T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ENDOSKOPİK GÖRÜNTÜLERDEN POLİP SEGMENTASYONU VE GASTROİNTESTİNAL BULGULARIN SINIFLANDIRILMASI: TRANSFORMERS VE HİBRİT MODELLERİN ETKİNLİĞİ

Bengisu Ungan Eker

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Haziran-2025
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Bengisu UNGAN EKER tarafından hazırlanan "ENDOSKOPİK GÖRÜNTÜLERDEN POLİP SEGMENTASYONU VE GASTROİNTESTİNAL BULGULARIN SINIFLANDIRILMASI: TRANSFORMERS VE HİBRİT MODELLERİN ETKİNLİĞİ" adlı tez çalışması 17/06/2025 tarihinde aşağıdaki jüri tarafından oy birliği ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Doç. Dr. Murat KARAKOYUN (Necmettin Erbakan Üniversitesi)	……..
Danışman Dr. Öğr. Üyesi Fatma Zehra SOLAK (Konya Teknik Üniversitesi)	……..
Üye Dr. Öğr. Üyesi Burak YILMAZ (Konya Teknik Üniversitesi)	……..

Yukarıdaki sonucu onaylarım.

Prof. Dr. Mevlüt UYAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

Bengisu Ungan Eker
Tarih:

ÖZET

YÜKSEK LİSANS TEZİ

Endoskopik Görüntülerden Polip Segmentasyonu ve Gastrointestinal Bulguların Sınıflandırılması: Transformers ve Hibrit Modellerin Etkinliği

Bengisu Ungan Eker

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Fatma Zehra SOLAK

2025, 63 Sayfa

Jüri
Doç. Dr. Murat KARAKOYUN
Dr. Öğr. Üyesi Burak YILMAZ

Gastrointestinal (GI) sistem hastalıkları, dünya genelinde sıkça görülen ve hem ölüme hem de yaşam kalitesinin düşmesine yol açan önemli sağlık sorunlarıdır. Bu nedenle, erken ve doğru teşhis tedavi başarısı için hayati öneme sahiptir; özellikle polip, ülser ve tümör gibi bulguların zamanında saptanması hastalığın ilerlemesini önlemede kritik rol oynar. Endoskopik görüntüler bu açıdan değerli bir tanı aracı olmakla birlikte, manuel değerlendirmeleri hem zaman alıcı hem de hata riski taşımaktadır. Bu yüzden, yapay zekâ destekli otomatik analiz yöntemleri, klinik karar süreçlerini desteklemek için giderek daha fazla önem kazanmaktadır.

Bu tez çalışması, Kvasir v2 veri seti kullanılarak endoskopik görüntüler üzerinden hem polip segmentasyonu hem de gastrointestinal bulguların sınıflandırılması görevlerinde Transformer tabanlı ve hibrit derin öğrenme mimarilerinin etkinliğini incelemektedir. Sınıflandırma görevinde Vision Transformer (ViT), Swin Transformer, ConvNeXt ve ConViT; segmentasyon görevinde ise SegFormer ve Swin Transformer + UPerNet mimarileri kullanılmıştır. Geniş kapsamlı veri önişleme adımlarının ardından, eğitim sürecinde rastgele veri ayrıştırma ve 5 katlı çapraz doğrulama stratejileri uygulanmış; modeller doğruluk, F1-skoru, Dice ve IoU gibi çeşitli metriklerle değerlendirilmiştir.

Sınıflandırma görevinde ConvNeXt modeli, rastgele ayrıştırma stratejisinde %98,50 doğruluk ve %98,46 F1-skoru ile en yüksek başarıyı göstermiş; çapraz doğrulama stratejisinde ise %98,13 doğruluk ve %98,25 F1-skoruna ulaşmıştır. Segmentasyon görevinde ise Swin Transformer + UPerNet modeli %97,24 doğruluk, 0,9025 Dice ve 0,8303 IoU skorları ile öne çıkarken; SegFormer modeli %96,20 doğruluk ve 0,8731 Dice skoruna ulaşmıştır.

Transformer mimarileri, özellikle uzun menzilli bağlam bilgilerini modelleme becerisi sayesinde tıbbi görüntülerdeki karmaşık yapıları analiz etmede önemli

avantajlar sunmakta; hibrit modeller ise bu yapıları konvolüsyonel önyargılarla destekleyerek genelleme kabiliyetini artırmaktadır. Elde edilen bulgular, Transformer tabanlı yaklaşımların tıbbi görüntüleme alanındaki uygulanabilirliğini ortaya koyarken, klinik karar destek sistemlerinin geliştirilmesine de katkı sunmaktadır.

Anahtar Kelimeler: Endoskopik Görüntüler, Hibrit Model, Gastrointestinal Bulgular, Polip Segmentasyonu, Transformers

ABSTRACT

MS THESIS

Polyp Segmentation and Classification of Gastrointestinal Findings from Endoscopic Images: The Effectiveness of Transformers and Hybrid Models

Bengisu Ungan Eker

Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering

Advisor: Assist Prof Dr. Fatma Zehra SOLAK

Year, 63 Pages

Jury
Assoc. Prof. Dr. Murat KARAKOYUN
Assist Prof Dr. Burak YILMAZ

Gastrointestinal (GI) system diseases are common worldwide and cause both death and a decline in quality of life. Therefore, early and accurate diagnosis is vital for successful treatment; timely detection of findings such as polyps, ulcers, and tumors plays a critical role in preventing disease progression. While endoscopic images serve as valuable diagnostic tools, their manual evaluation is time-consuming and prone to errors. For this reason, AI-supported automatic analysis methods are becoming increasingly important to assist clinical decision-making processes.

This thesis investigates the effectiveness of Transformer-based and hybrid deep learning architectures in polyp segmentation and GI finding classification using the Kvasir v2 dataset. Vision Transformer (ViT), Swin Transformer, ConvNeXt, and ConViT were used for classification tasks, while SegFormer and Swin Transformer + UPerNet were employed for segmentation. After comprehensive preprocessing, training was conducted using both random data splitting and 5-fold cross-validation strategies, and the models were evaluated using multiple metrics, including accuracy, F1-score, Dice, and IoU.

In classification, the ConvNeXt model achieved the highest performance with 98.50% accuracy and 98.46% F1-score under random splitting, and 98.13% accuracy and 98.25% F1-score with cross-validation. For segmentation, the Swin Transformer + UPerNet model stood out with 97.24% accuracy, a Dice score of 0.9025, and an IoU of 0.8303, while SegFormer reached 96.20% accuracy and a Dice score of 0.8731.

Transformer architectures provide notable advantages in analyzing complex structures in medical images due to their ability to model long-range contextual relationships. Hybrid models further enhance generalization by incorporating convolutional inductive biases. The findings demonstrate the applicability of

Transformer-based approaches in medical image analysis and contribute to the development of clinical decision support systems.

Keywords: Endoscopic Images, Hybrid Models, Gastrointestinal Findings, Polyp Segmentation, Transformer

ÖNSÖZ

Bu tez çalışması, gastrointestinal sistem hastalıklarının erken teşhisinde kritik rol oynayan endoskopik görüntü analizinin geliştirilmesine yönelik bir adım olarak ortaya çıkmıştır. Transformer mimarilerinin ve hibrit yaklaşımların tıbbi görüntü analizindeki potansiyelini keşfetmeyi amaçlayan bu araştırma, hem akademik hem de klinik uygulamalar açısından değerli katkılar sunmayı hedeflemektedir.

Lisansüstü eğitimim boyunca beni yönlendiren, destekleyen ve akademik gelişimime katkıda bulunan danışmanım Sayın Dr. Fatma Zehra SOLAK'a en içten teşekkürlerimi sunarım.

Bengisu Ungan Eker
KONYA-2025

İÇİNDEKİLER

SİMGELER VE KISALTMALAR

Simgeler

DSC: Dice Similarity Coefficient – Dice Benzerlik Katsayısı (Piksel bazlı uyum ölçütü)

ACC: Accuracy – Doğruluk (Genel sınıflandırma başarımı)

F1: F1-Skoru – Harmonik Ortalama (Kesinlik ve duyarlılığın dengesi)

TP: True Positive – Doğru Pozitif

TN: True Negative – Doğru Negatif

FP: False Positive – Yanlış Pozitif

FN: False Negative – Yanlış Negatif

IoU: Intersection over Union – Kesişim / Birleşim Oranı

Kısaltmalar

AI: Artificial Intelligence – Yapay Zekâ

CNN: Convolutional Neural Network – Evrimsel Sinir Ağı

ViT: Vision Transformer

U-Net: U-Net Segmentasyon Ağı

UPerNet: Unified Perceptual Parsing Network – Birleşik Algısal Ayrıştırma Ağı

FPN: Feature Pyramid Network – Özellik Piramit Ağı

PPM: Pyramid Pooling Module – Piramit Havuzlama Modülü

C.L.A.H.E.: Contrast Limited Adaptive Histogram Equalization – Kontrast Sınırlamalı Uyarlamalı Histogram Eşitleme

WCE: Wireless Capsule Endoscopy – Kablosuz Kapsül Endoskopi

GIS: Gastrointestinal Sistem

BDA: Binary Dragonfly Algorithm – İkili Yusufçuk Algoritması

ROC-AUC: Receiver Operating Characteristic – Area Under Curve

Grad-CAM: Gradient-weighted Class Activation Mapping – Sınıf Aktivasyon Görselleştirme Tekniği

MLP: Multi-Layer Perceptron – Çok Katmanlı Algılayıcı

1. GİRİŞ

Gastrointestinal sistem (GIS) hastalıkları, dünya genelinde yaygın olarak görülen ve bireylerin yaşam kalitesini ciddi biçimde etkileyen önemli sağlık sorunları arasında yer almaktadır. Özellikle kolorektal kanser gibi ölümcül sonuçlara yol açabilen patolojilerle ilişkili polip oluşumlarının erken evrede saptanması, tedavi başarısını artırmakta ve yaşam süresini uzatmaktadır. Bu kapsamda, endoskopik görüntüleme yöntemleri; yüksek çözünürlüklü görseller sağlayarak gastrointestinal anomalilerin tanı ve takibinde kritik bir rol oynamaktadır (Gunasekaran ve diğ., 2023). Öte yandan, bu görüntülerin manuel olarak değerlendirilmesi zaman alıcı, yorucu ve yorumlayıcının kişisel algı ve yorumlarına bağlı farklılıklar nedeniyle hata yapmaya açık bir süreçtir (Keleş, 2022).

Derin öğrenme temelli yapay zekâ yaklaşımları, manuel değerlendirmeye kıyasla daha hızlı, tutarlı ve genellenebilir çözümler sunarak tıbbi görüntü analizinde hızla yaygınlaşmıştır. Derin öğrenme modellerinin, ham tıbbi görüntülerden manuel özellik mühendisliğine gerek kalmadan anlamlı temsiller çıkarabilme yeteneği, sınıflandırma ve segmentasyon gibi tanısal görevlerde önemli başarılar sağlamıştır (Thakur ve diğ., 2024). Endoskopik görüntüler üzerinde yapılan çalışmalar da bu çerçevede hız kazanmış; polipler, ülserler ve kötü huylu (malign) yapılar gibi anomalilerin otomatik tespiti ve sınıflandırılması amacıyla derin öğrenme modellerinden yoğun şekilde faydalanılmıştır (Gunasekaran ve diğ., 2023).

Gerçekleştirilen tez çalışmasında, endoskopik görüntüler üzerinden iki temel görev ele alınmıştır: gastrointestinal bulguların sınıflandırılması ve poliplerin segmentasyonu. Bu görevler, farklı tanısal hedeflere odaklandıkları, farklı modelleme yaklaşımları ve çıktı türleri gerektirdikleri için teknik açıdan tamamen bağımsız biçimde yürütülmüştür. Her bir görevde, özgün modelleme süreçleri izlenmiş ve birinin çıktısı diğerinin girdisi olarak kullanılmamıştır. Sınıflandırma sürecinde, görüntüler sekiz farklı bulgu sınıfına ayrılmıştır: boyanmış kaldırılmış polipler (dyed-lifted-polyps), boyanmış rezeksiyon kenarları (dyed-resection-margins), özofajit (esophagitis), normal çekum (normal-cecum), normal pilor (normal-pylorus), polipler (polyps), ülseratif kolit (ulcerative-colitis) ve z çizgisi (z-line). Bu görevde modelin amacı, verilen bir görüntünün bu sınıflardan hangisine ait olduğunu doğru biçimde belirlemektir. Segmentasyon görevi ise yalnızca “polip” sınıfına ait görüntüler üzerinde

gerçekleştirilmiş; bu yapıların konumlarının ve sınırlarının piksel düzeyinde hassas biçimde çıkarılması hedeflenmiştir. Her iki görev, teknik olarak birbirinden tamamen bağımsız biçimde gerçekleştirilmiş olsa da, klinik açıdan birbirini tamamlayıcı niteliktedir: sınıflandırma genel tanısal kategorilerin belirlenmesine katkı sağlarken, segmentasyon belirli yapıların morfolojik detaylarının ortaya konmasına olanak tanımaktadır. Bu kapsamda elde edilen sonuçlar birlikte değerlendirilmiş ve geliştirilen modellerin tıbbi karar destek sistemlerine katkısı bütüncül bir yaklaşımla analiz edilmiştir.

Literatürde, derin öğrenme tabanlı yöntemlerin sınıflandırma ve segmentasyon görevlerinde yüksek doğrulukla çalıştığı birçok çalışma ile gösterilmiştir. Bu başarıda kullanılan mimarilerin rolü belirleyicidir. Uzun süredir tıbbi görüntüleme alanında yaygın biçimde kullanılan Konvolüsyonel Sinir Ağları (Convolutional Neural Network - CNN), yerel uzamsal özellikleri etkili biçimde yakalayabilmektedir (Ali ve diğ., 2021). Ancak bu yapılar, uzun menzilli bağlamsal ilişkileri modelleme konusunda sınırlılıklar taşımakta ve özellikle endoskopik görüntülerdeki dağınık patolojik yapılar karşısında yetersiz kalmaktadır (Li ve diğ., 2023). Ayrıca yüksek performanslı CNN modellerinin geniş, dengeli ve etiketli veri kümelerine olan gereksinimi, klinik uygulamalar açısından önemli bir engel teşkil etmektedir.

Bu kısıtlamaları aşmak adına, başlangıçta doğal dil işleme alanında geliştirilen Transformer mimarileri, son yıllarda görüntü işleme problemlerine başarıyla uyarlanmıştır. Öz-dikkat (self-attention) mekanizmaları sayesinde küresel bağlamı etkili biçimde öğrenebilen bu yapılar, görsel verilerdeki uzun mesafeli ilişkileri temsil etme konusunda önemli avantajlar sunmaktadır. Vision Transformer (ViT), Swin Transformer ve SegFormer gibi modeller, tıbbi görüntüleme alanında umut vadeden sonuçlara ulaşmıştır (Azat ve diğ., 2024; Pereira ve diğ., 2024; He ve diğ., 2023). Bununla birlikte, bu mimarilerin yüksek hesaplama maliyeti ve büyük miktarda etiketli veriye olan ihtiyacı, geniş ölçekli klinik uygulamaların önünde hâlâ bir engel oluşturmaktadır.

Bu nedenlerle, son yıllarda konvolüsyonel yapılar ile Transformer mimarilerinin güçlü yönlerini birleştiren hibrit modeller geliştirilmiştir (Khan ve diğ., 2022; d'Ascoli ve diğ., 2021). Sınıflandırma görevinde; ViT (Dosovitskiy ve diğ., 2020) ve Swin Transformer (Liu ve diğ., 2021) gibi saf Transformer mimarilerine ek olarak, konvolüsyonel önyargılarla zenginleştirilmiş ConViT modeli (d'Ascoli ve diğ., 2021) öne çıkmaktadır. Ayrıca, CNN tabanlı olup Transformer ilkelerinden esinlenen

ConvNeXt modeli (Liu ve diğ., 2022) de güncel ve etkili bir alternatif olarak değerlendirilmektedir.

Segmentasyon özelinde ise, Transformer mimarisini temel alan SegFormer (Xie ve diğ., 2021) başarılı sonuçlar üretirken, Swin Transformer ile UPerNet mimarisinin birleşiminden oluşan hibrit yapılar da yüksek doğruluk ve detay başarımı sunarak ön plana çıkmaktadır (Xiao ve diğ., 2018; Liu ve diğ., 2021).

Kvasir v2 veri seti kullanılarak gerçekleştirilen bu tez çalışmasında, yukarıda tanımlanan yöntemler doğrultusunda sınıflandırma ve segmentasyon görevlerinde Transformer tabanlı ve hibrit mimarilerin performansı ayrıntılı biçimde karşılaştırılmıştır. Değerlendirmelerde, sınıflandırma görevine yönelik modeller için doğruluk, kesinlik, duyarlılık, özgüllük ve F1 skoru; segmentasyon görevine yönelik modeller için ise bu metriklere ek olarak Dice katsayısı ve Kesişim/Birleşim Oranı (IoU) esas alınmıştır. Elde edilen bulgular, Transformer mimarilerinin ve hibrit yapıların endoskopik görüntü analizinde yüksek doğruluk, detay duyarlılığı ve genellenebilirlik sunduğunu göstermektedir.

Bu tez beş ana bölümden oluşmaktadır. İkinci bölümde, endoskopik görüntüler üzerinde yapılan çalışmalar ele alınmış; literatürde gastrointestinal bulguların sınıflandırılması ve polip segmentasyonu alanlarında kullanılan yöntemler iki ayrı alt başlık altında detaylı biçimde incelenmiştir. Üçüncü bölümde, çalışmada kullanılan veri seti, uygulanan önişleme adımları, sınıflandırma ve segmentasyon için tercih edilen derin öğrenme mimarileri, model eğitimi süreci ve performans değerlendirme kriterleri kapsamlı şekilde açıklanmıştır. Dördüncü bölümde, sınıflandırma ve segmentasyon görevlerinde elde edilen deneysel bulgulara yer verilmiş; her bir modelin sonuçları ayrı ayrı sunulmuş ve ardından kapsamlı karşılaştırmalar gerçekleştirilmiştir. Beşinci ve son bölümde ise, tez kapsamında ulaşılan genel sonuçlar özetlenmiş ve gelecekteki çalışmalara yönelik öneriler sunulmuştur.

2. KAYNAK ARAŞTIRMASI

Tıpta yapay zekâ ve derin öğrenme temelli yaklaşımların artan kullanımı, özellikle görüntü tabanlı tanı süreçlerinde önemli ilerlemelere yol açmıştır. Bu bağlamda, endoskopik görüntülerden elde edilen bilgilerle gastrointestinal sistem hastalıklarının otomatik olarak analiz edilmesine yönelik çok sayıda çalışma gerçekleştirilmiştir. Bu bölümde, çalışmanın odaklandığı iki temel problem olan gastrointestinal bulguların sınıflandırılması ve polip segmentasyonu üzerine yapılan güncel araştırmalar incelenmekte, kullanılan yöntemler, karşılaşılan zorluklar ve mevcut literatürdeki boşluklar değerlendirilerek alandaki gelişim yönleri ortaya konulmaktadır.

2.1. Endoskopik Görüntülerde Gastrointestinal Bulguların Sınıflandırılması

Noha ve diğ. (2019), DenseNet tabanlı Faster R-CNN modelini ve Gabor filtre özelliklerini kullanarak özofagus anormalliklerini tespit eden bir sistem geliştirmiştir. Çalışmada, Kvasir veri seti üzerinde %90,2 duyarlılık, %92,1 kesinlik ve %75,9 mAP değerleri elde edilmiş olup, Gabor filtrelerinin CNN modelleri ile birlikte kullanımının sınıflandırma doğruluğunu artırdığı belirtilmiştir.

Wang ve diğ. (2021), CapsNet (Capsule Networks) kullanarak Kvasir veri setindeki gastrointestinal bulguların sınıflandırılmasını gerçekleştirmiş ve kapsül ağlarının, geleneksel CNN modellerine kıyasla daha hassas öznitelik öğrenme yeteneğine sahip olduğunu ortaya koymuştur. Yapılan deneylerde Kvasir v2 veri setinde %94,83 doğruluk oranı elde edilmiştir.

Khan ve diğ. (2022), GastroNet adlı model ile gastrointestinal hastalıkların tespitinde saliency tahmini ve derin öğrenme özelliklerini birleştirerek önemli sonuçlar elde etmiştir. Hibrit ardışık füzyon yaklaşımı ile enfekte bölgelerin kontrastı artırılarak poliplerin daha belirgin hale getirilmesi sağlanmış ve Kvasir v1 veri setinde %98,25, Kvasir v2 veri setinde %98,02 doğruluk oranına ulaşılmıştır.

Regmi ve diğ. (2023), gastrointestinal görüntü sınıflandırması için Vision Transformer (ViT) mimarisini kullanarak Kvasir ve Kvasir-Capsule veri setleri üzerinde deneyler gerçekleştirmiştir. ViT mimarisi, özellikle büyük veri setleriyle iyi performans gösteren saf Transformer temelli bir görsel modeldir. Bu çalışmada model, 16x16 piksellik patch'lere bölünmüş giriş görüntülerini lineer bir şekilde embed ederek ve ardından çok katmanlı öz-dikkat mekanizmalarıyla işleyerek nihai sınıflandırmayı

gerçekleştirmiştir. Kvasir veri seti üzerinde yapılan testlerde, model %94,36 F1 skoru ve 0,97 ROC-AUC skoru elde etmiş ve klasik CNN tabanlı yaklaşımların üzerine çıkmayı başarmıştır.

Solak (2023), Kvasir v2 veri seti kullanarak endoskopik görüntülerden Özofajit, Polip ve Ülseratif Kolit sınıflarının otomatik olarak sınıflandırılmasını ele almıştır. Çalışmada yalnızca bu üç patolojik bulguya odaklanılmış ve önceden eğitilmiş ResNet-50 modelinin, özel olarak tasarlanmış bir CNN mimarisine kıyasla %99 doğruluğa ulaşarak üstün performans gösterdiği raporlanmıştır.

Yoshioka ve diğ. (2023), Kvasir veri seti kullanarak özofajit tespiti ve Z-line lokalizasyonuna yönelik CNN tabanlı mimarileri karşılaştırmalı olarak incelemiştir. Elde edilen sonuçlara göre, GoogLeNet modeli en yüksek F1 skorunu elde ederken, MobileNetV3 modeli ortalama doğru pozitif oranı açısından daha güvenilir sonuçlar vermiştir.

Auzine ve diğ. (2024), InceptionV3, Inception-ResNetV2 ve VGG16 mimarilerini birleştirerek gastrointestinal kanser sınıflandırması için bir topluluk modeli önermiştir. Kvasir-V2 veri seti üzerinde %93,17 doğruluk ve %97 F1 skoru elde edilmiştir. Ayrıca SHAP yöntemi ile model kararlarının açıklanabilirliği sağlanmıştır.

Cambay ve diğ. (2024), Kvasir, Kvasir-v2 ve WCE veri setleri kullanılarak ResNet50* temelli bir sınıflandırma modeli geliştirmiştir. Geliştirilen model, özellikle WCE veri seti üzerinde %99,13 doğruluk elde ederek güçlü genelleme kabiliyeti sergilemiştir.

Eker ve Solak (2024), ön işlenmiş Kvasir v2 veri seti üzerinde altı gastrointestinal sınıfın sınıflandırılmasını gerçekleştirmiştir. Sekiz farklı CNN mimarisi (Net-1–Net-8) ile yapılan çalışmada, Net-5 modeli %87,33 doğrulukla en iyi sonucu elde etmiş; F1 skoru, hassasiyet ve özgüllük değerleri de benzer düzeyde bulunmuştur. Ağ derinliğinin ve filtre sayısının arttırılmasının performansı artırdığı buna karşın belirli bir noktadan sonra bu etkinin sınırlı kaldığı gözlemlenmiştir.

Khan ve diğ. (2024), Darknet-53 ve Xception modellerinden elde edilen derin özellikleri ikili dragonfly algoritması (BDA) ile optimize ederek, bunları ensemble subspace k-en yakın komşu (ESKNN) sınıflandırıcısıyla birleştirmiştir. Kvasir-V2 ve CUI Wah veri setleri üzerinde yapılan deneylerde sırasıyla %98,25 ve %99,90 doğruluk oranlarına ulaşılmıştır.

