1ST INTERNATIONAL PEDIATRIC EMERGENCY CONGRESS MALATYA
28 - 30 August 2026
Malatya Turgut Özal University İkizce Campus

SÖZLÜ SUNUM • NO: M4461

Yapay Zekâ Çocuk Travmasına Aktarılabiliyor mu?

Derin Öğrenme Tabanlı Kırık Sınıflandırmada Out-of-Domain Stress Test

When Internal Performance Does Not Transfer: An Out-of-Domain Stress Test of Deep Learning-Based Fracture Classification in Pediatric Wrist Trauma

Dr. Mustafa Bulut – Dr. M. Furkan Darılmaz
Aksaray Üniversitesi Eğitim ve Araştırma Hastanesi
Ortopedi ve Travmatoloji

Sunum Akışı

- Klinik problem
- Araştırma sorusu ve amaç
- Yöntem
- Bulgular
- Klinik çıkarım
- Eve götürülecek mesaj

Klinik Problem

Çocuk grafisi küçük erişkin grafisi değildir

- Açık fizisler
- Ossifikasyon merkezleri
- Yaşa bağlı anatomik değişiklikler
- Farklı kırık paternleri

Akla gelen büyük soru:

Yapay zeka modelleri kırığı mı öğreniyor, yoksa eğitim aldığı veri setini mi?

Araştırma Sorusu ve Amaç

Araştırma sorusu:
Erişkin ağırlıklı, çok bölgeli radyografilerde eğitilmiş bir kırık sınıflandırıcısı pediatrik el bileği travmasına genellenebilir mi?

Amaç:
FracAtlas üzerinde eğitilen üç ayrı ResNet18 stratejisinin GRAZPEDWRI-DX pediatrik veri setindeki davranışını değerlendirmek.

Bu bir klinik external validation değil; out-of-domain stress testtir (Alan dışı test) .

A. FracAtlas: non-fractured



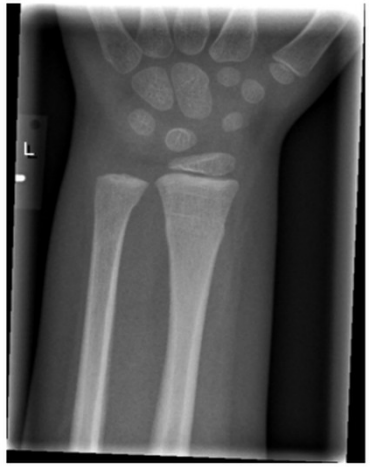
B. FracAtlas: fractured



C. GRAZPEDWRI-DX: non-fractured



D. GRAZPEDWRI-DX: fractured



FracAtlas → GRAZPEDWRI-DX

Değişenler:

- Yaş
- Anatomi
- İskelet maturitesi
- Görüntü özellikleri
- Preprocessing
- Label tanımları

Model başarısızlığının tek bir nedene bağlanması mümkün değildir.

Çalışma Tasarımı

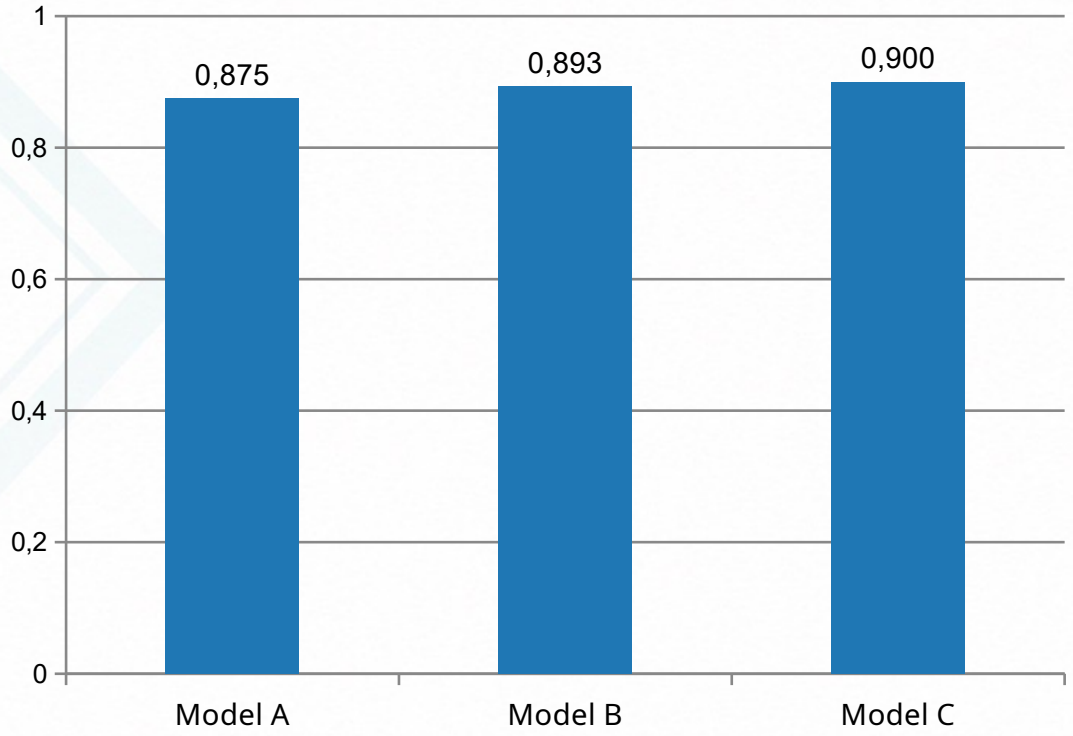
FracAtlas
4.024 radyografi
↓
ResNet18 model geliştirme
↓
Model A / B / C
↓
Internal test
↓
GRAZPEDWRI-DX
9.907 pediatrik el bileği radyografisi
↓
Out-of-domain stress test

