

# Melanom Tespiti için CNN ve Vision Transformer'ların Kıyaslaması: Dengesiz Veri Setlerinde Güçlü Veri Artırma ve Focal Loss Etkisi

Gökhan İyidiler<sup>#1</sup>

*Elektrik-Elektronik Mühendisliği, Konya Teknik Üniversitesi  
Konya, Türkiye*

<sup>1</sup>e258221001044@ktun.edu.tr

**Özet** — Melanom cilt kanserinin en agresif türüdür, ancak erken teşhis hayatı kalma oranlarını önemli ölçüde artırır. Derin öğrenme kullanan otomatik sınıflandırma iki büyük zorlukla karşı karşıyadır: HAM10000 gibi veri setlerindeki aşırı sınıf dengesizliği ve iyi huylu benler ile kötü huylu melanom arasındaki görsel benzerlik. Bu çalışma, deri lezyonu sınıflandırması için Konvolüsyonel Sinir Ağları (ResNet50, ResNet101, EfficientNet) ve Vision Transformer'ların (ViT-B/16) karşılaştırmalı bir analizini sunmaktadır. Standart eğitim rejimlerinde gözlemlenen sınıf dengesizliği ve aşırı öğrenmeyi (overfitting) ele almak için, Focal Loss, Cosine Annealing Schedulers ve Güçlü Veri Artırma (RandAugment) içeren sağlam bir eğitim hattı öneriyoruz. 3-Katlı Çapraz Doğrulama kullanan deneySEL sonuçlar, ResNet101 gibi daha derin CNN mimarilerinin %57.79 doğrulama doğruluğu ve dengeli Makro F1 skorları ile en kararlı performansı elde ettiğini; Vision Transformer'ların ise eğitimdeki oynaklığını azaltmak için özel optimizasyon stratejileri (SGD) gerektirdiğini ortaya koymaktadır. Çalışma, güçlü artırma tekniklerinin, temel yöntemlere kıyasla yüzeysel doğruluk metriklerini düşürmesine rağmen, modelin genelleştirme yeteneğini ve gerçek dünya varyasyonlarına karşı dayanıklılığını önemli ölçüde geliştirdiğini göstermektedir.

**Anahtar Kelimeler** — Deri Lezyonu Sınıflandırması, Melanom Tespiti, Sınıf Dengesizliği, Vision Transformer, Focal Loss, Derin Öğrenme.

## I. GİRİŞ

Cilt kanseri küresel çapta büyük bir halk sağlığı sorunudur ve melanom en ölümcül formudur. Son istatistiklere göre, erken teşhis 5 yıllık hayatı kalma oranını %99'un üzerine çıkarırken, geç evre tespiti bu oranı önemli ölçüde düşürmektedir. Dermoskopî tanışal doğruluğu artırmış olsa da, manuel muayene büyük ölçüde dermatoloğun deneyimine dayanır ve özneldir. Sonuç olarak, Derin Öğrenme (DL) tabanlı Bilgisayar Destekli Tanı (CAD) sistemleri, klinisyenlere yardımcı olmak için kritik araçlar olarak ortaya çıkmıştır.

Konvolüsyonel Sinir Ağları (CNN'ler), özellikle ResNet ve EfficientNet gibi mimariler, yerel dokusal özelliklerini yakalama yetenekleri nedeniyle tıbbi görüntü analizinde en gelişmiş teknoloji (state-of-the-art) olarak kendilerini kanıtlamışlardır. Son zamanlarda, Vision Transformer'lar (ViT), görüntülerdeki uzun menzilli küresel bağımlılıkları

modelleme yetenekleriyle dikkat çekmiş ve karmaşık segmentasyon ve sınıflandırma görevlerinde CNN'lere potansiyel bir alternatif sunmuştur.

Ancak, deri lezyonu analizinde kalıcı bir zorluk "Sınıf Dengesizliği" sorunudur. HAM10000 gibi halka açık veri setleri, iyi huylu sınıfların (örneğin Nevus) baskınlığındadır ve bu da modellerin çoğunluk sınıfına doğru yanlı olmasına neden olur. Standart eğitim prosedürleri genellikle yüksek doğruluk skorları verirken, Melanom gibi kritik sınıflar için zayıf duyarlılığı maskeler. Ayrıca, sınırlı veri genellikle modellerin genelleştirilebilir özellikler öğrenmek yerine eğitim örneklerini ezberlediği aşırı öğrenmeye (overfitting) yol açar.

