

Evrişimsel Sinir Ağları Ve Transformer Mimarilerinin Akciğer BT Görüntülerini Sınıflandırma Performans Analizi

Elif Tilki

Sude Nur Gülcan

Karun Acar

eliftlk34@gmail.com

sudenur.glcn@gmail.com

acarkarun@gmail.com



Danışman: Doç. Dr. Ayşe Berna Altınel Girgin

Giriş

Akciğer kanseri, dünyada en sık görülen ve en ölümcül kanser türlerinden biridir. BT görüntülerinin manuel analizi zaman alıcı ve hataya açık olabilir. Bu nedenle yapay zekâ destekli sistemlere ihtiyaç duyulmaktadır.

- Üç adet herkese açık veri setinin kullanımı
- CNN ve Transformer temelli modellerin karşılaştırılması
- Sınıflandırma metrikleri ve **T-testi** ile yapılan istatistiksel değerlendirme

Veri Setleri

IQ-OTH/NCCD: Toplam 1190 BT görüntüsü, 110 vakayı temsil eder [1].

Üç sınıf: Normal (416), Benign (120), Malignant (561).

SPIE-AAPM-NCI: Toplam 876 BT görüntüsü 65 vakayı

temsil eder [2].

iki sınıf: Benign (120), Malignant (561).

CT-Scan Images: Toplam 364 BT görüntüsü içerir [3].

iki sınıf: Cancerous (238), Non-cancerous (126).

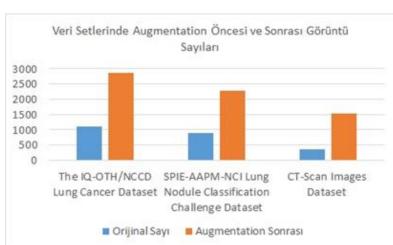
Şekil 1 Benign [2]

Çapraz Doğrulama ve Veri Artırımı

Katkılar

setleri, sınıf dağılımını Veri koruyarak 5-katlı Stratified K-Fold yöntemiyle bölünmüştür.

verilerine, Eğitim döndürme, kaydırma, ölçekleme, parlaklıkdeğişimi, kontrast Gauss gürültüsü, elastik bozulma gibi dönüşümler uygulanmıştır.

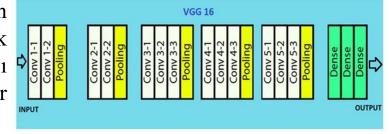


Modeller

VGG16

ResNet50

Oxford Visual Geometry Group tarafından geliştirilen CNN tabanlı bir modeldir. Küçük (3×3) konvolüsyon filtrelerini çok katmanlı olarak kullanarak hiyerarşik görsel özellikler öğrenir [4].



Şekil 2 VGG16 Katmanları [5]

CNN tabanlı bir model olup, artık bağlantılar sayesinde derin yapıda öğrenmeyi kolaylaştırır ve özellik çıkarımını güçlendirir. Beş aşamalı konvolüsyon bloklarından oluşur [6].

EfficientNetV2-S

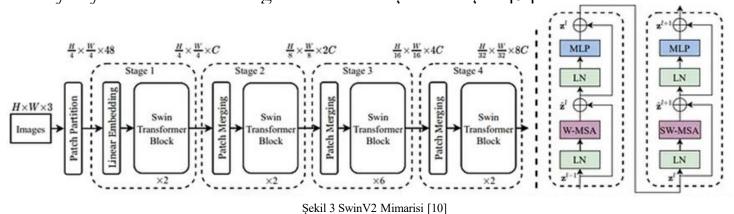
Gelişmiş ölçeklendirme yöntemleriyle tasarlanmış CNN tabanlı bir modeldir. Daha hızlı eğitim ve daha yüksek doğruluk için katman derinliği, genişliği ve çözünürlüğü optimize eder [7].

CoAtNet

CNN ve Transformer özelliklerini birleştiren hibrit bir modeldir. Evrişimsel genellemeyi self-attention ile küresel bilgi öğrenimiyle bütünleştirir [8].

SwinV2-CR-Small-224

Görsel tanımada yüksek başarı için geliştirilmiş bir Vision Transformer modelidir. Pencere kaydırmalı dikkat mekanizması ve çok ölçekli, hiyerarşik özellik çıkarımıyla yerel ve küresel bağlamları etkin şekilde işler [9].



BEiT-Base

Çift yönlü Vision Transformer modeli olup, BERT'in maskelenmiş dil modelleme yaklaşımıyla etiketlenmemiş verilerde görsel sembolleri maskeleyip tahmin ederek güçlü bağlamsal temsil öğrenir [11].

