



**TIBBİ GÖRÜNTÜ VERİ SETLERİ İÇİN EVRİŞİMLİ
SİNİR AĞLARI VE GÖRÜNTÜ DÖNÜŞTÜRÜCÜ
MİMARİLERİNE YÖNELİK HİBRİT SİSTEMLERİN
GELİŞTİRİLMESİ**

Hasan UCUZAL

BİYOİSTATİSTİK VE TIP BİLİŞİMİ ANABİLİM DALI

**Tez Danışmanı
Prof. Dr. Cemil ÇOLAK
Doktora Tezi – 2026**

**T.C.
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**TIBBİ GÖRÜNTÜ VERİ SETLERİ İÇİN EVRİŞİMLİ SİNİR AĞLARI VE
GÖRÜNTÜ DÖNÜŞTÜRÜCÜ MİMARİLERİNE YÖNELİK HİBRİT
SİSTEMLERİN GELİŞTİRİLMESİ**

Hasan UCUZAL

Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

Doktora Tezi

**Tez Danışmanı
Prof. Dr. Cemil ÇOLAK**

**Tez Jüri Üyeleri
Prof. Dr. Cemil ÇOLAK
Doç. Dr. Burcu Kayhan TETİK
Doç. Dr. Emek GÜLDOĞAN
Doç. Dr. Mehmet KIVRAK
Doç. Dr. Mehmet Onur KAYA**

**MALATYA
2026**

KABUL VE ONAY



T.C.
İNÖNÜ ÜNİVERSİTESİ
Sağlık Bilimleri Enstitüsü Müdürlüğüne

ETİK BEYANI

İnönü Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazık Kurallarına uygun olarak “Prof. Dr. Cemil ÇOLAK” danışmanlığında hazırlayıp sunduğum “Tıbbi Görüntü Veri Setleri İçin Evrişimli Sinir Ağları Ve Görüntü Dönüştürücü Mimarilerine Yönelik Hibrit Sistemlerin Geliştirilmesi” Doktora tezim için elde ettiğim verileri, bilgileri, belgeleri akademik ve etik kurallar çerçevesinde elde ettiğimi, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tezimde yararlandığım eserlere bilimsel kurallara uygun atıfta bulunarak kaynak gösterdiğimi, tezimin özgün olduğunu, tezimin çalışma ve yazımında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabul ettiğimi beyan ederim.

09/03/2026

Hasan UCUZAL, İmza

İÇİNDEKİLER

ÖZET	ix
ABSTRACT	x
SİMGELER VE KISALTMALAR DİZİNİ.....	x
ŞEKİLLER DİZİNİ	xii
TABLolar DİZİNİ.....	xiii
1. GİRİŞ.....	1
2. GENEL BİLGİLER.....	3
2.1. CNN Mimari Yapısı	4
2.1.1. Evrişimli Katmanlar	4
2.1.2. Havuzlama (Pooling) Katmanları.....	4
2.1.3. Aktivasyon Fonksiyonları	5
2.1.4. Tam Bağlı Katmanlar (Fully Connected Layers)	5
2.2. ViT Mimari Yapısı	5
2.2.1. Görüntüyü Yamalara Ayırma	5
2.2.2. Öğrenilebilir Gömmeler (Embeddings).....	6
2.2.1. Çok Kafalı Dikkat (Mekanizması) (Multi-head Attention).....	6
2.2.2. İleri Beslemeli Sinir Ağları (Feed-Forward Neural Networks).....	6
2.3. Veri ve Eğitim Süreci	6
2.3.1. CNN Veri ve Eğitim Süreci.....	6
2.4. Performans Karşılaştırması	7
2.4.1. CNN Performansı	7
2.4.2. ViT Performansı	7
2.5. Hesaplama Maliyetleri ve Verimlilik	7
2.5.1. CNN Hesaplama Maliyetleri	7
2.5.2. ViT Hesaplama Maliyetleri	8
2.6. Uygulama Alanları	8

2.6.1. CNN Uygulama Alanları	8
2.6.2. ViT Uygulama Alanları	8
2.7. Hibrit Mimarilerin Sınıflandırılması	9
2.7.1. Entegrasyon Stratejisine Göre	9
2.7.2. Mimari Tasarıma Göre	12
2.7.3. Uygulama Alanlarına Göre.....	16
2.7.4. Eğitim Stratejilerine Göre.....	18
2.8. Cnn, Vit ve Hibrit Mimariler.....	21
3. MATERYAL VE METOT	22
3.1. Troid Tümörü Veri Seti	22
3.2. Veri Ön İşleme	22
3.3. Veri Artırma	22
3.4. Veri Bölütleme	23
3.5. Modelleme	23
3.6. Model Mimarileri	24
3.6.1. CNN Tabanlı Mimariler	24
3.6.2. Vision Transformer Tabanlı Mimariler	25
3.6.3. Hibrit (CNN + Transformer) Mimariler	27
3.7. Eğitim Parametreleri.....	30
3.8. DeneySEL Kurulum.....	31
3.8.1. Donanım ve Özellikler	31
3.8.2. Yazılım Kütüphaneleri	32
3.8.3. Eğitim Süresi ve Kaynak Kullanımı.....	33
3.9. Performans Metrikleri	34
3.10. Model Açıklanabilirliği	36
3.11. Yazılım	37
3.11.1. Yazılım Mimarisi.....	37

3.11.2. Yazılım Özellikleri	37
3.11.3. Yazılım Arayüz Tasarımı ve İş Akışı	38
4. BULGULAR	43
4.1. Veri Seti Bulguları.....	43
4.2. Model Çıktılarının Analizi	45
4.3. Karmaşıklık Matrisi Analizi	46
4.4. ROC Eğrisi ve AUC Değeri	47
4.5. Doğruluk Analizi	48
4.6. Tahmin Performansı	49
4.7. Görsel Analiz.....	50
4.8. Isı Haritası (Heatmap) İstatistiksel Analizi	52
5. TARTIŞMA.....	53
6. SONUÇ VE ÖNERİLER	56
6.1. Sonuç	56
6.2. Öneriler.....	56
KAYNAKLAR.....	58
EKLER	63
EK-1. Özgeçmiş	63
EK-2. Etik Kurul Onayı.....	64

TEŐEKKÖR

Akademik kariyerimde en deęerli aőamalardan biri olan bu doktora alıőmamda, bilimsel vizyonu, engin birikimi ve deęerli rehberlięiyle yol gōsteren danıőmanım Prof. Dr. Cemil OLAK'a en iten teőekkōrlerimi sunarım. Kendisinin özverili yaklaőımı, akademik öngōrōsō ve sōrekli motivasyonu, bu zorlu sōrete en bōyōk destek kaynaęım olmuőtur.

Sadece bir danıőman olarak deęil, aynı zamanda bir akademik rol model olarak sergiledięi etik duruő ve bilimsel titizlik mesleki kariyerim boyunca bana rehber olacak deęerli ilkeler bırakmıőtır.

Doktora sōrecim boyunca karőılaőtıęım zorluklarda bana yardımcı olan ve desteęini esirgemeyen deęerli hocalarım Do. Dr. Mehmet Kıvrak ile Do. Dr. Ahmet Kadir ARSLAN'a ve bōlōmdeki dięer tōm hocalarıma da ayrıca teőekkōr ederim.

alıőmalarım boyunca maddi ve manevi anlamda desteęini hibir zaman esirgemeyen, sabırla yanımda olan sevgili eőime ve ocuklarıma ise borcumu hibir zaman ödeyemeyeceęimi bilerek, minnet ve őōkranlarımı sunarım. Onların sevgisi ve anlayıőı, bu zorlu sōrete en bōyōk dayanaęım olmuőtur.

Hasan UCUZAL

ÖZET

Tıbbi Görüntü Veri Setleri İçin Evrişimli Sinir Ağları Ve Görüntü Dönüştürücü Mimarilerine Yönelik Hibrit Sistemlerin Geliştirilmesi

Amaç: Bu çalışmanın amacı, tiroid nodülü ultrason görüntülerinin sınıflandırılmasında, CNN, ViT ve bu mimarilerin Hibrit kombinasyonlarının performansını kapsamlı bir şekilde karşılaştırmak ve her iki yaklaşımın güçlü yönlerini birleştiren hibrit bir modelin (HuViT) etkinliğini değerlendirmektir

Materyal ve Metot: Çalışmada, 6.905 tiroid ultrason görüntüsünden oluşan açık erişimli bir veri seti kullanılmıştır. Görüntülere yeniden boyutlandırma, normalizasyon ve çeşitli veri artırma teknikleri uygulanmıştır. Performans değerlendirmesi için ResNet-50, EfficientNet-B4, ConvNeXt, Swin Transformer, DeiT, PVTv2, XCiT, CoaT, MobileViT, ConViT ve özgün bir hibrit model olan HuViT olmak üzere 11 farklı derin öğrenme mimarisi eğitilmiştir. Modellerin karşılaştırılmasında doğruluk, kesinlik, duyarlılık, F1-skoru, özgüllük ve ROC-AUC metrikleri kullanılmıştır. Model kararlarının şeffaflığını artırmak için Grad-CAM, Score-CAM, Attention Rollout gibi açıklanabilir yapay zeka (XAI) yöntemleri entegre edilmiştir.

Bulgular: Deneysel sonuçlar, hibrit bir mimari olan ConViT modelinin tüm metriklerde en yüksek performansı sergilediğini göstermiştir (%83.30 doğruluk, %83.30 kesinlik, %83.26 duyarlılık, %83.30 F1-skoru, %89.94 AUC). ConvNeXt (%82.63 doğruluk, %90.02 AUC) ve DeiT (%81.18 doğruluk, %89.10 AUC) modelleri de yüksek ve rekabetçi performans göstermiştir. Özgün hibrit model HuViT ise %76.35 doğruluk elde etmiştir. XAI analizleri, modellerin karar mekanizmalarının radyolojik inceleme ile uyumlu olduğunu göstermiştir.

Sonuç: Bu çalışma, CNN'in yerel özellik çıkarımındaki gücü ile ViT'nin küresel bağlam modelleme kapasitesini birleştiren hibrit mimarilerin, tiroid ultrason görüntü analizinde saf mimarilere kıyasla üstün bir performans sergileyebileceğini kanıtlamıştır. Özellikle ConViT modelinin sergilediği yüksek ve dengeli başarı, bu tür hibrit yaklaşımların klinik karar destek sistemlerinde güvenilir bir araç olarak kullanılma potansiyelini ortaya koymaktadır.

Anahtar Kelimeler: Derin öğrenme, evrişimli sinir ağları, görüntü dönüştürücü, hibrit modeller, görüntü işleme, tiroid nodülü sınıflandırması, açıklanabilir yapay zeka.

ABSTRACT

Development of Hybrid Systems for Convolutional Neural Networks and Vision Transformer Architectures for Medical Image Datasets

Aim: The aim of this study is to comprehensively compare the performance of CNN, ViT, and their hybrid combinations in the classification of thyroid nodule ultrasound images, and to evaluate the effectiveness of a hybrid model (HuViT) that combines the strengths of both approaches.

Materials and Methods: A publicly available dataset consisting of 6,905 thyroid ultrasound images was used in the study. The images were resized, normalized, and augmented using various data augmentation techniques. Eleven different deep learning architectures were trained for performance evaluation, including ResNet-50, EfficientNet-B4, ConvNeXt, Swin Transformer, DeiT, PVTv2, XCiT, CoaT, MobileViT, ConViT, and an original hybrid model, HuViT. Performance comparison was conducted using accuracy, precision, recall, F1-score, specificity, and ROC-AUC metrics. To enhance the transparency of model decisions, explainable AI (XAI) methods such as Grad-CAM, Score-CAM, and Attention Rollout were integrated.

Results: Experimental results showed that the hybrid architecture ConViT model achieved the highest performance across all metrics (83.30% accuracy, 83.30% precision, 83.26% recall, 83.30% F1-score, 89.94% AUC). ConvNeXt (82.63% accuracy, 90.02% AUC) and DeiT (81.18% accuracy, 89.10% AUC) models also demonstrated high and competitive performance. The original hybrid model HuViT achieved 76.35% accuracy. XAI analyses confirmed that the decision mechanisms of the models were consistent with radiological examination.

Conclusion: This study demonstrated that hybrid architectures combining CNN's local feature extraction capability with ViT's global context modeling capacity can achieve superior performance in thyroid ultrasound image analysis compared to pure architectures. The high and balanced success of the ConViT model, in particular, highlights the potential of such hybrid approaches as reliable tools in clinical decision support systems.

Keywords: Deep learning, convolutional neural networks, vision transformer, hybrid models, image processing, thyroid nodule classification, explainable artificial intelligence.

SİMGELER VE KISALTMALAR DİZİNİ

AUC	: ROC Eğrisi Altında Kalan Alan
AI	: Artificial Intelligence (Yapay Zeka)
BT	: Bilgisayarlı Tomografi (Computerized Tomography)
CNN	: Convolutional Neural Network (Evrişimsel Sinir Ağı)
CPU	: Central Processing Unit (Merkezi İşlem Birimi)
CUDA	: Compute Unified Device Architecture
ConvSA	: Convolutional Gated Self-Attention (Evrişimli Kapılı Öz-Dikkat)
cuDNN	: CUDA Deep Neural Network
DL	: Deep Learning (Derin Öğrenme)
FN	: False Negative (Yanlış Negatif)
FP	: False Positive (Yanlış Pozitif)
GAT	: Graph Attention Network (Çizge Dikkat Ağı)
GCN	: Graph Convolutional Network (Çizge Evrişimli Ağ)
GELU	: Gaussian Error Linear Unit (Gauss Hata Doğrusal Birimi)
GNN	: Graph Neural Network (Çizge Sinir Ağı)
GPSA	: Graph-based Progressive Sampling Algorithm
GPU	: Graphics Processing Unit (Grafik İşlem Birimi)
LightGBM	: Light Gradient Boosting Machine
LSTM	: Long Short-Term Memory (Uzun Kısa-Vadeli Bellek)
MLP	: Multilayer Perceptron (Çok Katmanlı Algılayıcı)
MR	: Magnetic Resonance (Manyetik Rezonans)
NLP	: Natural Language Processing (Doğal Dil İşleme)

PCA	: Principal Component Analysis (Temel Bileşenler Analizi)
RAM	: Random Access Memory (Rastgele Erişimli Bellek)
ReLU	: Rectified Linear Unit (Düzeltilmiş Doğrusal Birim)
RGB	: Red, Green, Blue (Kırmızı, Yeşil, Mavi)
RNN	: Recurrent Neural Network (Yinelemeli Sinir Ağı)
SGD	: Stochastic Gradient Descent (Stokastik Gradyan İnişi)
SHAP	: SHapley Additive exPlanation
SVM	: Support Vector Machine (Destek Vektör Makineleri)
TN	: True Negative (Doğru Negatif)
TP	: True Positive (Doğru Pozitif)
t-SNE	: t-Distributed Stochastic Neighbor Embedding
ViT	: Vision Transformer (Görü Transformeri)
XAI	: Explainable Artificial Intelligence (Açıklanabilir Yapay Zeka)
XGBoost	: Extreme Gradient Boosting (Aşırı Gradyan Yükseltme)

ŞEKİLLER DİZİNİ

Şekil No	Sayfa No
Şekil 2.1. CNN mimarisi	4
Şekil 2.2. ViT mimarisi	5
Şekil 3.1. Veri seti yükleme.....	38
Şekil 3.2. Veri ön işleme	39
Şekil 3.3. Veri artırma	39
Şekil 3.4. Veri bölütleme	40
Şekil 3.5. Model eğitim	41
Şekil 3.6. Model değerlendirme	41
Şekil 3.7. Model açıklanabilirlik	42
Şekil 3.8. Sonuçlar ve raporlama.....	42
Şekil 4.1. Model performans karşılaştırması	44
Şekil 4.2. ROC eğrileri – Tüm modeller	45
Şekil 4.3. Karmaşık matrisi	46
Şekil 4.4. ConViT – ROC eğrisi.....	47
Şekil 4.5. En iyi modele ait kayıp değer grafiği	48
Şekil 4.6. En iyi modele ait doğruluk değeri grafiği	49
Şekil 4.7. Orijinal ultrason görüntüsü.....	50
Şekil 4.8. Grad-CAM yöntemiyle oluşturulmuş ısı haritası	51
Şekil 4.9. Grad-CAM ısı haritasının orijinal görüntü ile birleştirilmesi.....	51
Şekil 4.10. Modellerin Isı haritalarının karşılaştırılması	52

TABLÖLAR DİZİNİ

Tablo No	Sayfa No
Tablo 2.1. Hibrit Modeller Karşılaştırma Tablosu.....	20
Tablo 3.1. Model Eğitim Hiperparametreleri.....	31
Tablo 4.1. Görüntü Veri Seti Üzerindeki Performans Karşılaştırma Sonuçları.....	43
Tablo 4.2. ConvViT modeline ait sınıf bazlı performans metrikleri.....	45



1. GİRİŞ

Bilgisayarlı görü ve yapay zekâ alanındaki son dönemdeki hızlı gelişmeler, özellikle derin öğrenme tekniklerinin tıbbi görüntüleme alanında devrim niteliğinde yenilikler getirmesine olanak tanımıştır. Manyetik Rezonans Görüntüleme (MR), Bilgisayarlı Tomografi (BT), dijital patoloji ve fundus görüntüleri gibi yüksek çözünürlüklü ve karmaşık tıbbi veri setlerinin hacmi her geçen gün katlanarak artmakta; bu da bu verileri hassasiyet, güvenilirlik ve hızla analiz edebilecek otomatik sistemlere olan ihtiyacı kritik bir seviyeye taşımaktadır. Bu sistemlerin geliştirilmesi, hastalıkların erken ve doğru teşhisi, kişiselleştirilmiş tedavi stratejilerinin belirlenmesi ve klinisyenlerin teşhis yükünün hafifletilmesi noktasında büyük önem arz etmektedir. Başlangıçta yalnızca makine öğrenimi algoritmaları ile sınırlı olan bu yapay zekâ uygulamaları, günümüzde Derin Öğrenme (Deep Learning - DL), Görüntü Dönüştürücüler (Vision Transformers - ViT) ve Hibrit (Hybrid) model yaklaşımları gibi daha gelişmiş tekniklerle desteklenerek büyük bir dönüşüm geçirmiştir. Bu ileri teknolojiler, özellikle yüksek doğruluk ve hızlı analiz gerektiren teşhis süreçlerinde hekimlerin iş yükünü azaltmakta ve hastalara hızlı bir şekilde tanı konulmasına olanak tanımaktadır (1, 2).

Geleneksel olarak, tıbbi görüntü analizinde Evrişimli Sinir Ağları (Convolutional Neural Network - CNN) hakim mimari olmuştur. CNN'ler, yerel öznelikleri (kenar, doku, şekil) ve onların hiyerarşik kombinasyonlarını çıkarmadaki üstün yetenekleri sayesinde, sınıflandırma, segmentasyon ve nesne tespiti gibi birçok görevde insan seviyesine yakın performans sergilemiştir. Ancak, bu mimarilerin temel bir sınırlaması, uzun menzilli bağımlılıkları (long-range dependencies) ve görüntünün global bağlamını (global context) modellemede yetersiz kalmalarıdır. Evrişimsel işlemlerin doğası gereği, bir katmandaki bir nöron yalnızca bir önceki katmandaki komşu nöronlardan girdi alır (yerel reseptif alan). Global bir ilişkinin anlaşılması için çok sayıda katmanın birikimli etkisine güvenmek zorunda kalır, bu da hem hesaplama verimsizliğine yol açabilir hem de kritik global bağlam bilgisini yitirme riski taşır. Buna karşılık, doğal dil işleme alanında devrim etkisi yapan dönüştürücü (transformer) mimarisi, özellikle de ViT, öz-dikkat (self-attention) mekanizması aracılığıyla bu soruna yenilikçi bir çözüm sunar. ViT'ler, bir görüntüyü bir dizi yama (patch) olarak ele alır ve öz-dikkat sayesinde, görüntüdeki her bir yamanın diğer tüm yamalar ile olan ilişkisini doğrudan

modelleyebilir. Bu, modelin, bir tümörün farklı bölgeleri arasındaki ince ilişkiler veya bir organın global anatomik yapısı gibi uzun menzilli bağımlılıkları yakalama kapasitesini önemli ölçüde artırır. Ancak, ViT'ler genellikle çok büyük miktarda veriye ihtiyaç duyarlar ve CNN'lerin yerel öznelilik çıkarımındaki yerleşik verimliliğinden ve örneklem karmaşıklığından (sample efficiency) yoksundurlar. Bu durum, her iki mimarinin de güçlü yönlerini tek bir çatı altında birleştirerek sinerji yaratmayı hedefleyen hibrit mimarilere olan akademik ve pratik ilgiyi ateşlemiştir. Mevcut literatürde, hem CNN hem de ViT modellerinin performansını aşan hibrit yaklaşımlara dair umut verici çalışmalar bulunmakla birlikte, bu modellerin mimari tasarımı, optimal entegrasyon noktaları, hesaplama verimlilikleri ve farklı tıbbi görüntüleme yöntemleri üzerindeki genelleştirilebilirliği hala aktif bir araştırma alanı olarak önümüzde durmaktadır (3, 4).

Radyoloji (kitle bölütleme), patoloji (mitoz sayımı, kanser alt tiplerinin sınıflandırılması), dermatoloji (ben kanseri teşhisi) ve oftalmoloji (diyabetik retinopati taranması) gibi görüntüleme ağırlıklı tıbbi alanlarda yaygın olarak benimsenen bu hibrit modeller, aynı zamanda klinik karar destek sistemlerinde de devrim niteliğinde yenilikler sunmaktadır. Hekimlere nesnel, ölçeklenebilir ve yorulmaz bir ikinci görüş sağlayarak erken ve daha doğru tanıyı mümkün kılar, sonuçta hasta sonuçlarını iyileştirme ve klinik iş yükünü azaltma vaadi taşırlar. Verimlilikleri, sınırlı etiketli veriyle bile güçlü performans gösterebilmeleri sayesinde daha da artmaktadır (5).

Geleneksel makine öğrenimi algoritmalarının sunduğu temel analiz kapasitelerine ek olarak, derin öğrenme ve görüntü dönüştürücü teknolojilerin entegre edilmesi, tıbbi verilerin çok boyutlu özelliklerinin daha etkin bir şekilde ortaya çıkarılmasını sağlamaktadır. Bu gelişmeler sayesinde hastalıkların erken teşhis edilme ve tedaviye daha hızlı başlanma potansiyeli artmakta, böylece hasta sonuçları iyileşmektedir. Sonuç olarak, hibrit model yaklaşımlarının sunduğu yeniliklerin, tıbbi görüntü analizinde gelecekte daha yaygın bir şekilde kullanılacağı ve sağlık alanında kalıcı bir dönüşüm sağlayacağı öngörülmektedir.

Bu çalışmada, literatürdeki mevcut yöntemlerle rekabet edebilecek özgün yaklaşımlar sunmayı amaçlayarak CNN ve ViT modellerinin entegre edildiği hibrit yaklaşımlarının açık erişimli veri set(ler)inde kullanımını ve uygulamasını araştırmak amaçlanmıştır. Çalışmanın temel motivasyonu, her iki mimarinin de güçlü yönlerini (CNN'lerin yerel özellik çıkarımındaki verimliliği ve ViT'lerin global bağlamı modelleme kapasitesi) tek bir uyumlu çatı altında birleştirmek ve bu sinerjinin tıbbi görüntü analizi görevlerindeki etkinliğini ortaya koymak amaçlanmaktadır.

