

TRD YAPAY ZEKA UYGULAMALARI REHBERİ

1. Giriş
2. İlgili Tanımlar
3. Genel Prensipler
4. Uygulama Geliştirme Süreçlerinde Uyulması Beklenen İlkeler
 - 4.1. Klinik Problem ve Bağlam Belirleme
 - 4.2. Veri Toplama veri Yönetimi
 - 4.2.1. Veri Toplama Stratejileri
 - 4.2.2. Veri Kürasyonu
 - 4.2.3. Veri Gizliliğinin Sağlanması
 - 4.2.4. Görüntü Etiketleme ve Anotasyon
 - 4.3. Modelleme ve Geçerlilik Testi
 - 4.3.1. Performans Değerlendirmesi
 - 4.3.1.1. Ayırt Etme Performansı Değerlendirmesi
 - 4.3.1.2. Kalibrasyon Performansı Değerlendirmesi
 - 4.3.1.3. Klinik Fayda Analizi
 - 4.3.2. Açıklanabilirlik ve Yorumlanabilirlik
5. Klinik Uygulamalarda Uyulması Beklenen İlkeler
 - 5.1. Uygulama Seçerken
 - 5.2. Uygulama Kullanımında
 - 5.3. Uygulamalar Hakkında Hastaların Bilgilendirilmesi
 - 5.4. Uygulamaların Güvenirlik Düzeyleri
 - 5.5. Sürekli İyileştirme ve Kalite Denetimi
 - 5.6. Geri Ödeme
 - 5.7. Sorumluluklar
6. Yararlanılan Kaynaklar
7. Hazırlayanlar

TRD YAPAY ZEKA UYGULAMALARI REHBERİ

1. Giriş

Yapay zeka (YZ) kapsamında geliştirilen ve başarılı sonuçları olan uygulamaların giderek artması ve öğrenen yordamların (algoritmaların) kullanıma girmesi ile yaşamın birçok alanında olduğu gibi sağlıkta ve radyolojide de köklü değişikliklerin olacağı bir döneme girmiş bulunmaktayız. Kaçınılmaz olan değişimin emekleme döneminde olduğumuz kabul edilse bile değişimin beklenenden hızlı olacağı görülmektedir.

Derneğimiz, bu değişimin mesleğimize yönelik bir tehdit potansiyeli taşıdığını öngörmekle birlikte, YZ'nın oyun değiştirici bir teknoloji ürünü olarak, daha çok yeni fırsatlar sağlama gücüne odaklanmayı ve uygulamaların hasta yararına olacak şekilde ve etik kurallar çerçevesinde yürütülmesini sağlamayı hedeflemektedir. Bu rehber, meslektaşlarımıza YZ alanında karşılaşılabilecekleri sorunlarda yol göstermek, gerek yordam geliştirme aşamasında gerekse klinik uygulamalarda YZ'nın radyolojik değerlendirmelerin niteliğini artıracak şekilde kullanımına katkı vermek amacıyla geliştirilmiştir.

TRD Görüntüleme Bilişimi Çalışma Grubunun ortak bir ürünü olan rehberin sürekli geliştirilmesi ve yaygın kullanımı başlıca hedefimizdir.

2. İlgili Tanımlar

Bu bölümde radyoloji alanında YZ ve ilgili kavramların daha iyi anlaşılmasına yardımcı olabilmek amacı ile temel kaynaklarda sık kullanılan terimlerin tanımlamalarına yer verilmiştir. Terimler, İngilizce karşılıkları ile birlikte sunulmuştur. Daha ayrıntılı tanımlar için ayrıca <https://oasys.tk/trd-gbs/> adresindeki TRD Görüntüleme Bilişimi Sözlüğünden yararlanılabilir.

Açıklanabilir Yapay Zeka (Explainable AI- XAI): YZ modellerinin karar mekanizmalarının şeffaf hale getirilmesini ve kullanıcıların modelin nasıl çalıştığını anlamasını sağlayan yaklaşımlar bütünüdür.