Kulkarni ve diğ. (2024), DenseNet201, InceptionV3 ve ResNet50 mimarilerini birleştirerek GIT-NET adlı ağırlıklı topluluk modelini geliştirmiştir. Model, Kvasir-V2

veri seti üzerinde %95 doğrulukla en yüksek başarıyı sağlamış ve bireysel modellerin üzerinde performans sergilemiştir.

Oğuz ve Alkan (2024), Xception, ResNet-101 ve NASNet-Large mimarilerini birleştirerek ağırlıklı topluluk modeli oluşturmuş ve Kvasir v2 veri setinde %94,125 doğruluk sağlamıştır. Modelin karar mekanizmaları Grad-CAM ile açıklanmış, sınıflandırma performansına ek olarak yorumlanabilirlik de artırılmıştır.

Patil ve Giripunje (2024), EfficientNetB1 ve EfficientNetB0 mimarileri kullanarak gastrointestinal hastalık sınıflandırması gerçekleştirmiştir. Kvasir v2 veri seti üzerinde yapılan deneylerde sırasıyla %94,75 ve %94,31 doğruluk değerlerine ulaşılmış, sınıflar arası benzerliklerin model başarımını zorlaştırdığı vurgulanmıştır.

Pokuaa ve diğ. (2024), küçük tıbbi veri setleriyle çalışmaya uygun bir kapsül ağ (Capsule Network) modeli önermiştir. Yama temelli ve özellik güçlendirme mekanizmalarıyla geliştirilen model, Kvasir-V2 veri setinde %93,40 doğruluk sağlamış ve birçok SOTA modeli geride bırakmıştır. Modelin iç işleyişi detaylı görselleştirmelerle açıklanmıştır.

Sameh Abd El-Ghany ve diğ. (2024), model eğitimi süresince öğrenme oranını adaptif biçimde ayarlayan Akıllı Öğrenme Oranı Denetleyicisi (ILRC) adlı bir mekanizma önermiştir. EfficientNet-B0, ResNet101v2, InceptionV3 ve InceptionResNetV2 modelleri ile birlikte Kvasir-Capsule ve Kvasir-v2 veri setleri üzerinde yapılan deneylerde sırasıyla %99,906 ve %98,062 doğruluk değerlerine ulaşılmıştır.

Siddiqui ve diğ. (2024), CG-Net adını verdikleri CNN tabanlı çerçevede, özellik seçimi ile desteklenen bir sınıflandırma yöntemi sunmuşlardır. Kvasir v1 üzerinde %95 doğruluk edilerek birçok güncel modeli geride bırakmıştır. Ayrıca modelin çalışma süresi ve verimliliği de olumlu değerlendirilmiştir.

Tsai ve Lee (2024), Kvasir v2 veri seti üzerinde 11 farklı önceden eğitilmiş CNN modelini incelemiş ve en başarılı olanları ensemble yöntemle birleştirmiştir. DenseNet201, VGG19 ve InceptionV3 modellerini içeren topluluk, %91 doğrulukla en iyi sonucu elde etmiş ve Grad-CAM ile karar görselleştirmesi sağlanmıştır.

Varam ve diğ. (2024), düşük hesaplama gücüne sahip cihazlarda (edge devices) medikal görüntü sınıflandırması yapılabilirliğini değerlendirmek amacıyla Vision Transformer modellerini optimize etmiştir. Kvasir-Capsule veri seti kullanılarak gerçekleştirilen çalışmada, Transformer yapısı kompaktlaştırılarak hem enerji verimliliği sağlanmış hem de yüksek doğruluk korunmuştur. Geliştirilen model, %95,3

F1 skoru ile ViT mimarisinin edge uygulamalar için uyarlanabilir olduğunu göstermiştir.

Zhou ve diğ. (2024), ConvMixer mimarisine mekansal dikkat (spatial attention) katmanı entegre ederek, CNN'lerin yerel özellik çıkarım yeteneği ile dikkat mekanizmalarının küresel bağlam farkındalığını birleştiren hibrit bir yaklaşım önermiştir. Kvasir veri seti üzerinde yapılan deneyler sonucunda model, %93,37 doğruluk elde etmiştir. Modelin öne çıkan özelliği, hastalıklı bölgelerin konumunu daha doğru tespit etmesini sağlayan dikkat tabanlı filtreleme mekanizmasıdır.

Muhammed ve diğ. (2025), geleneksel CNN mimarisine Transformer katmanları entegre ederek yeni bir hibrit model önermiştir. Kvasir-V2 veri seti kullanılarak gerçekleştirilen çalışmada model, %95,3 doğruluk oranı ile literatürdeki pek çok mevcut yöntemi geride bırakmıştır. Derin özellik haritalarının görselleştirilmesiyle modelin öğrenme süreci detaylandırılmış ve klinik karar destek sistemleri için açıklanabilir yapısı vurgulanmıştır.

Tsai ve Lee (2025), aynı veri seti üzerinde bu kez dinamik ağırlıklarla optimize edilen (%0,4, %0,4, %0,2) topluluk modeli geliştirmiştir. Ensemble yapı, özellikle zorlayıcı sınıflar olan özofajit ve normal z-line gibi durumlarda bireysel modellerin ötesinde başarı sağlamış ve %91 doğruluk elde etmiştir.

Waheed ve diğ. (2025), EfficientNetB0, EfficientNetB2 ve ResNet101 modellerini Beta Normalizasyon Toplama yöntemiyle birleştirerek transfer öğrenme temelli bir topluluk modeli oluşturmuştur. İki farklı gastrointestinal veri setinde sırasıyla %97,88 ve %97,47 doğruluk elde edilmiştir. Modelin açıklanabilirliği Grad-CAM ile görsel olarak sunulmuştur.

2.2. Endoskopik Görüntülerde Polip Segmentasyonu

Fan ve diğ. (2020) tarafından geliştirilen PraNet modeli, paralel kısmi kod çözücüler ve ters dikkat mekanizmasını kullanarak polip sınırlarını belirginleştirmiş ve %89,8 Dice skoru ile %85,5 IoU oranına ulaşmıştır.

Srivastava ve diğ. (2021) tarafından sunulan DDANet modeli, yoğun bağlantılar ve çift dikkat mekanizması kullanarak hem yerel hem de global özellikleri öğrenmiş ve %90,2 Dice skoru ile %82,3 IoU değerine ulaşmıştır.

Srivastava ve diğ. (2022) tarafından önerilen MSRF-Net modeli, çoklu ölçekli özellik çıkarımı ve artık bağlantılar kullanarak farklı boyutlardaki poliplerin daha başarılı bir şekilde segmentasyonunu gerçekleştirmiş ve %92,7 Dice skoru elde etmiştir.

Chen ve diğ. (2024), CNN ve Transformer yapısını tek bir çatı altında birleştiren MixFormer modelini önermiştir. Model, Kvasir-SEG veri seti üzerinde test edilmiş ve %91,8 Dice skoru ile %87,9 IoU değeri elde etmiştir. Bu yapı, global dikkat yapısı ile ayrıntılı kenar tespiti sağlayan CNN katmanlarını birleştirerek daha bütüncül bir segmentasyon sağlamıştır.

Tang ve diğ. (2024), polip segmentasyonu için hibrit bir mimari olan HTC-Net'i geliştirmiştir. Model, yerel bilgi akışını sağlayan CNN yapıları ile küresel dikkat alanı sunan Transformer bloklarını birleştirmiştir. Kvasir-SEG veri seti kullanılarak %92,1 Dice skoru ve %86,6 IoU değeri elde edilmiştir. Modelin performansı, özellikle sınır belirsizliği olan bölgelerde yüksek doğrulukla segmentasyon yapabilmesi sayesinde öne çıkmıştır.

Wang ve diğ. (2025), etiketli verilerin az olduğu senaryolara uygun olarak yarı denetimli bir segmentasyon modeli geliştirmiştir. Kvasir-SEG veri seti üzerinde yapılan çalışmalarda, bilgi madenciliği (knowledge mining) ve pseudo-labeling tekniklerini kullanarak %89,5 Dice skoru elde edilmiştir. Özellikle düşük etiketli veri koşullarında modelin stabil ve güçlü sonuçlar verdiği vurgulanmıştır.

Yapılan kaynak incelemesi, gastrointestinal sistem hastalıklarının sınıflandırılması ve polip segmentasyonu gibi görevlerde derin öğrenme tabanlı yöntemlerin yaygın olarak kullanıldığını, özellikle CNN tabanlı mimarilerin yüksek başarı sağladığını ortaya koymuştur. Ancak, uzun menzilli bağıntıları modellemede yetersiz kalan bu yapılar, görsel bağlamı tam anlamıyla değerlendirememektedir. Son dönemde geliştirilen Transformer temelli yaklaşımlar, öz-dikkat mekanizmaları

sayesinde bu eksikliği gidermeye yönelik önemli bir potansiyel sunmakta; bazı çalışmalarda umut verici sonuçlar elde edilmesine rağmen, bu modellerin endoskopik görüntü analizinde kullanımı halen sınırlı kalmaktadır. Ayrıca, Transformer mimarileriyle hibrit şekilde tasarlanan yapılar üzerine gerçekleştirilen karşılaştırmalı çalışmaların sayısı da oldukça azdır. Bu bağlamda, söz konusu tez çalışması, Transformer ve hibrit yaklaşımların hem sınıflandırma hem de segmentasyon görevlerinde sistematik olarak değerlendirilmesini sağlayarak literatürdeki bu önemli boşluğu doldurmayı amaçlamaktadır.

3. MATERYAL VE YÖNTEM

3.1. Veri Seti

Gerçekleştirilen tez çalışmasında, açık erişimli Kvasir Versiyon 2 (Kvasir v2) kullanılarak endoskopik görüntülerden polip segmentasyonu ve gastrointestinal bulguların sınıflandırılması görevleri gerçekleştirilmiştir.

Kvasir v2 veri seti (Pogorelov ve diğ., 2017), gastrointestinal sistemdeki normal anatomik yapılar ve patolojik bulguları içeren toplamda sekiz sınıftan oluşmaktadır. Her sınıf, farklı tipte bulguları temsil ederek araştırmacılara ve klinik süreçlerde teşhis destek sistemlerine kapsamlı bir veri kaynağı sunmaktadır. Veri seti, toplamda 8000 görüntü içermektedir ve bu görüntülerin çözünürlükleri 720x576 ile 1920x1072 piksel arasında değişmektedir. Veri setindeki sınıflar aşağıda detaylı bir şekilde açıklanmıştır:

i. **Z-Çizgisi (Z Line)**: Z çizgisi, yemek borusu ile mide arasındaki geçiş bölgesinde bulunan bir yapıdır. Bu yapı, sağlıklı bireylerde gastroözofageal reflü veya diğer anormalliklerin olmadığını gösterir. Veri setinde bu sınıfa ait 1000 görüntü bulunmaktadır ve bu görüntüler, normal Z çizgisinin detaylı görsellerini içermektedir.

ii. **Normal Çekum (Normal Cecum)**: Çekum, kalın bağırsağın başlangıç kısmıdır ve kolonoskopi sürecinde gözlemlenmesi gereken önemli bir bölgedir. Sağlıklı çekumun doğru gözlemlenmesi, tam bir kolonoskopi gerçekleştirilmiş olduğunu göstermektedir. Veri setinde, sağlıklı çekum yapılarını gösteren 1000 görüntü bulunmaktadır.

iii. **Normal Pilor (Normal Pylorus)**: Pilor, mide ile ince bağırsak arasındaki geçiş bölgesinde bulunur ve bu yapı mide içeriğinin ince bağırsağa kontrollü geçişini sağlar. Sağlıklı pilor yapısının endoskopik görüntüleri, mide ve ince bağırsak arasındaki bu bölgenin normal olduğunu göstermektedir. Veri setinde bu sınıfa ait 1000 görüntü yer almaktadır.

iv. **Özofajit (Esophagitis)**: Özofajit, yemek borusunun iltihaplanması durumudur ve genellikle asit reflü, enfeksiyon veya alerjik reaksiyonlar gibi nedenlerle ortaya çıkabilir. Özofajitin farklı evrelerini temsil eden bu sınıfta 1000 görüntü yer almaktadır. Özofajit erken teşhis edilmediğinde yemek borusunun ciddi şekilde hasar görmesine neden olabilir, bu nedenle erken tanı büyük önem taşımaktadır.

v. **Polipler (Polyps)**: Polipler, bağırsak veya mide gibi iç organların yüzeyinde gelişen çıkıntılardır ve erken teşhis edilmediği takdirde kansere dönüşme riski taşımaktadır. Veri setinde poliplerin farklı boyut ve şekillerdeki örneklerini içeren 1000 görüntü bulunmaktadır. Poliplerin doğru ve erken teşhisi, özellikle kolon kanseri riskini azaltmada büyük önem taşımaktadır.

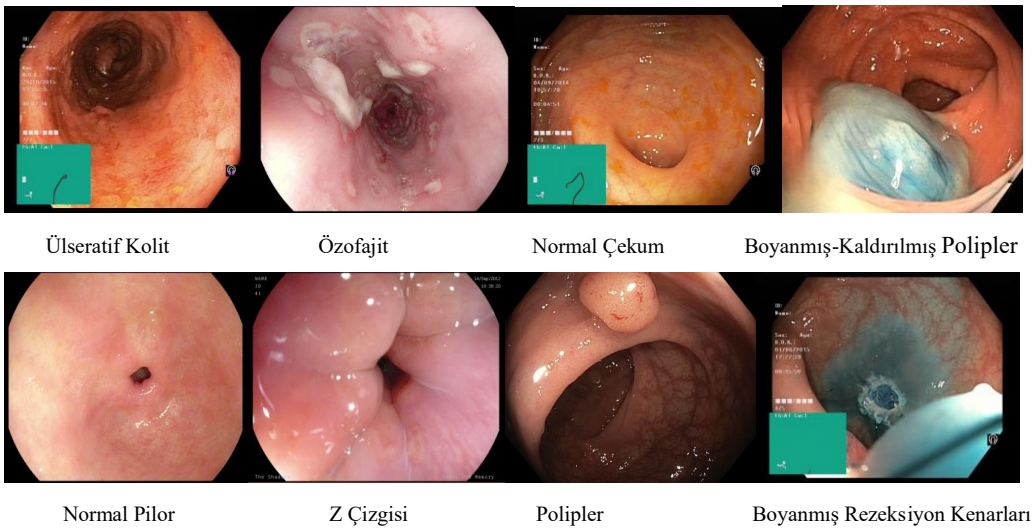
vi. **Ülseratif Kolit (Ulcerative Colitis)**: Ülseratif kolit, kalın bağırsağın kronik iltihaplanması ile karakterize bir hastalıktır ve bağırsak astarında ülserlerin oluşmasına neden olur. Bu sınıf, ülseratif kolitin farklı evrelerini temsil eden 1000 görüntü içermektedir. Erken teşhis ve uygun tedavi, hastaların yaşam kalitesini artırmada kritik rol oynar.

vii. **Boyanmış-Kaldırılmış Polipler (Dyed Lifted Polyps)**: Bu sınıf, poliplerin endoskopi sırasında özel boyalarla belirgin hale getirilmiş ve şişirilmiş hallerini içerir. Poliplerin boyanarak belirgin hale getirilmesi, doktorların poliplerin sınırlarını daha net bir şekilde görmesine ve bu sayede tam olarak çıkarılmasını sağlamasına yardımcı olur. Veri setinde bu sınıfı temsil eden 1000 görüntü yer almaktadır.

viii. **Boyanmış Rezeksiyon Kenarları (Dyed Resection Margins)**: Endoskopik işlemler sırasında poliplerin çıkarılmasının ardından kalan bölgelerin sınırları özel boyalarla belirgin hale getirilmektedir. Bu yöntem, çıkarılan alanın tam olarak kontrol edilmesine ve işlem sonrası kalan dokuların izlenmesine olanak tanır. Veri setinde, boyalı rezeksiyon sınırlarını gösteren 1000 görüntü bulunmaktadır.

Kvasir v2 veri seti, farklı anatomik ve patolojik yapıların ayrıntılı bir şekilde incelenmesi için araştırmacılara geniş bir veri kaynağı sağlamaktadır. Veri setinin sağladığı çeşitlilik, derin öğrenme modellerinin eğitilmesi ve test edilmesi için uygun bir yapıya sahiptir.

Şekil 3.1, çalışmada kullanılan farklı anatomik işaretçi ve patolojik bulgulara ait görüntüleri göstermektedir.



Şekil 3.1. Farklı anatomik işaretçi ve patolojik bulgulara ait görüntüler

3.2.Veri Önişleme

Model eğitimi öncesinde, veri kalitesini artırmak ve analiz sürecinde metodolojik tutarlılığı sağlamak amacıyla kapsamlı bir önişleme süreci uygulanmıştır. Bu süreç, Kvasir v2 veri setinde sıkça karşılaşılan görüntüsel artefaktların giderilmesine yönelik olarak Eker ve Solak (2024) tarafından önerilen önişleme stratejileri temel alınarak yapılandırılmış ve daha sistematik bir yaklaşımla genişletilmiştir. Gerçekleştirilen tüm önişleme adımları, aşağıda ilgili alt başlıklar altında ayrıntılı şekilde sunulmuştur.

3.2.1. Görüntülerdeki Yazıların Silinmesi

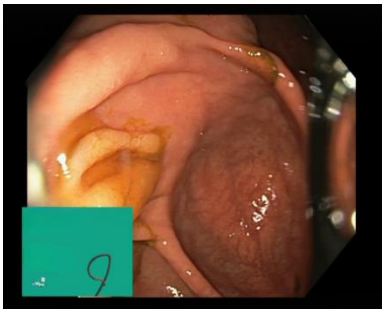
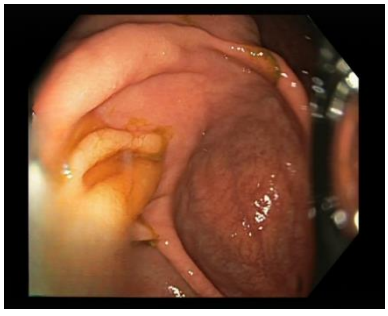
Görüntüler üzerinde yer alan yazılı ifadeler, hem sınıflandırma hem de segmentasyon görevlerinde modelin asıl görsel içerikten saparak dikkatini alakasız bölgelere yönlendirmesine neden olabileceği için sistematik biçimde veri setinden temizlenmiştir. Bu işlem, OpenCV kütüphanesi kullanılarak geliştirilen özel bir Python betiği aracılığıyla gerçekleştirilmiş; böylece yazıların model eğitimi sürecinde yanıltıcı veya yapay bir öğrenme kaynağı oluşturması engellenmiştir. Şekil 3.2, söz konusu yazı temizleme sürecine ilişkin örnek bir görsel sunmaktadır.

a.  b. 

Şekil 3.2. Görüntü üzerindeki yazıların silinmesinden önceki (a) ve silme işleminden sonraki (b) görüntü örneği

3.2.2. Görüntülerin Alt Köşelerinde Yer Alan Artefaktların Temizlenmesi

Endoskopik görüntülerin sol alt köşelerinde yer alan yeşil renkli artefaktlar, modelin dikkatini anlamlı anatomik yapılardan saptırarak hem sınıflandırma hem de segmentasyon görevlerinde öğrenme sürecini olumsuz yönde etkileyebilmektedir. Bu nedenle, söz konusu artefaktlar veri setinden sistematik olarak temizlenmiştir. Temizleme işlemi, yapay zekâ destekli çevrim içi bir platform olan Cleanup.pictures aracı kullanılarak görüntüler üzerinde manuel olarak gerçekleştirilmiştir. Bu araç, yeşil alanları doğal görsel dokuyla uyumlu biçimde kaldırarak görüntü bütünlüğünü korumakta ve modelin semantik olarak anlamlı bölgelere odaklanmasını sağlamaktadır. Şekil 3.3'te, bu önişleme adımına ilişkin bir görüntünün işlem öncesi ve sonrası halleri karşılaştırmalı olarak sunulmuştur.

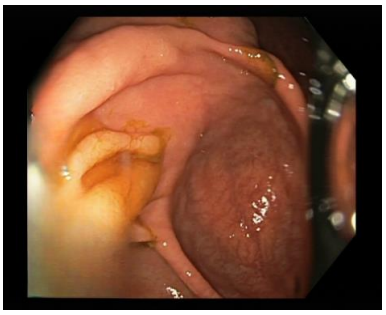
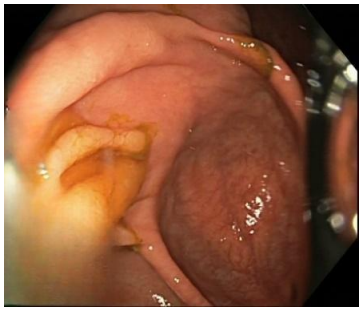
a.  b. 

Şekil 3.3. Sol alt köşedeki yeşil artefaktların temizlenmesinden önceki (a) ve temizleme işleminden sonraki (b) görüntü örneği.

3.2.3. Siyah Kenarların Kırpılması

Artefaktların temizlenmesinin ardından, görüntülerin etrafında bulunan ve tanısal açıdan herhangi bir bilgi içermeyen siyah kenarlıklar kırpılmıştır. Bu işlem,

modelin yalnızca anlamlı görsel bilgilere odaklanmasını sağlamak ve gereksiz hesaplama yükünü azaltmak amacıyla uygulanmıştır. Şekil 3.4, bu kırpma işlemine ait örnek bir görüntüyü göstermektedir.

a.  b. 

Şekil 3.4. Siyah kenarların kırpılmasından önceki (a) ve kırpma işleminden sonraki (b) görüntü örneği.

3.2.4. Görüntülerin Yeniden Boyutlandırılması

Temizleme ve kırpma işlemlerinin ardından, tüm görüntüler 224×224 piksel boyutlarına yeniden boyutlandırılmıştır. Bu sabit boyutlandırma, çalışmada kullanılan ViT (google/vit-base-patch16-224), Swin Transformer (microsoft/swin-tiny-patch4-window7-224), ConvNeXt, ConViT, SegFormer ve Swin Transformer + UPerNet gibi tüm modellerin giriş boyutu gereksinimlerini karşılamak amacıyla uygulanmıştır. Görüntü boyutlarının standartlaştırılması, hem mimariler arası tutarlı özellik çıkarımını sağlamakta hem de eğitim sürecinde hesaplama verimliliğini artırmaktadır.

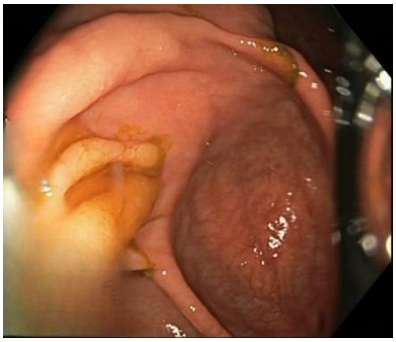
3.2.5. Kontrast İyileştirme: C.L.A.H.E. Uygulaması

Bu çalışmada görüntülerin kontrastını artırmak ve özellikle düşük aydınlatmalı, gölgeli ya da yapısal olarak belirsiz bölgelerdeki detayları ortaya çıkarmak amacıyla C.L.A.H.E. (Contrast Limited Adaptive Histogram Equalization - Kontrast Sınırlamalı Uyarlamalı Histogram Eşitleme) yöntemi kullanılmıştır. C.L.A.H.E., klasik histogram eşitleme yönteminin yerel olarak uygulanmış ve iyileştirilmiş halidir. Klasik histogram eşitleme, tüm görüntünün histogramını temel alarak kontrast artırımı yaptığı için, bazı bölgelerde detay kayıplarına ve görüntüde gürültü artışına neden olabilir. C.L.A.H.E. ise görüntüyü eşit boyutlarda küçük parçalara (örneğin 8x8 veya 16x16 piksellik

bloklara) ayırarak her blok için ayrı ayrı histogram eşitleme uygular. Bu sayede her bölgenin kendi iç dinamiklerine göre kontrastı artırılır ve yerel detaylar belirginleşir.

