T.C.

MALTEPE ÜNİVERSİTESİ TIP FAKÜLTESİ HASTANESİ

RADYOLOJİ ANABİLİM DALI

YAPAY ZEKADA VERİLERİN İŞLENMESİ, VERİ SETLERİNİN OLUŞTURULMASI VE MAMOGRAFİDE MEME DANSİTE UYGULAMALARI

UZMANLIK TEZİ

DR. DAVUT CAN GÜNER

Orcid No: 0000-0003-3141-6175

TEZ DANIŞMANI

DOÇ.DR. ÖMER ÖZÇAĞLAYAN

İSTANBUL - 2024

TEŞEKKÜRLER

Asistanlık eğitimim süresince destek ve ilgisini esirgemeyen, bilgi ve deneyimlerini aktaran değerli hocam Prof. Dr. Levent Çelik'e,
Asistanlık eğitimim boyunca her zaman yakın ilgi ve yardımlarını gördüğüm, bilgi ve deneyimlerini aktararak yetişmemi sağlayan, tez çalışmalarım sırasında bana çokça yardımcı olan Anabilim Dalı başkanımız değerli hocam Doç. Dr. Rahmi Çubuk'a,
Asistanlık eğitimim süresince destek ve ilgisini esirgemeyen, bilgi ve deneyimlerini aktaran değerli hocam Doç. Dr. Ömer Özçağlayan 'a teşekkürlerimi ve saygılarımı sunarım.

Asistanlık eğitimim boyunca destek ve yardımlarını esirgemeyen, beraber çalışmaktan zevk duyduğum, birçok güzel anılar paylaştığım, sevgili doktor, raportör, teknisyen ve sekreter arkadaşlarıma teşekkür ederim.

Sevgili dostlarıma, üzerimde sonsuz hak ve emekleri olan her zaman beni destekleyen sevgili ailem; eşim, annem, babam, abim ve ablama çok teşekkür ederim.

Dr. Davut Can Güner

İÇİNDEKİLER

TEŞEKKÜR i

İÇİNDEKİLER ii

KISALTMALAR iii

TABLOLAR LİSTESİ iv

ŞEKİLLER LİSTESİ vi

ÖZET vii

ABSTRACT viii

1. GİRİŞ VE AMAÇ 1
2. GENEL BİLGİLER 2
   - 2.1. YAPAY ZEKA 2
     - 2.1.1. Yapay Zekanın Genel Tanımı ve Tarihsel Gelişimi 2
     - 2.1.2. Yapay Zeka Eğitimi için Veri Toplama Metodları (Data Mining) 4
     - 2.1.3. Veri Setleri Nedir? Eğitim Amacıyla Nasıl Kullanılır? 6
       - 2.1.3.1. Training Set 6
       - 2.1.3.2. Validation Set 6
       - 2.1.3.3. Test Set 6
     - 2.1.4. Başarılı Bir Yapay Zeka Eğitiminde Önemli Noktalar 7
   - 2.2. MEME ANATOMİSİ 9
   - 2.3. MAMOGRAFİ VE MEME DANSİTESİ 11
     - 2.3.1. Mamografi Nedir? Ne için Kullanılır? 11
     - 2.3.2. Mamografide Meme Dansitesi 12
     - 2.3.3. Mamografide Meme Dansitesinin Önemi 12
3. GEREÇ VE YÖNTEM 15
   - 3.1. Hasta Seçimi 15
   - 3.2. Mamografi Çekim Protokolü 15
   - 3.3. Mamografi Görüntülerinin Değerlendirilmesi ve İstatistiksel Analiz 16
4. BULGULAR 17
5. OLGU ÖRNEKLERİ 40
6. TARTIŞMA 48
7. SONUÇ 51

KAYNAKÇA 52

ETİK KURUL ONAY SAYFASI 57

TEZ ONAY SAYFA 58

KISALTMALAR

YZ	**Yapay Zeka**
CNN	**Convolutional Neuronal Network (Konvolusyonel Nöronal Ağ)**
RNN	**Recurrent Neuronal Network (Tekrarlayan Sinir Ağı)**
GAN	**Generative Adversarial Network**
FDA	**Federal Drug Administration**
FFDM	**Full-Field Digital Mammography**
PACS	**Picture Archiving Communication Systems**

TABLOLAR LİSTESİ

Tablo 1. Çalışmaya dahil edilen olguların yaşlarına ait tanımlayıcı istatistikler

Tablo 2. Meme yoğunluğu açısından kıdemli radyolog, yapay zeka ve radyoloji asistanı tarafından yapılan sınıflama yönünden olguların frekans dağılımları

Tablo 3. Memenin yoğun olup olmaması açısından kıdemli radyolog, yapay zeka ve radyoloji asistanı tarafından yapılan sınıflama yönünden olguların frekans dağılımları

Tablo 4. Meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

Tablo 5. Memenin yoğun olup olmaması açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

Tablo 6. Her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka tarafından yapılan sınıflamaların klinik performansı

Tablo 7. Her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

Tablo 8. Her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle karşılaştırılması

Tablo 9. Meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyum durumları yönünden yaş grupları arasında yapılan karşılaştırmalar

Tablo 10. Her bir yaş grubu içerisinde memenin dense olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka tarafından yapılan sınıflamaların klinik performansı

Tablo 11. Her bir yaş grubu içerisinde memenin dense olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

Tablo 12. Her bir yaş grubu içerisinde memenin yoğun olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle karşılaştırılması

Tablo 13. Memenin yoğun olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyum durumları yönünden yaş grupları arasında yapılan karşılaştırmalar

ŞEKİLLER LİSTESİ

Şekil 1: Meme Anatomisi

Şekil 2: Kıdemli Radyolog ile Yapay Zekanın Meme Yoğunluğu Sınıflaması

Şekil 3: Kıdemli Radyolog ile Radyoloji Asistanının Meme Yoğunluğu Sınıflaması

Şekil 4: Kıdemli Radyolog ile Yapay Zekanın Dens-Nondens Meme Yoğunluğu Sınıflaması

Şekil 5: Kıdemli Radyolog ile Radyoloji Asistanının Dens-Nondens Meme Yoğunluğu Sınıflaması

Şekil 6: Olgu 1

Şekil 7: Olgu 2

Şekil 8: Olgu 3

Şekil 9: Olgu 4

Şekil 10: Olgu 5

Şekil 11: Olgu 6

Şekil 12: Olgu 7

Şekil 13: Olgu 8

ÖZET

Amaç: Yapay zeka uygulamaları günlük radyoloji pratiğinde kullanılmaktadır. Çalışmamızda yapay zeka eğitimi için doğru ve kaliteli veri toplamanın önemi, programının meme dansitesinin tahminindeki başarısı ve radyoloji asistanı ile performans karşılaştırılması ele alınmaktadır. Sonuç olarak yapay zekanın klinik olarak uygulanabilirliği değerlendirilmektedir.

Gereç ve yöntem: Ocak 2021 – Temmuz 2021 tarihleri kliniğimizde; meme kanseri taraması amacı ile mamografi yapılan olgular retrospektif olarak tarandı ve PACS 'ta 754 hasta kaydına ulaşıldı ve 282 hasta dışlama kriterlerine (artefakt, post-op vb) göre çalışmadan çıkarıldı. Kalan 472 hasta çalışmaya dahil edildi. Hastaların mamografi görüntüleri meme konusunda kıdemli radyolog, yapay zeka programı ve radyoloji asistanı tarafından değerlendirildi. Meme dansitesinin belirlenmesinde yapay zeka programı ile radyoloji asistanının performans analizi IBM SPSS Statistics ver. 25 (IBM Corporation, Armonk, NY, US) istatistik programı kullanılarak değerlendirildi.

Bulgular: Meme dansitesinin belirlenmesinde yapay zekanın kıdemli radyolog ile olan tutarlılığı %97,2 iken radyoloji asistanının kıdemli radyolog ile olan tutarlılığı %95,6 olup yapay zeka ile radyoloji asistanı arasında istatistiksel olarak anlamlı fark yoktu (p=0,057). Benzer şekilde <40 yaş, 40-49 yaş, 50-59 yaş ve ≥60 yaş grupları içerisinde memenin dense olup olmaması açısından kıdemli radyolog ile tutarlı sonuç verme durumu yapay zekanın lehine görece yüksek olmakla beraber yapay zeka ile radyoloji asistanı arasında istatistiksel olarak anlamlı herhangi bir fark görülmemiştir (p=0,500; p=0,453; p>0,999 ve p=0,500).

Sonuç: Meme dansitesinin belirlenmesinde yapay zeka programı ile radyoloji asistanı benzer performans göstermekle birlikte, yaş alt grupları ile birlikte değerlendirildiğinde yapay zeka bazı alt gruplarda radyoloji asistanından yüksek performans göstermektedir. Güncel pratikte meme dansitesinin belirlenmesinde yapay zeka programı başarılı ve meme konusunda kıdemli radyolog ile performansları koreledir.

ABSTRACT

Purpose: Artificial intelligence applications are used in daily radiology practice. In our study, the importance of collecting accurate and high-quality data for artificial intelligence training, its capacity in comprehensive breast density estimation, and its performance comparison with a radiology asistant are discussed. As a result, the clinical applicability of artificial intelligence is evaluated.

Materials and methods: January 2021 – July 2021 in our affiliated clinic; cases who underwent mammography for breast cancer screening were retrospectively scanned and 754 patient records were accessed in PACS and 282 patients were excluded from the study according to exclusion criteria (artifact, post-op, etc.). The remaining 472 patients were included in the study. The mammography rates of the patients were evaluated by the expert radiologist, artificial intelligence program and radiology asistant. Performance analysis of the radiology assistant and the artificial intelligence program in determining breast density: IBM SPSS Statistics ver. 25 (IBM Corporation, Armonk, NY, US) was evaluated using the statistical program.

Results: The study revealed that even though the consistency of artificial intelligence with the expert radiologist in determining breast density was 97.2%, and the consistency of the radiology assistant with the experted radiologist was 95.6%, there was no statistically significant difference between artificial intelligence and the radiology assistant ($p=0.057$). Similarly, artificial intelligence exhibited a tendency towards relatively higher consistency with the expert radiologist in determining breast density across age groups of <40 years, 40-49 years, 50-59 years, and ≥60 years, no statistically significant distinction was found between artificial intelligence and the radiology assistant ($p=0.500$; $p=0.453$; $p>0.999$ and $p=0.500$).

Conclusion: Although the artificial intelligence program and radiology asistant show similar performance in determining breast density, when evaluated together with age subgroups, artificial intelligence shows higher performance than radiology assistant in some subgroups.

1. GİRİŞ VE AMAÇ

Meme dansitesinin tespiti, meme sağlığı ve tıbbi görüntüleme bağlamında önemlidir (1,2). Meme dansitesi, memede bulunan lifli ve glandüler dokunun, yağ dokusuna göre oranını olarak tanımlanır (3).

Meme dansitesi, meme kanseri için bir risk faktörü olarak kabul edilir (4). Dens meme dokusuna sahip kadınlar, daha az yoğun meme dokusuna sahip olanlara göre meme kanseri riski daha yüksek olabilir. Meme dansitesini belirleme ve ölçme, bireyin meme kanseri geliştirme riskini değerlendirmede yardımcı olabilir (5). Yoğun meme dokusuna sahip kadınlar, standart mamografi ötesinde ek veya alternatif tarama yöntemlerinden fayda sağlayabilir. Yoğun meme dokusu, hem kanserli hem de yoğun dokunun beyaz göründüğü mamogramlarda anormallikleri tespit etmeyi zorlaştırabilir. Meme dansitesini doğru belirlemek hastaya doğru kişiselleştirilmiş tarama önerileri yapmaya yardımcı olur (6). Meme dansitesini doğru belirlemek, meme kanseri taramasının etkinliğini optimize etmek için önemlidir. Yoğun meme dokusuna sahip kadınlar için ultrason veya manyetik rezonans görüntüleme (MRG) gibi ek tarama yöntemleri, mamografide kaçırılabilecek potansiyel anormalliklerin tespitini artırmak için önerilebilir (7,8). Meme dansitesinin tespiti, devam eden bir araştırma alanıdır. Meme yoğunluğu ve meme kanseri riski arasındaki ilişkiyi incelemek, hastalığın daha iyi anlaşılmasına katkıda bulunabilir ve gelişmiş tarama stratejileri ile risk değerlendirme modellerinin geliştirilmesine rehberlik edebilir (9).

Meme dansitesi bilgileri, bireyin riskine göre tarama-takip stratejilerini uyarlamak için kullanılabilir. Daha yüksek meme dansitesine ve artmış kanser riskine sahip kadınlara daha sık veya gelişmiş görüntüleme yöntemleri önerilebilir.

Çalışmamızda yapay zeka eğitimi için doğru ve kaliteli veri toplamanın önemi, yapay zeka programının meme dansitesinin tahminindeki başarısı ve bunun klinik olarak uygulanabilirliği ele alınmaktadır.

2. GENEL BİLGİLER

2.1. YAPAY ZEKA

2.1.1. YAPAY ZEKA'NIN GENEL TANIMI VE TARİHSEL GELİŞİMİ

Yapay Zeka (YZ), insan zekasının taklit edildiği, öğrenme, problem çözme, algılama ve dil anlama gibi bilişsel fonksiyonları düşünmek ve taklit etmek üzere programlanmış makineleri ifade eder (10). Yapay Zekanın amacı, genellikle insan zekası gerektiren görevleri gerçekleştirebilen sistemler oluşturmaktır. Bu görevler arasında görsel algılama, konuşma tanıma, karar verme ve dil çevirisi gibi şeyler bulunur.

YAPAY ZEKA'NIN TARİHSEL GELİŞİMİ

1. Antik Kökenler:

İnsan benzeri zekaya sahip yapay varlıkların yaratma konsepti, antik mitoloji ve folklor kökenlerine dayandırılabilir. Ancak, Yapay Zekanın resmi olarak incelenmeye başlaması daha sonraki zamanlarda gerçekleşti (11).

2. Turing Testi (1950):

Modern Yapay Zeka tarihi genellikle Alan Turing'in "Hesap Makineleri ve Zeka" adlı makalesiyle başlar. Bu makalede, bir makinenin insandan ayırt edilemez bir şekilde akıllı davranma yeteneğini ölçmek için Turing Testi önerildi (12).

3. Dartmouth Konferansı (1956):

"Yapay Zeka" terimi, Yapay Zeka'nın bir araştırma alanı olarak doğduğu kabul edilen 1956'daki Dartmouth Konferansı'nda ortaya çıktı. Konferansı John McCarthy, Marvin Minsky, Nathaniel Rochester ve Claude Shannon düzenledi (13).

4. Erken Yapay Zeka Programları (1950'ler-1960'lar):

1950'ler ve 1960'lar boyunca araştırmacılar ve mantık teorisyenleri basit yapay zeka programları geliştirdi. Bu programlar, insan problem çözme ve akıl yürütme yeteneklerini taklit etmeyi amaçlıyordu.