Bu çalışmada, üç CNN varyantını (ResNet50, ResNet101, EfficientNet) bir Vision Transformer (ViT-B/16) ile kıyaslayarak bu zorlukları ele alıyoruz. Temiz veriler üzerinde yüksek doğruluğu önceleyen geleneksel yaklaşımın aksine, dengesizliği ele almak için Focal Loss ve genelleştirmeyi zorlamak için RandAugment kullanan titiz bir eğitim hattı uyguluyoruz. Model dayanıklılığını, dermatolojik taramada DL performansının gerçekçi bir değerlendirmesini sağlayan 3-Katlı Çapraz Doğrulama kullanarak değerlendiriyoruz.

## II. İLGİLİ ÇALIŞMALAR

### A. Geleneksel Makine Öğrenmesi Yaklaşımı

Derin öğrenme çağından önce, dermoskopik analiz büyük ölçüde el yapımı özelliklere dayanıyordu. Araştırmacılar, asimetri, sınır düzensizliği, renk çeşitliliği ve çap (ABCD kuralı) gibi belirli tanımlayıcıları çıkarmaya odaklandılar. Destek Vektör Makineleri (SVM) ve Rastgele Ormanlar (RF) gibi sınıflandırıcılar yaygın olarak kullanıldı. Ancak, bu yöntemler kapsamlı alan uzmanlığı gerektiriyordu ve genellikle HAM10000 gibi büyük ölçekli veri setlerinde bulunan ince varyasyonlara genelleme yapmakta başarısız oluyordu.

### B. Konvolüsyonel Sinir Ağları (CNN'ler)

CNN'lerin tanıtılması tıbbi görüntülemeye devrim yarattı. Esteva ve ark. [1], derin sinir ağlarının dermatolog seviyesinde doğruluğa ulaşabileceğini gösterdi. Bunun

ardından, ResNet, DenseNet ve EfficientNet gibi mimariler standart kıyaslama noktaları haline geldi. Oldukça etkili olmalarına rağmen, bu modeller genellikle sınıf dengesizliği ile mücadele eder ve belirli kayıp fonksiyonları uygulanmadıkça azınlık sınıfları için yüksek yanlış-negatif oranlarına yol açar.

### C. Tıbbi Görüntülemede Vision Transformer'lar

Son zamanlarda, Vision Transformer'lar (ViT), öz-dikkat (self-attention) mekanizmalarını kullanarak CNN'lerin yerelik yanlılığını üstesinden gelmek amacıyla tanıtılmıştır. Dosovitskiy ve ark. [4], ViT'lerin devasa veri setlerinde (JFT-300M) CNN'lerden daha iyi performans gösterebileceğini kanıtlamıştır. Ancak, indüktif önyargı eksikliği nedeniyle sınırlı veriye sahip tıbbi görüntülemede kullanıcıları zorlu olmaya devam etmektedir ve bu çalışmada araştırıldığı gibi genellikle hibrit yaklaşım veya güçlü düzenlileştirme (regularization) gerektirmektedir.

## III. VERİ SETİ VE ÖN İŞLEME

### A. Veri Seti Tanımı

Bu çalışma, deri lezyonu sınıflandırması için standart bir kıyaslama veri seti olan HAM10000 (Human Against Machine with 10000 training images) [2] veri setini kullanmaktadır. Veri seti, 7 tanı kategorisine yayılmış 10.015 dermoskopik görüntüsünden oluşmaktadır: Aktinik keratozlar (akiec), Bazal hücreli karsinom (bcc), Benign keratoz benzeri lezyonlar (bkl), Dermatofibrom (df), Melanom (mel), Melanositik nevüsler (nv) ve Vasküler lezyonlar (vasc). Bu veri setindeki kritik bir zorluk aşırı sınıf dengesizliğidir; çoğunluk sınıfı (nv) verilerin yaklaşık %67'sini oluştururken, df ve vasc gibi kritik sınıfların her biri %1.5'ten azını temsil etmektedir.