Ancak her bloğun kontrastını artırmak, bazı bölgelerde aşırı parlaklık ya da karanlık alanların oluşmasına yol açabilir. Bu durumu engellemek amacıyla C.L.A.H.E., bir "clip limit" parametresi ile çalışır. Bu parametre, bir yoğunluk değerinin histogramda ne kadar sıklıkla yer alabileceğini sınırlar. Eğer bu sınır aşılırsa, fazla yoğunluk değerleri komşu yoğunluklara dağıtılarak daha dengeli bir kontrast dağılımı sağlanır. Bu işlem sayesinde görüntüdeki ani parlaklık geçişleri, gölgeli alanlarda oluşabilecek bozulmalar ve gürültüler azaltılır. Bu çalışmada clipLimit=2.0 olarak belirlenmiş, böylece kontrast artışı dengelenmiştir.

Özellikle medikal görüntülerde, örneğin endoskopik görüntülerde, doku ve sınır detaylarının netliği tanı koyma ve model eğitimi açısından büyük önem taşır. Bu nedenle C.L.A.H.E., görüntülerin önişleme sürecine entegre edilmiş ve kontrastı artırılarak modelin polip gibi küçük ve düşük kontrastlı yapıları daha kolay öğrenmesi hedeflenmiştir. Yapılan ön gözlemler sonucunda clipLimit=2.0 değerleri ile en iyi sonuçlar elde edilmiş, böylece hem sınıflandırma hem de segmentasyon modellerinin doğruluk oranlarında artış gözlemlenmiştir. Şekil 3.5, C.L.A.H.E. uygulanmadan önceki ve sonraki bir görüntüyü karşılaştırmalı olarak göstermektedir.

a.  b. 

Şekil 3.5. C.L.A.H.E. uygulaması öncesi (a) ve sonrası (b) görüntü örneği.

3.2.6. Piksel Değerlerinin Normalizasyonu

Modelin giriş verileriyle uyumlu ve istikrarlı bir öğrenme süreci sağlanabilmesi için görüntülerin piksel değerlerine yönelik özel bir normalizasyon adımı uygulanmıştır. Bu çalışmada, model eğitimi öncesinde piksel değerleri doğrudan ImageNet istatistiklerine göre normalize edilmiştir (mean = [0.485, 0.456, 0.406]; std = [0.229, 0.224, 0.225]). Bu adım, önceden eğitilmiş model omurgalarıyla tam uyumluluk

sağlamak amacıyla gerçekleştirilmiş ve modellerin daha dengeli ve verimli bir şekilde öğrenmesini mümkün kılmıştır.

3.3. Sınıflandırma ve Segmentasyon İçin Kullanılan Derin Öğrenme Mimarileri

Bu tez kapsamında, sınıflandırma ve segmentasyon olmak üzere iki temel görev için çeşitli modern derin öğrenme mimarileri uygulanmıştır. Sınıflandırma görevinde, Vision Transformer (ViT), Swin Transformer, ConvNeXt ve ConViT modelleri değerlendirilmiş; segmentasyon görevinde ise SegFormer ile Swin Transformer + UPerNet birleşiminden oluşan hibrit yapı tercih edilmiştir. Kullanılan tüm modeller, saf Transformer mimarilerinin yanı sıra konvolüsyonel öncül bilginin entegre edildiği hibrit tasarımları da içermektedir. Bu bölümde, sınıflandırma ve segmentasyon modelleri, görev temelli alt başlıklar altında ayrıntılı olarak sunulmaktadır.

3.3.1. Vision Transformer (ViT)

Vision Transformer (ViT), Dosovitskiy ve diğ. (2020) tarafından önerilen ve bilgisayarla görü alanında çığır açan bir mimaridir. ViT, görüntüleri klasik konvolüsyonel işlemlerle analiz etmek yerine, doğal dil işleme alanından gelen Transformer mimarisine dayalı olarak işler. Bu model, giriş görüntüsünü sabit boyutlu parçalara (örneğin 16×16 piksellik patch'ler) ayırır ve her bir parçayı düzleştirerek sıralı bir girdi vektörüne dönüştürür. Daha sonra, bu parçalar pozisyonel kodlamayla birlikte çok katmanlı bir Transformer encoder'a beslenir. Bu yapı, tüm görüntü üzerindeki uzun menzilli bağıntıların paralel olarak öğrenilmesini sağlar. CNN'lerin tersine, ViT mimarisinde konvolüsyonel filtreler bulunmaz; bunun yerine öz-dikkat mekanizması, görüntü içindeki yapısal ilişkileri modellemenin ana aracıdır (Dosovitskiy ve diğ., 2020).

ViT'nin en büyük teorik katkılarından biri, modelin küresel bağlamı doğrudan modelleyebilmesidir. CNN'ler yerel pencereler üzerinden ilerlerken, ViT tüm yamalar arasında bağıntıları aynı anda ele alır. Bu durum, özellikle tıbbi görüntülerde görülen ince yapılar, sınır belirsizlikleri ve düşük kontrast alanlar gibi tanısal açıdan kritik bilgilerin bütüncül şekilde analiz edilmesini mümkün kılar (Shamshad ve diğ., 2022). Ayrıca, CNN'lerde görülen filtre bazlı "inductive bias" etkisinden kaçınılarak, verinin kendisinin yapısal özellikleri öğrenme sürecine yön vermesi sağlanır (Khan ve diğerleri., 2022).

Buna rağmen, ViT mimarisinin bazı önemli sınırlamaları da bulunmaktadır. İlk olarak, veri ihtiyacı oldukça yüksektir; ViT'nin başarılı şekilde eğitilebilmesi için genellikle milyonlarca görüntü içeren veri setlerine ihtiyaç duyulur. Küçük veri setlerinde, yeterli örnek sayısı sağlanmadığında modelin aşırı uyum (overfitting) eğilimi artar. Ayrıca, öz-dikkat işlemleri kare-zamanlı çalıştığı için ($O(n^2)$), hesaplama maliyeti çok yüksektir. Bu da büyük giriş görüntülerinin işlenmesini zorlaştırmakta ve ViT modellerinin genellikle yüksek bellek kapasiteli donanımlar gerektirmesine neden olmaktadır (Dosovitskiy ve diğerleri., 2020; Khan ve diğerleri., 2022).

Avantajlarına bakıldığında, ViT'nin; Global bağlamı etkin öğrenme yeteneği, Parametre sayısına rağmen yüksek genelleme gücü, Transfer öğrenme performansının CNN'lere göre daha yüksek olması gibi üstünlükleri bulunmaktadır (Azad ve diğ., 2023).

Dezavantajları arasında ise: Yüksek veri gereksinimi, Uzun eğitim süresi, Eğitim sırasında çok sayıda hiperparametre hassasiyeti yer almaktadır.

Bu çalışmada, Vision Transformer modeli özellikle polip ve benzeri küçük ama klinik olarak önemli yapıların küresel bağlama oturtularak sınıflandırılabilmesi amacıyla tercih edilmiştir. Kvasir v2 veri seti üzerinde yapılan deneylerde, ViT'nin bu avantajları sayesinde geleneksel CNN tabanlı modellere göre daha yüksek doğruluk ve daha stabil sonuçlar ürettiği gözlemlenmiştir.

3.3.2. Swin Transformer

Swin Transformer, Liu ve diğ. (2021) tarafından geliştirilen ve klasik Vision Transformer (ViT) mimarisinin hesaplama karmaşıklığını azaltmak ve görsel görevlerde daha verimli hale getirmek üzere tasarlanmış hiyerarşik bir Transformer mimarisidir. ViT'nin her yama arasında tüm ilişkileri modellemeye çalışmasının aksine, Swin Transformer bu işlemi daha ölçeklenebilir hale getirmek için kayan pencereler (shifted windows) kullanır. Bu yapı, öz-dikkat mekanizmasını yalnızca küçük yerel bölgelerde uygular, böylece kare-zamanlı karmaşıklık ($O(n^2)$) yerine doğrusal karmaşıklık ($O(n)$) elde edilir. Bununla birlikte, pencereler arasında kayan mekanizma sayesinde, küresel bağlam da aşamalı olarak öğrenilebilir hale gelir (Liu ve diğerleri., 2021).

Swin Transformer'ın mimarisi dört aşamalı bir yapıdan oluşur. Her aşama, sabit boyutlu pencereler içinde çoklu öz-dikkat katmanları içerir ve ardından patch merging ile uzamsal çözünürlük azaltılırken kanal sayısı artırılır. Bu, klasik CNN

mimarilerindeki havuzlama ve kanal genişletme işlemine benzer bir yapı sunar. Ayrıca bu hiyerarşik özellik öğrenimi sayesinde hem düşük seviyeli kenar bilgileri hem de yüksek seviyeli bağlamsal yapılar öğrenilebilir (Khan ve diğerleri., 2022).

Swin Transformer'ın avantajları çok yönlüdür. Öncelikle, yerel pencere tabanlı yapısı nedeniyle büyük görüntüler üzerinde çok daha verimli çalışır. Bu da yüksek çözünürlüklü tıbbi görüntülerin analizi için idealdir. İkinci olarak, hiyerarşik yapısı sayesinde çok ölçekli bilgi çıkarımı yapılabilir, bu da polip gibi küçük yapıları geniş bağlamda analiz edebilmek açısından önemlidir. Ayrıca, Transformer yapısından gelen parametrik esneklik sayesinde transfer öğrenme senaryolarında da yüksek başarı sağlamaktadır.

Ancak dezavantajları da bulunmaktadır. Örneğin, pencere tabanlı dikkat mekanizması her ne kadar verim sağlasa da, ilk katmanlarda küresel bağlamı doğrudan öğrenemez; bu ancak daha derin katmanlarda, kaydırmalı pencere yaklaşımıyla mümkün olur. Ayrıca, Swin Transformer klasik CNN'lere göre daha fazla bellek tüketir ve eğitim süreci sırasında dikkatli hiperparametre ayarı gerektirir (Shamshad ve diğ., 2022).

Bu çalışmada Swin Transformer, sınıflandırma görevlerinde lokal ve küresel bağlamı birlikte öğrenebilmesi, düşük ve yüksek seviyeli öznitelikleri bir arada sunabilmesi ve Kvasir v2 veri setinde daha kararlı sınıflandırma performansı göstermesi nedeniyle tercih edilmiştir. Özellikle polip gibi küçük ve düşük kontrastlı yapıların bulunduğu görüntülerde, lokal detayların öğrenimi modelin başarısını artırmıştır.

3.3.3. ConvNeXt

ConvNeXt, Liu ve diğ. (2022) tarafından geliştirilen ve klasik konvolüsyonel sinir ağlarını (CNN) Transformer mimarisine yaklaştırmak amacıyla yeniden tasarlanmış modern bir konvolüsyonel modeldir. ViT'nin ve Swin Transformer'ın başarısı, bilgisayarla görme alanında saf Transformer yapılarına olan ilgiyi artırmış; ancak bu modellerin yüksek hesaplama maliyeti ve eğitimde zorlukları bazı araştırmacıları CNN mimarilerini iyileştirerek benzer performanslara ulaşmaya yönlendirmiştir. ConvNeXt bu motivasyonla, özellikle ResNet mimarisini modern derin öğrenme trendleriyle güncelleyerek tasarlanmıştır (Liu ve diğerleri., 2022).

ConvNeXt'in mimarisi, temel olarak ResNet-50 ile benzer bir iskelet kullanır; buna karşın bazı önemli yenilikler içerir. Öncelikle, klasik ReLU yerine GELU

(Gaussian Error Linear Unit) aktivasyon fonksiyonu kullanılarak Transformer mimarilerinde görülen daha yumuşak aktivasyon davranışı sağlanmıştır. Ayrıca Batch Normalization yerine Layer Normalization tercih edilmiştir, bu da Transformer'lara özgü bir norm türüdür. Daha da önemlisi, konvolüsyon çekirdek boyutu 3×3 yerine 7×7 olarak belirlenmiş ve kanal sayıları arttırılarak daha geniş bir receptive field sağlanmıştır. Bu tasarımsal değişiklikler sayesinde ConvNeXt, saf Transformer modellerine yakın doğruluk değerlerine ulaşırken, klasik CNN'lerin eğitim kolaylığını ve donanım uyumluluğunu korumuştur (Khan ve diğerleri., 2022).

ConvNeXt'in avantajları çok yönlüdür. Öncelikle, eğitim süreci oldukça kararlıdır ve büyük veri kümelerine bağımlılığı düşüktür. Ayrıca, GPU donanımına optimize edilmiş konvolüsyon işlemleri sayesinde daha düşük eğitim süresi ve daha az bellek kullanımı sunar. Bu, özellikle tıbbi görüntüleme gibi sınırlı kaynaklarda çalışan sistemlerde önemli bir avantajdır. Ayrıca, ViT ve benzeri Transformer mimarilerine kıyasla, ConvNeXt modelleri genellikle daha az parametreyle benzer performans sergileyebilir.

Bununla birlikte bazı sınırlamaları da vardır. Örneğin, global bağlamı doğrudan öğrenemez; tüm görüntü ilişkileri küçük çekirdekler üzerinden yerel olarak çıkarılır. Bu, düşük kontrastlı veya bağlamsal bilgi gerektiren görevlerde Transformer mimarilerine göre dezavantaj yaratabilir. Ayrıca, her ne kadar modernize edilmiş olsa da, CNN yapısının doğasında bulunan lokal öznitelik ağırlıklı öğrenme yaklaşımı, bazı sınıflandırma görevlerinde yeterli esnekliği sağlayamayabilir Shamshad ve diğ., 2022).

Bu çalışmada ConvNeXt modeli, sınıflandırma görevlerinde yüksek doğrulukla birlikte hesaplama maliyetinin düşük olması, eğitim sürecindeki stabilitesi ve transfer öğrenmeye uygunluğu nedeniyle tercih edilmiştir. Kvasir v2 veri seti gibi nispeten küçük ve etiketli tıbbi görüntülerden oluşan bir veri yapısında, CNN temelli modellerin performansı Transformer mimarilerine çok yakın olabilir. ConvNeXt, polip gibi yapısal detayların güçlü şekilde öğrenilebildiği bir sınıflandırma çerçevesi sunmuştur.

3.3.4. ConViT

ConViT (Convolutional Vision Transformer), d'Ascoli ve diğ. (2021) tarafından geliştirilen, konvolüsyonel sinir ağlarının yerel bilgi çıkarımındaki başarısını ve Transformer mimarisinin küresel bağlam öğrenme kabiliyetini bir araya getirmeyi amaçlayan hibrit bir modeldir. Bu modelin temel çıkış noktası, klasik ViT yapılarının

sahip olduğu esnek öğrenme kapasitesine rağmen, konumsal ilişkileri öğrenmede yeterince güçlü ön kabullere sahip olmamasıydı. ConViT, bu problemi çözmek amacıyla Soft Convolutional Inductive Bias olarak adlandırılan yeni bir mekanizma önerir. Bu sayede model, hem CNN'lerdeki gibi yerel yapıların güçlü şekilde çıkarılmasını sağlar, hem de Transformer'ların global bağlamı öğrenme avantajlarını korur (d'Ascoli ve diğerleri., 2021).

ConViT mimarisi, Transformer encoder bloklarına konvolüsyonel ön işlemleme katmanları entegre ederek çalışır. Modelin ilk katmanlarında uygulanan girişe duyarlı convolutional gating mekanizması, giriş verisinden hangi alanların dikkatle işleneceğini yumuşak bir şekilde kontrol eder. Bu özellik, hem yerel detayları (örneğin kenar bilgisi ve doku yapısı gibi) hem de uzamsal konumu önemli olan küçük yapıları yüksek doğrulukla tespit etmeye olanak tanır. Aynı zamanda bu yaklaşım, Transformer'ın pozisyonel kodlamaya olan bağımlılığını azaltır, bu da görüntülerin döndürülmesi, ölçeklenmesi gibi işlemlere karşı daha dayanıklı hale gelmesini sağlar (Khan ve diğerleri., 2022).

ConViT'nin avantajları şunlardır; Yerel ve küresel öznitelik öğrenimi aynı anda yapılabilir. Bu, özellikle polip gibi detayların çevresel bağlamla birlikte anlamlandırılması gereken tıbbi görüntülerde önemlidir. ViT gibi büyük veri setlerine bağımlı olmaksızın, daha az veriyle daha kararlı öğrenme sağlar. Görsel dikkat mekanizmasını konvolüsyonel önyargılarla yönlendirerek daha doğru konumsal farkındalık sunar.Yapay olarak eklenen pozisyonel kodlamaya olan ihtiyacı azaltır, dolayısıyla mekânsal dönüşümler karşısında daha esnektir.

Ancak ConViT'nin bazı sınırlamaları da vardır. Modelin hibrit yapısı, saf ViT ya da CNN'lere kıyasla daha karmaşık bir mimari tasarımı gerektirir. Bu durum hem eğitim sırasında hiperparametre ayarlamasını daha hassas hale getirir hem de donanım uyumluluğunu kısıtlayabilir. Ayrıca, tüm avantajlarına rağmen dikkat mekanizmasının öğrenimi hâlâ büyük ölçüde veri kalitesine ve çeşitliliğine bağlıdır. Çok karmaşık görüntülerde dikkat dağılımı istenmeyen şekilde sapabilir (d'Ascoli ve diğ., 2021; Azad ve diğ., 2023).

Bu çalışmada ConViT modeli, özellikle polip gibi kenar tabanlı, düzensiz sınırlara sahip yapıların sınıflandırılmasında tercih edilmiştir. Modelin hem yerel ayrıntıları hem de genel bağlamı dikkate alması, Kvasir v2 veri setinde yüksek sınıflandırma doğruluğu ile sonuçlanmıştır. Ayrıca, klasik CNN yapılarının

öğrenemediği karmaşık doku yapılarında ConViT'nin daha üstün performans gösterdiği gözlemlenmiştir.

3.3.5. SegFormer

SegFormer, Xie ve diğ. (2021) tarafından geliştirilen ve görüntü segmentasyonu görevleri için optimize edilmiş hafif, modüler ve saf Transformer tabanlı bir mimaridir. Geleneksel segmentasyon modellerinde genellikle encoder kısmında CNN veya Transformer, decoder kısmında ise çok katmanlı upsampling mekanizmaları bulunurken, SegFormer bu süreci basitleştirmiştir. Modelin en büyük yeniliği, klasik konvolüsyonel önyargılardan arındırılmış bir encoder ile, oldukça hafif ve etkili bir MLP tabanlı decoder'ı bir araya getirmesidir. Böylece yüksek doğruluk ve gerçek zamanlı kullanım için uygunluk bir araya getirilmiştir (Xie ve diğerleri., 2021).

SegFormer mimarisi, çok seviyeli özellik çıkarımı yapabilen hierarchical Transformer encoder blokları içerir. Bu encoder, çeşitli çözünürlüklerdeki bilgileri paralel olarak işler ve segmentasyon için gerekli hem düşük seviyeli detayları hem de yüksek seviyeli bağlamsal öznitelikleri yakalayabilir. Diğer birçok Transformer modelinde olduğu gibi, SegFormer da öz-dikkat mekanizması kullanır, buna karşın CNN'lerin aksine, bu işlemi herhangi bir pozisyonel kodlama kullanmadan gerçekleştirir. Bu yönüyle, pozisyonel duyarlılığı azaltır ve modelin genel uzamsal esnekliğini artırır (Khan ve diğerleri., 2022).

SegFormer'ın başlıca avantajları şunlardır; Hafifliği ve hız avantajı: MLP tabanlı decoder sayesinde eğitim ve çıkarım süreci çok hızlıdır. Bu, medikal uygulamalarda düşük gecikmeli sistemler için önemlidir. Pozisyonel kodlama olmadan çalışabilme yeteneği, onu döndürme, ölçekleme gibi görüntü dönüşümlerine karşı daha esnek hale getirir. Çok seviyeli encoder yapısı, hem detaylı segmentasyon hem de genel bağlam öğrenimi sağlar. End-to-end eğitime uygundur ve herhangi bir özel önişleme veya tasarımsal modifikasyon gerektirmez.

Dezavantajları ise; CNN'lerin sunduğu güçlü lokal öznitelik çıkarımı yer yer eksik kalabilir, özellikle sınırlı sayıda örneğe sahip tıbbi veri kümelerinde düşük seviyeli detayları öğrenmekte zorlanabilir. Segmentasyon doğruluğu, kullanılan veri kalitesine ve örnek dengesine karşı hassastır. Decoder MLP yapısı her ne kadar verimli olsa da, bazı karmaşık geometrik yapılarda ince kenarların doğru tahmin edilmesini zorlaştırabilir (Azad ve diğ.,2023).

Bu çalışmada SegFormer modeli, özellikle polip segmentasyonu gibi tıbbi görüntü segmentasyonunda hız, hassasiyet ve çözünürlük dengesini ideal biçimde sunduğu için tercih edilmiştir. Kvasir v2 veri seti üzerinde yapılan deneylerde, SegFormer modelinin polip gibi küçük yapıların doğru şekilde ayrıştırılması açısından başarılı olduğu, sınıf dengesizliğinden kaynaklanan problemleri minimize ettiği ve inference süresinin oldukça düşük olduğu gözlemlenmiştir.

3.3.6. Swin Transformer + UPerNet

Swin Transformer + UPerNet, görüntü segmentasyonu alanında çok ölçekli bilgi entegrasyonu ve küresel bağlam öğrenimi gerektiren görevlerde kullanılan güçlü bir birleşik mimaridir. Bu yapı, encoder kısmında Swin Transformer mimarisini, decoder kısmında ise Unified Perceptual Parsing Network (UPerNet) mimarisini kullanarak çalışır. Böylece hem görsel dikkat (attention) mekanizmasıyla global bağlam öğrenimi sağlanır hem de farklı çözünürlüklerden gelen öznitelikler etkili şekilde birleştirilerek segmentasyon kalitesi artırılır (Xiao ve diğerleri., 2021; Liu ve diğerleri., 2021).

Swin Transformer, lokal pencere temelli öz-dikkat mekanizması ile yüksek çözünürlüklü görüntüleri verimli şekilde işlerken, kayan pencere stratejisi sayesinde global bağlamı da aşamalı olarak öğrenebilir. Bu özelliği, özellikle tıbbi görüntülerde polip gibi küçük ve detaylı yapıların hem lokal sınırlarının hem de bağlamsal ilişkilerinin öğrenilmesine olanak tanır. Swin'in hiyerarşik yapısı, çok seviyeli öznitelik çıkarımı ile UPerNet'in kullandığı çoklu ölçekli öznitelik birleştirme işlemini desteklemek açısından oldukça uygundur (Liu ve diğerleri., 2021).

UPerNet ise, segmentasyon çıktısını iyileştirmek için Feature Pyramid Network (FPN) ve Pyramid Pooling Module (PPM) gibi yapıların birleşiminden oluşur. Encoder'dan gelen çok seviyeli özellik haritaları FPN ile yukarı örneklenir ve bu katmanlar arası bilgiler PPM ile küresel bağlama dayalı olarak daha sağlam şekilde bütünleştirilir. Bu sayede hem düşük seviyeli detaylar hem de yüksek seviyeli semantik bilgi tek bir birleşik öznitelik haritasında toplanabilir (Xiao ve diğerleri., 2018).