5. Yapay Zeka Kışı – AI Winter (1970'ler-1980'ler):

1970'ler ve 1980'lerde, yapay zeka önemli zorluklarla karşılaştı ve "AI Winter" olarak bilinen bir döneme yol açtı. Beklentiler yüksekti, ancak ilerleme beklenenden daha yavaştı ve yapay zeka araştırmalarına yönelik finansman azaldı (14).

6. Uzman Sistemler – Expert Systems (1980'ler):

1980'lerde uzman sistemlere odaklanıldı; bunlar, belirli bir alandaki bir insan uzmanın karar alma yeteneğini taklit etmeyi amaçlayan kurallara dayalı sistemlerdi. Başarılarına rağmen, uzman sistemler belirsizlikleri iyi yönetemedi.

7. Makine Öğrenmesi Yeniden Canlanması (1990'lar-2000'ler):

Makine öğrenmesindeki ilerlemeler, özellikle sinir ağlarının yeniden ortaya çıkmasıyla, 1990'larda ve 2000'lerde yapay zekanın yeniden canlanmasına katkıda bulundu. Geriye yayılma gibi algoritmaların geliştirilmesi sinir ağlarının eğitimini iyileştirdi (15).

8. Büyük Veri ve Derin Öğrenme (2010'lar):

Büyük veri setlerinin ve artan bilgi işleme gücünün, makine öğrenmenin bir alt dalı olan derin öğrenmede çığır açıcı gelişmelere yol açtı. Özellikle konvolüsyonel sinir ağları (CNN'ler) ve tekrarlayan sinir ağları (RNN'ler), görüntü tanıma ve doğal dil işleme gibi görevlerde dikkat çekici performans sergiledi (16).

9. Mevcut Durum (2020'ler):

Yapay Zeka uygulamaları, sağlık, finans, otonom araçlar ve doğal dil işleme gibi çeşitli endüstrilerde yaygın olarak kullanılmaktadır. Yapay Zeka, aynı zamanda sanal asistanlar, öneri sistemleri ve otonom sistemler gibi teknolojilerin kritik bir bileşenidir (17).

Yapay zekanın gelişimi günümüzde devam etmektedir. Güncel araştırmalar takviyeli öğrenme, üretici modeller ve bu teknolojilerin etik boyutları gibi alanlarda gerçekleştirilmektedir. Bu alan dinamiktir ve ilerlemeler, yapay zeka alanının gelecekteki yönünü şekillendirecektir.

2.1.2. YAPAY ZEKA EĞİTİMİ İÇİN VERİ TOPLAMA METODLARI (VERİ MADENCİLİĞİ)

Radyolojide yapay zeka eğitimi için veri toplama, yapay zeka modellerini eğitmek ve değerlendirmek için çeşitli ve temsilci bir tıbbi görüntü veri seti toplamayı içerir (18). Veri setinin kalitesi ve çeşitliliği, güçlü ve genelleştirilebilir yapay zeka algoritmalarının geliştirilmesi için kritiktir. İşte radyoloji eğitiminde yapay zeka için yaygın veri toplama yöntemleri:

1. Halka Açık Veri Tabanları:

- The Cancer Imaging Archive
- Radiological Society of North America veri setleri
- Open Access Biomedical Image Search Engine
- National Cancer Institute Görüntü Veri Arşivi

2. Kurumsal Veri Tabanları:

Hastanelerle ve tıbbi kurumlarla işbirliği yaparak radyolojik veri tabanlarına kullanılabilir. Bu yöntemde etik standartlara, hasta gizliliği yasalarına uyulma zorunluluğu mevcuttur.

3. Veri Etiketleme:

Gözetimli öğrenme modellerini eğitmek için tıbbi görüntüleri ilgili etiketlerle (örneğin, tümör konumları, organ sınırları) etiketlebilir. Bu süreç, uzman radyologların görüntüleri manuel olarak etiketleme veya yarı otomatik araçlar kullanma sürecini içerir (19).

4. Sentetik Veri Üretimi:

Veri setini zenginleştirmek için sentetik tıbbi görüntüler üretilebilir. Generative Adversarial Networks (GANs) gibi teknikler, gerçek dünya verilerini destekleyebilen gerçekçi görüntüler oluşturmak için kullanılabilir (20).

5. Veri Gizliliği ve Güvenliği:

Toplanan verinin gizlilik ve güvenlik düzenlemelerine uygun olduğundan emin olunmalıdır. Hassas verileri korumak için hasta bilgileri de-identifiye edilmeli veya anonimleştirilmelidir (21).

6. Çok Merkezli İşbirliği:

Farklı coğrafi konumlardan, çeşitli hasta popülasyonlarından ve değişen görüntüleme ekipmanlarından veri toplamak için birden fazla tıbbi merkezle işbirliği yapılabilir. Bu, daha güçlü ve genelleştirilebilir bir yapay zeka modeli oluşturmaya yardımcı olur.

7. Longitudinal Veri:

Hastalıkların zaman içindeki ilerlemesini anlamak için longitudinal veri toplanabilir. Böylelikle yapay zeka modelinin farklı aşamalarda koşulları tahmin etme ve teşhis etme yeteneğini geliştirilebilir (22).

8. Klinik Çalışma Verileri:

Devam eden klinik deneme ve çalışmalarla işbirliği yaparak ilgili görüntü verilerine erişilebilir. Bu veriler, belirli koşullar hakkında değerli bilgiler sağlayabilir ve hedeflenmiş yapay zeka modellerinin geliştirilmesine katkıda bulunabilir (23,24).

9. Kalite Kontrol:

Toplanan verinin doğruluğunu ve güvenilirliğini sağlamak için sıkı kalite kontrol önlemleri uygulanmalıdır. Bu basamak, görüntü kalitesinin bozulduğu ileri derece artefaktlı görüntüyü bulmayı, yanlış etiketleme kontrol etmeyi ve model performansını etkileyebilecek diğer sorunları ele almayı içerir.

Tıbbi görüntü verilerini radyoloji alanındaki yapay zeka eğitimi için toplarken etik düşünceler, hasta gizliliği ve düzenleyici uyumluluk ön planda olmalıdır. Araştırmacılar, tıbbi verilere erişmeden önce etik kurallara uymalı ve gerekli onayları almalıdır.

2.1.3. VERİ SETLERİ NEDİR? EĞİTİM AMACIYLA NASIL KULLANILIR?

Makine öğreniminde; eğitim veri seti, doğrulama veri seti ve test veri seti terimleri, model geliştirme ve değerlendirme sürecinde farklı amaçlar için kullanılan veri setleri anlamına gelir. Bu veri setleri, bir makine öğrenme modelinin eğitimi, ayarlanması ve performansının değerlendirilmesi için kritiktir (25).

2.1.3.1. Eğitim Veri Seti:

Eğitim veri seti, makine öğrenme modelini eğitmek için kullanılan verinin bir bölümüdür. Model, bu veri seti içindeki desenleri ve ilişkileri öğrenir. Eğitim veri seti giriş-çıkış çiftlerinden oluşur, giriş verisi verinin özelliklerini veya niteliklerini temsil ederken çıkış, modelin tahmin etmeye çalıştığı karşılık gelen hedef veya etikettir. Model, eğitim veri setindeki bilgiye dayanarak parametrelerini ayarlar ve tahminleri ile gerçek hedef değerleri arasındaki farkı en aza indirmeye çalışır (26).

2.1.3.2. Doğrulama Veri Seti:

Doğrulama veri seti, modeli ayarlamak ve hiperparametrelerini optimize etmek için kullanılır. Bu, modelin eğitim sırasında görmediği ayrı bir veri seti olarak hareket eder (27). Modeli eğitirken, doğrulama veri setindeki performans izlenir. Bu, modelin eğitim verilerinde başarılı olmasına rağmen yeni ve görünmeyen verilere genelleme yapma yeteneğini kaybetmesini önlemenin bir yoludur. Modelin mimarisi veya hiperparametreleri, doğrulama veri setindeki performansına dayanarak ayarlanır. Bu süreç genellikle tatmin edici bir model elde edilene kadar tekrarlanır (28).

2.1.3.3. Test Veri Seti:

Test veri seti, modelin eğitim veya doğrulama sırasında hiç karşılaşmadığı tamamen bağımsız bir veri setidir. Bu, modelin yeni, görülmemiş verilere ne kadar iyi genelleme yaptığını değerlendirmek için kullanılan son bir aşamadır (29). Test veri seti, modelin performansının tarafsız bir tahminini sağlar ve gerçek dünya verileri üzerinde doğru tahminler yapma yeteneğini belirlemede yardımcı olur. Test veri

setindeki sonuçlar, modelin genel performansını anlamak ve pratik uygulamalarda güvenilirliğini sağlamak için önemlidir.

Verinin bu üç sete (eğitim, doğrulama ve test) ayrılması, güçlü ve genelleştirilebilir makine öğrenme modelleri oluşturmak için önemlidir. Bu, modelde görülen overfitting gibi sorunları tanımlamaya ve ele almaya, ince ayar yapmaya ve bir modelin yeni veriler üzerindeki performansını güvenilir bir şekilde değerlendirmeye yardımcı olur.

2.1.4. BAŞARILI BİR YAPAY ZEKA EĞİTİMİNDE ÖNEMLİ NOKTALAR

Radyoloji uygulamaları için yapay zeka modellerini eğitmek karmaşık bir süreçtir ve çeşitli faktörleri dikkate almayı gerektirir. Radyolojide başarılı bir yapay zeka programı için düşünülmesi gereken bazı önemli noktalar:

1. Verinin Kalitesi ve Miktarı:

- Çeşitli ve Temsilci Veri: Eğitim veri seti, modelin karşılaşacağı nüfusu çeşitli ve temsilci olarak yansıtmalıdır. Bu, yapay zeka sisteminin farklı senaryolara iyi cevap vermesine yardımcı olur.

- Etiketlenmiş Veri: Deneyimli radyologlar tarafından yapılan yüksek kaliteli etiketlemeler, gözetimli öğrenme için hayati önem taşır (30). Doğru etiketlemeler, modeli eğitmek için temel gerçeklik görevi görür.

2. Veri Ön İşleme:

- Normalizasyon ve Standartlaştırma: Farklı görüntüleme modaliteleri arasında piksel değerlerinde ve yoğunluk seviyelerinde tutarlılık sağlanmalıdır.

- Veri Artırma: Döndürme, çevirme ve ölçeklendirme gibi teknikleri kullanarak eğitim veri setinin boyutunu yapay olarak artırılabilir (31).

3. Model Mimarisi:

- Özel Mimariler: Yapay zeka görevinin özel gdereksinimlerine göre sinir ağı mimarisi tasarlanmalıdır. Özellikle görüntü odaklı görevlerde başarılı olan konvolüsyonel sinir ağı (CNN) gibi mimarileri ön plana çıkmaktadır (32).

4. Transfer Öğrenme:

- Önceden Eğitilmiş Modeller: Büyük veri setlerinde önceden eğitilmiş modeller kullanılabilir. Böylelikle yeni veri seti ile model üzerinde bazı değişiklikler yapılarak daha önce öğrenilmiş özelliklerden faydalanılabilir (33).

5. Doğrulama ve Değerlendirme:

- Güçlü Doğrulama: Veri seti; eğitim, doğrulama ve test setlerine bölünebilir. Doğrulama setini hiperparametre ayarı ve aşırı uyumunu önleme amacıyla kullanılır.

- Klinik İlişkili Metrikler: Klinik hedeflerle uyumlu değerlendirme metrikleri modeli değerlendirmede önemlidir. Örneğin; duyarlılık, özgüllük ve ROC eğrisi altındaki alan.

6. Yorumlanabilirlik:

- Açıklanabilirlik: Karar alma sürecine içgörü sağlayan modelleri hedeflenmelidir. Bu, şeffaflığın esas olduğu tıbbi uygulamalarda hayati öneme sahiptir.

7. Düzenleyici Uyumluluk:

- Standartlara Uygunluk: Yapay zeka modeli, tıbbi görüntüleme için belirlenen ilgili düzenleyici standartlara ve kurallara uygun olmalıdır. Bu, FDA veya diğer bölgesel sağlık otoriteleri tarafından belirlenen standartları içerebilir.

8. Sürekli Öğrenme:

- Adaptasyon: Modellerin, zamanla güncellenen yeni veri ile güncellenebilecek şekilde geliştirilmesi önemlidir. Sürekli öğrenme, modele yeni desenlere ve tıbbi verideki değişikliklere adapte olma yeteneğine yardımcı olabilir.

9. Etik Düşünceler:

- Hasta Gizliliği: Hasta gizliliğini korumak ve veri koruma düzenlemelerine uymak için güçlü önlemler uygulanmalıdır.

- Önyargı ve Adillik: Veride ve model tahminlerinde potansiyel önyargıları ele alarak tüm hasta grupları için adil ve önyargısız sonuçlar sağlanmalıdır.

10. Klinik Entegrasyon ve Doğrulama:

- Klinik Denemeler: Gerçek dünya senaryolarında yapay zeka modelinin etkinliğini ve güvenliğini doğrulamak için klinik denemeler yapılmalıdır.

- İş Akışına Entegrasyon: Yapay zeka araçlarını radyoloji iş akışına sorunsuz bir şekilde entegre ederek verimlilik ve etkinlik artırılmalıdır.