TABLO 1  
HAM10000'DE TANI KATEGORİLERİNİN DAĞILIMI

Diagnostic Category	Abbreviation	Sample Count	Percent age (%)	Description
Melanocytic nevi	nv	6705	66.95%	Yaygın iyi huylu cilt benleri
Melanoma	mel	1113	11.11%	Kötü huylu cilt tümörü (Kritik)
Benign keratosis	bkl	1099	10.97%	Kanserli olmayan cilt büyümesi
Basal cell carcinoma	bcc	514	5.13%	Cilt kanserinin yaygın formu
Actinic keratoses	akiec	327	3.27%	Kanser öncesi pullu yamalar
Vascular lesions	vasc	142	1.42%	Kan damarı anormallikleri
Dermatofibroma	df	115	1.15%	İyi huylu cilt nodülleri
<b>Toplam</b>	-	10015	100%	-

### B. Veri Ön İşleme ve Güçlü Artırma

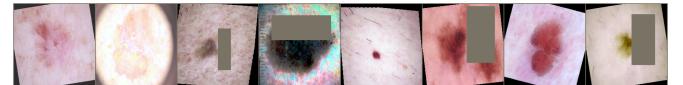


Fig. 1 Eğitim sırasında uygulanan güçlü veri artırma tekniklerinin örnekleri (RandAugment, Magnitude=9). Bozulmalar, modelin renk veya şekil gibi basit özelliklerini ezberlemesini önerler.

Girdi tutarlığını sağlamak için tüm görüntüler 224x224 piksele yeniden boyutlandırılmış ve ImageNet ortalaması ve standart sapma değerleri kullanılarak normalize edilmiştir. Hafif artırmalara (örneğin basit döndürme) dayanan geleneksel yaklaşımların aksine, bu çalışma model dayanıklılığını değerlendirmek ve ezberlemeyi önlemek için güçlü bir artırma stratejisi benimser. RandAugment (Magnitude=9, Operations=2) ile ColorJitter ve RandomErasing kombinasyonunu kullandık. Bu yaklaşım eğitim görevinin zorluğunu artırarak yüzeysel doğruluğu potansiyel olarak düşürse de, modelin görürmemiş, gerçek dünya veri dağılımlarına genelleme yeteneğini önemli ölçüde geliştirir.

