

DERİN ÖĞRENME İLE MEME KANSERİ TESPİTİ¹

BREAST CANCER DETECTION WITH DEEP LEARNING

Duhan KÖSALİ

1904040013

duhanksali@gmail.com

* Beykoz Üniversitesi Yazılım Mühendisliği Bölümü

ÖZ

Derin öğrenme, meme kanseri tespitinde önemli bir rol oynayabilen bir yapay zeka teknolojisidir. Derin öğrenme, karmaşık veri setlerini analiz etmek ve desenleri tespit etmek için çok katmanlı yapay sinir ağları kullanır. Meme kanseri taramasında derin öğrenme algoritmaları, mamogram görüntülerini inceleyerek anormal bölgeleri tanımlayabilir ve kanser tespitinde yardımcı olabilir. Derin öğrenme yöntemleri, mamogramları daha hassas bir şekilde değerlendirebilir, kanserli lezyonları diğer normal dokulardan ayırt edebilir ve erken teşhis imkanı sunabilir. Bu teknoloji, doğruluk oranını artırarak yanlış pozitif ve yanlış negatif sonuçları azaltabilir. Derin öğrenme, hızlı ve otomatik bir şekilde büyük miktardaki veriyi analiz edebilir ve doktorlara daha güvenilir ve hızlı bir şekilde sonuçlar sunabilir.

Anahtar Kelimeler: Meme Kanseri tespiti, Derin öğrenme, yapay zeka

ABSTRACT

Deep learning, plays a significant role in breast cancer detection. It is an artificial intelligence technology that utilizes multi-layered neural networks to analyze complex datasets and identify patterns. In breast cancer screening, deep learning algorithms can examine mammogram images and identify abnormal regions, aiding in cancer detection. Deep learning methods can evaluate mammograms more accurately, distinguish cancerous lesions from normal tissues, and provide early diagnosis opportunities. This technology can increase accuracy, reducing false positives and false negatives. Deep learning can analyze large amounts of data rapidly and automatically, providing doctors with more reliable and faster results.

Keywords: deep learning, breast cancer detection, mammogram, artificial intelligence

1. GİRİŞ

"Meme kanseri", meme dokusunda başlayan ve hücrelerin kontrolsüz bir şekilde çoğalması sonucunda oluşan bir kanser türüdür. Bu kanser türü, meme dokusundaki hücrelerde meydana gelen genetik değişiklikler sonucu ortaya çıkar. Meme kanseri genellikle kadınlarda görülse de, nadir durumlarda erkeklerde de ortaya çıkabilir.

İlk olarak meme kanseri hakkında tarihsel bir bilgi vermek gerekirse, meme kanserine ilişkin en eski kayıtlar M.Ö. 1600'lü yıllara kadar uzanmaktadır. Eski Mısırlılar ve Antik Yunanlılar meme kanserine dair bazı

belirtileri ve tedavi yöntemlerini bilmekteydi. Ancak modern anlamda meme kanseri tanısı ve tedavisi 20. yüzyılın başlarında gelişmiştir. Meme kanseri konusunda önemli ilerlemeler, meme ultrasonu, mamografi ve diğer tarama yöntemlerinin kullanıma girmesiyle gerçekleşmiştir.

Ductal Carcinoma, meme dokusunun kanallarında (süt kanalları) başlayan kanser türüdür. Lobular Carcinoma ise meme dokusunun lobül adı verilen süt bezlerinde ortaya çıkan kanser türüdür. Mucinous Carcinoma (mukoepidermoid kanseri olarak da bilinir), kanser hücrelerinin mukus adı verilen yapışkan bir madde salgıladığı bir tür

meme kanseridir. Papillary Carcinoma ise meme dokusunda tümörün parmak şeklinde çıkıntılar oluşturduğu bir kanser türüdür.

