çalışmada hem duruş tahmini hem de hareket tanıma çalışması için kullanılmıştır. Ayrıca veri setinden sınıflar içerisindeki kesin olmayan hareketler çıkarılmıştır. Bu çıkarılmalar sonrasında veri setinde 15033 görüntü kalmış, hareketleri oluşturan karelerin bir araya getirilmesi sonucunda veri seti üzerinde veri artırımı uygulanmadan 387 eğitim 123 test görüntüsü olmak üzere 510 görüntü, veri artırımı ve veri seti dengesizliği giderilmesi işlemi sonrasında 1974 eğitim 123 test görüntüsü olmak üzere 2097 görüntü çalışmada kullanılmıştır. Görüntüler kullanılan görüntü sınıflandırma modellerinin önerilen girdi boyutuna göre 224x224, 299x299, 331x331 ve 380x380 boyutlarına yeniden boyutlandırılarak kullanılmıştır. Model eğitimi sırasında kaçma, j-dönüşü, c-dönüşü, r-dönüşü, yüzeye çıkma, dibe dalma, yüzeye çıkma, zikzak, kıvrınma ve donma hareketlerine sırasıyla 0, 1, 2, 3, 4, 5, 6, 7, 8 olmak üzere toplamda 9 etiket kullanılmış, dibe dalma ve yüzeye çıkma hareketlerinin çıkarıldığı durumlarda aynı sıra ile 0, 1, 2, 3, 4, 5, 6 olacak şekilde 7 etiket kullanılmıştır.

3.1.2. Performans Metrikleri

Makine öğrenmesi modelleri gibi derin öğrenme modellerinin de eğitim sonrasında değerlendirilmeleri gerekmektedir. Bu çalışmada model değerlendirmeleri yapılırken doğruluk, kesinlik, duyarlılık, f1-skor kullanılmıştır. Bu metriklerin hesaplanmasında karmaşıklık matrisi kullanılır. Şekil 3.1’de örnek karmaşıklık matrisi görülmektedir. Karmaşıklık matrisi kullanılarak doğruluk, kesinlik, duyarlılık ve f1-score metriklerinin

hesaplanma formülleri sırasıyla Denklem 3.1, Denklem 3.2, Denklem 3.3 ve Denklem 3.4'te, verilmiştir.

		Tahmin Edilen Sınıf	
		1	0
Asıl Sınıf	1	Gerçek Pozitif (GP)	Yanlış Negatif (YN)
	0	Yanlış Pozitif (YP)	Gerçek Negatif (GN)

Şekil 3.1. Örnek karmaşıklık matrisi (Özden, 2024)

Doğruluk, modelin doğru olarak tahmin ettiği verilerin tüm verilere oranıdır (Erdogan, 2023). Doğruluk metriği değerlendirme için kullanılabilmesine karşın, dengeli olmayan veri setleri için tek başına yeterli bir metrik değildir (Öğündür, 2024).

$$\text{Doğruluk} = \frac{GP + GN}{GP + GN + YP + YN} \quad (3.1)$$

Kesinlik değeri pozitif olarak tahmin edilen değerlerin içerisinde gerçekte kaç tanesinin pozitif olduğunu gösteren değerdir (Öğündür, 2024). Bu sebeple modelin pozitif sınıfı ne kadar doğru sınıflandırdığını ölçen metriktir (Erdogan, 2023).

$$\text{Kesinlik} = \frac{GP}{GP + YP} \quad (3.2)$$

Duyarlılık, gerçekte pozitif olan yani pozitif olarak tahmin etmemiz gereken örneklerin model sonucunda ne kadarını pozitif olarak tahmin ettiğimizin çıktısını veren metriktir (Erdogan, 2023; Öğündür, 2024).

$$\text{Duyarlılık} = \frac{GP}{GP + YN} \quad (3.3)$$

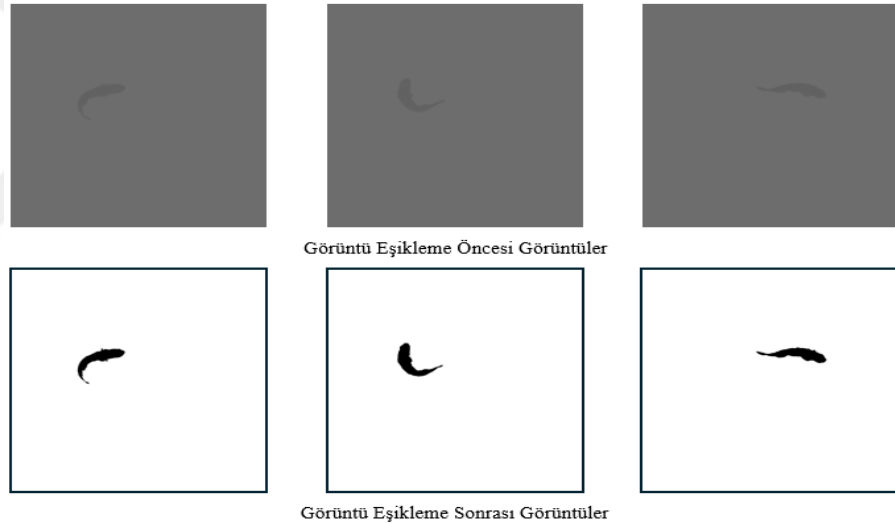
F1-skor, modelin sınıf bazında performansına odaklanarak modelin tahmin becerisini değerlendiren klasik değerlendirme metriklerine alternatif bir metriktir. Kesinlik ve duyarlılık metriklerini kullanarak bu değerlerden yeni bir değer ortaya çıkarır (Kundu, 2022).

$$F1 = \frac{2 \times \text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.4)$$

3.2. Yöntem

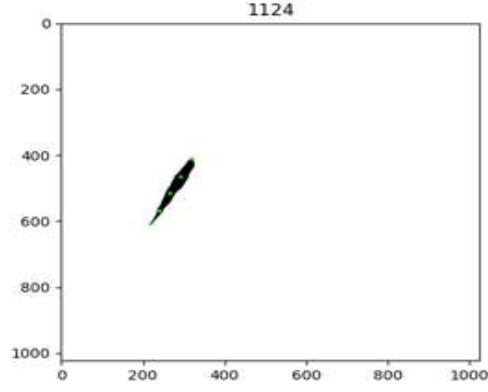
3.2.1. Görüntü Ön İşleme Teknikleri ve Veri Setinin Hazırlanması

Görüntü ön işleme, görüntü işleme yapılmadan önce görüntülerdeki gürültüyü (Eze ve Berkaya, 2024) ve bulanıklığı ortadan kaldırma, görüntülerin kontrastını, parlaklığını ve ayrıntılarını iyileştirme çalışmaları olarak kategorize edilebilir (Archana ve Jeevaraj, 2024). Bu işlem sayesinde nesne tespiti ve sınıflandırma gibi çalışmalarda daha doğru ve güvenilir sonuçlara ulaşabiliriz. Bu çalışmada görüntü ön işleme tekniği olarak görüntü eşikleme kullanarak duruş tahmini için anahtar nokta tespitinin daha güvenilir olması amaçlanmıştır. Eşikleme sonucunda görüntüler siyah beyaz hale getirilmiştir. Şekil 3.2’de veri seti içerisinde eşikleme işlemi uygulanmış örnekler görülmektedir.



Şekil 3.2. Görüntü eşikleme uygulanarak siyah beyaz yapılmış görüntüler

Görüntü eşikleme sonrasında elde edilen görüntüler YOLOv8 ile balığın anahtar noktalarının tespit edilmesi için kullanılmıştır. Şekil 3.3’te anahtar nokta tespiti sonrası örnek görüntü mevcuttur.



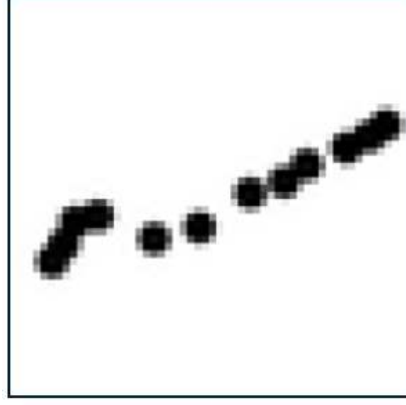
Şekil 3.3. Anahtar nokta tespiti

Anahtar baş, üç noktada gövde ve kuyruk noktaları tespit edilen balık görüntüsünün balığın yaptığı hareketin rotası belirlenmesi ve bu rota sayesinde transfer öğrenme yöntemleri kullanılarak görüntü sınıflandırması ile hareketinin tanınması için veri setinde belirli olan aralıklardaki görüntülerdeki tüm balıkların baş noktaları tek bir düzleme aktarılmıştır. Kullanılan noktaların düzleme aktarılmış örneği Şekil 3.4’te mevcuttur.



Şekil 3.4. Hareketi oluşturan balık baş noktalarının düzleme aktarılması

Noktaların düzleme aktarılmasından sonra veri setinin modeller tarafından daha anlaşılır olması amacıyla siyah olan noktalar referans alınarak fazlalık beyaz alanlar kırpılarak ortadan kaldırılmış, modellerin en iyi çalışma çözünürlüklerine göre yeniden boyutlandırılmıştır. Yapılan bu işlem sonrası ortaya çıkan görüntü örneği Şekil 3.5’te verilmiştir.



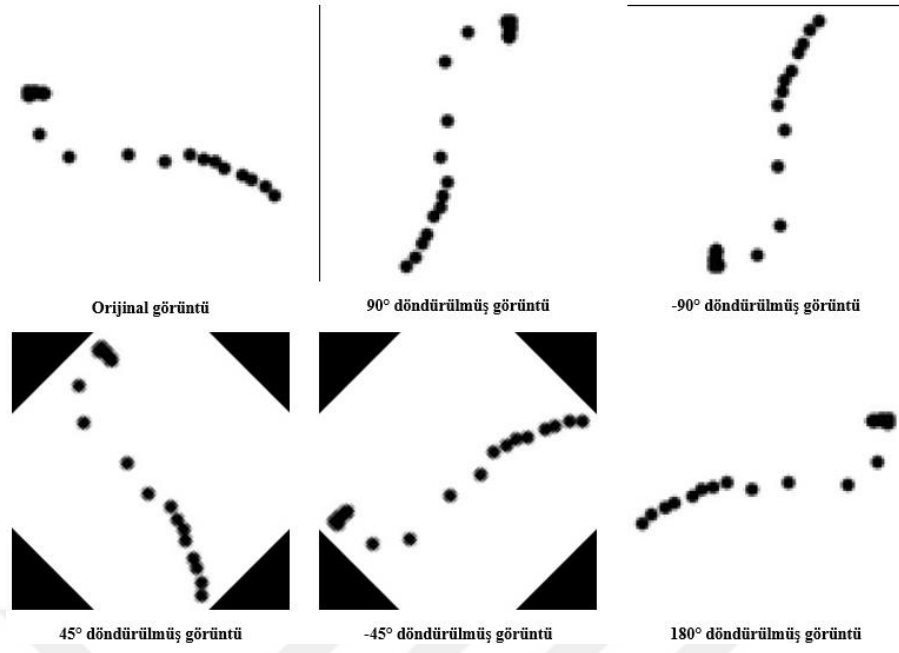
Şekil 3.5. Model eğitimi için kırılıp yeniden boyutlandırılmış görüntü

Son olarak veri setindeki görüntülerin sayısının az olması, veri setinin dengesiz olması gibi sebeplerden dolayı model performansında problem yaşanmaması adına veri seti üzerinde veri artırımı ile görüntü sayıları artırılmış ve dengelenmiştir.

3.2.2. Veri Artırma

Derin öğrenme modelleri görüntü sınıflandırma alanında muazzam bir başarı göstermiştir (Naveed vd., 2024) ancak derin öğrenme modellerinde milyonlarca parametre bulunması sebebiyle eğitilirken veri setinde yeterli görüntü olmaması model performansında sorunlara yol açabilir (Kebaili vd., 2023; Naveed vd., 2024). Bu durumda görüntü ön işleme tekniklerinden birisi olan veri artırma teknikleri kullanılabilir. Veri artırma mevcut veri örneklerinin üzerinde değişiklikler yaparak yüksek kaliteli yapay veriler üretilmesine olanak sağlayan bir tekniktir. Bu tekniklerden yararlanarak içerisinde az veri bulunan ya da dengesiz veri setlerinin iyileştirilmesi sonucunda derin öğrenme modellerinde büyük ölçüde performans artışı sağlanabilir (Zeng, 2024).

Döndürme sırasında istenmeyen piksel kaybı olmaması ve modelin değerlendirme başarısını farklı perspektiflerden oluşacak görüntüler artırabileceği için (Zeng, 2024) bu çalışmada veri artırma yöntemi olarak veri kümesindeki sınıflardaki veri azlığı ve dengesizliği dikkate alınarak döndürme kullanılmıştır. Döndürme işlemi gerçekleştirilirken $+90^\circ$, -90° , $+45^\circ$, -45° ve 180° olmak üzere 5 farklı derece göz önünde bulundurulmuştur. Şekil 3.6'da bu döndürme işlemleri sonucunda oluşan örnek görüntüler gösterilmiştir. Çizelge 3.1'de yedi sınıflı Çizelge 3.2'de dokuz sınıflı eğitim veri setindeki görüntülerin veri artırımı öncesi ve sonrası görüntü sayıları verilmiştir.