2. GENEL BİLGİLER

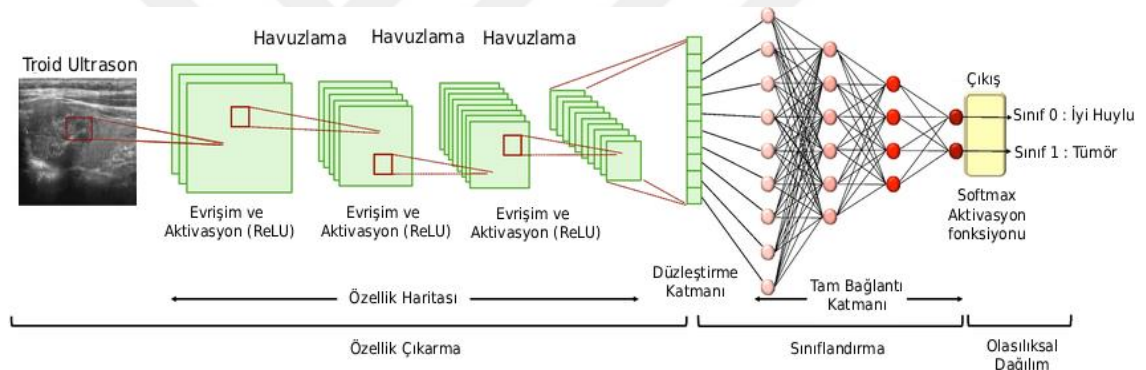
Son yıllarda görüntü işleme ve bilgisayarlı görü alanında kaydedilen ilerlemeler, DL tabanlı mimarilerin yaygınlaşmasını beraberinde getirmiştir. Özellikle CNN'ler, görüntü sınıflandırma, nesne tespiti ve bölütleme gibi görevlerde yüksek başarı oranlarına ulaşmış ve bu alanda bir standart haline gelmiştir (6). CNN'lerin bu başarısının ardındaki temel faktör, yerel bağlantılar, ağırlık paylaşımı ve havuzlama (pooling) katmanları aracılığıyla oluşturdıkları öznitelik hiyerarşisi ve uzamsal özellikleri verimli bir şekilde öğrenme kabiliyetleridir. Ancak, evrimsel işlemlerinin doğası gereği sahip oldukları yerel reseptif alan (local receptive field) sınırlaması, bir katmandaki bir nöronun yalnızca bir önceki katmandaki komşu nöronlardan girdi almasına neden olmaktadır. Bu durum, uzun menzilli bağımlılıkların (long-range dependencies) ve küresel bağlamın (global context) ancak çok sayıda katmanın birikimli etkisiyle modellenebileceği anlamına gelir. Bu da, hem hesaplama verimsizliğine yol açabilir hem de modelin görüntünün global yapısını tam olarak kavramasını engelleyerek kritik bağlamsal bilgilerin kaybolmasına neden olabilir (7).

Bu sınırlamaları aşmak amacıyla geliştirilen Görüntü Dönüştürücü ViT mimarileri, doğal dil işleme (NLP) alanında devrim yaratan Transformer yapısını görüntü verilerine uyarlamaktadır. Transformer mimarilerinin çekirdeğini oluşturan çok kafalı öz-dikkat (multi-head self-attention) mekanizması, bir dizideki (sequence) her bir ögenin (örneğin, bir görüntü yamasının - image patch) diğer tüm öğelerle olan ilişkisini doğrudan hesaplama olanağı tanır. Bu mekanizma, ViT'lerin uzun menzilli ve küresel ilişkileri (global dependencies) doğuştan (innately) ve yüksek doğrulukla modellemesine olanak sağlar, bu da onları özellikle yapısal olarak karmaşık tıbbi görüntülerin analizinde ümit verici bir araç haline getirir. Ancak ViT'lerin bu üstünlüğü, önemli bir trade-off (denge ödünleşimi) ile gelir. İyi generalize olabilmeleri ve öğrenme yeteneklerini tam olarak sergileyebilmeleri için çok büyük miktarda etiketli eğitim verisine ve yoğun hesaplama gücüne ihtiyaç duyarlar. Özellikle tıbbi görüntüleme alanında olduğu gibi veri sayısının nispeten sınırlı ve veri toplama/etiketlemenin maliyetli ve zaman alıcı olduğu domain'lerde, saf ViT modelleri yetersiz örneklem verimliliği (sample inefficiency) nedeniyle beklenen performansı gösteremeyebilir ve aşırı öğrenme (overfitting) eğiliminde olabilirler (8, 9).

Söz konusu iki mimarinin CNN'lerin yerel öznitelik çıkarımındaki yapısal verimliliği ile ViT'lerin küresel bağlamı modellemedeki üstün kapasitesinin bir araya getirilmesi fikri, bu çalışmanın temel motivasyonunu oluşturmaktadır. Literatürde, bu hibrit yaklaşımların, her iki dünyanın da en iyi yönlerini alarak hem verimliliği hem de performansı artırabileceği yönünde güçlü bir kanıt bulunmaktadır.

2.1. CNN Mimari Yapısı

CNN, görsel verilerde yerel özellikleri öğrenmeye odaklanmış derin öğrenme modelleridir. CNN'ler, bir görüntüdeki yerel özellikleri tespit etmek için evrişimli katmanlar kullanır. Bu katmanlar, görüntü üzerinde kayarak özellikleri çıkaran filtreler (kernels) uygular. CNN'ler, özellikle görsel verilerdeki uzamsal hiyerarşiyi öğrenme konusunda etkilidir ve nesne tanıma, görüntü sınıflandırma gibi görevlerde kullanılır (10, 11). CNN mimarisi Şekil 2.1.'de gösterildi.



Şekil 2.1. CNN mimarisi

2.1.1. Evrişimli Katmanlar

Bu katmanlar, görüntülerdeki yerel özellikleri çıkarır. Evrişimli işlemler, küçük filtreler (kernels) kullanarak görsel veriyi işler. Görüntü üzerinde filtreler (kernels) kaydırılarak yerel özellikler (kenarlar, dokular, renk geçişleri) çıkarılır. Her filtre, belirli bir özelliği tespit etmeyi öğrenir (örneğin, yatay kenarlar için bir filtre). Stride (kayma adımı) ve padding (kenar genişletme) parametreleri, çıktı boyutunu ve bilgi kaybını kontrol eder (11).

2.1.2. Havuzlama (Pooling) Katmanları

Havuzlama, özellik haritalarının boyutlarını küçültür ve verinin boyutunu azaltırken önemli bilgileri korur. En yaygın havuzlama yöntemi max pooling'dir.

Maksimum Pooling veya Ortalama Pooling ile öznitelik haritalarının boyutu azaltılır, hesaplama verimliliği artar ve aşırı uyum (overfitting) riski düşer (12).

2.1.3. Aktivasyon Fonksiyonları

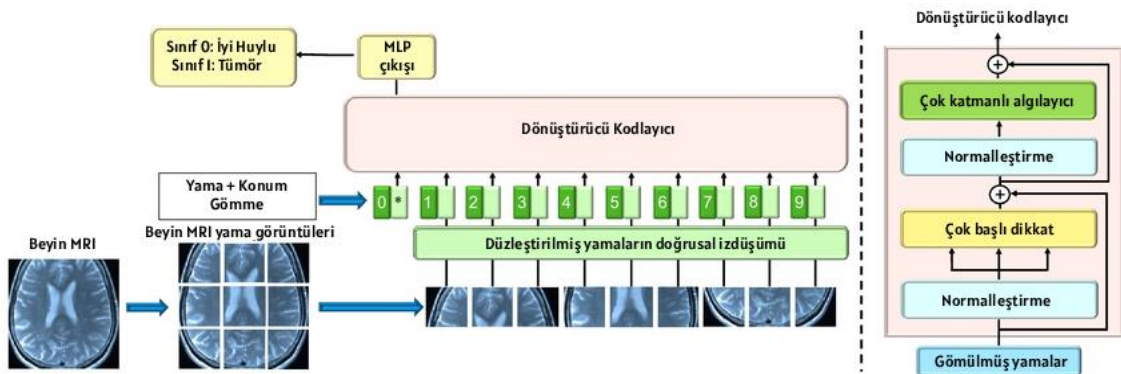
ReLU (Rectified Linear Unit), negatif değerleri sıfırlayarak modelin doğrusal olmayan ilişkileri öğrenmesini sağlar (13).

2.1.4. Tam Bağlı Katmanlar (Fully Connected Layers)

Konvolüsyon ve pooling ile çıkarılan özellikler, sınıflandırma veya regresyon için bu katmanlarda işlenir. Derin katmanlar tarafından öğrenilen özellikleri kullanarak nihai tahmin veya sınıflandırma kararını verir (14).

2.2. ViT Mimari Yapısı

ViT, görsel verileri işlemek için Transformer mimarisini kullanan bir modeldir. ViT, görüntüleri küçük parçalara böler ve her bir parçayı ayrı bir vektöre dönüştürür. Bu parçalar daha sonra öz-dikkat mekanizması ile birbirleriyle olan ilişkilerini öğrenir. ViT, özellikle uzun menzilli bağıntıları öğrenme yeteneği ile öne çıkar, en iyi performansı büyük veri setleriyle çalışırken gösterir (8). ViT mimarisi Şekil 2.2.'de gösterildi.



Şekil 2.2. ViT mimarisi

2.2.1. Görüntüyü Yamalara Ayırma

Görüntü, küçük parçalara (yamalar - patches) bölünür ve her bir yama (patch), doğrusal olarak bir vektöre (düzeltilmiş bir dizi sayıya) dönüştürülür. Bu işlem, genellikle 16x16 veya 32x32 piksel gibi sabit bir boyuttaki yamalar (patches) kullanılarak, görüntünün bir ızgara deseni halinde token'lara (dil işlemedeki kelimelerin eşdeğeri) ayrıştırılması şeklinde gerçekleştirilir (8).

2.2.2. Öğrenilebilir Gömmeler (Embeddings)

Bu yama gömmeleri (patch embeddings), öğrenilebilir bir doğrusal izdüşüm (linear projection) veya bir evrişim katmanı aracılığıyla modelin işleyeceği boyutta bir gömme uzayına (embedding space) taşınır. Sınıflandırma (classification) görevleri için, bu yama dizisinin en başına, tüm görüntüyü temsil etmesi beklenen özel bir sınıf token'ı ([class] token) eklenir. Son olarak, Transformer mimarisinin dizideki öğelerin sırasını (konumunu) anlayabilmesi için, her bir tokena (sınıf token'ı dahil), önceden tanımlanmış veya öğrenilmiş bir konumsal gömme (positional embedding) eklenir. Bu konumsal gömmeler, yamaların orijinal görüntüdeki uzamsal düzenine ilişkin bilgiyi koruyarak modelin bağlamı doğru bir şekilde öğrenmesini sağlar (15).

2.2.3 Dönüştürücü Kodlayıcı (Transformer Encoder)

Her bir vektör, öz-dikkat katmanlarından geçirilerek, her parçanın diğer parçalarla olan ilişkisini öğrenir. Bu sayede, uzun menzilli bağıntılar öğrenilmiş olur. MLP (Çok Katmanlı Algılayıcı) blokları, öznitelikleri dönüştürür (15).

2.2.1. Çok Kafalı Dikkat (Mekanizması) (Multi-head Attention)

Öz-dikkat, birden fazla dikkat başlığı kullanılarak paralel şekilde işlenir. Modelin her bir parçayı farklı açılardan öğrenmesine olanak tanır (15).

2.2.2. İleri Beslemeli Sinir Ağları (Feed-Forward Neural Networks)

Dikkat katmanının ardından gelen feed-forward ağlar, modelin daha karmaşık temsiller öğrenmesini sağlar.

2.3. Veri ve Eğitim Süreci

2.3.1. CNN Veri ve Eğitim Süreci

Derin öğrenme modelleri genellikle etiketlenmiş büyük veri setlerine ihtiyaç duyar. Bu veriler, döndürme, ölçekleme ve parlaklık değişimi gibi veri artırma teknikleri ile çeşitlendirilebilir. Eğitim süreci, genellikle geri yayılım algoritması ve optimizasyon yöntemleriyle, örneğin stokastik gradyan inişi (SGD) veya Adam kullanılarak gerçekleştirilir. Bu süreçte, modelin ağırlıkları, hata sinyali geri yayılımı ile güncellenir. Model, başlangıçta yerel özellikleri öğrenmeye odaklanırken, derinleştikçe daha karmaşık soyutlamalar yapar (10, 11).

2.3.2. ViT Veri ve Eğitim Süreci

ViT, büyük veri kümelerine ihtiyaç duyar ve yeterli veri sağlandığında daha etkili çalışır. Model genellikle ön eğitim (pre-training) süreci ile başlar; bu süreçte geniş bir veri seti üzerinde genel özellikler öğrenilir. Eğitim süreci, önceden büyük veri kümesinde yapılan eğitimden sonra, ince ayar (fine-tuning) ile devam eder. İnce ayar, modelin spesifik görevler için optimize edilmesini sağlar. ViT, transfer öğrenme (transfer learning) yöntemini kullanarak eğitildiğinde, daha küçük veri kümelerinde bile etkili sonuçlar elde edilebilir (8, 15).

2.4. Performans Karşılaştırması

2.4.1. CNN Performansı

Küçük veri kümelerinde, CNN modelleri hızlı bir şekilde öğrenebilir ve yüksek performans sergileyebilir. Özellikle yerel özelliklerin çıkarılması konusunda oldukça etkilidir. CNN'ler, genellikle daha hızlı eğitilebilir çünkü yerel bağımlılıkları öğrenir ve parametre sayısı daha azdır. Bu özellikleri sayesinde, görüntü sınıflandırma, nesne tanıma ve bölütleme gibi görsel görevlerde güçlü bir performans sergiler (16).

2.4.2. ViT Performansı

ViT, genellikle büyük veri kümelerinde daha iyi performans sergiler. Uzun menzilli bağıntıları öğrenme kabiliyeti, büyük veri setlerinde güçlüdür. Ancak, küçük veri kümelerinde ViT genellikle daha düşük performans gösterir. Bu nedenle, küçük veri setlerinde etkili olabilmesi için transfer learning gibi tekniklerle büyük veri setlerinde önceden eğitilmesi gerekir. ViT, uzun menzilli ilişkileri öğrenme yeteneği sayesinde daha karmaşık görevlerde güçlüdür ve özellikle büyük veri setlerinde daha iyi sonuçlar elde edilir (17).

2.5. Hesaplama Maliyetleri ve Verimlilik

2.5.1. CNN Hesaplama Maliyetleri

CNN'ler, daha az parametreyle çalışarak görüntüdeki yerel bağımlılıkları öğrenir. Bu özellik, daha az bellek ve işlem gücü kullanmasını sağlar. Evrişimli katmanlar, yerel bağımlılıkları öğrendiği için eğitim süreci daha hızlıdır ve daha az hesaplama gücü gerektirir. Bu sayede, küçük veri kümeleriyle etkili bir şekilde çalışabilirler. Ayrıca,

CNN'ler, anlık uygulamalarda ve düşük donanımda hızlı sonuçlar verebilir, bu da onları çeşitli pratik kullanım alanları için ideal hale getirir (16).

2.5.2. ViT Hesaplama Maliyetleri

ViT, büyük veri kümelerinde etkili olsa da, her bir parçayı işlerken yüksek hesaplama gücü ve bellek gerektirir. Özellikle öz-dikkat mekanizması, işlem gücü açısından oldukça yoğun olup, büyük kaynaklar tüketir. Bu nedenle, ViT'nin eğitim ve tahmin süreleri, CNN'ye kıyasla genellikle daha uzundur. Büyük veri kümelerinde yüksek performans sağlansa da, genellikle güçlü GPU'lar ve büyük donanımlar gereklidir. Ayrıca, ViT küçük veri setlerinde düşük verimlilik gösterirken, büyük veri setlerinde verimlilik artar (17) .

2.6. Uygulama Alanları

2.6.1. CNN Uygulama Alanları

Görüntü Sınıflandırma: Görüntüleri belirli kategorilere ayırmak. Örneğin, yüz tanıma, el yazısı tanıma, tıbbi görüntü analizi.

Nesne Tanıma ve İzleme: Görsel verilerdeki nesneleri tanımak ve takip etmek. Bu, güvenlik kameraları, otonom araçlar gibi uygulamalarda kullanılır.

Görüntü Bölütleme: Görüntülerdeki belirli bölgeleri ayırmak. Örneğin, tıbbi görüntü bölütleme, şehir planlama.

Video Analizi: Hareketli görüntülerde nesnelerin tanımlanması ve takip edilmesi.

2.6.2. ViT Uygulama Alanları

Büyük Veri Kümesiyle Görüntü Tanıma: ViT, büyük veri kümeleriyle daha verimli çalışır. Örneğin, ImageNet gibi büyük veri setlerinde görüntü sınıflandırma görevlerinde kullanılır.

Uzun Menzilli Bağlantılar Gerektiren Görevler: ViT, uzun menzilli ilişkilerin öğrenilmesinin kritik olduğu görevlerde öne çıkar. Örneğin, görüntü anlamlandırma, görsel anlam çıkarma.

Yüksek Performans Gerektiren Uygulamalar: ViT, büyük veri kümeleri üzerinde yüksek performans gösterdiği için gelişmiş görsel analiz ve artırılmış gerçeklik gibi uygulamalarda tercih edilir.

2.7. Hibrit Mimarilerin Sınıflandırılması

Hibrit yapılar, tek bir modelin sınırlı öğrenme kapasitesini aşmak, farklı mimarilerin sinerjistik güçlü yanlarını birleştirerek daha doğru, sağlam (robust) ve genellenebilir (generalizable) sonuçlar üretmek amacıyla geliştirilmiştir. Özellikle görüntü işleme (image processing), doğal dil işleme (natural language processing - NLP) ve biyomedikal veri analizi gibi disiplinler arası alanlarda öne çıkan bu yapılar, modern yapay öğrenme sistemlerinde vazgeçilmez bir değer hâline gelmiştir.

Bu hibrit yaklaşımların temel mantığı, birleşik bir model çerçevesi (unified model framework) içinde, farklı algoritmaların veya mimarilerin tamamlayıcı yeteneklerinden (complementary capabilities) faydalanmaktır. Bu mimarilerin stratejik olarak bir araya getirilmesiyle oluşturulan hibrit sistemler, tek başına kullanıldıklarında ortaya çıkan örtük önyargı (inductive bias) sınırlamalarını azaltır ve modelin verideki karmaşık desenleri çok boyutlu ve çok ölçekli (multi-scale) olarak kavramasını sağlar. Özellikle tıbbi görüntüleme gibi gürültülü (noisy), sınırlı etiketli ve yüksek boyutlu (high-dimensional) veri kümeleriyle çalışılırken, hibrit modeller daha düşük bir varyans (variance) ve daha yüksek bir örneklem verimliliği (sample efficiency) sergilerler. Bu da onları, klinik karar destek sistemleri gibi yüksek riskli ve yüksek doğruluk gerektiren gerçek dünya uygulamaları için son derece uygun kılar (18).

Sonuç olarak, hibrit yapılar sadece performans metriklerini iyileştirmekle kalmaz, aynı zamanda XAI için de yeni olanaklar sunar.

2.7.1. Entegrasyon Stratejisine Göre

Sıralı (Sequential) Hibritler

Sıralı (sequential) hibrit modeller, yapay zekâ ve makine öğrenmesi alanlarında farklı algoritmaların veya modelleme tekniklerinin belirli bir işlem sırasına göre bir araya getirilmesiyle oluşturulan yapay öğrenme sistemleridir. Bu modellerde her bir bileşen, belirli bir veri işleme görevini üstlenmekte ve elde ettiği çıktıyı bir sonraki modelin girdisi olarak aktarmaktadır. Bu yönüyle sıralı hibrit yapılar, bir işlem hattı (pipeline) mimarisi çerçevesinde organize edilir (19).

Sıralı hibrit modellerin uygulama süreçleri genellikle çok aşamalı olarak planlanmaktadır. Örneğin, tıbbi görüntü analizi alanında sıklıkla karşılaşılan bir uygulama senaryosunda; ilk aşamada bir derin öğrenme modellerinden (örneğin, CNN, LSTM, vb.) ile görüntüden özellik çıkarımı gerçekleştirilir. İkinci aşamada boyut

indirgeme algoritmaları (Principal Component Analysis - PCA veya t-Distributed Stochastic Neighbor Embedding - t-SNE gibi) kullanılarak verinin boyutu ve karmaşıklığı azaltılır. Son aşamada ise bir makine öğrenmesi algoritması (örneğin, Destek Vektör Makineleri – SVM, Rastgele Ormanlar – Random Forest, Gradyan Artırmalı Ağaçlar – XGBoost) ile sınıflandırma işlemi gerçekleştirilir. Bu ardışık yapı sayesinde her model, veriyi daha anlamlı ve işlenebilir hale getirerek sistemin genel başarımını artırır (20).

Sıralı hibrit modellemenin öne çıkan avantajları arasında:

- Farklı öğrenme algoritmalarının sinerjik biçimde kullanılması,
- Görev odaklı bileşenlerin entegrasyonu,
- Özellik mühendisliği, boyut indirgeme ve sınıflandırma gibi süreçlerin ayrıştırılarak optimize edilebilmesi bulunmaktadır.

Ancak bu modellerin geliştirilmesinde dikkat edilmesi gereken bazı sınırlılıklar da mevcuttur. Model yapısının giderek karmaşıklaşması, sistemin yorumlanabilirliğini zorlaştırabilir; ayrıca ardışık yapının doğası gereği, erken aşamalarda oluşan hatalar sistem genelinde hata artışına neden olabilir. Bunun yanı sıra, özellikle derin öğrenme bileşenleri içeren yapılarda yüksek işlem gücü ve kaynak gereksinimi, uygulama açısından dikkate alınması gereken unsurlardandır. Sıralı hibrit modeller; görüntü işleme, doğal dil işleme, biyomedikal veri analizi, finansal modelleme ve eğitim teknolojileri gibi çok sayıda alanda başarıyla uygulanmaktadır. Doğal dil işleme örneğinde, kelime gömme (word embedding) katmanının ardından sıralı bir LSTM modeli ve sonrasında bir sınıflayıcı entegre edilerek metin sınıflandırma işlemi gerçekleştirilebilmektedir (21).

Paralel Hibritler

Paralel hibrit modeller, yapay zekâ ve makine öğrenmesi alanlarında birden fazla modelin veya algoritmanın aynı veri kümesi üzerinde eşzamanlı ve bağımsız biçimde çalıştırılması ve bu modellerden elde edilen çıktıların belirli birleştirme stratejileri doğrultusunda sentezlenmesiyle oluşturulan bütünleşik öğrenme yaklaşımlarıdır. Bu yöntem, farklı model türlerinin öğrenme kapasitelerinden eşzamanlı olarak faydalanarak sistemin genel başarımını ve kararlılığını artırmayı amaçlamaktadır.

Bu modelleme yaklaşımında, her bir öğrenme algoritması aynı veri setini bağımsız olarak işler ve elde ettiği çıktıyı üretir. Daha sonra bu çıktılar, karar düzeyinde füzyon (decision-level fusion) yöntemleri kullanılarak birleştirilir. Bu füzyon işlemi çoğunluk oylaması (majority voting), ağırlıklı oylama (weighted voting), ortalama alma

(averaging) veya daha gelişmiş meta-öğrenme stratejileri ile gerçekleştirilebilir. Bu süreçteki temel hedef, bireysel modellerin sınırlılıklarını telafi edecek şekilde, modeller arası tamamlayıcılıktan faydalanarak daha yüksek doğruluk ve genelleme başarımı elde etmektir (22).

Paralel hibrit modelleme, özellikle heterojen, yüksek boyutlu ve karmaşık veri yapılarında etkili sonuçlar üretebilmektedir. Örneğin, biyomedikal veri analizi alanında, aynı biyobelirteç veri seti üzerinde farklı sınıflandırma algoritmalarının (örneğin Random Forest, SVM, XGBoost, LightGBM) paralel olarak çalıştırılması ve sonuçların birleştirilmesi, sistemin genel güvenilirliğini ve hata toleransını artırmaktadır. Paralel hibrit modellerin temel avantajları arasında model çeşitliliğinin sunduğu öğrenme gücü, aşırı öğrenme (overfitting) riskinin azalması, farklı öğrenme perspektiflerinin entegre edilebilmesi ve çıktının daha dengeli şekilde üretilmesi yer almaktadır. Ancak, sistemin eşzamanlı olarak birden fazla modeli çalıştırması nedeniyle hesaplama maliyetinin yüksek olması, uygulama karmaşıklığının artması ve çıktı birleştirme stratejisinin uygun şekilde yapılandırılmaması hâlinde sistemin beklenen düzeyde performans gösterememesi gibi sınırlılıklar da söz konusudur (23).

