



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

Elif Tilki  
eliftlk34@gmail.com

Sude Nur Gülcan  
sudenur.glcen@gmail.com

Karun Acar  
acarkarun@gmail.com



**Danışman:** Doç. Dr. Ayşe Berna Altınel

## 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.

## 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]

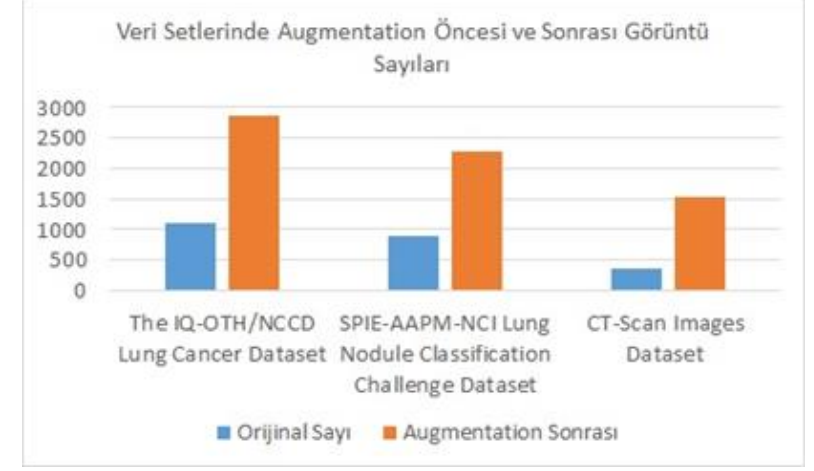
## Katkılar

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

## Çapraz Doğrulama ve Veri Artırımı

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

Eğitim verilerine, **döndürme, kaydırma, ölçekleme, parlaklık-kontrast değişimi, Gauss gürültüsü, elastik bozulma** gibi dönüşümler uygulanmıştır.



## Modeller

### VGG16

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]

### ResNet50

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

### EfficientNetV2-S

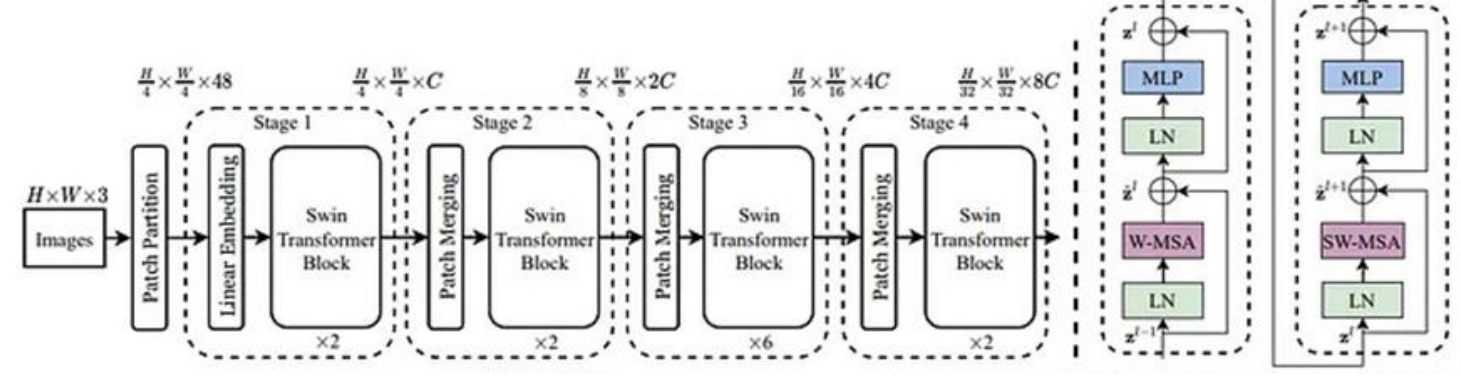
Gelişmiş ölçeklendirme yöntemleriyle tasarlanmış CNN tabanlı bir modeldir. Daha hızlı eğitim ve daha yüksek doğruluk için katman derinliği, genişliği ve çözünürlüğü optimize eder [7].

### CoAtNet

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

### SwinV2-CR-Small-224

Görsel tanımda 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].