2. MODEL GELİŞTİRME VE BAŞARI ORANINI ARTTIRMA

Bir yapay sinir ağı modelinin başarısını arttırmak için çeşitli yöntemler bulunmaktadır. Yapay sinir ağı modelinin genelleme yeteneğini artırmak için daha fazla etiketli veri toplanabilir. Bu, modelin daha çeşitli örüntüleri öğrenmesine ve daha iyi sonuçlar üretmesine yardımcı olabilir. Modelin karmaşıklığı, aşırı uyum riskini azaltmak için dikkatli bir şekilde ayarlanmalıdır. Aşırı karmaşık bir model, eğitim verilerine aşırı uyum sağlayabilir, ancak genelleme performansını düşürebilir. Model karmaşıklığı, düğüm sayısı, katman sayısı veya düşürme katmanları gibi yapılandırılmalarla kontrol edilebilir. Veri setinin düzgün bir şekilde ön işlenmesi ve normalizasyonu, modelin daha iyi sonuçlar üretmesine yardımcı olabilir. Bu, örneğin verileri standartlaştırmak, eksik verileri doldurmak veya gürültüyü temizlemek gibi işlemleri içerebilir.

2.1. AŞIRI UYUMUN (OVERFITTING) ENGELLENMESİ

Aşırı uyum (overfitting), bir modelin eğitim verilerine aşırı uyum sağladığı durumdur ve genelleme yeteneğini kaybedebilir. Düzenleştirme teknikleri, aşırı uyumu azaltmaya yardımcı olabilir. Bunlar arasında L1 ve L2 düzenleştirme, dropout ve toplu normalleştirme gibi teknikler bulunur. Bu teknikler, modelin ağırlıklarını düzenleyerek aşırı uyumu kontrol altına alabilir. Çapraz doğrulama, modelin performansını daha iyi değerlendirmek için kullanılır. Veri seti, farklı alt kümelerde bölünerek modelin her bir alt küme üzerinde nasıl performans gösterdiği değerlendirilir. Bu, modelin gerçek dünya verilerinde genelleme yapabilme yeteneğini daha iyi anlamamıza yardımcı olur.

2.2. DATA AUGMENTATION

Data augmentation, veri artırma veya veri çoğaltma olarak da bilinen bir tekniktir. Bu teknik, mevcut veri setinin yapısını değiştirerek yeni ve çeşitli veri örnekleri oluşturmayı amaçlar. Veri augmentasyonu, özellikle sınırlı miktarda veriyle çalışılan makine öğrenimi ve derin öğrenme modellerinde kullanılır.

Data augmentation, örneklem sayısını artırarak modelin genelleme yeteneğini artırabilir ve aşırı uyumu azaltabilir. Veri augmentasyonu, mevcut veri örnekleri üzerinde dönüşümler, ölçeklendirme, döndürme, yansıtma, kesme, parlaklık değişimi gibi değişiklikler yapılır. Örneğin, bir görüntüyü yatay olarak yansıtarak yeni bir görüntü elde edebiliriz. Böylece model, daha farklı açılardan ve varyasyonlarda veriyle eğitilmiş olur.

Data augmentation, modelin daha genel bir görüntü anlayışı geliştirmesine yardımcı olabilir ve veri setindeki dengesizlikleri giderme potansiyeline sahiptir. Bu teknik sayesinde daha çeşitli ve temsil edici bir veri seti elde edilerek modelin daha iyi performans göstermesi hedeflenir.

3. MODELİN DEĞERLENDİRİLMESİ VE DOĞRULUK ANALİZİ

Derin öğrenme modellerinin değerlendirilmesi ve doğruluk analizi, eğitim sonrası modelin performansını objektif bir şekilde değerlendirmeyi sağlar. Bunun için aşağıdaki adımlar izlenebilir:

- **Test Veri Seti Oluşturma:** Bağımsız bir test veri seti, modelin eğitim ve doğrulama veri setleri dışında tutulan verilerden oluşmalıdır. Bu veri seti, modelin gerçek dünya verilerine nasıl genelleştiğini değerlendirmek için kullanılır.
- **Metriklerin Hesaplanması:** Modelin performansını değerlendirmek için kayıp (loss) ve doğruluk (accuracy) gibi metrikler hesaplanır. Test veri seti üzerinde modeli çalıştırarak, kayıp ve doğruluk değerlerini elde edebiliriz. Kayıp değeri, modelin

BREAST CANCER DETECTION

tahminlerinin gerçek değerlerden ne kadar sapma gösterdiğini ifade ederken, doğruluk değeri doğru sınıflandırma oranını yüzde olarak gösterir.

