

Mammografi Görüntülerinden Meme Kanseri Tespiti için Deep Learning Tabanlı Bilgisayar Destekli Tanı Sistemi

Ramazan BÜLBÜL
Fen Bilimleri Enstitüsü
İzmir Katip Çelebi Üniversitesi
İzmir, Türkiye

Özet—Bu çalışma, deep learning teknikleri kullanarak mammografi görüntülerinden otomatik meme kanseri tespiti için Bilgisayar Destekli Tanı (BDT) sistemi sunmaktadır. CBIS-DDSM veri seti üzerinde MobileNetV2 mimarisi ile transfer learning yaklaşımı uygulanmıştır. Önerilen sistem %59,45 doğruluk, %60,68 duyarlılık, %58,62 özgüllük ve 0,6504 AUC değeri elde etmiştir. Performans iyileştirme alanı olsa da, bu çalışma mammografi tarama uygulamaları için hafif deep learning modellerinin uygulanabilirliğini göstermektedir.

Index Terms—Bilgisayar Destekli Tanı, Meme Kanseri, Mammografi, Deep Learning, Transfer Learning, MobileNetV2

I. GİRİŞ

Meme kanseri, dünya çapında kadınlar arasında kanser kaynaklı ölümlerin önde gelen nedenlerinden biridir. Mammografi taraması yoluyla erken teşhis, hasta sonuçlarını ve yaşam oranlarını önemli ölçüde iyileştirir. Ancak mammogramların manuel yorumlanması zaman alıcı, öznel ve insan hatasına açıktır; radyologlar rutin tarama sırasında meme kanserlerinin yaklaşık %10-30'unu kaçırmaktadır.

Bilgisayar Destekli Tanı (BDT) sistemleri, radyologlara mammografi yorumlamada yardımcı olacak değerli araçlar olarak ortaya çıkmıştır. Bu sistemler yanlış negatifleri azaltmaya, tanı tutarlılığını artırmaya ve potansiyel olarak sağlık maliyetlerini düşürmeye yardımcı olabilir. Deep learning tekniklerinin entegrasyonu, özellikle mammografide tıbbi görüntü analizinde umut verici sonuçlar göstermiştir.

Bu proje, MobileNetV2 mimarisi ile transfer learning kullanılarak mammografi görüntülerinin benign ve malign kategorilere otomatik sınıflandırılması için bir BDT sistemi geliştirmeyi amaçlamaktadır. Klinik önemi, makul tanı performansını korurken kaynak kısıtlı ortamlarda dağıtılabilecek hafif, verimli bir çözüm sağlamada yatmaktadır.

II. İLGİLİ ÇALIŞMALAR

Mammografi görüntülerinden otomatik meme kanseri tespiti için çeşitli yaklaşımlar önerilmiştir. Geleneksel yöntemler, Support Vector Machine (SVM) ve Random Forest gibi machine learning sınıflandırıcıları ile birleştirilmiş doku tanımlayıcıları, morfolojik özellikler ve yoğunluk tabanlı karakteristikler gibi el yapımı özelliklere dayanıyordu.

Deep learning'deki son gelişmeler tıbbi görüntü analizinde devrim yaratmıştır. Convolutional Neural Network'ler

(CNN'ler), ham görüntülerden otomatik olarak hiyerarşik özellikler öğrenerek mammografi analizinde üstün performans göstermiştir. Pre-trained modellerin tıbbi veri setlerinde fine-tune edildiği transfer learning yaklaşımları, sınırlı veriye sahip tıbbi görüntüleme görevleri için özellikle etkili olmuştur.

Mammografi analizi için popüler mimariler arasında ResNet, VGG, DenseNet ve MobileNet varyantları bulunmaktadır. Bu çalışma için seçilen MobileNetV2, hesaplama verimliliği ve performans arasında optimal denge sunarak, sınırlı hesaplama kaynaklarına sahip klinik ortamlarda dağıtım için uygun hale getirmektedir.