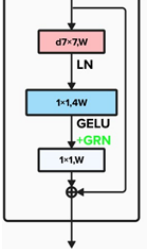
Şekil 3 SwinV2 Mimarisi [10]

### 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 yorumudur. Tamamen evrişimsel katmanlardan 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	EfficientNetV2S	SwinV2	BEiT-base	ConvNeXtV2	CoAtNet-1
<b>VGG16</b>	-	0/0/3	0/0/3	1/2/0	1/2/0	0/3/0	0/3/0
<b>ResNet50</b>	3/0/0	-	0/2/1	3/0/0	2/1/0	3/0/0	3/0/0
<b>EfficientNetV2S</b>	3/0/0	1/2/0	-	3/0/0	3/0/0	3/0/0	3/0/0
<b>SwinV2</b>	0/2/1	0/0/3	0/0/3	-	0/3/0	0/3/0	0/2/1
<b>BEiT-base</b>	0/2/1	0/1/2	0/0/3	0/3/0	-	0/3/0	0/2/1
<b>CoNvneXtV2</b>	0/3/0	0/0/3	0/0/3	0/3/0	0/3/0	-	0/3/0
<b>CoAtNet-1</b>	0/3/0	0/0/3	0/0/3	1/2/0	1/2/0	0/3/0	-
<b>T. Kazanma</b>	6	1	0	8	7	6	6
<b>T. Beraberlik</b>	10	3	2	10	11	12	10
<b>T. Kaybetme</b>	2	14	16	0	0	0	2

**Tablo 2.** Veri Artırımı Uygulanmış T-Test Tablosu

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

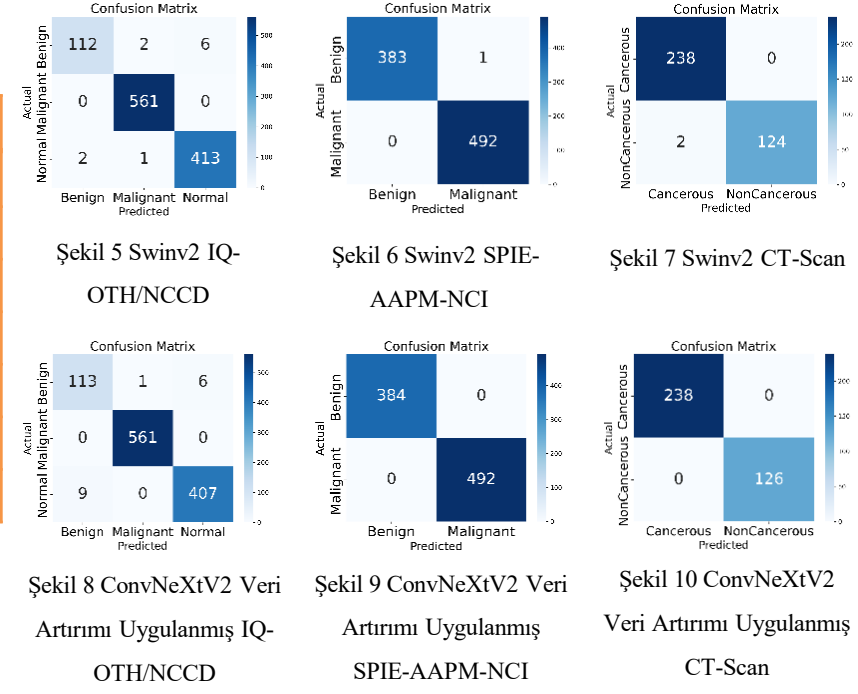
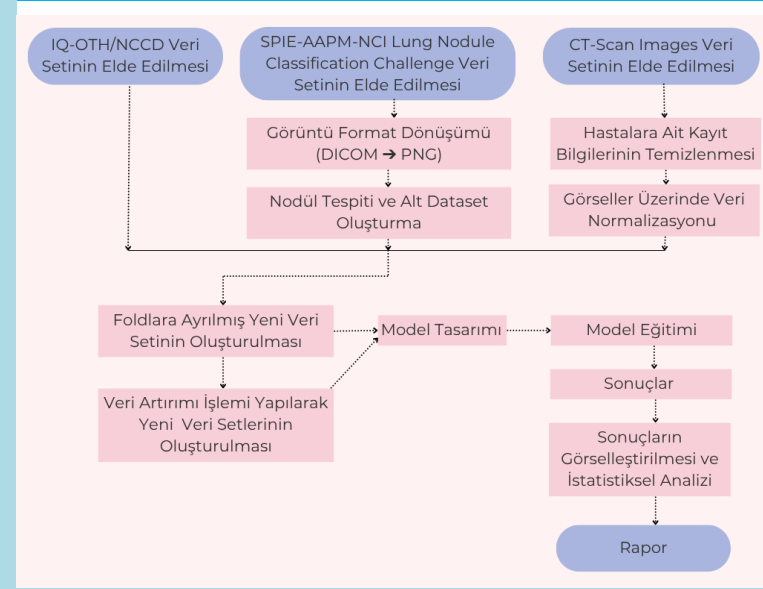
**Tablo 3.** Veri Artırımı Uygulanmamış Accuracy Sonuçları