Aktarmalı Öğrenme (Transfer Learning): Sınıflandırma, regresyon, kestirim için problemin çözümünü öğrendikten sonra yapay öğrenme algoritmasının bağlantı katsayılarının ve aktivasyon değerlerinin tümünü ya da bir kısmını başka problemlerin çözümü için kullanmaya verilen addır.

Aşırı Uyum (Overfitting): Bir YZ modelinin eğitim verilerine çok fazla uyum sağlaması nedeniyle yeni veriler üzerinde düşük performans göstermesi durumu.

Bölütleme (Segmentation): İstenilen bir bölge ya da organın çeşitli algoritmalarla ana yapıdan ayrıştırılmasına denir.

Büyük Dil Modelleri (Large Language Models): Çok geniş çaplı metin verileri ile eğitilmiş, insan benzeri metin oluşturabilen, anlam çıkarabilen ve çok çeşitli dil görevlerini yerine getiren yapay zeka modelleridir. Büyük dil modelleri, dil yapı ve bağlamını öğrenerek metin tamamlama

Çekişmeli Üretici Ağlar (Generative Adversarial Networks): Üretilmiş verilerin gerçek görüntüyle karşılaştırılması ve alınan geri bildirimlerle gerçek görüntüye benzer üretimler yapabilen ağlardır.

Derin Öğrenme (Deep Learning): Bir veya daha fazla gizli katman içeren yapay sinir ağları ve benzer algoritmalarla işleyen makina öğrenmesi çalışma alanıdır.

Doğal Dil İşleme (Natural Language Processing): Genellikle düz metin halindeki yazıların belirli yordamlara göre anlamlandırılmasını sağlayan işlemler bütününe verilen addır.

Doku analizi (Texture analysis): Görüntüde, seçilen bir bölgedeki uzamsal heterojenitenin matematiksel olarak değerlendirilmesidir. Görüntüdeki piksel ve voksellerin matematiksel diziliminden anlam çıkarımı yapılır.

Eğitim Seti (Training Set): Makine öğrenimi ve derin öğrenme modellerinin öğrenme sürecinde kullanılan veri kümesidir. Model, bu veriler üzerinde çalışarak örüntüleri öğrenir, parametrelerini ayarlar ve belirli görevleri yerine getirme yetisini geliştirir. Eğitim seti, genellikle büyük ve çeşitli olmalıdır, böylece model farklı durumları öğrenebilir ve genelleme yapabilir. Ancak, yalnızca eğitim verisine aşırı uyum sağlarsa (overfitting), model test verisi üzerinde düşük performans gösterebilir.

Etiketleme (Labeling): Bilişim uygulamalarında nesnelerin işaretleme araçlarıyla isimlendirilmesi ya da numaralandırılması işlemine verilen isimdir.

Evrişimsel Sinir Ağları (Convolutional Neural Networks) : Katmanlarının en az bir tanesinde matris çarpımı yerine konvolüsyon işleminin kullanıldığı çok katmanlı yapay sinir ağlarının bir türüdür.

Geçerlilik Test Seti (Validation Set): Makine öğrenimi modellerinin eğitim sürecinde, modelin performansını izlemek ve hiperparametre ayarlamalarını yapmak için kullanılan veri kümesidir. Eğitim verisinden ayrı tutulur ve modelin aşırı uyum (overfitting) yapmadan genelleme yeteneğini değerlendirmek için kullanılır. Validasyon seti, modelin optimize edilmesine yardımcı olur ancak nihai değerlendirme için kullanılmaz.

Gözetimli Öğrenme (Supervised Learning): Makine öğrenmesinde görüntü özelliklerinin verilmesini gerektiren öğrenme yöntemidir.

Gözetimsiz Öğrenme (Unsupervised Learning): Makine öğrenmesinde görüntü değerlendirmesi için kullanılan özelliklerin kullanılan algoritmalarca çıkarıldığı öğrenme yöntemidir.