Şekil 3.6. Örnek veri artırımı görüntüleri

Çizelge 3.1. Yedi sınıflı veri setinin görüntü sayıları

Sınıf	Veri Artırımı Öncesi Görüntü Sayısı (Eğitim)	Veri Artırımı Sonrası Görüntü Sayısı (Eğitim)	Test Veri Setindeki Görüntü Sayısı
0 (Kaçma)	53	282	21
1 (J-Dönüşü)	94	282	13
2 (C-Dönüşü)	51	282	8
3 (R- Dönüşü)	88	282	18
4 (Zikzak)	31	282	17
5 (Kıvrانma)	34	282	18
6 (Donma)	36	282	28
Toplam	387	1974	123

Çizelge 3.2. Dokuz sınıflı veri setinin görüntü sayıları

Sınıf	Veri Artırımı Öncesi Görüntü Sayısı (Eğitim)	Veri Artırımı Sonrası Görüntü Sayısı (Eğitim)	Test Veri Setindeki Görüntü Sayısı
0 (Kaçma)	53	282	21
1 (J-Dönüşü)	94	282	13
2 (C-Dönüşü)	51	282	8
3 (R- Dönüşü)	88	282	18
4 (Yüzeye Çıkma)	28	282	24
5 (Dibe Dalma)	11	282	26
6 (Zikzak)	31	282	17
7 (Kıvrانma)	34	282	18
8 (Donma)	36	282	28
Toplam	426	2538	173

3.2.3. Nesne Tespiti ve Anahtar Nokta Tespiti

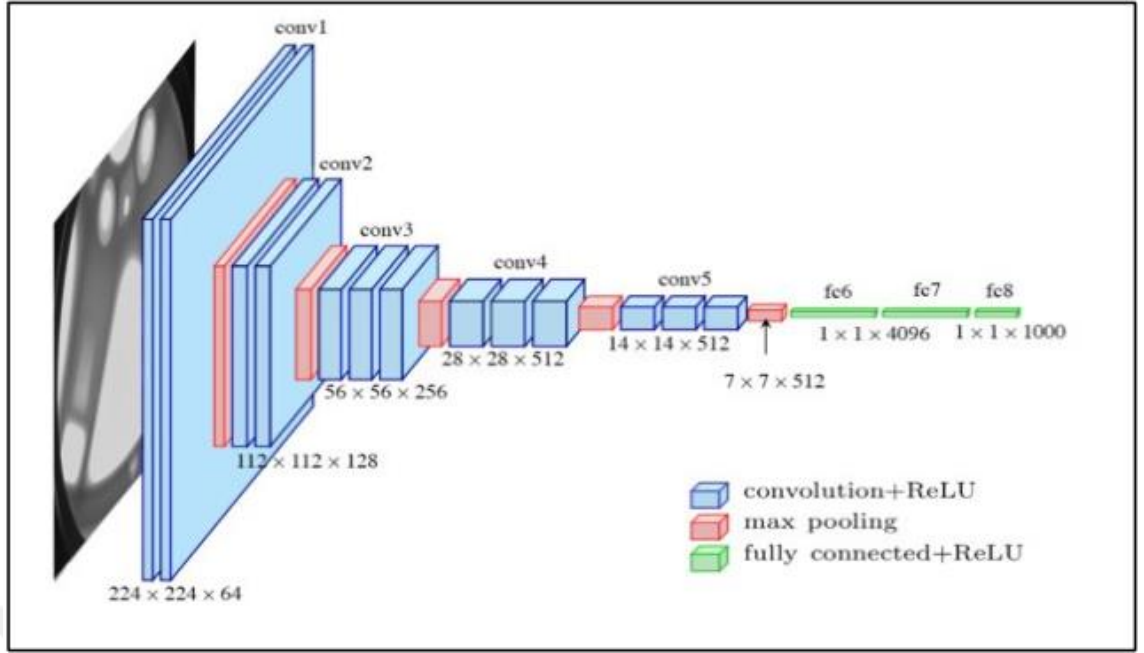
Duruş tahmini çalışmalarının yapılabilmesi için daha önceden etiketlenmiş bir veri seti yardımı ile görüntüdeki nesnenin tespit edilip işaretlenmiş anahtar noktalara göre bir duruş tahmini algoritmasının eğitilmesi gerekmektedir. YOLO serisi (Bochkovskiy vd., 2020; C. Li vd., 2022; C.-Y. Wang vd., 2023; H. Wang vd., 2022) son yıllarda görsel anlama alanında en popüler algoritmalarından birisi olup, bilgisayarlı görüde çeşitli alanlarda kullanılmaktadır. Önceki nesiller ile karşılaştırıldığı zaman duruş tahmini çalışmalarında doğruluk ve hız açısından YOLOv8 daha iyi bir performans göstermektedir (C. Dong ve Du, 2024). Bu sebeple bu çalışmada duruş tahmini ve anahtar noktaların tespiti için YOLOv8 kullanılmıştır.

3.2.4. Transfer Öğrenme Yöntemleri

Transfer öğrenme belirli bir amaç için geliştirilen bir modelin, farklı ancak ilgili yeni bir amaç için oluşturulan modele başlangıç noktası olarak kullanılması yöntemidir. Özellikle veri setinin kısıtlı olduğu durumlarda (Jaiswal vd., 2021; Weiss vd., 2016) transfer öğrenme yöntemleri daha başarılı sonuçlar elde etmektedir (Iman vd., 2023). Modeller ön bir eğitimden geçtiği için daha az veri ve daha kısa eğitim süresi ile daha iyi performans verebilmektedir (Farahani vd., 2020).

Bu çalışmada kullanılan veri seti büyük bir veri seti olmadığı için daha önceden ImageNet veri seti ile eğitilmiş ESA'lerden ResNet50, ResNet101, NasNetLarge, Inception v3, DenseNet169, Xception, VGG 16 ve EfficientNet-B4 modelleri kullanılmıştır. Bu model seçimleri ile ResNet50 ve ResNet101 gibi aynı modelin farklı derinlikteki versiyonları karşılaştırılabilecek, VGG 16 ve NasNetLarge gibi yüksek parametrelili, EfficientNet-B4, Xception ve DenseNet169 gibi düşük parametrelili modellerin performansları değerlendirilebilecek, Inception v3 ve Xception gibi birbiri ile ilişkili modellerin performansları karşılaştırılabilecektir.

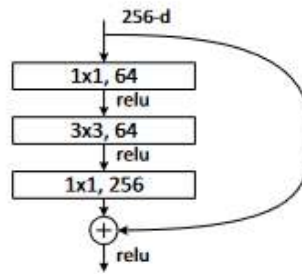
2014 yılında ImageNet yarışmasında Oxford Üniversitesi Geometri grubu tarafından önerilen ve dereceye giren (Qassim vd., 2018) VGG-16 modeli Simonyan ve Zisserman tarafından yapılan çalışmada (Simonyan ve Zisserman, 2014) tanıtılmış, 13 evrişim katmanı, 3 tam bağlantı katmanı ve 5 adet maksimum havuzlama katmanı içeren ESA mimarisidir. 224x224 çözünürlüğünde RGB görüntüleri kabul eder. Tüm katmanları 3x3 filtrelerle sahiptir. Evrişim ve tam bağlantı katmanlarında Rectified Linear Unit (ReLU) aktivasyon fonksiyonu kullanır. Şekil 3.7'de VGG-16 mimarisi görülmektedir.



Şekil 3.6. VGG-16 mimarisi (Younis vd., 2022)

Model karmaşıklığına bakıldığında diğer ESA'lara göre daha basit bir yapısı olmasına rağmen büyük veri setlerinde de başarılı sonuçlar verebilmektedir (Simonyan ve Zisserman, 2014).

ResNet50 ve ResNet101 He vd. (He vd., 2016) tarafından yapılan çalışmada 18, 34, 50, 101 ve 152 katmanlı olarak 5 farklı şekilde önerilmiş olan temeli artık bloklara dayanan bir ESA'dır. 50, 101 ve 152 katmanlı olan katmanlar "bottleneck" adını verdikleri aktivasyon fonksiyonu olarak ReLU kullanan bloklardan oluşmakta olup şekil 3.7'de verilmiştir. Sınıflandırma (Targ vd., 2016; Zhang, 2022), nesne tespiti (Du vd., 2021; Loey vd., 2021), anlamsal bölümlendirme (Gómez-Flores ve de Albuquerque Pereira, 2020) ve video analizi (Manjula ve Lakshmi, 2022) gibi alanlarda sıklıkla kullanılan ResNet ağları önerilen boyut olarak 224×224 çözünürlüğünde görüntüler kullanılır.



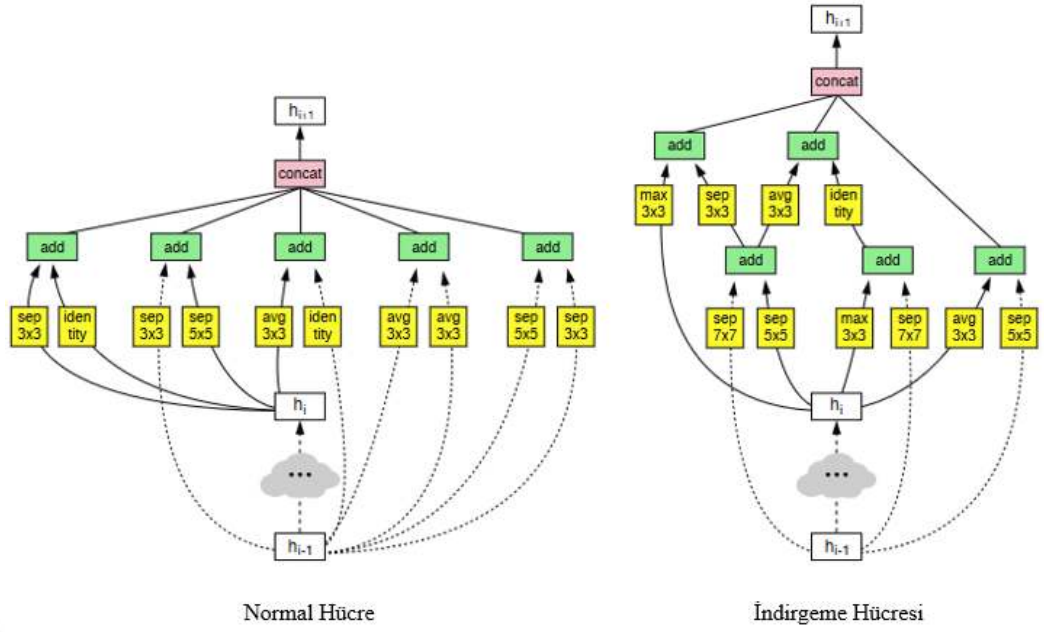
Şekil 3.7. "Bottleneck" bloğu

ResNet50 ağında sırasıyla 1 7x7 evrişim katmanı, 1 maksimum havuzlama katmanı, 3, 4, 6, 3 olmak üzere 16 bottleneck bloğu (Mukti ve Biswas, 2019), 1 ortalama havuzlama katmanı ve 1 tam bağlantı katmanı mevcuttur. ResNet101 ağına ResNet50'den farkı ise 6 bottleneck bloğunun olduğu kısımda 23 bottleneck bloğu kullanarak toplamda 33 blok yapısı oluştuğu için 101 katmandan oluşmasıdır. Şekil 3.8'de bahsedilen mimariler verilmiştir. Katman sayıları fazla olmasına rağmen en yüksek karmaşıklığa sahip olan ResNet152 ağı bile 11,3 milyar flop'a sahipken VGG16-19 15,3/16,9 milyar flop'a sahiptir. Bu sebeple ResNet ağları için VGG ağlarından daha az karmaşıklığa sahip diyebiliriz (He vd., 2016).

50-layer	101-layer
7×7, 64, stride 2	
3×3 max pool, stride 2	
$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
average pool, 1000-d fc, softmax	

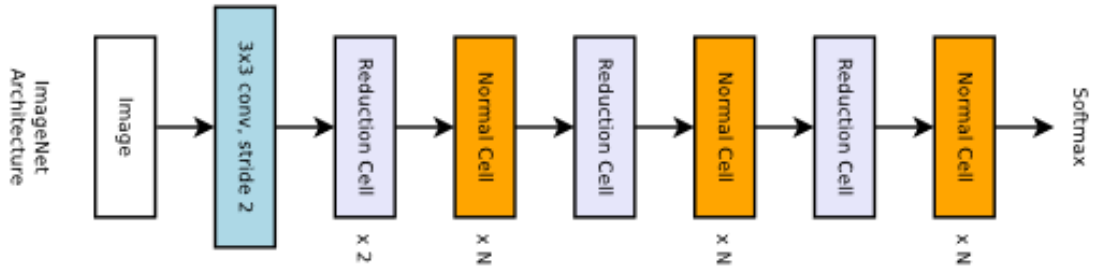
Şekil 3.8. ResNet50 ve ResNet101'in mimari yapıları (Simonyan ve Zisserman, 2014)

NasNet Google çalışanı olan Zoph vd. (Zoph vd., 2018) tarafından önerilmiş olan görüntü sınıflandırma (Shrivastava vd., 2021; G. Yang vd., 2020) ve diğer bilgisayarlı görü alanlarında kullanılmak üzere dizayn edilmiş bir modeldir. Mimarisi şekil 3.9'da da görüldüğü üzere birbirini tekrarlayan normal ve indirgeme hücrelerinden (reduction cells) oluşmaktadır (Zoph vd., 2018). Normal hücreler görüntünün boyutlarını indirgmeden özellik çıkarımı yaparken, indirgeme hücreleri hesaplama yükünü adım adım azaltmak için özellik çıkarımını görüntünün boyutunu indirgeyerek yapar.



Şekil 3.9. Normal hücre ve indirgeme hücresinin yapısı (Zoph vd., 2018)

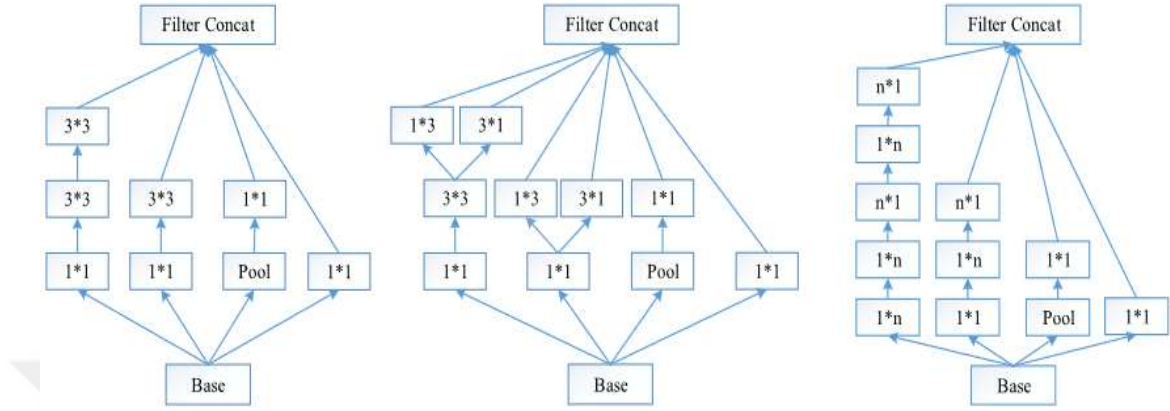
Bu çalışmada kullanılan NasNetLarge modelinin önerilen görüntü girdi çözünürlüğü 331x331 olmalıdır (Shrivastava vd., 2021). Bu görüntü şekil 3.10’da basitleştirilmiş NasNet mimarisinde görülmekte olan 3x3 bir evrişim katmanından geçtikten sonra, tekrarlayan hücre yapılarında özellik haritaları çıkarıldıktan sonra sınıflandırma işlemi yapılır.