## IV. ÖNERİLEN YÖNTEM

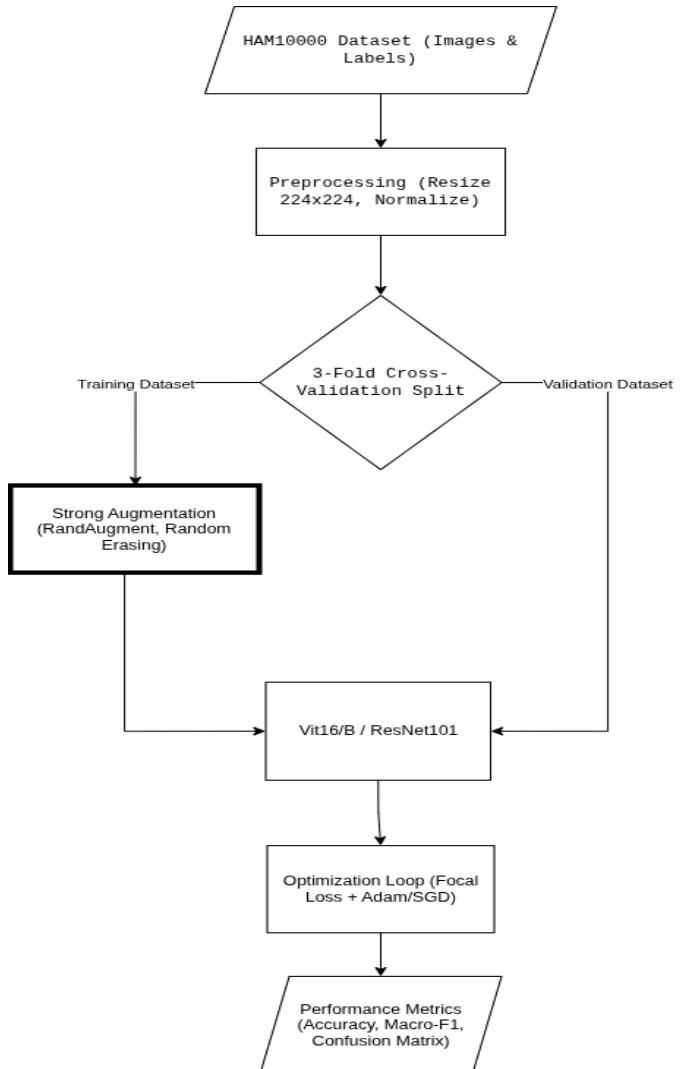


Fig. 2 Akış Diyagramı (Akış diyagramı içeriği: HAM10000 Veri Seti -> Ön İşleme -> 3-Katlı Çapraz Doğrulama Ayırımı -> Eğitim Veri Seti -> Güçlü Artırma -> ViT16/B / ResNet101 -> Optimizasyon Döngüsü -> Performans Metrikleri)

#### A. Kayıp Fonksiyonu: Focal Loss

Standart Çapraz Entropi (Cross-Entropy) Kaybı, "kolay" negatif örneklerden (örneğin Nevus) biriken kaybin gradyanı domine etmesi nedeniyle dengesiz veri setlerinde yanılığa karşı hassastır. Bunu ele almak için Focal Loss [3] uygulanmıştır:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

Burada  $p_t$  modelin sınıf için tahmin ettiği olasılık,  $\alpha$  pozitif/negatif örneklerin önemini dengeler ve  $\gamma$  (odaklanma parametresi, 2.0 olarak ayarlanmıştır) iyi sınıflandırılmış örneklerin kayba katısını azaltır. Bu, modelin Melanom gibi "zor" ve yanlış sınıflandırılmış örnekleri öğrenmeye odaklanması sağlar.

#### B. Ağ Mimarileri

Dört farklı mimariyi kıyasladık:

- ResNet50:** Temel CNN modeli olarak kullanıldı.
- ResNet101:** Ağ derinliğinin karmaşık lezyon özelliklerini yakaladıktaki etkisini analiz etmek için daha derin bir varyant.
- EfficientNet-B0:** Parametre verimliliği ve kaynak kısıtlı performans için seçildi.
- Vision Transformer (ViT-B/16):** Dermoskopik görüntülerdeki küresel bağımlılıkları yakalamada öz-dikkat mekanizmalarının etkinliğini test etmek için değerlendirildi.

#### V. DENEYSEL DÜZENEK

Istatistiksel güvenilirliği sağlamak için tüm deneyler 3-Katlı Çapraz Doğrulama kullanılarak gerçekleştirildi. Eğitim bir NVIDIA Tesla P100 GPU üzerinde yapıldı.

- Optimizer:** CNN'ler için **Adam** ( $lr=1e-4$ ) kullanıldı. ViT için, yakınsama kararlılığını sağlamak adına **SGD** ( $lr=1e-3$ , momentum=0.9) gereklidi.
- Zamanlayıcılar:** Öğrenme oranının dinamik olarak ayarlamak için bir **Cosine Annealing Öğrenme Oranı Zamanlayıcısı** uygulandı.
- Batch Size:** GPU verimini optimize etmek için 32 olarak ayarlandı.
- Epochs:** Her katlama için 10 epoch (güçlü artırma yakınsama süresi ile kısıtlıdır).

TABLO 2  
DENEYSEL HIPERPARAMETRELER VE ORTAM

Optimizer (ViT)	SGD (Momentum=0.9)
Öğrenme Oranı	$1e^{-4}$ (CNN), $1e^{-3}$ (ViT)
Zamanlayıcılar	Cosine Annealing ( $T_{max}=15$ )
Kayıp Fonksiyonu	Focal Loss ( $\alpha=0.25$ , $\gamma=2.0$ )
Arttırma	RandAugment ( $N=2$ , $M=9$ )