İşbirlikli Öğrenme (Federated Learning) : Makine öğrenmesinde verilerin merkezi bir yerde toplanarak kullanılması yerine farklı uç noktalardaki verileri kullanarak algoritmayı eğitme yöntemidir.

Makine Öğrenmesi (Machine Learning): Yapay zeka uygulamalarının bir alt grubunu oluşturan ve deneyimlerden öğrenme ya da örneklerden bilgi çıkarımı yapmayı amaçlayan etkinlik alanıdır.

Model Eğitimi (Model Training): Yapay zeka modelinin, verilen eğitim verileri üzerinde çalışarak belirli bir problemi çözme yeteneğini geliştirdiği süreçtir.

Niteleyici (Attribute): Görüntülerde yer alan özelliklerin genel adıdır. “Feature” olarak bilinen kavramla eşdeğerdir. Hacim, yüzey görünüm, kenar, dansite, intensite, yerleşim ve çevre ilişkileri birer niteleyicidir.

Radyogenomik (Radiogenomics): Tıbbi görüntü ve genetik bilgilerden elde edilen verileri birlikte kullanarak kişiselleştirilmiş tanı ve tedavi yöntemlerinin geliştirilmesini amaçlayan alandır.

Radyomiks (Radiomics): Veri karakterize etme algoritmaları kullanılarak tıbbi görüntülerden, insan gözünün farkedemeyeceği çok sayıda nicel özellik çıkarmayı ve bu özelliklerin klinik karşılıklarını bulmayı amaçlayan bir çalışma alanıdır. İnsan gözünün çıkaramayacağı görüntülerin elde edilmesiyle ayırıcı tanı ve tanı yapılması hedeflenir.

Sinir Ağı (Neural Networks): Nöron davranışını taklit ederek çözüm üreten bilgisayar programı ya da sistemi. Özellikle sinir bağlantıları üzerinden aksiyon potansiyellerinin toplanması, bekleme ve eşik değerlerinde ateşleme süreçleri taklit edilmektedir.

Sözdizimsel çözümleme (Parsing): Doğal dil işleme çalışmalarında metnin dil bilgisidilbilgisi kurallarına göre ayrıştırılması ve sözcükler arası ilişkilerin belirlenmesi işlemine verilen addır.

Test Seti (Test Set): Bir modelin eğitim ve validasyon aşamaları tamamlandıktan sonra, gerçek dünyada nasıl performans göstereceğini ölçmek için kullanılan bağımsız veri kümesidir. Modelin daha önce görmediği bu veri kümesi, genel başarımını ve genelleme yeteneğini değerlendirmek için kullanılır. Test seti, modelin klinik uygulamadaki doğruluğunu belirlemek açısından kritik bir role sahiptir.

Uzman sistemler (Expert systems): Bir sorunu çözmek için geliştirilmiş, uzman bilgisinden yararlanılarak oluşturulmuş ve sıklıkla öyleyse budur (if/then) şeklinde kuralların kullanıldığı bilgisayar programıdır.

Veri Madenciliği (Data Mining): Büyük verinin içindeki örüntülerden anlamlı sonuçlar çıkarma uğraşına verilen addır.

Yapay Zeka (Artificial Intelligence): İnsana özgü olduğu düşünülen ve genel olarak iletişim kurma ve problem çözme amacıyla kullanılan becerilerin makinalar tarafından gerçekleştirilmesini amaçlayan çalışmalara verilen isimdir. Makina öğrenmesi YZ çalışmalarını önemli bölümünü içermektedir. Derin öğrenme ve onun bir alt alanı olan evrimsel sinir ağıları makina öğrenmesinin kapsamındadır.

