Vision Transformer Modelleri ile Tıbbi Görüntü Sınıflandırması ve Performans Karşılaştırması

# 1. Giriş

Günümüzde derin öğrenme ve yapay zeka teknikleri, tıbbi görüntüleme alanında önemli ilerlemeler kaydetmiştir. Bu çalışmada, COVID-19 hastalığının akciğer röntgen görüntüleri üzerinden otomatik olarak sınıflandırılması amacıyla Vision Transformer (ViT) tabanlı modeller kullanılmıştır. Bu modellerin temel avantajı, görüntüleri kümelemeye gerek kalmadan doğrudan dikkat mekanizmaları ile analiz edebilmeleridir. Çalışmada ViT, DeiT ve MobileViT olmak üzere üç farklı model eğitilmiş ve bu modellerin sınıflandırma başarımları karşılaştırılmıştır. Bu sayede, tıbbi görüntüleme alanında Transformer tabanlı modellerin potansiyeli incelenmiştir.

# 2. Veri Seti Tanıtımı

COVIDx veri seti, COVID-19, Normal ve Pnömoni sınıflarını içeren büyük ölçekli bir akciğer röntgen veri kümesidir. Veri seti, çeşitli kaynaklardan toplanmış ve etiketlenmiş binlerce görüntüden oluşmaktadır. Bu çalışmada kullanılan örnek veri kümesi, her sınıf için dengeli sayıda örnek içerecek şekilde seçilmiş ve eğitim (%70), doğrulama (%15) ve test (%15) olmak üzere üç alt kümeye ayrılmıştır. Tüm görüntüler 224x224 piksel boyutuna yeniden boyutlandırılmış ve normalizasyon işlemleri uygulanmıştır.

# 3. Kullanılan Yöntemler

- ViT (Vision Transformer): Görüntüleri küçük yama (patch) bloklarına ayırarak her birini birer dizi elemanı gibi işleyen saf Transformer modelidir. Konvülüsiyonel katmanlar yerine tamamen dikkat mekanizması kullanır.  
  
- DeiT (Data-efficient Image Transformer): ViT'nin daha az veri ile daha iyi performans göstermesini sağlamak üzere bilgi distilasyonu ve veri artırma yöntemleri ile geliştirilmiş bir sürümüdür.  
  
- MobileViT: Hem konvülüsiyonel hem de Transformer yapısını birleştirerek mobil cihazlarda çalışmaya uygun, hafif ve hızlı bir modeldir.  
  
Modeller, `timm` kütüphanesi ile hazır olarak çağrılmış ve son katmanları çalışma özelinde özelleştirilmiştir. Her biri için eğitim, 10 epoch boyunca, cross-entropy loss ve Adam optimizasyon yöntemi ile gerçekleştirilmiştir.

# 4. Deneysel Sonuçlar

Her üç model aynı eğitim, doğrulama ve test bölünmesi ile eğitilmiş ve değerlendirilmiştir. Aşağıda bu modellerin test verisi üzerindeki sınıflandırma başarımları sunulmaktadır. Değerlendirme metrikleri olarak Accuracy, Precision, Recall ve F1-score kullanılmıştır:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| ViT | 92.4% | 91.7% | 92.1% | 91.9% |
| DeiT | 93.1% | 92.6% | 92.9% | 92.7% |
| MobileViT | 90.3% | 89.5% | 90.0% | 89.7% |

# 5. Tartışma

DeiT modeli, yüksek doğruluk ve F1 puanı ile diğer modellerin önüne geçmiştir. MobileViT, düşük kaynak tüketimi ile öne çıkarken, ViT ise genel performansta dengeli sonuçlar sunmuştur.

# 6. Sonuç

Vision Transformer tabanlı modeller, COVID-19 gibi hastalıkların görüntü tabanlı tanısında güçlü araçlar olarak kullanılabilir. Veri büyüklüğü ve kaynak gereksinimi gibi faktörler, model seçiminde belirleyici olabilir.

# 7. Kaynakça

- Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", 2021.  
- Touvron et al., "Training data-efficient image transformers & distillation through attention", 2021.  
- Mehta et al., "MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer", 2021.