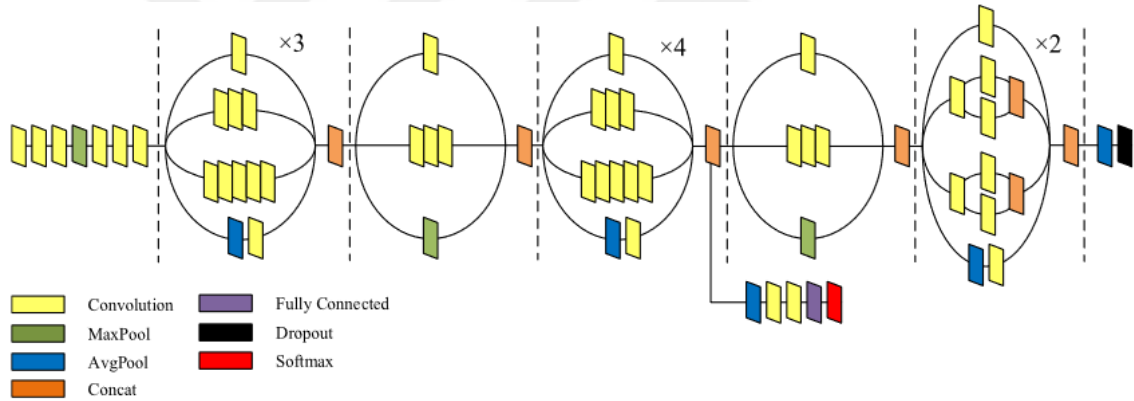
Şekil 3.10. Basitleştirilmiş NasNet mimarisi (Zoph vd., 2018)

Szegedy vd. tarafından yapılan çalışmada (Szegedy vd., 2016) önerilen Inceptionv3 modeli görüntü sınıflandırma (C. Wang vd., 2019; Xia vd., 2017), nesne tespiti (Sun vd., 2020) vb. alanlarda kullanılmaktadır. Inception modelinin getirdiği yeniliklerden birisi evrişim katmanlarında 5x5 gibi filtreler yerine iki tane 3x3 gibi küçük filtreler kullanmasıdır. Bu sebeple hem performanstan kayıp olmamakta hem de hesaplama maliyeti düşmektedir. Bunun yanında geleneksel simetrik nxn evrişim katmanları yerine

modelde $1 \times n$ ve $n \times 1$ boyutunda asimetrik evrişim katmanları modelin daha verimli olması için tercih edilmiştir (Szegedy vd., 2016). Bu katmanların modüller içerisindeki kullanımı Şekil 3.11’de Inceptionv3 modelinin genel mimarisi Şekil 3.12’de verilmiştir. Modelde önerilen görüntü çözünürlüğü 299×299 ’dur.



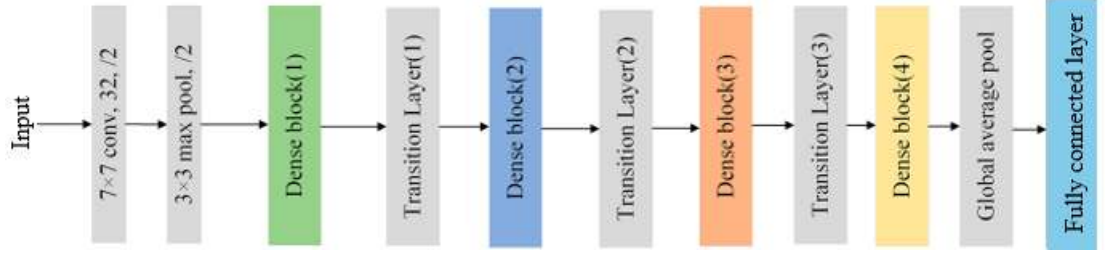
Şekil 3.11. Inception modülleri (N. Dong vd., 2020)



Şekil 3.12. Inception v3 model mimarisi (N. Dong vd., 2020)

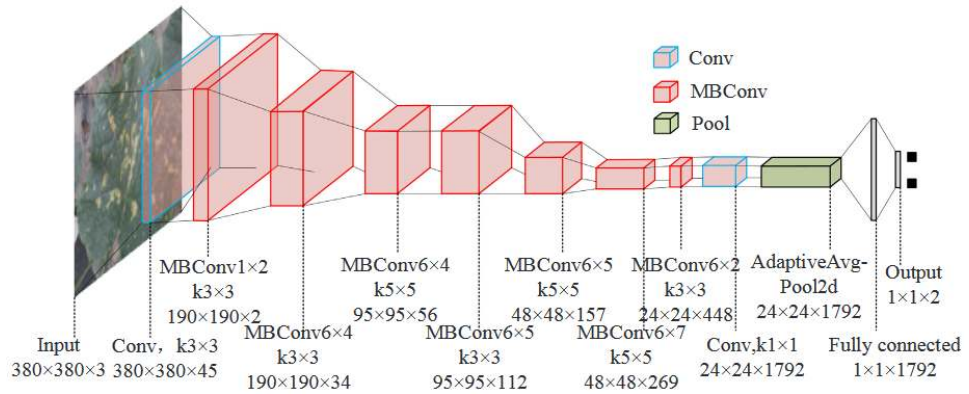
DenseNet169, DenseNet model ailesinin 169 katmanlı versiyonudur. Katmanlar arasındaki yenilikçi bağlantı sistemi ile bilinen DenseNet Huang vd. tarafından yapılan çalışmada (G. Huang vd., 2017) önerilmiştir. Şekil 3.13’te mimarisi görülen DenseNet ağı adını aldığı dense blokları her katmanın diğer katmanlara ileri beslemeli olarak bağlandığı bloklardır. Bu sayede katmanlar arasındaki bilgi akışı en üst düzeye çıkarılır ve daha verimli bir model oluşturulmuş olur. Model içerisinde girdi özellik haritalarının azaltılması için dense katmanlarında 1×1 darboğaz katmanları bulunur. Dense bloklarının arasında 1×1 evrişim katmanı ile 2×2 ortalama havuzlama katmanı bulunur bu yapıya geçiş katmanları denir. Geçiş katmanları ile ağıın dense katmanları arasındaki bilgi akışı en üst düzeyde sağlanmaya çalışılmaktadır. Modelin önerilen girdi çözünürlüğü

224x224'tür. DenseNet169 modeli görüntü sınıflandırma (Manna vd., 2021; Q. Zhang vd., 2021) alanındaki başarısından dolayı bu çalışmada tercih edilen modellerden birisi olmuştur.



Şekil 3.13. DenseNet169 mimarisi (Jiang vd., 2021)

EfficientNet-B4 Tan ve Le tarafından yapılan çalışmada (Tan ve Le, 2019) önerilen doğruluk ve hesaplama verimliliği arasındaki denge ile dikkat çeken görüntü sınıflandırma (Atıla vd., 2021; D. Zhang vd., 2020) çalışmalarında tercih edilen bir ESA modelidir. EfficientNet modelleri derinliği ve çözünürlüğü eşit şekilde ölçeklendiren bir ölçeklendirme yöntemi kullanır ve bu yöntem sayesinde hesaplama verimliliği korunurken model performansının artırılması amaçlanır. Mobile Inverted Bottleneck Convolution (MBConv) EfficientNet'in temel birimidir. MBConv parametre sayısını ve hesaplama maliyetini önemli ölçüde azaltmaya yardımcı olur. Girdi çözünürlüğü için model önerisi 380x380'dir. Şekil 3.14'te model mimarisi görülmektedir.

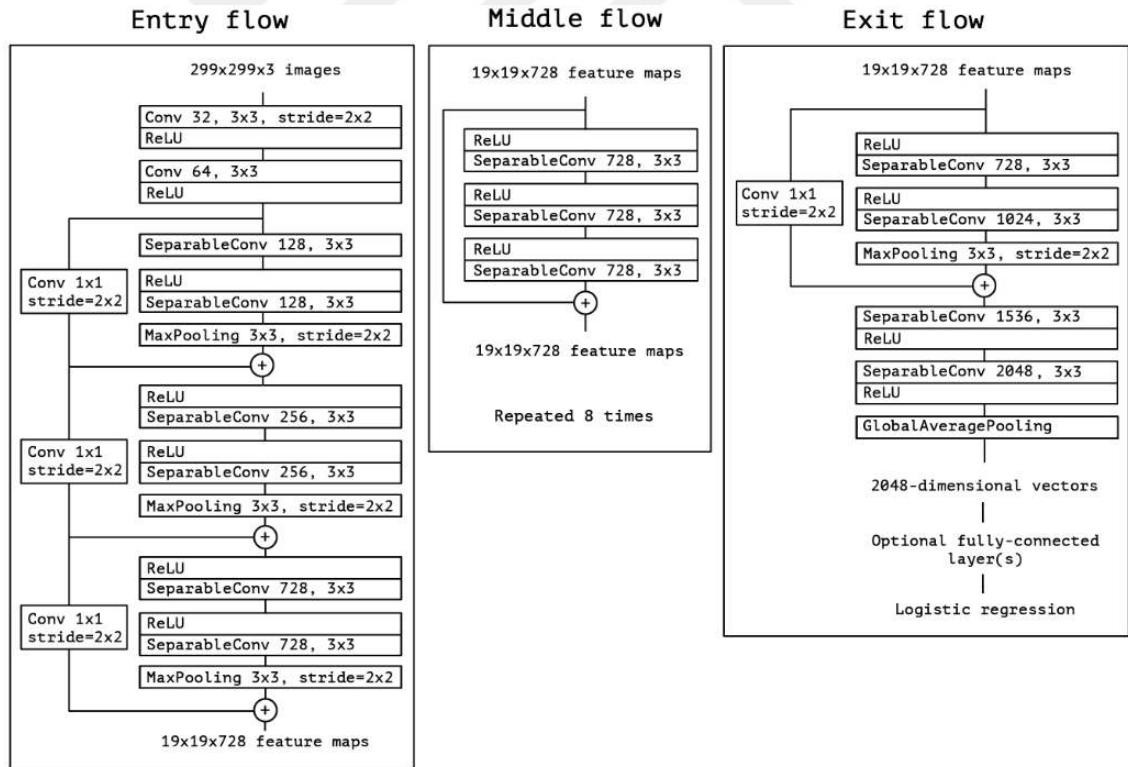


Şekil 3.14. EfficientNet-B4 model mimarisi (P. Zhang vd., 2020)

Adı “Extreme Inception”dan gelen Xception modeli (Kılıçarslan vd., 2024), Inception’da bulunan normal veya derin evrişimler yerine derinlemesine ayrılabilir evrişimler (ayrılabilir evrişimler) kullanan Inception mimarisinin bir çeşidi olan ESA’dır (Isık vd., 2023). Chollet tarafından yapılan çalışmada (Chollet, 2017) önerilmiştir.

Yapılan tasarım değişimi sayesinde hem parametre kullanımı hem hesaplama maliyeti açısından daha verimli bir model oluşturulmuşken, modelin performans kaybına uğramaması da amaçlanmıştır (Chollet, 2017). ImageNet sınıflandırma işlerinde mükemmelliğini kanıtlayan Xception, derin öğrenmede etkili ve verimli model tasarımlarının gelişimine önemli ölçüde katkı sağlayarak çeşitli görevlere uygulanabilen temel bir model haline gelmiştir (Chollet, 2017). Xception modelinin görüntü sınıflandırma (Benbrahim ve Behloul, 2021; Xizhi Wu vd., 2020), nesne tespiti (Dhillon ve Verma, 2020; Zou vd., 2023), yüz tanıma gibi alanlarda kullanılabilmesi sebebiyle bu çalışmada da kullanılmıştır.

Şekil 3.15'te mimarisi görülen Xception modeli entry flow (giriş akışı), middle flow (orta akış), exit flow (çıkış akışı) olmak üzere 3 parçadan oluşur. Giriş kısmında 2 normal 6 ayrılabilir evrişim katmanları, orta kısımda 8 kere tekrar eden 3'lü ayrılabilir evrişim katmanı ve çıkış kısmında ise 4 ayrılabilir evrişim katmanı ile toplamda 36 katmandan oluşmaktadır. Katmanlarla beraber aktivasyon fonksiyonu olarak ReLU kullanılmıştır. Modelin önerilen görüntü çözünürlüğü 299x299'dur (Chollet, 2017).



Şekil 3.15. Xception model mimarisi

3.2.5. Hiper-Parametre Optimizasyonu

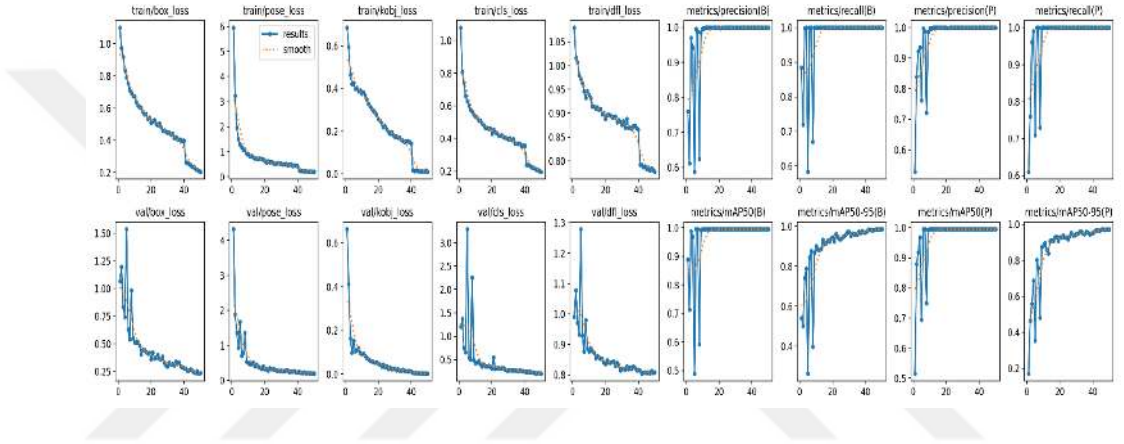
Hiper-parametreler verilerden öğrenilmeyen, eğitim öncesi ayarlanan öğrenme hızı (learning rate) ve katman başına birim sayısı (units) gibi parametrelerdir. Hiper-parametre optimizasyonu bir modelin performansını artırmak için en uygun parametreleri bulma yöntemidir (Yu ve Zhu, 2020). Bu yöntem için ızgara arama (grid search), rastgele arama (random search), bayesian optimizasyonu gibi teknikler kullanılmaktadır.