Bu mimarinin avantajları şunlardır: Yerel ve küresel öznitelik öğrenimi birlikte yapılır: Swin'in dikkat mekanizması ve UPerNet'in çok seviyeli yapı entegrasyonu birleşerek güçlü bir segmentasyon performansı sunar. Karmaşık yapıları ayırt edebilme: Özellikle düzensiz kenarlara sahip küçük poliplerin segmentasyonu gibi zorlayıcı görevlerde başarılıdır. Esnek ve yeniden kullanılabilir yapı: Encoder ve decoder

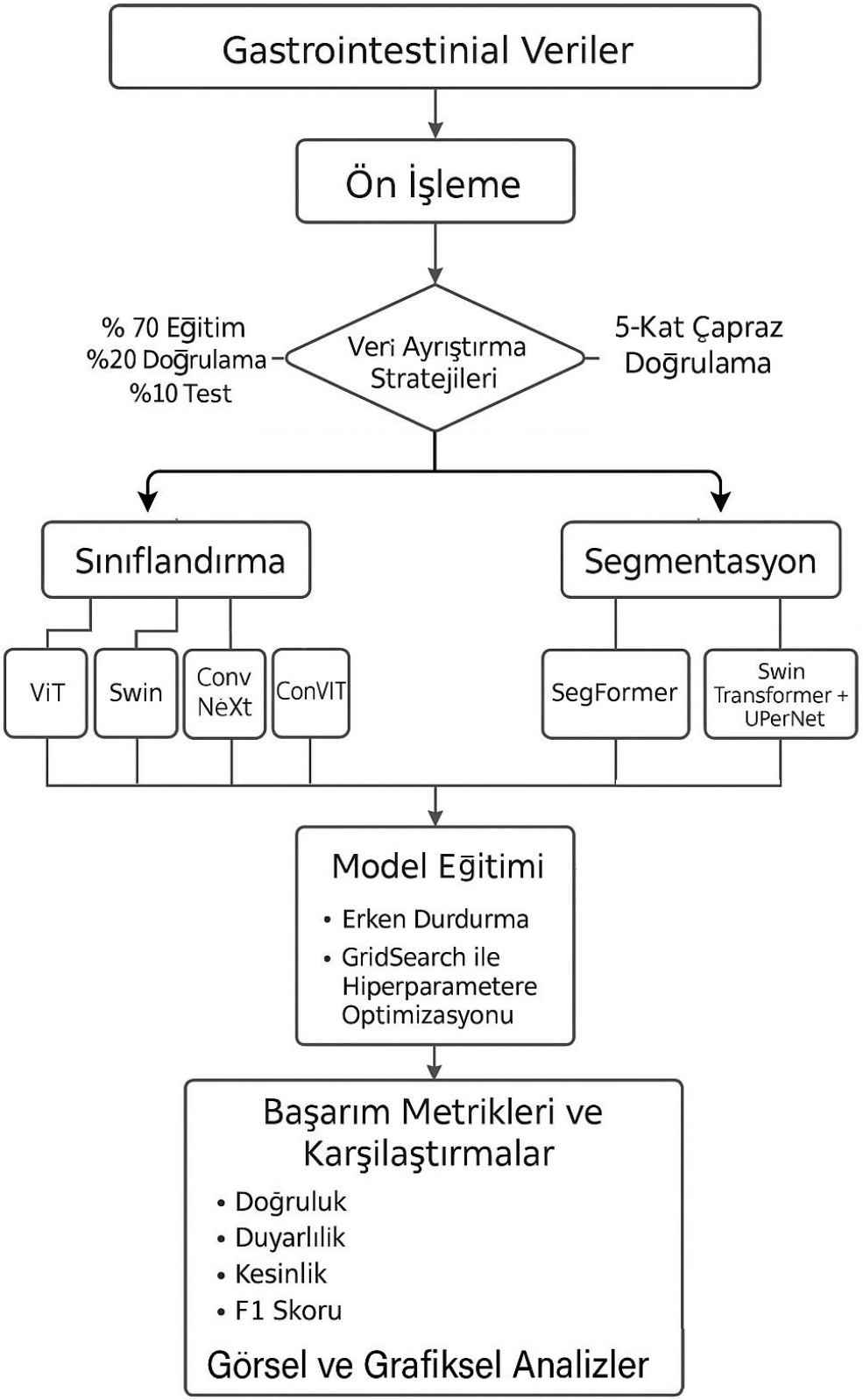
bağımsız olarak değiştirilebilir, bu da modülerlik ve transfer öğrenme açısından avantaj sağlar.

Dezavantajları ise: Swin Transformer ve UPerNet her ikisi de karmaşık mimarilere sahip olduğundan, birleşik model yüksek hesaplama gücü ve daha fazla GPU belleği gerektirir. Eğitim süresi, daha sade mimarilere göre uzundur ve hiperparametre ayarı karmaşıklaşabilir. Özellikle düşük veri senaryolarında modelin aşırı öğrenme (overfitting) riski artabilir.

Bu çalışmada Swin + UPerNet mimarisi, özellikle detaylı dokuya sahip, sınırlı veriyle eğitilmiş medikal görüntüler üzerinde üstün segmentasyon başarısı sağladığı için tercih edilmiştir. Kvasir v2 veri setinde yapılan deneylerde, bu mimari yüksek Dice ve IoU skorları ile dikkat çekmiş, segmentasyon sınırlarında daha az hata üretmiş ve çoklu çözünürlükte bilgi işleyebilme yeteneği ile diğer modellere göre daha stabil sonuçlar vermiştir.

3.4. Model Eğitimi ve Optimizasyon Süreci

Sınıflandırma ve segmentasyon görevlerine yönelik model geliştirme süreci, endoskopik görüntülere uygulanan önişleme adımlarıyla başlamış; ardından görüntüler, %70 eğitim, %20 doğrulama ve %10 test oranlarında rastgele veri ayrıştırma yöntemiyle ve 5 kat çapraz doğrulama stratejisiyle eğitim ve değerlendirme amaçlı ayrılmıştır. Eğitim aşamasında, farklı derin öğrenme mimarileri kullanılarak modeller oluşturulmuş; hiperparametre optimizasyonu için Grid Search algoritması uygulanmış ve aşırı öğrenmeyi önlemek amacıyla erken durdurma (early stopping) mekanizması devreye alınmıştır. Model performansları, doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metriklerin yanı sıra grafiksel analizlerle değerlendirilmiştir. Şekil 3.6’da sunulan diyagram, çalışmada izlenen bu genel süreci görsel olarak özetlemektedir.



Şekil 3.6. Çalışmada izlenen yöntemsel adımları özetleyen genel iş akışı diyagramı.

Bu çalışmada kullanılan tüm sınıflandırma ve segmentasyon modelleri, iki temel eğitim stratejisiyle değerlendirilmiştir: rastgele veri ayrıştırma (random split) ve 5 kat çapraz doğrulama (k-fold cross validation).

Rastgele ayrıştırma stratejisi kapsamında, önişleme adımlarının ardından kalite ve yapısal tutarlılığı artırılmış olan endoskopik görüntüler, eğitim, doğrulama ve test alt kümelerine rastgele ayrılmıştır. Bu süreçte, Kvasir v2 veri setindeki toplam 8000 görüntüden 5600'ü eğitim, 1600'ü doğrulama ve 800'ü test için kullanılmıştır. Bu oranlar sırasıyla %70–%20–%10 olarak belirlenmiştir. Rastgele örnekleme sürecinden kaynaklı olarak sınıflar arasında küçük farklılıklar gözlenmiş olsa da, her bir sınıfın tüm alt kümelerde yeterli sayıda örnekle temsil edilmesi sağlanmıştır. Böylece modelin dengeli bir şekilde öğrenmesi ve her sınıf özelinde güvenilir performans değerlendirmesi yapılması amaçlanmıştır. Eğitim, doğrulama ve test kümelerine ait sınıf dağılımları Çizelge 3.1'de sunulmuştur.

Çizelge 3.1. Eğitim, doğrulama ve test kümeleri için sınıf bazlı örnek dağılımı

Sınıf	Eğitim	Doğrulama	Test
Boyanmış kaldırılmış polipler	708	197	95
Boyanmış rezeksiyon kenarları	687	206	107
Özofajit	691	216	93
Normal sekum	701	200	99
Normal pilor	696	198	106
Polipler	691	221	88
Ülseratif kolit	728	167	105
Z çizgisi	698	195	107

İkinci strateji olarak uygulanan 5 kat çapraz doğrulama, veri setinin her bir alt bölümde farklı kombinasyonlarla eğitilmesini sağlayarak modelin genellenebilirliğini artırmayı ve aşırı öğrenme riskini azaltmayı amaçlamaktadır. Bu yöntemde, veri seti eşit büyüklükte beş alt kümeye ayrılmakta; her iterasyonda bu alt kümelerden biri doğrulama (veya test) için kullanılırken kalan dört alt küme eğitim için kullanılmaktadır. Bu işlem beş kez tekrarlanarak her alt küme bir kez doğrulama verisi olarak değerlendirilmiş olur. Böylece model, tüm veri üzerinde eğitilip test edilme şansı bulur ve sonuçlar daha istikrarlı, güvenilir ve genellenebilir hale gelir. Özellikle küçük ve dengesiz veri kümelerinde, modelin tüm örnekleri öğrenme sürecine dahil etmesini sağladığı için doğruluktan ödün vermeden değerlendirme yapma imkânı sunar.

Model başarımının yalnızca veri bölme stratejilerine değil, aynı zamanda hiperparametre seçimlerine de duyarlı olması nedeniyle, tüm mimariler için hiperparametre optimizasyonu süreci büyük önem taşımaktadır. Bu kapsamda, modellerin eğitiminde kullanılan öğrenme oranı, batch boyutu ve optimizasyon algoritması gibi hiperparametreler, Grid Search algoritması kullanılarak sistematik biçimde değerlendirilmiştir. Böylece, çapraz doğrulama ile birlikte Grid Search'ın bütüncül uygulanması, modellerin her biri için hem eğitim koşullarının optimize edilmesini hem de karşılaştırmalı analizlerin objektif ve tutarlı biçimde gerçekleştirilmesini mümkün kılmıştır.

Grid Search, her hiperparametrenin önceden tanımlanmış bir değer aralığına göre tüm olası kombinasyonlarını sistematik olarak tarayan ve her bir yapılandırmayı doğrulama seti üzerindeki performansa göre test eden bir yöntemdir. Bu algoritma, modelin başarımını en üst düzeye çıkaran hiperparametre konfigürasyonunu belirlemek amacıyla her olasılığı sırasıyla dener ve sonuçları performans metrikleri üzerinden değerlendirir. Böylece, her model için parametre alanı boyunca otomatik bir tarama gerçekleştirilerek en iyi sonuçları veren yapılandırmalar belirlenmiş olur.

Gerçekleştirilen tez çalışmasında, her model, belirlenen hiperparametre aralıklarında sistematik olarak eğitilmiş ve doğrulama seti üzerinde elde edilen sonuçlara göre değerlendirilmiştir. Bu süreç sonunda, en yüksek doğruluk, F1 skoru veya Dice katsayısı sağlayan yapılandırmalar, ilgili modelin nihai parametre kombinasyonu olarak seçilmiştir. Bu yaklaşım, her mimarinin en iyi performans gösterdiği koşullarda değerlendirilmesini ve karşılaştırmalı analizlerin adil ve tekrarlanabilir biçimde yapılmasını mümkün kılmıştır.

Eğitim sürecinde erken durdurma (early stopping) mekanizması uygulanmış; kriter olarak doğrulama kaybı (validation loss) esas alınmıştır. Doğrulama kaybında art arda 5 epok boyunca herhangi bir iyileşme gözlemlenmediğinde eğitim süreci otomatik olarak sonlandırılmıştır. Bu yöntem, aşırı öğrenmenin (overfitting) önlenmesi ve eğitim süresinin optimize edilmesi amacıyla tercih edilmiştir. Ayrıca, görev bazlı olarak farklı kayıp fonksiyonları kullanılmıştır: sınıflandırma görevlerinde Cross Entropy Loss, segmentasyon görevlerinde ise Dice Loss fonksiyonu uygulanmıştır. Bu seçimler, her görevin doğasına uygun hata hesaplama yaklaşımının benimsenmesini sağlamıştır.

3.5. Performans Değerlendirme Metrikleri

Gerçekleştirilen tez çalışmasında her modelin performansı, belirli görev türlerine uygun değerlendirme metrikleri kullanılarak analiz edilmiş ve sonuçlar hem rastgele veri bölme (Random Split) hem de 5 katlı çapraz doğrulama (K-Fold Cross Validation) stratejileri için ayrı ayrı raporlanmıştır. Şekil 3.7, değerlendirme metriklerinin kullanıldığı çok sınıflı bir konfüzyon matrisi yapısını göstermektedir (Markoulidakis ve diğ. 2021).

		Tahmin Edilen Sınıf			
		C1	C2	…	CN
Gerçek Sınıf	C1	C1,1	FP	…	C1,N
	C2	FN	TP	…	FN
	…	…	…	…	…
	CN	CN,1	FP	…	CN,N

Şekil 3.7. Çok sınıflı konfüzyon matrisi

Şekil 3.7'de görülen konfüzyon matrisi, çok sınıflı bir sınıflandırma problemi için genel bir yapıdır. Bu yapı temel alınarak değerlendirme metrikleri hesaplanır.

- TP (Doğru Pozitif- True Positive): Sınıfın doğru tahmin edilen örnek sayısı.
- FP (Yanlış Pozitif- False Positive): Sınıfa ait olmadığı halde o sınıfa tahmin edilen örnek sayısı.
- FN (Yanlış Negatif- False Negative): Sınıfa ait olduğu halde başka bir sınıfa tahmin edilen örnek sayısı.
- TN (Doğru Negatif- True Negative): Diğer tüm doğru tahminler.

Doğruluk (Accuracy), modelin tüm sınıflar üzerindeki genel doğruluğunu gösterir. Tüm doğru tahminlerin (hem pozitif hem negatif) toplam tahmin sayısına oranı olarak Denklem 3.1'de görüldüğü gibi hesaplanır.

$$\textbf{\textit{Doğruluk}} = \frac{\sum_{i=1}^{N} TP\ (Ci)}{\sum_{i=1}^{N}\sum_{j=1}^{N} Ci,j} \tag{3.1}$$

Kesinlik (Precision), bir sınıf için yapılan pozitif tahminlerin ne kadarının gerçekten doğru olduğunu gösterir. Yanlış pozitiflerin sayısı fazla ise bu metrik düşük değer alacaktır. Denklem 3.2'de görüldüğü gibi hesaplanır.

$$\textbf{\textit{Kesinlik}}_i = \frac{TP(Ci)}{TP(Ci)+FP(Ci)} \tag{3.2}$$

Duyarlılık (Recall), gerçek pozitif örneklerin ne kadarının doğru şekilde tahmin edildiğini gösterir. Kaçırılan pozitifler (FN) fazlaysa bu değer düşer. Denklem 3.3, duyarlılık metriğinin nasıl hesaplandığını göstermektedir.

$$\textbf{\textit{Duyarlılık}}_i = \frac{TP(Ci)}{TP(Ci)+FN(Ci)} \tag{3.3}$$

F1-Skoru, kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır. İki metrik arasında denge kurar ve ikisinin de yüksek olduğu durumlarda yüksek çıkar. Bu metrik Denklem 3.4'te görüldüğü gibi hesaplanır.

$$\textbf{\textit{F1}}_i = 2 \times \frac{Duyarlılık_{\text{i}}\ x\ Kesinlik_{\text{i}}}{Duyarlılık_{\text{i}} + Kesinlik_{\text{i}}} \tag{3.4}$$

Bu metrikler, tüm sınıflar üzerinden hesaplanan makro ortalama değerleriyle raporlanmış, böylece sınıf dengesizliklerinden kaynaklanabilecek yanlılıkların önüne geçilmiştir. Makro ortalama, her bir sınıf için ayrı ayrı hesaplanan metriklerin (örneğin doğruluk, duyarlılık, özgüllük, F1 skoru gibi) aritmetik ortalamasının alınmasıyla elde edilir. Bu yöntem, her sınıfı eşit derecede önemseyerek değerlendirme yapar ve özellikle dengesiz veri setlerinde daha adil bir genel performans ölçümü sağlar. Kesinlik, duyarlılık ve F1 skoru değerlerinin makro ortalama formülleri sırasıyla Denklem 3.5, Denklem 3.6 ve Denklem 3.7'de gösterilmektedir.

$$\textbf{\textit{Kesinlik macro}} = \frac{1}{N}\sum_{i=1}^{N} Kesinlik_{\text{i}} \tag{3.5}$$

$$\textbf{\textit{Duyarlılık macro}}=\frac{1}{N}\sum_{i=1}^{N} Duyarlılık_{\mathrm{i}} \tag{3.6}$$

$$\textbf{\textit{F1 macro}}=2\times\frac{Duyarlılık\ macro\ x\ Kesinlik\ macro}{Duyarlılık\ macro+\ Kesinlik\ macro} \tag{3.7}$$

Segmentasyon görevleri kapsamında, model çıktıları Denklem 3.8'de verilen Dice Katsayısı ve Denklem 3.9'da görülen Kesişim / Birleşim Oranı (Intersection over Union - IoU) metrikleri ile değerlendirilmiştir. Bu metrikler, özellikle ikili (binary) segmentasyon problemlerinde yaygın olarak kullanılmakta olup, modelin tahmin ettiği nesne bölgelerinin gerçek (ground truth) bölgelerle olan örtüşme derecesini nicel olarak ifade eder.

Bu değerlendirme sürecinde, her görüntü için tahmin edilen segmentasyon maskesi ile gerçek maske karşılaştırılarak, 2×2 boyutunda ikili bir konfüzyon matrisi oluşturulmaktadır. Şekil 3.8, polip segmentasyonu için kullanılan bu konfüzyon matrisini göstermektedir.

	Tahmin: Polip (1)	Tahmin: Arka Plan (0)
Gerçek: Polip (1)	True Positive (TP)	False Negative (FN)
Gerçek: Arka Plan (0)	False Positive (FP)	True Negative (TN)

Şekil 3.8. İkili konfüzyon matrisi

Şekil 3.8'de sunulan ikili konfüzyon matrisinde, her bir bileşen aşağıdaki şekilde tanımlanmaktadır:

- TP (Doğru Pozitif - True Positive): Gerçekten polip olan piksellerin, model tarafından doğru bir şekilde polip olarak tahmin edilmesi durumudur.
- FP (Yanlış Pozitif - False Positive): Gerçekte arkaplan olan piksellerin, model tarafından hatalı bir şekilde polip olarak sınıflandırılmasıdır.
- FN (Yanlış Negatif - False Negative): Gerçekte polip olan piksellerin, model tarafından yanlışlıkla arkaplan olarak değerlendirilmesidir.
- TN (Doğru Negatif - True Negative): Gerçekten arkaplan olan piksellerin, model tarafından doğru bir şekilde arkaplan olarak tahmin edilmesidir.

Dice katsayısı, modelin tahmin ettiği segmentasyon maskesi ile gerçek maskenin benzerliğini ölçer (Denklem 3.8). 0 ile 1 arasında değer alır. 1 değeri tam örtüşme, 0 değeri hiç örtüşme olmadığını ifade eder.

$$\textbf{\textit{Dice Katsayısı}} = (2 \times TP) / (2 \times TP + FP + FN) \quad (3.9)$$

Yüksek Dice Katsayı değeri, segmentasyon maskesinin doğru şekilde sınırlandığını gösterir. Tıbbi görüntülerde sık kullanılır çünkü küçük bölgelerin sınırları kritik olabilir.

IoU, tahmin ve gerçek segmentasyon bölgelerinin kesişim alanının, birleşim alanına oranıdır. 0 ile 1 arasında değer alır. 1 değeri tam örtüşmeyi gösterir. IoU daha katı bir metriktir. Genelde 0,5 üzeri kabul edilebilir, 0,75 üzeri güçlü sonuç olarak değerlendirilir. Segmentasyon yarışmalarında standart olarak kullanılır.

$$\textbf{\textit{IoU}} = TP / (TP + FP + FN) \quad (3.10)$$

Sayısal değerlendirmelere ek olarak, modellerin çıktıları aşağıda belirtilen görsel ve grafiksel analizlerle desteklenmiştir:

- **Konfüzyon matrisi görselleri**,
- **Segmentasyon sonuçlarına ilişkin örnek çıktı görselleri**

Sınıflandırma işlemlerinin sonucunda elde edilen çok sınıflı konfüzyon matrisleri tüm sınıflar için sabit bir sayısal etiketleme (0–7) kullanılarak oluşturulmuş ve her sınıf farklı modellerde aynı sayısal gösterimle ifade edilmiştir. Bu sayede, sınıflandırma çıktılarının görsel olarak karşılaştırılabilirliği sağlanmış ve sınıf bazında hataların analiz edilmesi mümkün hale getirilmiştir. Sayısal sınıf etiketleri, Kvasir v2 veri setindeki kategorilere şu şekilde karşılık gelmektedir:

0 – Boyanmış-kaldırılmış polipler,

1 – Boyanmış-rezeksiyon kenarları,

2 – Özofajit,

3 – Normal çekum,

4 – Normal pilor,

5 – Polipler,

6 – Ülseratif kolit,

7 – Z-çizgisi

Bu bütüncül değerlendirme yaklaşımı, her modelin hem nicel hem de nitel yönlerden detaylı biçimde incelenmesini ve karşılaştırmalı olarak yorumlanmasını sağlamıştır.

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Sınıflandırma Modellerinin Deneysel Performans Analizi

Bu çalışmada kullanılan Vision Transformer (ViT), Swin Transformer, ConvNeXt, TransUNet ve ConViT modelleri için en uygun hiperparametre kombinasyonlarının belirlenmesinde Grid Search yöntemi kullanılmıştır. Bu yöntem sayesinde her model, farklı öğrenme oranları, batch boyut değerleri, epok sayıları ve optimizasyon algoritmaları türleriyle sistematik olarak eğitilmiş; böylece her model için en yüksek doğrulama başarımını sağlayan yapılandırmalar titizlikle belirlenmiştir.

Grid Search kapsamında değerlendirilen hiperparametre aralıkları şu şekildedir:

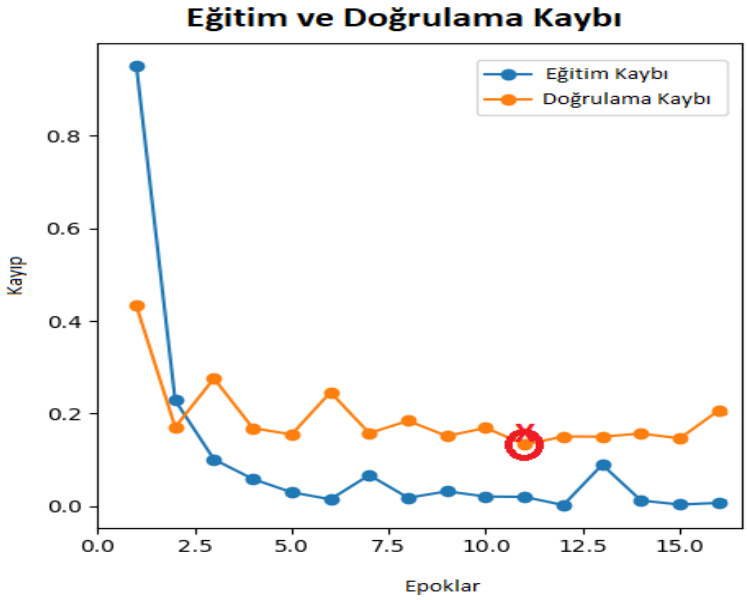
- Öğrenme Oranı: {1e-3, 5e-4, 1e-4, 5e-5, 1e-5}
- Batch Boyutu: {16, 32, 64}
- Optimizasyon algoritması: {Adam, Sgdm, RMSprop}

Bu kapsamda, her bir model için $5 \times 3 \times 3 = 45$ farklı hiperparametre kombinasyonu test edilmiştir. Her kombinasyon, belirli bir öğrenme oranı, batch boyutu ve optimizasyon algoritmasından oluşmaktadır. Bu 45 farklı yapılandırmanın her biri için model eğitimi, başlangıçta maksimum 50 epok üzerinden başlatılmış; öte yandan aşırı öğrenmenin önlenmesi ve eğitim süresinin verimli kullanımı amacıyla her bir kombinasyona erken durdurma (early stopping) stratejisi uygulanmıştır. Bu stratejiye göre, doğrulama kaybı (validation loss) art arda 5 epok boyunca iyileşme göstermediğinde eğitim süreci otomatik olarak durdurulmuştur.