5.440 non-fractured
4.467 fractured

Strateji	Augmentation	Class weighting	Threshold
Model A	Hayır	Hayır	0.500
Model B	Evet	Hayır	0.500
Model C	Evet	Evet	0.500
Model C – Youden	Evet	Evet	0.441

0.441 yeni bir model değildir; Model C'nin ikinci operating point'idir.

İç Doğrulama: FracAtlas üzerinde performans iyi görünüyor



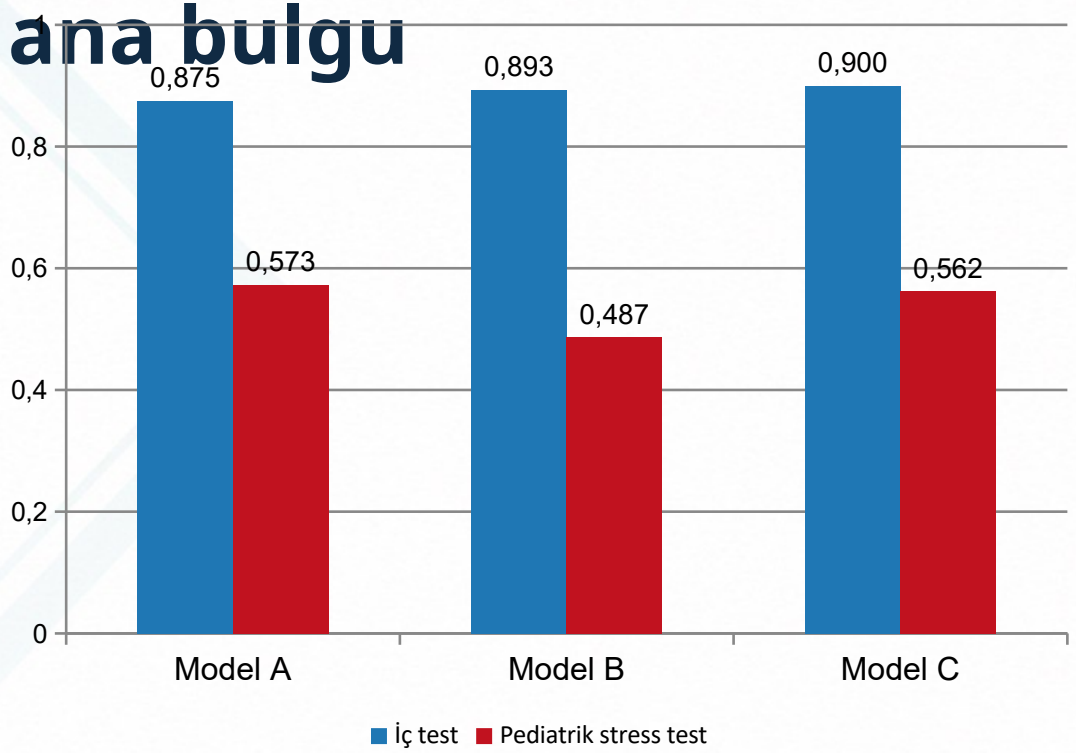
0.900
En yüksek iç AUROC
Model C

0.889
En yüksek iç accuracy
Model B

0.869
En yüksek iç sensitivity
Model C,
0.441

İç test sonuçları tek başına klinik transfer güvenilirliği göstermez.
Model C'de threshold değişimi AUROC'yi değiştirmez; sadece karar noktasını değiştirir