Bu çalışmada hiper-parametre optimizasyonu için Li vd. tarafından yapılan çalışmada (L. Li vd., 2018) önerilen hyperband algoritması kullanılmıştır. Bu algoritma birkaç epochta birçok parametrenin denenmesini sağlar (L. Li vd., 2018). Başlangıçta küçük epoch sayıları ile parametreleri test eder en düşük başarıya sahip olan parametreleri eleyerek epoch sayılarını artırarak eğitime devam eder. Algoritmaya parametre olarak girilen epoch sayısına ulaştığında performansı en yüksek olan parametreleri seçerek modelin başarımını artıracak parametrelerin seçimi için yardımcı olur. Bu çalışmada dense katmanın için ayarlanması gereken ‘units’ ve ‘learning rate’ parametreleri için hyperband kullanılarak hiper-parametre optimizasyonu yapılmıştır.

4. ARAŞTIRMA BULGULARI VE TARTIŞMA

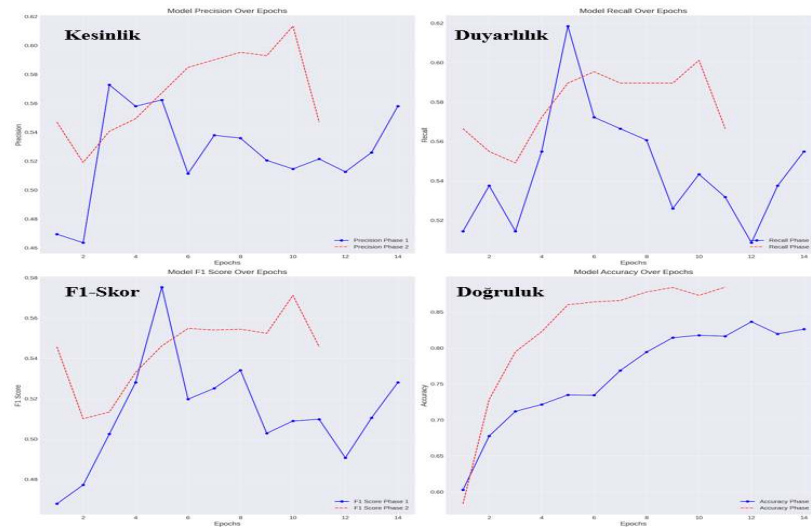
4.1. Araştırma Bulguları

Çalışmada hareket tanıma çalışmasını 2 boyutlu düzlem üzerinde yapabilmek için balıkların YOLOv8 ile anahtar noktalarının tespitine ilişkin grafikler Şekil 4.1’de verilmiştir. Grafikler dikkate alındığında loss (kayıp) ile ilgili grafiklerde düşüş olduğu, diğer başarı metrikleri ile ilgili grafiklerde sürekli bir yükseliş olduğu görülmektedir. Bu grafiklere bakarak anahtar noktalarının doğru bir şekilde tespit edildiğinden bahsedebiliriz.

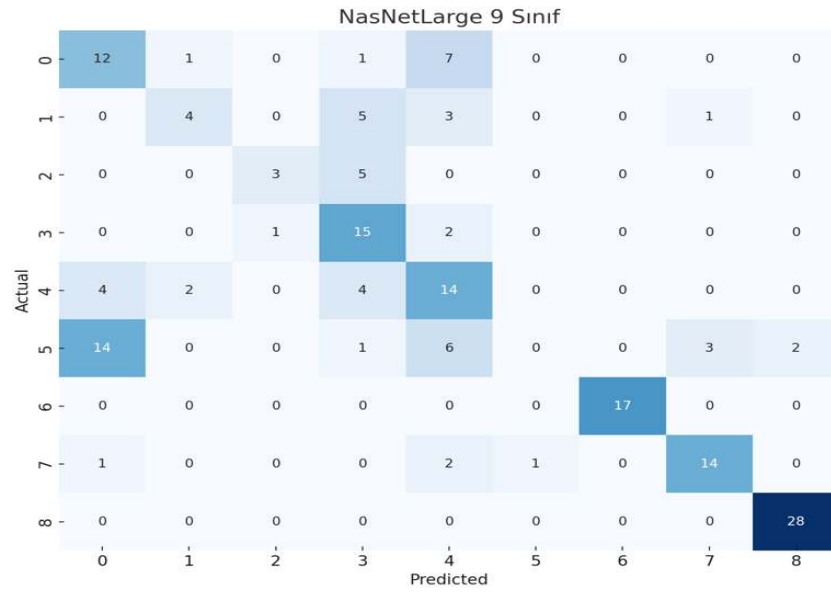


Şekil 4.1. YOLOv8 sonuçları

Çalışmada veri seti içerisindeki tüm sınıflar yalnızca NasNetLarge ile kullanılmıştır. Şekil 4.2 ve Şekil 4.3’te NasNetLarge modeli için 9 sınıf kullanılarak yapılan eğitimin sonuçları görülmektedir. Önerilen girdi boyutu 331x331 kullanılmıştır.



Şekil 4.2. NasNetLarge 9 sınıf değerlendirme grafikleri



Şekil 4.3. NasNetLarge 9 sınıf karmaşıklık matrisi

Çizelge 4.1’de karmaşıklık matrisi göz önüne alınarak metriklerin detaylı analizi görülmektedir.

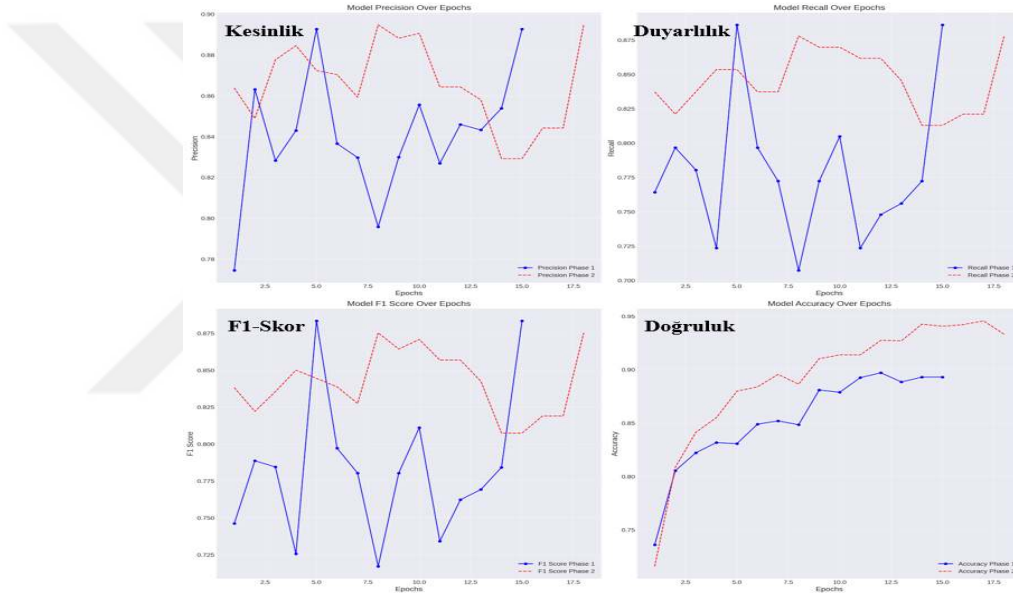
Çizelge 4.1. NasNetLarge 9 sınıflı sınıflandırma raporu

Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	0,3871	0,5714	0,4615	21
1 (J- Dönüşü)	0,5714	0,3077	0,4000	13
2 (C- Dönüşü)	0,7500	0,3750	0,5000	8
3 (R- Dönüşü)	0,4839	0,8333	0,6122	18
4 (Yüzeye Çıkma)	0,4118	0,5833	0,4828	24
5 (Dibe Dalma)	0	0	0	26
6 (Zikzak)	1	1	1	17
7 (Kıvrınma)	0,7778	0,7778	0,7778	18
8 (Donma)	0,9333	1	0,9655	28
Ağırlıksız Ortalama	0,5906	0,6054	0,5778	-
Ağırlıklı Ortalama	0,5623	-	0,5753	-
Doğruluk	0,6185			

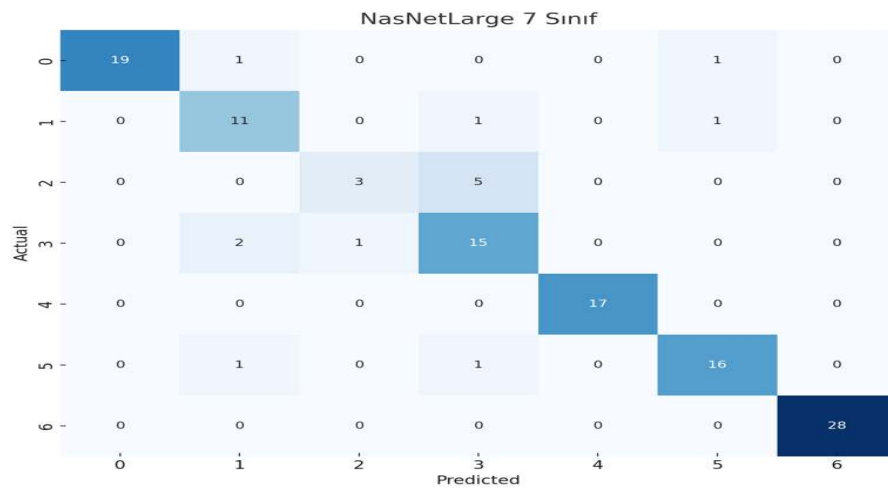
Bütün modeller 40’ar epoch eğitilmiş ancak modellerde eğitim uydurmanın (overfitting) önüne geçebilmek ve daha stabil tahmin sonuçlarına ulaşabilmek için iki aşamalı ve erken durdurma (early stopping) kullanılarak yapılmıştır. 9 sınıflı NasNetLarge için hiper-parametre optimizasyonu sonucunda en iyi ‘units’ ve ‘learning rate’ değeri sırasıyla 896 ve 0,001 olarak bulunmuştur. Grafikler ve çizelgelerden de görülebildiği ve çalışmada bölüm 3.1.1’de değinildiği gibi görüntülerin iki boyutlu düzlem üzerine aktarıldığı durumda yüzeye çıkma (4. sınıf) ve dibe dalma (5.sınıf)

hareketleri performansta düşümlere yol açmaktadır. 9 sınıflı NasNetLarge modeli ile deneme yapılmış, grafiklerde doğruluk oranında düzgün bir artış olmasına rağmen 4. ve 5. Sınıfların tahmini bozmasından dolayı test doğruluk oranı %61,85 bulunmuştur. 9 sınıflı değerlendirmede test doğruluk oranı düşüklüğü sebebi ile geriye kalan modellerde sadece 7 sınıflı değerlendirme yapılmıştır.

NasNetLarge modelinin 7 sınıf üzerinde eğitim yapılması sonucunda oluşan grafikler Şekil 4.3'te karmaşıklık matrisi Şekil 4.4'te görülmektedir. 9 sınıflı eğitim gibi 331x331 çözünürlüğünde görüntüler girdi olarak kullanılmıştır. 7 sınıflı NasNetLarge için hiper-parametre optimizasyonu sonucunda en iyi 'units' ve 'learning rate' değeri sırasıyla 1024 ve 0,001 olarak bulunmuştur.



Şekil 4.4. NasNetLarge 7 sınıf değerlendirme grafikleri



Şekil 4.5. NasNetLarge 7 sınıf karmaşıklık matrisi

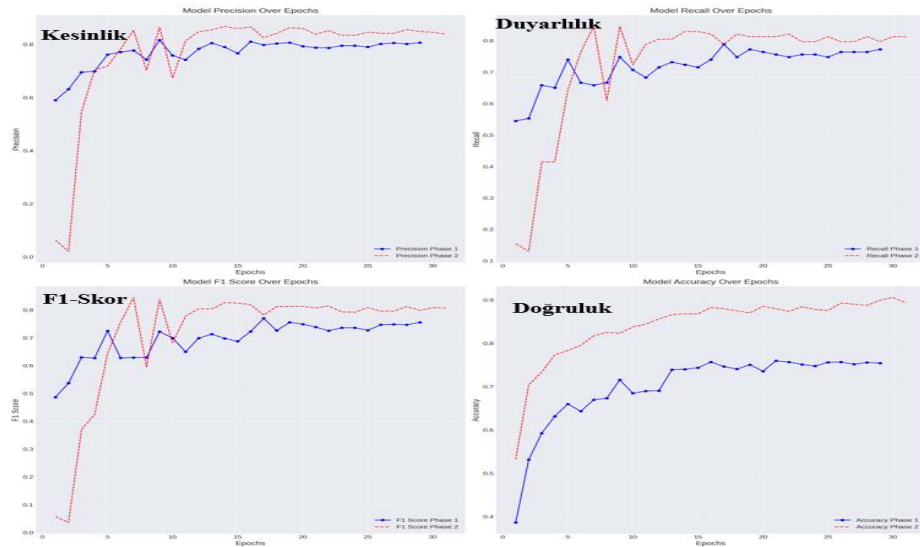
Test sonrası detaylı metrik değerleri Çizelge 4.2’de verilmiştir.

Çizelge 4.2. NasNetLarge 7 sınıflı sınıflandırma raporu

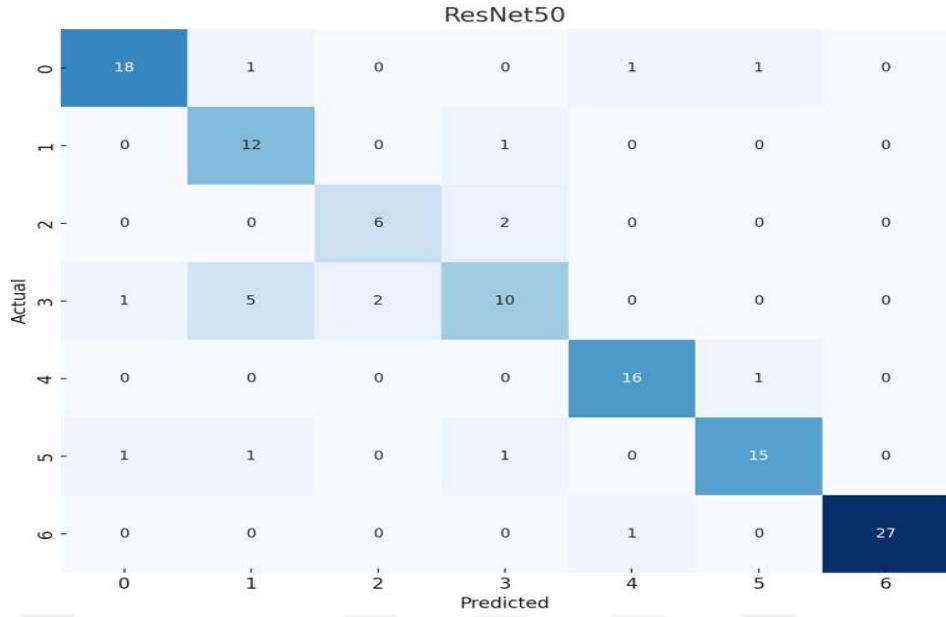
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	1	0,9048	0,9500	21
1 (J- Dönüşü)	0,7333	0,8462	0,7857	13
2 (C- Dönüşü)	0,7500	0,3750	0,5000	8
3 (R- Dönüşü)	0,6818	0,8333	0,7500	18
4 (Zikzak)	1	1	1	17
5 (Kıvrınma)	0,8889	0,8889	0,8889	18
6 (Donma)	1	1	1	28
Ağırlıksız Ortalama	0,8649	0,8354	0,8392	-
Ağırlıklı Ortalama	0,8927	-	0,8834	-
Doğruluk	0,8862			

Grafiklerden ve karmaşıklık matrisinden de görülebildiği gibi kullanılan veri seti ve yöntem için 7 sınıflı bir değerlendirme yapılmasının daha doğru olduğundan bahsedilebilir. İki aşamalı bir eğitim yapıldığı için ikinci eğitim daha kararlı olmuş ve erken durdurma sayesinde uydurmanın önüne geçilmeye çalışılmıştır. NasNetLarge modeli 7 sınıflı test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %88,62, %89,27, %88,62, %88,34 oranlarına ulaşılmıştır.