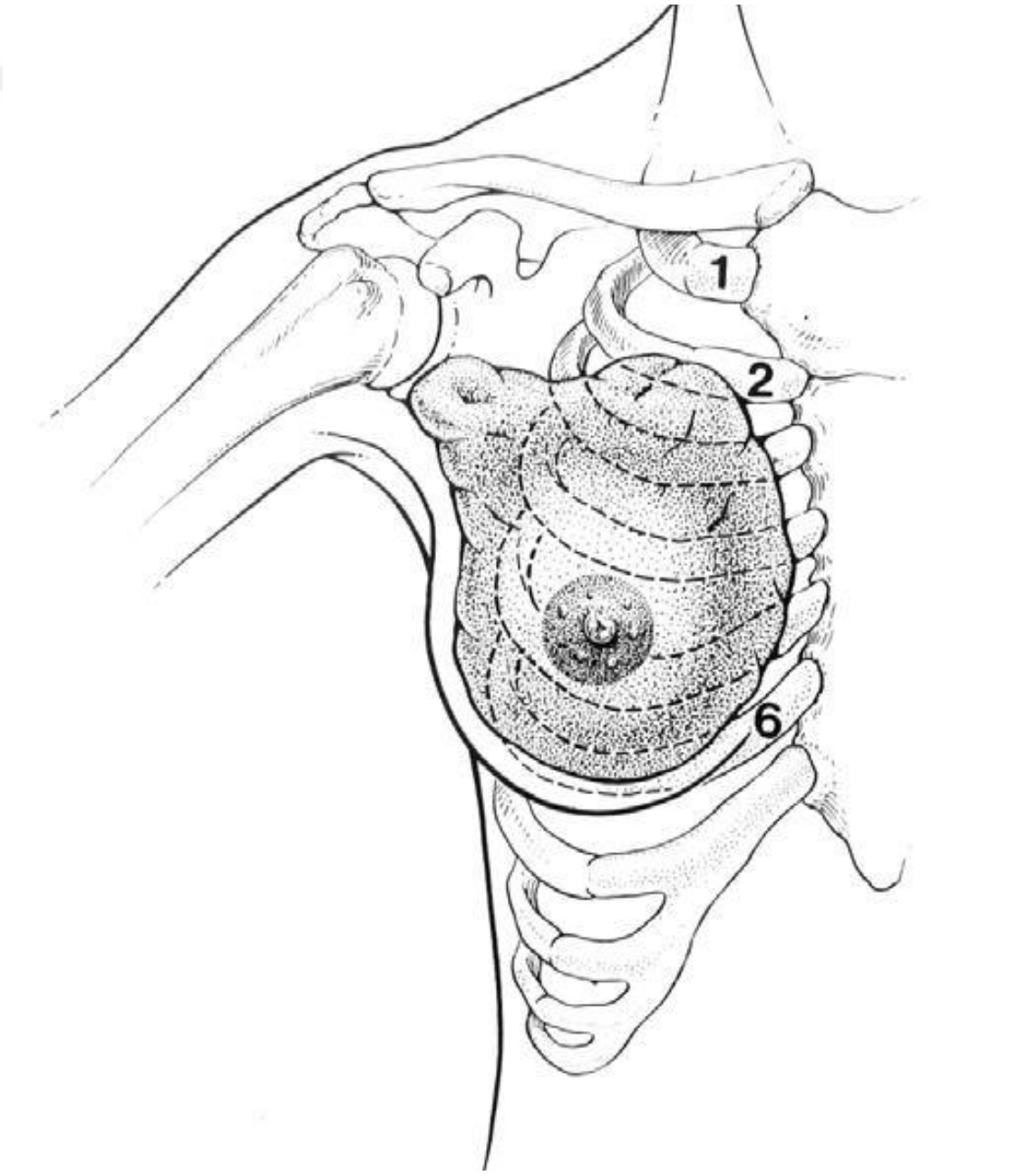
Radyolojide başarılı bir yapay zeka eğitimi, bilgisayar bilimcileri, radyologlar ve diğer sağlık profesyonelleri arasında çok disiplinli bir yaklaşım gerektirir. Ayrıca, düzenleyici kurumlarla sürekli işbirliği ve etik standartlara uyum, yapay zekanın klinik ortamlarda başarılı bir şekilde kullanımı için esastır.

2.2. MEME ANATOMİSİ

Meme dokusu cilt, cilt altı yağ dokusu, meme parankimi ve stromal dokudan meydana gelir. Üst dış kadranda meme dokusu daha yoğun olarak izlenmektedir. Meme dokusunun yaklaşık dörtte üçü pektoralis major kası üzerinde yerleşmektedir. Medialde rektus kılıfının üst kısmını, inferiorda kısmen serratus anterior ve eksternal oblik kasını, lateralde serratus anterior kasını örter (34). Meme dokusu göğüs ön duvarında en önde bulunan süperfisyal faysa içerisinde yerleşiktir. Bu süperfisyal fasya yüzeyel ve derin olmak üzere ikiye ayrılmaktadır. Meme dokusunun posteriorunu superfisyel fasyanın derin kısmı, anteriorunu yüzeyel fasyanın üst kısmı örter. Midaksiller çizgiyle memenin lateral kenarı, sternum ile de medial kenarı sınırlanır. Memenin superior kısmında belirgin olan, meme başına ve deriye doğru uzanan yüzeyel fasyanın üst kısmında, fibröz bantlardan oluşan fibröz lifler mevcuttur. Cooper ligamanları denilen bu yapılar meme dokusunu septasyonlara

ayırmakta ve parankim dokusuna destek vermektedir. Malign meme lezyonlarındaki meme derisinde çekintiler bu yapıların infiltre olması ile gelişen fibrozisinden kaynaklanır. Cooper ligamanlarındaki bu infiltrasyon memede oluşan portakal kabuğu görünümünün nedenidir (35,36,37). Meme dokusu genellikle her iki tarafta simetrik şekilde ve posteriorunda bağ dokusu nedeniyle hareket edebilir şekildedir. 2. kostadan başlar ve 6-7. interkostal alana kadar uzanır.

Şekil 1: Meme Anatomisi



2.3. MAMOGRAFİ VE MEME DANSİTESİ

2.3.1. MAMOGRAFİ NEDİR VE NE İÇİN KULLANILIR?

Mammografi, özel olarak meme dokusunun incelenmesi için tasarlanmış bir radyografik görüntüleme tekniğidir. Meme kanseri tanısında temel bir araç olarak kullanılan bu yöntem, hem rutin tarama hem de tanısal incelemeler için kullanılır. Temel amaç, meme iç yapısının detaylı görüntülerini oluşturarak tümör veya kalsifikasyon gibi patolojik özelliklerin tanımlanmasını kolaylaştırmaktır.

Mammografi, düşük dozda X ışınları kullanarak meme dokusunun radyografik görüntülerini oluşturmayı içerir. Meme dokusunun kalınlığını azaltmak ve görüntü netliğini artırmak için meme, iki plaka arasında sıkıştırılır. Hem geleneksel film tabanlı mammografi hem de dijital mammografi teknikleri yaygın olarak kullanılır.

Meme Kanseri Taraması:

Mammografi, özellikle semptomsuz popülasyonlarda meme kanseri için bir tarama aracı olarak rutin olarak kullanılır. Hedef, potansiyel malignitelerin erken tespiti ve böylece zamanında müdahale ve geliştirilmiş prognoz sağlamaktır.

Tanısal Görüntüleme:

Anormalliklerin tespit edildiği durumlarda, tarama veya klinik belirtiler nedeniyle, tanısal mammografi yapılır. Tanısal görüntüleme, daha ayrıntılı ve odaklı görüntüler sağlayarak tanımlanan anormalliklerin daha fazla karakterizasyonu ve değerlendirmesine yardımcı olur.

Mammografi, meme kanseri ile mücadelede temel bir tanı aracı olarak önemli bir rol oynar ve bu yaygın kanserin erken tespiti ve yönetimine önemli katkılarda bulunur. Rutin tarama programlarına entegrasyonu, meme kanseri kaynaklı morbidite ve mortaliteyi azaltmaya yönelik kamu sağlığı girişimlerinde kritik bir rol oynar.

2.3.2. MAMOGRAFİDE MEME DANSİTESİ NEDİR? NASIL SINIFLANDIRILIR?

Meme dansitesi, mammografi aracılığıyla değerlendirilen, X-ışınlarının çeşitli bileşenler tarafından farklı derecelerde soğurulmasına dayanan meme dokusunun radyografik görünümünü ifade eder. Mammografi görüntüleri, X-ışınlarının memedeki dokularla etkileşimi sonucunda oluşur; bu dokular arasında yağlı doku, glandüler doku ve fibroz stroma bulunur.

Yağlı Doku:

Adipöz doku, radyolusenttir, bu da X-ışınlarının daha kolay nüfuz etmesine olanak tanır ve mammogramda daha koyu alanlara neden olur.

Glandüler ve Fibroz Doku:

Glandüler ve fibroz dokular daha yoğun olup X-ışınlarını daha fazla soğururlar, mammogramda daha beyaz görünen radiopak bölgeler olarak ortaya çıkar.

MAMOGRAFİDE MEME DANSİTE SINIFLAMASI

Meme dansitesi, glandüler ve fibroz doku oranının yağlı dokuya göre olan oranına dayalı olarak dört sınıfa ayrılır:

1. **Neredeyse Tamamen Yağlı – Tip A Meme**

Radyolusent yağlı dokunun hakim olduğu, çevre yapılarla kontrastın arttığı lezyonların tanımlanmasını kolaylaştıran bir durum.

2. **Dağınık Fibroglandüler Dansite Yapısı – Tip B Meme**

Artmış dansiteye sahip dağınık alanların varlığı, meme boyunca glandüler ve fibroz dokunun karışımını gösterir.

3. **Heterojen Dansite – Tip C Meme**

Glandüler ve fibroz dokunun yüksek oranda bulunduğu, genel radyopaklığın artması nedeniyle normal ve patolojik yapılar arasındaki ayrımın zorlaştığı durum.

4. İleri Derece Yoğun Meme – Tip D Meme

Yoğun glandüler ve fibroz dokunun hakim olduğu, anormalliklerin çevresel dokularla karıştığı, meme kanseri riskinin arttığı bir durum (38,39).

2.3.3. MAMOGRAFİDE MEME DANSİTESİNİN ÖNEMİ

Meme dansitesi, bir mamogramda görülen meme dokusunun, yağ ve bez dokusu gibi farklı yoğunluktaki bileşenlerin oranını ifade eder. Meme dansitesi önemlidir çünkü meme kanseri riskini değerlendirmeye yardımcı olabilir ve mamogramın yorumlanmasında önemli bir faktördür.

Meme Dansitesi ve Mamografi İlişkisi:

Mamogramlar, meme dokusunun yoğunluğunu gösteren bir ölçüm sunar. Yoğun meme dokusu, beyaz renkte görünürken, yağ dokusu daha gri renkte görünür. Meme dansitesi, bir kadının meme dokusundaki yoğun ve düşük yoğunluktaki alanların oranını ifade eder.

Yoğun Meme Dokusu ve Risk:

Yoğun meme dokusuna sahip olmak, meme kanseri gelişme riskini artırabilir. Yoğun meme dokusu, mamogramda kanser lezyonlarını tespit etmeyi zorlaştırabilir, çünkü kanser lezyonları da beyaz renkte görünür.

Düşük Yoğunluklu Meme Dokusu ve Mamogramın Etkinliği:

Düşük yoğunluklu meme dokusuna sahip bireylerde, mamogramlar daha etkili olabilir. Daha düşük yoğunluktaki dokular arasında kanser lezyonları daha belirgin olabilir, bu da erken tespit ve doğru teşhis için avantaj sağlayabilir.

Risk Değerlendirmesi ve Takip:

Yoğun meme dokusuna sahip bireylerde, ek tarama yöntemleri veya daha sık mamografi taramaları gibi önlemler alınabilir. Bu, meme kanseri riskinin daha dikkatli bir şekilde değerlendirilmesine ve takip edilmesine yardımcı olabilir.

Sonuç olarak, meme dansitesi, mamografi sonuçlarını değerlendirmede önemli bir faktördür ve meme kanseri riskini değerlendirmek için kullanılır. Ancak, yoğunluk ve risk arasındaki ilişki karmaşıktır, ve bu nedenle bireylerin sağlık profesyonelleriyle bu konuda iletişim kurmaları önemlidir. Sağlık uzmanları, bireyin özel durumunu değerlendirerek en uygun tarama ve takip stratejilerini belirleyebilirler.

3. GEREÇ VE YÖNTEM

3.1. Hasta Seçimi

Ocak 2021 – Temmuz 2021 tarihleri arasında İstanbul Maltepe Üniversitesi Tıp Fakültesi Hastanesi Radyoloji Anabilim Dalında; meme kanseri taraması amacı ile mamografi yapılan olgular retrospektif olarak tarandı ve PACS 'ta 754 hasta kaydına ulaşıldı. Kayıtlardan 41 tanesi aynı hastaya ait olduğu belirlendi (sağ, sol meme girişi vb.) ve bu hastalar tek vaka olarak kabul edildi. Kalan hastaların mamografi görüntüsünde; 13 tanesinde artefakt, 29 tanesinde implant, 110 tanesinde patoloji izlendi ve çalışmadan dışlandı. Ayrıca 89 hastadanın mamografi görüntüsünde post-op değişiklikler izlenmiş olup bu hastalar çalışmadan dışlanmıştır. Kalan 472 hasta çalışmaya dahil edildi. Olguların yaşları 30-88 arasında, ortalama yaş 51'idi. Tüm olgular kliniğimizde FFDM (Full Field Dijital Mamografi) tekniği ile 4 pozisyon (2CC, 2 MLO) olarak yapıldı. Bu görüntüler meme konusunda uzman radyolog, radyoloji asistanı ve (LUNİT Insight MMG) tarafından değerlendirildi ve veriler kaydedildi.

Çalışma için Marmara Üniversitesi Tıp Fakültesi Bilimsel Araştırma ve Bilim Kurulundan, 12.01.2024 tarihli, 09.2024.150 protokol kodlu etik kurul onayı alındı.

3.2. Mamografi Çekim Protokolü

Tüm mamografi çekimleri FFDM- Full Field Dijital Mamografi (Mammomat 1000, Siemens Medical Systems, Germany) cihazında yapılmıştı. Mamografi görüntüleri 2 Kraniokaudal ve 2 Mediolateral Oblik olmak üzere 4 poz standart olarak alınmış ve PACS (Picture Archiving and Communication Systems) sistemine kaydedilmişti.

3.3. Mamografi Görüntülerinin Değerlendirilmesi ve İstatistiksel Analiz

Verilerin analizi IBM SPSS Statistics *ver.* 25 (IBM Corporation, Armonk, NY, US) paket programında yapıldı. Tanımlayıcı istatistikler; kantitatif değişkenler için ortalama ± standart sapma biçiminde ifade edilirken kalitatif değişkenler ise olgu sayısı ve (%) şeklinde gösterildi. Kategorik verilerin analizlerinde aksi belirtilmedikçe Pearson'un χ^2 testi kullanıldı. 2x2'lik çapraz tablolarda gözelerin en az ¼'ünde beklenen frekansın 5'in altında olması durumunda söz konusu kategorik veriler Fisher'in kesin sonuçlu olasılık testiyle değerlendirilirken beklenen frekansın 5-25 arasında olduğu durumlarda Süreklilik düzeltmeli χ^2 testi kullanıldı. RxC (satır ya da kolondaki kategorik değişkenlerden en az birinin ikiden fazla sonuçlu olması durumunda) çapraz tablolarda gözelerin en az ¼'ünde beklenen frekansın 5'in altında olması durumunda söz konusu kategorik veriler Fisher Freeman Halton testi ile incelendi. Meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların kıdemli radyolog ile olan uyumu Kappa (κ) katsayısı ve önemlilik düzeyi hesaplanarak incelendi. Kappa katsayısının 0.00-0.20 arasında değişmesi gözlemciler arasında uyumun olmadığı, 0.21-0.39 aralığı gözlemciler arasında çok zayıf uyumun olduğu, 0.40-0.59 aralığı gözlemciler arasında zayıf uyumu, 0.60-0.79 arasında olması orta düzeyde uyumun olduğu, 0.80-0.90 arasında yer alması gözlemciler arasında güçlü uyumu gösterirken 0.90'ın üzerinde olması çok güçlü uyumu ifade etmektedir[i].