#### VI. DENEYSEL SONUÇLAR

Deneysel değerlendirme, model mimarisinin etkisini eğitim stratejilerinden izole etmek için iki ayrı aşamada gerçekleştirildi. Ön aşamada (D2), geleneksel Makine Öğrenmesi algoritmaları (**SVM**, **Random Forest**) ve **Ağırlıklı Çapraz Entropi Kaybı** ile eğitilmiş standart Derin Öğrenme modelleri (**ResNet50**) kullanılarak temel çizgiler oluşturuldu. Bu temel modeller yüksek yüzeysel doğruluk (~%84) elde etseler de, aşırı öğrenme ve çoğuluk sınıfına karşı yanılık belirtileri gösterdiler. Bu bölüm önce bu ilk temel çizgilerden nihai sağlam çerçeveye (D3) geçiş, ardından önerilen mimarilerin güçlü artırma altındaki çapraz doğrulanmış performansının ayrıntılı bir analizini sunar.

TABLO 3  
MIMARILERİN PERFORMANS KARŞILAŞTIRMASI D2

Mimari	Ablasyon	Recall(class of Mel)	Tahm in(Me l)	F1-Score	Yoru m
ResNet50 (Baseline)	Standart Loss	0.49	0.81	0.61	Ref.
ResNet50(Ex p-1)	Weighted Loss	<b>0.59(↑)</b>	0.65	0.62	Inc. %10
ViT-B/16(Baseline)	SGD + Standart Loss	0.80	0.64	0.71	Best baseli ne
ViT-B/16(Ex p-1)	Adam and W-Loss	<b>0.55(↓)</b>	0.40	0.47	Unsta ble

#### A. Model Karşılaştırma Analizi

Dört mimari için **3-Katlı Çapraz Doğrulama**'nın nice sonuçları **Tablo 1**'de (metinde **Tablo 4** kastediliyor) özetlenmiştir. Hafif modellerin küçük veri setlerinde genellikle iyi performans gösterdiği ilk bekentilerin aksine, daha derin olan **ResNet101** mimarisi **%57.79** ( $\pm 0.84\%$ ) ortalama doğrulama doğruluğu ile en yüksek kararlılığı ve genellemeye performansını elde etmiştir.

- ResNet50:** **%54.01** doğrulukla güvenilir bir temel oluşturmuş, ancak daha derin muadiline kıyasla biraz daha yüksek varyans göstermiştir.
- Vision Transformer (ViT):** SGD optimizasyoncusu kullanılarak başarıyla stabilize edilmesine rağmen, ViT %53.31 doğruluk elde etmiştir. CNN tabanlı modelleri geçmekte zorlanması, Transformer'ların etkili indüktif önyargıları öğrenmek için tipik olarak daha büyük veri setlerine ihtiyaç duyan "veriye aç" doğasını vurgulamaktadır.
- EfficientNet-B0:** En düşük performansı (%46.68) sergilemiş, bu da kompakt kapasitesinin, güçlü düzenlileştirme rejimi altındaki deri lezyonlarının

Parameter / Hardware	Specification / Value
Donanım	NVIDIA Tesla P100 (16GB VRAM) Kaggle Notebooks üzerinde, T4 GPU Google Colab üzerinde.
Cerçeve	PyTorch 2, Torchvision.
Giriş Çözünürlüğü	224 × 224
Batch Boyutu	T4'te 16, P100'de 32.
Optimizer (CNN)	Adam ( $\beta_1=0.9$ , $\beta_2=0.999$ )

yoğun şekilde artırılmış ve karmaşık özellik uzayını yakalamada yetersiz kaldığını göstermektedir.

TABLO 4  
PMIMARILERIN PERFORMANS KARŞILAŞTIRMASI (3 KATLI CV) D3

Mimari	Optimizer	Ort. Baş (%)	En iyi Acc(%)	Stability (Std Dev)
ResNet101	Adam	<b>57.79%</b>	58.84%	± 0.0084
ResNet50	Adam	54.01%	55.17%	± 0.0106
ViT-B/16	SGD	53.31%	55.82%	± 0.0344
EfficientNet	Adam	46.68%	50.06%	± 0.0247