ResNet50 modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 256 ve 0,001 olarak bulunmuş. Model eğitiminde görüntüler 224x224 çözünürlüğünde kullanılmıştır. Şekil 4.6’da modele ait grafikler Şekil 4.7’de karmaşıklık matrisi görülmektedir.



Şekil 4.6. ResNet50 değerlendirme grafikleri



Şekil 4.7. ResNet50 karmaşıklık matrisi

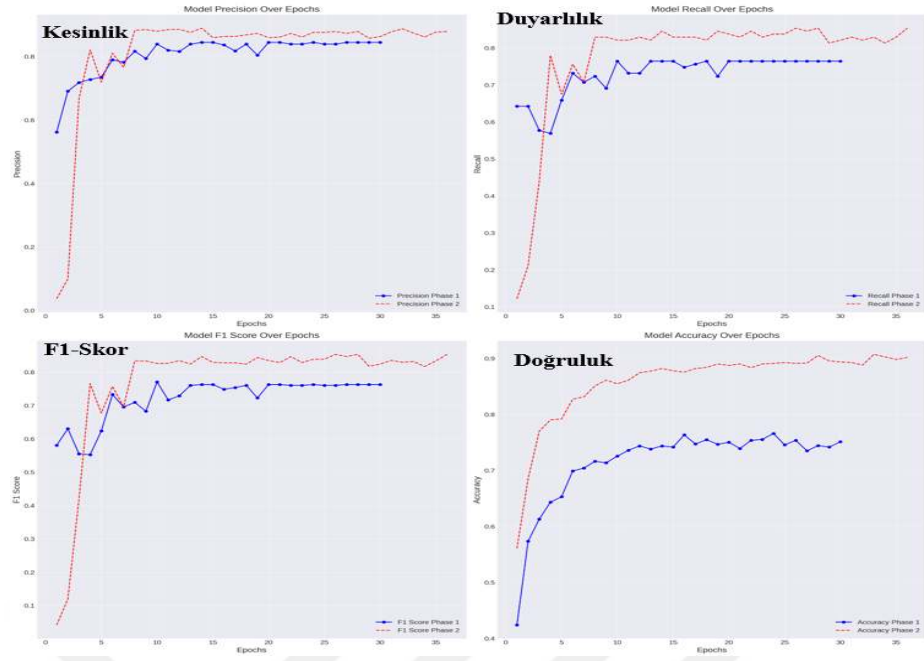
Test sonrası metriklerin detaylı analizi Çizelge 4.3'te verilmiştir.

Çizelge 4.3. ResNet50 sınıflandırma raporu

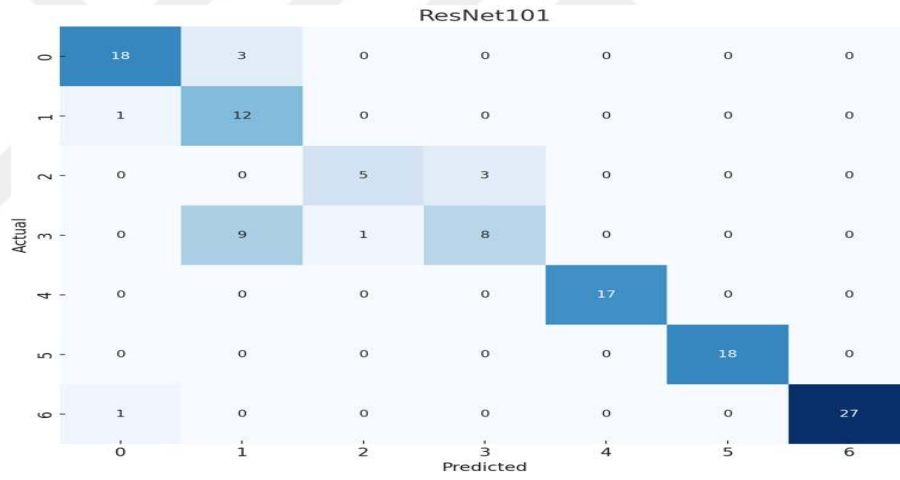
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	0,9000	0,8571	0,8780	21
1 (J- Dönüşü)	0,6316	0,9231	0,7500	13
2 (C- Dönüşü)	0,7500	0,7500	0,7500	8
3 (R- Dönüşü)	0,7143	0,5556	0,6250	18
4 (Zikzak)	0,8889	0,9412	0,9143	17
5 (Kıvrınma)	0,8824	0,8333	0,8571	18
6 (Donma)	1	0,9643	0,9818	28
Ağırlıksız Ortalama	0,8239	0,8321	0,8223	-
Ağırlıklı Ortalama	0,8533	-	0,8447	-
Doğruluk	0,8455			

İki aşamalı bir eğitim yapıldığı için ikinci eğitim daha kararlı olmuş ve performansta artış meydana gelmiştir. ResNet50 modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %84,55, %85,33, %84,55, %84,47 oranlarına ulaşılmıştır.

ResNet101 modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 640 ve 0,001 olarak bulunmuş. Model eğitiminde görüntüler 224x224 çözünürlüğünde kullanılmıştır. Şekil 4.8’de modele ait grafikler Şekil 4.9’da karmaşıklık matrisi görülmektedir. Test sonrası metriklerin analizleri Çizelge 4.4’te verilmiştir.



Şekil 4.8. ResNet101 değerlendirme grafikleri



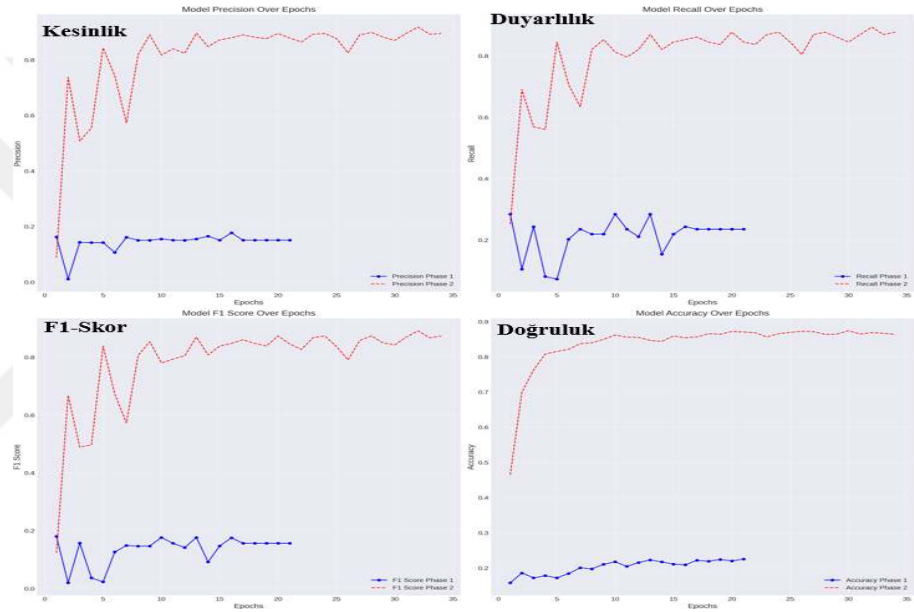
Şekil 4.9. ResNet101 karmaşıklık matrisi

Çizelge 4.4. ResNet101 sınıflandırma raporu

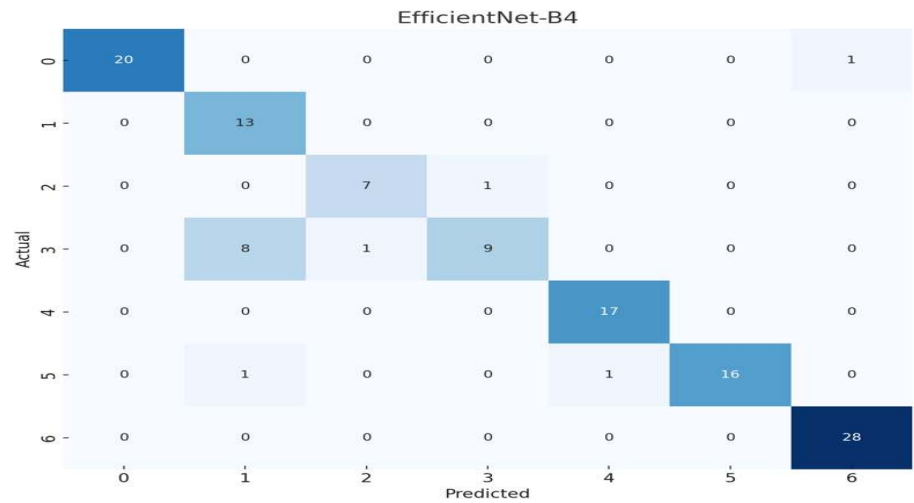
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	0,9000	0,8571	0,8780	21
1 (J- Dönüşü)	0,5000	0,9231	0,6486	13
2 (C- Dönüşü)	0,8333	0,6250	0,7143	8
3 (R- Dönüşü)	0,7273	0,4444	0,5517	18
4 (Zikzak)	1	1	1	17
5 (Kıvrınma)	1	1	1	18
6 (Donma)	1	0,9643	0,9818	28
Ağırlıksız Ortalama	0,8515	0,8306	0,8249	-
Ağırlıklı Ortalama	0,8793	-	0,8537	-
Doğruluk	0,8537			

Model grafikleri incelendiğinde metriklerde sapmanın olmadığı düzgün bir artış görülmektedir. İki aşamalı bir eğitim yapıldığı için ikinci eğitim daha kararlı olmuş ve performansta artış meydana gelmiştir. ResNet101 modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %85,37, %87,93, %85,37, %85,37 oranlarına ulaşılmıştır.

EfficientNet-B4 modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 384 ve 0,0001 olarak bulunmuş. Model eğitiminde görüntüler 380x380 çözünürlüğünde kullanılmıştır. Şekil 4.11’de modele ait grafikler Şekil 4.10’da karmaşıklık matrisi görülmektedir. Test sonrası metriklerin analizleri Çizelge 4.5’te verilmiştir.



Şekil 4.10. EfficientNet-B4 değerlendirme grafikleri



Şekil 4.11. EfficientNet-B4 karmaşıklık matrisi

Çizelge 4.5. EfficientNet-B4 sınıflandırma raporu

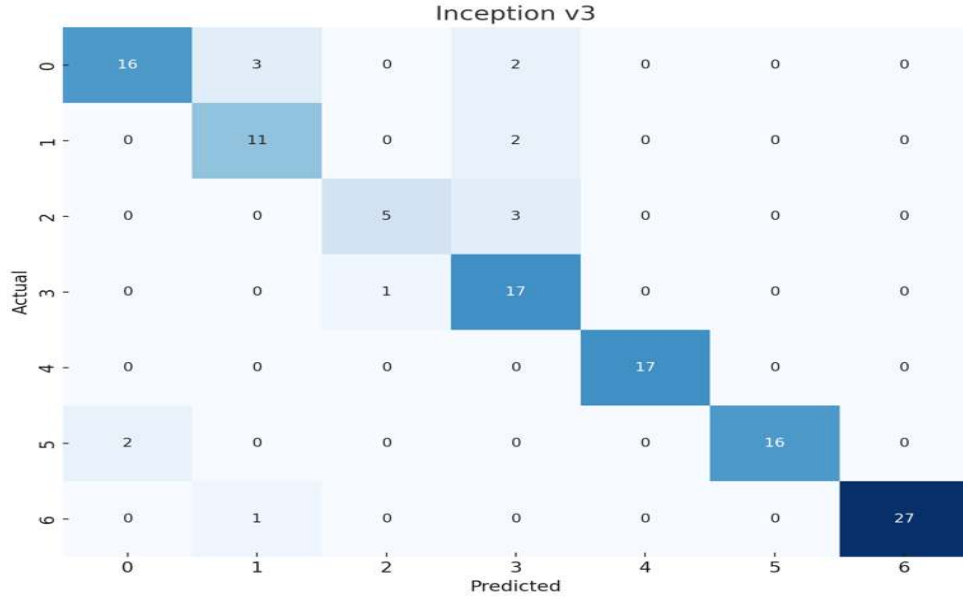
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	1	0,9524	0,9756	21
1 (J- Dönüşü)	0,5909	1	0,7429	13
2 (C- Dönüşü)	0,8750	0,8750	0,8750	8
3 (R- Dönüşü)	0,9000	0,5000	0,6429	18
4 (Zikzak)	0,9444	1	0,9714	17
5 (Kıvrınma)	1	0,8889	0,9412	18
6 (Donma)	0,9655	1	0,9825	28
Ağırlıksız Ortalama	0,8966	0,8880	0,8759	-
Ağırlıklı Ortalama	0,9185	-	0,8917	-
Doğruluk	0,8943			

Model grafikleri incelendiğinde birinci aşamada modelin katmanları dondurulmuşken eğitim metriklerinin düşük olduğu, son katman harici eğitime dahil edilen ikinci aşamada ise başlangıçta sapmalar olsa da orta ve son bölümlerde artışın düzgün ve sapmaların düşük seviyede olduğu gözlemlenmektedir. İki aşamalı bir eğitim yapıldığı için ikinci eğitim daha kararlı olmuş ve performansta artış meydana gelmiştir. EfficientNet-B4 modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %89,43, %91,85, %89,43, %89,17 oranlarına ulaşılmıştır.