Dönüşümsel Hibritler

Dönüşümsel hibrit modeller, yapay zekâ ve makine öğrenmesi bağlamında birden fazla modelin ardışık biçimde entegre edildiği ve her bir bileşenin yalnızca bir çıktıyı üretmekle kalmayıp aynı zamanda verinin temsilini anlamsal ve yapısal olarak dönüştürdüğü bütünsel modelleme yaklaşımlarıdır. Bu yapı, klasik sıralı hibrit modellerden farklı olarak, veri akışını sadece bir modelden diğerine aktarmakla yetinmeyip, her aşamada verinin içeriksel anlamını yeniden tanımlayarak bir üst düzey soyutlama düzlemine taşır (24).

Dönüşümsel hibrit modellerin temel varsayımı, her bir bileşenin, yalnızca görev odaklı bir çıktı üretmekten öte, sonraki modelin öğrenme kapasitesini artıracak nitelikte yeniden yapılandırılmış bir özellik uzayı üretmesidir. Bu yaklaşım, özellikle çok boyutlu, heterojen veya yapılandırılmamış veri tiplerinin işlendiği alanlarda önemli avantajlar sunmaktadır. Örneğin, doğal dil işleme alanında, metin verileri ilk aşamada kelime gömme (word embedding) yöntemleriyle vektörel temsillere dönüştürülmekte, sonrasında bu temsiller dikkat mekanizmaları (attention) ve dönüştürücü mimariler (transformer) kullanılarak daha üst düzey anlamsal yapılara dönüştürülmekte ve nihai

olarak sınıflandırma işlemi gerçekleştirilmektedir. Bu tür bir ardışık yapıda her model, yalnızca veri aktarımı değil, aynı zamanda temsil dönüşümü gerçekleştirmektedir (15).

Dönüşümsel hibrit modellerin uygulama alanları oldukça geniştir. Biyoenformatik (örneğin, genomik verilerin çok aşamalı öznitelik dönüşümleriyle analiz edilmesi), doğal dil işleme (word embedding → attention → transformer zincirleri), görüntü işleme (görsel özelliklerin ViT gibi yapılarda yeniden temsil edilmesi) ve öğrenme analitiği (öğrenci davranışlarının çok düzeyli modellenmesi) gibi alanlarda başarıyla kullanılmaktadır. Bu modeller, veri temsillerinin sürekli dönüşerek daha soyut ve anlamlı biçimlere ulaşması yoluyla öğrenme sürecinin derinleştirilmesine katkı sağlar (25).

2.7.2. Mimari Tasarıma Göre

CNN Temelli Hibritler

CNN'ler, özellikle görsel verilerin işlenmesinde gösterdiği üstün performans sayesinde, derin öğrenme literatüründe yaygın olarak tercih edilen model yapılarından biri hâline gelmiştir. CNN'ler, yerel örüntüleri yakalayabilme, çok katmanlı temsiller oluşturma ve özellik çıkarımını otomatik olarak gerçekleştirme yetenekleri sayesinde yüksek boyutlu ve karmaşık veri yapılarında etkili bir öğrenme mekanizması sunmaktadır. Bununla birlikte, CNN tabanlı modellerin tek başına karar verme süreçlerinde sınırlı kalabilmesi, bu yapının diğer makine öğrenmesi algoritmalarıyla bütünleştirilerek hibrit bir mimari içerisinde değerlendirilmesine olanak tanımıştır. Bu bağlamda geliştirilen CNN temelli hibrit modeller, evrimsel öğrenmenin güçlü özellik çıkarım yeteneğini farklı karar sistemleriyle birleştirerek daha yüksek doğruluk, genellenebilirlik ve yorumlanabilirlik sunmayı hedefleyen bütünlük yapay öğrenme sistemleridir.

CNN temelli hibrit modeller genel olarak iki temel bileşenden oluşmaktadır. İlk aşamada yer alan CNN mimarisi, girdi veriyi (örneğin görüntü, sinyal, metin temsilleri) çok katmanlı evrimsel filtreler aracılığıyla işleyerek, ham veriden ayrıştırıcı nitelikte öznitelikler (feature maps) üretmektedir. Bu temsiller, modelin ikinci aşamasında yer alan karar birimine aktarılmakta ve burada sınıflandırma, regresyon veya sıralama gibi görevler yerine getirilmektedir. Karar birimi olarak çoğunlukla geleneksel makine öğrenmesi algoritmaları (SVM, Random Forest, XGBoost) veya farklı derin öğrenme

yapıları (örneğin Tekrarlayan Sinir Ağları – RNN, Uzun Kısa Süreli Bellek – LSTM, Dönüştürücü Mimariler – Transformer) tercih edilmektedir (26).

Bu tür hibrit yapıların öne çıkan avantajları arasında, manuel özellik mühendisliği gereksinimini azaltması, çeşitli karar sistemleriyle esnek biçimde entegre edilebilmesi, CNN tarafından çıkarılan temsillerin farklı bağlamlarda yeniden işlenebilmesi ve performans artışı sağlaması sayılabilir. Ayrıca, CNN'in yorumlanabilirliğini artırmaya yönelik geliştirilen yöntemler (örneğin Grad-CAM, SHAP) ile birlikte kullanıldığında, sistemin genel yorumlanabilirliği de yükseltilmektedir (27, 28).

CNN temelli hibrit modellerin uygulama alanları oldukça geniştir. Tıbbi görüntüleme sistemlerinde, örneğin tümörlerin segmentasyonu ve sınıflandırılmasında; uzaktan algılama uygulamalarında uydu görüntülerinden arazi örtüsü çıkarımında; biyoinformatik çalışmalarında mikroskobik hücre görüntülerinin analizi gibi alanlarda başarıyla kullanılmaktadır. Ayrıca eğitim teknolojilerinde öğrenci davranışlarının görsel analizine dayalı modellerin geliştirilmesi ve tarımsal görüntülerden hastalık tespiti gibi uygulamalarda da yaygın biçimde yer bulmaktadır (29).

ViT Temelli Hibritler

ViT, görsel veri analizi alanında son dönemde öne çıkan ve geleneksel CNN'den farklı olarak, transformer mimarisine dayanan bir modeldir. ViT, görüntüleri sabit boyutlu yama (patch) bloklarına ayırarak bu blokları sıralı vektör temsillerine dönüştürmekte, ardından bu temsilleri çok başlı dikkat (multi-head attention) mekanizmaları yoluyla işleyerek küresel bağlamı öğrenmektedir. Özellikle uzun menzilli bağıntıları modelleyebilme yeteneği ile dikkat çeken ViT modelleri, yüksek çözünürlüklü görsellerde anlamlı temsiller üretme konusunda güçlü bir alternatif sunmaktadır. Bununla birlikte, ViT modelleri yüksek veri gereksinimi, hesaplama maliyeti ve sınırlı yorumlanabilirlik gibi nedenlerle, tek başına tüm karar verme sürecini kapsayacak şekilde yeterli olmayabilir. Bu durum, ViT'nin farklı algoritmalarla bütünleştirilerek hibrit yapılarda değerlendirilmesi gerekliliğini ortaya koymuştur.

ViT temelli hibrit modeller, ViT bileşeninin öznetelik çıkarımı işleviyle, diğer makine öğrenmesi algoritmalarının veya yapay sinir ağı bileşenlerinin karar üretme süreçlerini desteklediği çok katmanlı sistemler olarak tanımlanabilir. Bu modellerin temel amacı, ViT'nin küresel temsil öğrenme yeteneğini, başka algoritmaların yerel örüntüleri yakalama veya karar düzeyinde sınıflandırma becerileriyle birleştirmektir. Söz konusu hibrit yapılar, çoğunlukla iki aşamalı olarak kurgulanmaktadır. İlk aşamada ViT mimarisi

aracılığıyla giriş verisinden üst düzey temsiller elde edilir. İkinci aşamada ise bu temsiller, klasik makine öğrenmesi algoritmaları (örneğin Destek Vektör Makineleri, Rastgele Ormanlar, Gradyan Artırmalı Ağaçlar), derin öğrenme tabanlı yapılar (örneğin CNN, LSTM) veya istatistiksel modelleme yöntemleri ile analiz edilerek karar verme süreci tamamlanır. Alternatif bir hibrit yaklaşımda, ViT ve CNN bileşenleri paralel biçimde yapılandırılarak aynı veri üzerinde hem küresel hem yerel temsillerin eşzamanlı olarak öğrenilmesi hedeflenmektedir (30).

ViT temelli hibrit modeller, özellikle tıbbi görüntü analizi, uzaktan algılama, endüstriyel kalite kontrol, güvenlik ve savunma teknolojileri, biyomedikal görüntüleme gibi yüksek hassasiyet ve yorumlanabilirliğin kritik olduğu alanlarda etkili biçimde kullanılmaktadır. Örneğin, bir ViT modeli aracılığıyla tümör bölgesi küresel ölçekte haritalanabilirken, aynı yapının CNN bileşeni lokal sınırların hassas biçimde belirlenmesinde görev alabilmektedir. Aynı şekilde ViT ile çıkarılan temsiller, geleneksel sınıflandırıcılarla birleştirilerek düşük örneklem sayısına sahip veri kümelerinde daha kararlı performanslar sergileyebilmektedir (31).

GNN Temelli Hibritler

Graph Neural Network (GNN) temelli hibrit modeller, yapay öğrenme sistemlerinde özellikle yapısal verilerin analizi açısından öne çıkan bütünsel yaklaşımardan biridir. GNN'ler, verileri düğüm (node) ve kenar (edge) bileşenleri aracılığıyla grafik yapısı içerisinde temsil ederek, yalnızca bireysel özniteliklerin değil, aynı zamanda bu öznitelikler arasındaki ilişkisel örüntülerin de öğrenilmesini mümkün kılmaktadır. GNN tabanlı hibrit modellerde genel yaklaşım, grafik biçiminde yapılandırılan verilerden anlamlı temsil vektörlerinin elde edilmesi ve bu temsillerin diğer model bileşenleri ile ardışık ya da paralel biçimde işlenmesidir. Bu modeller, grafik verilerden çıkarılan temsilleri geleneksel makine öğrenmesi algoritmalarına (örneğin SVM, Random Forest, XGBoost) ya da derin öğrenme bileşenlerine aktararak çok düzeyli karar mekanizmaları kurmaktadır (32).

Uygulamalı örnekler incelendiğinde, GCN (Graph Convolutional Network) + CNN mimarilerinin tıbbi görüntüleme ve moleküler yapı analizinde, GAT (Graph Attention Network) + Transformer yapılandırmalarının sosyal ağ analizleri ve metinsel verilerde, GNN + XGBoost gibi geleneksel makine öğrenmesi algoritmalarıyla oluşturulan birleşik sistemlerin ise klinik karar destek ve öneri sistemleri gibi alanlarda etkin biçimde kullanıldığı görülmektedir. Bu sistemler, GNN'lerin sunduğu yapısal

temsil gücünü diğer modellerin karar üretim yetenekleriyle birleştirerek daha yüksek doğruluk, açıklanabilirlik ve kararlılık düzeyleri sunabilmektedir. GNN tabanlı hibrit yapılar, özellikle çok kaynaklı, heterojen ve ilişkisel bağlam içeren veri setlerinin analizinde tercih edilmekte ve giderek artan biçimde yapay öğrenme uygulamalarında yer bulmaktadır (33).

RNN / LSTM Temelli Hibritler

RNN ve LSTM ağları, sıralı ve zamana bağımlı verilerin modellenmesinde öne çıkan derin öğrenme yaklaşımlarıdır. Bu modeller, zaman içindeki ardışık ilişkileri dikkate alarak öğrenme gerçekleştirebilmekte; bu özellikleri sayesinde doğal dil işleme, konuşma tanıma, biyomedikal sinyal işleme ve finansal zaman serisi analizi gibi pek çok alanda başarıyla uygulanmaktadır. RNN/LSTM tabanlı hibrit mimariler ise, bu ağların sıralı veri işleme kapasitesini diğer öğrenme mimarilerinin güçlü yönleriyle bütünleştirerek daha kapsamlı ve bağlama duyarlı yapay öğrenme sistemlerinin geliştirilmesini amaçlamaktadır. Bu tür hibrit modellerde RNN veya LSTM bileşeni, zaman bağımlı örüntülerin yakalanması görevini üstlenmekte; diğer bileşenler (örneğin CNN, ViT, GNN veya destek vektör makineleri gibi geleneksel makine öğrenmesi algoritmaları) ise verinin mekânsal, yapısal veya öznelik düzeyinde analizini gerçekleştirmektedir. Model mimarisi, sıralı verilerden elde edilen temsillerin farklı karar bileşenlerine aktarılması ya da paralel olarak işlenen farklı veri türlerinden gelen temsillerin birleştirilerek bütüncül bir karar mekanizması oluşturulması esasına dayanmaktadır (34).

RNN/LSTM temelli hibrit modellerin sunduğu başlıca avantajlar arasında, uzun süreli bağımlılıkların etkin biçimde modellenebilmesi, sıralı verinin bağlamsal temsillerinin çıkarılabilmesi ve farklı öğrenme yaklaşımlarının sinerjik şekilde entegre edilebilmesi yer almaktadır. Özellikle CNN tabanlı mimarilerle birlikte kullanıldığında, mekânsal örüntüler ile zamansal ilişkiler aynı anda öğrenilebilmekte; bu da modelin genel başarımını ve genellenebilirliğini artırmaktadır. Buna ek olarak, LSTM'nin hücre durumu ve unutma kapısı gibi mekanizmaları, bilgi aktarımının daha kontrollü gerçekleştirilmesine olanak tanımakta ve eğitim sürecinde ortaya çıkabilecek gradyan sönümlenmesi (vanishing gradient) gibi sorunların önüne geçmektedir (35).

Hiyerarşik Hibritler

Hiyerarşik hibrit modeller, yapay zekâ ve makine öğrenmesi alanında farklı algoritmaların katmanlı ve çok düzeyli bir yapıda entegre edilerek organize edildiği gelişmiş modelleme yaklaşımlarıdır. Bu modellerde, her bir katman belirli bir alt görevi gerçekleştirmek üzere özelleştirilmiş olup, elde edilen çıktılar bir üst düzey bileşene aktarılır. Böylece sistem, görevleri hiyerarşik bir sırayla ele alarak daha derinlemesine analiz ve karar verme süreçleri gerçekleştirebilir. Bu yapılar, özellikle çok aşamalı karar süreçlerinin gerektiği durumlarda yüksek doğruluk, esneklik ve yorumlanabilirlik sunmaktadır (36).

Hiyerarşik hibrit modeller, hem görev temelli hem de veri temsili düzeyinde organize edilebilir. Görev temelli yapılandırmalarda alt katmanlar veri ön işleme, öznitelik seçimi veya ara sınıflandırmalar gibi işlemleri üstlenirken, üst katmanlar daha kapsamlı analiz ve nihai karar mekanizmalarını yürütür. Veri temsiline dayalı hiyerarşik yapılar ise, veriyi basitten karmaşığa doğru işleyerek farklı soyutlama düzeylerinde bilgi çıkarımı gerçekleştirir. Bu bağlamda, sistem içerisinde kullanılan algoritmalar görevlerin gereksinimlerine göre seçilir. Alt katmanlarda genellikle düşük maliyetli ve hızlı çalışan yöntemler (örneğin karar ağaçları, kural tabanlı sistemler) tercih edilirken, üst katmanlarda daha karmaşık ve bağlamsal öğrenme yeteneği yüksek olan yöntemler (örneğin derin öğrenme modelleri, CNN, Transformer) kullanılır (37).

Bu yapı sayesinde, sistemin hem verimliliği hem de genellenebilirliği artırılabilir. Ayrıca modüler yapısı sayesinde her katman ayrı ayrı optimize edilebilir, yeniden eğitilebilir veya uygulamaya göre özelleştirilebilir. Modelin çok düzeyli ve görev ayrımlı oluşu, hata ayıklama ve sistem güncellemeleri gibi süreçleri de kolaylaştırır. Hiyerarşik hibrit modellerin başlıca avantajları arasında ölçeklenebilirlik, modülerlik, görev uyarlılığı ve verimli kaynak kullanımı yer almaktadır. Örneğin, daha basit katmanlar aracılığıyla ön değerlendirme yapılarak, yalnızca karmaşık örneklerin üst katmanlara yönlendirilmesi sağlanabilir. Böylece işlem yükü azaltılırken, yüksek doğruluk elde edilmesi de mümkün hâle gelir (38).

2.7.3. Uygulama Alanlarına Göre

Görüntü Sınıflandırma: DeiT (Data-efficient Image Transformer)

DeiT (Data-efficient Image Transformer), temelinde ViT mimarisini barındıran ancak eğitim verimliliğini artırmak üzere CNN tabanlı bilgi damıtımı (distillation) stratejisiyle desteklenen bir hibrit modeldir. Model, ViT'nin öğretici olarak CNN'den

öğrendiği bir damıtma süreci ile eğitilir. Böylece ViT'nin geniş veri gereksinimi azaltılırken, CNN'in yerel örüntü tanıma kapasitesi modele entegre edilmiş olur. DeiT, veri açısından daha verimli çalışarak daha küçük veri setlerinde dahi yüksek sınıflandırma başarımı sağlar. Ayrıca, eğitimin daha hızlı gerçekleşmesine ve modelin hata oranlarının düşürülmesine katkı sunar.

Bu yapı özellikle genel görüntü sınıflandırma görevlerinde kullanılmak üzere tasarlanmıştır. Büyük hacimli veri setlerine erişimin sınırlı olduğu durumlarda veya düşük kaynaklı ortamlarda ViT mimarisinin performansını artırmak amacıyla tercih edilmektedir. ImageNet gibi büyük ölçekli benchmark veri kümeleri üzerinde rekabetçi sonuçlar elde etmektedir (39).

Nesne Tespiti: DETR (DEtection TRansformer)

DETR (Detection Transformer), nesne tespiti görevleri için geliştirilen ilk tam transformer tabanlı model olma özelliği taşımaktadır. Yapısında bir CNN tabanlı omurga (örneğin ResNet) kullanılarak görüntüden öznitelikler çıkarılır. Bu öznitelikler, transformer mimarisine aktarılır ve burada encoder-decoder yapısı ile hem nesne sınıflandırması hem de konumlandırması (bounding box prediction) eşzamanlı olarak öğrenilir. DETR'nin yenilikçi yönü, nesne tespiti için geleneksel anchor box tabanlı yaklaşımları ortadan kaldırması ve end-to-end öğrenme yapısına olanak tanımasıdır.

Modelin temel avantajı, karmaşık ön işleme veya elle belirlenen bölge önerileri gerektirmemesi, dolayısıyla sistem tasarımını sadeleştirmesidir. Ayrıca dikkat (attention) mekanizmaları sayesinde uzun menzilli bağlam ilişkilerini etkili biçimde modelleyebilir. DETR, özellikle otonom sürüş sistemleri, güvenlik kameraları, trafik sahneleri analizi ve endüstriyel kalite kontrol sistemleri gibi çok nesneli, yüksek çözünürlüklü görüntülerde başarılı sonuçlar vermektedir (40).

Görüntü Segmentasyonu: TransUNet

TransUNet, özellikle tıbbi görüntü bölütleme gibi yüksek hassasiyet gerektiren alanlar için tasarlanmış bir ViT-CNN hibrit modelidir. Modelin temel mimarisi, U-Net yapısına benzer şekilde CNN tabanlı bir encoder ile başlamakta; ardından ViT ile özniteliklerin küresel düzeyde yeniden kodlanması sağlanmaktadır. Son olarak decoder katmanları aracılığıyla bölütleme haritası oluşturulur. Bu yapı, bölütleme sürecinde hem yerel detayları (kenarlar, şekiller) hem de küresel bağlam ilişkilerini aynı anda modelleyerek, daha isabetli sonuçlar üretir. TransUNet'in en önemli avantajı, düşük veri

miktarı ile eğitildiğinde bile güçlü bağlamsal ilişki kurabilmesi ve klinik doğruluğa sahip çıktılar verebilmesidir. TransUNet, özellikle manyetik rezonans (MR) ve bilgisayarlı tomografi (BT) gibi tıbbi görüntülerde organ, tümör, damar ve doku bölütleme gibi görevlerde kullanılmaktadır. Klinik karar destek sistemleri, hastalık takibi ve ameliyat planlama gibi kritik alanlarda tercih edilmektedir (41).

Hafif Yapılar (Lightweight Modeller): Mobile-Former

Mobile-Former, kaynak kısıtlı ortamlarda (örneğin mobil cihazlar ve gömülü sistemler) çalıştırılmak üzere geliştirilen hafif yapıya sahip bir ViT-CNN hibrit mimaridir. Modelde, yerel örüntüleri hızlı biçimde öğrenmek için MobileNet benzeri CNN katmanları kullanılmakta; buna ek olarak, küresel bilgi çıkarımı için dikkat temelli hafif transformer blokları entegre edilmektedir. CNN ve transformer bileşenleri bu modelde paralel biçimde çalışmakta, aralarındaki bilgi alışverişi ise özel bağlantı köprüleri (bridges) ile sağlanmaktadır. Bu modelin en büyük avantajı, çok düşük parametre sayısı ve işlem yükü ile gerçek zamanlı görevleri gerçekleştirebilmesidir. Aynı zamanda, hem yerel hem küresel düzeyde anlamlı özellik çıkarımı yapılmasını sağlayarak doğruluk-performans dengesini korur. Mobile-Former, mobil cihazlarda artırılmış gerçeklik uygulamaları, akıllı kamera sistemleri, düşük enerji tüketimli nesne tanıma, gerçek zamanlı yüz tanıma ve el hareketi izleme gibi uygulama alanlarında etkili biçimde kullanılmaktadır (42).

2.7.4. Eğitim Stratejilerine Göre

Transfer Öğrenme (Transfer Learning)

Transfer öğrenme, özellikle büyük hacimli veri setlerine erişimin sınırlı olduğu senaryolarda yaygın olarak kullanılan bir yaklaşımdır. Bu stratejide, genellikle bir CNN tabanlı omurga (backbone) model, geniş kapsamlı bir kaynak veri kümesi üzerinde önceden eğitilir. Elde edilen öznitelik temsilleri daha sonra ViT bileşenine aktarılır ve model, hedef göreve yönelik daha küçük bir veri kümesi ile ince ayar (fine-tuning) sürecine tabi tutulur. Bu yöntem, CNN'in düşük seviyeli öznitelik çıkarımı konusundaki başarısını ViT'nin küresel bağlam öğrenme kapasitesi ile birleştirmeyi amaçlar.

Transfer öğrenmenin başlıca avantajı, eğitim sürecini hızlandırması ve sınırlı veri ile yüksek doğruluk elde edilmesine olanak tanımasıdır. Özellikle tıbbi görüntü analizi, uzaktan algılama ve biyoinformatik gibi veri etiketlemenin maliyetli olduğu alanlarda tercih edilmektedir (43).

Çok Görevli Öğrenme (Multi-task Learning)

Çok görevli öğrenme, bir hibrit modelin birden fazla bağımsız ya da ilişkili görevi eşzamanlı olarak öğrenmesini sağlayan bir stratejidir. Bu yaklaşım, öğrenme sürecinin daha verimli olmasını, bilgi paylaşımını ve genel model performansının artmasını hedefler. Hibrit yapılarda çok görevli öğrenme sayesinde, örneğin bir CNN bileşeniyle özellik çıkarılırken, bir ViT veya transformer bileşeniyle farklı öğrenme görevleri (sınıflandırma, konum tahmini, öznitelik analizleri vb.) şekillendirilebilir. Bu strateji, veri verimliliğini artırmakla kalmaz; aynı zamanda farklı görevlerin birbirini desteklemesini sağlar. Özellikle medikal görüntüleme gibi alanlarda, hem tümör saptama hem de organ bölütleme gibi görevlerin aynı modelde çözülmesi bu stratejiyle mümkün olur (44).

Kendiliğinden Denetimli Öğrenme (Self-supervised Learning)

Kendiliğinden denetimli öğrenme, çok az ya da hiç etiketli veri kullanmadan, verinin kendi yapısından faydalanarak modelin öğrenmesini sağlayan bir yaklaşımdır. Hibrit sistemlerde bu strateji genellikle ön eğitim için kullanılır. Örneğin, bir ViT bileşeni üzerinde uygulanacak "patch tahmini" veya "jigsaw puzzle" tipi ödevler, modelin görsel temsilleri daha iyi öğrenmesini sağlar. Daha sonra, bu ön eğitilmiş model bir CNN yapısıyla birleştirilerek hedef görev (sınıflandırma, bölütleme vb.) için ince ayar (fine-tuning) yapılabilir. Bu strateji, etiketleme maliyeti yüksek alanlarda (tıbbi görüntüler, biyoinformatik) önemli avantajlar sağlar. Ayrıca transfer öğrenmeyle birleştirildiğinde çok az veriyle bile yüksek performans elde edilebilir (44).