Bu yöntemler, modelin performansını değerlendirmek için yaygın olarak kullanılır. Ancak, diğer metrikler de kullanılabilir, özellikle veri setinin özelliğine ve problemin doğasına bağlı olarak değişebilir.

4. MEME KANSERİ TESPİTİ UYGULAMASI

Meme kanseri, dünya genelinde kadınlar arasında en sık görülen kanser türüdür. Erken teşhis, hastaların hayatta kalma oranını büyük ölçüde artırabilir. Bu bağlamda, bu projede, histopatolojik meme kanseri görüntülerinin otomatik sınıflandırılmasını sağlamak için derin öğrenme yöntemleri kullanılmıştır.

Projenin ana veri seti olarak "Breast Cancer Histopathological Database (BreakHis)" kullanılmıştır. Bu veri setinde hem iyi huylu (benign) hem de kötü huylu (malignant) tümör görüntüleri bulunmaktadır. Ancak bu çalışma, kötü huylu tümörlerin sınıflandırılmasına odaklandığı için, benign olan görüntüler veri setinden çıkarılmıştır. Kötü huylu tümörler arasında dört farklı tür sınıflandırılmıştır: ductal_carcinoma, lobular_carcinoma, mucinous_carcinoma, papillary_carcinoma.

Bu proje için derin öğrenme modeli olarak ResNet101 seçilmiştir. ResNet101, genişleyen özellik hiyerarşileri ile birlikte gelişmiş bir performans sunabilen bir derin sinir ağı modelidir. Modelin ilk katmanları görüntülerin daha basit özelliklerini (örneğin kenarlar ve renk blokları) yakalar, sonraki katmanlar daha karmaşık özellikleri yakalar. Bu, karmaşık görsel sınıflandırma görevlerinde genellikle çok etkilidir.

Veri setindeki görüntüler, modeli eğitmek için kullanılan "ImageDataGenerator" fonksiyonu ile ön işleme tabi tutulmuştur. Bu aşamada, "data augmentation" adı verilen bir teknik kullanılmıştır. Data augmentation, mevcut eğitim verilerinden yapay olarak daha fazla eğitim verisi

üretmek için bir dizi rastgele dönüşümler uygulayarak modelin genelleştirme yeteneğini artırmayı amaçlar.

İlgili modelde "epoch" değeri 10 olarak tanımlanmıştır. Yani tüm eğitim veri setinin modelden 10 kez geçirileceği ve her geçişte modelin ağırlıklarının güncelleneceği anlamına gelir.

Bir yapay zeka modelinde epoch sayısı oldukça önemlidir. Çünkü bu sayı az tutulduğunda yetersiz uyum (underfitting), fazla tutulduğu durumlarda ise aşırı uyum (overfitting) yaşanabilir.

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1	2.4629	0.6271	27.9375	0.6548
2	0.7581	0.7049	4.6070	0.6548
3	0.5221	0.8104	1.1215	0.6548
4	0.3461	0.8784	1.9336	0.6548
5	0.2570	0.9133	9.8815	0.6340
6	0.2370	0.9268	3.0047	0.6475
7	0.1589	0.9406	0.4685	0.8279
8	0.0904	0.9604	0.8065	0.7059

5. SONUÇ

Sonuç olarak, modelimizin eğitim süreci boyunca sürekli olarak doğruluk oranını artırdığı görülüyor, fakat genel anlamda modelin genelleme yeteneğinin geliştirilmesi gerektiği görülüyor. Bu da, aşırı uyumu önlemek ve modelin genelleme yeteneğini artırmak için daha fazla tekniklerin uygulanması gerektiğini gösteriyor.

Kaynakça

Cancer, B. (tarih yok). *Breast Cancer*. Cancer: <https://www.cancer.org/cancer/types/breast-cancer.html> adresinden alındı

Chollet, F. (2017). *The limitations of deep learning*.

Luis Perez, J. W. (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*.

Xiangyu Zhang, S. R. (2016). *Deep Residual Learning for Image Recognition*. IEEE.