Yordam (Algorithm): Bir işlemin gerçekleşmesi için izlenen adımları ve hedefi belirleyen aritmetiksel yol için kullanılan terimdir. Belirli bir görevi yerine getirmek veya bir problemi çözmek için veri analizi, örüntü tanıma ve öğrenme süreçlerini içeren matematiksel ve istatistiksel yöntemlerden oluşan işlem dizisidir. Algoritmalar, büyük veri kümelerinden anlamlı bilgiler çıkarmak, tahminler yapmak veya karar verme süreçlerini otomatikleştirmek için tasarlanır.

Yordamsal Önyargı (Algorithmic Bias): YZ modellerinin, eğitim verisindeki dengesizlikler nedeniyle belirli hasta grupları veya hastalık türleri için yanlış veya yanıltıcı tahminlerde bulunması.

3. Genel Prensipler

Yapay zeka uygulamalarının geliştirilmesi, ulaşılabilirliği, ticareti, kullanımı ve sorumluluk paylaşımı hakkındaki genel prensipler, radyoloji alanı için geliştirilecek uygulamalarda da geçerli ve benzerdir. Uluslararası ilgili kurumların önerileri doğrultusunda aşağıdaki ilkelere dikkat edilmesi uygun olacaktır.

- Sağlık alanında YZ modellerinin kullanımının hasta üzerindeki etkileri, örneğin tanı ve tedavi kararında oynadığı rol, açıklanabilir olmalı ve hasta ile paylaşılmalıdır.
- Yapay zeka modellerinin kullanımına devam edilip edilmeyeceği kararı yetkili kişilerce verilmeli ve gerekli görülen durumlarda modeller devre dışı bırakılabilir.
- Yaptırım gücünü elinde bulunduran kuruluşlarca tüm paydaşların görüşleri alınarak ve ülkelerin kendine özgü güçlü ve zayıf yönleri göz önünde bulundurularak, rehber ilkeler belirlenmelidir.
- Yapay zeka çalışmalarında kullanılacak kişiye özel verilerin korunması birinci öncelik olmalıdır. Hangi verilerin, ne amaçla ve nasıl kullanılacağı konusunda kişiler bilgilendirilmelidir.
- Yapay zeka uygulamalarının geliştirilme aşamaları yakından izlenerek faydalarının yanı sıra olası tehlikeleri konusunda bilinçli olunmalı ve erken dönemde gerekli önlemler alınmalıdır.
- Yapay zeka modellerinin kullanımı öncesi radyoloji uzmanlarının ve diğer kullanıcıların sistemin çalışma koşulları, yarar ve olası zararları, hangi durumlarda tercih edilip ne zaman yetersiz kalabileceği gibi temel konularda eğitim alması sağlanmalıdır.
- Toplumun tüm kesimlerinin sağlık alanındaki YZ modellerinin olası yararlarından eşit ve adil şekilde faydalanabilmesi garanti edilmelidir.
- Akademik çalışmalarda YZ modellerinden destek alındığı durumlarda, elde olunan sonuçlar titizlikle kontrol edilmeli, son ürünün etik sorumluluğunu yazarlar

yüklenmelidir. Telif hakları gözetilmeli, YZ modelinden elde edilen yarar belirtilmeli, veri güvenliği ve gizliliği sağlanmalıdır.