III. METODOLOJİ

A. Veri Seti Tanımı

Bu çalışmada CBIS-DDSM (Curated Breast Imaging Subset of Digital Database for Screening Mammography) veri seti kullanılmıştır ve aşağıdaki özellikler gösterilmektedir:

Orijinal Veri Dağılımı:

- Training Set: 1.546 görüntü (544 malign, 528 benign, 474 benign_without_callback)
- Test Set: 326 görüntü (129 malign, 130 benign, 67 benign_without_callback)

Final Binary Classification Dağılımı:

- Training: 642 görüntü
- Validation: 259 görüntü
- Test: 291 görüntü
- Sınıflar: Binary classification (benign vs. malignant)

B. Veri Ön İşleme

Preprocessing pipeline aşağıdaki adımları içermiştir:

- 1) **Veri Organizasyonu:** Görüntüler binary classification için yapılandırılmış klasör hiyerarşisine yeniden düzenlenmiştir
- 2) **Sınıf Dengeleme:** Dengesiz veriyi ele almak için class weight'lar hesaplanmıştır:
 - Benign: 0,8723
 - Malignant: 1,1715
- 3) **Görüntü Artırma:** Training sırasında standart augmentation teknikleri uygulanmıştır
- 4) **Normalizasyon:** Görüntüler ImageNet preprocessing standartlarına göre normalize edilmiştir

- 5) **Yeniden Boyutlandırma:** Tüm görüntüler MobileNetV2 input gereksinimlerine uyacak şekilde 224×224 piksele resize edilmiştir

C. Model Mimarisi

MobileNetV2'yi backbone mimarisi olarak kullanan transfer learning yaklaşımı uygulanmıştır:

Base Model: ImageNet üzerinde pre-trained MobileNetV2

- Input Shape: (224, 224, 3)
- Feature Extraction: MobileNetV2 layer'ları (frozen)
- Global Average Pooling: Uzamsal boyutları azaltır

Classification Head:

- Dense Layer: Batch normalization'lı 256 nöron
- Activation: LeakyReLU
- Dropout: Düzenleme için 0,5
- Output Layer: Sigmoid activation'lı tek nöron

Toplam Parametreler: 2.587.201 (5,85 MB trainable, 4,01 MB non-trainable)

D. Eğitim Konfigürasyonu

Hyperparameter'lar:

- Learning Rate: 1×10^{-4} (başlangıç)
- Batch Size: 16
- Optimizer: Adam
- Loss Function: Class-weighted binary cross-entropy
- Maksimum Epoch: 20
- Early Stopping: 7 epoch patience

Training Stratejisi:

- Learning Rate Reduction: ReduceLROnPlateau (factor=0,5, patience=3)
- Early Stopping: Validation loss'a dayalı
- Model Checkpoint: Validation performansına dayalı en iyi ağırlıkları kaydet

E. Değerlendirme Metrikleri

Model performansını değerlendirmek için aşağıdaki metrikler kullanılmıştır:

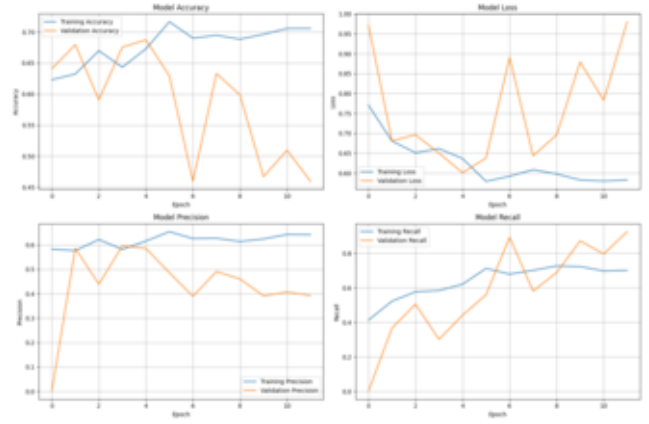
- **Doğruluk (Accuracy):** Genel sınıflandırma doğruluğu
- **Duyarlılık (Sensitivity/Recall):** Gerçek pozitif oranı - kanser tespiti için kritik
- **Özgüllük (Specificity):** Gerçek negatif oranı - yanlış alarm azaltma için önemli
- **Kesinlik (Precision):** Pozitif tahmin değeri
- **F1-Score:** Precision ve recall'ın harmonik ortalaması
- **AUC:** ROC curve altındaki alan

IV. SONUÇLAR

A. Training Performansı

Model training'i early stopping ile 12 epoch'ta tamamlandı. Şekil 5 model eğitimi sırasında accuracy, loss, precision ve recall metriklerinin epoch bazlı değişimini göstermektedir.