Meme yoğunluğu açısından kıdemli radyolog ile olan tutarlılık düzeyleri yönünden yapay zeka ile radyoloji asistanı arasında istatistiksel olarak anlamlı fark olup olmadığı ise McNemar testiyle değerlendirildi. $p<0,05$ için sonuçlar istatistiksel olarak anlamlı kabul edildi.

[1] McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276-282.

4. BULGULAR

Mevcut çalışmada yaşları 30 ile 88 arasında değişen 472 olgunun verileri değerlendirilmiştir. Olguların yaş ortalaması 50,2±9,7 (yıl) olup 43'ü (%9,1) 40 yaş altı, 214'ü (%45,3) 40-49 yaş arası, 127'si (%26,9) 50-59 yaş arası, 64'ü (%13,6) 60-69 yaş arası, 24'ü (%5,1) ise 70 yaş ve üzerindeydi.

Tablo 1. Çalışmaya dahil edilen olguların yaşlarına ait tanımlayıcı istatistikler

	n=472
Yaş (yıl)	50,2±9,7
Yaş aralığı (yıl)	30-88
Yaş Grupları	
<40 yıl	43 (%9,1)
40-49 yıl	214 (%45,3)
50-59 yıl	127 (%26,9)
60-69 yıl	64 (%13,6)
≥70 yıl	24 (%5,1)

Tablo 2'de meme yoğunluğu açısından kıdemli radyolog, yapay zeka ve radyoloji asistanı tarafından yapılan sınıflama yönünden olguların frekans dağılımları gösterilmiştir.

Tablo 2. Meme yoğunluğu açısından kıdemli radyolog, yapay zeka ve radyoloji asistanı tarafından yapılan sınıflama yönünden olguların frekans dağılımları

	n=472
Kıdemli Radyolog	
Yağlı	74 (%15,7)
Dağılmış	189 (%40,0)
Heterojen	183 (%38,8)
Yoğun	26 (%5,5)
Yapay Zeka	
Yağlı	82 (%17,4)
Dağılmış	192 (%40,7)
Heterojen	167 (%35,3)
Yoğun	31 (%6,6)
Radyoloji Asistanı	
Yağlı	75 (%15,9)
Dağılmış	189 (%40,0)
Heterojen	150 (%31,8)
Yoğun	58 (%12,3)

Tablo 3'te ise memenin yoğun olup olmaması açısından kıdemli radyolog, yapay zeka ve radyoloji asistanı tarafından yapılan sınıflama yönünden olguların frekans dağılımları verilmiştir.

Tablo 3. Memenin yoğun olup olmaması açısından kıdemli radyolog, yapay zeka ve radyoloji asistanı tarafından yapılan sınıflama yönünden olguların frekans dağılımları

	n=472
Kıdemli Radyolog	
Nondense	263 (%55,7)
Dense	209 (%44,3)
Yapay Zeka	
Nondense	274 (%58,1)
Dense	198 (%41,9)
Radyoloji Asistanı	
Nondense	264 (%55,9)
Dense	208 (%44,1)

Tablo 4'te meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı incelenmiştir.

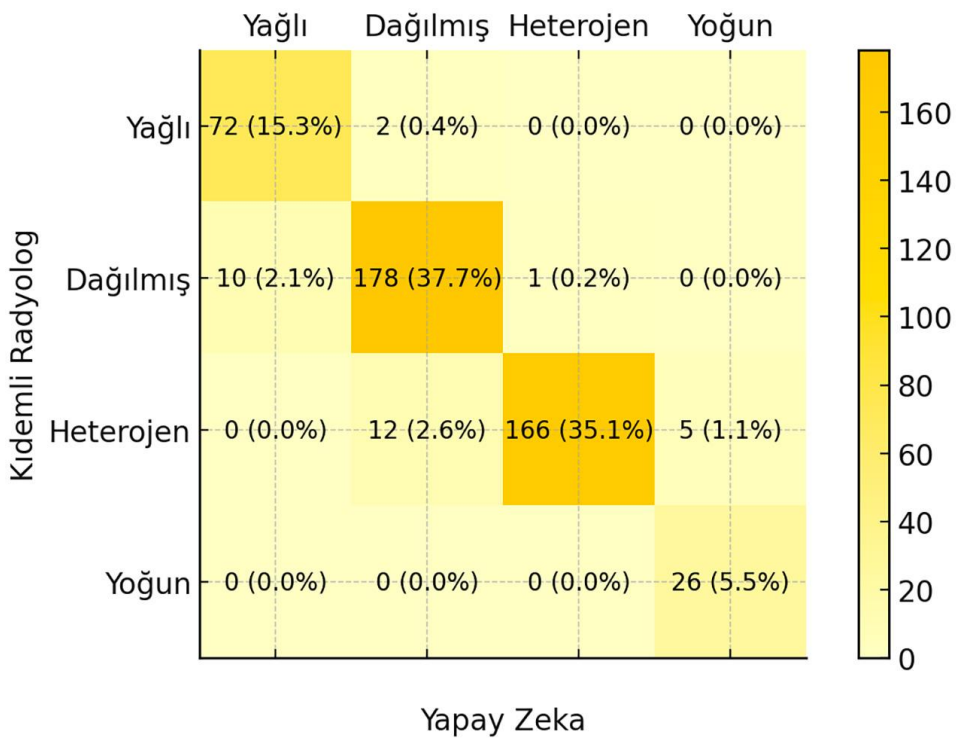
Tablo 4. Meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

	Kıdemli Radyolog					κ katsayısı †	p-değeri
	Yağlı	Dağılmış	Heterojen	Yoğun	Toplam		
Yapay Zeka						0,905±0,017	**<0,001**
Yağlı	72 (%15,3)	10 (%2,1)	-	-	82 (%17,4)		
Dağılmış	2 (%0,4)	178 (%37,7)	12 (%2,6)	-	192 (%40,7)		
Heterojen	-	1 (%0,2)	166 (%35,1)	-	167 (%35,3)		
Yoğun	-	-	5 (%1,1)	26 (%5,5)	31 (%6,6)		
Toplam	74 (%15,7)	189 (%40,0)	183 (%38,8)	26 (%5,5)	472 (%100,0)		
Asistan						0,783±0,023	**<0,001**
Yağlı	66 (%14,0)	9 (%1,9)	-	-	75 (%15,9)		
Dağılmış	8 (%1,7)	170 (%36,0)	11 (%2,3)	-	189 (%40,0)		
Heterojen	-	10 (%2,1)	140 (%29,7)	-	150 (%31,8)		
Yoğun	-	-	32 (%6,8)	26 (%5,5)	58 (%12,3)		
Toplam	74 (%15,7)	189 (%40,0)	183 (%38,8)	26 (%5,5)	472 (%100,0)		

† Uyum sonuçlarına ait istatistikler; κ katsayısı ve ± std.hata biçiminde verilmiştir.

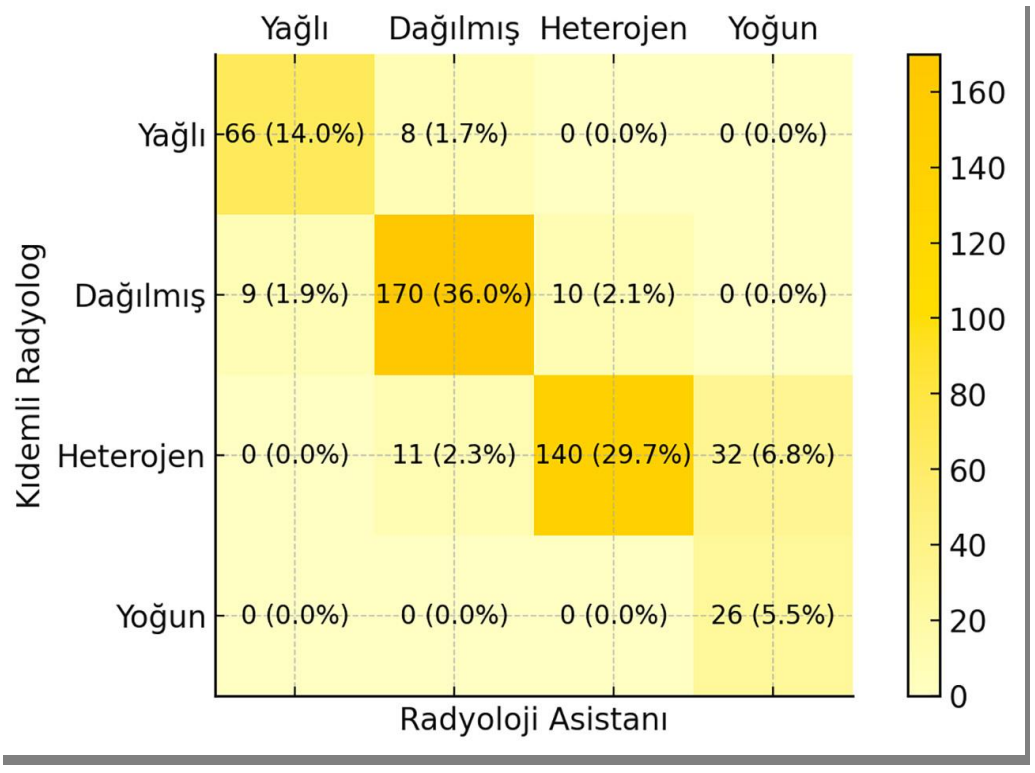
Kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,905; Std.Hata=0,017 ve p<0,001). Mevcut değerlendirmeye göre kıdemli radyoloğun yağlı olarak sınıfladığı 2 vaka (%0,4) YZ tarafından yanlışlıkla dağılmış olarak sınıflanırken kıdemli radyoloğun dağılmış olarak sınıfladığı vakalardan 10 tanesini (%2,1) YZ yağlı olarak 1'ini (%0,2) ise heterojen olarak sınıflamıştır. Son olarak kıdemli radyoloğun heterojen olarak sınıfladığı 12 vakayı (%2,6) YZ dağılmış olarak 5'ini (%1,1) ise yoğun olarak sınıflamıştır. Geriye kalan diğer tüm durumlarda kıdemli radyolog ile YZ bire-bir uyum göstermiştir (bkz Şekil 2).

Şekil 2: Kıdemli Radyolog ile Yapay Zekanın Meme Yoğunluğu Sınıflaması



Kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile radyoloji asistanının yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan kabul edilebilir düzeyde uyumlu olduğu görüldü (κ=0,783; Std.Hata=0,023 ve p<0,001). Mevcut değerlendirmeye göre kıdemli radyoloğun yağlı olarak sınıfladığı 8 vaka (%1,7) radyoloji asistanı tarafından yanlışlıkla dağılmış olarak sınıflanırken kıdemli radyoloğun dağılmış olarak sınıfladığı vakalardan 9 tanesini (%1,9) radyoloji asistanı yağlı olarak 10 tanesini (%2,1) ise heterojen olarak sınıflamıştır. Son olarak kıdemli radyoloğun heterojen olarak sınıfladığı 11 vakayı (%2,3) radyoloji asistanı dağılmış olarak 32'sini (%6,8) ise yoğun olarak sınıflamıştır. Geriye kalan diğer tüm durumlarda kıdemli radyolog ile radyoloji asistanı bire-bir uyum göstermiştir (bkz Şekil 3).

Şekil 3: Kıdemli Radyolog ile Radyoloji Asistanının Meme Yoğunluğu Sınıflaması



Tablo 5'te ise memenin yoğun olup olmaması açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı değerlendirilmiştir.

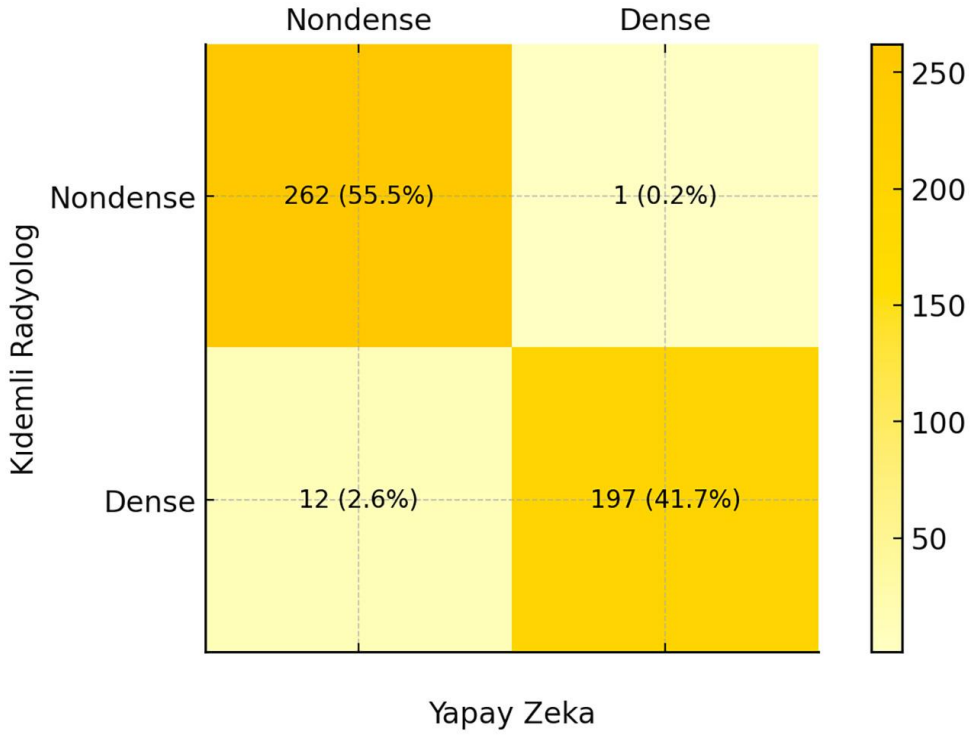
Tablo 5. Memenin yoğun olup olmaması açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

	Kıdemli radyolog			**κ katsayısı †**	**p-değeri**
	Nondense	Dense	Toplam		
Yapay Zeka				0,944±0,015	**<0,001**
Nondense	262 (%55,5)	12 (%2,6)	274 (%58,1)		
Dense	1 (%0,2)	197 (%41,7)	198 (%41,9)		
Toplam	263 (%55,7)	209 (%44,3)	472 (%100,0)		
Asistan				0,910±0,019	**<0,001**
Nondense	253 (%53,5)	11 (%2,4)	264 (%55,9)		
Dense	10 (%2,2)	198 (%41,9)	208 (%44,1)		
Toplam	263 (%55,7)	209 (%44,3)	472 (%100,0)		

† Uyum sonuçlarına ait istatistikler; κ katsayısı ve ± std.hata biçiminde verilmiştir.