Bilgi Damıtımı (Knowledge Distillation)

Bilgi damıtımı, büyük ve hesaplama açısından yoğun bir öğretici modelden (genellikle önceden eğitilmiş bir ViT) daha küçük ve verimli bir öğrenci modele (örneğin ViT-CNN hibriti) temsil bilgisi ve karar stratejisi aktarımı sürecini ifade eder. Öğrenci model, yalnızca gerçek etiketler üzerinden değil, aynı zamanda öğretici modelin çıktıları ve ara temsilleri doğrultusunda eğitilir. Bu strateji, model sıkıştırma (model compression) ve donanım kısıtlı ortamlara uygun hale getirme açısından önemli avantajlar sunar.

Damıtma yöntemiyle eğitilen hibrit modeller, daha düşük parametre sayısına sahip olmalarına rağmen öğretici modellerin performansına yakın sonuçlar elde edebilir. Bu yaklaşım, özellikle mobil uygulamalar ve gömülü sistemler için tasarlanmış hafif modellerin geliştirilmesinde kullanılmaktadır. Örnek olarak, DeiT (Data-efficient Image Transformer) modeli bu stratejinin başarılı bir uygulamasıdır (45).

Uçtan Uca Eğitim (End-to-End Training)

Uçtan uca eğitim, hibrit modelin tüm bileşenlerinin (örneğin CNN ve ViT) aynı anda ve ortak bir kayıp fonksiyonu aracılığıyla optimize edildiği eğitim stratejisidir. Bu yöntem sayesinde modelin tüm katmanları eş zamanlı olarak güncellenmekte; bu da CNN ve ViT bileşenleri arasında temsil düzeyinde derin uyum ve bilgi etkileşimi sağlanmasına olanak tanımaktadır.

Bu strateji, özellikle çok katmanlı ve karmaşık görevleri (örneğin nesne tespiti, çok görevli öğrenme ve bölütleme) içeren senaryolarda tercih edilmektedir. Bununla birlikte, uçtan uca eğitim genellikle daha yüksek hesaplama gücü gerektirir ve optimizasyon süreci, bileşenler arası etkileşim nedeniyle daha karmaşık bir hâl alabilir. DETR (DEtection TRansformer) ve TransUNet gibi hibrit modellerde bu strateji başarıyla uygulanmaktadır (40).

Tablo 2.1. Hibrit Modeller Karşılaştırma Tablosu

Hibrit Model	Yapısal Özellik	Avantajları	Sınırlılıkları	Kullanım Alanları
Sıralı Hibrit	Model çıktıları ardışık şekilde aktarılır	Modülerlik, özelleştirilebilirlik, yüksek doğruluk	Hata yayılımı, karmaşık yapı, yüksek kaynak ihtiyacı	Tıbbi görüntüleme, NLP, eğitim analitiği
Paralel Hibrit	Model çıktıları eşzamanlı üretilir ve birleştirilir	Model çeşitliliği, overfitting azaltımı, karar füzyonu	Yüksek hesaplama maliyeti, çıktı birleştirme zorluğu	Biyomedikal analiz, NLP, görüntü işleme
Dönüşümsel Hibrit	Her bileşen veriyi anlamsal olarak dönüştürür	Derin soyutlama, bağlamsal temsil, çok yönlülük	Yorumlanabilirlik düşüşü, bilgi kaybı, kaynak ihtiyacı	Genomik analiz, NLP, öğrenme analitiği
CNN Temelli	CNN öznitelik çıkarımı, ML veya DL karar bileşeni	Yüksek özellik çıkarımı, farklı karar sistemleriyle entegrasyon	Hesaplama yoğunluğu, karmaşık yapı, veri gereksinimi	Tümör analizi, tarımsal hastalık tespiti
ViT Temelli	ViT öznitelik çıkarımı, CNN veya ML karar bileşeni	Küresel bağlam öğrenimi, düşük veriyle güçlü performans	Yüksek veri ve kaynak gereksinimi, yorumlanabilirlik sorunu	Tıbbi görüntüleme, kalite kontrol, savunma
GNN Temelli	GNN veriyi grafik olarak işler, diğerleriyle birleşir	İlişkisel öğrenme, heterojen veri analizi	Graf dönüşümü zorluğu, entegrasyon karmaşıklığı	Moleküler analiz, sosyal ağlar, öneri sistemleri
RNN/LSTM Temelli	Zamansal veriler LSTM ile, diğer öznitelikler ayrı analiz edilir	Zamansal örüntü öğrenimi, bağlamsal analiz	Uzun dizilerde hesaplama maliyeti, hiperparametre zorluğu	Eğitim analitiği, finansal tahmin, sinyal işleme
Hiyerarşik Hibrit	Katmanlı yapı, her katman farklı görev üstlenir	Çok düzeyli görev çözümü, görev ayrımı, esneklik	Tasarım karmaşıklığı, alt katman hatalarının etkisi	Tıbbi tanı, eğitim verisi analizi, endüstriyel otomasyon

2.8. Cnn, Vit ve Hibrit Mimariler

CNN'ler bilgisayarlı görü alanında temel yapı taşlarından biri olmuştur. Özellikle görüntü sınıflandırma, nesne tanıma ve segmentasyon gibi görevlerdeki başarısı, CNN'lerin yerel örüntüleri etkili bir şekilde yakalayabilme yeteneğinden kaynaklanmaktadır (11). Ancak son yıllarda, Görüntü Dönüştürücüler ViT adı verilen ve öz-dikkat (self-attention) mekanizmasına dayanan modellerin ortaya çıkışı, bu alandaki dinamikleri önemli ölçüde değiştirmiştir. ViT'ler, görüntüleri sabit boyutlu yama (patch) dizileri olarak işler ve küresel (global) ilişkileri doğrudan modelleyerek daha yüksek temsil gücüne ulaşabilir. Bu sayede daha karmaşık ve büyük ölçekli veri kümelerinde başarıyla genelleme yapabilirler (8).

Günümüzdeki en başarılı ViT mimarilerinin büyük bir kısmı, aslında CNN ve Transformer yapılarının bir hibriti olarak geliştirilmektedir. Bu hibrit mimariler, her iki yaklaşımın güçlü yönlerini bir araya getirerek hem verimliliği hem de doğruluğu artırmayı hedeflemektedir. Bu yapıların başarısında iki temel tasarım yaklaşımı öne çıkmaktadır:

CNN'den esinlenen hiyerarşik mimari: Bu tür yapılarda ağ derinleştikçe özellik boyutları (feature dimensions) artarken, mekânsal boyutlar küçülür. Böylece piramitsel bir özellik haritası elde edilir. Bu mimari, yerel alıcı alanlarla birleştirildiğinde hem modelin hesaplama karmaşıklığını azaltır hem de yerel örüntüleri yakalama başarısını korur (11).

Transformer'dan gelen öz-dikkat mekanizması: Küresel bağıntıları modelleyebilen bu yapı, modelin ifade gücünü önemli ölçüde artırır. Böylece daha iyi genelleme ve daha yüksek doğruluklar elde edilmesine katkı sağlar (33).

ViT'lerin tümüyle CNN'lerin yerini aldığına dair genel bir kanıya varmak şu an için mümkün değildir. Uygulamanın niteliği, modelin eğitileceği veri setinin büyüklüğü, kullanılacak donanım ve performans beklentileri gibi kriterler, hangi modelin tercih edilmesi gerektiğini belirlemede belirleyici olacaktır.

Sonuç olarak, hem CNN hem de Transformer tabanlı yaklaşımların güçlü yanlarını bir araya getiren hibrit mimariler, bilgisayarlı görü alanında geleceğin öncelikli çözümleri arasında yer almaya aday görünmektedir.

3. MATERYAL VE METOT

Bu bölümde, CNN, ViT ve Hybrid mimarilerin yapısı, mimariyi değerlendirmek amacıyla seçilen tıbbi görüntü veri seti, uygulanan veri ön işleme ve artırma teknikleri, eğitim protokolleri ve performans değerlendirme metrikleri kapsamlı olarak açıklanmaktadır. Ayrıca, model kararlarının şeffaflığını sağlamak için entegre edilen XAI yöntemlerine yer verilmektedir.

3.1. Troid Tümörü Veri Seti

Tiroid nodülü ultrason görüntülerinden oluşan kamuya açık bir veri seti oluşturuldu. Sınıflandırma amacıyla hazırlanan bu veri setinde toplam 6.905 görüntü yer aldı. 3.331'i benign (iyi huylu) ve 3.574'ü malignant (kötü huylu) olarak etiketlendi. Görüntüler JPEG ve PNG resim formatlarında saklandı. Farklı klinik merkezlerden ve çeşitli ultrason cihazlarından toplanan veriler, gerçek klinik uygulamalardaki çeşitliliği yansıtan heterojen bir yapı gösterdi. Bu özelliğiyle veri seti, geliştirilen modellerin genelleme kabiliyetini artırdı ve klinik pratiğe daha uygun sonuçların üretilmesine katkı sağladı (46, 47).

3.2. Veri Ön İşleme

Veri setinde bulunan 6.905 adet tiroid ultrason görüntüsü, farklı çözünürlüklerde olup 224x224 piksel boyutuna yeniden ölçeklendirildi. Kullanılan derin öğrenme mimarilerinin (CNN, ViT ve Hybrid) ImageNet veri seti üzerinde ön eğitilmiş (pre-trained) modeller olması nedeniyle, normalizasyon işlemi için ImageNet standardı tercih edildi. Bu kapsamda normalizasyon parametreleri ortalama [0.485, 0.456, 0.406] ve standart sapma [0.229, 0.224, 0.225] değerleriyle uygulandı. Görüntülerin renk uzayı zaten gri tonlamalı (grayscale) olduğundan, RGB'den gri tonlamaya ek bir dönüşüm işlemi yapılmadı.

3.3. Veri Artırma

Derin öğrenme modellerinde aşırı uyumu (overfitting) önlemek ve modelin genelleme kabiliyetini artırmak amacıyla veri artırma teknikleri uygulandı. Ultrason görüntülerinin anatomik anlamlılığını koruyan ve prob hareketlerinin yaratabileceği gerçekçi varyasyonları taklit eden aşağıdaki teknikler, belirtilen parametrelerle kullanıldı.

- Yatay Çevirme: Görüntülerin yatay ekseninde çevrilmesi
- Küçük Açılı Döndürme: Görüntülerin $\pm 10^\circ$ aralığında rastgele döndürülmesi
- Parlaklık Değişimi: Görüntü parlaklığının $\pm 30\%$ oranında değiştirilmesi
- Kontrast Değişimi: Görüntü kontrastının $\pm 30\%$ oranında ayarlanması
- Rastgele Kırpma: Orijinal görüntünün $80-100\%$ aralığında rastgele kırpılarak yeniden boyutlandırılması
- Gürültü Ekleme: Görüntülere Gauss gürültüsü eklenmesi

Tüm artırma işlemleri eğitim sırasında gerçek zamanlı (on-the-fly) ve rastgele olarak uygulandı. Böylece modelin her epoch'ta farklı varyasyonlar görmesi sağlandı. Bu yöntemlerle veri setinin etkin boyutu ve çeşitliliği artırıldı, modelin sağlamlığı (robustness) geliştirildi ve aşırı uyum riski minimize edildi.

3.4. Veri Bölütleme

Toplam 6.905 ultrason görüntüsünden oluşan veri seti, model eğitimi ve değerlendirme sürecinin sağlıklı yürütülebilmesi için üç ayrı kümeye ayrıldı. Veri dağılımı, 70% eğitim, 15% doğrulama ve 15% test olacak şekilde, 42 random seed değeri kullanılarak rastgele bölütleme yöntemiyle gerçekleştirildi. Bu sayede sonuçların tekrarlanabilirliği garanti edildi.

Bölütleme sonucunda;

- Eğitim seti: 4833 örnek
- Doğrulama seti: 1036 örnek
- Test seti: 1036 örnek

Eğitim seti, modelin öğrenme süreci için, doğrulama seti hiperparametre ayarlaması ve erken durdurma için, test seti ise modelin nihai performans değerlendirmesi için ayrıldı. Tüm veri kümeleri orijinal dağılımlarıyla korundu ve modelin gerçek dünya verileri üzerindeki performansının objektif bir şekilde ölçülebilmesi amaçlandı.

3.5. Modelleme

Bu çalışmada, derin öğrenme tabanlı sınıflandırma performansını kapsamlı bir şekilde değerlendirmek amacıyla, çeşitli state-of-the-art mimarilerden oluşan geniş kapsamlı bir model seti özellikle seçildi. Seçilen bu modeller, CNN mimarilerini, ViT

tabanlı yenilikçi yaklaşımları ve her iki tekniğin avantajlarını birleştiren Hybrid mimarileri kapsayacak şekilde çeşitlendirildi ve genişletildi.

3.6. Model Mimarileri

3.6.1. CNN Tabanlı Mimariler

ResNet-50: ResNet-50, derin sinir ağlarının eğitimi sırasında karşılaşılan gradyan kaybolması sorununa bir çözüm olarak geliştirilmiş bir evrişimli sinir ağı mimarisidir. He ve ark. (48) tarafından önerilen bu model, ağ içerisine özdeşlik kısayol bağlantıları ekleyerek artık öğrenme (residual learning) prensibini uygular. Bu bağlantılar, bir katman bloğunun girdisinin, bloğun çıktısına doğrudan eklenmesini sağlar. Bu sayede, ağırlık katmanlarının öğrenmesi gereken fonksiyon basitleşir ve gradyanların ağ boyunca daha etkili bir şekilde akması mümkün olur (48).

Mimari, "bottleneck" tasarımına sahip artık bloklardan oluşmaktadır. Her bir blok, 1x1, 3x3 ve tekrar 1x1 boyutlarında üç evrişimsel katmanı içerir. İlk 1x1 katmanı, öznitelik haritasının boyutunu düşürerek hesaplama yükünü azaltır, 3x3 katmanı uzamsal bilgiyi işler ve son 1x1 katmanı ise boyutu tekrar eski haline getirir. Bu yapı, 50 katman derinliğindeki bir modelin verimli bir şekilde eğitilmesine olanak tanır.

ResNet-50, sunduğu bu çözümle derin modellerin eğitiminde önemli bir ilerleme sağlamıştır. Yüksek sınıflandırma doğruluğunun yanı sıra, nesne tespiti ve anlamsal bölütleme gibi görevler için güçlü bir özellik çıkarıcı olarak kullanılmış ve sonraki birçok mimariye temel oluşturmuştur.

EfficientNet-B4: EfficientNet-B4, model ölçeklendirme yaklaşımını yeniden ele alan EfficientNet model ailesinin bir üyesidir. Tan ve Le (49) tarafından önerilen bu çalışma, bir modelin derinlik, genişlik ve çözünürlük boyutlarının izole şekilde değil, dengeli bir şekilde ölçeklendirilmesi gerektiği fikrine dayanır. Bileşik ölçeklendirme (compound scaling) adı verilen bu yöntem, her bir boyut için sabit bir ölçeklendirme katsayıları seti tanımlar.

Mimarinin geliştirilmesi iki aşamada gerçekleşmiştir. İlk aşamada, sinir ağı mimarisi arama (NAS) tekniği kullanılarak hafif ve temel bir ağ olan EfficientNet-B0 elde edilmiştir. İkinci aşamada, bu temel ağ, bileşik ölçeklendirme yöntemiyle $\phi=1.6$ katsayısı kullanılarak EfficientNet-B4 modeline genişletilmiştir. Bu yöntem, model kapasitesinin geleneksel yollara kıyasla daha verimli artırılmasını sağlamaktadır.

EfficientNet-B4, parametre verimliliği ile sınıflandırma doğruluğu arasında dengeli bir performans sunar. Daha küçük EfficientNet modellerine kıyasla daha yüksek doğruluk sağlarken, daha büyük modellere göre hesaplama maliyeti açısından daha uygundur. Bu özellikleriyle pratik uygulamalarda sıklıkla tercih edilen bir model haline gelmiştir (49).

ConvNeXt: ConvNeXt, modernize edilmiş bir evrimsel sinir ağı mimarisidir. Liu ve ark. (50) tarafından geliştirilen bu model, ViT'lerin tasarım ilkelerinden esinlenerek saf bir evrimsel mimarinin bu tür modellerle benzer performans sergileyebileceğini göstermiştir. Çalışma, standart bir ResNet mimarisinin adım adım iyileştirilmesiyle ConvNeXt'in elde edildiği sistematik bir yaklaşım sunar.

İyileştirme adımları arasında, yamuk çıkarımı için daha büyük evrim çekirdekleri kullanılması, aktivasyon fonksiyonu sayısının azaltılması, Batch Normalization yerine Layer Normalization'a geçilmesi ve derinlik-yönlü evrimlerin kullanılması yer almaktadır. Bu değişiklikler, modelin ViT benzeri bir özellik öğrenme yeteneği kazanmasını sağlarken mimarinin tamamen evrimsel kalmasına olanak tanır.

ConvNeXt, evrimsel mimarilerin doğru tasarım seçimleriyle güncel görevlerde etkili bir şekilde kullanılabileceğini göstermiştir. Hem sınıflandırma hem de yoğun tahmin görevlerinde iyi performans sergileyen bu model, mimari seçeneklerini genişleten bir katkı sağlamıştır (50).

3.6.2. Vision Transformer Tabanlı Mimariler

Swin Transformer: Swin Transformer, hiyerarşik bir yapıya sahip olacak şekilde tasarlanmış bir ViT modelidir. Liu ve ark. (51) tarafından önerilen bu mimari, öz-dikkat mekanizmasının hesaplama karmaşıklığını azaltmak için kaydırılmış pencereler (shifted windows) yöntemini kullanır. Bu yaklaşım, modelin yüksek çözünürlüklü görüntülerle ve yoğun tahmin görevleriyle daha verimli çalışmasını sağlar.

Mimari, öz-dikkat hesaplamalarını örtüşmeyen yerel pencereler içinde yapar. Ardışık katmanlarda bu pencerelerin konumu kaydırılarak, farklı pencereler arasında bilgi akışı sağlanır. Aynı zamanda, öznitelik haritalarının çözünürlüğü kademeli olarak düşürülürken kanal derinliği artar. Bu hiyerarşik yapı, geleneksel evrimli sinir ağlarının piramit yapısına benzerlik gösterir.

Swin Transformer, bu özellikleriyle sadece sınıflandırma değil, aynı zamanda nesne tespiti ve anlamsal bölütleme gibi görevlerde de başarılı sonuçlar vermiştir.

Hesaplama verimliliği ve hiyerarşik tasarımı, onu görüntü işleme alanında yaygın olarak kullanılan bir model haline getirmiştir (51).

DeiT: DeiT (Data-efficient Image Transformer), ViT modellerinin büyük ölçekli veri kümelerine olan bağımlılığını azaltmayı amaçlayan bir modeldir. Touvron ve ark. (39) tarafından önerilen bu çalışma, bir öğretmen-öğrenci damıtma stratejisi kullanarak ViT modellerinin daha küçük veri kümeleri üzerinde etkili bir şekilde eğitilmesini mümkün kılar.

Mimarinin temelinde, sınıf token'ine ek olarak bir damıtma token'i (distillation token) kullanılması fikri yatar. Damıtma token'i, bir öğretmen modelin (genellikle bir CNN) çıktılarını öğrenmekten sorumludur. Her iki token de öz-dikkat katmanları boyunca birlikte çalışarak modelin hem gerçek etiketlerden hem de öğretmen modelin yumuşak etiketlerinden aynı anda öğrenmesini sağlar.

DeiT, bu yaklaşımla ViT modellerinin ImageNet-1k gibi standart veri kümeleri üzerinde ek veriye ihtiyaç duyulmadan yüksek doğrulukla eğitilebileceğini göstermiştir. Bu durum, ViT modellerinin uygulanabilirliğini ve erişilebilirliğini önemli ölçüde artırmıştır (39).

PVTv2: PVTv2 (Pyramid Vision Transformer v2), özellikle yoğun tahmin görevleri için tasarlanmış hiyerarşik bir Transformer mimarisidir. Wang ve ark. (52) tarafından önerilen bu model, orijinal PVT'nin sınırlamalarını ele almak üzere geliştirilmiştir. Temel katkıları, örtüşen yamuk gömme, evrimsel ileri besleme ağları ve doğrusal karmaşıklığa sahip dikkat mekanizmalarıdır.

Örtüşen yamuk gömme yöntemi, komşu yamuklar arasındaki yerel sürekliliği koruyarak daha zengin öznitelik gösterimi sağlar. Evrimsel ileri besleme ağları ise derinlik-yönlü evrimler içererek yerel bağlamsal bilginin modellenmesini iyileştirir. Bu iyileştirmeler, modelin hem sınıflandırma hem de yoğun tahmin görevlerindeki performansını artırmıştır.

PVTv2, bu özellikleriyle ölçeklenebilir ve verimli bir ViT mimarisi sunar. Hiyerarşik yapısı ve geliştirilmiş tasarımı, onu çeşitli görüntü işleme görevleri için uygun bir seçenek haline getirmiştir (52).

XCiT: XCiT (Cross-covariance Image Transformer), geleneksel öz-dikkat mekanizmasının hesaplama karmaşıklığı ve bellek talebi sorunlarını çözmek amacıyla geliştirilmiş bir dönüştürücü mimarisidir. Ali ve ark. (53) tarafından önerilen bu model,

dikkat (attention) hesaplamasını uzamsal boyutlar (token'lar) yerine özellik kanalı boyutu üzerinden yeniden tanımlayarak devrim niteliğinde bir yenilik sunar: çapraz-kovaryans dikkat mekanizması.

Bu çapraz-kovaryans dikkat mekanizması, anahtar (key) ve değer (value) matrislerini önce kanal boyutunda projeler ve dikkati token'lar arasında değil, kanal grupları arasındaki kovaryans üzerinden hesaplar. Bu, hesaplama karmaşıklığını token sayısından (N) kanal boyutuna (D) kaydırarak, yüksek çözünürlüklü görüntüleri işlerken son derece verimli hale gelmesini sağlar. Ayrıca, yerel olarak duyarlı özellikleri korumak için bu global attention katmanının üzerine bir Yerel İletişim Bloğu (Local Patch Interaction Block) ekler.

XCiT, bu özellikleriyle ViT modellerinin ölçeklenebilirliğini ve verimliliğini büyük ölçüde artıran bir yaklaşım sunmuştur. Bu durum, özellikle yüksek çözünürlüklü görüntü ve video işleme gibi senaryolarda model seçeneklerini genişletmiştir (53).

3.6.3. Hibrit (CNN + Transformer) Mimariler

ConViT: ConViT, Vision Transformer (ViT) mimarisinin veri açısından verimsiz olma ve iyi özellikler öğrenmek için büyük miktarda veriye ihtiyaç duyma sorununu çözmek için tasarlanmış melez bir vision mimarisidir. d'Ascario ve diğerleri (54) tarafından önerilen bu model, evrişimli sinir ağlarının yerel özyapısal (locally inductive) ön kabulleri ile Transformer'ların dinamik ve ölçeklenebilir küresel modelleme kapasitesini birleştirmeyi amaçlar. Mimari, ViT'lerin daha az veriyle daha iyi ve daha hızlı öğrenmesini sağlamak için optimize edilmiştir.

Temel bileşeni, Yumuşak Özyapısal Özdeşlik (Soft Inductive Bias) adı verilen ve öğrenilebilir bir mekanizma olan GPSA (Gated Positional Self-Attention) bloklarıdır. Bu bloklar, başlangıçta bir evrişim katmanı gibi davranır (yerel alana odaklanır) ve zamanla eğitim ilerledikçe öğrenerek klasik bir Transformer'ın küresel öz-dikkat (global self-attention) mekanizmasına dönüşür. Bu tasarım, modelin evrişimlerin güçlü ön kabullerinden faydalanarak daha hızlı yakınsamasını ve daha az veriyle genelleme yapabilmesini sağlar.