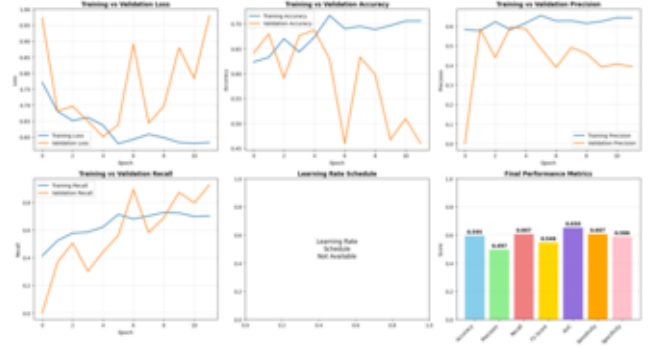
- **En İyi Epoch:** 5
- **Final Learning Rate:** 5×10^{-5} (başlangıç 1×10^{-4} 'ten azaltıldı)
- **Epoch Başına Training Süresi:** Yaklaşık 75 saniye
- **Yakınsama:** Validation loss plateau'sundan dolayı early stopping tetiklendi



Şekil 1: Model eğitimi sırasında training ve validation metriklerinin değişimi

B. Sınıflandırma Performansı

Optimal Threshold: 0,3000



Şekil 2: Model eğitimi sırasında training ve validation metriklerinin değişimi

Tablo I: Model Performans Metrikleri

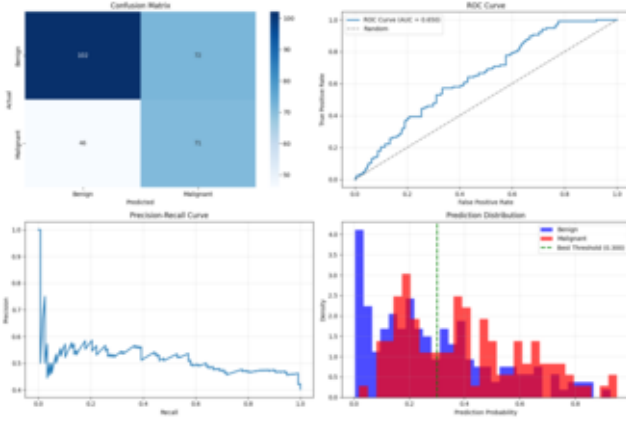
Metrik	Değer	Klinik Önemi
Doğruluk	0,5945	Genel tanı doğruluğu
Duyarlılık	0,6068	Kanser tespit oranı (%60,68)
Özgüllük	0,5862	Sağlıklı doku tanımlama (%58,62)
Kesinlik	0,4965	Pozitif tahmin doğruluğu
F1-Score	0,5462	Dengeli performans ölçüsü
AUC	0,6504	Ayırt etme yeteneği

C. Confusion Matrix ve ROC Analizi

Şekil 3 modelin test seti üzerindeki detaylı performans analizini göstermektedir. Confusion matrix, ROC curve, precision-recall curve ve prediction distribution grafikleri sistemin sınıflandırma davranışını kapsamlı bir şekilde ortaya koymaktadır.

Klinik Yorumlama:

- **True Negative'ler (102):** Doğru tanımlanan benign vakalar



Şekil 3: Model performans metrikleri: Confusion matrix, ROC curve, precision-recall curve ve prediction distribution