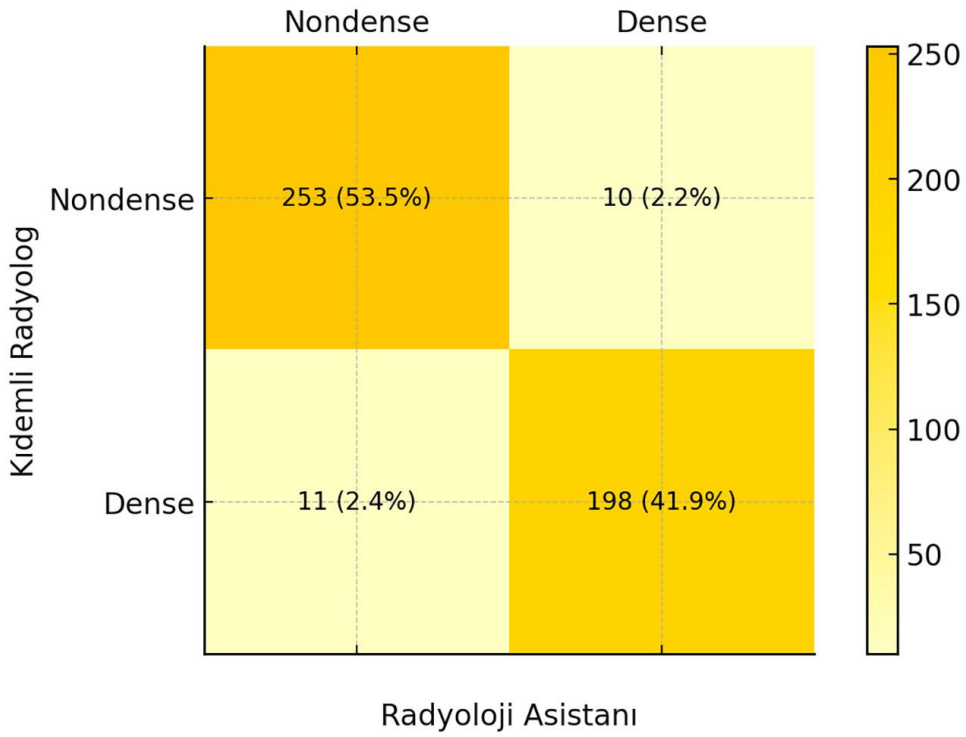
Kıdemli radyolog tarafından memenin yoğun olup olmama durumu ile yapay zekanın (YZ) yapmış olduğu sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,944; Std.Hata=0,015 ve p<0,001). Mevcut değerlendirmeye göre kıdemli radyoloğun nondense olarak sınıfladığı 1 vaka (%0,2) YZ tarafından yanlışlıkla dense olarak sınıflanırken kıdemli radyoloğun dense olarak sınıfladığı vakalardan 12'sinin (%2,6) ise YZ tarafından yanlışlıkla nondense olarak sınıflandığı görülmüştür (bkz Şekil 4).

Şekil 4: Kıdemli Radyolog ile Yapay Zekanın Dens-Nondens Meme Yoğunluğu Sınıflaması



Kıdemli radyolog tarafından memenin yoğun olup olmama durumu ile radyoloji asistanının yapmış olduğu sınıflamasının da istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,910; Std.Hata=0,019 ve p<0,001). Mevcut değerlendirmeye göre kıdemli radyoloğun nondense olarak sınıfladığı 10 vaka (%2,2) radyoloji asistanı tarafından yanlışlıkla dense olarak sınıflanırken kıdemli radyoloğun dense olarak sınıfladığı vakalardan 11'inin (%2,4) ise radyoloji asistanı tarafından yanlışlıkla nondense olarak sınıflandığı görülmüştür (bkz Şekil 5).

Şekil 5: Kıdemli Radyolog ile Radyoloji Asistanının Dens-Nondens Meme Yoğunluğu Sınıflaması



Tablo 6'da her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka tarafından yapılan sınıflamaların klinik performansı yer almaktadır.

Tablo 6. Her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka tarafından yapılan sınıflamaların klinik performansı

Y.Zeka	**Kıdemli radyolog**					**κ katsayısı †**	**p-değeri**
	Yağlı	Dağılmış	Heterojen	Yoğun	Toplam		
<40 yıl						0,883±0,065	**<0,001**
Yağlı	-	-	-	-	-		
Dağılmış	-	10 (%23,3)	1 (%2,3)	-	11 (%25,6)		
Heterojen	-	1 (%2,3)	22 (%51,2)	-	23 (%53,5)		
Yoğun	-	-	1 (%2,3)	8 (%18,6)	9 (%20,9)		
Toplam	-	11 (%25,6)	24 (%55,8)	8 (%18,6)	43 (%100,0)		
40-49 yıl						0,890±0,027	**<0,001**
Yağlı	20 (%9,3)	4 (%1,9)	-	-	24 (%11,2)		
Dağılmış	1 (%0,5)	64 (%29,9)	7 (%3,2)	-	72 (%33,6)		
Heterojen	-	-	99 (%46,3)	-	99 (%46,3)		
Yoğun	-	-	3 (%1,4)	16 (%7,5)	19 (%8,9)		
Toplam	21 (%9,8)	68 (%31,8)	109 (%50,9)	16 (%7,5)	214 (%100,0)		
50-59 yıl						0,923±0,031	**<0,001**
Yağlı	18 (%14,2)	2 (%1,6)	-	-	20 (%15,8)		
Dağılmış	1 (%0,8)	63 (%49,5)	2 (%1,6)	-	66 (%51,9)		
Heterojen	-	-	38 (%29,9)	-	38 (%29,9)		
Yoğun	-	-	1 (%0,8)	2 (%1,6)	3 (%2,4)		
Toplam	19 (%15,0)	65 (%51,1)	41 (%32,3)	2 (%1,6)	127 (%100,0)		
≥60 yıl						0,881±0,047	**<0,001**
Yağlı	34 (%38,6)	4 (%4,5)	-	-	38 (%43,1)		
Dağılmış	-	41 (%46,6)	2 (%2,3)	-	43 (%48,9)		
Heterojen	-	-	7 (%8,0)	-	7 (%8,0)		
Yoğun	-	-	-	-	-		
Toplam	34 (%38,6)	45 (%51,1)	9 (%10,3)	-	88 (%100,0)		

† Uyum sonuçlarına ait istatistikler; κ katsayısı ve ± std.hata biçiminde verilmiştir.

40 yaş altı bireylerde kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,883; Std.Hata=0,065 ve p<0,001). 40-49 yaşları arasındaki bireylerde de kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,890; Std.Hata=0,027 ve p<0,001). 50-59 yaşları arasındaki bireylerde kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,923; Std.Hata=0,031 ve p<0,001). Son olarak 60 yaş ve üzeri (olguların ileri yaş gruplarındaki sayıları yetersiz olduğundan 70 yaş ve üzeri grup ile 60-69 yaşları arasındaki grup birleştirilerek 60 yaş ve üzeri şeklinde değerlendirme yapılmıştır) bireylerde de kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,881; Std.Hata=0,047 ve p<0,001).

Tablo 7'de ise her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı incelenmiştir.

Tablo 7. Her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

Asistan	**Kıdemli radyolog**					**κ katsayısı †**	**p-değeri**
	Yağlı	Dağılmış	Heterojen	Yoğun	Toplam		
<40 yıl						0,694±0,098	**<0,001**
Yağlı	-	-	-	-	-		
Dağılmış	-	8 (%18,6)	1 (%2,3)	-	9 (%20,9)		
Heterojen	-	3 (%7,0)	19 (%44,2)	-	22 (%51,2)		
Yoğun	-	-	4 (%9,3)	8 (%18,6)	12 (%27,9)		
Toplam	-	11 (%25,6)	24 (%55,8)	8 (%18,6)	43 (%100,0)		
40-49 yıl						0,717±0,039	**<0,001**
Yağlı	18 (%8,4)	4 (%1,9)	-	-	22 (%10,3)		
Dağılmış	3 (%1,4)	60 (%28,0)	6 (%2,8)	-	69 (%32,2)		
Heterojen	-	4 (%1,9)	79 (%36,9)	-	83 (%38,8)		
Yoğun	-	-	24 (%11,2)	16 (%7,5)	40 (%18,7)		
Toplam	21 (%9,8)	68 (%31,8)	109 (%50,9)	16 (%7,5)	214 (%100,0)		
50-59 yıl						0,861±0,039	**<0,001**
Yağlı	17 (%13,4)	2 (%1,6)	-	-	19 (%15,0)		
Dağılmış	2 (%1,6)	61 (%48,0)	1 (%0,8)	-	64 (%50,4)		
Heterojen	-	2 (%1,6)	36 (%28,3)	-	38 (%29,9)		
Yoğun	-	-	4 (%3,1)	2 (%1,6)	6 (%4,7)		
Toplam	19 (%15,0)	65 (%51,1)	41 (%32,3)	2 (%1,6)	127 (%100,0)		
≥60 yıl						0,800±0,059	**<0,001**
Yağlı	31 (%35,2)	3 (%3,4)	-	-	34 (%38,6)		
Dağılmış	3 (%3,4)	41 (%46,6)	3 (%3,4)	-	47 (%53,4)		
Heterojen	-	1 (%1,1)	6 (%6,9)	-	7 (%8,0)		
Yoğun	-	-	-	-	-		
Toplam	34 (%38,6)	45 (%51,1)	9 (%10,3)	-	88 (%100,0)		

† Uyum sonuçlarına ait istatistikler; κ katsayısı ve ± std.hata biçiminde verilmiştir.

40 yaş altı bireylerde kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile radyoloji asistanının yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan orta derecede uyumlu olduğu görüldü (κ=0,694; Std.Hata=0,098 ve p<0,001). 40-49 yaşları arasındaki bireylerde de kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile radyoloji asistanının yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan orta derecede uyumlu olduğu görüldü (κ=0,717; Std.Hata=0,039 ve p<0,001). 50-59 yaşları arasındaki bireylerde ise kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile radyoloji asistanının yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,861; Std.Hata=0,039 ve p<0,001). Son olarak 60 yaş ve üzeri (olguların ileri yaş gruplarındaki sayıları yetersiz olduğundan 70 yaş ve üzeri grup ile 60-69 yaşları arasındaki grup birleştirilerek 60 yaş ve üzeri şeklinde değerlendirme yapılmıştır) bireylerde kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile radyoloji asistanının yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan kabul edilebilir düzeyde uyumlu olduğu görüldü (κ=0,800; Std.Hata=0,059 ve p<0,001).

Tablo 8'de her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle olan karşılaştırılması verilmiştir.

Tablo 8. Her bir yaş grubu içerisinde meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle karşılaştırılması

Asistan	**Yapay zeka**			**p-değeri †**
	Uyumlu	Uyumlu değil	Toplam	
<40 yıl				0,063
Uyumlu	35 (%81,4)	-	35 (%81,4)	
Uyumlu değil	5 (%11,6)	3 (%7,0)	8 (%18,6)	
Toplam	40 (%93,0)	3 (%7,0)	43 (%100,0)	
40-49 yıl				**<0,001**
Uyumlu	171 (%79,9)	2 (%0,9)	173 (%80,8)	
Uyumlu değil	28 (%13,1)	13 (%6,1)	41 (%19,2)	
Toplam	199 (%93,0)	15 (%7,0)	214 (%100,0)	
50-59 yıl				0,125
Uyumlu	115 (%90,6)	1 (%0,8)	116 (%91,4)	
Uyumlu değil	6 (%4,7)	5 (%3,9)	11 (%8,6)	
Toplam	121 (%95,3)	6 (%4,7)	127 (%100,0)	
≥60 yıl				0,219
Uyumlu	77 (%87,5)	1 (%1,1)	78 (%88,6)	
Uyumlu değil	5 (%5,7)	5 (%5,7)	10 (%11,4)	
Toplam	82 (%93,2)	6 (%6,8)	88 (%100,0)	
Genel				**<0,001**
Uyumlu	398 (%84,3)	4 (%0,9)	402 (%85,2)	
Uyumlu değil	44 (%9,3)	26 (%5,5)	70 (%14,8)	
Toplam	442 (%93,6)	30 (%6,4)	472 (%100,0)	

† McNemar testi.

Mevcut değerlendirmede kıdemli radyoloğun yapmış olduğu sınıflamada yağlı, dağılmış, heterojen ve yoğun olarak sınıflanan bireylerin gerek YZ gerekse radyoloji asistanı tarafından bire bir aynı kategoride sınıflanması dikkate alınmıştır. Tüm olgular içerisinde yapay zekanın kıdemli radyolog ile olan tutarlılığı %93,6 iken radyoloji asistanının kıdemli radyolog ile olan tutarlılığı %85,2 olup YZ'ya karşı radyoloji asistanının istatistiksel anlamlı olarak kıdemli radyolog ile daha az tutarlılık gösterdiği saptanmıştır ($p<0,001$). Benzer durum 40-49 yaş grubunda da geçerli olup bu yaş grubu içerisinde yapay zekanın kıdemli radyolog ile olan tutarlılığı %93,0 iken radyoloji asistanının kıdemli radyolog ile olan tutarlılığı %80,8 düzeyinde kalmıştır ($p<0,001$). 40 yaş altı bireylerde de YZ'nin lehine bir durum söz konusu olmakla beraber bu yaş grubundaki olgu sayısı görece kısıtlı olduğundan YZ ile radyoloji asistanı arasında istatistiksel olarak anlamlı fark görülmemiştir ($p=0,063$). Son olarak gerek YZ'nın gerekse radyoloji asistanının kıdemli radyolog ile olan tutarlılıklarının peak yaptığı 50-59 yaş ve ≥60 yaş grupları içerisinde YZ ile radyoloji asistanı arasında kıdemli radyolog ile olan tutarlılıklarının istatistiksel olarak benzer olduğu görüldü ($p=0,125$ ve $p=0,219$).

Tablo 9'da meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyum durumları yönünden yaş grupları arasında yapılan karşılaştırmalar bulunmaktadır.

Tablo 9. Meme yoğunluğu açısından kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyum durumları yönünden yaş grupları arasında yapılan karşılaştırmalar

	<40 yıl (n=43)	**40-49 yıl (n=214)**	**50-59 yıl (n=127)**	**≥60 yıl (n=88)**	**p-değeri †**
Yapay zeka					0,854
Uyumlu	40 (%93,0)	199 (%93,0)	121 (%95,3)	82 (%93,2)	
Uyumlu değil	3 (%7,0)	15 (%7,0)	6 (%4,7)	6 (%6,8)	
Asistan					**0,040**
Uyumlu	35 (%81,4)	173 (%80,8)[a]	116 (%91,4)[a]	78 (%88,6)	
Uyumlu değil	8 (%18,6)	41 (%19,2)[a]	11 (%8,6)[a]	10 (%11,4)	

† Pearson'un χ^2 testi, a: 40-49 yıl ile 50-59 yıl arasındaki fark istatistiksel olarak anlamlı (p=0,014).

Yapay zekanın kıdemli radyolog ile olan uyum durumu yönünden yaş grupları arasında istatistiksel olarak anlamlı herhangi bir fark yoktu (p=0,854). Öte yandan radyoloji asistanının kıdemli radyolog ile olan uyum durumu yönünden yaş grupları arasında istatistiksel olarak anlamlı fark olup (p=0,040) söz konusu farka neden olan durum; 40-49 yaş grubuna göre 50-59 yaş grubunda radyoloji asistanının kıdemli radyolog ile olan uyumunun daha fazla olması idi (p=0,014).