Inception v3 modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 768 ve 0,001 olarak bulunmuş. Model eğitiminde görüntüler 299x299 çözünürlüğünde kullanılmıştır. Şekil 4.12’de modele ait grafikler Şekil 4.13’te karmaşıklık matrisi görülmektedir.



Şekil 4.12. Inception v3 değerlendirme grafikleri



Şekil 4.13. Inception v3 karmaşıklık matrisi

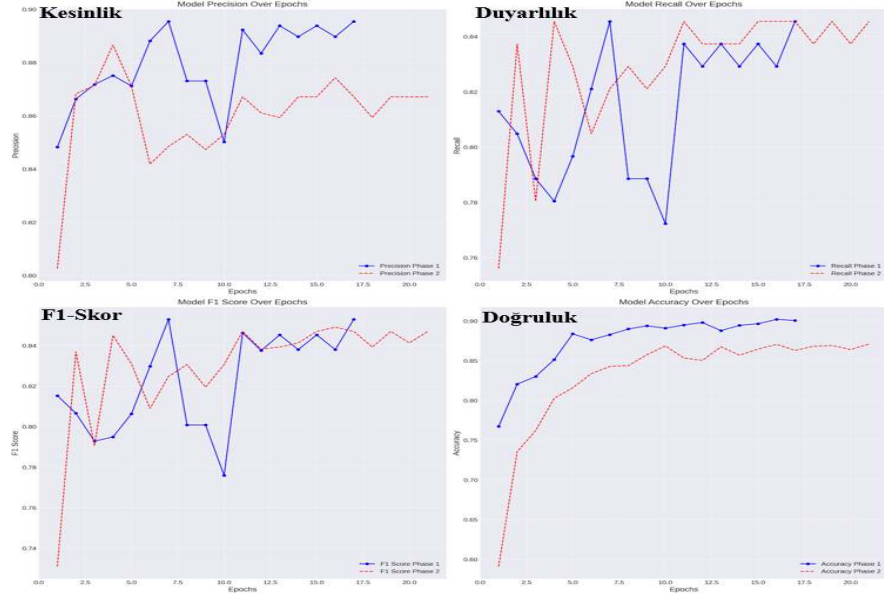
Test sonrası metriklerin analizleri Çizelge 4.6’da verilmiştir.

Çizelge 4.6. Inception v3 sınıflandırma raporu

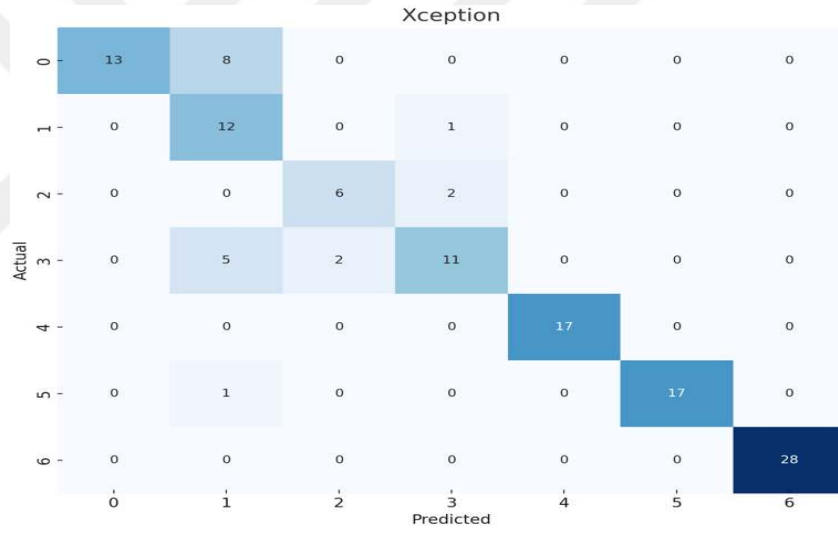
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	0,8889	0,7619	0,8205	21
1 (J- Dönüşü)	0,7333	0,8462	0,7857	13
2 (C- Dönüşü)	0,8333	0,6250	0,7143	8
3 (R- Dönüşü)	0,7083	0,9444	0,8095	18
4 (Zikzak)	1	1	1	17
5 (Kıvrınma)	1	0,8889	0,9412	18
6 (Donma)	1	0,9643	0,9818	28
Ağırlıksız Ortalama	0,8806	0,8615	0,8647	-
Ağırlıklı Ortalama	0,8993	-	0,8875	-
Doğruluk	0,8862			

Model grafikleri incelendiğinde birinci aşamada model genelinde dalgalanmalar olduğu ancak ikinci aşamada bu dalgalanmaların düzelip daha güvenilir bir sonuç verdiği görülmektedir. İkinci aşamada metrik değerleri daha düşük olsa da bu modelde ikinci aşama eğitimin daha düzgün olmasını sağlamıştır. Inception v3 modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %88,62, %89,93, %88,62, %88,75 oranlarına ulaşılmıştır.

Xception modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 896 ve 0,001 olarak bulunmuş. Model eğitiminde görüntüler 299x299 çözünürlüğünde kullanılmıştır. Şekil 4.14’te modele ait grafikler Şekil 4.15’te karmaşıklık matrisi ve Çizelge 4.7’de metriklerin analizi görülmektedir.



Şekil 4.14. Xception değerlendirme grafikleri



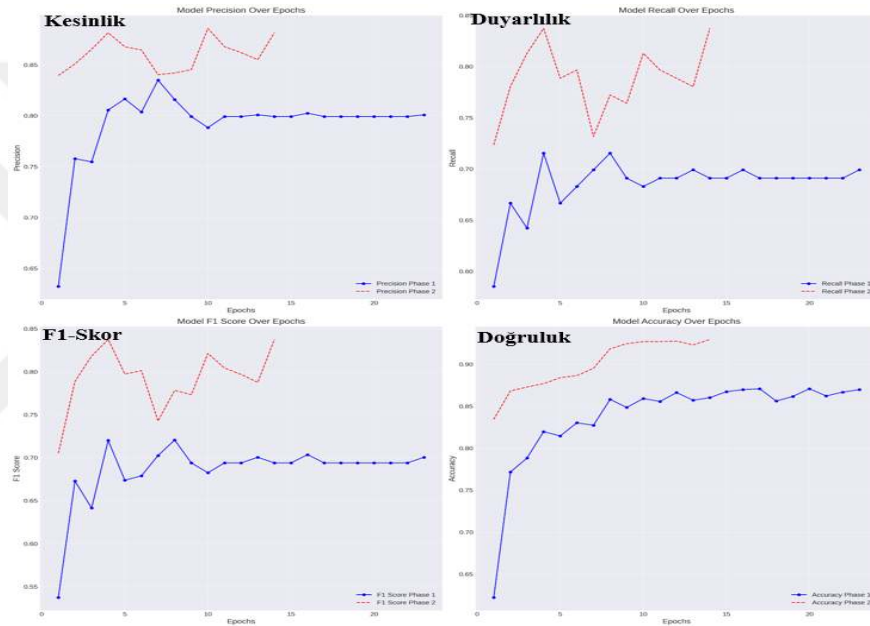
Şekil 4.15. Xception karmaşıklık matrisi

Çizelge 4.7. Xception sınıflandırma raporu

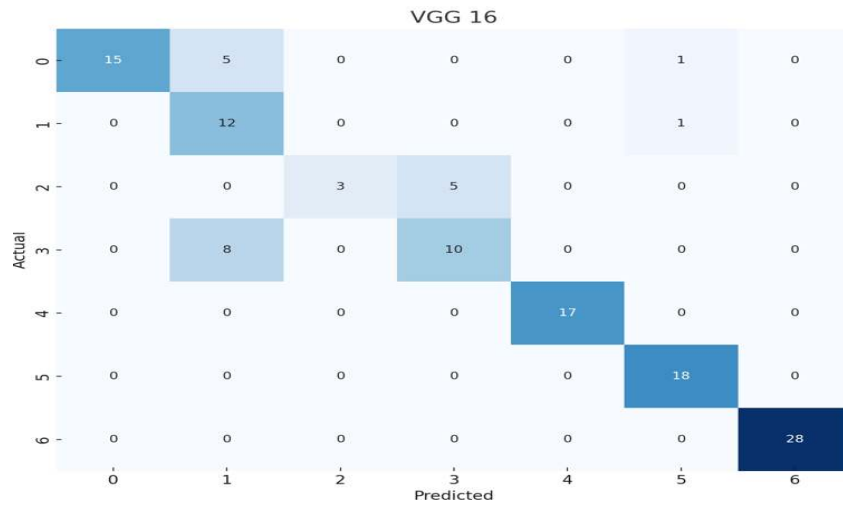
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	1	0,6190	0,7647	21
1 (J- Dönüşü)	0,4615	0,9231	0,6154	13
2 (C- Dönüşü)	0,7500	0,7500	0,7500	8
3 (R- Dönüşü)	0,7857	0,6111	0,6875	18
4 (Zikzak)	1	1	1	17
5 (Kıvrınma)	1	0,9444	0,9714	18
6 (Donma)	1	1	1	28
Ağırlıksız Ortalama	0,8568	0,8354	0,8270	-
Ağırlıklı Ortalama	0,8955	-	0,8530	-
Doğruluk	0,8455			

Model grafikleri incelendiğinde modelin iki aşamasında da dalgalanmalar olduğu ancak ikinci aşamada bu dalgalanmaların daha düzenli olduğu görülmektedir. İki aşamalı eğitim bu modelde eğitimin daha sağlıklı olmasını sağlamıştır. Xception modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %84,55, %85,68, %84,55, %85,30 oranlarına ulaşılmıştır.

VGG 16 modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 384 ve 0,001 olarak bulunmuş. Model eğitiminde görüntüler 224x224 çözünürlüğünde kullanılmıştır. Şekil 4.16’da modele ait grafikler Şekil 4.17’de karmaşıklık matrisi ve Çizelge 4.8’de metriklerin analizi görülmektedir.



Şekil 4.16. VGG-16 değerlendirme grafikleri



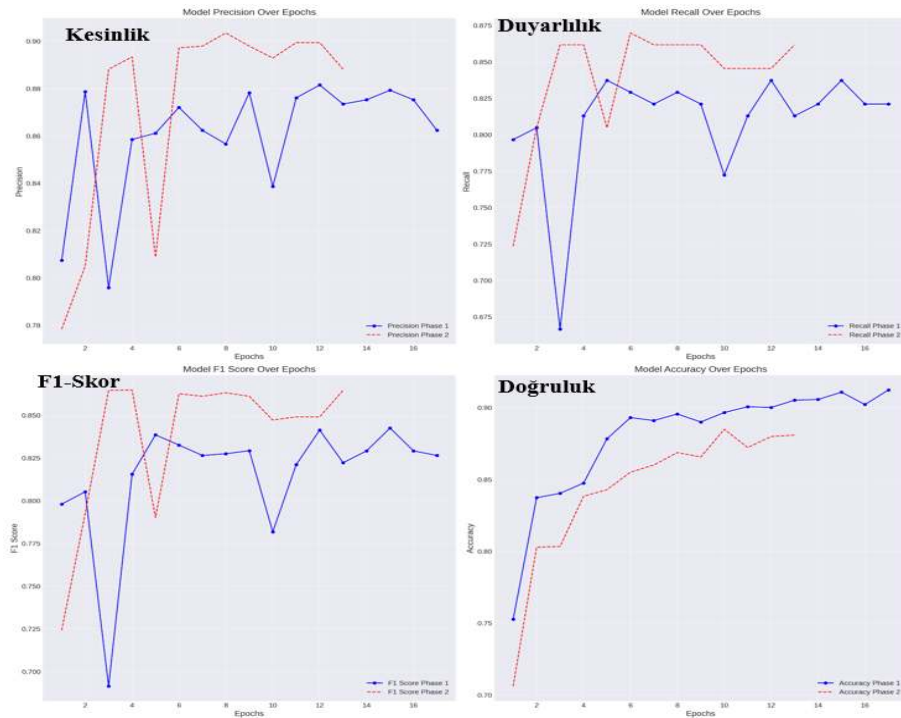
Şekil 4.17. VGG 16 karmaşıklık matrisi

Çizelge 4.8. VGG 16 sınıflandırma raporu

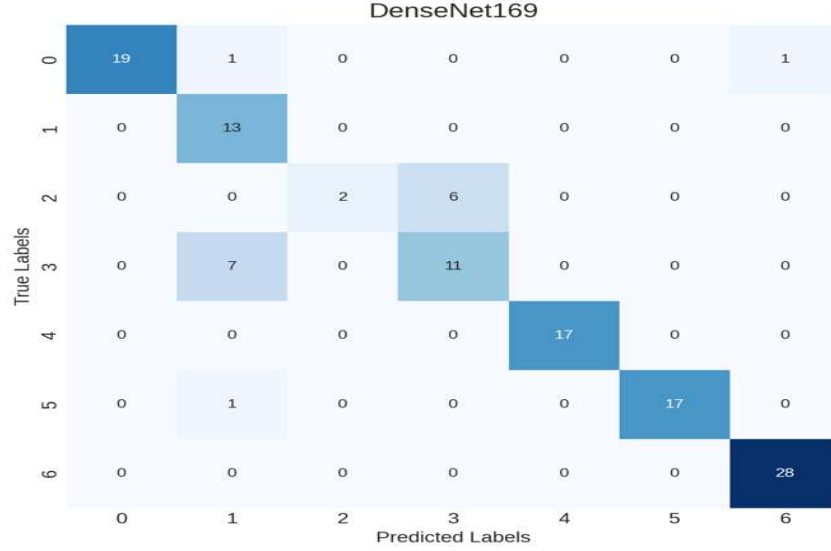
Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	1	0,7143	0,8333	21
1 (J- Dönüşü)	0,4800	0,9231	0,6316	13
2 (C- Dönüşü)	1	0,3750	0,5455	8
3 (R- Dönüşü)	0,6667	0,5556	0,6061	18
4 (Zikzak)	1	1	1	17
5 (Kıvrınma)	0,9000	1	0,9474	18
6 (Donma)	1	1	1	28
Ağırlıksız Ortalama	0,8638	0,7954	0,7948	-
Ağırlıklı Ortalama	0,8816	-	0,8377	-
Doğruluk	0,8374			

Model grafikleri incelendiğinde modelin iki aşamasında da dalgalanmalar olduğu ancak ikinci aşamada modelin başarısı arttığı görülmektedir. İki aşamalı eğitim bu modelde tahmin doğruluğunun artmasını sağlamıştır. VGG 16 modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %83,74, %88,16, %83,74, %83,77 oranlarına ulaşılmıştır.