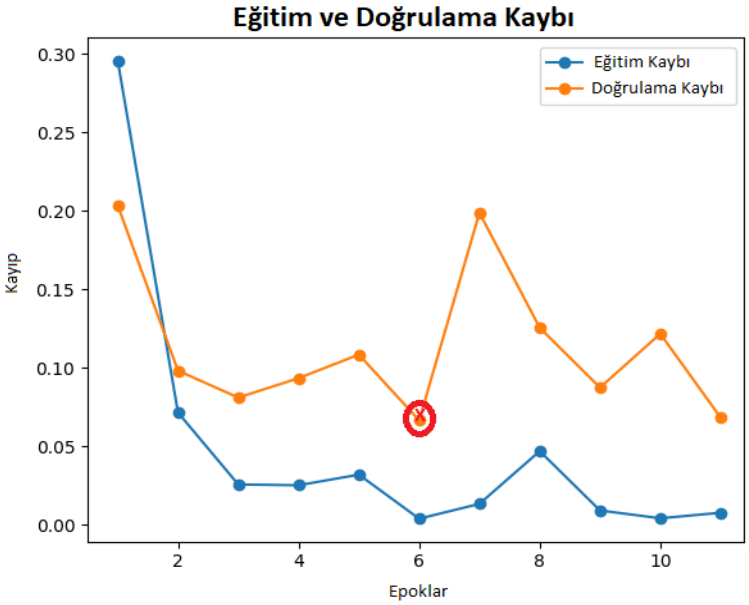
Bu yaklaşım sayesinde, epok sayısı sabit bir hiperparametre olarak değil, modelin yakınsama hızına bağlı olarak dinamik bir şekilde belirlenmiştir. Dolayısıyla her hiperparametre kombinasyonu için eğitim süresi, modelin performansına göre otomatik olarak şekillenmiş ve modeller manuel müdahaleye gerek kalmaksızın kendi optimum epok sayılarında eğitimi tamamlamıştır.

Eğitim sürecinde elde edilen eğitim ve doğrulama kayıpları detaylı şekilde izlenmiş, değerlendirmeler Grid Search süreci sonunda doğrulama başarımı en yüksek olan hiperparametre kombinasyonlarıyla yapılan eğitimler üzerinden yapılmıştır. Bu kombinasyonlar sadece doğruluk açısından değil, eğitim süresinin etkinliği bakımından da en uygun sonuçları sağlamıştır.

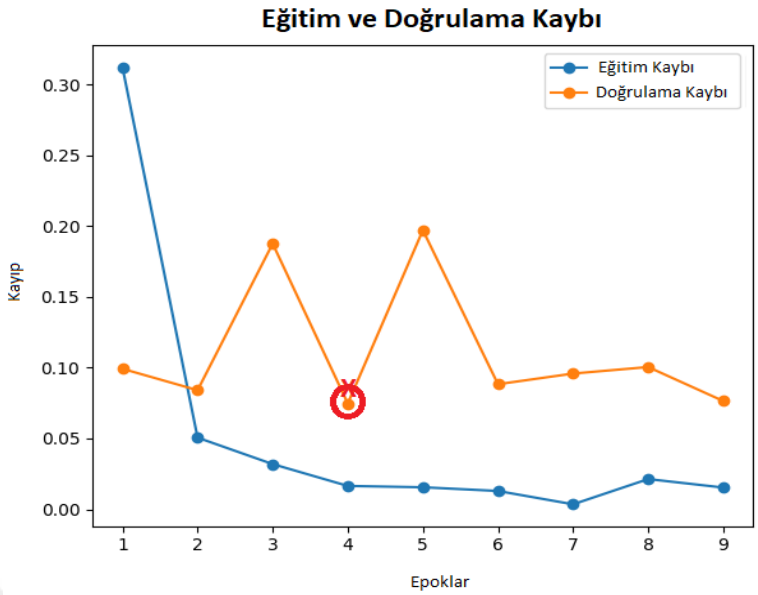
Bu bağlamda, Şekil 4.1, Şekil 4.2, Şekil 4.3 ve Şekil 4.4, sırasıyla ViT, Swin Transformer, ConvNeXt ve ConViT modellerinin Çizelge 4.1'de belirtilen optimum hiperparametrelerle eğitildiği durumları göstermektedir. Her bir grafik, modelin doğrulama başarımının en yüksek olduğu yapılandırmaya ait eğitim sürecini yansıtmakta; aynı zamanda erken durdurmanın hangi epokta devreye girdiğini de açıkça ortaya koymaktadır. Bu grafikler sayesinde, elde edilen başarıların rastlantısal değil, 45 farklı yapılandırma üzerinde uygulanan sistematik bir arama ve değerlendirme süreci sonucunda ortaya çıktığı somut biçimde gösterilmiştir.



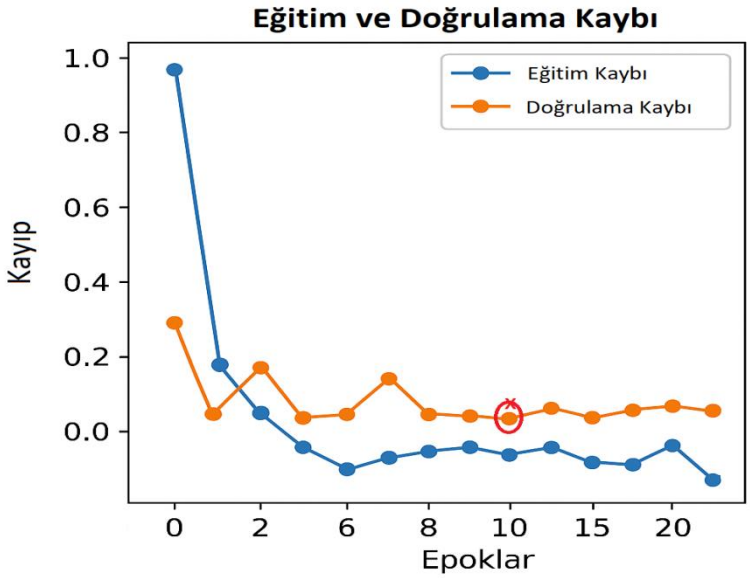
Şekil 4.1. ViT modeline ait eğitim ve doğrulama kaybı grafiği



Şekil 4.2. Swin modeline ait eğitim ve doğrulama kaybı grafiği



Şekil 4.3. ConvNeXt modeline ait eğitim ve doğrulama kaybı grafiği



Şekil 4.4. ConVİT modeline ait eğitim ve doğrulama kaybı grafiği

Bu kapsamda, modellerin eğitiminin sonlandığı optimum epok sayıları, ilgili en iyi kombinasyonlar için şu şekilde gözlemlenmiştir:

- ViT: 11 epok
- Swin Transformer: 6 epok
- ConvNeXt: 4 epok
- ConViT: 11 epok

Grid Search süreci sonunda en iyi hiperparametre kombinasyonları Çizelge 4.1'de özetlenmiştir:

Çizelge 4.1. Sınıflandırma modelleri için optimum hiperparametre ayarları

Model	Öğrenme Oranı	Batch Boyutu	Epok Sayısı	Optimizasyon Algoritması
VİT	1e-4	32	11	RMSprop
Swin	5e-5	16	6	Adam
ConvNeXt	1e-4	16	4	Adam
ConVİT	1e-4	16	11	Adam

Bu parametreler kullanılarak gerçekleştirilen eğitimlerde modellerin başarımı, hem Rastgele Ayrıştırma hem de 5-Kat Çapraz Doğrulama ile ölçülmüş ve ayrıntılı şekilde analiz edilmiştir.

4.1.1. Vision Transformer (ViT) Modeli ile Elde Edilen Sonuçlar

ViT modeli, sınıflandırma görevinde hem rastgele veri ayrıştırma hem de 5 kat çapraz doğrulama stratejileri ile değerlendirilmiştir. Model, daha önce belirlenen optimum hiperparametrelerle eğitilmiş ve her iki strateji altında elde edilen performans metrikleri Çizelge 4.2'de sunulmuştur.

Çizelge 4.2. ViT modeli için performans sonuçları

Değerlendirme Yöntemi	**Doğruluk(%)**	**Kesinlik(%)**	**Duyarlılık(%)**	**F1-Skoru (%)**
Rastgele Ayrıştırma	97,63	97,64	97,60	97,59
5 Katlı Çapraz Doğrulama	97,70	97,63	97,75	97,75

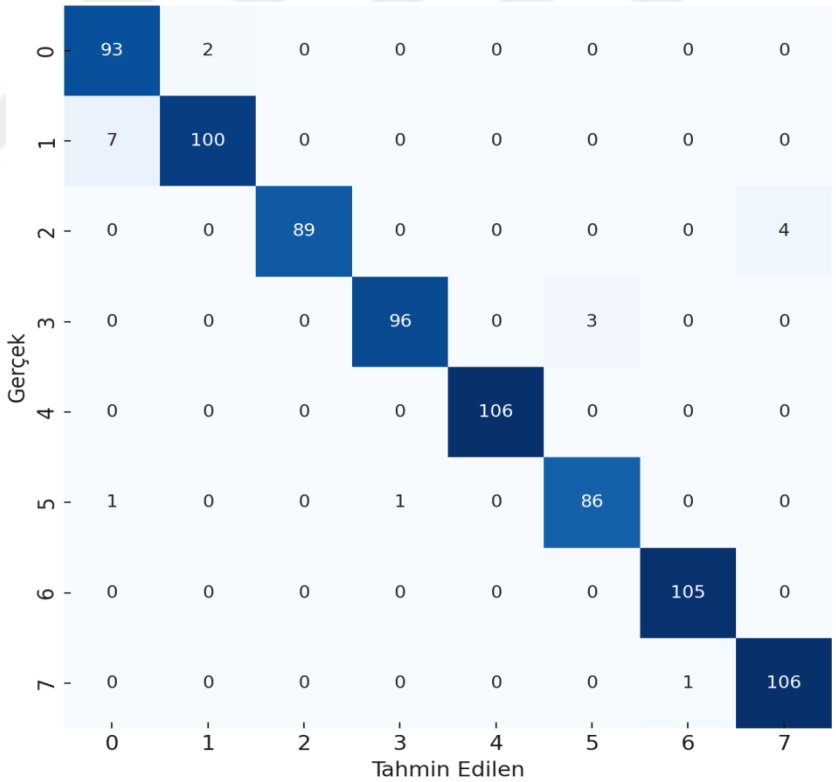
Çizelge 4.2'de sunulan sonuçlara göre, Vision Transformer (ViT) modeli her iki değerlendirme stratejisi altında da yüksek sınıflandırma performansı sergilemiştir. Rastgele ayrıştırma yöntemi ile elde edilen doğruluk (%97,63), kesinlik (%97,64), duyarlılık (%97,60) ve F1-skoru (%97,59), modelin test setinde genel olarak istikrarlı ve başarılı tahminler yaptığını göstermektedir. Bu sonuçlar, modelin sınıflar arasında dengeli bir başarı sergilediğini ve yanlış sınıflandırma oranının oldukça düşük olduğunu ortaya koymaktadır.

Öte yandan, 5 kat çapraz doğrulama ile elde edilen tüm metrikler (yaklaşık olarak %98,00) modelin farklı veri alt kümelerinde de tutarlı sonuçlar verdiğini

göstermektedir. Bu durum, ViT modelinin genellenebilirliğinin güçlü olduğunu ve öğrenilen temsillerin farklı örnek gruplarına başarıyla aktarılabildiğini göstermektedir.

Sonuçlar ayrıca, ViT mimarisinin öz-dikkat mekanizması sayesinde uzun menzilli bağıntıları başarılı şekilde modelleyebildiğini ve bu yeteneğin, özellikle çok sınıflı tıbbi görüntü sınıflandırma görevlerinde önemli bir avantaj sağladığını ortaya koymaktadır.

Şekil 4.5'deki konfüzyon matrisi, Vision Transformer (ViT) modelinin rastgele veri ayrıştırma stratejisiyle eğitildikten sonra test verisi üzerindeki sınıflandırma performansını özetlemektedir. Matrisin satırları gerçek sınıfları, sütunları ise model tarafından tahmin edilen sınıfları temsil etmektedir. Sayısal sınıf etiketleri 0–7 arasında numaralandırılmıştır ve her bir hücrede modelin ilgili sınıf için kaç örneği doğru veya yanlış tahmin ettiğine dair bilgiler yer almaktadır. Bu yapı, modelin hangi sınıflarda yüksek doğrulukla tahmin yaptığını ve hangi sınıflarda karışıklık yaşadığını görsel olarak sunmaktadır.



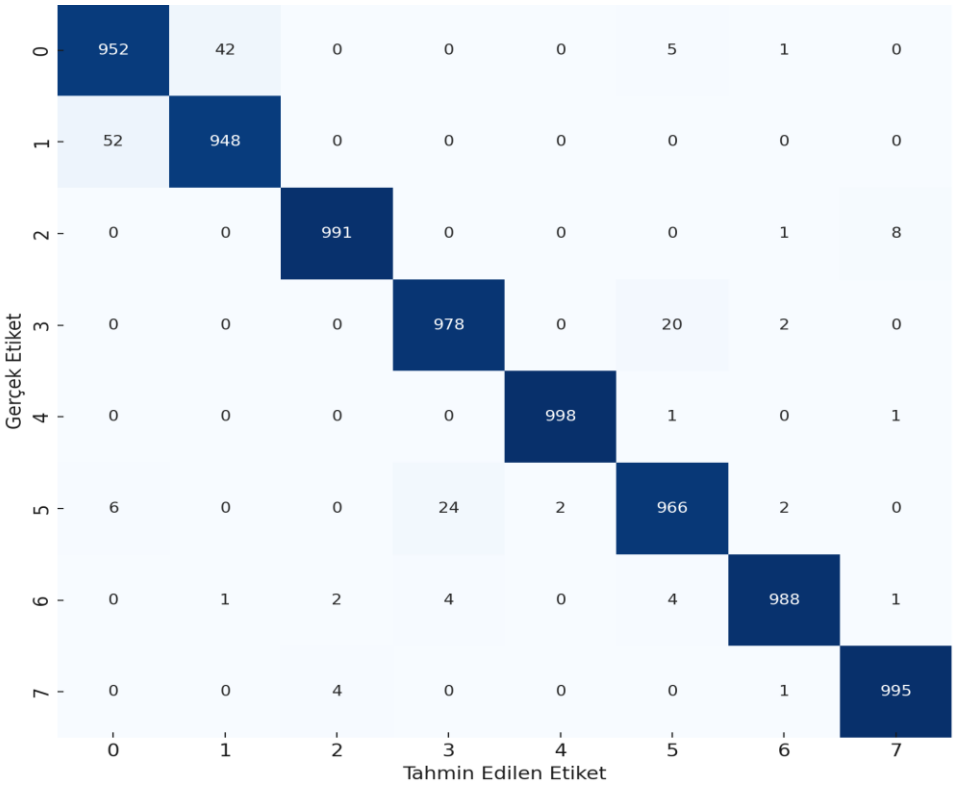
Şekil 4.5. ViT modeli için rastgele ayrıştırma yöntemine göre gösterilen konfüzyon matrisi

Şekil 4.5'de görüldüğü üzere, diyagonal üzerindeki yüksek değerler, modelin büyük çoğunlukla doğru sınıflamalar yaptığını ortaya koymakta ve yüksek genel

doğruluğu desteklemektedir. Özellikle “Normal Pilor” (etiket 4), “Ülseratif Kolit” (etiket 6) ve “Z-line” (etiket 7) sınıflarına ait örneklerin tamamına yakını doğru şekilde sınıflandırılmıştır. Bu durum, modelin bu sınıfları yüksek ayırt edicilikle tanıyabildiğini göstermektedir.

Bununla birlikte, sınıflar arası görsel benzerlik nedeniyle bazı sınıflarda sınırlı çapta karışıklıklar gözlemlenmiştir. Örneğin, “Boyanmış Kaldırılmış Polipler” (etiket 0) ile “Boyanmış Rezeksiyon Kenarları” (etiket 1) sınıfları arasında belirgin bir karışıklık gözlemlenmiştir. Bu iki sınıfa ait görüntüler, benzer renk ve doku özelliklerine sahip oldukları için model tarafından sıklıkla birbirine karıştırılmıştır. Etiket 0’daki bazı örnekler etiket 1 olarak tahmin edilmiş; benzer şekilde, etiket 1’e ait örneklerin bir kısmı da etiket 0 olarak sınıflandırılmıştır. Bu durum, özellikle bu iki sınıf arasındaki görsel benzerliğin modelin ayırt etme yetisini zorladığını göstermektedir. Ayrıca “Özofajit” (etiket 2) sınıfına ait 4 görüntü, “Z-line” (etiket 7) sınıfına yanlış sınıflandırılmıştır. Bu tür hatalar, sınıflar arasındaki görsel örtüşmelerin modelin ayrım gücünü zorladığını göstermektedir. Genel olarak ise ViT modeli, sekiz sınıfın tamamında güçlü ve dengeli bir sınıflandırma performansı sergilemiştir.

Şekil 4.6’da da bu modelin 5-kat çapraz doğrulma ile değerlendirmesine ait konfüzyon matrisini göstermektedir.



Şekil 4.6. ViT modeli için 5 kat çapraz doğrulama yöntemine göre konfüzyon matrisi

Şekil 4.6'da sunulan ViT modeline ait konfüzyon matrisi, 5 katlı çapraz doğrulama yöntemiyle elde edilen sınıflandırma performansını özetlemektedir. Matris genel olarak yüksek doğru sınıflandırma oranlarını yansıtırken, özellikle "0" (Boyanmış Kaldırılmış Polipler) ve "1" (Boyanmış Rezeksiyon Kenarları) etiketli sınıflar arasında gözlenen karşılıklı karışıklık dikkat çekmektedir. Bu iki sınıf, sırasıyla 42 ve 52 örnekle birbirine karıştırılmış olup, görsel benzerlik veya sınıflar arası sınırların belirsizliği bu hatalara neden olabilir. Diğer sınıflarda ise hatalı tahmin sayısı oldukça düşüktür ve diyagonal üzerindeki değerlerin baskın olması, modelin çoğu sınıf için yüksek doğrulukla tahmin gerçekleştirdiğini göstermektedir. Bu sonuçlar, ViT modelinin genel olarak güçlü bir sınıflandırma başarımı sunduğunu göstermektedir. Buna rağmen, belirli sınıflar arasında ayrımın daha da iyileştirilebileceği açıkça görülmektedir.

4.1.2. Swin Transformer Modeli ile Elde Edilen Sonuçlar

Swin Transformer modelinin sınıflandırma performansı, hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejileri kullanılarak incelenmiş olup, elde edilen değerlendirme metrikleri Çizelge 4.3'te özetlenmiştir.

Çizelge 4.3. Swin transformer modeli için performans sonuçları

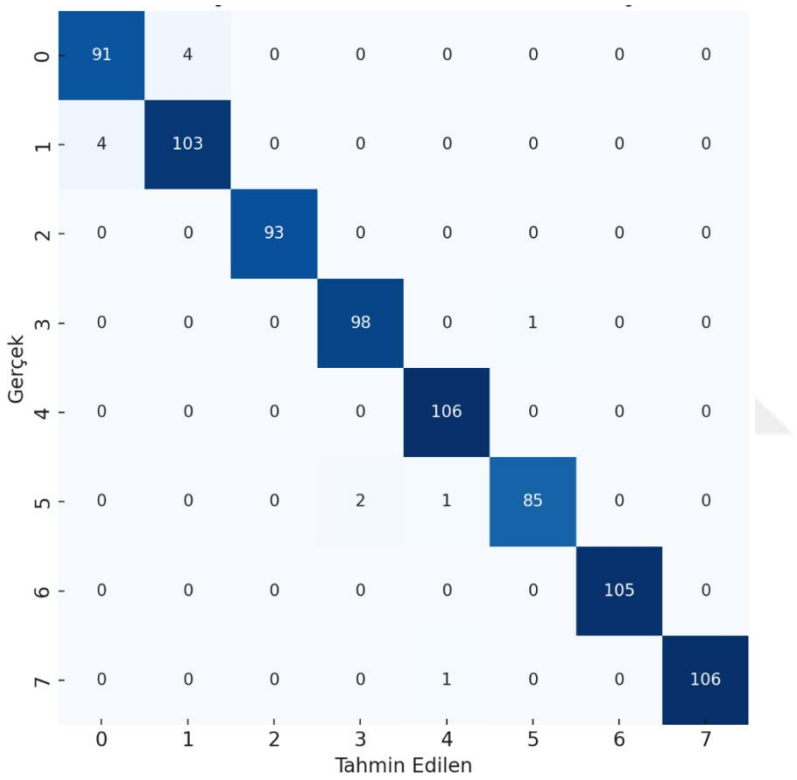
Değerlendirme Yöntemi	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1-Skoru (%)
Rastgele Ayrıştırma	98,38	98,38	98,34	98,36
5 Katlı Çapraz Doğrulama	97,97	97,88	98,00	98,00

Çizelge 4.3'te sunulan Swin Transformer modeline ait sınıflandırma sonuçları, her iki değerlendirme stratejisi altında yüksek doğruluk ve tutarlılık sergilediğini ortaya koymaktadır. Rastgele veri ayrıştırma yönteminde model, %98,38 doğruluk, %98,38 kesinlik, %98,34 duyarlılık ve %98,36 F1-skoru ile oldukça başarılı bir performans göstermiştir. Bu sonuçlar, modelin farklı sınıflar arasında güçlü ayırt edici özellikler öğrenebildiğini ve yüksek doğrulukla sınıflandırma yapabildiğini göstermektedir.

5 katlı çapraz doğrulama yöntemiyle elde edilen sonuçlar ise, modelin genellenebilirliğini değerlendirmek açısından önemlidir. Bu yöntemde modelin %97,97 doğruluk, %97,88 kesinlik, %98,00 duyarlılık ve %98,00 F1-skoru elde ettiği görülmektedir. Elde edilen bu skorlar, Swin Transformer'ın farklı veri bölünmeleri

karşısında da kararlı bir şekilde yüksek performansını koruduğunu ve aşırı öğrenme eğilimi göstermediğini ortaya koymaktadır.

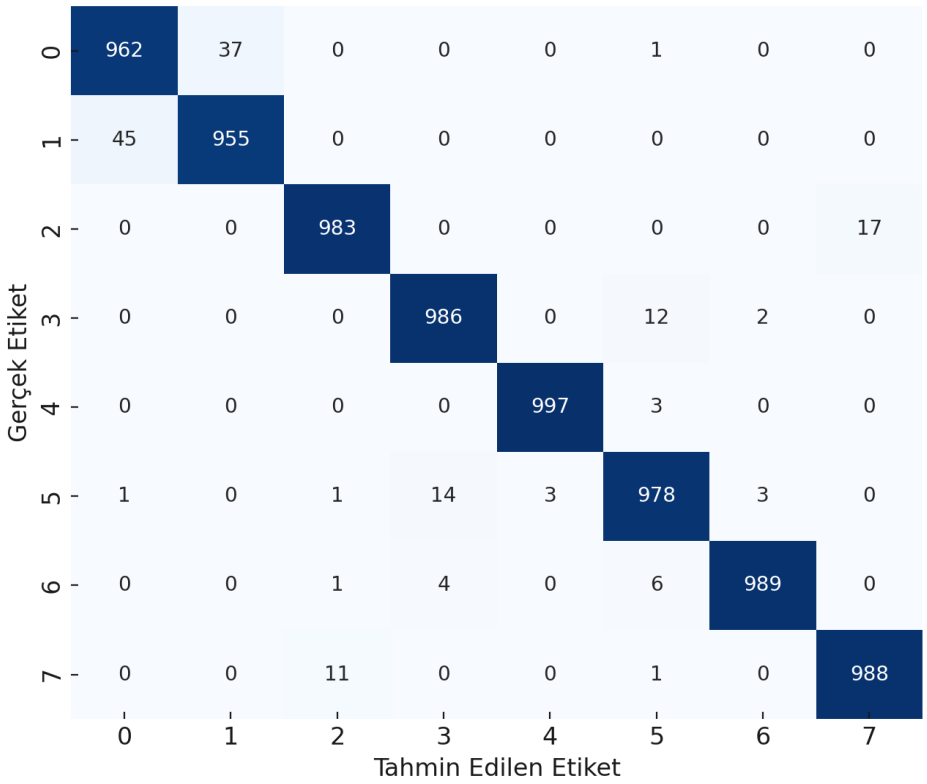
Bu doğrultuda, Swin Transformer modelinin sınıflandırma performansına ilişkin sınıf bazlı hata dağılımlarını görsel olarak sunmak amacıyla, her iki eğitim stratejisine ait konfüzyon matrisleri de ayrı ayrı verilmiştir. Rastgele veri ayrıştırma yöntemiyle elde edilen konfüzyon matrisi Şekil 4.7'de**,** 5 katlı çapraz doğrulama yöntemiyle elde edilen konfüzyon matrisi ise Şekil 4.8'de sunulmaktadır. Bu matrisler, modelin hangi sınıflarda başarılı olduğunu ve olası karışıklıkların hangi sınıflar arasında yoğunlaştığını ayrıntılı şekilde analiz edebilme imkânı sağlamaktadır.