Tablo 10'da her bir yaş grubu içerisinde memenin dense olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka tarafından yapılan sınıflamaların klinik performansı gösterilmiştir.

Tablo 10. Her bir yaş grubu içerisinde memenin dense olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka tarafından yapılan sınıflamaların klinik performansı

Yapay Zeka	**Kıdemli radyolog**			**κ katsayısı †**	**p-değeri**
	Nondense	Dense	Toplam		
<40 yıl				0,878±0,084	**<0,001**
Nondense	10 (%23,3)	1 (%2,3)	11 (%25,6)		
Dense	1 (%2,3)	31 (%72,1)	32 (%74,4)		
Toplam	11 (%25,6)	32 (%74,4)	43 (%100,0)		
40-49 yıl				0,933±0,025	**<0,001**
Nondense	89 (%41,6)	7 (%3,3)	96 (%44,9)		
Dense	-	118 (%55,1)	118 (%55,1)		
Toplam	89 (%41,6)	125 (%58,4)	214 (%100,0)		
50-59 yıl				0,964±0,025	**<0,001**
Nondense	84 (%66,1)	2 (%1,6)	86 (%67,7)		
Dense	-	41 (%32,3)	41 (%32,3)		
Toplam	84 (%66,1)	43 (%33,9)	127 (%100,0)		
≥60 yıl				0,863±0,095	**<0,001**
Nondense	79 (%89,8)	2 (%2,2)	81 (%92,0)		
Dense	-	7 (%8,0)	7 (%8,0)		
Toplam	79 (%89,8)	9 (%10,2)	88 (%100,0)		

† Uyum sonuçlarına ait istatistikler; κ katsayısı ve ± std.hata biçiminde verilmiştir.

40 yaş altı bireylerde memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile yapay zekanın (YZ) yapmış olduğu sınıflamanın istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,878; Std.Hata=0,084 ve p<0,001). 40-49 yaşları arasındaki bireylerde de memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile yapay zekanın (YZ) yapmış

olduğu sınıflamanın istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,933; Std.Hata=0,025 ve p<0,001). 50-59 yaşları arasındaki bireylerde memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile yapay zekanın (YZ) yapmış olduğu sınıflamanın istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,964; Std.Hata=0,025 ve p<0,001). Son olarak 60 yaş ve üzeri (olguların ileri yaş gruplarındaki sayıları yetersiz olduğundan 70 yaş ve üzeri grup ile 60-69 yaşları arasındaki grup birleştirilerek 60 yaş ve üzeri şeklinde değerlendirme yapılmıştır) bireylerde de memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile yapay zekanın (YZ) yapmış olduğu sınıflamanın istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,863; Std.Hata=0,095 ve p<0,001).

Tablo 11'de ise her bir yaş grubu içerisinde memenin dense olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı incelenmiştir.

Tablo 11. Her bir yaş grubu içerisinde memenin dense olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde radyoloji asistanı tarafından yapılan sınıflamaların klinik performansı

Asistan	**Kıdemli radyolog**			**κ katsayısı †**	**p-değeri**
	Nondense	Dense	Toplam		
<40 yıl				0,740±0,122	**<0,001**
Nondense	8 (%18,6)	1 (%2,3)	9 (%20,9)		
Dense	3 (%7,0)	31 (%72,1)	34 (%79,1)		
Toplam	11 (%25,6)	32 (%74,4)	43 (%100,0)		
40-49 yıl				0,904±0,030	**<0,001**
Nondense	85 (%39,7)	6 (%2,8)	91 (%42,5)		
Dense	4 (%1,9)	119 (%55,6)	123 (%57,5)		
Toplam	89 (%41,6)	125 (%58,4)	214 (%100,0)		
50-59 yıl				0,948±0,030	**<0,001**
Nondense	82 (%64,5)	1 (%0,8)	83 (%65,3)		
Dense	2 (%1,6)	42 (%33,1)	44 (%34,7)		
Toplam	84 (%66,1)	43 (%33,9)	127 (%100,0)		
≥60 yıl				0,725±0,130	**<0,001**
Nondense	78 (%88,7)	3 (%3,4)	81 (%92,1)		
Dense	1 (%1,1)	6 (%6,8)	7 (%7,9)		
Toplam	79 (%89,8)	9 (%10,2)	88 (%100,0)		

† Uyum sonuçlarına ait istatistikler; κ katsayısı ve ± std.hata biçiminde verilmiştir.

40 yaş altı bireylerde memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile radyoloji asistanının yapmış olduğu sınıflamanın istatistiksel açıdan orta derecede uyumlu olduğu görüldü (κ=0,740; Std.Hata=0,122 ve p<0,001). 40-49 yaşları arasındaki bireylerde de memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile radyoloji asistanının yapmış

olduğu sınıflamanın istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,904; Std.Hata=0,030 ve p<0,001). 50-59 yaşları arasındaki bireylerde memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile radyoloji asistanının yapmış olduğu sınıflamanın istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,948; Std.Hata=0,030 ve p<0,001). Son olarak 60 yaş ve üzeri (olguların ileri yaş gruplarındaki sayıları yetersiz olduğundan 70 yaş ve üzeri grup ile 60-69 yaşları arasındaki grup birleştirilerek 60 yaş ve üzeri şeklinde değerlendirme yapılmıştır) bireylerde de memenin dense olup olmamasına göre kıdemli radyolog tarafından yapılan sınıflaması ile radyoloji asistanının yapmış olduğu sınıflamanın istatistiksel açıdan orta derecede uyumlu olduğu görüldü (κ=0,725; Std.Hata=0,130 ve p<0,001).

Tablo 12'de her bir yaş grubu içerisinde memenin yoğun olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle karşılaştırılması yapılmıştır.

Tablo 12. Her bir yaş grubu içerisinde memenin yoğun olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle karşılaştırılması

Asistan	**Yapay zeka**			**p-değeri †**
	Uyumlu	Uyumlu değil	Toplam	
<40 yıl				0,500
Uyumlu	39 (%90,7)	-	39 (%90,7)	
Uyumlu değil	2 (%4,7)	2 (%4,7)	4 (%9,3)	
Toplam	41 (%95,3)	2 (%4,7)	43 (%100,0)	
40-49 yıl				0,453
Uyumlu	202 (%94,4)	2 (%0,9)	204 (%95,3)	
Uyumlu değil	5 (%2,3)	5 (%2,3)	10 (%4,7)	
Toplam	207 (%96,7)	7 (%3,3)	214 (%100,0)	
50-59 yıl				>0,999
Uyumlu	123 (%96,8)	1 (%0,8)	124 (%97,6)	
Uyumlu değil	2 (%1,6)	1 (%0,8)	3 (%2,4)	
Toplam	125 (%98,4)	2 (%1,6)	127 (%100,0)	
≥60 yıl				0,500
Uyumlu	84 (%95,4)	-	84 (%95,4)	
Uyumlu değil	2 (%2,3)	2 (%2,3)	4 (%4,6)	
Toplam	86 (%97,7)	2 (%2,3)	88 (%100,0)	
Genel				0,057
Uyumlu	448 (%94,9)	3 (%0,7)	451 (%95,6)	
Uyumlu değil	11 (%2,3)	10 (%2,1)	21 (%4,4)	
Toplam	459 (%97,2)	13 (%2,8)	472 (%100,0)	

† McNemar testi.

Mevcut değerlendirmede kıdemli radyoloğun yapmış olduğu sınıflamada nondense (yağlı veya dağılmış) ve dense (heterojen veya yoğun) olarak sınıflanan bireylerin gerek YZ gerekse radyoloji asistanı tarafından bire bir aynı kategoride sınıflanması dikkate alınmıştır. Tüm olgular içerisinde yapay zekanın kıdemli radyolog ile olan tutarlılığı %97,2 iken radyoloji asistanının kıdemli radyolog ile olan tutarlılığı %95,6 olup YZ ile radyoloji asistanı arasında istatistiksel olarak anlamlı fark yoktu (p=0,057). Benzer şekilde <40 yaş, 40-49 yaş, 50-59 yaş ve ≥60 yaş grupları içerisinde memenin dense olup olmaması açısından kıdemli radyolog ile tutarlı sonuç verme durumu YZ'nin lehine görece yüksek olmakla beraber YZ ile radyoloji asistanı arasında istatistiksel olarak anlamlı herhangi bir fark görülmemiştir (p=0,500; p=0,453; p>0,999 ve p=0,500).

Son olarak Tablo 13'te memenin yoğun olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyum durumları yönünden yaş grupları arasında yapılan karşılaştırmalar verilmiştir.

Tablo 13. Memenin yoğun olup olmamasına göre kıdemli radyoloğun yapmış olduğu sınıflama altın standart olarak kabul edildiğinde yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyum durumları yönünden yaş grupları arasında yapılan karşılaştırmalar

	<40 yıl (n=43)	40-49 yıl (n=214)	50-59 yıl (n=127)	≥60 yıl (n=88)	p-değeri †
Yapay zeka					0,649
Uyumlu	41 (%95,3)	207 (%96,7)	125 (%98,4)	86 (%97,7)	
Uyumlu değil	2 (%4,7)	7 (%3,3)	2 (%1,6)	2 (%2,3)	
Asistan					0,271
Uyumlu	39 (%90,7)	204 (%95,3)	124 (%97,6)	84 (%95,4)	
Uyumlu değil	4 (%9,3)	10 (%4,7)	3 (%2,4)	4 (%4,6)	

† Fisher Freeman Halton testi.

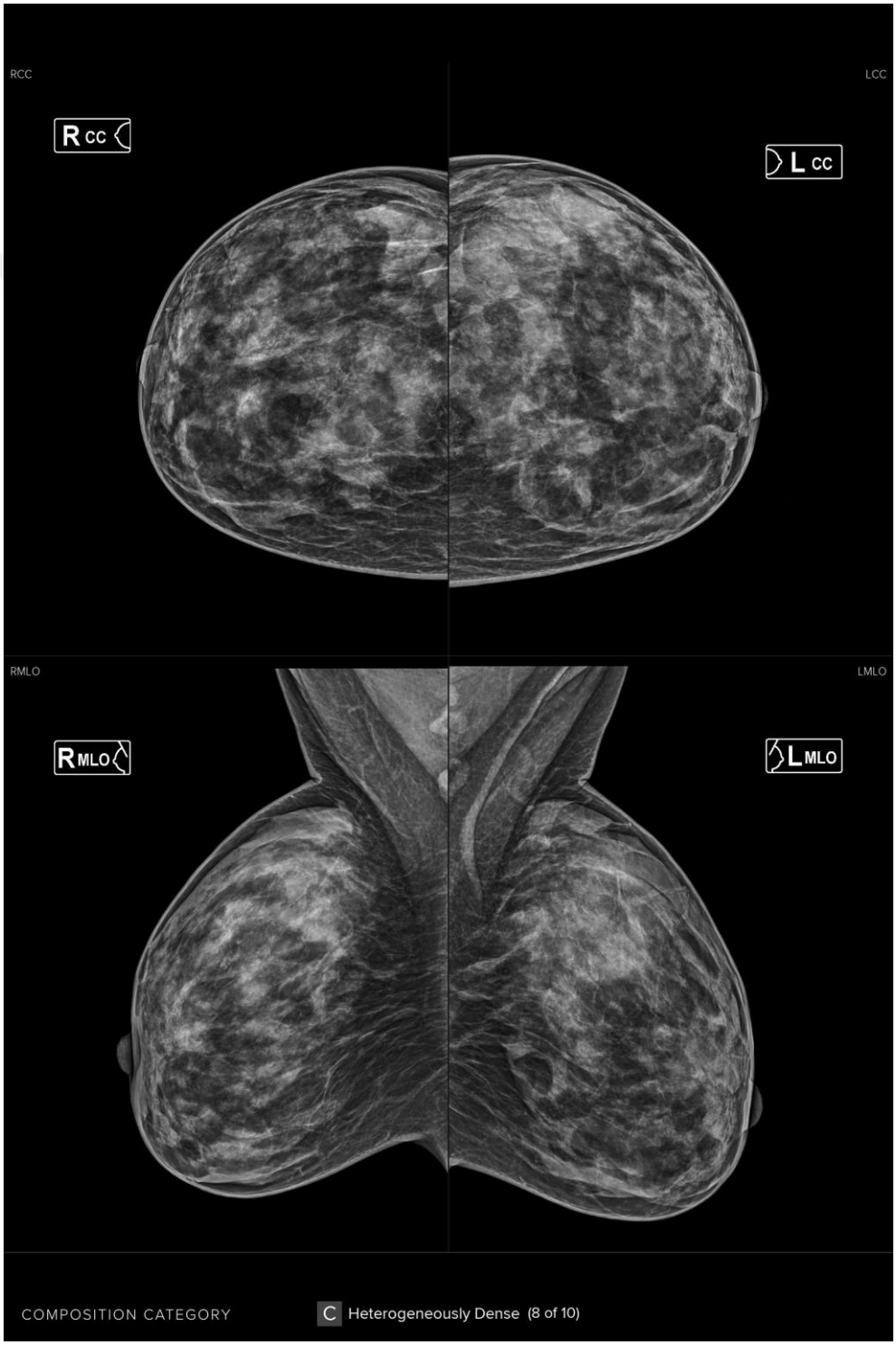
Memenin dense olup olmaması açısından yapay zekanın kıdemli radyolog ile olan uyumu yaş grupları arasında istatistiksel olarak anlamlı herhangi bir değişiklik göstermemiştir (p=0,649). Benzer şekilde memenin dense olup olmaması açısından radyoloji asistanının kıdemli radyolog ile olan uyumu da yaş grupları arasında istatistiksel olarak anlamlı herhangi bir değişiklik göstermemiştir (p=0,271).