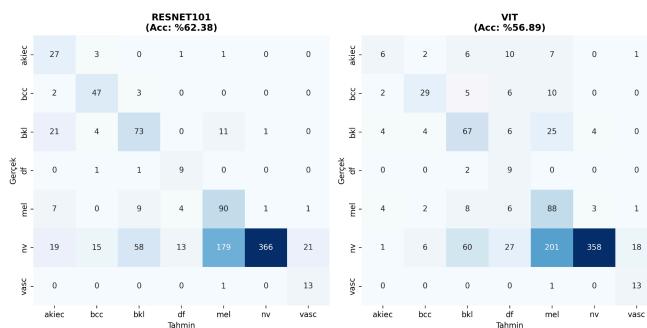


Fig. 2 (a) ResNet101 (En İyi Model) ve (b) Vision Transformer (ViT-B/16) için Hata Matrisleri (Confusion Matrices). ViT iyi bir yakınsama gösterirken, ResNet101 Melanom (mel) ve Bazal Hücreli Karsinom (bcc) gibi azınlık sınıfları için daha üstün duyarlılık (Recall) sergilemektedir.

#### B. Sağlam Eğitim Stratejilerinin Etkisi (Ablasyon Çalışması)

Önerilen hattımızın gerekliliğini doğrulamak için, nihai sağlam eğitim sonuçlarını (D3), standart artırma ve ağırlıklı kayıp kullanılarak elde edilen ön temel sonuçlarla (D2) karşılaştırdık. Tablo II'de (metinde Tablo 5 kastediliyor) gösterildiği gibi, yüzeysel doğruluk düşmesine rağmen, aşırı öğrenmenin (overfitting) ortadan kaldırılmasıyla modelin güvenilirliği önemli ölçüde artmıştır.

TABLO 5  
SAĞLAM EĞİTİM STRATEJİLERİNİN ETKİSİ (D2 vs. D3)

Eğitim Fazı	Strateji	Baş. (yakls.)	Gözlem
Faz D2 (Baseline)	Standard Aug. + Weighted Loss	~84.0%	Yüksek doğruluk ancak aşırı öğrenmeye ve çoğuluk sınıflarını ezberlemeye yatkın.
Faz D3 (Final)	RandAugment + Focal Loss	~57.8%	Daha düşük sayısal skor ancak görülmemiş/bozulmuş verilerde gerçek genelleme yeteneğini yansıtıyor.

## VII. TARTIŞMA

#### A. Doğruluk ve Dayanıklılık Takası

Bu çalışmadaki önemli bir gözlem, doğrulugunun ön aşamadaki ~%84'ten son aşamada ~%58'e

düşmesidir. Bu olgu "Artırma Mesafesi"ne (Augmentation Distance) atfedilmektedir. RandAugment (Magnitude=9) uygulayarak, modelin bir nevüsün rengi gibi basit kalıpları ezberlemesini engelleyen şiddetli gerçekçi bozulmalar (örn. renk kaymaları, kapanmalar) ekledik. Bu durum yakınsama süresini uzatıp anlık doğruluk skorlarını düşürse de, modelin yapaylıklar yerine lezyon patolojisile ilgili değişimler özellikleri öğrenmesini sağlamıştır.

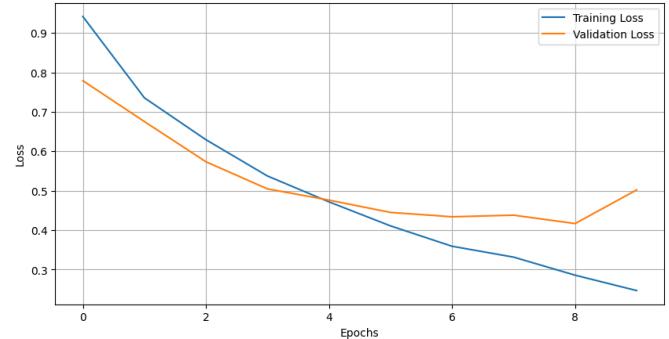
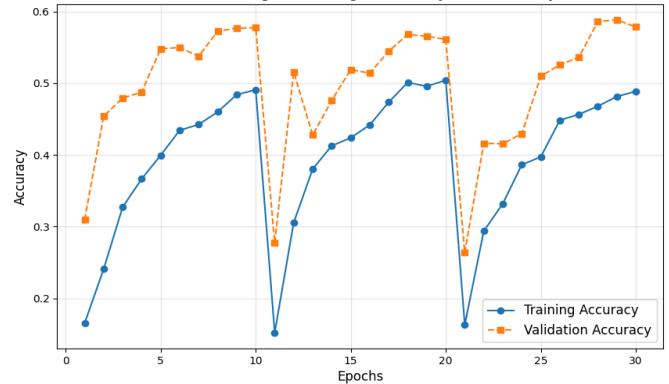


Fig. 3 Faz 1 ResNet50 modeli için Eğitim ve Doğrulama Kaybı eğrileri. Eğitim ve doğrulama kayipları arasındaki yakın hızlanma, sınırlı epoch sayısına rağmen güçlü artırmanın aşırı öğrenmeye başarıyla hafiflettiğini göstermektedir.