Şekil 4.7. Swin Transformer modeli için rastgele ayrıştırma yöntemine göre konfüzyon matrisi

Şekil 4.7'de yer alan Swin Transformer modelinin rastgele ayrıştırma yöntemine göre değerlendirilmesi sonucunda elde edilen konfüzyon matrisi, genel anlamda modelin güçlü bir sınıflandırma performansı sergilediğini göstermektedir. Doğru sınıflandırma oranları tüm sınıflarda oldukça yüksek olmakla birlikte, sınıf bazlı bazı karışıklıklar da gözlemlenmiştir. Özellikle "Boyanmış Kaldırılmış Polipler" (etiket 0) ve "Boyanmış Rezeksiyon Kenarları" (etiket 1) sınıfları arasında birkaç örnekte yanlış sınıflandırma meydana gelmiş; bu durum bu iki sınıfın görsel benzerliklerinin model

üzerinde ayrıştırma zorluğu oluşturduğunu düşündürmektedir. Ayrıca “Polipler” (etiket 5) sınıfında, sınıfın kendi toplam örnek sayısına oranla nispeten daha yüksek yanlış sınıflandırma oranı gözlenmiş ve bu sınıf diğer sınıflarla görsel örtüşme nedeniyle karıştırılmıştır. Genel olarak ise model, sekiz sınıf arasında dengeli bir performans göstermiştir.



Şekil 4.8. Swin Transformer modeli için 5 kat çapraz doğrulama yöntemine göre konfüzyon matrisi

Şekil 4.8’de görüldüğü üzere, Swin Transformer modelinin 5 kat çapraz doğrulama yöntemiyle elde edilen konfüzyon matrisi, modelin genel doğruluk oranının oldukça yüksek olduğunu göstermektedir. Bununla birlikte, etiket 0 (Boyanmış Kaldırılmış Polipler) ve etiket 1 (Boyanmış Rezeksiyon Kenarları) sınıfları arasında karşılıklı karışıklıklar dikkat çekmektedir. Bu durum, her iki sınıfın görsel benzerliğinin modelin ayırt ediciliğini zorladığını düşündürmektedir. Ayrıca, etiket 2 (Özofajit) ile etiket 7 (Z-line) sınıfları arasında gözlenen sınırlı düzeydeki karışıklıklar da, modelin bu alt sınıflara ait ayırt edici özellikleri tam olarak öğrenememiş olabileceğini göstermektedir. Bununla birlikte, genel olarak yanlış sınıflandırma oranlarının düşük seyretmesi, modelin sınıflar arasında yüksek düzeyde ayrım yapabilme ve güçlü bir genelleme yetisine sahip olduğunu ortaya koymaktadır.

4.1.3. ConvNeXt Modeli ile Elde Edilen Sonuçlar

ConvNeXt modeli, sınıflandırma görevinde modern CNN mimarisiyle değerlendirilmiş; hem rastgele veri ayrıştırma hem de 5 kat çapraz doğrulama stratejileriyle test edilmiştir. Bu iki yöntem kapsamında elde edilen performans metrikleri, Çizelge 4.4'te özetlenmiştir.

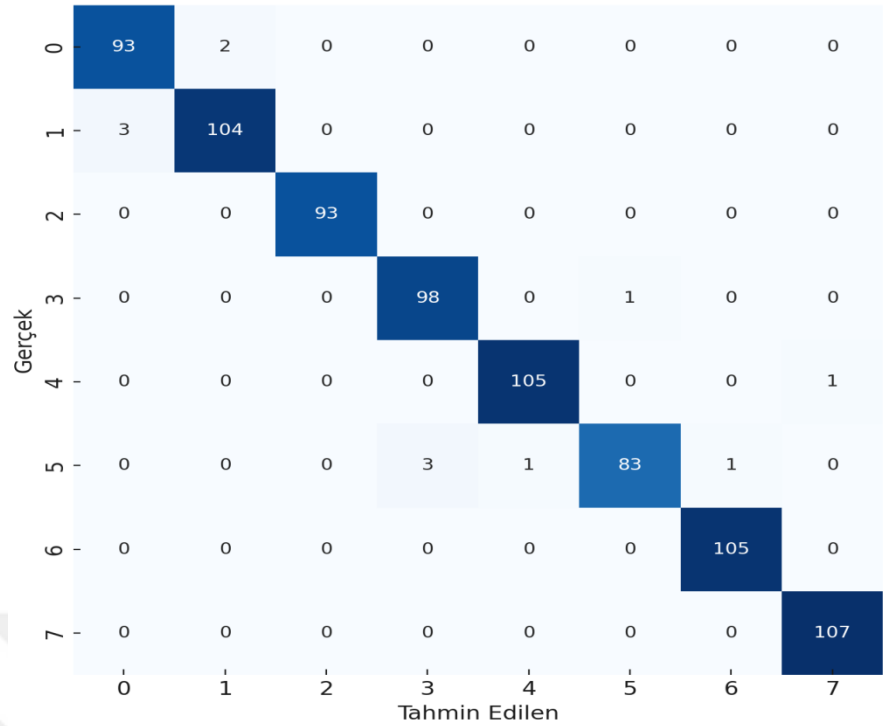
Çizelge 4.4. ConvNeXt modeli için performans sonuçları

Değerlendirme Yöntemi	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1-Skoru (%)
Rastgele Ayrıştırma	98,50	98,50	98,43	98,46
5 Katlı Çapraz Doğrulama	98,13	98,25	98,13	98,25

Çizelge 4.4'te sunulan sonuçlara göre ConvNeXt modeli, hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejilerinde oldukça yüksek sınıflandırma başarımı sergilemiştir. Rastgele ayrıştırma yöntemiyle elde edilen %98,50 doğruluk oranı, modelin eğitim ve test veri setleri arasında güçlü bir genelleme yeteneği geliştirdiğini göstermektedir. Bu doğruluk değeriyle birlikte %98,46 F1 skoru, modelin hem hassasiyet (duyarlılık) hem de kesinlik açısından dengeli performans sergilediğini ortaya koymaktadır.

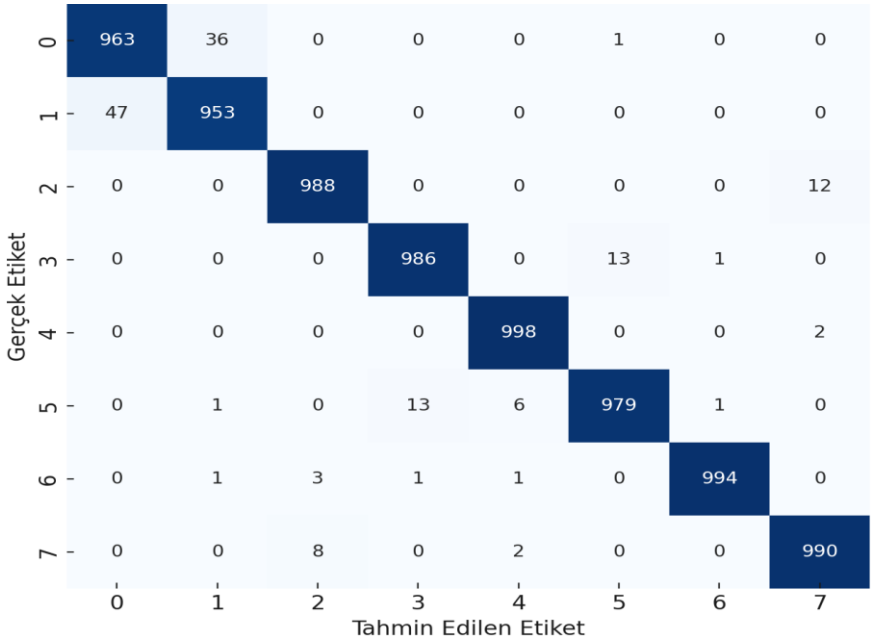
5 kat çapraz doğrulama sonuçları da benzer şekilde yüksek düzeydedir; %98,13 doğruluk ve %98,25 F1 skoru ile modelin farklı veri alt kümeleri üzerindeki kararlılığı teyit edilmiştir. Özellikle F1 skorundaki tutarlılık, ConvNeXt'in sınıflar arası dengesizlikten etkilenmeden doğru sınıflandırma yapabildiğini göstermektedir. Bu bulgular, ConvNeXt modelinin modern CNN mimarisine dayalı yapısının, endoskopik görüntü sınıflandırmasında güçlü ve istikrarlı bir alternatif sunduğunu göstermektedir.

ConvNeXt modeline ilişkin sınıflandırma başarımının görsel olarak daha ayrıntılı analiz edilebilmesi amacıyla, hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejilerine ait konfüzyon matrisleri sırasıyla Şekil 4.9 ve Şekil 4.10'da sunulmuştur. Bu görseller, modelin her sınıf üzerindeki doğru ve hatalı tahminlerini detaylı biçimde göstermekte ve sınıflar arası ayrım yeteneğinin değerlendirilmesine olanak tanımaktadır.



Şekil 4.9. ConvNeXt modeli için rastgele ayrıştırma yöntemine göre konfüzyon matrisi

Şekil 4.9'da sunulan ConvNeXt modeline ait konfüzyon matrisi, modelin çoğu sınıfı yüksek doğrulukla ayırt ettiğini göstermektedir. Özellikle 2 (Özofajit), etiket 6 (Ülseratif Kolit) ve etiket 7 (Z-line) sınıflarında tam doğruluk sağlanırken, en fazla hata etiket 5 (Polipler) sınıfında gözlenmiştir. Ayrıca etiket 0 ve 1 arasında sınırlı düzeyde karışıklıklar bulunmakta olup, modelin genel sınıflandırma başarısı oldukça yüksektir.



Şekil 4.10. ConvNeXt modeli için 5 kat çapraz doğrulama yöntemine göre konfüzyon matrisi

Şekil 4.10'da sunulan konfüzyon matrisi, ConvNeXt modelinin 5 kat çapraz doğrulama yöntemiyle elde edilen sınıflandırma sonuçlarını göstermektedir. Her sınıf için genel olarak yüksek doğru sınıflandırma oranları elde edilmiş olup, özellikle etiket 0 (Boyanmış Kaldırılmış Polipler) ve etiket 1 (Boyanmış Rezeksiyon Kenarları) sınıfları arasında bazı çift yönlü karışıklıklar gözlemlenmiştir. Bununla birlikte, yanlış sınıflandırmaların genel oranı düşüktür ve bu durum, modelin farklı sınıflar arasında yüksek ayırt edicilik yeteneğine sahip olduğunu göstermektedir.

4.1.4. ConVİT Modeli ile Elde Edilen Sonuçlar

ConViT modeli, sınıflandırma görevinde hibrit bir yaklaşım sunarak hem evrişimli sinir ağı yapılarının yerel özellik çıkarım gücünden hem de Transformer mimarisinin küresel bağlam öğrenme kapasitesinden faydalanmaktadır. Bu doğrultuda, modelin performansı hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejileri altında değerlendirilmiş ve elde edilen sonuçlar Çizelge 4.5'te sunulmuştur.

Çizelge 4.5. ConVİT modeli için performans sonuçları

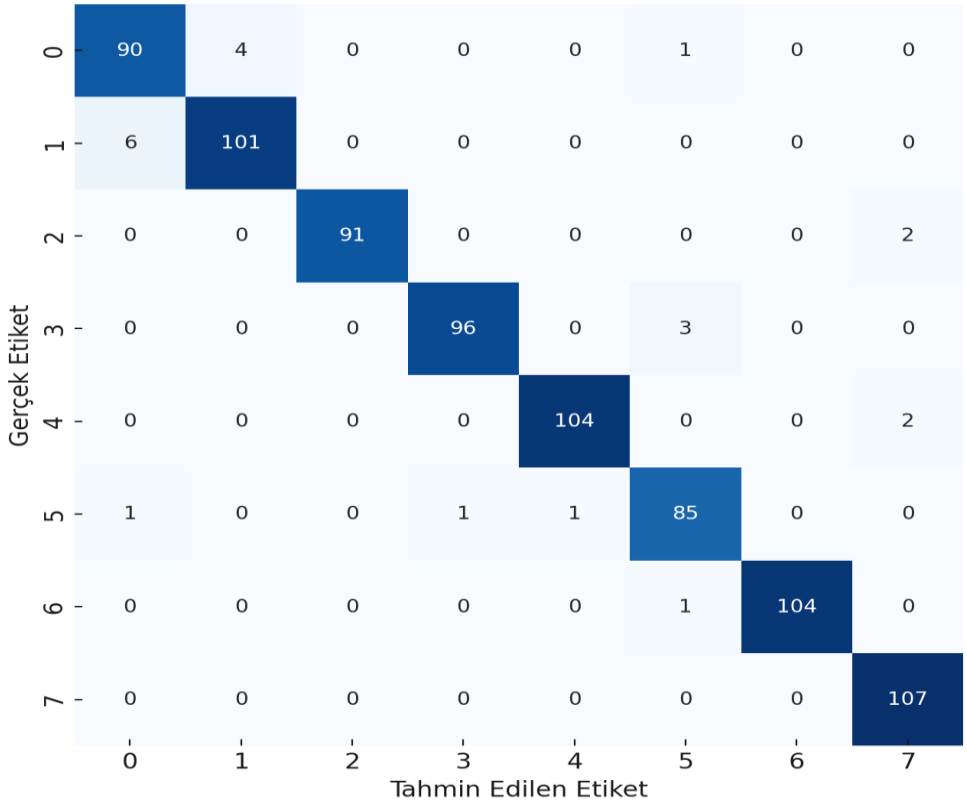
Değerlendirme Yöntemi	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1-Skoru (%)
Rastgele Ayrıştırma	97,25	97,12	97,25	97,37
5 Katlı Çapraz Doğrulama	98,07	98,13	98,00	98,13

Çizelge 4.5'te sunulan sonuçlar, ConViT modelinin hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejileri altında oldukça yüksek sınıflandırma başarımı sergilediğini ortaya koymaktadır. Rastgele ayrıştırma stratejisinde model, %97,25 doğruluk ve %97,37 F1-skoru ile güçlü bir genel performans göstermiştir. Kesinlik (%97,12) ve duyarlılık (%97,25) değerlerinin birbirine yakın olması, modelin sınıflar arasında dengeli tahminler yaptığını ve hem yanlış pozitif hem de yanlış negatif oranlarının düşük olduğunu göstermektedir.

Öte yandan, 5 katlı çapraz doğrulama stratejisi ile elde edilen sonuçlar modelin genellenebilirliğini daha açık şekilde ortaya koymaktadır. Bu yöntemde doğruluk %98,07'ye, F1-skoru ise %98,13'e yükselmiştir. Kesinlik (%98,13) ve duyarlılık (%98,00) oranlarının bu seviyelerde olması, modelin farklı veri alt kümeleri üzerinde tutarlı ve kararlı bir performans gösterdiğini ifade etmektedir.

Bu bulgular, ConViT modelinin hem lokal detayları hem de küresel bağlamı etkili biçimde yakalayarak sınıflandırma görevinde oldukça başarılı olduğunu göstermektedir. Özellikle çapraz doğrulama stratejisinde ulaşılan yüksek metrikler, modelin overfitting'e karşı dirençli olduğunu ve farklı veri senaryolarında güvenilir sonuçlar sunabileceğini desteklemektedir.

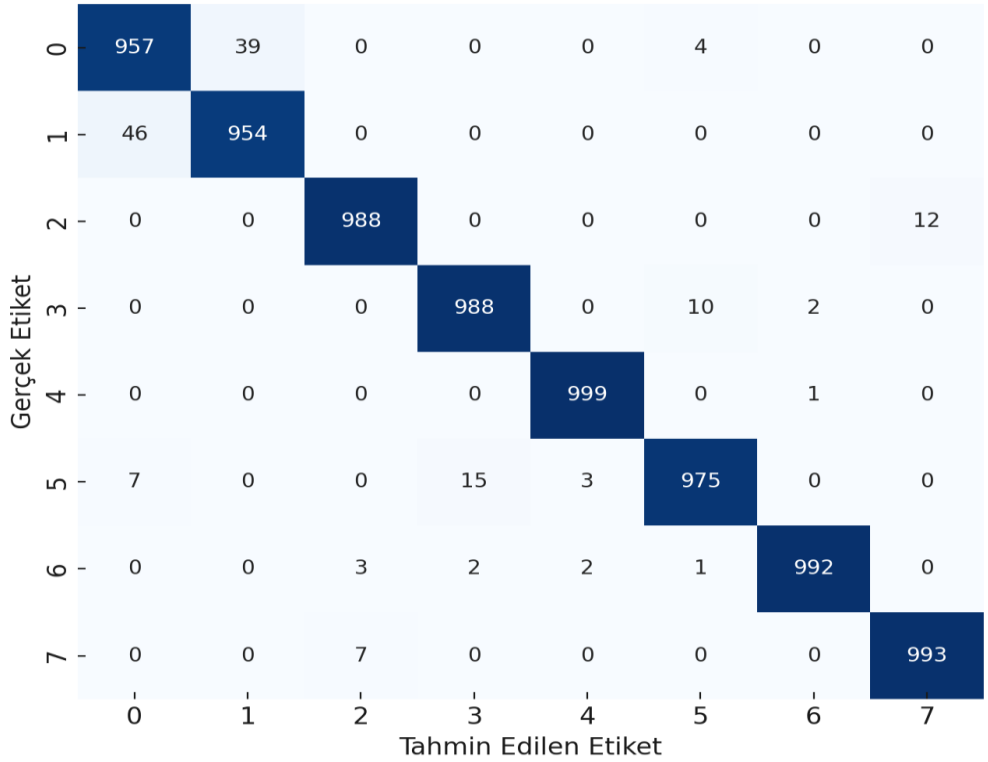
Şekil 4.11 ve Şekil 4.12, sırasıyla ConViT modelinin rastgele veri ayrıştırma ve 5 kat çapraz doğrulama yöntemlerine göre elde edilen konfüzyon matrisi sonuçlarını göstermektedir.



Şekil 4.11. ConVİT modeli için rastgele ayrıştırma yöntemine göre konfüzyon matrisi

Şekil 4.11'de sunulan matrise göre, model genel olarak yüksek doğrulukla sınıflar arasında başarılı ayrımlar yapabilmiştir. En fazla karışıklık, etiket 0 (Boyanmış Kaldırılmış Polipler) ile etiket 1 (Boyanmış Rezeksiyon Kenarları) arasında gözlenmiş olup, bu durum ilgili sınıfların görsel benzerliklerinden kaynaklanıyor olabilir. Ayrıca, etiket 5 (Polipler) sınıfına ait birkaç örneğin diğer sınıflarla karıştığı görülmektedir. Bununla birlikte, çoğu sınıfta yanlış sınıflandırma oranları düşük düzeyde kalmış ve bu

da ConViT modelinin genel olarak başarılı bir genelleme performansı sergilediğini göstermektedir.



Şekil 4.12. ConVİT modeli için 5 kat çapraz doğrulama yöntemine göre konfüzyon matrisi

Şekil 4.12'de gösterilen ConViT modeline ait konfüzyon matrisi, 5 katlı çapraz doğrulama sonucunda elde edilen sınıflandırma performansını ortaya koymaktadır. En belirgin hatalar, etiket 0 (Boyanmış Kaldırılmış Polipler) ile etiket 1 (Boyanmış Rezeksiyon Kenarları) arasında gerçekleşmiş; bu sınıflara ait sırasıyla 39 ve 46 örnek karşılıklı olarak yanlış sınıflandırılmıştır. Ayrıca, etiket 5 (Polipler) sınıfında da dikkat çekici miktarda hata gözlemlenmiş, bu sınıftaki 25 örnek diğer sınıflarla karıştırılmıştır. Etiket 2 (Özofajit) sınıfındaki 12 örneğin etiket 7 (Z-line) ile karışması, bu iki sınıf arasındaki görsel benzerlikten kaynaklanıyor olabilir. Etiket 3 (Normal Sekum) ve etiket 6 (Ülseratif Kolit) için de az sayıda çapraz hata mevcuttur. Bununla birlikte, etiket 4 (Normal Pilor) ve etiket 7 (Z-line) sınıflarında model yüksek doğruluk sergilemiştir. Genel olarak düşük çapraz sınıflandırma oranları, ConViT modelinin sınıflar arası ayrımı büyük ölçüde başarılı şekilde gerçekleştirdiğini göstermektedir.

4.1.5. Sınıflandırma Modellerinin Başarım Analizi ve Kapsamlı Karşılaştırması

Sınıflandırma görevinde kullanılan dört farklı modelin temel Transformer mimarileri olan ViT ve Swin Transformer, modern bir evrimsel CNN yapısı olan ConvNeXt, ve konvolüsyonel öncül bilgiyi Transformer yapısıyla bütünleştiren hibrit ConViT performansları karşılaştırmalı olarak analiz edilmiştir. Değerlendirmeler, rastgele veri ayrıştırma ve 5 katlı çapraz doğrulama stratejileri altında gerçekleştirilmiş, böylece modellerin farklı veri bölme yaklaşımlarına karşı gösterdiği tutarlılık irdelenmiştir.

Karşılaştırma sürecinde doğruluk, kesinlik, duyarlılık ve F1-skoru gibi yaygın olarak kabul gören metrikler esas alınmıştır. Mimari farklılıkların sınıflandırma performansına etkisi, yalnızca genel başarı düzeyleri üzerinden değil, aynı zamanda istikrarlılık ve genelleme kapasitesi açısından da değerlendirilmiştir. Transformer tabanlı modellerin küresel bağlamı yakalama yeteneği, CNN'lerin yerel özellikleri öğrenmedeki başarısı ve hibrit yaklaşımların bu iki yapının avantajlarını birleştirme potansiyeli, elde edilen bulguların yorumlanmasında temel belirleyicilerdir.

Modellere ait karşılaştırmalı performans sonuçları Çizelge 4.6'da sunulmuştur.

Çizelge 4.6. ViT, Swin Transformer, ConvNeXt ve ConViT modellerinin rastgele veri ayrıştırma ve 5 katlı çapraz doğrulama stratejilerine göre sınıflandırma performans karşılaştırması (% değerler)

Model	Değerlendirme Yöntemi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skoru (%)
ViT	Rastgele Ayrıştırma	97,63	97,64	97,60	97,59
	5 Kat Çapraz Doğrulama	97,70	97,63	97,75	97,75
Swin Transformer	Rastgele Ayrıştırma	98,38	98,38	98,34	98,36
	5 Kat Çapraz Doğrulama	97,97	97,88	98,00	98,00
ConvNeXt	**Rastgele Ayrıştırma**	**98,50**	**98,50**	**98,43**	**98,46**
	5 Kat Çapraz Doğrulama	**98,13**	**98,25**	**98,13**	**98,25**
ConViT	Rastgele Ayrıştırma	97,25	97,12	97,25	97,37
	5 Katlı Çapraz Doğrulama	98,07	98,13	98,00	98,13

Çizelge 4.6'da sunulan değerlendirme sonuçlarına göre, dört farklı modelin sınıflandırma performansları hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejileri kapsamında analiz edilmiştir. Genel doğruluk ve F1-skoru

metrikleri açısından değerlendirildiğinde, ConvNeXt modeli her iki stratejide de en yüksek başarıyı sergileyerek %98,50 doğruluk ve %98,46 F1-skoru ile dikkat çekmiştir. Swin Transformer modeli de özellikle rastgele ayrıştırma senaryosunda %98,38 doğruluk ve %98,36 F1-skoru ile ConvNeXt'i yakından takip etmiş, Transformer mimarilerinin klinik görüntü sınıflandırmalarında etkili alternatifler sunabileceğini göstermiştir. ConViT modeli, çapraz doğrulamada %98,13 doğruluk ile ViT'yi geride bırakarak hibrit mimarilerin veri çeşitliliği içeren senaryolarda güçlü bir genel performans sergilediğini ortaya koymuştur.