5. OLGU ÖRNEKLERİ

OLGU 1: 43 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: C / Yapay Zeka: C / **Radyoloji Asistanı: D**

Şekil 6: Olgu 1



OLGU 2: 50 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: B / Yapay Zeka: B / **Radyoloji Asistanı: C**

Şekil 7: Olgu 2



OLGU 3: 42 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: A / Yapay Zeka: A / **Radyoloji Asistanı: B**

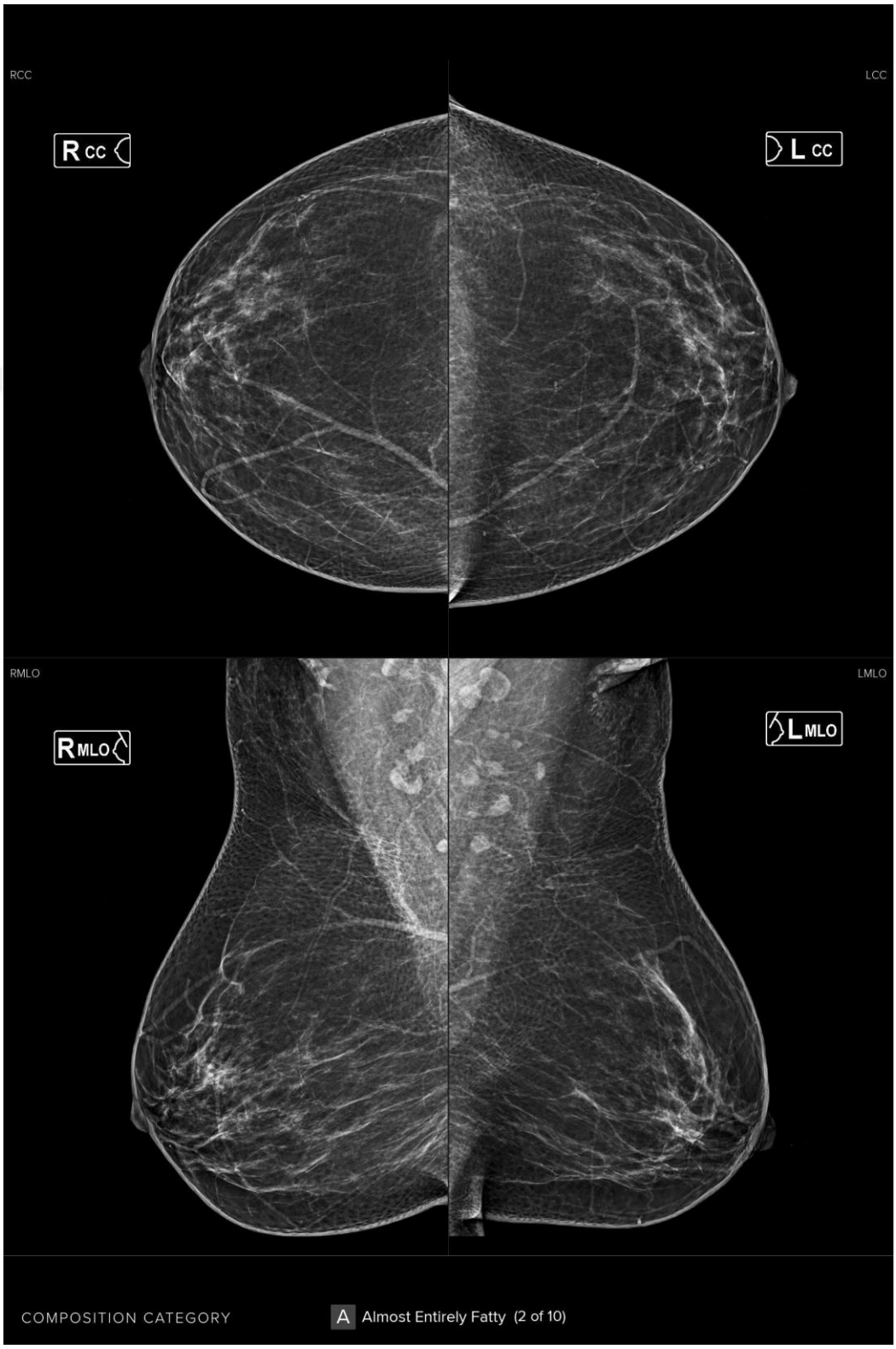
Şekil 8: Olgu 3



OLGU 4: 54 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: A / Yapay Zeka: A / **Radyoloji Asistanı: B**

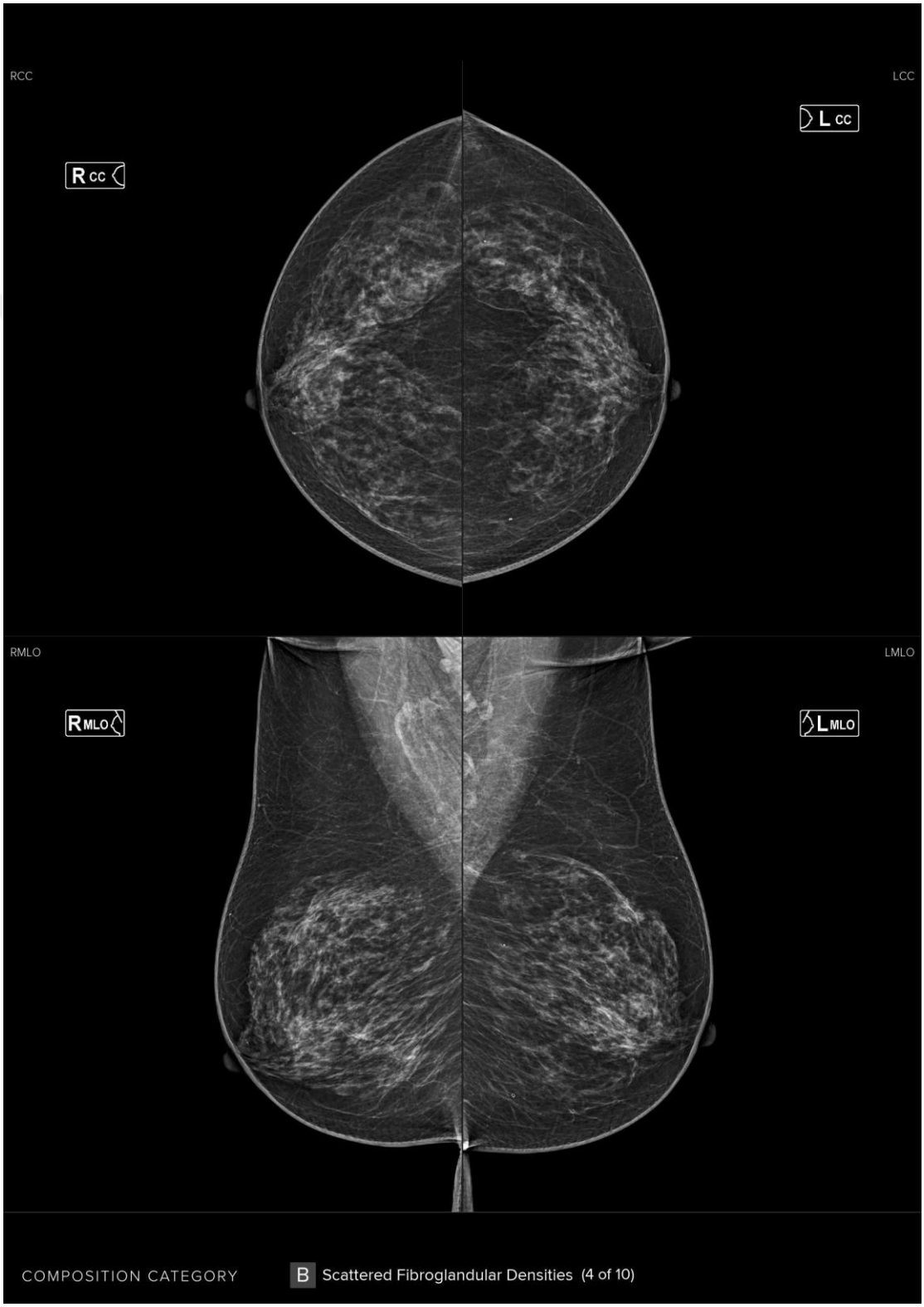
Şekil 9: Olgu 4



OLGU 5: 66 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: C / Yapay Zeka: B / Radyoloji Asistanı: B

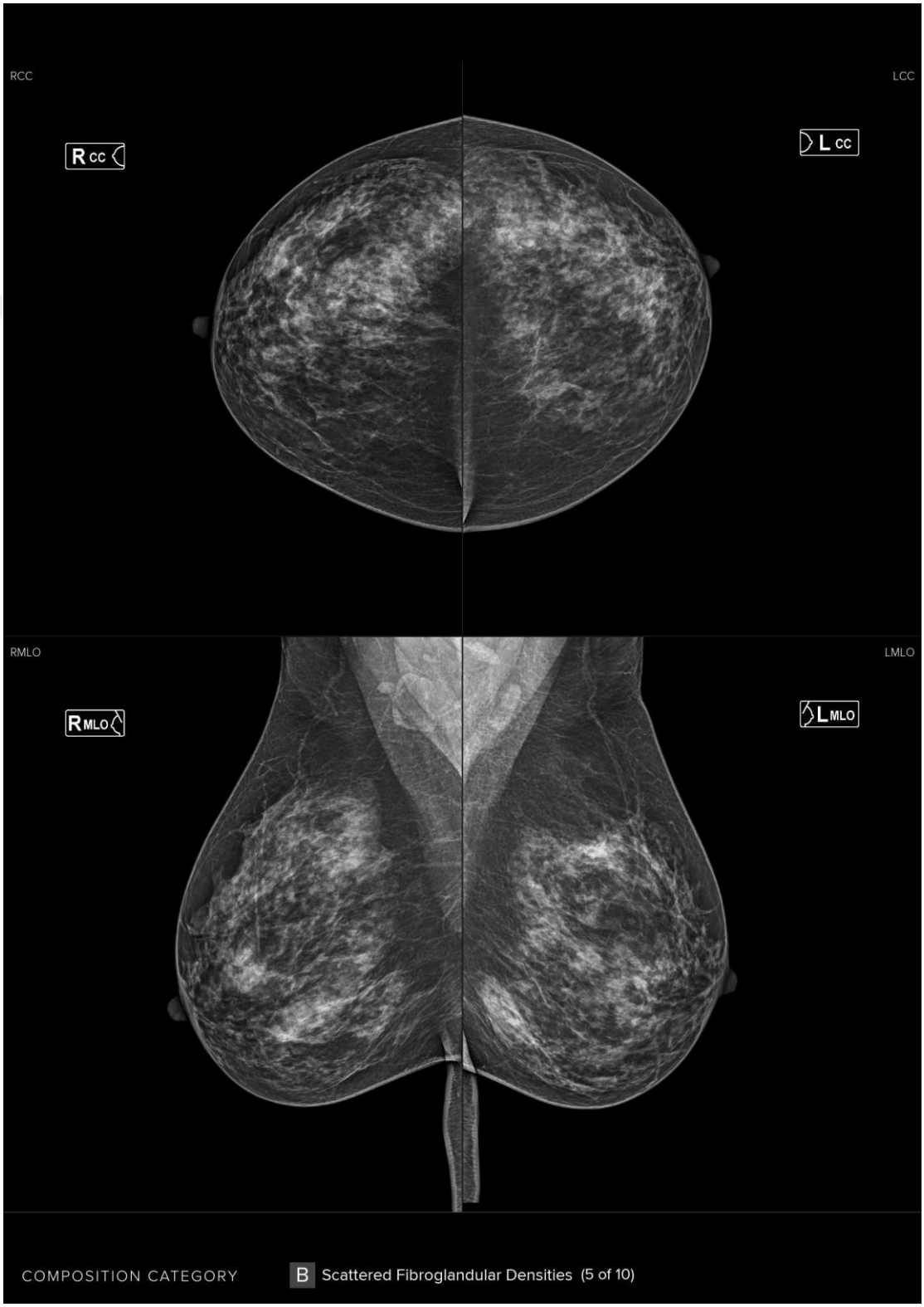
Şekil 10: Olgu 5



OLGU 6: 66 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: C / Yapay Zeka: B / Radyoloji Asistanı: B

Şekil 11: Olgu 6



OLGU 7: 60 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: B / **Yapay Zeka: A** / Radyoloji Asistanı: B

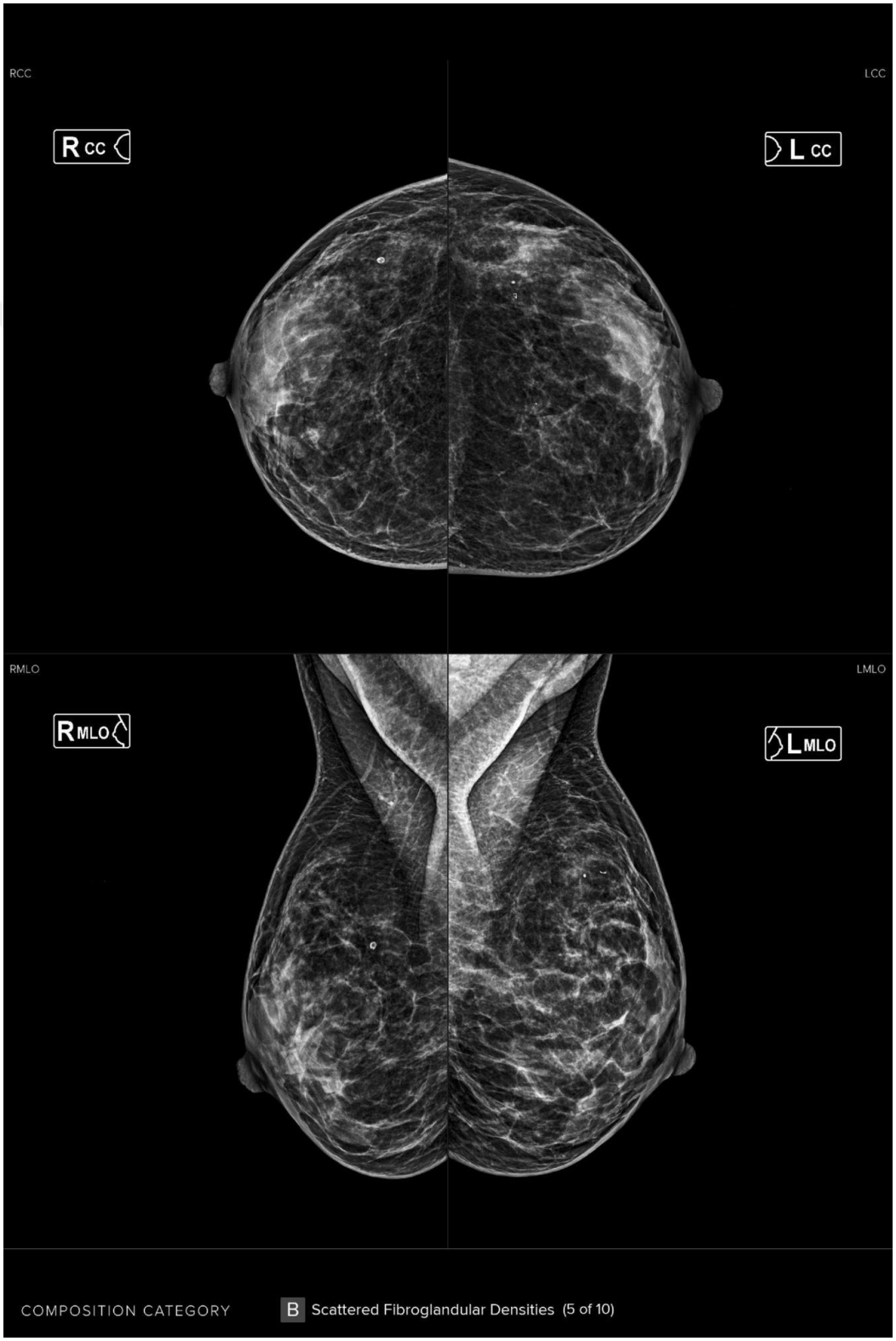
Şekil 12: Olgu 7



OLGU 8: 41 yaşında kadın olguda meme dansite sınıflaması

Kıdemli Radyolog: C / **Yapay Zeka: B** / Radyoloji Asistanı: C

Şekil 13: Olgu 8



6. TARTIŞMA

Meme kanseri, dünya genelinde kadınları etkileyen en önemli dört kanser türünden biridir ve meme yoğunluğu meme kanseri riskini artıran bağımsız bir faktördür (40). Mamografi ise meme kanserinin erken teşhis ve tedavisinde kritik rol oynamaktadır. Mamografi görüntülerinin doğru ve tutarlı değerlendirilmesi önemlidir. Bu nedenle meme dansitesinin değerlendirilmesinde oluşan tutarsızlıklar hastada kaygıya neden olmakta ve gereksiz prosedürlere yol açmaktadır (41). Meme dansitesinin doğru belirlenmesi önem arz etmektedir.