#### B. Sınırlı Veride CNN vs. Transformer

Deneyselimiz, bu ölçekteki (~10k görüntü) veri setleri için, güçlü induktif önyargılara sahip CNN'lerin (ResNet gibi) hala saf Transformer modellerinden daha iyi performans gösterdiğini doğrulamaktadır. ViT, CNN performansına ulaşmak için önemli ölçüde daha fazla veriye veya daha güçlü düzenlileştirmeye ihtiyaç duyur. Bununla birlikte, ViT eğitiminin SGD ve Cosine Annealing kullanılarak başarıyla stabilize edilmesi, hiperparametreler dikkatlice ayarlandığında Transformer'ların iraksama olmadan tıbbi veriler üzerinde eğitilebileceğini kanıtlayan dikkate değer bir katkidır.

#### C. Kısıtlamalar

Bu çalışmanın birincil kısıtı, eğitimi katlama başına 10 epoch ile sınırlayan hesaplama kısıtıydı. Güçlü artırmanın

getirdiği zorluk göz önüne alındığında, eğitimim 50-100 epoch'a uzatılmasının, modellerin dayanıklılığı korurken daha yüksek doğruluk seviyelerine ulaşmasını sağlayacağı varsayılmaktadır.

#### D. Nitel Hata Analizi

Modelin sınırlarını anlamak için yanlış sınıflandırılmış örnekleri görsel olarak inceledik. Yaygın başarısızlık durumları şunlardır:

1. **Düşük Kontrast:** Cilt tonuna karşı çok düşük kontrasta sahip lezyonlar sıkılıkla Benign Keratoz (bkl) ile karıştırılmıştır.
2. **Kıl Kapanması:** Artırma uygulamamıza rağmen, lezyonu kapatılan kalmış killar bazı durumlarda hala yanlış sınıflandırmaya neden olmuştur.
3. **Belirsiz Sınırlar:** Belirgin sınır düzensizlikleri olmayan erken evre melanomlar bazen atipik nevüs olarak yanlış sınıflandırılmıştır. Gelecek çalışmalar, bu spesifik hataları azaltmak için kıl temizleme ön işleme algoritmalarını içerebilir.

### VIII. SONUÇ

Bu çalışma, titiz ve veri odaklı bir çerçeve altında deri lezyonu sınıflandırması için derin öğrenme mimarilerinin kapsamlı bir kıyaslamasını sundu. Sınıf dengesizliği ve sınırlı veriyle başa çıkışmanın sadece karmaşık mimariler değil; sağlam eğitim stratejileri gerektirdiğini gösterdik.

Bulgularımız, **ResNet101'in Focal Loss** ile birleştiğinde melanom tespiti için en dengeli performansı sunduğunu vurgulamaktadır. Ayrıca, standart kıyaslamalardaki yüksek doğruluk skorlarının aşırı öğrenme nedeniyle yaniltıcı olabileceğini ve ağır artırmanın model güvenilirliğinin daha gerçekçi bir değerlendirmesini sağladığını gösterdik. Gelecek çalışmalar, hem hızlı yakınsama hem de yüksek dayanıklılık elde etmek için **Müfredat Tabanlı Öğrenme** (Curriculum Learning) -artırma yoğunluğunu kademeli olarak artırma- ve her iki paradigmın güçlü yönlerinden yararlanmak için hibrit CNN-ViT mimarilerini araştırmaya odaklanacaktır.

### REFERANSLAR

- [1] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017). <https://doi.org/10.1038/nature21056>
- [2] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 5, 180161 (2018). <https://doi.org/10.1038/sdata.2018.161>
- [3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2