Pediatrik veri setinde AUROC düşüşü: ana bulgu

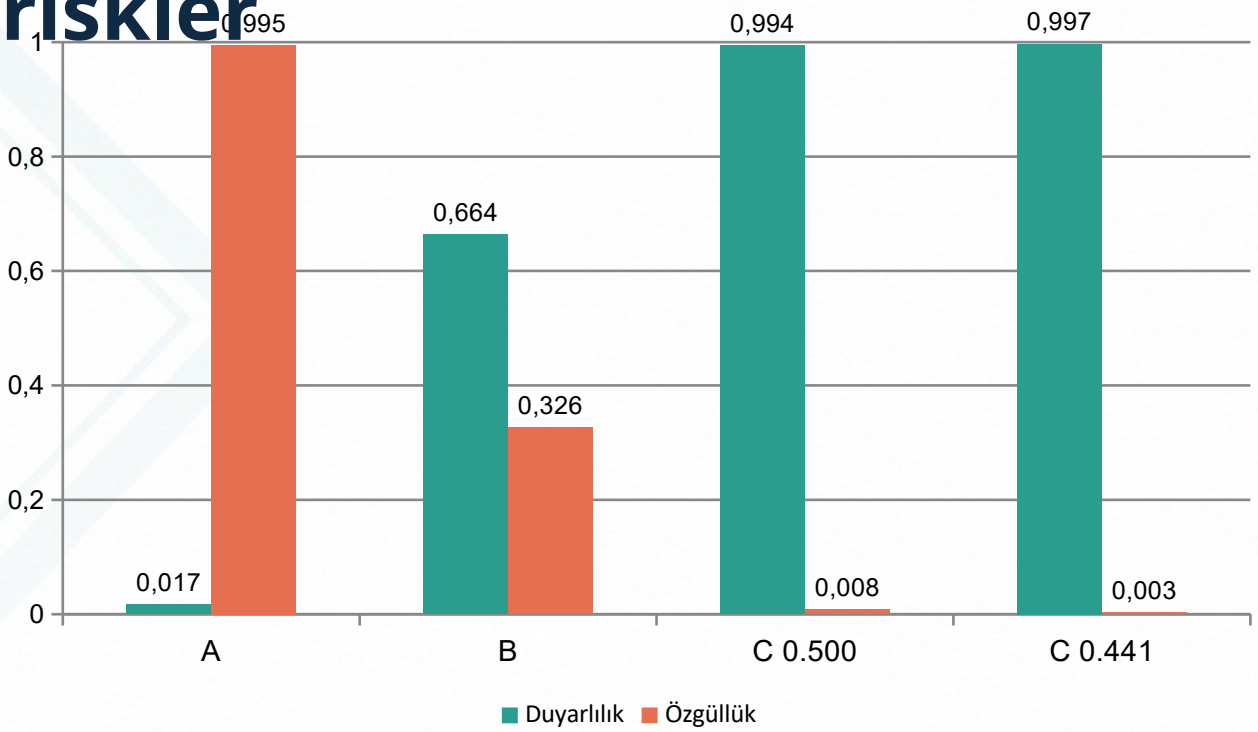


İç AUROC 0.875–0.900

Pediatrik AUROC 0.487–0.573

En iyi iç performans, çocuk el bileğinde korunmadı.

Hata paterni: aynı başarısızlık değil, farklı klinik riskler



Model A
4.390 / 4.467 kırığı kaçırdı

Model C
5.397 / 5.440 non-fractured görüntüyü kırık dedi

Klinik yorum: biri "kaçırma" riski, diğeri "aşırı alarm" riski üretir.

Neden “external validation” değil, “stress test”?

Klinik dış doğrulama;

- Aynı klinik görev
- Benzer hedef popülasyon
- Uyumlu etiket tanımı
- Bağımsız ama klinik olarak karşılaştırılabilir test kohortu

Bu çalışma;

- Yaş grubu değişti
- Anatomik bölge değişti
- İskelet maturitesi değişti
- Preprocessing ve image encoding farklı
- Etiketleme / filtreleme uyumlu değil

Sonuç: Bu analiz klinik kullanılabilirlik iddiası değil; modelin alan kayması altında kırılganlığını gösterir.

Limitasyonlar ve sonraki adım

Limitasyonlar

- Veri setleri kasıtlı olarak eşleştirilmedi; tekil domain faktörü izole edilemedi.
- FracAtlas image-level split: olası ilişkili görüntü sızıntısı dışlanamaz.
- Tek mimari: ResNet18; her strateji tek eğitim koşusuyla değerlendirildi.
- Kalibrasyon, localization ve reader comparison yapılmadı.

Sonraki adım

- Aynı anatomik bölge ve yaş grubunda hasta/exam-level split
- Pediatrik fine-tuning ve hedef-domain kalibrasyonu
- Tekrarlı seed analizi ve model değişkenliği
- Grad-CAM / hata analizi ile klinik failure audit

Bu sunumun değeri “yeni model yüksek başarı sağladı” değil; başarısızlığın güvenilir şekilde gösterilmesidir.

Sonuç: pediatrik kırık yapay zekâsında ayrı doğrulama şart

Take-home message

İç doğrulamadaki iyi AUROC, pediatrik el bileği travmasına taşınmadı.

Augmentation, class weighting ve validasyon-türetilmiş eşik seçimi domain shift'i çözmedi.

Pediatrik kırık yapay zekâ sistemleri; popülasyon-spesifik geliştirme, kalibrasyon ve bağımsız doğrulama olmadan klinik uygulamaya taşınmamalıdır.