Çalışmamızda meme dansitesinin belirlenmesinde meme konusunda kıdemli radyolog görüşü gold standart kabul edilmiş olup, yapay zeka programı ve radyoloji asistanının performansı değerlendirilmiştir. Kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının istatistiksel açıdan oldukça uyumlu olduğu görüldü (κ=0,905; Std.Hata=0,017 ve $p<0,001$). Hem kıdemli radyoloğun hem de yapay zekanın geniş vaka serileri ile eğitim görmesi performanslarının yüksek olmasına neden olmaktadır. Ayrıca Magni ve arkadaşları (42) geliştirdikleri yapay zeka programı ile meme dansitesinin doğru belirlenmesinde %90 başarıya ulaşmıştır. Bu da güncel pratikte yapay zeka programlarının meme dansitesinin belirlenmesinde kullanılabileceğini göstermektedir.

Alt yaş gruplarına göre kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile yapay zekanın (YZ) yapmış olduğu meme yoğunluk sınıflamasının görece en iyi sonuç verdiği grup 50-59 yaşları arasındaki bireyler olup görece en zayıf uyumun olduğu gruplar ise sırasıyla; ≥70 yaş ile <40 yaş grupları idi. Ayrıca mevcut bulgulara göre kıdemli radyolog tarafından yapılan meme yoğunluğu sınıflaması ile radyoloji asistanının yapmış olduğu meme yoğunluk sınıflamasının görece en iyi sonuç verdiği grup 50-59 yaşları arasındaki bireyler olup görece en zayıf uyumun olduğu grup ise <40 yaş altı bireylerin yer aldığı grup idi. Benzer bulgular kıdemli radyolog ile yapay zeka programının meme yoğunluk sınıflamasında da mevcuttur. Meme dansitesi dens ve non-dens olarak sınıflandırıldığında ise memenin dense olup olmaması açısından kıdemli radyolog tarafından yapılan

sınıflaması ile yapay zekanın (YZ) yapmış olduğu sınıflamasının görece en iyi sonuç verdiği gruplar sırasıyla; 50-59 ve 40-49 yaşları arasındaki bireyler olup görece en zayıf uyumun olduğu gruplar ise sırasıyla; ≥70 yaş ile <40 yaş grupları idi. Ayrıca mevcut bulgulara göre memenin dense olup olmaması açısından kıdemli radyolog tarafından yapılan sınıflaması ile radyoloji asistanının yapmış olduğu sınıflamasının görece en iyi sonuç verdiği gruplar sırasıyla; 50-59 ve 40-49 yaşları arasındaki bireyler olup görece en zayıf uyumun olduğu gruplar ise sırasıyla; ≥70 yaş ile <40 yaş grupları idi. Benzer bulgular kıdemli radyolog ile yapay zeka programının meme yoğunluk sınıflamasında da mevcuttur. Checka ve arkadaşları (43) yaptıkları araştırmada yaş ile mamografik yoğunluk arasında genel olarak ters doğrusal bir ilişki olduğunu göstermiştir. Orta yaşlardaki bireylerde ise meme yoğunluğu ara kategorilerde değişkenlik göstermektedir ve çalışmamızda bu yaş gruplarında yapay zekanın diğer yaş gruplarına göre performansı yüksek bulunmuştur. Bu da yapay zeka programlarının meme dansitesinin belirlenmesinde kritik bir rol üstlendiğinin göstergesidir.

Her bir yaş alt grubuna bakıldığında yapay zeka ile radyoloji asistanı klinik açıdan meme dansitesinin belirlenmesinde başarılıdır. Ancak her bir yaş grubu içerisinde meme yoğunluğu açısından yapay zeka ve radyoloji asistanının kıdemli radyolog ile olan uyumlarının birbirleriyle olan karşılaştırılmasında tüm olgular içerisinde yapay zekanın kıdemli radyolog ile olan tutarlılığı %93,6 iken radyoloji asistanının kıdemli radyolog ile olan tutarlılığı %85,2 olup yapay zekaya karşı radyoloji asistanının istatistiksel anlamlı olarak kıdemli radyolog ile daha az tutarlılık gösterdiği saptanmıştır ($p<0,001$). Benzer durum 40-49 yaş grubunda da geçerli olup bu yaş grubu içerisinde yapay zekanın kıdemli radyolog ile olan tutarlılığı %93,0 iken radyoloji asistanının kıdemli radyolog ile olan tutarlılığı %80,8 düzeyinde kalmıştır ($p<0,001$). Son olarak gerek YZ'nın gerekse radyoloji asistanının kıdemli radyolog ile olan tutarlılıklarının peak yaptığı 50-59 yaş ve ≥60 yaş grupları içerisinde YZ ile radyoloji asistanı arasında kıdemli radyolog ile olan tutarlılıklarının istatistiksel olarak benzer olduğu görüldü ($p=0,125$ ve $p=0,219$). Bu performans farklılığı radyoloji asistanının meme dansitesi konusunda kısa dönem deneyime sahipken

yapay zekanın daha geniş ve büyük veri setleri ile eğitilmiş olduğundan kaynaklanabilir. Radyoloji asistanının düşük yaş gruplarında performansının daha düşük olması eğitim sürecindeki hasta popülasyonunun ve bu popülasyonun yaş dağılımının inhomojen olmasından kaynaklanabilir.

Bu çalışmada limitasyonlarımız vaka sayımızın az ve inhomojen olmasıydı. Bunun nedeni ise çalışmanın tek merkezde gerçekleştirilmiş olmasıydı. Multimerkez ve daha fazla vaka ile daha doğru sonuçlar elde edilebilir.

7. SONUÇ

Mamografide meme dansitesinin belirlenmesinde meme konusunda tecrübe önemli bir faktördür. Radyoloji asistanı eğitim süreci boyunca genel olarak tek merkezde belirli sayıda mamografi görüntüsünü değerlendirmektedir ancak yapay zeka programının eğitiminde kullanılan veri çeşitliliği daha fazladır. Buna ragmen genel olarak bakıldığında yapay zeka programı ile radyoloji asistanı benzer performans göstermekle birlikte, yaş alt grupları ile birlikte değerlendirildiğinde yapay zeka bazı alt gruplarda radyoloji asistanından yüksek performans göstermektedir. Güncel pratikte meme dansitesinin belirlenmesinde yapay zeka programı başarılı ve meme konusunda kıdemli radyolog ile performansları koreledir.

KAYNAKÇA

1- Lee, C., Chen, L. & Elmore, J. (2017). Risk-based Breast Cancer Screening: Implications of Breast Density.. The Medical clinics of North America.

2- Weigel, S., Heindel, W., Heidrich, J., Hense, H. & Heidinger, O. (2017). Digital mammography screening: sensitivity of the programme dependent on breast density. European Radiology.

3- Mohamed, A., Berg, W., Peng, H., Luo, Y., Jankowitz, R. & Wu, S. (2018). A deep learning method for classifying mammographic breast density categories. Medical Physics (Lancaster).

4- McCormack, V. & Silva, I. (2006). Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis. Cancer Epidemiology, Biomarkers and Prevention.

5- Lee, C., Chen, L. & Elmore, J. (2017). Risk-based Breast Cancer Screening: Implications of Breast Density.. The Medical clinics of North America.

6- Dembrower, K., Liu, Y., Azizpour, H., Eklund, M., Smith, K., Lindholm, P. & Strand, F. (2019). Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction.. Radiology.

7- McDonald, E., Hammersley, J., Chou, S., Rahbar, H., Scheel, J., Lee, C., Liu, C., Lehman, C. & Partridge, S. (2016). Performance of DWI as a Rapid Unenhanced Technique for Detecting Mammographically Occult Breast Cancer in Elevated-Risk Women With Dense Breasts.. AJR. American journal of roentgenology.

8- Krammer, J., Stepniewski, K., Kaiser, C., Brade, J., Riffel, P., Schönberg, S. & Wasser, K. (2017). Value of Additional Digital Breast Tomosynthesis for Preoperative Staging of Breast Cancer in Dense Breasts.. Anticancer Research.

9- Rhodes, D., Breitkopf, C., Ziegenfuss, J., Jenkins, S. & Vachon, C. (2015). Awareness of breast density and its impact on breast cancer detection and risk.. Journal of Clinical Oncology.

10- Siontis, K., Noseworthy, P., Attia, Z. & Friedman, P. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology.

11- Bera, K., Braman, N., Gupta, A., Velcheti, V. & Madabhushi, A. (2021). Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nature Reviews Clinical Oncology.

12- Park, H., Bae, H., Hong, G., Kim, M., Yun, J., Park, S., Chung, W. & Kim, N. (2021). Realistic High-Resolution Body Computed Tomography Image Synthesis by Using Progressive Growing Generative Adversarial Network: Visual Turing Test. JMIR Medical Informatics.

13- Dzobo, K., Adotey, S., Thomford, N. & Dzobo, W. (2020). Integrating Artificial and Human Intelligence: A Partnership for Responsible Innovation in Biomedical Engineering and Medicine.. Omics.

14- Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V., Tanwar, S., Sharma, G., Bokoro, P. & Sharma, R. (2022). Explainable AI for Healthcare 5.0: Opportunities and Challenges. IEEE Access.

15- Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S. et al. (2020). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence.

16- Rawat, W. & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation.

17- Busnatu, Ș., Niculescu, A., Bolocan, A., Petrescu, G., Păduraru, D., Năstasă, I., Lupușoru, M., Geantă, M., Andronic, O., Grumezescu, A. & Martins, H. (2022). Clinical Applications of Artificial Intelligence—An Updated Overview. Journal of Clinical Medicine.

18- Kung, T., Cheatham, M., Medenilla, A., Sillos, C., Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J. & Tseng, V. (2022). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. medRxiv.

19- Liao, Z., Hu, S., Xie, Y. & Xia, Y. (2021). Modeling annotator preference and stochastic annotation error for medical image segmentation.. Medical Image Analysis.

20- Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S. & Vercauteren, T. (2018). Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing.

21- Lee, C., Ho, K. & Lee, W. (2011). A Novel Key Management Solution for Reinforcing Compliance With HIPAA Privacy/Security Regulations. IEEE Transactions on Information Technology in Biomedicine.

22- Britton, A., Ben-Shlomo, Y., Benzeval, M., Kuh, D. & Bell, S. (2015). Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. BMC Medicine.

23- Lambin, P., Leijenaar, R., Deist, T., Peerlings, J., Jong, E., Timmeren, J., Sanduleanu, S., Larue, R., Even, A., Jochems, A., Wijk, Y., Woodruff, H., et al (2017). Radiomics: the bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology.

24- Yang, J., Nie, J., Ma, X., Wei, Y., Peng, Y. & Wei, X. (2019). Targeting PI3K in cancer: mechanisms and advances in clinical trials. Molecular Cancer.

25- Xiao, H., Rasul, K. & Vollgraf, R. (2017). Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. arXiv.org.

26- Rudin, C. (2018). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence.

27- Clift, A., Coupland, C., Keogh, R., Diaz-Ordaz, K., Williamson, E., et al. (2020). Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. British medical journal.

28- Zhang, M., Cui, Z., Neumann, M. & Chen, Y. (2018). An End-to-End Deep Learning Architecture for Graph Classification. AAAI Conference on Artificial Intelligence.

29- Kohli, M., Summers, R. & Geis, J. (2017). Medical Image Data and Datasets in the Era of Machine Learning—Whitepaper from the 2016 C-MIMI Meeting Dataset Session. Journal of digital imaging.

30- Fox, B., Hendry, J. & Selikoff, D. (1979). Radiobiology for the Radiologist. British Journal of Cancer.

31- Lashgari, E., Liang, D. & Maoz, U. (2020). Data augmentation for deep-learning-based electroencephalography. Journal of Neuroscience Methods.

32- Woźniak, M., Siłka, J. & Wieczorek, M. (2021). Deep neural network correlation learning mechanism for CT brain tumor detection. Neural computing & applications (Print).

33- Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience.

34- Moore KL, Dalley AF. Clinically Oriented Anatomy. In: Thorax: Breast. Moore KL. 4th edition. Lippincott Williams&Wilkins, Philadelphia, 1999; 72–79.

35- April EW. Clinically Anatomy. In: Introduction to Clinically Anatomy: Breast. 3rd edition. Williams&Wilkins A Waverly Company, 1996; 12–14.

36- Kopans DB. Breast Imaging. 3rd Edition. Philedelphia Lipincott Williams&Wilkins 2007

37- Rosen PP. Breast pathology. 4th Edition. Philadelphia, Lipincott Williams Wilkins, 2002

38- Warrier, S., Tapia, G., Goltsman, D. & Beith, J. (2015). An Update in Breast Cancer Screening and Management. Women's Health.

39- Tice, J., Miglioretti, D., Li, C., Vachon, C., Gard, C. & Kerlikowske, K. (2015). Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer.. Journal of Clinical Oncology.

40- Badawy, E. R., ElNaggar, R., Soliman, S. a. M., & Elmesidy, D. S. (2023c). Performance of AI-aided mammography in breast cancer diagnosis: Does breast density matter? Egyptian Journal of Radiology and Nuclear Medicine, 54(1).

41- Lehman, C., Yala, A., Schuster, T., Dontchos, B. N., Bahl, M., Swanson, K., & Barzilay, R. (2019b). Mammographic breast density Assessment using Deep Learning: Clinical implementation. Radiology, 290(1), 52–58.

42- Magni, V., Interlenghi, M., Cozzi, A., Alì, M., Salvatore, C., Azzena, A. A., . . . Sardanelli, F. (2022b). Development and validation of an AI-driven mammographic breast density classification tool based on radiologist consensus. Radiology, 4(2).

43- Checka, C. M., Chun, J., Schnabel, F., Lee, J., & Toth, H. K. (2012). The relationship of mammographic density and age: Implications for breast cancer screening. American Journal of Roentgenology, 198(3), W292–W295.