Model bazlı genel başarıya ek olarak, konfüzyon matrislerinden elde edilen sınıf düzeyindeki analizler, her bir modelin farklı gastrointestinal bulgular üzerindeki ayırt etme kabiliyetine dair önemli bulgular sunmaktadır. Özellikle "Normal Pilor" ve "Normal Sekum" gibi anatomik sınıflar, tüm modeller tarafından yüksek doğrulukla sınıflandırılmış; bu sınıflarda duyarlılık ve kesinlik değerleri %99 düzeyine ulaşmıştır. Öte yandan, "Boyanmış Kaldırılmış Polipler" (etiket 0) ile "Boyanmış Rezeksiyon Kenarları" (etiket 1) sınıfları arasında her modelde dikkat çeken çift yönlü karışıklıklar gözlemlenmiştir. Bu durum, söz konusu iki sınıfın benzer görsel özellikler taşıması nedeniyle modellerin ayrıştırma gücünü sınamakta olup, hem konvolüsyonel hem de Transformer tabanlı yapılar için ortak bir zorluk teşkil etmektedir.

Sınıf bazlı ayırt edicilik açısından, ViT modeli "Özofajit" sınıfında çapraz doğrulama stratejisi kapsamında en yüksek doğruluğa ulaşmıştır. Öte yandan, Swin Transformer ve ConvNeXt modelleri, rastgele veri ayrıştırma stratejisinde aynı sınıfı %100 doğrulukla sınıflandırarak dikkat çekici bir başarı sergilemiştir. Bununla birlikte, bu iki modelin "Z-Line" ve "Polipler" gibi klinik açıdan kritik öneme sahip diğer sınıflarda da yüksek ayırt edicilik sunduğu görülmüştür.

Sonuç olarak, Transformer ve hibrit mimariler, geleneksel CNN tabanlı yöntemlere güçlü alternatifler sunarken; modern CNN yapıları (ConvNeXt), özellikle veri miktarının sınırlı olduğu durumlarda daha stabil ve yüksek performanslı çözümler ortaya koyabilmektedir. Bu bulgular, klinik uygulamalar açısından değerlendirildiğinde, söz konusu modellerin gastrointestinal bulguların doğru sınıflandırılmasında etkili birer karar destek aracı olarak kullanılabileceğini göstermektedir. Sınıf düzeyindeki güçlü ayrımlar, özellikle erken teşhis ve tedavi planlamasında model tabanlı karar sistemlerinin uygulanabilirliğini desteklemektedir.

Elde edilen bu sonuçlar, mevcut literatürde yer alan 8 sınıflı gastrointestinal bulguların sınıflandırıldığı çalışmalarla da karşılaştırılmış ve kullanılan modellerin

gösterdiği üstünlükler değerlendirilmiştir. Bu kapsamda, yapılan literatür karşılaştırması Çizelge 4.7'de sunulmuştur.

Çizelge 4.7. Literatürdeki yöntemlerle sınıflandırma performans karşılaştırması

Çalışma	Kullanılan Model	Veri Seti	Performans
Noha ve diğ. (2019)	DenseNet + Faster R-CNN	Kvasir	%90,2 duyarlılık, %92,1 kesinlik, %75,9 mAP
Khan ve diğ. (2022)	GastroNet	Kvasir v2	%98,02 doğruluk
Wang ve diğ. (2021)	CapsNet	Kvasir v2	%94,83 doğruluk
Cambay ve diğ. (2024)	ResNet-50*	Kvasir v2	%94.08 doğruluk ve %94.08 F1 Skoru
Muhammed ve diğ. (2025)	CNN + Transformer (Hibrit)	Kvasir v2	%95,30 doğruluk
Regmi ve diğ. (2023)	Vision Transformer (ViT)	Kvasir v2	%94,36 F1-skoru, 0,97 ROC-AUC
Tsai ve Lee (2024)	DenseNet201 + VGG19 + InceptionV3 (Ensemble)	Kvasir v2	%91,00 doğruluk
Tsai ve Lee (2025)	Dinamik Ağırlıklı Ensemble	Kvasir v2	%91,00 doğruluk
Oğuz ve Alkan (2024)	Xception + ResNet-101 + NASNet-Large	Kvasir v2	%94,13 doğruluk
Auzine ve diğ. (2024)	InceptionV3 + ResNetV2 + VGG16	Kvasir v2	%93,17 doğruluk, %97,00 F1-skoru
Waheed ve diğ. (2025)	EfficientNetB0 + B2 + ResNet101	Kvasir v2	%97,88 doğruluk
Patil ve Giripunje (2024)	EfficientNetB0 / B1	Kvasir v2	%94,75 / %94,31 doğruluk
Kulkarni ve diğ. (2024)	GIT-NET (DenseNet201 + ResNet50)	Kvasir v2	%95,00 doğruluk
Siddiqui ve diğ. (2024)	CG-Net	Kvasir v2	%95 doğruluk
Zhou ve diğ. (2024)	ConvMixer + Spatial Attention	Kvasir	%93,37 doğruluk
Khan ve diğ. (2024)	Darknet-53 + Xception + BDA + ESKNN	Kvasir v2	%98,25 doğruluk
Sameh Abd El-Ghany ve diğ. (2024)	EfficientNet + ILRC	Kvasir v2	%98,06 doğruluk
Bu çalışma	**ViT**	**Kvasir v2**	**%97,63 doğruluk, %97,59 F1 (Rastgele Veri Ayrışımı)**
Bu çalışma	**Swin Transformer**	**Kvasir v2**	**%98,38 doğruluk, %98,36 F1 (Rastgele Veri Ayrışımı)**
Bu çalışma	**ConvNeXt**	**Kvasir v2**	**%98,50 doğruluk, %98,46 F1 (Rastgele Veri Ayrışımı)**
Bu çalışma	**ConViT**	**Kvasir v2**	**%97,25 doğruluk, %97,37 F1 (Rastgele Veri Ayrışımı)**
Bu çalışma	**ViT**	**Kvasir v2**	**%97,70 doğruluk, %97,75 F1 (5-Kat Çapraz Doğrulama)**
Bu çalışma	**Swin Transformer**	**Kvasir v2**	**%97,97 doğruluk, %98,00 F1 (5-Kat Çapraz Doğrulama)**
Bu çalışma	**ConvNeXt**	**Kvasir v2**	**%98,13 doğruluk, %98,25 F1 (5-Kat Çapraz Doğrulama)**
Bu çalışma	**ConViT**	**Kvasir v2**	**%98,07 doğruluk, %98,13 F1 (5-Kat Çapraz Doğrulama)**

Çizelge 4.7 incelendiğinde, literatürde kullanılan çeşitli derin öğrenme mimarilerinin endoskopik görüntülerden gastrointestinal bulguları sınıflandırma başarımı açısından farklı seviyelerde sonuçlar sunduğu görülmektedir. Bu çalışmaların büyük çoğunluğunun klasik konvolüsyonel sinir ağı (CNN) tabanlı yapılar üzerine kurulu olduğu; özellikle ResNet (Cambay ve diğ., 2024; Kulkarni ve diğ., 2024), DenseNet (Noha ve diğ., 2019; Tsai ve Lee, 2024), InceptionV3 (Tsai ve Lee, 2024; Auzine ve diğ., 2024), EfficientNet (Waheed ve diğ., 2025; Patil ve Giripunje, 2024; El-Ghany ve diğ., 2024), VGG16/VGG19 (Tsai ve Lee, 2024; Auzine ve diğ., 2024), Xception (Oğuz ve Alkan, 2024), NASNet (Oğuz ve Alkan, 2024) gibi mimarilerin yaygın biçimde tercih edildiği dikkat çekmektedir. Bu modeller yüksek doğruluk oranları sunsa da, genellikle sınırlı bağlamsal öğrenme kapasitesi ve yüksek veri çeşitliliğine karşı genelleme güçlüğü gibi dezavantajlara sahiptir.

Bu tez çalışmasında ise, literatürde baskın olan CNN temelli yaklaşımların ötesine geçilerek, Transformer tabanlı ve hibrit mimarilerin sınıflandırma görevlerindeki etkinliği araştırılmış ve bu yaklaşımların sağladığı performans avantajları ortaya konmuştur.

Bu kapsamda uygulanan ConvNeXt modeli, sekiz sınıflı çok sınıflı sınıflandırma görevinde hem Rastgele Veri Ayrışımı hem de 5-Katlı Çapraz Doğrulama stratejilerinde %98,50 ve %98,13 doğruluk değerleri; %98,46 ve %98,25 F1 skoru değerleri ile literatürdeki tüm yöntemlerle doğrudan rekabet edebilecek seviyede bir başarı ortaya koymuştur. ConvNeXt'in bu başarısı, modern konvolüsyonel sinir ağı prensiplerini (örneğin: Layer Normalization, geniş receptive field, GELU aktivasyonu) Transformer mimarilerine benzer şekilde kullanması sayesinde elde edilmiştir. Bununla birlikte, Transformer mimarilerinin sahip olduğu yüksek veri ihtiyacı gibi sınırlamaları yaşamadan daha az kaynakla eğitilebilmiş olması, onu literatürdeki birçok modele göre daha verimli bir seçenek hâline getirmektedir.

Swin Transformer modeli ise özellikle lokal ve global öznitelikleri birlikte öğrenebilme kabiliyeti sayesinde %98,38 ve %97,97 doğruluk; %98,36 ve %98,00 F1 skoru ile ikinci en yüksek performansı sergilemiştir. Bu mimarinin hiyerarşik yapısı ve kayan pencere (shifted window) mekanizması, görüntüdeki küçük ve karmaşık yapıları daha etkili şekilde öğrenmesini sağlamıştır. Bu yönüyle, saliency map ya da dikkat haritalarıyla zenginleştirilmiş modeller (örneğin: Khan ve diğ., 2022 – GastroNet) ile benzer doğruluklara ulaşmakla kalmamış, açıklanabilirlik açısından da avantaj sağlamıştır.

ViT modeli, Transformer yapısının küresel bağlamı doğrudan modelleme kabiliyeti sayesinde %97,63 ve %97,70 doğruluk; %97,59 ve %97,75 F1 skoru ile güçlü bir temel sunmuştur. Buna karşın yüksek veri ihtiyacı ve eğitim sürecinde hiperparametre hassasiyeti nedeniyle bazı durumlarda ConvNeXt kadar stabil sonuçlar vermemiştir. Benzer şekilde, ConViT modeli ise konvolüsyonel ön yargıların Transformer içerisine entegre edilmesiyle daha esnek bir öğrenme süreci sunmuş ve literatürdeki klasik CNN modellerinden daha iyi performans sağlamıştır. %97,25 ve %98,07 doğruluk ile %97,37 ve %98,13 F1 skorları, bu mimarinin pratikteki güçlü performansını desteklemektedir.

Sonuç olarak, bu çalışmada önerilen modeller özellikle veri çeşitliliği yüksek, çok sınıflı problemler için literatürdeki yöntemlere göre daha genellenebilir, dengeli ve yüksek doğruluk sağlayan çözümler üretmiştir. Hem saf Transformer modelleri hem de modern CNN'ler ile yapılan kapsamlı değerlendirme, bu tez çalışmasının literatüre önemli katkılar sunduğunu ortaya koymakta; özellikle Transformer tabanlı ve hibrit yaklaşımların, geleneksel CNN mimarilerine kıyasla gastrointestinal bulguların sınıflandırılmasında daha etkin ve açıklanabilir çözümler sunduğunu güçlü bir şekilde ortaya koymaktadır.

4.2. Segmentasyon Modellerinin Deneysel Performans Analizi

Gastrointestinal endoskopik görüntülerde yer alan poliplerin piksel düzeyinde hassas biçimde bölütlenmesi (segmentasyonu), erken evrede tanı konulabilmesi ve doğru tedavi planlaması açısından önemli bir adımdır. Bu tez çalışmasında, polip segmentasyon görevini gerçekleştirmek amacıyla SegFormer ve Swin Transformer + UPerNet mimarileri değerlendirilmiştir. Model eğitim süreçlerinden önce Grid Search yöntemi uygulanarak her iki model için optimizasyon algoritması ve epok sayısı gibi hiperparametreler sistematik biçimde taranmış ve en uygun kombinasyonlar belirlenmiştir. Grid Search sonucunda elde edilen optimal parametre ayarları ve en yüksek doğrulama Dice Katsayı skorları Çizelge 4.8'de sunulmaktadır. Segmentasyon başarımı yalnızca metriksel olarak değil, aynı zamanda model çıktılarının görsel doğruluğu açısından da değerlendirilmiş ve takip eden alt bölümlerde detaylandırılmıştır.

Çizelge 4.8. Segmentasyon modelleri için optimum hiperparametre ayarları

Model	Epok	Optimizasyon Algoritması	Doğrulama Dice Skorları (%)
SegFormer	13	RMSprop	90,09
Swin Transformer + UPerNet	10	RMSprop	87,75

4.2.1. SegFormer Modeli ile Elde Edilen Sonuçlar

SegFormer modeli, polip segmentasyonu görevinde hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejileri kullanılarak değerlendirilmiştir. Her iki yöntem kapsamında elde edilen performans metrikleri detaylı biçimde hesaplanmış ve modelin genel başarımı karşılaştırmalı olarak incelenmiştir. Elde edilen değerlendirme sonuçları Çizelge 4.9'da özetlenmektedir.

Çizelge 4.9. SegFormer modeli için performans sonuçları

Değerlendirme Yöntemi	**Doğruluk(%)**	**Kesinlik(%)**	**Duyarlılık(%)**	**F1-Skoru (%)**	**Dice**	**IoU**
Rastgele Ayrıştırma	96,20	90,65	84,52	87,47	0.8731	0.7751
5 Katlı Çapraz Doğrulama	88,32	88,95	87,09	87,38	0,8756	0,7807

Çizelge 4.9'a göre, rastgele ayrıştırma stratejisinde model, %96,20 doğruluk, %90,65 kesinlik, %84,52 duyarlılık ve %87,47 F1-skoru ile genel anlamda başarılı bir performans sunmuştur. Bu değerler, modelin hem genel doğruluk hem de dengesiz verilerdeki tahmin kabiliyeti açısından yeterli düzeyde olduğunu göstermektedir. Dice katsayısının 0,8731 ve IoU değerinin 0,7751 olması, modelin piksel bazlı segmentasyon doğruluğu açısından da yeterli olduğuna işaret etmektedir.

Öte yandan 5 katlı çapraz doğrulama stratejisi kullanıldığında, doğruluk değerinin %88,32'ye düştüğü gözlemlenmiştir. Bu, modelin tüm veri üzerindeki genel performansının rastgele ayrıştırmaya göre daha düşük olduğunu ortaya koymaktadır. Yine de %87,38 F1-skoru, 0,8756 Dice Katsayısı ve 0,7807 IoU değerleri, modelin belirli bir tutarlılık gösterdiğini ifade etmektedir. Bu durum, SegFormer'ın kaynak kullanımı açısından verimli ve makul performanslı bir alternatif olabileceğini göstermektedir.

4.2.2. Swin Transformer + UPerNet Modeli ile Elde Edilen Sonuçlar

Swin Transformer + UPerNet modeli, bu tez çalışmasında polip segmentasyonu görevinde değerlendirilen ikinci derin öğrenme tabanlı yaklaşımdır. Bu yapı, Swin Transformer'ın çok katmanlı dikkat mekanizmasını UPerNet'in çok ölçekli özellik birleştirme yeteneğiyle bütünleştiren hibrit bir mimari sunmaktadır. Modelin segmentasyon performansı, hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejileri çerçevesinde analiz edilmiş ve sonuçlar, temel değerlendirme metrikleri üzerinden karşılaştırılmıştır. Elde edilen performans değerleri Çizelge 4.10'da sunulmaktadır.

Çizelge 4.10. Swin transformer + UPerNet modeli için performans sonuçları

Değerlendirme Yöntemi	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1-Skoru (%)	Dice	IoU
Rastgele Ayrıştırma	97,24	89,06	92,01	90,51	0.9025	0.8303
5 Katlı Çapraz Doğrulama	96,11	88,47	87,37	87,77	0,87	0,7905

Çizelge 4.10 göre söylenebilir ki, rastgele ayrıştırma yönteminde model %97,24 doğruluk, %92,01 duyarlılık, %90,51 F1-skoru, 0,9025 Dice Katsayı değeri ve 0,8303 IoU değerleri ile son derece başarılı bir performans sergilemiştir. Bu durum, modelin özellikle medikal görüntüler gibi karmaşık yapılara sahip verilerde bile başarılı çıktılar verebildiğini göstermektedir.

5 katlı çapraz doğrulama senaryosunda da Swin Transformer + UPerNet modelinin doğruluk değeri %96,11 olup, bu değer modelin tüm veri seti üzerinde de tutarlı bir performans gösterdiğini ortaya koymaktadır. %87,77 F1-skoru, 0,87 Dice Katsayı ve 0,7905 IoU gibi değerler, rastgele ayrıştırmaya göre bir miktar düşüş olsa da genel olarak yüksek performansın korunduğuna işaret etmektedir. Bu model, hem bireysel tahmin kabiliyeti hem de genel genellenebilirlik açısından son derece güvenilir bir alternatif olarak öne çıkmaktadır.

4.2.3. Segmentasyon Modellerinin Başarım Analizi ve Kapsamlı Karşılaştırması

Bu tez çalışmasında, SegFormer ve Swin Transformer + UPerNet modellerinin polip segmentasyonundaki başarımı hem sayısal metriklerle hem de görsel çıktılarla kapsamlı şekilde değerlendirilmiştir. Sayısal sonuçlar, doğruluk düzeyini ve maskelerin

ne derece isabetli üretildiğini nesnel biçimde yansıtırken; görsel segmentasyon örnekleri ise modellerin tıbbi açıdan kritik bölgelerdeki sınırları ne kadar hassas tespit edebildiğini ortaya koymaktadır

Karşılaştırmalı sayısal sonuçlar Çizelge 4.11'de sunulmuştur. Bu çizelgede, her iki modelin hem rastgele veri ayrıştırma hem de 5 katlı çapraz doğrulama stratejilerine göre elde ettiği Dice Katsayısı ve IoU (Intersection over Union) değerleri yer almaktadır.

Çizelge 4.11. SegFormer ve Swin Transformer + UPerNet modellerinin rastgele veri ayrıştırma ve 5 katlı çapraz doğrulama stratejilerine göre segmentason performans karşılaştırması

Model	Değerlendirme Yöntemi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skoru (%)	Dice	IoU
SegFormer	Rastgele Ayrıştırma	96,20	90,65	84,52	87,47	0.8731	0.7751
	5 Katlı Çapraz Doğrulama	88,32	88,95	87,09	87,38	0,8756	0,7807
Swin Transformer + UPerNet	**Rastgele Ayrıştırma**	**97,24**	**89,06**	**92,01**	**90,51**	**0.9025**	**0.8303**
	5 Katlı Çapraz Doğrulama	**96,11**	**88,47**	**87,37**	**87,77**	**0,87**	**0,7905**

Çizelge 4.11’deki metrikler üzerinden yapılan analiz, her iki modelin de farklı veri ayrıştırma stratejilerinde ne derece tutarlı ve başarılı olduğunu ortaya koymaktadır.

Rastgele Veri Ayrıştırma Stratejisi senaryoda en yüksek performansı, Swin Transformer + UPerNet modeli sergilemiştir. %97,24 doğruluk ve %90,51 F1-Skoru ile SegFormer modeline (%96,20 doğruluk, %87,47 F1-Skoru) göre daha iyi bir genel performans ortaya koymuştur. Aynı zamanda Dice katsayısı (0,9025) ve IoU (0,8303) değerleri ile Swin + UPerNet modeli segmentasyon kalitesi bakımından daha başarılı görülmektedir. Bu, modelin polip gibi kritik yapıların ayrımında daha isabetli tahminler yaptığını gösterir.

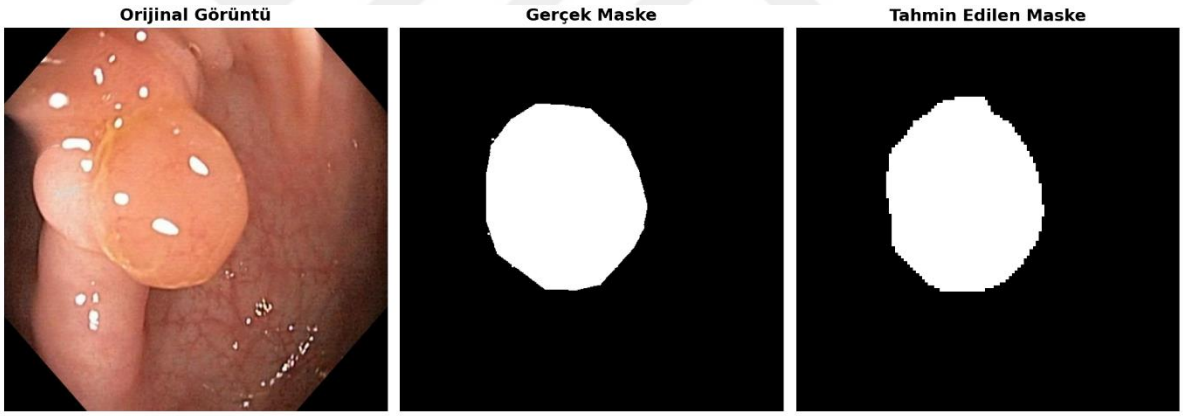
5 Katlı Çapraz Doğrulama Stratejisi daha genellenebilir ve adil bir değerlendirme yöntemi olup modelin tüm veri üzerindeki tutarlı performansını gösterir. Burada Swin + UPerNet modelinin doğruluğu %96,11 iken, SegFormer modelinde %88,32 olarak gözlemlenmektedir. Benzer farklar F1-Skoru, Dice ve IoU gibi metriklerde de belirgindir. Swin modeli %87,77 F1, 0.87 Dice, 0,7905 IoU ile, SegFormer'a (%87,38 F1, 0,8756 Dice, 0,7807 IoU) oranla hem genel başarıda hem de piksel düzeyindeki tutarlılıkta bir adım öndedir.

Her iki değerlendirme yöntemi de dikkate alındığında, Swin Transformer + UPerNet modeli segmentasyon için daha üstün performans göstermiştir. Bu modelin

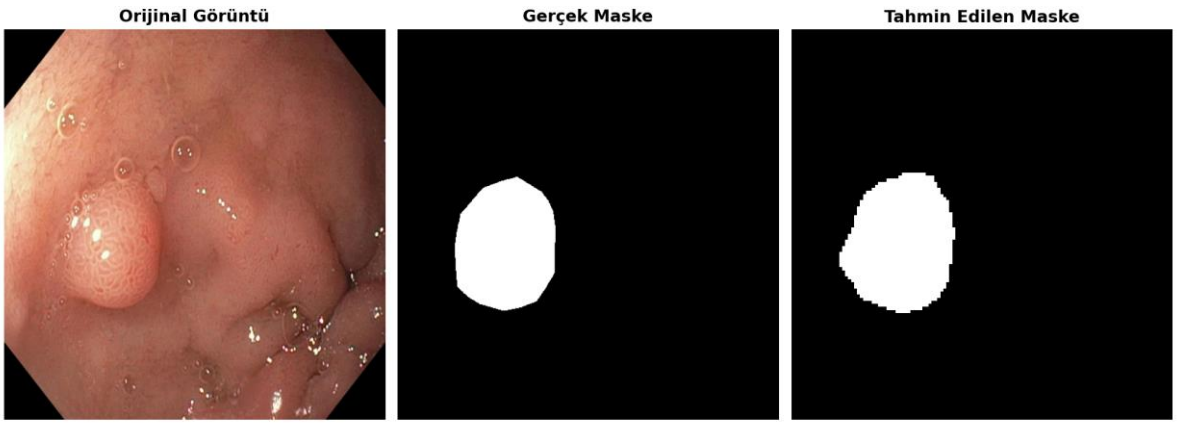
hem duyarlılık (recall) hem de Dice ve IoU değerleri daha yüksek olup, karmaşık medikal görüntülerde bile çıktıların tutarlı ve klinik açıdan anlamlı olduğunu göstermektedir. Bu da Swin + UPerNet modelinin, gerçek dünya uygulamalarında tercih edilmesi için güçlü bir aday olduğunu ortaya koymaktadır.