ConViT, özellikle orta ölçekli veri kümelerinde (örneğin ImageNet-1k) saf ViT modellerinden daha üstün performans sergilemiştir. Bu özellikleriyle, ViT'ler evrişimsel ön kabuller olmadan öğrenmedeki zorluklarını gideren bir model seçeneği sunmaktadır (54).

CoaT: CoaT (Co-Scale Conv-Attentional Transformer), evrişim ve öz-dikkat mekanizmalarını entegre eden bir ViT ailesidir. Xu ve ark. (55) tarafından önerilen bu mimari, çoklu ölçeklerdeki öznitelik haritaları arasında etkileşimi teşvik eden bir tasarım sunar. Temel bileşenleri arasında derinlik-yönlü evrişimler ve eş-ölçekli mekanizma yer almaktadır.

Derinlik-yönlü evrişimler, öz-dikkat katmanlarına entegre edilerek örtülü konumsal kodlama sağlar. Eş-ölçekli mekanizma ise farklı ölçeklerdeki öznitelik haritalarını birbirine bağlayarak çapraz-ölçek etkileşimini mümkün kılar. Bu tasarım, modelin hem yerel detayları hem de küresel bağlamı aynı anda işlemesine olanak tanır.

CoaT, bu özellikleriyle hem sınıflandırma hem de yoğun tahmin görevlerinde iyi performans sergilemiştir. Mimari, evrişimsel ve öz-dikkat mekanizmalarının entegrasyonu için bir çerçeve sunmaktadır (55).

MobileViT: MobileViT, mobil ve kaynak kısıtlı cihazlar için tasarlanmış hafif bir melez vision mimarisidir. Mehta ve Rastegari (56) tarafından önerilen bu model, evrişimli sinir ağlarının yerel işleme yeteneği ile Transformer'ların küresel modelleme kapasitesini birleştirmeyi amaçlar. Mimari, mobil cihazlarda verimli çalışacak şekilde optimize edilmiştir.

Temel bileşenleri arasında mobil birimler ve MobileViT blokları yer almaktadır. MobileViT blokları, hem evrişimsel hem de Transformer benzeri özellikler sergiler. Bu bloklar, yerel olarak bilgiyi işlerken aynı zamanda küresel bağlamı da modelleyebilir. Bu tasarım, parametre verimliliği ve düşük hesaplama maliyeti sağlar.

MobileViT, hafif olmasının yanı sıra çeşitli vision görevlerinde iyi performans sergilemiştir. Bu özellikleriyle mobil dağıtım için uygun bir model seçeneği sunmaktadır (56).

HuViT: HuViT, bu çalışma kapsamında geliştirilen, CNN'lerin güçlü yerel özellik çıkarma yeteneği ile Transformer'ların üstün küresel ilişki modelleme kapasitesini entegre eden özgün bir iki modlu (Hybrid) vision mimarisidir. Bu model, mevcut hibrit yaklaşımlardan farklı olarak önceden eğitilmiş bir CNN backbone ile çok katmanlı Transformer encoder'ını ardışık ve optimize edilmiş bir pipeline şeklinde birleştirmektedir.

Mimari, ResNet50 CNN backbone ve çok katmanlı Transformer encoder temel bileşenleri üzerine inşa edilmiştir. Tasarımın özgün yönü, CNN katmanlarının

görüntünün yerel özelliklerini ve hiyerarşik yapısını çıkarması, ardından Transformer encoder'ın bu özellikler arasındaki küresel bağımlılıkları ve uzamsal ilişkileri modellemesi prensibine dayanmaktadır. Bu iki aşamalı yaklaşım, hem yerel detayların korunmasını hem de global bağlamın anlaşılmasını sağlamaktadır.

HuViT'nin temel avantajları:

- Transfer öğrenme için optimize edilmiş yapı
- Parametre verimliliği (CNN katmanları isteğe bağlı dondurulabilir)
- Ölçeklenebilir mimari (Transformer katman sayısı ve boyutları ayarlanabilir)
- Esnek backbone entegrasyonu (farklı CNN mimarileriyle uyumluluk)

HuViT Mimari Yapısı

HuViT mimarisi dört ana bileşenden oluşan ardışık bir pipeline şeklinde tasarlanmıştır.

CNN Özellik Çıkarıcı Katmanı

Önceden eğitilmiş ResNet50 modelinin son iki katmanı çıkarılarak modifiye edilmiştir. Giriş görüntülerinden hiyerarşik özellik haritaları çıkarır. Çıktı olarak [B, 2048, H, W] boyutunda özellik haritaları üretir. Bu katman isteğe bağlı olarak dondurulabilir veya ince ayar için serbest bırakılabilir.

Özellik Dönüşüm ve Hazırlık Katmanı

Adaptive Average Pooling ile özellik haritaları sabit [7, 7] boyutuna ölçeklendirilir. Özellikler [B, 2048, 49] boyutunda yeniden düzenlenir. Linear projeksiyon katmanı ile CNN özellikleri Transformer boyutuna [512] dönüştürülür.

Transformer Encoder Katmanı

4 katmanlı Transformer encoder bloğu kullanılır. 8 başlı multi-head self-attention mekanizması içerir. Her attention katmanı 512 boyutunda giriş alır ve işler. Positional encoding eklenerek uzamsal bilgi korunur.

Sınıflandırma ve Çıktı Katmanı

Transformer çıktıları üzerinde global average pooling uygulanır. [B, 512] boyutunda sabit uzunluklu vektörler elde edilir. Linear sınıflandırıcı ile son sınıf sayısına projeksiyon yapılır. Softmax aktivasyonu ile son sınıflandırma skorları üretilir.

Teknik Özellikler:

- Giriş boyutu: 224x224x3
- CNN çıktı kanal sayısı: 2048
- Embedding dimension: 512
- Transformer head sayısı: 8
- Transformer katman sayısı: 4
- Toplam parametre sayısı: ~25M

Bu özellikleriyle HuviT, özellikle tıbbi görüntü analizi, nesne tanıma ve yüksek doğruluk gerektiren bilgisayarlı görü görevlerinde etkili bir çözüm sunmaktadır.

Ancak, önemle belirtmek gerekir ki HuviT'nin performansı doğrudan kullanılan veri setinin özelliklerine, boyutuna ve kalitesine bağlıdır. Model her veri setinde aynı performansı göstermeyebilir. Performansı etkileyen temel faktörler arasında veri seti büyüklüğü ve çeşitliliği, veri kalitesi ve ön işleme süreçleri, hiperparametre optimizasyonu ve domain uyumu bulunmaktadır.

Büyük, dengeli ve yüksek kaliteli veri setlerinde model yüksek performans gösterme potansiyeline sahipken, küçük veya dengesiz veri setlerinde performans düşüklüğü yaşanabilir. Gürültülü veya düşük kaliteli verilerde ise ek ön işleme teknikleri ve domain adaptasyonu gerekmektedir.

3.7. Eğitim Parametreleri

Model karşılaştırmalarının sistematik ve tutarlı olabilmesi amacıyla, seçilen tüm mimariler aynı hiperparametre konfigürasyonu ile eğitildi. Çalışmada transfer öğrenme yaklaşımı benimsendi ve tüm mimariler ImageNet veri seti üzerinde önceden eğitilmiş ağırlıklarla başlatıldı. Önceden eğitilmiş olan temel özellik çıkarıcı katmanların ağırlıkları dondurularak sabitlendi; sadece her mimariye özgü son sınıflandırıcı head kısmının ağırlıkları eğitime açık bırakıldı.

Bu yaklaşım, sınırlı veri seti ile yürütülen bu çalışmada aşırı uyum riskini en aza indirmek ve önceden öğrenilmiş genel özelliklerin korunarak model performansının optimize edilmesini sağlamak amacıyla uygulandı. Kullanılan ortak hiperparametreler Tablo 3.1'de özetlendi.

Tablo 3.1. Model Eğitim Hiperparametreleri

Parametre	Değer/Ayarlar	Açıklama
Öğrenme Oranı	0.001	Optimizasyon sürecindeki adım
Yığın Boyutu	32	Bir iterasyonda işlenen örnek sayısı
İterasyon Sayısı	50	Veri setinde yapılacak geçiş sayısı
Optimizasyon	AdamW	Optimizasyon algoritması
Kayıp Fonksiyonu	Odak Kaybı	Çoklu sınıf kayıp fonksiyonu
Ağırlık Çürümesi	0.0001	L2 regularizasyon katsayısı
Momentum	0.9	Optimizasyon için momentum değeri
Rastgelelik Tohumu	42	Tekrarlanabilirlik için sabit tohum değeri

Tüm eğitim süreci, sonuçların bilimsel tekrarlanabilirliğini sağlamak amacıyla deterministik yaklaşımla yürütüldü ve random seed değeri 42 olarak sabitlendi. Bu kapsamda:

- Pytorch kütüphanesinin deterministik algoritma ayarları etkinleştirildi,
- CUDA convolution determinism mode aktive edildi,
- Tüm rastgele sayı üreticileri (Python, NumPy, PyTorch) sabit seed değeri ile yazılıma dahil edildi,
- Veri Yükleyicilerin çalışma prensibi yeniden üretilebilecek şekilde yapılandırıldı.

Alınan bu teknik önlemler sayesinde, farklı çalıştırma koşullarında dahi aynı hiperparametreler ve veri kümesi ile yapılan eğitimler aynı sonuçları verdi. Bu yaklaşım, deneysel sonuçların güvenilirliğini ve karşılaştırılabilirliğini bilimsel standartlara uygun şekilde destekledi.

3.8. Deneysel Kurulum

3.8.1. Donanım ve Özellikler

Tüm deneysel çalışmalar ve model eğitim süreçleri, yüksek performanslı bir hesaplama sunucusu üzerinde gerçekleştirildi. Kullanılan donanım altyapısının teknik özellikleri şu şekildedir.

İşlemci Birimi (CPU)

- Intel Xeon Gold 5320 işlemci
- 2.20 GHz temel saat hızı (2.19 GHz baz çekirdek frekansı)
- 26 çekirdek ve 52 iş parçacığı
- 35.75 MB akıllı önbellek

Bellek (RAM)

- 512 GB DDR4 bellek kapasitesi
- 2933 MHz bellek hızı
- 8 bellek kanalı desteği

Grafik İşlemci Birimi (GPU)

- NVIDIA Tesla A40 ekran kartı
- 10,752 CUDA çekirdek
- 48 GB GDDR6 bellek
- 696 GB/s bellek bant genişliği
- 336 Tensor çekirdek

Depolama Birimi

- 2.5 TB kapasiteli NVMe SSD depolama
- 7000 MB/s okuma, 5500 MB/s yazma hızı
- RAID 0 disk konfigürasyonu

3.8.2. Yazılım Kütüphaneleri

Bu çalışmada, derin öğrenme modelinin geliştirilmesi, eğitimi ve değerlendirilmesi için aşağıdaki yazılım kütüphaneleri ve framework'leri kullanıldı.

- **Streamlit:** Etkileşimli web arayüzü oluşturmak için kullanılan açık kaynak kodlu bir Python framework'üdür.
- **PyTorch:** Derin öğrenme modellerinin geliştirilmesi ve eğitimi için kullanılan temel framework'tür.

- **TorchVision:** Bilgisayarlı görü uygulamaları için veri kümeleri, model mimarileri ve görüntü dönüşümleri sağlar.
- **Pillow (PIL) :** Görüntü işleme ve manipülasyon operasyonları için kullanılmıştır.
- **NumPy:** Bilimsel hesaplamalar ve çok boyutlu dizi operasyonları için temel kütüphanedir.
- **OpenCV (cv2) :** Gelişmiş görüntü işleme ve bilgisayarlı görü algoritmaları için kullanılmıştır.
- **Scikit-learn:** Makine öğrenmesi metrikleri, veri bölütleme ve model değerlendirme için kullanılmıştır.
- **Pandas:** Veri manipülasyonu ve analizi için kullanılmıştır.
- **Matplotlib:** Görselleştirme ve grafik oluşturma için kullanılmıştır.
- **Seaborn:** İstatistiksel veri görselleştirme için kullanılmıştır.
- **TIMM:** Çeşitli önceden eğitilmiş vision modellerine erişim sağlayan kütüphanedir.
- **Python:** Tüm çalışmanın geliştirildiği programlama dilidir.

Ek olarak, CUDA ve cuDNN kütüphaneleri GPU hızlandırma için kullanıldı. Tüm kütüphane versiyonları, çalışmanın tekrarlanabilirliğini sağlamak amacıyla sabitlendi.

3.8.3. Eğitim Süresi ve Kaynak Kullanımı

Bu çalışmada kullanılan tüm derin öğrenme mimarileri, önceden belirlenen ortak hiperparametreler ve sabit donanım konfigürasyonu altında eğitildi. Eğitim süreleri ve kaynak kullanımı, model mimarilerinin hesaplama karmaşıklığına ve parametre sayılarına bağlı olarak değişkenlik gösterdi.

Genel olarak, CNN tabanlı mimarilerin (ResNet-50, EfficientNet-B4, ConvNeXt) daha düşük hesaplama kaynağı ile daha kısa sürede eğitildiği gözlemlendi. Buna karşılık, ViT tabanlı mimariler (Swin Transformer, DeiT, PVTv2) ve özellikle Hybrid yapılar daha yüksek bellek kapasitesi ve daha uzun eğitim süresi gerektirdi.

Tüm modellerin eğitimi sırasında sistem kaynakları istikrarlı bir şekilde izlendi, bellek taşması (out-of-memory) veya diğer donanım kısıtlamaları nedeniyle herhangi bir eğitim kesintisi yaşanmadı. GPU bellek kullanımı model karmaşıklığına paralel olarak artış gösterdi; özellikle büyük ölçekli transformer modelleri için daha yüksek bellek ayırımı gerekti.

3.9. Performans Metrikleri

Doğruluk (Accuracy)

Doğruluk, modelin tüm tahminlerinin doğruluğunu ölçen temel bir metriktir. Genellikle sınıflandırma problemlerinde modelin genel başarısını değerlendirmek için kullanılır. Doğruluk, doğru tahminlerin toplam örnek sayısına oranı olarak hesaplanır ve modelin genel performansını hızlıca anlamaya yardımcı olur. Ancak dengesiz veri setlerinde yanıltıcı olabilir, çünkü azınlık sınıfın doğru tahmin edilmesi göz ardı edilebilir. Özellikle sınıfların dengeli olduğu durumlarda doğruluk önemli bir göstergedir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN}$$

Kesinlik (Precision)

Kesinlik metriği, modelin pozitif olarak işaretlediği tahminlerin ne kadarının gerçekten doğru olduğunu ölçer. Bu metrik, tahminlerin güvenilirliği ve kalitesi hakkında önemli bilgiler verir. Kesinlik, gerçek pozitiflerin, toplam pozitif tahminlere (gerçek pozitifler ve yanlış pozitiflerin toplamı) oranı olarak hesaplanır.

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

Yüksek bir kesinlik değeri, modelin pozitif etiketi verdiği durumlara yüksek güvenle inanabileceğimizi gösterir. Bu metrik özellikle yanlış pozitif hatalarının maliyetinin yüksek olduğu senaryolarda kritik öneme sahiptir. Örneğin tıbbi teşhis sistemlerinde, yanlış pozitif sonuçlar gereksiz tıbbi müdahalelere ve hasta stresine yol açabilir. Kesinlik metriği, bu tür yanlış alarmları minimize etmeye odaklanır ve modelin pozitif sınıfı ne kadar "kesin" şekilde tanımlayabildiğini ölçer.

Duyarlılık - Hassasiyet (Recall - Sensitivity)

Duyarlılık metriği, gerçekte pozitif olan tüm örneklerin ne kadarının model tarafından doğru bir şekilde tespit edildiğini ölçer. Bu metrik, modelin bir hastalığı veya nadir bir olayı yakalama yeteneğini değerlendirir. Duyarlılık, gerçek pozitiflerin, toplam gerçek pozitifler ve yanlış negatiflerin toplamına oranı olarak hesaplanır.

$$\text{Duyarlılık} = \frac{TP}{TP + FN}$$

Yüksek bir duyarlılık değeri, modelin pozitif sınıfa ait örnekleri atlamadığını ve neredeyse tamamını doğru şekilde tanımladığını gösterir. Bu metrik, yanlış negatif hatalarının sonuçlarının ağır olduğu durumlarda en önemli metrik haline gelir. Örneğin güvenlik tarama sistemlerinde veya kritik tıbbi teşhislerde, gerçek tehditlerin veya hastalıkların atlanması ciddi sonuçlar doğurabilir. Duyarlılık metriği, bu tür kritik hataları önlemeye odaklanır.

F1 Skoru

F1 skoru, Precision ve Recall arasındaki dengeyi ölçer ve özellikle dengesiz veri setlerinde veya kritik hataların önemli olduğu durumlarda önemli hale gelir. Bu metrik, doğru pozitiflerin yanı sıra yanlış pozitifleri de göz önünde bulundurarak modelin doğruluğunu daha kapsamlı bir şekilde değerlendirir. Modelin sınıflandırmadaki başarısını hem doğru pozitifler açısından hem de yanlış pozitifler açısından dengeleyerek sunar. F1 skoru, özellikle yanlış negatiflerin ve yanlış pozitiflerin önemli olduğu tıbbi görüntüleme gibi alanlarda yaygın olarak kullanılır.

$$F1 = \frac{2 * Kesinlik * Hassasiyet}{Kesinlik + Hassasiyet}$$

Özgüllük (Specificity)

Özgüllük metriği, gerçekte negatif olan örneklerin ne kadarının model tarafından doğru şekilde tanımlandığını ölçer. Bu metrik, duyarlılığın negatif sınıf için olan karşılığı olarak düşünülebilir. Özgüllük, gerçek negatiflerin, toplam gerçek negatifler ve yanlış pozitiflerin toplamına oranı olarak hesaplanır.

$$\text{Özgüllük} = \frac{TN}{TN + FP}$$

Yüksek bir özgüllük değeri, modelin negatif sınıfı doğrulamada çok başarılı olduğunu gösterir. Bu metrik, bir hastalığın veya durumun yokluğunu teyit etmek istediğinizde kritik öneme sahiptir. Özgüllük ve duyarlılık genellikle ters orantılıdır; birini artırmak genellikle diğlerinden ödün verilmesine neden olur. Bu nedenle, bu iki metriği birlikte değerlendirmek önemlidir.

ROC-AUC (Area Under the ROC Curve)

ROC-AUC, modelin sınıflandırma performansını değerlendiren bir metriktir. ROC eğrisinin altında kalan alanı ölçerek, modelin farklı karar sınırlarında nasıl performans gösterdiğini analiz eder. Yüksek AUC, modelin genel başarı oranını yansıtarak modelin doğru sınıflandırmalar yapma yeteneğini gösterir. Bu metrik, sınıflandırma ve anomali tespiti gibi görevlerde kullanılır ve modelin doğruluğunu değerlendirmek için kapsamlı bir bakış açısı sunar.

3.10. Model Açıklanabilirliği

Modellerin karar mekanizmalarının şeffaflığını artırmak ve tahminlerin güvenilirliğini değerlendirmek amacıyla çeşitli XAI yöntemleri kullanılmıştır. Bu yöntemler, özellikle tıbbi tanı uygulamalarında kritik önem taşıyan karar verme süreçlerinin anlaşılabilirliğini artırmaktadır.

Kullanılan Yöntemler

Grad-CAM (Gradient-weighted Class Activation Mapping): Hedef sınıf için hesaplanan gradyanların, son evrimsel katmandaki özellik haritalarına geri yayılımı temel alır. Belirli bir sınıf için en çok katkıda bulunan bölgeleri görselleştirir. Yerelleştirme yeteneği yüksektir ve hesaplaması nispeten hızlıdır. Özellikle ResNet, EfficientNet gibi geleneksel CNN mimarileri için etkili sonuçlar vermiştir.

Score-CAM (Score-weighted Class Activation Mapping): Grad-CAM'den farklı olarak gradyan bağımlılığını ortadan kaldırarak, özellik haritalarının aktivasyon skorları üzerinden ağırlıklandırma yapar. Gradyan gürültüsünden etkilenmez ve daha tutarlı ısı haritaları üretir. Özellikle gradyan akışının zayıf olduğu durumlarda daha güvenilir sonuçlar sağlamıştır.

Attention Rollout: ViT mimarilerindeki öz-dikkat mekanizmalarının katmanlar boyunca birikimli olarak analiz edilmesine dayanır. Dikkat ağırlıklarının çarpımı ile global bir dikkat haritası oluşturur. Transformer tabanlı modellerin karmaşık dikkat mekanizmalarını görselleştirmede etkilidir. Swin Transformer, DeiT, PVTv2 gibi mimarilerde başarıyla uygulanmıştır.

Layer-wise Relevance Propagation (LRP): Modelin çıktısını girdi uzayına geri yayarak, her bir pikselin nihai karara olan katkısını hesaplar. İleri yayılım kurallarının tersi işletilir. Pixel-level detay sağlar ve oldukça hassas lokalizasyon yapabilir. Özellikle sınır bölütleme önemli olduğu durumlarda kullanılmıştır.

Her bir XAI yöntemi, farklı mimariler için optimize edilmiş parametrelerle uygulandı. Isı haritalarının nicel değerlendirilmesi için maksimum aktivasyon, ortalama aktivasyon ve aktivasyon alanı metrikleri hesaplandı. Aktivasyon alanı hesaplamalarında, 0.5'lik bir eşik değeri kullanılarak anlamlı aktivasyon bölgeleri belirlendi. Bu görselleştirmeler, model kararlarının doğrulanmasına ve potansiyel hatalı pozitif/negatif durumların tespit edilmesine olanak sağladı. Ayrıca, modelin anatomik olarak anlamlı bölgeleri öğrenip öğrenmediğinin doğrulanması yapıldı. Kullanılan çoklu XAI yöntemleri, modellerin karar süreçlerine ilişkin kapsamlı bir anlayış sundu ve farklı mimarilerin farklı XAI yöntemlerine verdiği yanıtlar karşılaştırmalı olarak analiz edildi.

3.11. Yazılım

Bu çalışmada geliştirilen yazılım mimarisi, sınıflandırma görevini gerçekleştirmek üzere tasarlanmış modüler ve ölçeklenebilir bir yapıya dayanmaktadır. Sistem, veri ön işleme, model eğitimi, değerlendirme ve görselleştirme süreçlerini kapsayan entegre bir pipeline şeklinde yapılandırıldı.

3.11.1. Yazılım Mimarisi

Mimari, veri yönetim katmanı, model katmanı, eğitim katmanı ve değerlendirme katmanı olmak üzere dört temel bileşenden oluşmaktadır. Veri yönetim katmanında veri okuma, ön işleme, veri artırma (augmentation) ve batch tabanlı yükleme mekanizmaları bulunmaktadır. Model katmanı, ön eğitilmiş CNN ve ViT tabanlı modelleri, Hybrid mimari entegrasyonunu ve model kaydetme-yükleme modüllerini içermektedir. Eğitim katmanı hiperparametre yapılandırma, eğitim döngüsü optimizasyonu, doğrulama süreçleri ve model checkpointing mekanizmasıyla desteklenmiştir. Değerlendirme katmanı ise metrik hesaplama modülleri, görselleştirme araçları ve sonuç raporlama sisteminden oluşmaktadır.

3.11.2. Yazılım Özellikleri

Mimari tasarımın öne çıkan özellikleri arasında modüler yapısı sayesinde bileşenlerin bağımsız güncellenebilmesi, yeniden kullanılabilir kod yapısı, esnek konfigürasyon yönetimi, otomatik loglama ve monitoring desteği ile GPU eğitim imkânı yer almaktadır. Bu sayede sistem, deneylerin tekrarlanabilirliğini ve sonuçların güvenilirliğini sağlarken, farklı model mimarilerinin kolayca entegre edilmesine de olanak tanımaktadır.

3.11.3. Yazılım Arayüz Tasarımı ve İş Akışı

Başlangıç ve Kurulum Aşaması

Sistem başlatıldığında, öncelikle seed değerleri ayarlanarak deterministik bir ortam sağlanır. GPU/CPU tespiti gerçekleştirilir, oturum durumu (session state) değişkenleri dahil edilir ve bellek temizleme işlemleri uygulanır. Ardından kullanıcı arayüzü yüklenir; kenar çubuğu menüsü oluşturulur, ana ekran bileşenleri hazırlanır ve kullanıcı etkileşimi için gerekli olay dinleyicileri aktif hale getirilir.