- Geliştirilecek algoritmalar insan ve hasta haklarına aykırı, erişebilirliği kısıtlayan yan tutucu unsurlardan arınmış olmalıdır.
- Özellikle geniş dil modelleri gibi hızla toplum tarafından kullanılmaya başlanan modellerin geliştiricileri sağlık alanındaki etik değerlere sahip olmalıdır.
- Yapay zeka uygulamalarının kullanımı arttıkça ve yaygınlaştıkça karşılaşılabilecek olası etik ve yasal sorunlar konusunda tekrarlayan aralıklı çalışmalar gerçekleştirilmelidir.
- Tıbbi görüntülerin işlenmesi ve analizinde yapay zekânın kullanım amacı uzmanların görüntüleri daha iyi yorumlamasını ve daha doğru tanı koymasını sağlamak, klinik yararlanımı artırmak, yapılan işi zaman ve maliyet açısından daha ekonomik hale getirmektir. Yapay zeka, radyologların rakibi veya alternatifi olmayıp aksine onların kullandığı yardımcı ve faydalı bir araç olacaktır. Çok çeşitli ve telafisi olmayan tıbbi hatalara, sistemsel aksaklıklara ve kaosa yol açmamak için yapay zeka uygulamaları radyolog denetimi olmadan devreye sokulmamalıdır.
- Yapay zeka çalışmaları radyologların işini elinden alacak hedefler gütmemelidir.
- Radyoloji uzmanları ve diğer sağlık çalışanları YZ konusunda bilgi ve yeteneklerini geliştirerek olası yeni iş bölümlerine açık olmalıdır.
- Yapay zeka uygulamalarının geliştirilmesi ve iş akışına dahil edilmesi aşamalarında kararlar öncelikli gereksinimler belirlenerek karar alınmalıdır.
- Farklı klinik senaryolarda YZ uygulamalarının rolü, iş akışındaki yeri ve YZ-insan iş birliğinin nasıl olacağı iş-birliği, iş akışında YZ uygulamalarının yeri belirlenmelidir.
- Yapay zeka uygulamaları kullanılmaya başlanmadan önce kurumun alt yapı olanakları gözden geçirilmeli ve deneme sürümleri talep edilmelidir.
- Hasta-hekim iletişimini geliştirmek, hekimler arasındaki fikir alış-verişlerini zenginleştirmek, görüntüleme birimlerinin çalışma verimini arttırmak, sürekli mesleki gelişim ve uzmanlık öğrencilerinin eğitimi ile katma değer üretebilecek araştırmalara yönlendirmek YZ uygulamaları ile ek zaman kazanan radyoloji uzmanlarının hedefleri arasında olmalıdır.
- Yapay zeka ile ilgili yapılacak çalışmaların hastanın sağlık düzeyini koruyacak, artıracak ve bozulan sağlık durumunu iyileştirecek amaçlar taşıması gereklidir.
- Yapay zeka modellerinin, kullanıma geçilmeden önce, hedef kitle üzerindeki yararlılık ve sınırlılıkları test edilmelidir. İş akışını kesintiye uğratan, zaman alıcı ve rutin uygulamada fark yaratacak fayda sağlamayan uygulamalar konusunda dikkatli olunmalıdır.
- Yapay zeka destekli karar modellerinin kullanımında hastanın kendisi ile ilgili konularda karar verme özgürlüğünün kısıtlanmamasına özen gösterilmelidir.
- Kaliteli veri üretmek ve kullanmak ana hedef olmalıdır. Veri kalitesini değerlendirmede İngilizce karşılıkları ile 10v kuralı örnek alınabilir: Belirsizlik (vagueness), Değişkenlik (variability), Köken (venue), Çeşitlilik (variety), Değer (value), Miktar (volume), Doğruluk (veracity), Kalite (validity), Kelime içeriği (vocabulary), Artış hızı (velocity)
- Yapay zeka uygulaması geliştirme amacıyla kullanılan veri dizilerinin çeşitliliği, görüntü kalitesi, varsa etiketlerin doğruluğu kontrol edilmelidir.

4. Uygulama Geliştirme Süreçlerinde Uyulması Beklenen İlkeler

4.1. Klinik Problem ve Bağlam Belirleme

Yapay zeka destekli bir model geliştirmeden önce ilk yapılması gereken klinik problemi ortaya koymak olmalıdır. Pratikte eksik, karmaşık, çok sayıda hesaplama gerektiren, doğruluk sorunu yaşanan durumlar saptandıktan sonra