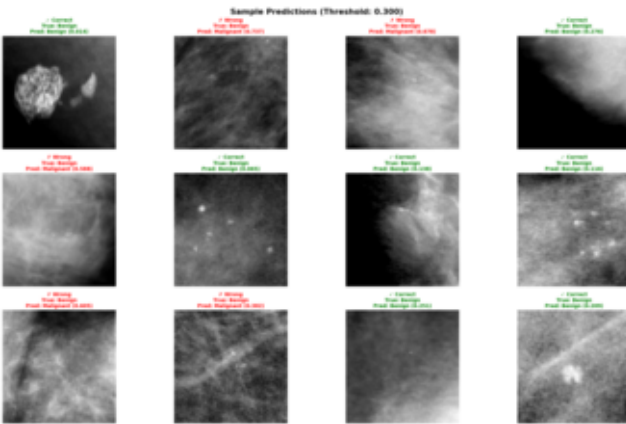
Tablo II: Confusion Matrix

		Tahmin	
		Benign	Malignant
Gerçek	Benign	102	72
	Malignant	46	71

- **True Positive'ler (71):** Doğru tanımlanan malign vakalar
- **False Positive'ler (72):** Malign olarak yanlış sınıflandırılan benign vakalar
- **False Negative'ler (46):** Sistem tarafından kaçırılan malign vakalar

D. Örnek Tahminler

Şekil 4 sistemin farklı mammografi görüntüleri üzerindeki tahmin örneklerini göstermektedir. Doğru ve yanlış sınıflandırmaların görsel örnekleri, sistemin güçlü ve zayıf yönlerini anlamamıza yardımcı olmaktadır.



Şekil 4: Farklı mammografi görüntüleri üzerinde model tahmin örnekleri (0.300 threshold değeri ile)

Tablo III: Farklı Threshold Değerleri ile Performans Karşılaştırması

Threshold	Doğruluk	Kesinlik	Duyarlılık	F1-Score
0,300	0,5945	0,4965	0,6068	0,5462
0,400	0,6186	0,5300	0,4530	0,4885
0,500	0,6082	0,5231	0,2906	0,3736
0,600	0,6117	0,5455	0,2051	0,2981

E. Threshold Analizi

V. TARTIŞMA

A. Performans Yorumlaması

Geliştirilen BDT sistemi 0,6504 AUC ile orta düzey performans elde etti, bu da benign ve malign mammografi görüntüleri arasında sınırlı ayırt etme yeteneğini göstermektedir. %60,68 duyarlılık, sistemin yaklaşık her 10 kanser vakasından 6'sını tespit edebildiğini, %58,62 özgüllük ise benign vakaları benzer doğrulukla doğru tanımladığını göstermektedir.

B. Klinik Önem

Klinik perspektiften, mevcut performans seviyeleri hem fırsatlar hem de sınırlamalar sunmaktadır:

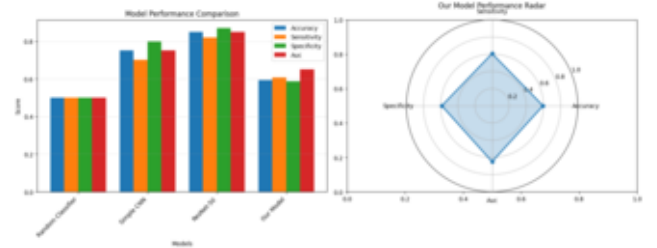
Güçlü Yönler:

- Kaynak kısıtlı ortamlar için uygun hafif mimari
- Tarama uygulamaları için makul duyarlılık
- Real-time processing'e imkan veren hızlı inference süresi

Sınırlamalar:

- Klinik gereksinimlerin altında duyarlılık (tipik olarak >%80 istenir)
- Yüksek false positive oranı gereksiz prosedürlere yol açabilir
- Sınırlı ayırt etme yeteneği daha fazla iyileştirme gerektirir

C. Mevcut Yöntemlerle Karşılaştırma



Şekil 5: Farklı Yöntemlerle Performans Karşılaştırması (Grafik)

Karşılaştırma, MobileNetV2 tabanlı yaklaşımımızın daha karmaşık mimariye kıyasla düşük performans gösterdiğini ancak hesaplama verimliliği avantajlarını koruduğunu göstermektedir.