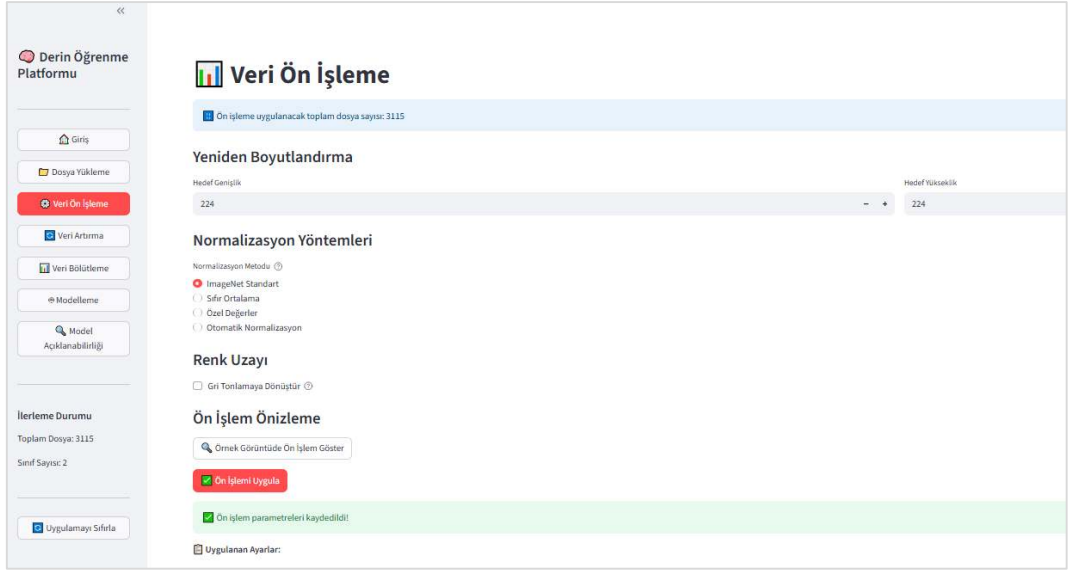
Veri Toplama ve Hazırlama Aşaması

Kullanıcı, sınıf etiketlerini girerek veri sınıflandırmasını başlatır. Etiketler sınıf listesine eklenir ve ilgili oturum durumu değişkenleri güncellenir. Dosya yükleme arayüzü üzerinden kullanıcı görüntüleri sisteme aktarır. Dosya formatı kontrol edilir, PIL Image formatına dönüştürülür ve oturum durumuna kaydedilir. Ardından önizleme görüntüleri hazırlanır ve temel istatistikler hesaplanır.

Şekil 3.1. Veri seti yükleme

Veri Ön İşleme Aşaması

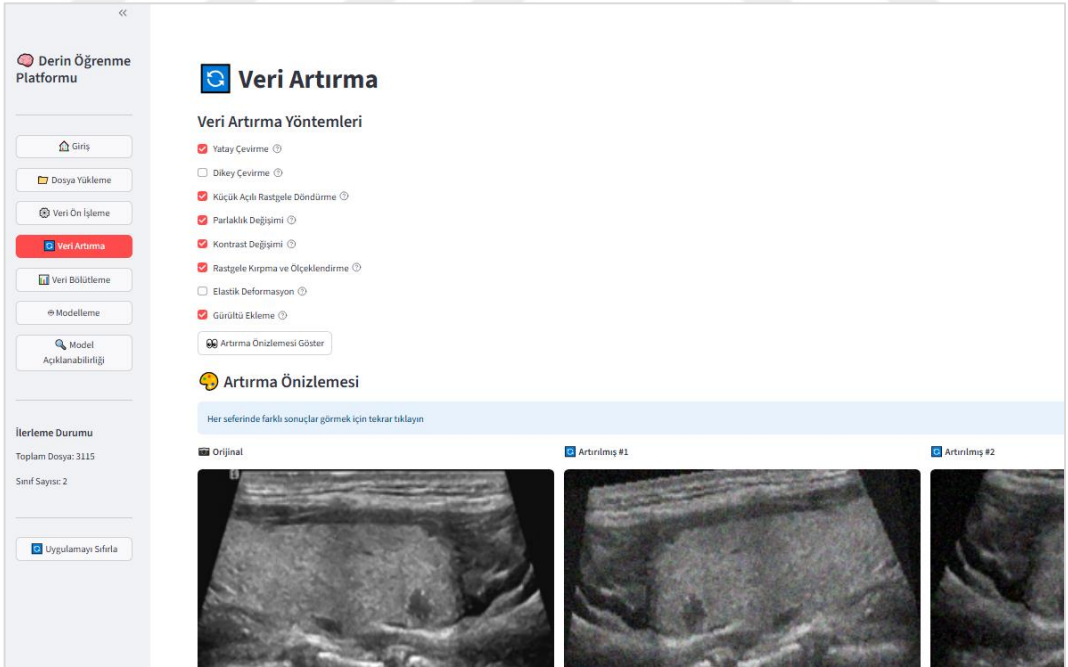
Bu aşamada kullanıcı, hedef boyut, normalizasyon yöntemi ve renk uzayı gibi parametreleri belirler. Seçilen parametrelere uygun dönüşüm pipeline'ı oluşturularak tüm görüntülere uygulanır. Normalizasyon istatistikleri hesaplanır ve oturum durumuna kaydedilir. Böylece veriler sonraki işlemler için hazır hale getirilir.



Şekil 3.2. Veri ön işleme

Veri Artırma (Augmentation) Aşaması

Kullanıcı, uygulanacak veri artırma yöntemlerini ve parametrelerini belirler. Rastgele artırılmış örneklerin önizlemesi sunulur. Seçilen yöntemlerle eğitim ve test dönüşümleri ayrı ayrı oluşturulur ve oturum durumuna kaydedilir.



Şekil 3.3. Veri artırma

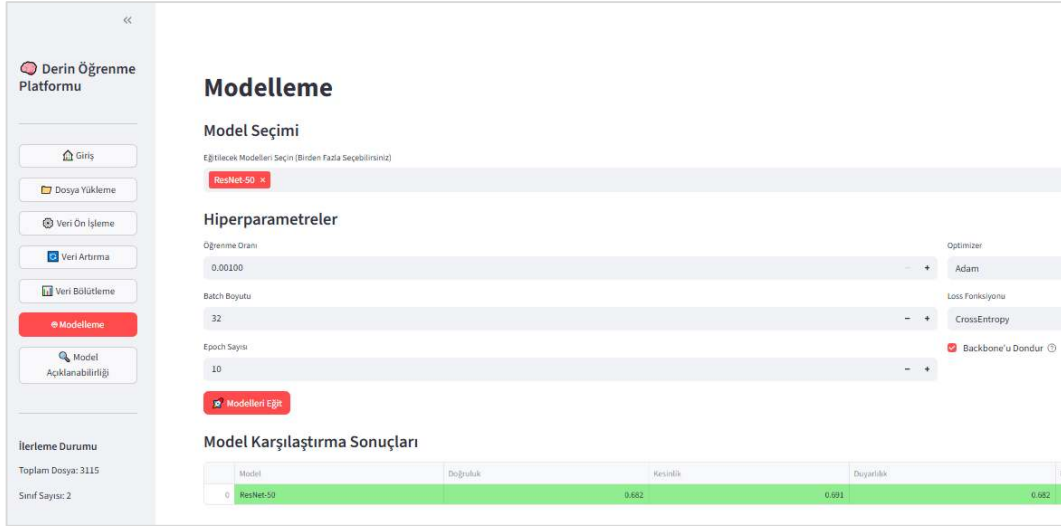
Veri Bölütleme Aşaması

Bu aşamada kullanıcı, veri setinin eğitim, doğrulama ve test alt kümelerine ayrılması için oranları ve bölütleme stratejisini belirler. Sınıf dengesizliği ve random seed ayarları yapılır. Görseller ve etiketler birleştirildikten sonra veri seti bölütlenir, istatistikler hesaplanarak kullanıcıya sunulur ve oturum durumuna kaydedilir.

Şekil 3.4. Veri bölütleme

Model Eğitim Aşaması

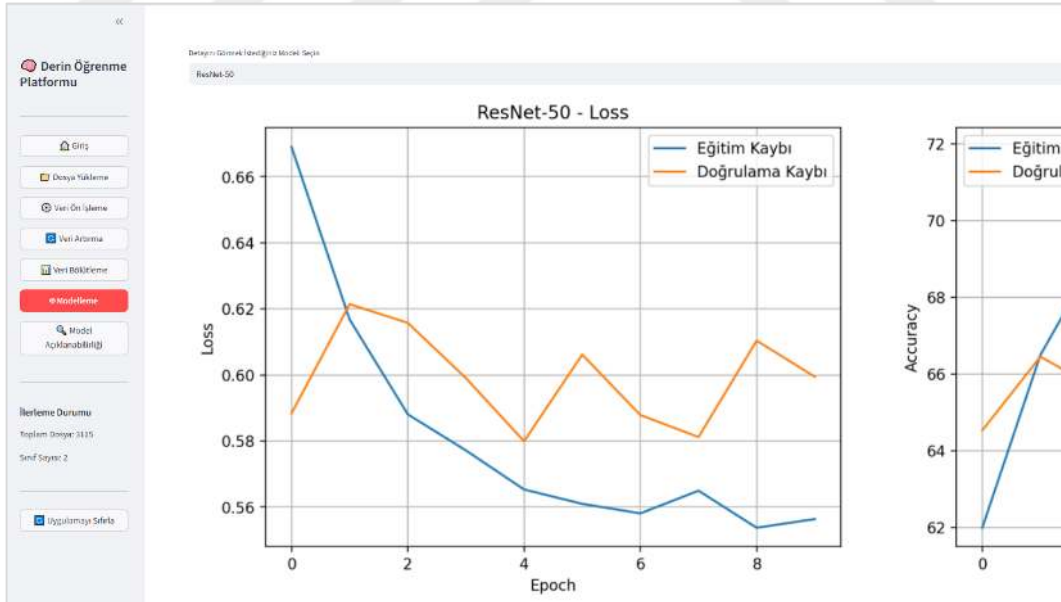
Kullanıcı, eğitilecek modelleri ve hiperparametrelerini seçer. Optimizer ve loss fonksiyonu belirlenir. Seçilen modeller için mimari oluşturulur, ağırlıklar dahil edilir, veri yükleyiciler hazırlanır ve eğitim döngüsü başlatılır. Eğitim sürecinde ilerleme durumu gerçek zamanlı gösterilir, Kayıp değer ve doğruluk grafikleri çizilir ve model ağırlıkları kaydedilir.



Şekil 3.5. Model eğitim

Model Değerlendirme Aşaması

Test seti üzerinde metrikler hesaplanır, karmaşıklık matrisi ve sınıf bazlı performans analizleri yapılır. Kayıp Değer ve Doğruluk eğrileri çizilir ve tüm sonuçlar Oturum durumuna kaydedilir. Modellerin performansları karşılaştırılır, görsel grafiklerle desteklenir ve en iyi performans gösteren model belirlenir.



Şekil 3.6. Model değerlendirme

Model Açıklanabilirlik Aşaması

Kullanıcı, açıklanacak modeli ve yöntemi seçer. Belirlenen görüntü üzerinde ısı haritası tabanlı görselleştirme yapılır, sonuçlar orijinal görüntüyle birleştirilir ve yan yana

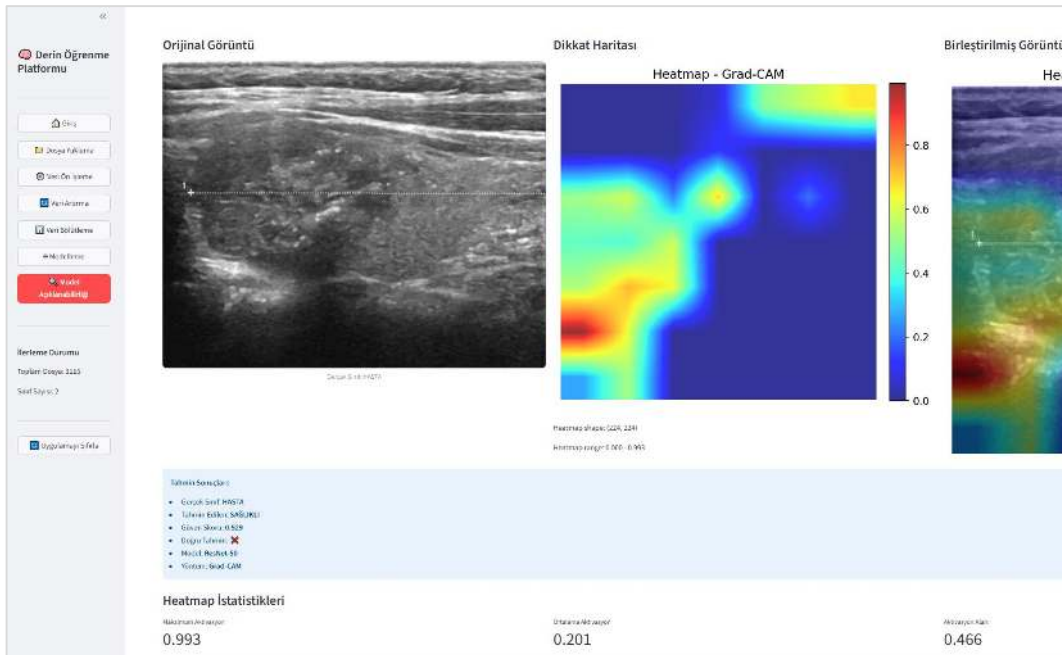
sunulur. Farklı modellerin aynı görüntü üzerindeki açıklamaları karşılaştırılarak tahminler, güven skorları ve performans metrikleri birlikte değerlendirilir.



Şekil 3.7. Model açıklanabilirlik

Sonuçlar ve Raporlama Aşaması

Tüm süreç özetlenir, en iyi model ve performans metrikleri vurgulanır. Detaylı sonuç tabloları, öneriler ve iyileştirme noktaları sunulur. Kullanıcıya model ağırlıklarını indirme, metrik raporlarını dışa aktarma ve ısı haritası görsellerini kaydetme seçenekleri sağlanır.



Şekil 3.8. Sonuçlar ve raporlama

4. BULGULAR

4.1. Veri Seti Bulguları

Bu çalışma için geliştirilen yazılım ile tiroid ultrason görüntü veri seti üzerinde derin öğrenme mimarilerinin karşılaştırmalı performans analizi gerçekleştirildi. CNN, Transformer ve hibrit mimariler olmak üzere üç farklı kategoride toplam on bir derin öğrenme modelinin kapsamlı değerlendirmesi yapıldı. Modellerin performans ölçümü için doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1-skoru, özgüllük (specificity) ve ROC-AUC değeri olmak üzere altı farklı performans metriği kullanıldı; elde edilen sonuçlar Tablo 4.1’de sunuldu.

Tablo 4.1. Görüntü Veri Seti Üzerindeki Performans Karşılaştırma Sonuçları

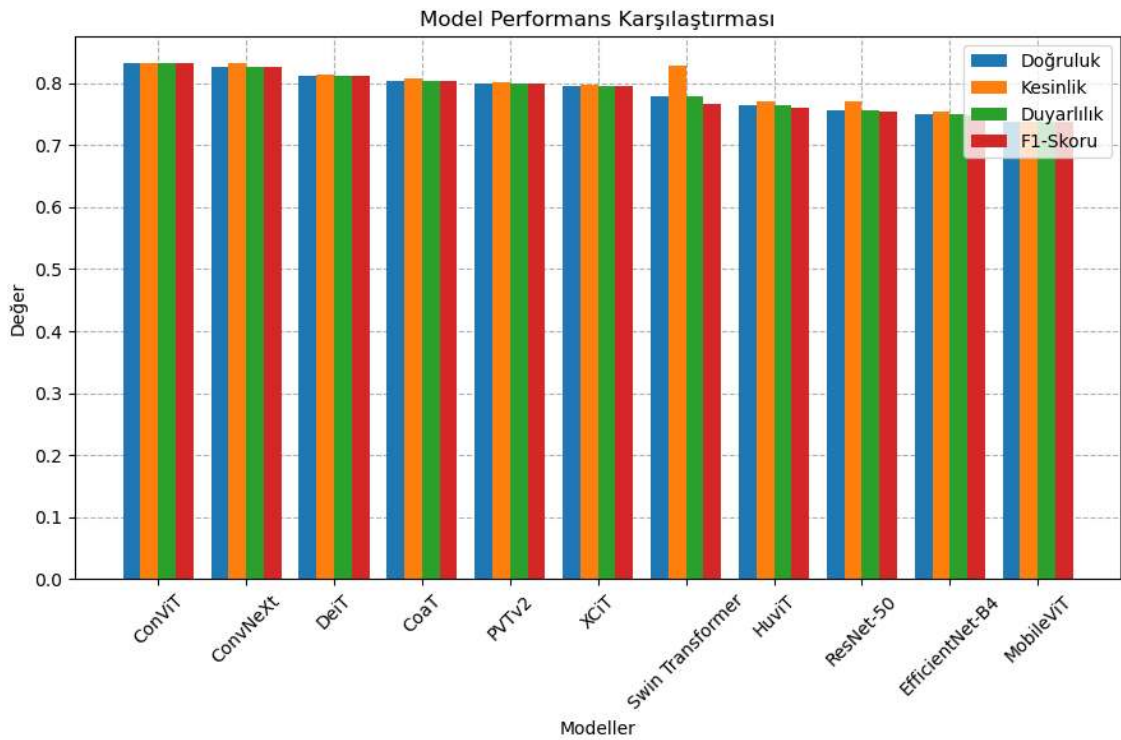
Model	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	Özgüllük	ROC-AUC
ConViT	0.8330	0.8330	0.8326	0.8330	0.8326	0.8994
ConvNeXt	0.8263	0.8312	0.8238	0.8251	0.8238	0.9002
DeiT	0.8118	0.8134	0.8101	0.8112	0.8101	0.8910
CoaT	0.8041	0.8084	0.8016	0.8028	0.8016	0.8785
PVTv2	0.8002	0.8015	0.7986	0.7996	0.7986	0.8753
XCiT	0.7954	0.7964	0.7938	0.7948	0.7938	0.8604
Swin Transformer	0.7780	0.8271	0.7709	0.7673	0.7709	0.8849
HuViT	0.7635	0.7709	0.7601	0.7608	0.7601	0.8386
ResNet-50	0.7568	0.7711	0.7602	0.7548	0.7602	0.8771
EfficientNet-B4	0.7490	0.7530	0.7462	0.7471	0.7462	0.8105
MobileViT	0.7375	0.7379	0.7360	0.7369	0.7360	0.8118

Deneysel bulgular, ConViT mimarisinin tüm performans metriklerinde en yüksek değerlere ulaşarak üstün bir performans sergilediğini gösterdi. Model, %83.3 doğruluk, %83.3 F1-Skoru ve %89.9 ROC-AUC değeri elde etti. Bu sonuçlar, ConViT’in tiroid ultrason görüntülerinin sınıflandırılmasında istikrarlı şekilde en yüksek performansı sağladığını ortaya koydu.

Evrişimsel mimariler arasında, ikinci sırayı %82.6 doğruluk ve %90.0 ROC-AUC değeri ile ConvNeXt modeli aldı. Transformer tabanlı mimariler arasında ise DeiT, %81.2 doğruluk ve %89.1 ROC-AUC değeri ile en yüksek performansı sergiledi.

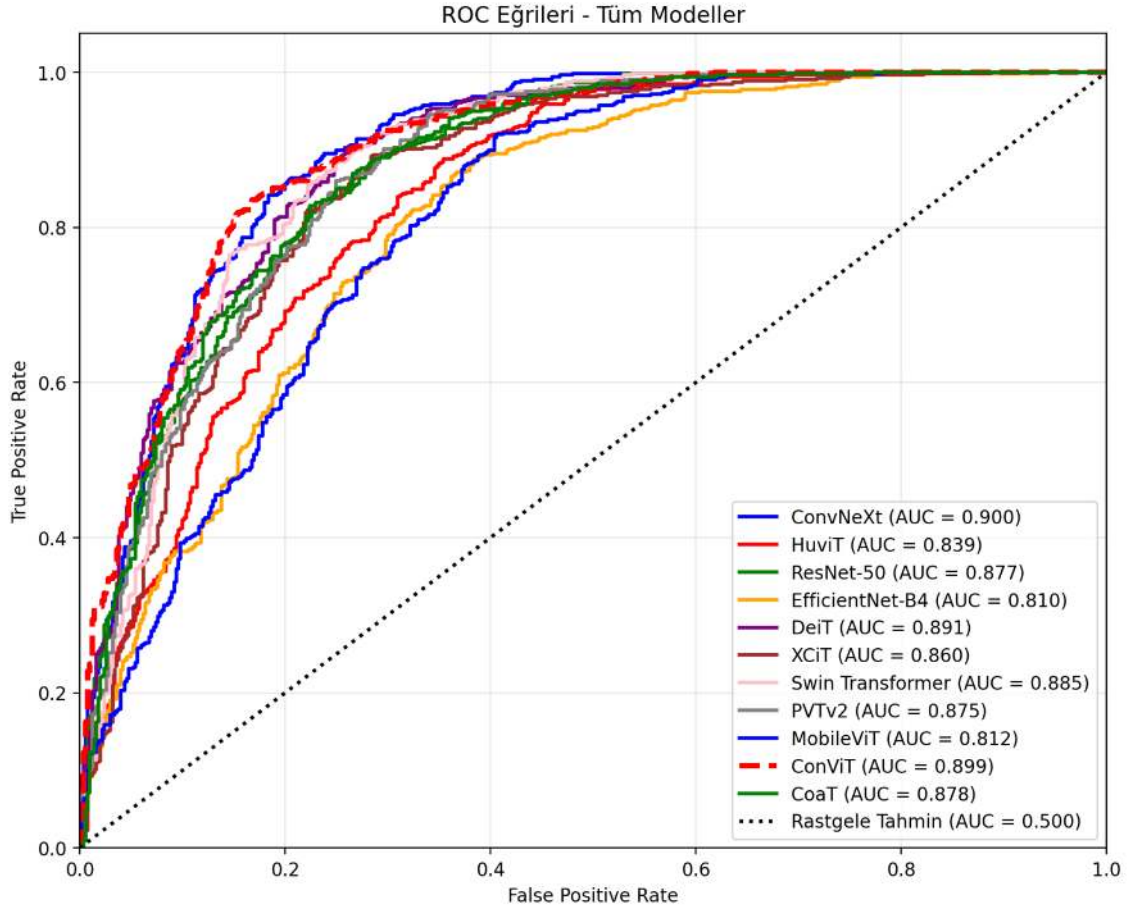
Hibrit mimarilerin kendi içindeki karşılaştırmasında, ConViT açık ara en başarılı performansı sergiledi. ConViT'i %80.4 doğruluk ve %87.9 ROC-AUC değeri ile CoaT modeli takip etmiş, onu %76.4 doğruluk değeri ile HuViT modeli izledi. MobileViT modeli ise %73.8 doğruluk değeri ile hibrit modeller arasında en düşük performansı gösterdi.

Performans değerlendirmesi, ConViT gibi hibrit bir mimarinin performansının, diğer tüm CNN, ViT ve hibrit mimarilerden daha üstün olduğunu gösterdi. Bununla birlikte, ConvNeXt ve DeiT gibi modern mimariler de yüksek ve birbirine yakın performans değerleri elde etti.



Şekil 4.1. Model performans karşılaştırması

Şekil 4.1 incelendiğinde, tüm modellerde doğruluk, kesinlik, duyarlılık ve F1-skoru değerlerinin birbirine yakın olduğu gözlemlendi. Bu durum, modellerin sınıflandırma performanslarının dengeli olduğunu ve herhangi bir sınıfın diğerine göre belirgin şekilde önceliklendirilmediğini ortaya koydu.



Şekil 4.2. ROC eğrileri – Tüm modeller

Şekil 4.2.'de ROC-AUC değerlerinin tüm modellerde doğruluk değerlerinden yaklaşık %5-10 daha yüksek olması, modellerin ayırt edicilik gücünün genel olarak yüksek olduğuna işaret etmektedir.