var olan YZ uygulamaları değerlendirilir. Buna göre saptanan eksiklikler doğrultusunda yeni bir model geliştirilmesi planlanır. Teknik ve klinik başarı kriterleri (duyarlılık/özellik hedefleri performans ölçütleri, tetkik yorumlama süresinin kısaltılması, hasta sağ kalımının iyileştirilmesi vb.) önceden belirlenmeli, geliştirilecek YZ destekli modelin sağlık sistemine beklenen faydası net olarak ortaya konmalıdır. Modelle doğabilecek negatif sonuçların maliyetine göre yararlarının ağır basıp basmadığı değerlendirilmelidir. Ürüne dönüştürmede sorunlar yaşamamak için hedef pazara ait düzenleyici kurumların rehber ve yönetmelikleri (örn. Amerika Birleşik Devletleri için Software as a Medical Device (SAMd): Clinical Evaluation, Avrupa Birliği için EU Artificial Intelligence Act) titizlikle gözden geçirilmelidir. The American College of Radiology (ACR) Data Science Institute (DSI), Define-AI Directory aracılığıyla klinik uygulamaya uygun ve efektif modellerin geliştirilmesi için tanımlanan hedef kullanım senaryolarından da yararlanılabilir.

4.2. Veri Toplama ve Veri Yönetimi

4.2.1. Veri Toplama Stratejileri

Model geliştirme sürecinde kullanılan klinik bilgiler, laboratuvar bulguları, radyolojik görüntüler, DICOM meta-verileri veya uzman etiketlemeleri veri olarak adlandırılır. Veri kaynakları PACS (Picture Archiving and Communication Systems), elektronik sağlık kayıtları, radyoloji raporları, patoloji raporları, araştırma ve görüntüleme veri tabanları olabilir. Yapay zeka modellerinin performansı veri büyüklüğünün yanında veri çeşitliliği ve kalitesinden de etkilenmektedir.

Model performansının veri miktarıyla doğru orantılı olduğu yaygın bir gözlemdir. Ancak, başarılı ve genellenebilir bir YZ modelinin elde edilebilmesi için ideal örneklem büyüklüğünü net olarak tahmin edebilmek şu anda mümkün değildir. Model geliştirmek için alınması gereken minimum hasta sayısı, büyük ölçüde çözmeye çalıştığımız sorunun karmaşıklığına, hedeflenen model performansına ve kullanılacak modelleme yöntemine bağlıdır. Örneğin, hastalık saptama problemleri tedavi yanıtının tahmin edilmesine dayalı prognostik problemlerden, iki gruplu sınıflama problemleri, üç gruplu sınıflama problemlerinden, radyomiks yöntemleri yapay sinir ağlarından daha basit kabul edildiğinden daha az hasta kümesi ile planlanabilir.

Tamamen istatistiksel bir bakış ile ikili bir sınıflandırma probleminde, radyomiks imzasına katılan her bir özellik için eğitim kümesinde 10 – 15 hasta gereklidir. Yapay sinir ağlarının eğitiminde kullanılacak veri kümesinin belirlenmesinde kılavuz olarak kullanılabilir birkaç temel yaklaşım tanımlanmıştır. Birinci yaklaşım, eğitim veri kümesinin sınıf sayısına göre belirlenmesi olup bu durumda önerilen minimum veri sayısı sınıf sayısının 50 – 1000 katıdır. İkinci yaklaşım, kullanılan toplam öznitelik sayısına dayanır. Burada ise önerilen minimum veri sayısı, öznitelik sayısının 10 – 100 katı arasındadır. Ancak, en yaygın olarak kullanılan kural genellikle ağdaki ağırlıkların sayısını baz alır. Bu kurala göre, önerilen minimum veri sayısı, ağırlıkların sayısının en az 10 katı olmalıdır. Ancak, bu kurallar genellikle gerçek dünya uygulamalarında kullanım kolaylığı sağlamak için basitleştirilmiş yaklaşımlardır ve spesifik durumlarda örneklem büyüklüğünün başarı oranına karşı çizilen öğrenme eğrisi ile optimize edilmesi gerekir.