DenseNet169 modelinin eğitiminde hiper-parametre optimizasyonu sonrasında ‘units’ ve ‘learning rate’ değerleri sırasıyla 1024 ve 0,001 olarak bulunmuş. Model eğitiminde görüntüler 224x224 çözünürlüğünde kullanılmıştır. Şekil 4.18’de modele ait grafikler Şekil 4.19’da karmaşıklık matrisi ve Çizelge 4.9’da metriklerin analizi görülmektedir.



Şekil 4.18. DenseNet169 değerlendirme grafikleri



Şekil 4.19. DenseNet169 karmaşıklık matrisi

Çizelge 4.9. DenseNet169 sınıflandırma raporu

Sınıf	Kesinlik	Duyarlılık	F1- Skor	Ağırlık
0 (Kaçma)	1	0,9048	0,9500	21
1 (J- Dönüşü)	0,5909	1	0,7429	13
2 (C- Dönüşü)	1	0,2500	0,4000	8
3 (R- Dönüşü)	0,6471	0,6111	0,6286	18
4 (Zikzak)	1	1	1	17
5 (Kıvrınma)	1	0,9444	0,9714	18
6 (Donma)	0,9655	1	0,9825	28
Ağırlıksız Ortalama	0,8862	0,8158	0,8108	-
Ağırlıklı Ortalama	0,8973	-	0,8627	-
Doğruluk	0,8699			

Model grafikleri incelendiğinde modelin iki aşamasında da dalgalanmalar olduğu ancak ikinci aşamada dalgalanmaların seviyesi azaldığı görülmektedir. İki aşamalı eğitim bu modelde güvenilirliğinin artmasını sağlamıştır. DenseNet169 modeli test veri setindeki sınıfların ağırlıkları dikkate alındığında, test sonucunda doğruluk, kesinlik, duyarlılık ve f1-skor metriklerinde sırasıyla %86,99, %89,73, %86,99, %86,27 oranlarına ulaşılmıştır.

Çizelge 4.10’da modellerin performans karşılaştırmaları verilmiştir. Model karşılaştırması yapıldığında EfficientNet-B4 modelinin %89,43 doğruluk, %91,85 kesinlik, %89,43 duyarlılık, %89,17 f1-skor oranlarıyla en başarılı model olduğu görülmektedir.

Çizelge 4.10. Modellerin performans karşılaştırması

Model Adı	Doğruluk	Kesinlik	Duyarlılık	F1-Skor
NasNetLarge (9)	0,6185	0,5623	0,6054	0,5753
NasNetLarge (7)	0,8862	0,8927	0,8354	0,8834
ResNet50	0,8455	0,8533	0,8321	0,8447
ResNet101	0,8537	0,8793	0,8306	0,8537
VGG 16	0,8374	0,8816	0,7954	0,8377
Inception v3	0,8862	0,8993	0,8615	0,8875
Xception	0,8455	0,8955	0,8354	0,8530
DenseNet169	0,8699	0,8973	0,8158	0,8627
EfficientNet-B4	0,8943	0,9185	0,8880	0,8917

Çizelge 4.11. Literatür karşılaştırması

Yapılan Çalışma	Başarı Yüzdesi
(Xu vd., 2017)	%91,33
(Måløy vd., 2019)	%80
(G. Wang vd., 2021)	%95,79
Tez çalışması	%89,43

4.2. Tartışma

Bölüm 4.1. Çizelge 4.10 dikkate alındığında EfficientNet-B4 modeli en yüksek performansı vermiştir. Birbirine benzer modeller arası performans karşılaştırması yapılacak olursa ResNet101 modeli ResNet50 modelinden daha fazla performans göstermiş, Xception modeli Inception v3 modelinin gerisinde kalmıştır. Kullanılan ESA modellerinde çalışmadaki veri setinin 9 sınıf performansının düşük olduğu, gerekli sınıflar çıkarıldığında modellerden daha iyi performans alındığı görülmektedir. Modellerde kullanılan iki aşamalı eğitim her modelde aynı etkiyi göstermese de genel anlamda modelin performansını ya da tahmin güvenilirliğini artırmaktadır. Modellerin karmaşıklık matrisleri incelendiğinde test sırasında 2. ve 3. sınıf etiketli görüntülerin tahminlerinde problem yaşadığı görülmüştür. Performans sonuçlarına bakıldığında modellerdeki parametre sayılarının fazla ya da az olmasının model başarısına direkt bir etki yaratmadığı görülmektedir.

Bölüm 4.1. Çizelge 4.11. dikkate alındığında yapılmış olan tez çalışması literatürdeki önemli balık hareket tanıma çalışmalarına kıyasla rekabetçi bir doğruluk oranı ortaya koymuştur.

5. SONUÇ VE ÖNERİLER

Balıkların besin değerinin yüksek olması, nüfus artışına bağlı olarak balığa olan talebin artması ve gelecekte daha da artacak olması sebebiyle balıklar insan yaşamında önemli bir noktadadır ancak balıklar üzerinde yapılan çalışmalarda birtakım sınırlılıklar mevcuttur. Nesnenin içinde yer aldığı su kütesinin derinliği, ışık alıp-almaması (Alaba vd., 2022), nesnenin arka ve ön planının hareketlilik durumu, suyun bulanıklığı ve düşük görüntü kalitesi gibi faktörler su altındaki nesneyi belirlemede zorlaştıracı etmenlerdir. Statik arka plan koşullarında arka plan durağan kalırken, dinamik arka plan koşullarında hem arka hem de ön planının hareketli olması durumunda nesneleri birbirinden ayırt etmede zorluk çıkabilmektedir. Yani, hem arka planın hem de ön plandaki nesnenin ayırt edilmesi dinamik koşullarda statik koşullara kıyasla çok daha zordur (Pagire vd., 2024).

Su altı görüntüleri, normal görüntü görüntülemesinden farklıdır. Bunun nedeni ise ışığın farklı dalga boylarının iletişim/geçişme esnasında farklı enerji zayıflama/güç yitirme oranlarına sahip olmasıdır. Dalga boyu ne kadar uzunsa, zayıflama o kadar hızlı olur. Örneğin, en uzun dalga boyuna sahip kırmızı ışık daha hızlı bozunurken, mavi ve yeşil ışık göreceli olarak yavaş bozunur. Bu nedenle su altı görüntüleri çoğunlukla mavi-yeşil eğilimlidir (B. Liu vd., 2024).

Bahsedilen zorluklar ve bu zorluklar sebebiyle meydana gelen veri seti azlığından dolayı literatürde balıkların üzerinde yapılan duruş tahmini ve hareket tanıma çalışmalarının sayısı çok düşüktür. Bu çalışmayı yaparak hem literatüre bir katkı vermek hem de bu çalışmalara farklı bir bakış açısı kazandırmak amaçlanmıştır.

Çalışmada 16528 görüntülü 9 sınıflı hazır bir veri setinin (Xu vd., 2017) dibe dalma ve yüzeye çıkma hareketlerini bulunduran sınıfları çıkarılarak kullanılmıştır. Literatürdeki hareket tanıma çalışmalarının neredeyse tamamı balık görüntüsü ya da balık videoları üzerinden yapılan çalışmalardır. Bu çalışmada farklı bir bakış açısı oluşturmak için hareket tanıma çalışması ile duruş tahmini çalışması balık görüntülerindeki hareketlerin 2 boyutlu düzleme aktarılması için kullanılmıştır. Görüntü işleme yöntemleri ile anahtar nokta tespiti için hazırlanan veri setinde YOLOv8 yardımıyla anahtar noktalar tespit edilmiştir. Anahtar noktaların tespiti ile veri seti içerisindeki tanımlanan hareketleri oluşturan görüntülerden ayrı ayrı balığın baş noktası 2 boyutlu düzleme aktarılmıştır. Düzleme aktarılan görüntüler tekrar görüntü işleme kullanılarak daha önceden imajenet veri seti ile eğitilmiş transfer öğrenme modelleri tarafından daha anlaşılabilir olması için istenilen boyuta indirilmiştir. Modellerde uydurma yaşanmaması için hiper-parametre

optimizasyonu, model eğitimini erken kesme algoritmaları ve modeli iki aşamalı eğitim yöntemleri kullanılmıştır. Çalışmanın sonucunda günümüzde yaygın olarak kullanılan ResNet50, ResNet101, DensNet169, Xception, Inception v3, EfficientNet-B4, NasNetLarge, VGG 16 modelleri ile eğitim yapılmış, eğitimlerin sonucunda EfficientNet-B4 modeli en yüksek başarıyı elde etmiştir.

Çalışmada kullanılan yöntem balıkların yönelim hareketlerini doğru tahmin edebilme yeteneğine sahiptir ancak dibe dalma ve yüze çıkma gibi üç boyutta yapılan hareketleri tanıma konusunda yüksek başarılar elde edememektedir. Ancak özel bir veri seti hazırlanması ya da ikinci bir kamera ya da bir sensör ile derinlik verisini elde edilmesi durumunda üçüncü bir boyutta da tahmin yapılabilmesi mümkün olabilir. Veri setinde gerekli veriler bulunmadığı için beslenme ve çiftleşme gibi karmaşık hareketleri tespit edebilme konusunda modeller test edilememiştir.

Çalışma literatür için yeni sayılabilecek yöntemde bir çalışma olduğu için gelecek çalışmalarda bu yöntem geliştirilerek daha yüksek performanslar alınabilir. Farklı bir yöntem ile duruş tahmini yapılarak düzleme aktarılan noktalar daha kesinleştirilebilir ve tahmin doğruluğu daha yüksek modeller elde edilebilir. Deney ortamı balık sayısı, kamera sayısı ve balık için daha doğal bir yaşam ortamı olacak şekilde geliştirilerek farklı tahmin sınıflarının da dahil olduğu kapsamlı bir veri seti elde edilerek modelin dezavantajlarından olan tahmin yeteneğinin zayıf olduğu sınıflar için daha güvenilir tahmin yapması sağlanabilir. Transfer öğrenme yöntemlerinin sadece özellik çıkarımı için kullanıldığı hibrit modeller oluşturularak daha performanslı modeller elde edilebilir. Çalışma bu yönüyle hem amacına ulaşmış literatürü destekleyici bir noktada bulunmaktadır hem de devam ettirilmesi mümkün yeni bir yol açmıştır.

KAYNAKLAR

- Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.
- Ajmal, H., Rehman, S., Farooq, U., Ain, Q. U., Riaz, F., ve Hassan, A. (2018). Convolutional neural network based image segmentation: a review. *Pattern Recognition and Tracking XXIX*, 10649, 191-203.
- Akgül, İ., Kaya, V., ve Zencir Tanır, Ö. (2023). A novel hybrid system for automatic detection of fish quality from eye and gill color characteristics using transfer learning technique. *Plos one*, 18(4), e0284804.
- Alaba, S. Y., Nabi, M., Shah, C., Prior, J., Campbell, M. D., Wallace, F., . . . Moorhead, R. (2022). Class-aware fish species recognition using deep learning for an imbalanced dataset. *Sensors*, 22(21), 8268.
- Archana, R., ve Jeevaraj, P. E. (2024). Deep learning models for digital image processing: a review. *Artificial Intelligence Review*, 57(1), 11.
- Atila, Ü., Uçar, M., Akyol, K., ve Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, 61, 101182.
- Benbrahim, H., ve Behloul, A. (2021). *Fine-tuned Xception for image classification on tiny Imagenet*. Paper presented at the 2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP).
- Bochkovskiy, A., Wang, C.-Y., ve Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Cao, Z., Simon, T., Wei, S.-E., ve Sheikh, Y. (2017). *Realtime multi-person 2d pose estimation using part affinity fields*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Dang, Q., Yin, J., Wang, B., ve Zheng, W. (2019). Deep learning based 2d human pose estimation: A survey. *Tsinghua Science and Technology*, 24(6), 663-676.
- Devlin, J., Chang, M.-W., Lee, K., ve Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dhillon, A., ve Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2), 85-112.

- Dong, C., ve Du, G. (2024). An enhanced real-time human pose estimation method based on modified YOLOv8 framework. *Scientific Reports*, 14(1), 8012.
- Dong, N., Zhao, L., Wu, C.-H., ve Chang, J.-F. (2020). Inception v3 based cervical cell classification combined with artificially extracted features. *Applied Soft Computing*, 93, 106311.
- Du, X., Zoph, B., Hung, W.-C., ve Lin, T.-Y. (2021). Simple training strategies and model scaling for object detection. *arXiv preprint arXiv:2107.00057*.
- Duda, A., Schwendner, J., Stahl, A., ve Rundtop, P. (2015). *Visual pose estimation for autonomous inspection of fish pens*. Paper presented at the OCEANS 2015-Genova.
- Er, M. B. (2021). Önceden Eğitilmiş Derin Ağlar İle Göğüs Röntgeni Görüntüleri Kullanarak Pnömoni Sınıflandırılması. *Konya Journal of Engineering Sciences*, 9(1), 193-204.
- Erdogan, M. (2023). Makine Öğrenmesi Sınıflandırma Modelleri: Accuracy, Precision, Recall, F1-score, Log Loss and Micro-Macro-Weighted Avg. Erişim adresi <https://medium.com/academy-team/makine-%C3%B6%C4%9Frenmesi-s%C4%B1n%C4%B1fland%C4%B1rma-modelleri-accuracy-precision-recall-f1-score-log-loss-and-e2c37fbcf285> (Son Erişim Tarihi: 10.07.2024)
- Eze, Ş. E., ve Berkaya, S. K. (2024). Görüntü ön işleme teknikleri ve derin öğrenme ile bitki zararlılarının sınıflandırılması. *Mühendislik Bilimleri ve Tasarım Dergisi*, 12(2), 455-465.
- Farahani, A., Pourshojae, B., Rasheed, K., ve Arabnia, H. R. (2020). *A concise review of transfer learning*. Paper presented at the 2020 international conference on computational science and computational intelligence (CSCI).
- Fisher, R. B., Chen-Burger, Y.-H., Giordano, D., Hardman, L., ve Lin, F.-P. (2016). *Fish4Knowledge: collecting and analyzing massive coral reef fish video data* (Vol. 104): Springer.
- Gómez-Flores, W., ve de Albuquerque Pereira, W. C. (2020). A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound. *Computers in Biology and Medicine*, 126, 104036.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., . . . Cai, J. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377.
- Gupta, N. (2017). A Literature Survey on Artificial Intelligence. *International Journal of Engineering Research & Technology (IJERT)*, 5(19), 1-5.
- Habe, H., Takeuchi, Y., Terayama, K., ve Sakagami, M.-a. (2021). Pose estimation of swimming fish using NACA airfoil model for collective behavior analysis. *Journal of Robotics and Mechatronics*, 33(3), 547-555. doi:<https://doi.org/10.20965/jrm.2021.p0547>