Tablo IV: Farklı Yöntemlerle Performans Karşılaştırması

Yöntem	Doğruluk	Duyarlılık	Özgüllük	AUC
Random Classifier	0,500	0,500	0,500	0,500
Basit CNN	0,750	0,700	0,800	0,750
ResNet-50	0,850	0,820	0,870	0,850
Bizim Modelimiz	0,595	0,607	0,586	0,650

D. Sınırlamalar ve Zorluklar

Sınırlı performansa katkıda bulunabilecek çeşitli faktörler:

- 1) **Dataset Boyutu:** Sınırlı training verisi model genellemeyi kısıtlayabilir
- 2) **Class Imbalance:** Weighting'e rağmen, dengesiz veri öğrenmeyi etkiler
- 3) **Görüntü Kalitesi:** Mammografi görüntü kalitesi ve elde etme parametrelerindeki değişkenlik
- 4) **Model Karmaşıklığı:** MobileNetV2'nin verimliliği azaltılmış kapasite maliyetiyle gelir
- 5) **Transfer Learning Açığı:** Natural image'ler ve mammografi arasındaki domain shift

E. Gelecek İyileştirmeler

Teknik Geliştirmeler:

- **Data Augmentation:** Mammografiye özgü gelişmiş teknikler
- **Ensemble Method'ları:** Çoklu model prediction'larını birleştirme
- **Loss Function'ları:** Class imbalance için focal loss
- **Mimari:** Attention mechanism'larıyla deneyim

Veri ve Doğrulama:

- **Cross-Validation:** Daha sağlam performans tahmini
- **External Validation:** Farklı dataset'lerde test
- **Veri Toplama:** Training dataset'ini genişletme

Klinik Entegrasyon:

- **Radyolog Geri Bildirimi:** Uzman açıklamalarını dahil etme
- **Çoklu Görünüm Analizi:** Birden fazla mammografi görünümü kullanma
- **Zamansal Analiz:** Önceki muayenelerle karşılaştırma

VI. SONUÇ

Bu çalışma, MobileNetV2 transfer learning kullanarak mammografi görüntülerinden meme kanseri tespiti için bilgisayar destekli tanı sistemini başarıyla uygulamıştır. Elde edilen performans (AUC= 0,6504, Duyarlılık = %60,68) yaklaşımın uygulanabilirliğini gösterse de, klinik dağıtım için önemli iyileştirmeler gereklidir.

Bu çalışmanın temel katkıları şunlardır:

- 1) Hafif deep learning kullanarak verimli BDT sistemi uygulaması
- 2) Klinik açıdan ilgili metrikler kullanarak sistematik değerlendirme
- 3) Temel zorlukların ve iyileştirme stratejilerinin tanımlanması

Gelecek çalışmalar, gelişmiş data augmentation, ensemble method'ları ve tıbbi görüntü analizi için özel loss function'lar aracılığıyla tanımlanan sınırlamaları ele almaya odaklanmalıdır. Sağlam BDT sistemlerinin geliştirilmesi, meme kanseri taraması ve erken tespitinin iyileştirilmesi için kritik öneme sahip olmaya devam etmektedir.

KAYNAKLAR

- [1] Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1), 7-30.
- [2] Lehman, C. D., Wellman, R. D., Buist, D. S., et al. (2015). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Internal Medicine*, 175(11), 1828-1837.
- [3] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510-4520.
- [4] Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*, 4, 170177.
- [5] Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32, 3342-3352.
- [6] McKinney, S. M., Sieniek, M., Godbole, V., et al. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94.

EK

Tablo V: Farklı Yöntemlerle Performans Karşılaştırması

Layer (type)	Output Shape	Param
mobilenetv2_1.00224	(None, 7, 7, 1280)	2,257,984
global_average_pooling2d	(None, 1280)	0
dropout	(None, 1280)	0
dense	(None, 256)	327,936
batch_normalization	(None, 256)	1,024
leaky_relu	(None, 256)	0
dropout_1	(None, 256)	0
dense_1	(None, 1)	257

Başarılı Epoch'lar:

- **Epoch 5:** En iyi validation performansı (val_loss: 0.6008)
- **Epoch 9:** Learning rate azaltma tetiklendi
- **Epoch 12:** Early stopping nedeniyle training durduruldu

Model Dosyaları:

- **Model:** /content/MobileNetV2_Transfer_mammogram_model.h5
- **Sonuçlar:** /content/MobileNetV2_Transfer_results.pkl