ConvNeXtV2

Meta AI tarafından geliştirilen, modern CNN mimarisinin güçlü bir evrişimsel katmanlardan yorumudur. Tamamen oluşarak Transformerların avantajlarını CNN yapısında yeniden tasarlar [12]. Şekil 4 ConvNeXt Mimarisi [13]

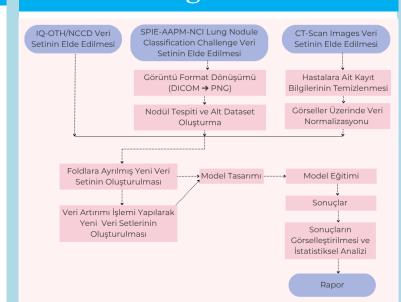
Model Eğitim Performansları

Tablo 1. Veri Artırımı Uygulanmamış T-Test Tablosu							
	VGG16	ResNet50	EfficientNetVS	SwinV2	BEiT-base	ConvNeXtV2	CoAtNet-1
VGG16	-	0/0/3	0/0/3	1/2/0	1/2/0	0/3/0	0/3/0
ResNet50	3/0/0	-	0/2/1	3/0/0	2/1/0	3/0/0	3/0/0
EfficientNetV2S	3/0/0	1/2/0	-	3/0/0	3/0/0	3/0/0	3/0/0
SwinV2	0/2/1	0/0/3	0/0/3	-	0/3/0	0/3/0	0/2/1
BEiT-base	0/2/1	0/1/2	0/0/3	0/3/0	-	0/3/0	0/2/1
CoNvneXtV2	0/3/0	0/0/3	0/0/3	0/3/0	0/3/0	-	0/3/0
CoAtNet-1	0/3/0	0/0/3	0/0/3	1/2/0	1/2/0	0/3/0	-
T. Kazanma	6	1	0	8	7	6	6
T. Beraberlik	10	3	2	10	11	12	10
T. Kaybetme	2	14	16	0	0	0	2

Tablo 3. Veri Artırımı Uygulanmamış Accuracy Sonuçları

	IQ-OTH/NCCD	SPIE-AAPM-NCI	CT-Scan	
VGG-16	0.9681	0.9989	0.9945	
ResNet50	0.8687	0.8733	0.9753	
EfficientNetV2S	0.8523	0.8242	0.9560	
SwinV2	0.9900	0.9989	0.9945	
BEiT-base	0.9900	0.9966	0.9890	
ConvNeXtV2	0.9809	1.000	0.9973	
CoAtNet-1	0.9754	1.000	0.9945	

Model Eğitim Süreci



Tablo 2. Veri Artırımı Uygulanmış T-Test Tablosu

	VGG16	ResNet50	EfficientNetV2S	SwinV2	BEiT-base	ConvNeXtV2	CoAtNet-1
VGG16	-	2/1/0	3/0/0	0/3/0	0/3/0	0/2/1	1/2/0
ResNet50	0/1/2	-	2/1/0	0/1/2	0/1/2	0/1/2	0/1/2
EfficientNetV2S	0/0/3	0/1/2	-	0/0/3	0/0/3	0/0/3	0/0/3
SwinV2	0/3/0	2/1/0	3/0/0	-	1/2/0	0/3/0	2/1/0
BEiT-base	0/3/0	2/1/0	3/0/0	0/2/1	-	0/2/1	0/3/0
ConvNeXtV2	1/2/0	2/1/0	3/0/0	0/3/0	1/2/0	-	2/1/0
CoAtNet-1	0/2/1	2/1/0	3/0/0	0/1/2	0/3/0	0/1/2	-
T. Kazanma	6	2	0	8	5	9	5
T. Beraberlik	11	6	1	10	11	9	8
T. Kaybetme	1	10	17	0	2	0	5

Tablo 4. Veri Artırımı Uygulanmış Accuracy Sonuçları				
	IQ-OTH/NCCD	SPIE-AAPM-NCI	CT-Scan	
VGG-16	0.9818	1.000	0.9918	
ResNet50	0.8842	0.8973	0.9918	
EfficientNetV2S	0.8496	0.8425	0.9780	
SwinV2	0.9900	1.000	0.9973	
BEiT-base	0.9699	0.9989	0.9973	
ConvNeXtV2	0.9854	1.000	1.000	
CoAtNet-1	0.9654	0.9932	0.9945	

Şekil 5 Swinv2 IQ Şekil 6 Swinv2 SPIE-Şekil 7 Swinv2 CT-Scan OTH/NCCD AAPM-NCI Şekil 10 ConvNeXtV2 Şekil 9 ConvNeXtV2 Veri Şekil 8 ConvNeXtV2 Veri Veri Artırımı Uygulanmış Artırımı Uygulanmış IQ-Artırımı Uygulanmış

SPIE-AAPM-NCI

Sonuçlar

Yapılan analizler, Transformer tabanlı modeller (SwinV2, BEiT) ile hibrit/modern CNN mimarileri (CoAtNet-1, ConvNeXt V2) kullanan modellerin, geleneksel CNN'lere kıyasla daha yüksek ve tutarlı doğruluk oranları sağladığını ortaya koymuştur. SwinV2, veri artırımı uygulanmamış veri setlerinde sırasıyla %99.0, %99.89 ve %99.45 doğruluk oranlarına ulaşırken; veri artırımı yapılmış veri setlerinde en başarılı performansı %98.54, %100 ve %100 doğruluk oranlarıyla ConvNeXt V2 göstermiştir. Yapılan t-testi analizleri de bu gelişmiş modellerin VGG16, ResNet50 ve EfficientNetV2-S gibi geleneksel modellere kıyasla istatistiksel olarak anlamlı şekilde üstün olduğunu doğrulamaktadır. Veri artırımı, düşük performans sergileyen modellerin başarısını artırırken; güçlü modellerin doğruluğunu stabilize etmiş ve genel kararlılığı güçlendirmiştir. Bu bulgular, medikal görüntü sınıflandırmasında hem model mimarisinin hem de doğru veri artırımı stratejisinin kritik rol oynadığını göstermektedir.





OTH/NCCD





CT-Scan