Toplanan verilerin hedef kitleyi temsil etmesi, gerçek dünyadaki heterojenite ve çeşitliliği olabildiğince yansıtmaları önemlidir. Yapay zeka sistemleri eğitim verilerine dayanır ve bağlamdan yoksundur. Yapay zeka sistemini eğitmek için kullanılan verilerin sistemin kullanılacağı

hasta popülasyonunu temsil etmesi önemlidir. Diğer bir deyişle, yüksek performanslı bir model geliştirmek isteniyorsa çeşitli hastalıkları, anatomik varyasyonları, görüntüleme protokollerini, rekonstrüksiyon yöntemlerini, tarayıcı modellerini ve demografik faktörleri kapsayan çok çeşitli vakaların yer aldığı bir eğitim kümesi şarttır. Eğitim kümesi çeşitlilikten yoksunsa YZ modeli yeni vakalar ile karşılaştığında performansı düşecektir. Algoritma geliştirme sürecinde verilerin temsiliyeti kapsamlı şekilde değerlendirilmelidir. Toplanan veriler ile hedef popülasyon özellikleri arasındaki olası farklar (örneğin, hangi grupların temsil edilmediği veya temsiliyetinin az olduğu) belirlenerek raporlanmalıdır. Bu şekilde verilerin doğru yorumlanabilmesi ve sonuçların geçerliliği daha iyi değerlendirilebilir.

Eğitim kümesinin büyüklüğü ve temsiliyeti çok merkezli çalışmalar veya açık kaynak görüntü arşivleri kullanılarak artırılabilir. Ancak çok merkezli araştırma tasarımları, özellikle veri gizliliği ve güvenliği açısından zorlukları da beraberinde getirmektedir. Açık kaynak görüntü arşivleri ise özellikle nadir hastalıkların görüntülerine ulaşılması için yetersiz kalabilmektedir. Nadir hastalıklardan oluşan sınıfların dışlanması seçim yanlılığını beraberinde getireceğinden bu hastalıkların tümü "diğer" sınıfı altında gruplandırılarak, verinin hedef popülasyon temsiliyeti artırılabilir ve sınıf dengesizliği azaltılabilir.

Büyük ölçekli tıbbi görüntüleme verisi toplama ve paylaşımında birçok engel bulunmaktadır. DICOM meta etiketleri ve görüntü tetkiklerinin isimlendirilmesindeki farklılıklar, veri bütünlüğünü etkileyen temel sorunlardandır. Aynı vücut bölgesi farklı kliniklerde veya cihazlarda farklı başlıklarla etiketlenebilir. İnsan hataları ve veri yönetimi konusundaki eğitim eksiklikleri de veride değişkenlik ve hata yaratabilir. Bu sorunların çözümü, cihaz üreticileri arasında iş birliğini artırmayı ve veri yönetimi farkındalığını güçlendirmeyi gerektirir. Ayrıca, elektronik tıbbi kayıt sistemlerine giren verilerin analiz edilebilir hale getirilmesi tüm sağlık sistemlerinin hedefi olmalıdır. Serbest metin radyoloji raporlarından yapılandırılmış veri elde etmenin zorluğu doğal dil işleme teknikleri ile bir ölçüde aşılabılır.

Hasta gizliliği endişeleri ve kaliteli etiketlenmiş veriye ulaşımın zor olması gibi nedenlerle her zaman geniş veri kümeleri oluşturmak mümkün olmayabilir. Kısıtlı veriyle yüksek performanslı modeller geliştirmek için üç temel strateji kullanılmaktadır: (i) *veri artırma (data augmentation)*: Döndürme, gürültü ekleme, ölçeklendirme, kırpmaya, parlaklık ve kontrast ayarlama gibi yöntemlerle sentetik görüntüler oluşturularak veri kümesi büyütülür; (ii) *yarı denetimli öğrenme*: Tam etiketlemenin maliyetli olduğu durumlarda, kısmen eğitilmiş bir model tarafından "yalancı etiketleme" yapılan görüntüler de eğitime dahil edilir; (iii) *transfer öğrenimi*: büyük bir veri kümesi üzerinde önceden eğitilmiş bir model, başlangıç noktası olarak kullanılır ve ilgili göreve uyarlamak için ince ayar yapılır.