4.2. Model Çıktılarının Analizi

Modellerin performans analizi için sınıf bazlı metrik değerler, karmaşıklık matrisi ve ROC eğrisi görselleştirmeleri gerçekleştirildi. ConViT modelinin sınıf bazlı performans metrikleri Tablo 4.2'de sunuldu.

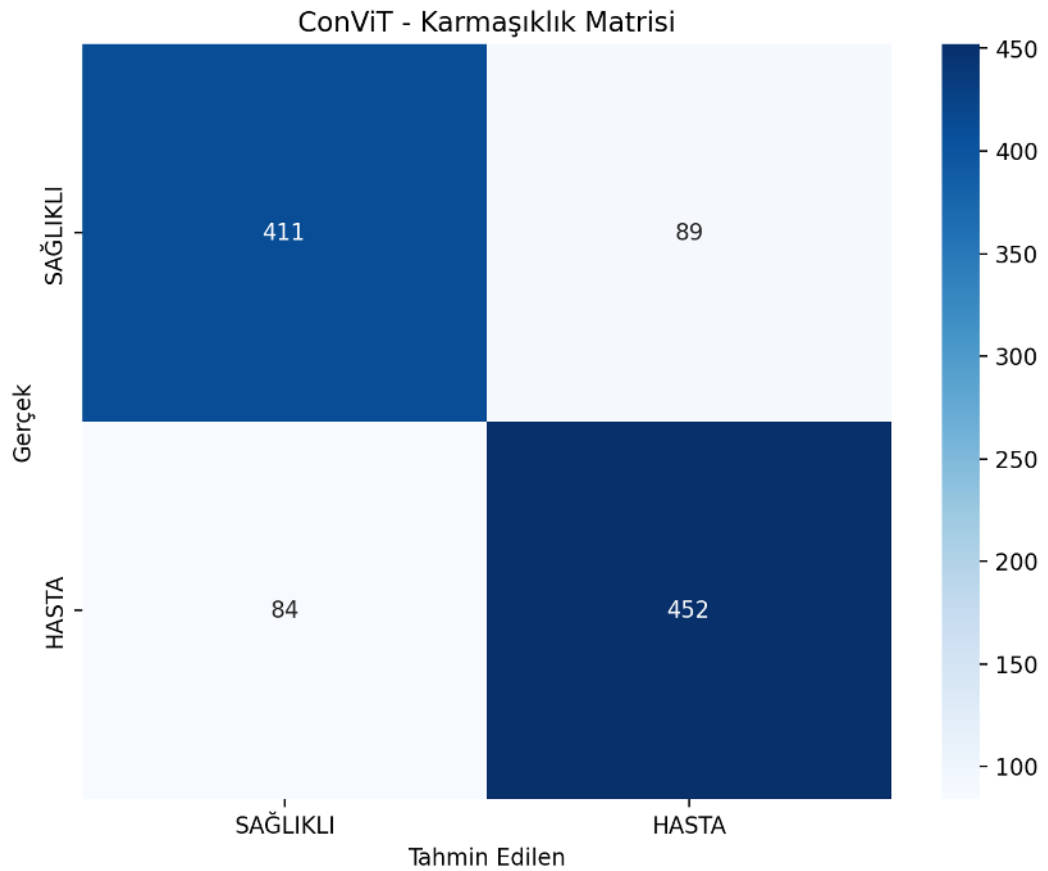
Tablo 4.2. ConViT modeline ait sınıf bazlı performans metrikleri

Sınıf	Kesinlik	Duyarlılık	F1-Skoru	Örnek
SAĞLIKLI	0.830	0.822	0.827	500
HASTA	0.836	0.843	0.839	536

Tablo 4.2 incelendiğinde, ConViT modelinin her iki sınıf için de dengeli ve yüksek performans sergilediği gözlemlendi. Sağlıklı sınıfta %83.0 kesinlik ve %82.2 duyarlılık oranlarına ulaşan model, bu sınıfa yönelik tahminlerinin yüksek güvenilirlikte olduğunu ve sınıfın vakalarının büyük çoğunluğunu doğru şekilde tanımlayabildiğini gösterdi. Hasta sınıfında ise %83.6 kesinlik ve %84.3 duyarlılık değerleri elde edildi. Modelin her iki sınıfta da benzer ve oldukça yüksek duyarlılık değerleri, pozitif vakaları tespit etmede güçlü bir performans sergilediğine işaret etti. Her iki sınıf için F1-skoru değerlerinin (%82.7 ve %83.9) birbirine yakın ve yüksek olması, modelin genel dengede son derece başarılı olduğunu ortaya koydu.

4.3. Karmaşıklık Matrisi Analizi

Şekil 4.3'te sunulan karmaşıklık matrisi, modelin sınıflandırma performansının detaylı bir dökümünü sağlamaktadır. Matris analizi, modelin hasta sınıfındaki örneklerin 452'sini doğru tahmin ederken (TP), 84'ünü yanlışlıkla sağlıklı olarak sınıflandırdığını (FN) gösterdi. Sağlıklı sınıfta ise 411 doğru tahmin (TN) ve 89 yanlış tahmin (FP) bulundu.

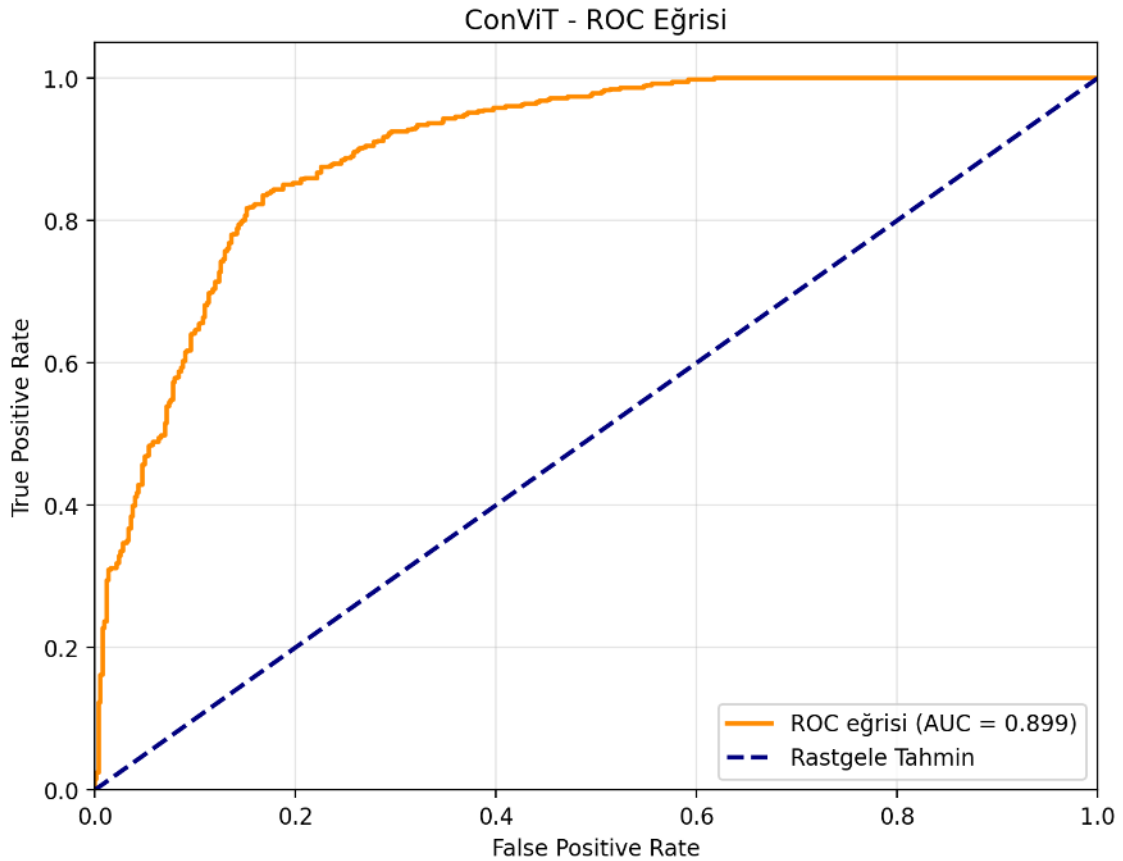


Şekil 4.3. Karmaşıklık matrisi

Bu dağılım, modelin özellikle hasta sınıfını tanımada daha başarılı olduğunu, bu sınıftaki vakaların büyük bir çoğunluğunu (%84.4) doğru bir şekilde tespit ettiğini ortaya koymaktadır. Sağlıklı sınıfındaki doğruluk oranı ise %82.2'dir.

4.4. ROC Eğrisi ve AUC Değeri

Şekil 4.4'te görülen ROC eğrisi, modelin sınıflandırma performansının eşik değerinden bağımsız bir değerlendirmesini sunmaktadır. ConViT modelinin 0.899 AUC değeri, modelin pozitif ve negatif sınıfları ayırt etmede oldukça başarılı olduğunu gösterdi. Eğrinin sol üst köşeye yakın seyri, modelin yüksek gerçek pozitif oranını düşük yanlış pozitif oranı ile elde edebildiğinin bir göstergesidir.

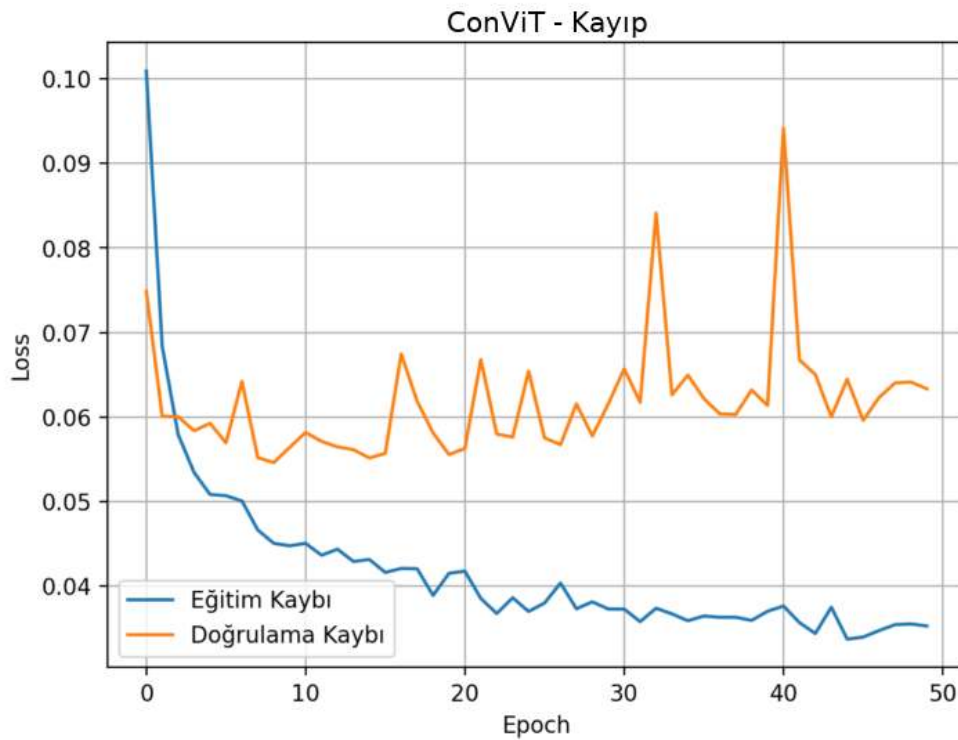


Şekil 4.4. ConViT – ROC eğrisi

Şekil 4.5'teki grafik, 50 dönem (epoch) boyunca ConViT modelinin kayıp değerlerini göstermektedir. Eğitim Kaybı ve Doğrulama Kaybı şeklinde iki farklı eğri sunulmaktadır.

Eğitim kaybı (mavi çizgi), yaklaşık 0.100'den başlayıp 50. dönemde 0.400'ün altına kadar istikrarlı bir düşüş eğilimi gösterdi. Bu durum, modelin eğitim verilerinden etkili bir şekilde öğrendiğini ve zaman içinde hatasını azalttığını ortaya koydu.

Doğrulama kaybı (turuncu çizgi) ise genel olarak 0.750 ile 0.550 aralığında seyretti ve eğitim kaybına kıyasla daha dalgalı bir seyir izledi. Özellikle 32. ve 40. dönemler civarında geçici artışlar gözlemlendi. Eğitim ve doğrulama kayıp eğrileri arasındaki bu davranış farkı, modelin öğrenme sürecindeki doğal bir varyasyonu yansıttı ve derin öğrenme modellerinde sıklıkla karşılaşılan bir durum olarak değerlendirildi.



Şekil 4.5. En iyi modele ait kayıp değer grafiği

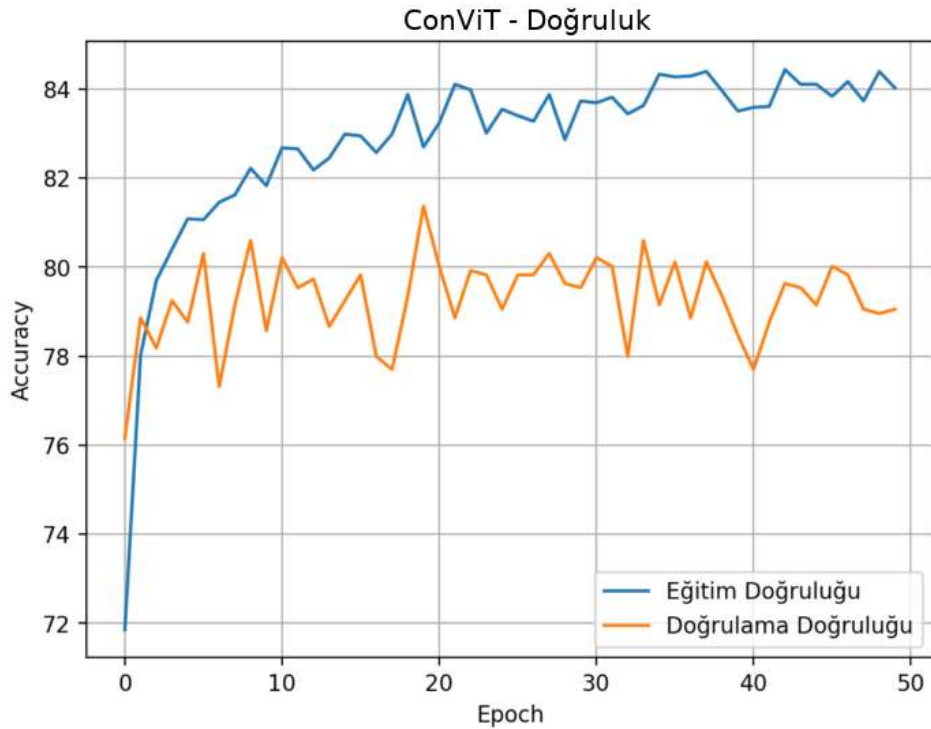
4.5. Doğruluk Analizi

Şekil 4.6'daki grafik, 50 dönem boyunca ConViT modelinin doğruluk değerlerini göstermektedir. Eğitim Doğruluğu ve Doğrulama Doğruluğu şeklinde iki farklı eğri sunulmaktadır.

Eğitim doğruluğu (mavi çizgi), ilk dönemlerde hızlı bir artış göstererek ilk 10 dönemde yaklaşık %72'den %83'ün üzerine çıktı. Daha sonra kademeli olarak iyileşmeye devam ederek 50. dönemde %84'i aşan bir seviyeye ulaştı. Bu durum, modelin eğitim verilerini doğru bir şekilde sınıflandırmada giderek daha yetkin hâle geldiğini ortaya koydu.

Doğrulama doğruluğu (turuncu çizgi) da başlangıçta hızlı bir artış göstererek ilk 10 dönemde yaklaşık %80'e ulaştı. Doğrulama doğruluğu, eğitim doğruluğuna kıyasla daha fazla değişkenlik gösterdi ve yaklaşık %78 ile %82 arasında dalgalandı. Genel olarak eğitim doğruluğunun artış eğilimini takip etmekle birlikte, iki eğri arasında belirli dönemlerde küçük farklar gözlemlendi. Bu davranış, modelin öğrenme sürecindeki doğal varyasyonları yansıtmakta olup, derin öğrenme modellerinde gözlemlenen tipik bir örüntüdür.

Özetle, ConViT modeli eğitim veri kümesinde güçlü öğrenme yetenekleri sergilerken, hem eğitim hem de doğrulama doğruluk değerlerinin yüksek seviyelerde seyretmesi ve birbirine yakın kalması modelin iyi bir genelleme performansı sağladığını gösterdi.



Şekil 4.6. En iyi modele ait doğruluk değeri grafiği

4.6. Tahmin Performansı

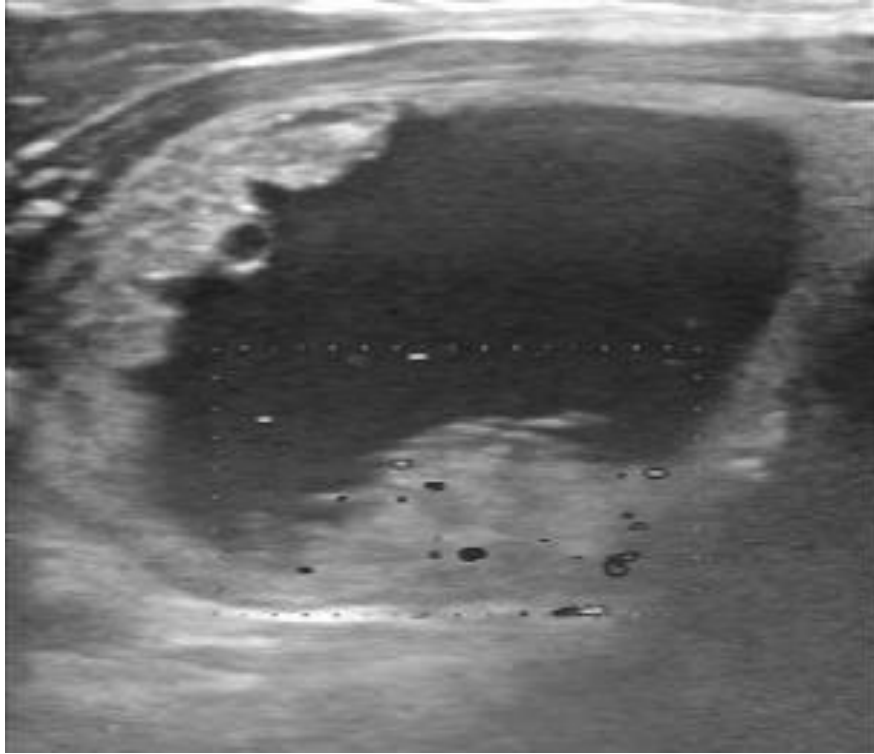
Çalışmada başarılı modellerin tahmin performanslarını detaylı olarak incelemek amacıyla örnek vaka analizleri gerçekleştirildi. Örnek model olarak ConvNeXt seçildi. Aşağıda, bu modelin tipik bir doğru tahmin örneği sunulmaktadır.

Örnek Tahmin Çıktısı:

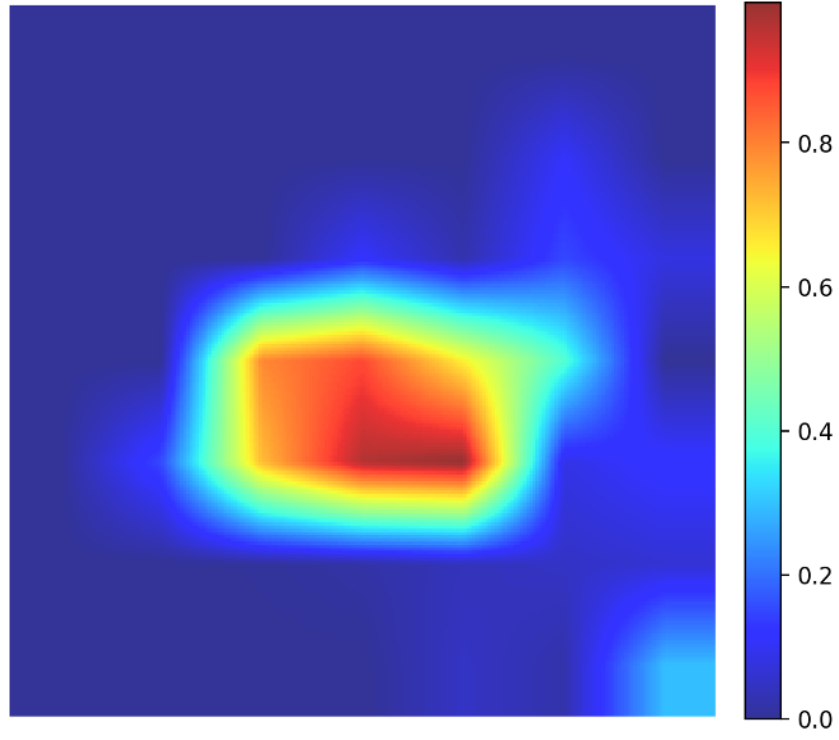
- **Gerçek Sınıf:** SAĞLIKLI
- **Tahmin Edilen Sınıf:** SAĞLIKLI
- **Güven Skoru:** 1.000
- **Tahmin Durumu:** Doğru Tahmin
- **Kullanılan Model:** ConvNeXt
- **Görselleştirme Yöntemi:** Grad-CAM

4.7. Görsel Analiz

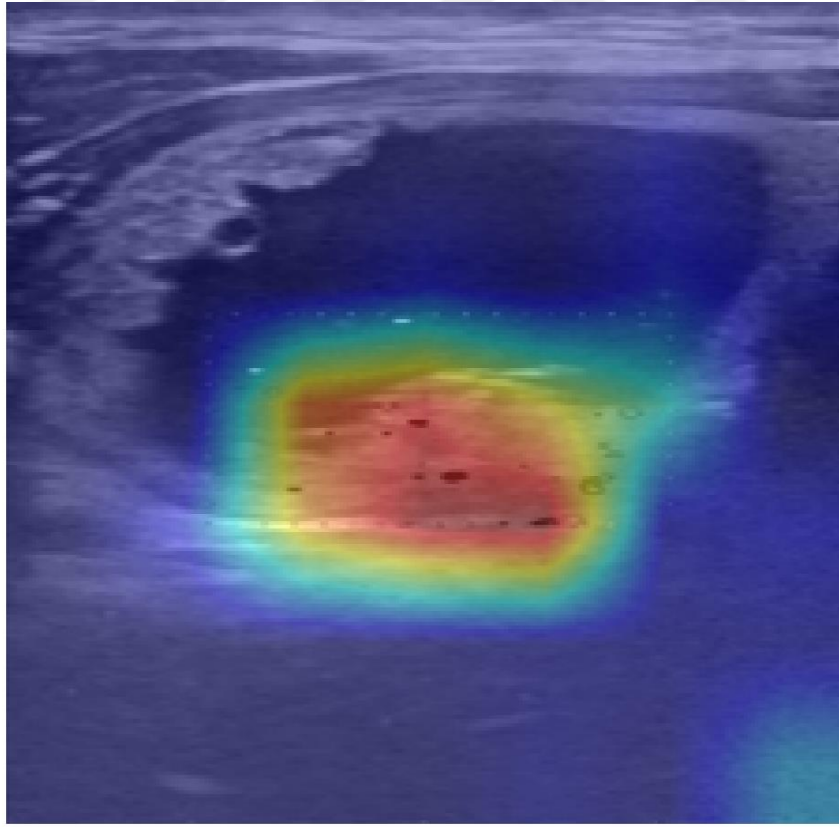
Şekil 4.7'de örnek için orijinal ultrason görüntüsü, Şekil 4.8'de Grad-CAM yöntemiyle oluşturulmuş ısı haritası (heatmap), ve Şekil 4.9'da Grad-CAM ısı haritasının orijinal görüntü ile birleştirilmiş hali sunulmaktadır. Görselleştirme sonuçları, modelin kararını oluştururken patolojik bulguların yoğun olduğu bölgelere odaklandığını göstermektedir.