Buna karşılık SegFormer modeli, daha sade bir yapıya sahip olmasının getirdiği avantajla daha düşük hesaplama maliyetine karşılık makul bir performans sunmuştur. Bu da kaynak sınırlaması olan sistemlerde SegFormer'ı mantıklı bir alternatif haline getirebilir.

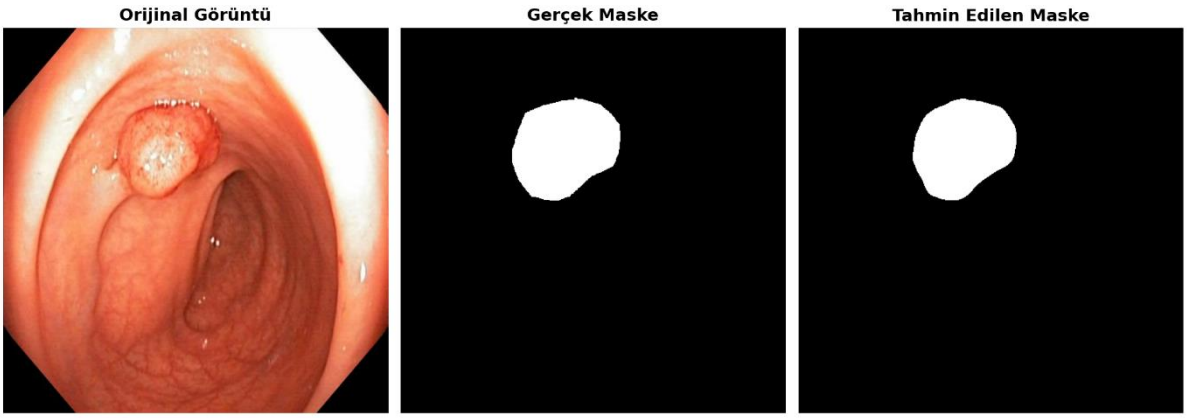
Sayısal analizleri desteklemek amacıyla, Şekil 4.13–4.18 arasında yer alan örnek segmentasyon çıktıları verilmiştir. Bu görsellerde, gerçek maske ile tahmin edilen maskeler karşılaştırmalı olarak sunulmuş ve modellerin segmentasyon doğruluğu görsel olarak da değerlendirilmiştir. Böylece, yalnızca istatistiksel değil, aynı zamanda görsel düzeyde de modellerin ayırt ediciliği ve tıbbi geçerliliği yorumlanabilir hâle getirilmiştir.



Şekil 4.13. Segformer modelinin görsel üzerinde gerçek maskesi ve tahmini

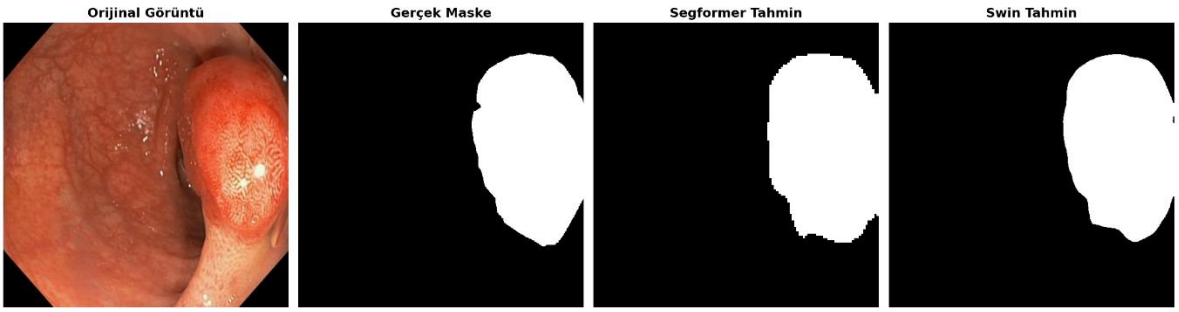


Şekil 4.14. Segformer modelinin görsel üzerinde gerçek maskesi ve hahmini

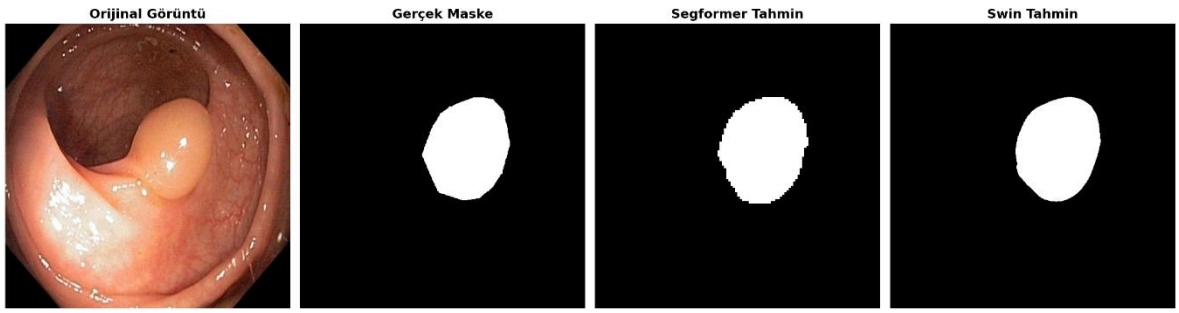


Şekil 4.15. Swin+UperNet modelinin görsel üzerinde gerçek maskesi ve tahmini

Şekil 4.16. Swin+UperNet modelinin görsel üzerinde gerçek maskesi ve tahmini



Şekil 4.17. Swin+UperNet ve segformer modellerinin tek görsel üzerinde gerçek maskesi ve tahmini karşılaştırılması



Şekil 4.18. Swin+UperNet ve segformer modellerinin tek görsel üzerinde gerçek maskesi ve tahmini karşılaştırılması

Bu görseller, segmentasyon modellerinin (SegFormer ve Swin+UPerNet) orijinal endoskopik görüntüler üzerindeki tahminlerini, yer gerçeği (ground truth) maskeleriyle birlikte sunmaktadır. Her bir örnekte, modelin segmentasyon maskesinin yer gerçeğiyle olan benzerliği değerlendirilebilir. Genel olarak her iki model de polip bölgesini başarılı bir şekilde tespit edebilmiş olsa da, Swin+UPerNet modelinin segmentasyon çıktılarının yer gerçeği maskelerine daha yakın olduğu, sınırların daha doğru belirlendiği ve şekil uyumunun daha iyi sağlandığı görülmektedir. Özellikle düzensiz şekilli poliplerde SegFormer modelinde sınırlarda daha belirgin sapmalar gözlemlenirken, Swin+UPerNet modeli daha stabil bir yapı sergilemektedir. Bu görsel çıktılar, nicel metriklerle elde edilen performans farklarını görsel olarak da desteklemektedir.

Elde edilen segmentasyon sonuçları, aynı problem üzerine yapılan literatür çalışmaları ile karşılaştırılmıştır. Bu karşılaştırmada, Dice Katsayısı ve Intersection over Union (IoU) metrikleri temel alınarak, kullanılan modellerin daha önce aynı veri kümesi ve görev üzerinde raporlanan yöntemlerle gösterdiği performans değerlendirilmiştir. Karşılaştırmalı sonuçlar Çizelge 4.12'de ayrıntılı olarak sunulmuştur.

Çizelge 4.12'de verilen segmentasyon performansları, polip gibi düzensiz yapıya sahip tıbbi nesnelerin piksel düzeyinde doğru biçimde ayrıştırılması açısından literatürde önerilmiş mimarilerle karşılaştırmalı olarak sunulmuştur.

PraNet ve MSRF-Net gibi mimariler genellikle klasik U-Net yapılarının üzerine çeşitli çok ölçekli dikkat veya ters kodlayıcı stratejileri ekleyerek Dice ve IoU skorlarını artırmayı hedeflemiştir. Özellikle MSRF-Net %91,5 Dice ve %84,7 IoU ile öne çıkarken, PraNet %89,8 Dice ile detaylı sınır çizimi açısından rekabetçi bir performans sergilemiştir. Ancak bu modeller, Swin gibi Transformer tabanlı mimarilerin global kontekst bilgisiyle sunduğu avantajları sunamamaktadır.

Çizelge 4.12. Literatürdeki yöntemlerle segmentasyon performans karşılaştırması

Çalışma	Model	Veri Kümesi	Dice Skoru	IoU
Fan ve diğ. (2020)	PraNet	Kvasir-SEG	89,8	85,5
Srivastava ve diğ. (2022)	MSRF-Net	Kvasir-SEG	91,5	84,7
Chen ve diğ. (2024)	MixFormer	Kvasir-SEG	91,8	87,9
Tang ve diğ. (2024)	HTC-Net	Kvasir-SEG	92,4	88,7
Srivastava ve diğ. (2021)	DDANet	Kvasir-SEG	90,2	82,3
Wang ve diğ. (2025)	Semi-supervised	Kvasir-SEG	89,5	-
Bu çalışma	**SegFormer**	**Kvasir-SEG**	**87,31**	**77,51**
Bu çalışma	**SegFormer (5-Fold CV)**	**Kvasir-SEG**	**87,56**	**78,07**
Bu çalışma	**Swin + UPerNet**	**Kvasir-SEG**	**90,25**	**83,03**
Bu çalışma	**Swin + UPerNet (5-Fold CV)**	**Kvasir-SEG**	**87**	**79,05**

Bu tez çalışmasında uygulanan Swin Transformer + UPerNet birleşimi, segmentasyon görevinde dikkat mekanizması ile çok ölçekli öznitelik birleştirme yapısının sinerjisinden yararlanmıştır. %90,25 Dice ve %83,03 IoU skorları ile literatürdeki en güçlü yöntemlerle yarışabilecek düzeyde sonuçlar sunmuştur. Özellikle Swin Transformer'ın hiyerarşik yapısı sayesinde hem düşük seviyeli detaylar hem de üst seviyeli semantik bilgiler etkili şekilde yakalanmış, UPerNet'in feature pyramid ağı ile bu bilgiler güçlü çıktılara dönüştürülmüştür. Bu yapı, özellikle poliplerin kenarlarının düzensiz olduğu ve sınıf dengesizliğinin bulunduğu senaryolarda çok daha stabil sonuçlar üretmiştir.

SegFormer modeli, daha hafif bir Transformer tabanlı çözüm sunması nedeniyle segmentasyon hızında avantaj sağlamıştır. %87,31 Dice ve %77,51 IoU gibi oldukça başarılı sayılabilecek skorlar elde etse de, bu model daha basit bir MLP tabanlı decoder kullandığı için bazı karmaşık kenar yapılarında Swin+UPerNet kadar hassas segmentasyon yapamamıştır. Bununla birlikte, SegFormer modelinin ortalama inference süresi 0,0316 saniye olarak ölçülmüş, bu değer Swin+UPerNet modelinin 0,0625

saniyelik süresine kıyasla yaklaşık iki kat daha hızlıdır. Bu fark, SegFormer'ı özellikle gerçek zamanlı sistemlerde kullanım açısından öne çıkarmaktadır ve modelin pratikteki uygulanabilirliğini artırmaktadır.

Literatürdeki MixFormer (%91,8 Dice, %87,9 IoU), HTC-Net (%92,4 Dice, %88,7 IoU), DDANet (%90,2 Dice, %82,3 IoU) gibi modeller her ne kadar yüksek skorlar sunsa da genellikle eğitim maliyeti yüksek, çok fazla parametre içeren ve karmaşık yapıların kullanıldığı modellerdir. Bu bağlamda, tez çalışmasında kullanılan modeller hem sade yapılarıyla hem de klinik tutarlılık sağlayan skorlarıyla dikkat çekmektedir.

Sonuç olarak, bu tez çalışmasında kullanılan Transformer tabanlı ve hibrit modeller, literatürdeki birçok yönteme kıyasla daha dengeli, açıklanabilir ve yüksek doğruluklu segmentasyon sonuçları sunarak klinik karar destek sistemlerine entegrasyon potansiyelini güçlendirmiştir.

5. SONUÇLAR VE ÖNERİLER

5.1. Sonuçlar

Bu çalışmada, gastrointestinal endoskopik görüntüler üzerinde hem sınıflandırma hem de segmentasyon görevlerinde derin öğrenme temelli modern mimarilerin (ViT, Swin Transformer, ConvNeXt, ConViT, SegFormer ve Swin+UPerNet) başarımı değerlendirilmiştir. Sınıflandırma görevinde dört farklı mimari rastgele veri ayrıştırma ve 5 katlı çapraz doğrulama stratejilerine göre test edilmiş; benzer şekilde, segmentasyon görevinde de iki farklı model aynı stratejilere tabi tutularak performans karşılaştırmaları gerçekleştirilmiştir.

Sınıflandırma görevinde, özellikle ConvNeXt modeli %98,50 doğruluk ve %98,46 F1-skoru ile en yüksek başarıyı rastgele veri ayrıştırma stratejisinde göstermiştir. Swin Transformer ise %98,38 doğruluk ile ConvNeXt'i yakından takip etmiştir. Hibrit mimari yapısına sahip ConViT modeli, 5 katlı çapraz doğrulamada %98,07 doğruluk ile genellenebilirlik açısından dikkat çekici sonuçlar sunmuştur

Segmentasyon görevlerinde ise en başarılı performans Swin Transformer + UPerNet modeli ile elde edilmiştir. Bu model, rastgele ayrıştırma stratejisinde %97,24 doğruluk, %90,51 F1-skoru, 0,9025 Dice ve 0,8303 IoU değerleri ile en yüksek metrikleri sağlamıştır. 5 katlı çapraz doğrulama stratejisi uygulandığında, doğruluk %96,11'e, F1-skoru %87,77'ye gerilemiş olsa da performans seviyesi genel olarak yüksek kalmıştır

SegFormer modeline ait segmentasyon çıktıları ise özellikle rastgele ayrıştırma senaryosunda %96,20 doğruluk, %87,47 F1-skoru, 0,8731 Dice ve 0,7751 IoU değerleri ile tatmin edici sonuçlar vermiştir. Ancak, çapraz doğrulama stratejisinde bu modelin doğruluğunun %88,32'ye düşmesi, genel performansının veri çeşitliliği karşısında biraz daha dalgalı olduğunu göstermektedir

Tüm bu sonuçlar; Transformer ve hibrit modellerin, klinik açıdan gastrointestinal bulguların doğru sınıflandırılması ve poliplerin segmentasyonunda yüksek doğruluk ve genellenebilirlik sunduğunu göstermektedir. Sınıflandırmada ConvNeXt modelinin, segmentasyonda ise özellikle Swin Transformer + UPerNet yapısının hem görsel hem de sayısal metrikler açısından öne çıktığı gözlemlenmiştir.

5.2. Öneriler

Bu tez çalışması, modern derin öğrenme mimarilerinin endoskopik görüntülerde polip tespiti ve sınıflandırması üzerine etkili performanslar sergileyebileceğini ortaya koymuştur. Ancak elde edilen bulgulara dayanarak aşağıdaki öneriler sunulmaktadır:

- ***Veri Miktarı ve Çeşitliliği Artırılmalı:*** Özellikle sınıf dengesizlikleri ve bazı benzer anatomik yapılar (örneğin; etiket 0 ve 1) modellerin ayırt etme gücünü zorlamaktadır. Daha dengeli ve genişletilmiş veri setleriyle eğitilen modellerin performansı daha da artırılabilir.
- ***Ensemble Yaklaşımlar Geliştirilmeli:*** Farklı model mimarilerinin çıktılarının birleştirildiği ensemble yapılar, özellikle sınıflandırma doğruluğu ve segmentasyon kalitesinde sinerji yaratabilir.
- ***Klinik Uyumun Güçlendirilmesi:*** Klinik uygulamalar için gerçek zamanlı ve kullanıcı dostu arayüzlerle entegre sistemler geliştirilmeli; tıbbi uzmanların kullanımına sunulabilecek prototipler oluşturulmalıdır.
- ***Hata Analizine Dayalı Yeni Modifikasyonlar:*** Özellikle karışıklık yaşanan sınıflarda (örneğin; "Boyanmış Kaldırılmış Polipler" ve "Boyanmış Rezeksiyon Kenarları") sınıf içi varyasyonları öğrenmeye yönelik veri artırma işlemleri (data augmentation) ve sınıf odaklı kayıp fonksiyonları (örneğin, focal loss) entegre edilebilir.
- ***Zaman Serisi ve Video Tabanlı Genişletme:*** Sabit görüntüler dışında, endoskopik video dizilerini dikkate alan zamansal modeller (örneğin, ConvLSTM veya video transformer yapıları) ile çalışmalar genişletilmelidir.
- ***Klinik Test ve Geri Bildirim Süreçleri:*** Modellerin gerçek klinik ortamlarda test edilmesi, uzman hekimlerden alınacak geri bildirimlerle model performansının optimize edilmesi gerekmektedir.

Bu çalışma, gelecekteki araştırmalar için güçlü bir temel sunmakta olup, ileri düzey yapay zeka uygulamalarıyla klinik karar destek sistemlerinin geliştirilmesine katkı sağlamayı amaçlamaktadır.

KAYNAKLAR

Ali, S., ve diğerleri, 2021, Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy, Medical Image Analysis, 70, 102002.

Auzine, M.M., Khan, M.H.-M., Baichoo, S., Sahib, N.G., Gao, X. ve Bissoonauth-Daiboo, P., 2023, Classification of gastrointestinal cancer through explainable AI and ensemble learning, 2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU), IEEE.

Azad, R., et al., 2024, Advances in medical image analysis with vision transformers: a comprehensive review, Medical Image Analysis, 91, 103000.

Cambay, V.Y., et al., 2024, Automated detection of gastrointestinal diseases using ResNet50-based explainable deep feature engineering model with endoscopy images, Sensors, 24(23), 7710.

Chen, Q., et al. (2024). MixFormer: A Mixed CNN-Transformer for Medical Image Segmentation. ResearchGate.

d'Ascoli, S., et al. (2021). ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases. arXiv:2105.03817.

Dosovitskiy, A., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929.

Eker, B.U. ve Solak, F.Z., 2024, Deep learning approaches for gastrointestinal disease classification, 10th Int. European Conf. on Interdisciplinary Scientific Research, Zurich.

El-Ghany, S.A., et al., 2024, An accurate deep learning-based computer-aided diagnosis system for GI disease detection using capsule endoscopy, Applied Sciences, 14(22), 10243.

Fan, D.-P., Ji, G.-P., Zhou, T., Chen, G., Fu, H., Shen, J., & Shao, L. (2020). PraNet: Parallel Reverse Attention Network for Polyp Segmentation. arXiv:2006.11392.

Ghatwary, N., Ye, X., & Zolgharni, M. (2019). Esophageal abnormality detection using DenseNet based Faster R-CNN with Gabor features.

Gunasekaran, H., et al., 2023, GIT-Net: an ensemble deep learning-based GI tract classification of endoscopic images, Bioengineering, 10(7), 809.

He, K., ve diğerleri, 2023, Transformers in medical image analysis, Intelligent Medicine, 3(1), 59–78.

Keleş, H., 2022, Tipta Yapay Zeka Uygulamaları, Kırıkkale Üniversitesi Tıp Fakültesi Dergisi, 24(3), 604-613.

Khan, M. A., Sahar, N., Khan, W. Z., Alhaisoni, M., Tariq, U., Zayyan, M. H., Kim, Y. J., & Chang, B. (2022)

Khan, S., et al. (2022). Transformers in Vision: A Survey. ACM Computing Surveys, 54(10s).

Khan, Z.F., et al., 2024, Deep CNNs for GI tract syndromes, CMC-Computers, Materials & Continua, 78(1), 1207–1225.

Kulkarni, S., et al., 2024, Hybrid deep learning for GI tract classification, 2nd WCONF, IEEE.

Li, J., et al., 2023, Transforming medical imaging with Transformers? Medical Image Analysis, 85, 102762.

Li, Y., Tang, H., Zhang, Z., Liu, X., Chen, H., & Wang, Y. (2024). HTC-Net: A hybrid CNN–Transformer framework for medical image segmentation.

Liu, Z., et al. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv:2103.14030.

Liu, Z., et al. (2022). A ConvNet for the 2020s. arXiv:2201.03545.

Muhammed, N., et al. (2025). AI-Based Classification of Medical Images using Hybrid CNN-Transformer Model. Scientific Reports.

Oğuz, F.E. ve Alkan, A., 2024, Weighted ensemble for GI classification in colonoscopy images, Displays, 85, 102874.

Patil, R.M. ve Giripunje, S., 2024, GI tract classification in endoscopy images, 2nd IDICAIEI, IEEE.

Pereira, G.A. ve Hussain, M., 2024, Review of Transformer-based models in computer vision, arXiv:2408.15178.

Pokuaa, H.A., et al., 2024, Patch-and-amplify capsule network for GI diseases, Scientific African, 25, e02277.

Regmi, S., et al. (2023). Vision Transformer for Chest X-ray and GI Image Classification. arXiv:2304.11529.

Shamshad, F., Khan, S., Zamir, S. W., Hira, S., Khan, F. S., & Shao, L. (2022). *Transformers in medical imaging: A survey*.

Siddiqui, S., Akram, T., Ashraf, I., Raza, M., Khan, M. A., & Damaševičius, R. (2024). CG-Net: A novel CNN framework for gastrointestinal tract diseases classification. International Journal of Imaging Systems and Technology,

Solak, F.Z., 2023, Automatic diagnosis of GI findings using deep learning, 14th Int. Istanbul Scientific Research Congress.

Srivastava, A., Jha, D., Chanda, S., Pal, U., Johansen, H., Johansen, D., Riegler, M., Ali, S., & Halvorsen, P. (2022). MSRF-Net: A multi-scale residual fusion network for biomedical image segmentation

Thakur, G.K., et al., 2024, Deep learning approaches for medical image diagnosis, Cureus, 16(5).

Tsai, C. M., & Lee, J.-D. (2024, October 29–November 1). Grad-CAM visualization and ensemble learning for improved gastrointestinal disease classification using CNNs. In 2024 IEEE 13th Global Conference on Consumer Electronics (GCCE)

Tsai, C. M., & Lee, J.-D. (2025). Dynamic ensemble learning with gradient-weighted class activation mapping for enhanced gastrointestinal disease classification.

Waheed, Z., Gui, J., Amjad, K., Waheed, I., & Asif, S. (2025). An ensemble approach of deep CNN models with Beta normalization aggregation for gastrointestinal disease detection.

Wang, Z., Xie, Y., Li, H., & Sun, J. (2022). Convolutional-capsule network for gastrointestinal endoscopy image classification.

Wang, J., et al. (2025). Semi-Supervised Medical Image Segmentation via Knowledge Mining. arXiv:2503.06816.

Xiao, T., et al. (2018). Unified Perceptual Parsing for Scene Understanding. arXiv:1807.10221.

Xiao, T., et al. (2021). Early Convolutions Help Transformers See Better. arXiv:2106.14881.

Xie, E., et al. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. arXiv:2105.15203.

Varam, D., Khalil, L., & Shanableh, T. (2024). On-Edge Deployment of Vision Transformers for Medical Diagnostics Using the Kvasir-Capsule Dataset. *Applied Sciences

Yağmur, E., ve Kesen, S.E., 2023, Sustainability in production-transportation scheduling, Int. J. Production Research, 61(3), 774-795.

Yoshioka, K., et al., 2023, Detection of esophagitis using deep learning on Kvasir dataset, arXiv:2301.02390.

Zhou, H., et al. (2024). Spatial-Attention ConvMixer for GI Disease Classification. Springer.