- He, K., Zhang, X., Ren, S., ve Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Hindarto, D. (2023). Battle Models: Inception ResNet vs. Extreme Inception for Marine Fish Object Detection. *Sinkron: jurnal dan penelitian teknik informatika*, 7(4), 2819-2826.
- Huang, G., Liu, Z., Van Der Maaten, L., ve Weinberger, K. Q. (2017). *Densely connected convolutional networks*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Huang, P. X., Boom, B. J., ve Fisher, R. B. (2013). *Underwater live fish recognition using a balance-guaranteed optimized tree*. Paper presented at the Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I 11.
- Iman, M., Arabnia, H. R., ve Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, 11(2), 40.
- Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., ve Schiele, B. (2016). *Deepercut: A deeper, stronger, and faster multi-person pose estimation model*. Paper presented at the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14.
- Isik, A. H., Özmen, Ö., Eskicioğlu, Ö. C., ve Isik, N. (2023). Classification and diagnosis of mammary tumors in dogs using deep learning techniques.
- Jäger, J., Rodner, E., Denzler, J., Wolff, V., ve Fricke-Neuderth, K. (2016). *SeaCLEF 2016: Object Proposal Classification for Fish Detection in Underwater Videos*. Paper presented at the CLEF (working notes).
- Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., ve Kaur, M. (2021). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*, 39(15), 5682-5689.
- Jia, R., Lv, B., Chen, J., Liu, H., Cao, L., ve Liu, M. (2023). Underwater Object Detection in Marine Ranching Based on Improved YOLOv8. *Journal of Marine Science and Engineering*, 12(1), 55.
- Jiang, T., Hu, X.-j., Yao, X.-h., Tu, L.-p., Huang, J.-b., Ma, X.-x., . . . Xu, J.-t. (2021). Tongue image quality assessment based on a deep convolutional neural network. *BMC Medical Informatics and Decision Making*, 21(1), 147.
- Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.-P., . . . Fisher, B. (2015). *LifeCLEF 2015: multimedia life species identification challenges*. Paper presented at the Experimental IR Meets Multilinguality, Multimodality, and Interaction: 6th International Conference of the CLEF Association, CLEF'15, Toulouse, France, September 8-11, 2015, Proceedings 6.

- Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., ve Whidden, C. (2022). Automated detection, classification and counting of fish in fish passages with deep learning. *Frontiers in Marine Science*, 8, 823173.
- Kebaili, A., Lapuyade-Lahorgue, J., ve Ruan, S. (2023). Deep learning approaches for data augmentation in medical imaging: a review. *Journal of Imaging*, 9(4), 81.
- Khalifa, N. E. M., Taha, M. H. N., ve Hassanien, A. E. (2018). *Aquarium family fish species identification system using deep neural networks*. Paper presented at the International Conference on Advanced Intelligent Systems and Informatics.
- Kılıçarslan, S., Çiçekliyurt, M. M. H., ve Kılıçarslan, S. (2024). Fish Freshness Detection Through Artificial Intelligence Approaches: A Comprehensive Study. *Turkish Journal of Agriculture-Food Science and Technology*, 12(2), 290-295.
- Konovalov, D. A., Saleh, A., Bradley, M., Sankupellay, M., Marini, S., ve Sheaves, M. (2019). *Underwater fish detection with weak multi-domain supervision*. Paper presented at the 2019 international joint conference on neural networks (ijcnn).
- Krizhevsky, A., Sutskever, I., ve Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Kundu, R. (2022). F1 Score in Machine Learning: Intro & Calculation. Erişim adresi <https://www.v7labs.com/blog/f1-score-guide> (Son Erişim Tarihi: 10.07.2024)
- LeCun, Y., Bengio, Y., ve Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., . . . Nie, W. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*.
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., ve Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185), 1-52.
- Li, Z., Liu, F., Yang, W., Peng, S., ve Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999-7019.
- Lin, B., Jiang, K., Xu, Z., Li, F., Li, J., Mou, C., . . . Duan, X. (2021). Feasibility research on fish pose estimation based on rotating box object detection. *Fishes*, 6(4), 65-85. doi:<https://doi.org/10.3390/fishes6040065>
- Lin, W., Sun, M.-T., Poovandran, R., ve Zhang, Z. (2008). *Human activity recognition for video surveillance*. Paper presented at the 2008 IEEE international symposium on circuits and systems (ISCAS).

- Liu, B., Yang, Y., Zhao, M., ve Hu, M. (2024). A Novel Lightweight Model for Underwater Image Enhancement. *Sensors*, 24(10), 3070.
- Liu, Q., Zhang, N., Yang, W., Wang, S., Cui, Z., Chen, X., ve Chen, L. (2017). *A review of image recognition with deep convolutional neural network*. Paper presented at the Intelligent Computing Theories and Application: 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part I 13.
- Loey, M., Manogaran, G., Taha, M. H. N., ve Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable cities and society*, 65, 102600.
- Lopez, M. M., ve Kalita, J. (2017). Deep Learning applied to NLP. *arXiv preprint arXiv:1703.03091*.
- Lu, M., Hu, Y., ve Lu, X. (2020). Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals. *Applied Intelligence*, 50(4), 1100-1111.
- Måløy, H., Aamodt, A., ve Misimi, E. (2019). A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture. *Computers and Electronics in Agriculture*, 167, 105087.
- Manjula, S., ve Lakshmi, K. (2022). *Human abnormal activity pattern analysis in diverse background surveillance videos using SVM and ResNet50 model*. Paper presented at the IoT and Analytics for Sensor Networks: Proceedings of ICWSNUCA 2021.
- Manna, A., Kundu, R., Kaplun, D., Sinitca, A., ve Sarkar, R. (2021). A fuzzy rank-based ensemble of CNN models for classification of cervical cytology. *Scientific Reports*, 11(1), 14538.
- McCarthy, J. (2007). What is artificial intelligence.
- Mukti, I. Z., ve Biswas, D. (2019). *Transfer learning based plant diseases detection using ResNet50*. Paper presented at the 2019 4th International conference on electrical information and communication technology (EICT).
- Naveed, H., Anwar, S., Hayat, M., Javed, K., ve Mian, A. (2024). Survey: Image mixing and deleting for data augmentation. *Engineering Applications of Artificial Intelligence*, 131, 107791.
- O'shea, K., ve Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Öğündür, G. (2024). Doğruluk (Accuracy) , Kesinlik(Precision) , Duyarlılık(Recall) ya da F1 Score ? Erişim adresi <https://medium.com/@gulcanogundur/dogruluk-accuracy-kesinlik-precision-duyarlilik-recall-ya-da-f1-score-300c925feb38> (Son Erişim Tarihi: 10.07.2024)

- Özden, S. (2024). Confusion Matrix (Karışıklık Matrisi). Erişim adresi <https://medium.com/@serapozden922/confusion-matrix-karışıklık-matrisi-62c43b8ad2b0> (Son Erişim Tarihi: 10.07.2024)
- Pagire, V., Phadke, A., ve Hemant, J. (2024). A deep learning approach for underwater fish detection. *Journal of Integrated Science and Technology*, 12(3), 765-765.
- Papandreou, G., Zhu, T., Kanazawa, N., Toshev, A., Tompson, J., Bregler, C., ve Murphy, K. (2017). *Towards accurate multi-person pose estimation in the wild*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Pareek, P., ve Thakkar, A. (2021). A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. *Artificial Intelligence Review*, 54(3), 2259-2322.
- Qassim, H., Verma, A., ve Feinzimer, D. (2018). *Compressed residual-VGG16 CNN model for big data places image recognition*. Paper presented at the 2018 IEEE 8th annual computing and communication workshop and conference (CCWC).
- Rashid, M., Gu, X., ve Jae Lee, Y. (2017). *Interspecies knowledge transfer for facial keypoint detection*. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Rawat, W., ve Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9), 2352-2449.
- Rum, S. N. M., ve Nawawi, F. A. Z. (2021). FishDeTec: a fish identification application using image recognition approach. *International Journal of Advanced Computer Science and Applications*, 12(3), 102-106.
- Salehi, H., ve Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering structures*, 171, 170-189.
- Schellewald, C., Stahl, A., ve Kelasidi, E. (2021). Vision-based pose estimation for autonomous operations in aquacultural fish farms. *IFAC-PapersOnLine*, 54(16), 438-443.
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., . . . Blake, A. (2011). *Real-time human pose recognition in parts from single depth images*. Paper presented at the CVPR 2011.
- Shrivastava, V. K., Pradhan, M. K., ve Thakur, M. P. (2021). *Application of pre-trained deep convolutional neural networks for rice plant disease classification*. Paper presented at the 2021 international conference on artificial intelligence and smart systems (ICAIS).
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., . . . Bolton, A. (2017). Mastering the game of go without human knowledge. *nature*, 550(7676), 354-359.

- Simonyan, K., ve Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sun, C., Zhan, W., She, J., ve Zhang, Y. (2020). Object detection from the video taken by drone via convolutional neural networks. *Mathematical Problems in Engineering*, 2020(1), 4013647.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., ve Wojna, Z. (2016). *Rethinking the inception architecture for computer vision*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Tan, M., ve Le, Q. (2019). *Efficientnet: Rethinking model scaling for convolutional neural networks*. Paper presented at the International conference on machine learning.
- Targ, S., Almeida, D., ve Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.
- Vianna, G. M., Zeller, D., ve Pauly, D. (2020). Fisheries and policy implications for human nutrition. *Current Environmental Health Reports*, 7, 161-169.
- Walther, D., Edgington, D. R., ve Koch, C. (2004). *Detection and tracking of objects in underwater video*. Paper presented at the Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.
- Wang, C.-Y., Bochkovskiy, A., ve Liao, H.-Y. M. (2023). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Wang, C., Chen, D., Hao, L., Liu, X., Zeng, Y., Chen, J., ve Zhang, G. (2019). Pulmonary image classification based on inception-v3 transfer learning model. *IEEE Access*, 7, 146533-146541.
- Wang, G., Muhammad, A., Liu, C., Du, L., ve Li, D. (2021). Automatic recognition of fish behavior with a fusion of RGB and optical flow data based on deep learning. *Animals*, 11(10), 2774.
- Wang, H., Jin, Y., Ke, H., ve Zhang, X. (2022). DDH-YOLOv5: improved YOLOv5 based on Double IoU-aware Decoupled Head for object detection. *Journal of Real-Time Image Processing*, 19(6), 1023-1033.
- Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., ve Shao, L. (2021). Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*, 210, 103225.

- Wei, S.-E., Ramakrishna, V., Kanade, T., ve Sheikh, Y. (2016). *Convolutional pose machines*. Paper presented at the Proceedings of the IEEE conference on Computer Vision and Pattern Recognition.
- Weiss, K., Khoshgoftaar, T. M., ve Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3, 1-40.
- Wu, X., Huang, J., ve Wang, L. (2024). Pose estimation-based visual perception system for analyzing fish swimming. *IEEE Sensors Journal*, 24(8), 13293 - 13303. doi:10.1109/JSEN.2024.3374236
- Wu, X., Liu, R., Yang, H., ve Chen, Z. (2020). *An xception based convolutional neural network for scene image classification with transfer learning*. Paper presented at the 2020 2nd international conference on information technology and computer application (ITCA).
- Xia, X., Xu, C., ve Nan, B. (2017). *Inception-v3 for flower classification*. Paper presented at the 2017 2nd international conference on image, vision and computing (ICIVC).
- Xu, C., Govindarajan, L. N., Zhang, Y., ve Cheng, L. (2017). Lie-x: Depth image based articulated object pose estimation, tracking, and action recognition on lie groups. *International Journal of Computer Vision*, 123, 454-478.
- Yang, G., He, Y., Yang, Y., ve Xu, B. (2020). Fine-grained image classification for crop disease based on attention mechanism. *Frontiers in Plant Science*, 11, 600854.
- Yang, L., Liu, Y., Yu, H., Fang, X., Song, L., Li, D., ve Chen, Y. (2021). Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: a review. *Archives of Computational Methods in Engineering*, 28, 2785-2816.
- Yasin, E. T., Ozkan, I. A., ve Koklu, M. (2023). Detection of fish freshness using artificial intelligence methods. *European Food Research and Technology*, 249(8), 1979-1990.
- Younis, A., Qiang, L., Nyatega, C. O., Adamu, M. J., ve Kawuwa, H. B. (2022). Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches. *Applied Sciences*, 12(14), 7282.
- Yu, T., ve Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *arXiv preprint arXiv:2003.05689*.
- Zeng, W. (2024). Image data augmentation techniques based on deep learning: A survey. *Mathematical Biosciences and Engineering*, 21(6), 6190-6224. doi:10.3934/mbe.2024272
- Zhang, D., Liu, Z., ve Shi, X. (2020). *Transfer learning on efficientnet for remote sensing image classification*. Paper presented at the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE).

- Zhang, P., Yang, L., ve Li, D. (2020). EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment. *Computers and Electronics in Agriculture*, 176, 105652.
- Zhang, Q. (2022). A novel ResNet101 model based on dense dilated convolution for image classification. *SN Applied Sciences*, 4, 1-13.
- Zhang, Q., Yang, Q., Zhang, X., Bao, Q., Su, J., ve Liu, X. (2021). Waste image classification based on transfer learning and convolutional neural network. *Waste Management*, 135, 150-157.
- Zhang, T., Xiao, J., Li, L., Wang, C., ve Xie, G. (2021). Toward Coordination Control of Multiple Fish-Like Robots: Real-Time Vision-Based Pose Estimation and Tracking via Deep Neural Networks. *IEEE CAA J. Autom. Sinica*, 8(12), 1964-1976. doi:10.1109/JAS.2021.1004228
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., ve Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57(4), 99.
- Zheng, Z., Guo, C., Zheng, X., Yu, Z., Wang, W., Zheng, H., . . . Zheng, B. (2018). *Fish recognition from a vessel camera using deep convolutional neural network and data augmentation*. Paper presented at the 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO).
- Zhou, Z., Yang, X., Ji, H., ve Zhu, Z. (2023). Improving the classification accuracy of fishes and invertebrates using residual convolutional neural networks. *ICES Journal of Marine Science*, 80(5), 1256-1266.
- Zhuang, P., Xing, L., Liu, Y., Guo, S., ve Qiao, Y. (2017). *Marine Animal Detection and Recognition with Advanced Deep Learning Models*. Paper presented at the CLEF (Working Notes).
- Zoph, B., Vasudevan, V., Shlens, J., ve Le, Q. V. (2018). *Learning transferable architectures for scalable image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Zou, Z., Chen, K., Shi, Z., Guo, Y., ve Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3), 257-276.