Şekil 4.7. Orijinal ultrason görüntüsü

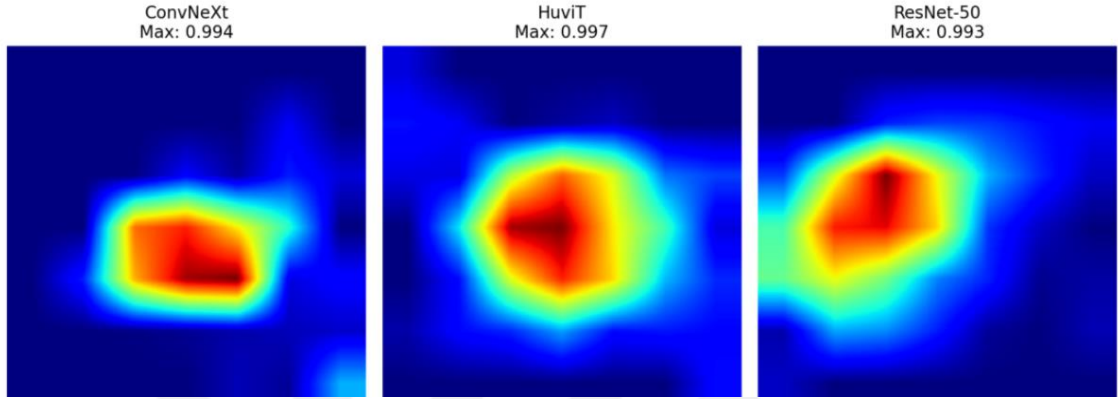


Şekil 4.8. Grad-CAM yöntemiyle oluşturulmuş ısı haritası



Şekil 4.9. Grad-CAM ısı haritasının orijinal görüntü ile birleştirilmesi

Şekil 4.9'da seçilen örnek vaka için, Tüm modellere ait Grad-CAM ısı haritaları aynı eksen üzerinde görselleştirilmiş olup, modellerin sınıflandırma kararlarını hangi görsel özniteliklere dayanarak verdiği karşılaştırmalı olarak analiz edildi. Her bir modelin ilgili sınıfa ait tahmin skorları (ConvNext: 0.989, HuviT: 0.996, ResNet-50: 0.991) ile birlikte sunulurken, model güvenilirliği ve odaklanma kalitesi arasındaki ilişki de değerlendirildi.



Şekil 4.10. Modellerin Isı haritalarının karşılaştırılması

4.8. Isı Haritası (Heatmap) İstatistiksel Analizi

Modelin kararını destekleyen Grad-CAM ısı haritasının kantitatif analizi aşağıdaki metrikleri ortaya koydu.

- **Maksimum Aktivasyon Değeri:** 0.994
- **Ortalama Aktivasyon Değeri:** 0.139
- **Aktivasyon Alanı Oranı:** 0.309

Seçilen örnekte, ConvNeXT modelinin yüksek bir güven skoru (%93) ile doğru tahmin yaptığı gözlemlendi. Grad-CAM ısı haritası analizi, modelin kararını oluştururken ilgili anatomik bölgelerde güçlü ve odaklanmış aktivasyonlar gösterdiğini ortaya koydu. Maksimum aktivasyon değerinin 0.994 gibi oldukça yüksek olması, modelin belirleyici patolojik bulguları yüksek kesinlikle tespit ettiğine işaret etti. Aktivasyon alanı oranının 0.309 olması ise modelin karar için görüntünün yaklaşık %30.9'lük bir bölümünü kullandığını ve oldukça odaklanmış bir inceleme yaptığını göstermektedir.

5. TARTIŞMA

Bu çalışma, tiroid nodülü sınıflandırması gibi kritik bir tıbbi görüntüleme problemini ele alarak, üç farklı derin öğrenme paradigmasını (CNN, Transformer ve Hibrit) kapsayan on bir modern mimarinin kapsamlı bir karşılaştırmasını sunmuştur. Elde edilen deneysel bulgular, hibrit bir mimari olan Convolutional Vision Transformer (ConViT) modelinin, tüm değerlendirilen performans metriklerinde tutarlı bir şekilde en üstün başarıyı sergilediğini ortaya koymaktadır. ConViT'in %83.3'lük doğruluk, kesinlik, duyarlılık ve F1-Skoru ile %89.94'lük ROC-AUC değeri, bu mimarinin tiroid ultrasonografi görüntülerinin analizinde sahip olduğu olağanüstü potansiyelin bir göstergesidir. Bu sonuç, hibrit mimarilerin, evrimsel operasyonların yerel özellik çıkarımındaki gücü ile transformerların küresel bağlam modelleme kapasitesini birleştirerek, tıbbi görüntü analizinde başarılı bir performans seviyesi belirleyebileceğini güçlü bir şekilde desteklemektedir.

ConViT'in bu başarısının arkasındaki temel akıl yürütme, mimarinin iki dünyanın en iyi yanlarını sentezlemesidir. Evrimsel katmanlar, görüntülerdeki yerel desenler, dokular ve sınırlar gibi düşük seviyeli özellikleri yakalamada son derece verimlidir; bu özellikler tiroid nodüllerinin değerlendirilmesinde (örneğin, mikrolobüle veya belirsiz kenar varlığı) hayati öneme sahiptir. Transformer bileşeni ise, bu yerel özellikler arasındaki uzun menzilli bağımlılıkları ve global bağlamı modelleyerek, görüntünün farklı bölgelerindeki ilişkileri anlamlandırabilir. Bu simbiyotik ilişki, modelin sadece küçük bir nodül lezyonunu değil, aynı zamanda bu lezyonun çevresindeki doku yapısıyla olan etkileşimini de değerlendirebilmesine olanak tanır, bu da daha doğru ve güvenilir bir sınıflandırma sağlar.

En yüksek performansı gösteren ConViT modelinin başarısı, literatürdeki son eğilimlerle uyumludur. ConViT, evrimsel operasyonların yerel özellik çıkarımındaki üstünlüğü ile Transformer'ların küresel bağlamı modelleme kapasitesini birleştirir. Bu hibrit yaklaşım, tıbbi görüntülemede hem ince ayrıntıların hem de genel yapısal bağlamın kritik öneme sahip olduğu göz önüne alındığında özellikle avantajlıdır. Benzer şekilde, farklı tıbbi görüntüleme alanlarında yapılan çalışmalar da hibrit mimarilerin saf CNN veya Dönüştürücü modellerine kıyasla daha iyi performans gösterebileceğini öne sürmektedir (41, 51). Bizim çalışmamızda ConViT'in gösterdiği üstün performans, bu

bulguyu destekler niteliktedir ve hibrit mimarilerin tiroid ultrasonografi analizinde umut vaat ettiğini göstermektedir.

Çalışmamızda, saf Transformer mimarileri (DeiT, PVTv2, XCiT) genel olarak geleneksel ve modern CNN mimarilerinden (ResNet-50, EfficientNet) daha iyi performans göstermiştir. Bu, ViT'lerin yeterli veri ile eğitildiğinde güçlü özellik temsili kapasitesini doğrular (8). Ancak, en iyi performans gösteren saf dönüştürücü modeli (DeiT) bile en iyi hibrit modele (ConViT) kıyasla ~%2.1 daha düşük doğruluğa sahiptir. Bu, yerel bağlamsal bilginin önemini vurgulamaktadır. Modern bir CNN mimarisi olan ConvNeXt, DeiT dahil olmak üzere tüm saf dönüştürücü mimarilerini geride bırakarak ikinci sırada yer almıştır. ConvNext, görüntü sınıflandırma görevlerinde performansı artırmak için transformatör tasarım ilkelerinden ilham alan evrişimli sinir ağı (CNN) mimarisinde önemli bir evrimi temsil eder. Bu modernize edilmiş ResNet mimarisi, görme transformatörleri ile etkili bir şekilde rekabet etmesini sağlayan özellikleri entegre ederek tıbbi görüntüleme ve birkaç çekim öğrenme dahil olmak üzere çeşitli uygulamalarda üstün doğruluk elde eder. ConvNext, verimli kanal dikkati ve değiştirilmiş artık bloklar gibi transformatörden ilham alan özellikleri içererek hesaplama verimliliğini ve performansını artırır. ConvNext, ImageNet gibi veri kümelerinde geleneksel CNN'lerden ve görme dönüştürücülerinden daha iyi performans gösteren son teknoloji sonuçları göstermiştir (57, 58). Bu sonuç, doğru tasarım ve modernizasyonla CNN mimarilerinin hâlâ son derece rekabetçi olabildiğini ve Vision Transformer'lar karşısında geçerliliklerini koruduğunu göstermektedir. Bu bulgu, Radosav ve diğerlerinin (2022) ConvNeXt'in genel computer vision görevlerindeki etkinliğine ilişkin savını tıbbi görüntüleme alanında da desteklemektedir.

Çalışmanın bir diğer çarpıcı bulgusu, geleneksel CNN mimarilerinin hâlâ son derece rekabetçi olabilmesidir. Saf transformer mimarilerinin çoğunu geride bırakarak ikinci sırada yer alan ConvNeXt modeli, %82.63 doğruluk ve tüm modeller arasında en yüksek ROC-AUC değeri olan %90.02 ile dikkat çekicidir. Bu durum, modernize edilmiş ve iyileştirilmiş evrişimsel mimarilerin, özellikle nispeten daha küçük ölçekli tıbbi görüntü veri setlerinde, karmaşık dikkat mekanizmalarına sahip modellere kıyasla avantaj sağlayabileceğini düşündürmektedir. Transformer modellerinin genellikle çok büyük veri kümeleri üzerinde eğitildiklerinde en yüksek performansı sergilemeleri göz önüne alındığında, ConvNeXt'in bu başarısı, veri hacmi sınırlı olduğunda özenle tasarlanmış CNN'lerin hâlâ en güçlü seçeneklerden biri olabileceğine işaret etmektedir.

Performans metriklerinin tüm modellerde (örneğin, Doğruluk, Kesinlik, Duyarlılık, F1-Skoru) birbirine çok yakın değerler göstermesi, modellerin sınıflandırma performansının dengeli olduğunu ve belirgin bir sınıf bazlı yanlılık (bias) göstermediğini kanıtlamaktadır. Bu dengenin en somut kanıtı, ConViT modelinin sınıf bazlı metriklerinde görülebilir. Hasta sınıfı için %83.6 Kesinlik / %84.3 Duyarlılık ve Sağlıklı sınıfı için %83.0 Kesinlik / %82.2 Duyarlılık. Bu dengeli performans, klinik uygulamada son derece önemlidir çünkü hem yanlış negatif (hasta bir vakayı gözden kaçırma) hem de yanlış pozitif (sağlıklı bir vakaya gereksiz invazif işlem yapılmasına neden olma) hataları minimize edilmiş olur.

ROC-AUC değerlerinin tüm modellerde doğruluk değerlerinden yaklaşık %5-10 daha yüksek olması, modellerin genel ayırt edicilik gücünün yüksek olduğunu gösterir. AUC değeri, modelin pozitif ve negatif sınıfları birbirinden ne kadar iyi ayırt edebildiğinin eşik bağımsız bir ölçüsüdür. 0.90'a yakın bir AUC değeri (ConViT ve ConvNeXt için), modelin mükemmele çok yakın bir sınıflandırma yeteneği olduğu şeklinde yorumlanabilir. Grad-CAM gibi görselleştirme teknikleriyle desteklendiğinde, bu yüksek ayırt edicilik gücünün modelin kararını, radyologların odaklandığı anatomik olarak anlamlı bölgelerden (patolojik bulguların yoğun olduğu alanlar) aldığı gösterilmiştir. Bu "açıklanabilir yapay zeka" (Explainable AI - XAI) yaklaşımı, model çıktılarına olan güveni artırarak klinisyenler tarafından bir "ikinci görüş" veya karar destek aracı olarak benimsenmesinin önünü açabilecek kritik bir unsurdur.

Son olarak, eğitim sürecinin analizi (Kayıp ve Doğruluk grafikleri), ConViT modelinin genelleme performansının yüksek olduğunu göstermiştir. Eğitim ve doğrulama eğrileri arasında makul bir uyum olması ve doğrulama doğruluğunun yüksek seviyelerde seyretmesi, modelin eğitim verilerine aşırı öğrenme (overfitting) yapmadığını ve görülmemiş verilere iyi genelleme yapabildiğini kanıtlamaktadır.

6. SONUÇ VE ÖNERİLER

6.1. Sonuç

Bu çalışma, tiroid ultrason görüntülerinin sınıflandırılması amacıyla on bir derin öğrenme modelinin sistematik bir karşılaştırmasını gerçekleştirmiştir. Yapılan kapsamlı değerlendirmeler neticesinde, ConViT hibrit mimarisinin, %83.3'lük bir doğruluk ve %89.94'lük AUC değeri ile tüm rakip mimarilere (CNN, Transformer ve diğer hibrit modeller) kıyasla en üstün performansı sergilediği kesin olarak tespit edilmiştir. Bu sonuç, hibrit yaklaşımların tıbbi görüntülemedeki üstünlüğünü vurgulamakta ve küresel bağlam anlayışı ile yerel özellik çıkarımının sinerjik birleşiminin bu başarının temel nedeni olduğunu ortaya koymaktadır. Ayrıca, ConvNeXt ve DeiT gibi modellerin de yüksek ve dengeli performansları, bu mimarilerin de pratik uygulamalar için güçlü adaylar olduğunu göstermiştir. Modellerin sınıflar arasında dengeli bir performans sergilemesi ve Grad-CAM ile doğrulanan odaklanma becerileri, geliştirilen modellerin klinik ortamlarda bir karar destek sistemi olarak kullanılma potansiyelini son derece güçlü kılmaktadır.

6.2. Öneriler

Elde edilen bu umut verici sonuçlar ışığında, gelecek çalışmalar için aşağıdaki öneriler sunulabilir.

Veri Setinin Genişletilmesi ve Çeşitlendirilmesi: Mevcut çalışma, model performanslarını güvenilir bir şekilde değerlendirmek için yeterli olsa da, daha büyük ve daha çeşitli bir veri seti (farklı cihazlardan, farklı hasta popülasyonlarından ve daha fazla nodül alt tipini içeren) üzerinde eğitim, modellerin genelleme yeteneğini daha da artıracak ve klinik geçerliliğini güçlendirecektir.

Model Optimizasyonu ve Hiperparametre Ayarı: ConViT ve ConvNeXt gibi en başarılı modeller üzerinde daha kapsamlı hiperparametre optimizasyonu (öğrenme oranı zamanlamaları, farklı optimizasyon algoritmaları, veri artırma teknikleri) ve model mimarisi ince ayarı (katman derinliği, dikkat başlığı sayısı) yapılarak performansın daha da yukarı taşınabileceği düşünülmektedir.

Açıklanabilir Yapay Zeka (XAI) Üzerine Odaklanma: Grad-CAM gibi yöntemlerin ötesinde, Layer-wise Relevance Propagation (LRP) veya SHAP gibi daha gelişmiş XAI tekniklerinin entegrasyonu, modellerin karar süreçlerini daha şeffaf hale

getirecek ve radyologlar ile model arasındaki güven ve iş birliğini güçlendirecektir. Modelin güven skoru ile XAI çıktılarının tutarlılığı üzerine nicel çalışmalar yapılabilir.

Klinik Entegrasyon için Çalışmalar: Bir sonraki aşama, bu modelleri gerçek zamanlı bir klinik iş akışına entegre etmek üzere bir prototip yazılım veya eklenti geliştirmek olmalıdır. Bu tür bir sistemin, radyologların tanı doğruluğunu ve iş yükü hızını ne ölçüde artırdığını ölçmek için prospektif klinik çalışmalar yapılması, bu teknolojinin nihai kabulü için en kritik adım olacaktır.

Çoklu-Modalite ve Çoklu-Sınıf Yaklaşımı: Gelecek çalışmalar, yalnızca B-mod ultrason yerine, Doppler ultrasonografi veya sonoelastografi gibi farklı modaliteleri de içerecek şekilde genişletilebilir. Ayrıca, ikili sınıflandırma (sağlıklı/hasta) yerine, TI-RADS veya Bethesda gibi standart sınıflandırma sistemlerine uygun çoklu sınıf sınıflandırması (örneğin, benign, malign, şüpheli) üzerine odaklanması, klinik uygulanabilirliği doğrudan artıracaktır.



KAYNAKLAR

1. Abraham S, Joseph S. Medical Imaging and Artificial Intelligence: Transforming the Nature of Diagnostics and Treatment. Intelligent Systems and IoT Applications in Clinical Health: IGI Global; 2025. p. 127-58.
2. Gejji S, Supreeth B, Khan PF, Venkat S, Pradeepa K, Kaliappan S, editors. Deep Learning Approaches for Solving Challenges in Computer Vision Applications and Exploring Evolving Patterns. 2024 7th International Conference on Contemporary Computing and Informatics (IC3I); 2024: IEEE.
3. Rangel G, Cuevas-Tello JC, Nunez-Varela J, Puente C, Silva-Trujillo AG. A survey on convolutional neural networks and their performance limitations in image recognition tasks. Journal of sensors. 2024;2024(1):2797320.
4. Naseer MM, Ranasinghe K, Khan SH, Hayat M, Shahbaz Khan F, Yang M-H. Intriguing properties of vision transformers. Advances in Neural Information Processing Systems. 2021;34:23296-308.
5. Galić I, Habijan M, Leventić H, Romić K. Machine learning empowering personalized medicine: A comprehensive review of medical image analysis methods. Electronics. 2023;12(21):4411.
6. Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sens. 2021;13:4712.
7. Wei J, Chen J, Wang Y, Luo H, Li W. Improved deep learning image classification algorithm based on Swin Transformer V2. PeerJ Computer Science. 2023;9.
8. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020.
9. Elharrouss O, Himeur Y, Mahmood Y, Alrabaei S, Ouamane A, Bensaali F, et al. ViTs as backbones: Leveraging vision transformers for feature extraction. Inf Fusion. 2025;118:102951.
10. Khan A, Sohail A, Zahoor U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review. 2019;53:5455-516.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.

12. Lee C-Y, Gallagher PW, Tu Z, editors. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. Artificial intelligence and statistics; 2016: PMLR.
13. Sharma S, Sharma S, Athaiya A. Activation functions in neural networks. Towards Data Sci. 2017;6(12):310-6.
14. Rawat W, Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural computation. 2017;29(9):2352-449.
15. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.
16. Bhatt D, Patel C, Talsania H, Patel J, Vaghela R, Pandya S, et al. CNN variants for computer vision: History, architecture, application, challenges and future scope. Electronics. 2021;10(20):2470.
17. Zhang H, Duan J, Xue M, Song J, Sun L, Song M, editors. Bootstrapping ViTs: Towards liberating vision transformers from pre-training. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022.
18. Li Y, Liu L-P. Incorporating Inductive Biases to Energy-based Generative Models. 2025.
19. Hajirahimi Z, Khashei M. Weighted sequential hybrid approaches for time series forecasting. Physica A: Statistical Mechanics and its Applications. 2019.
20. Thakur N, Karmakar S, Shrivastava R. Hybrid deep learning algorithms for forecasting air quality index using dimension reduction technique in search of precise results. International Journal of Information Technology. 2023;15:3181-7.
21. Adamuthe A. Improved Text Classification using Long Short-Term Memory and Word Embedding Technique. International Journal of Hybrid Information Technology. 2020.
22. Wang Z, Ren H, Lu R, Huang L, editors. Stacking based lightgbm-catboost-randomforest algorithm and its application in big data modeling. 2022 4th international conference on data-driven optimization of complex systems (DOCS); 2022: IEEE.
23. Tolk A, editor Hybrid Modeling Integrating Artificial Intelligence and Modeling & Simulation Paradigms. 2024 Winter Simulation Conference (WSC); 2024 15-18 Dec. 2024.

24. Sharma D, Garg A, Kumar A. Ensemble learning in machine learning: Integrating multiple models for improved predictions. *International Journal of Applied Research*. 2018.
25. Zhang S, Fan R, Liu Y, Chen S, Liu Q, Zeng W. Applications of transformer-based language models in bioinformatics: a survey. *Bioinformatics Advances*. 2023;3(1):vbad001.
26. Wang P, Fan E, Wang P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern recognition letters*. 2021;141:61-7.
27. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*; 2017.
28. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30.
29. Rasool M, Ismail N, Boulila W, Ammar A, Samma H, Yafooz W, et al. A Hybrid Deep Learning Model for Brain Tumour Classification. *Entropy*. 2022;24.
30. Haruna Y, Qin S, Adama Chukkol AH, Yusuf AA, Bello I, Lawan A. Exploring the synergies of hybrid convolutional neural network and Vision Transformer architectures for computer vision: A survey. *Engineering Applications of Artificial Intelligence*. 2025;144:110057.
31. Chai B. Research on the performance of hybrid vision models based on ViT. *Applied and Computational Engineering*. 2024.
32. Yu T, He S, Song Y-Z, Xiang T. Hybrid Graph Neural Networks for Few-Shot Learning. *ArXiv*. 2021;abs/2112.06538.
33. Sun C, Li C, Lin X, Zheng T, Meng F, Rui X, et al. Attention-based graph neural networks: a survey. *Artificial Intelligence Review*. 2023;56(2):2263-310.
34. Yu Y, Si X, Hu C, Zhang J-X. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*. 2019;31:1235-70.
35. Tan J, Ramos A, Abante M, Tadeo R, Lansigan R. A Performance Review of Recurrent Neural Networks Long Short-Term Memory (LSTM). 2022 3rd International Conference for Emerging Technology (INCET). 2022:1-5.
36. Zhou X, Liang W, Wang K, Yang L. Deep Correlation Mining Based on Hierarchical Hybrid Networks for Heterogeneous Big Data Recommendations. *IEEE Transactions on Computational Social Systems*. 2021;8:171-8.

37. Salakhutdinov R, Tenenbaum J, Torralba A. Learning with Hierarchical-Deep Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013;35:1958-71.
38. Seera M, Lim CP. A hybrid intelligent system for medical data classification. *Expert Systems with Applications*. 2014;41(5):2239-49.
39. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H, editors. Training data-efficient image transformers & distillation through attention. *International conference on machine learning*; 2021: PMLR.
40. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S, editors. End-to-end object detection with transformers. *European conference on computer vision*; 2020: Springer.
41. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:210204306*. 2021.
42. Chen Y, Dai X, Chen D, Liu M, Dong X, Yuan L, et al., editors. Mobile-former: Bridging mobilenet and transformer. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2022.
43. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*. 2009;22(10):1345-59.
44. Caruana R. Multitask learning. *Machine learning*. 1997;28(1):41-75.
45. Passalis N, Tzelepi M, Tefas A. Knowledge distillation. *Deep Learning for Robot Perception and Cognition: Elsevier*; 2022. p. 165-86.
46. Thyroid Cancer Ultrasound Dataset. Available from: <https://www.kaggle.com/datasets/diveshzz/thyroid-cancer-classification-ultrasound-dataset/data>. Son Erişim Tarihi 10 Mayıs 2025
47. Zhang H, Liu Q, Han X, Niu L, Sun W. TN5000: An Ultrasound Image Dataset for Thyroid Nodule Detection and Classification. *Scientific Data*. 2025;12(1):1437.
48. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016.
49. Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning*; 2019: PMLR.

50. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S, editors. A convnet for the 2020s. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022.
51. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al., editors. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision; 2021.
52. Wang W, Xie E, Li X, Fan D-P, Song K, Liang D, et al. Pvt v2: Improved baselines with pyramid vision transformer. Computational visual media. 2022;8(3):415-24.
53. Ali A, Touvron H, Caron M, Bojanowski P, Douze M, Joulin A, et al. Xcit: Cross-covariance image transformers. Advances in neural information processing systems. 2021;34:20014-27.
54. d'Ascoli S, Touvron H, Leavitt ML, Morcos AS, Biroli G, Sagun L, editors. Convit: Improving vision transformers with soft convolutional inductive biases. International conference on machine learning; 2021: PMLR.
55. Xu W, Xu Y, Chang T, Tu Z, editors. Co-scale conv-attentional image transformers. Proceedings of the IEEE/CVF international conference on computer vision; 2021.
56. Mehta S, Rastegari M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:211002178. 2021.
57. Todi A, Narula N, Sharma M, Gupta U, editors. Convnext: A contemporary architecture for convolutional neural networks for image classification. 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT); 2023: IEEE.
58. Lao C-X, Tsoi AC, Bugiolacchi R. ConvNeXt-ECA: An Effective Encoder Network for Few-Shot Learning. IEEE Access. 2024.

EKLER

EK-1. Özgeçmiş



EK-2. Etik Kurul Onayı