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

**Tablo 4.** Veri Artırımı Uygulanmış Accuracy Sonuçları

	IQ-OTH/NCCD	SPIE-AAPM-NCI	CT-Scan
<b>VGG-16</b>	0.9818	1.000	0.9918
<b>ResNet50</b>	0.8842	0.8973	0.9918
<b>EfficientNetV2S</b>	0.8496	0.8425	0.9780
<b>SwinV2</b>	0.9900	1.000	0.9973
<b>BEiT-base</b>	0.9699	0.9989	0.9973
<b>ConvNeXtV2</b>	<b>0.9854</b>	<b>1.000</b>	<b>1.000</b>
<b>CoAtNet-1</b>	0.9654	0.9932	0.9945

## Model Eğitim Süreci



## 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.

## Referanslar

- [1] "The IQ-OTH/NCCD lung cancer dataset," <https://www.kaggle.com/datasets/banubalala/the-iq-oth-nccd-lung-cancer-dataset>
- [2] The "SPIE-AAPM-NCI Lung Nodule Classification Challenge Dataset," <https://www.consenting.org/nci-collection/spie-aapm-nci-challenge/> (accessed Jan. 14, 2025).
- [3] N. Masah, "CT-Scan Images," vol. 3, Jan. 2020, doi: <https://doi.org/10.13140/RG224024026263>.
- [4] K. Shomeyri and A. Zia, "Very Deep Convolutional Networks for Lung-Scale Image Recognition," [arXiv.org](https://arxiv.org/abs/1409.1258), Apr. 10, 2015, <https://arxiv.org/abs/1409.1258>.
- [5] M. Tanaka, K. Sugita, S. N. Alotaibi, and N. Z. Boudi, "Transfer Learning Approach for the Classification of Lung Cancers," *Healthcare*, vol. 10, no. 6, p. 1055, Jun. 2022, doi: <https://doi.org/10.3390/healthcare10061055>.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," [arXiv.org](https://arxiv.org/abs/1512.03385), Dec. 10, 2015, <https://arxiv.org/abs/1512.03385>.
- [7] M. Tanaka, Y. Li, "EfficientNetV2: Scalable Models and Fast Training," [arXiv.org](https://arxiv.org/abs/2104.00297), Jan. 2021, Available: <https://arxiv.org/abs/2104.00297>.
- [8] Z. Du, H. Lu, Q. V. Le, and M. Tan, "CoAtNet: Marrying Convolution and Attention for All Data Scales," [arXiv](https://arxiv.org/abs/2106.04801), Sep. 2021, Available: <https://arxiv.org/abs/2106.04801>.
- [9] Z. Du, et al., "Swin Transformer V2: Scaling Up Vision Transformer Blockwise," [arXiv](https://arxiv.org/abs/2106.01548), Apr. 2022, Available: <https://arxiv.org/abs/2106.01548>.
- [10] K. Shomeyri, "Very Deep Residual Learning for Lung-Scale Image Recognition," [arXiv](https://arxiv.org/abs/1409.1258), Apr. 10, 2015, <https://arxiv.org/abs/1409.1258>.
- [11] S. H. Tong, "RoBERTa: A Robustly Optimized BERT Pre-Trained Representation," *MedRxiv*, Sep. 12, 2021, <https://doi.org/10.1101/2021.09.12.21260000>.
- [12] S. H. Tong, "RoBERTa: A Robustly Optimized BERT Pre-Trained Representation," *MedRxiv*, Sep. 12, 2021, <https://doi.org/10.1101/2021.09.12.21260000>.
- [13] M. A. Rahman, A. Joy, A. T. Ahsan, and T. Shamsuzzaman, "Unlocking the Power of Open-Source Transformers in Medical Imaging: Insights from a Benchmark," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, p. 2024, Jul. 2024, doi: <https://doi.org/10.14569/IJACSA.2024.01.07120>.

## Kullanılan Teknolojiler