4.2.2. Veri İyileştirme (Veri Kürasyonu)

Daha geniş veri kümeleri kullanmak, her zaman daha yüksek performanslı modeller elde edilmesini garanti etmez. Çünkü, kalitesiz veya hatalı etiketlenmiş veriler modelin hata oranlarını da artırabilir. Yüksek performanslı modeller elde etmenin ön koşullarından birisi de iyi düzenlenmiş ve doğru etiketlenmiş veri kümeleri kullanmaktır. Ancak, emek yoğun etiketleme süreçleri nedeniyle bu tür veri kümelerinin oluşturulmasında zorluklar yaşanmaktadır.

Veri k rasyonu, verilerin kalitesini, tutarlılıđını ve genel b t nl đ n  sađlama y ntemlerini ifade eder. Veri tabanındaki tutarsızlıkları ve kopyaları belirlemek; eksik verileri, farklı g r nt  formatındaki verileri saptamak; d ř k kaliteli, yanlış inceleme penceresi veya g r nt leme protokol  ile planlanmış g r nt leri bulmak gibi faaliyetleri i erir.

Veri kalitesinin detaylı řekilde deđerlendirilmesi sıklıkla  nerilmektedir. Her bir deđerřkenin nasıl, ne zaman ve hangi ara larla  l  d đ  net bir řekilde belirtilmelidir. Verinin kalitesi konusundaki sınırlılıklar mutlaka raporlanmalıdır. Veri kalitesinin ve yanlış etiketleme kontrol  a ısından veriler m mk nse rastgele alt gruplar se ilerek yanlış etiketleme a ısından manuel olarak deđerlendirilmeli ve hata oranı raporlanmalıdır.

Kaynak alarda  zellikle referans standardı verilerinin kalitesine  nemli bir vurgu yapılmaktadır. Verilerin nasıl toplandıđı (biyopsi, radyoloji raporları, laboratuvar testleri vb.) ve toplama sırasında oluřabilecek sorunlar tartıřılmalıdır. Manuel etiketleme yapıldıysa etiketleme yapanların deneyimleri mutlaka belirtilmelidir. Model performansının deđerlendirilmesinde  n yargının ortadan kalkabilmesi i in etiketlemenin bađımsız olarak yapılması ve uzmanların model performansının deđerlendirilmesine direkt olarak katılmaması  nemlidir. Etiketlerin kalitesinin  l  lmesi i in g zlemciler arası farklılık hesaplanarak raporlanmalıdır.

Sık kullanılan  n iřleme teknikleri arasında yapay veri artırımı (data augmentation), aykırı deđerlerin ayıklanması, deđerřkenlerin d n řt r lmesi ( rneđin;  l ekleme, standardizasyon), eksik verilere tahmini deđer ataması yapma (imputation) yer alabilir. Verilere uygulanan d n ř mlerin  n iřleme basamaklarının nedenlerinin ve kullanılan yazılım ile birlikte a ık a belirtilmesi  nemlidir.

Kayıp verilerle bařa  ıkmak i in genellikle iki   z m  nerilir: eksik veriler i in tahmini deđer atama ve dıřlama. Bunun yanı sıra, kayıp verilerle  alıřabilen makine  đrenmesi y ntemlerinin kullanılması da bir se enektir. Genellikle, tahmini deđer ataması, verinin tamamen dıřarıda bırakılmasına g re tercih edilen bir   z m y ntemidir.

Farklı tarayıcılar, g r nt leme protokolleri veya rekonstr ksiyon parametrelerinin kullanılmasına bađlı g r nt   zellikleri deđerřkenlik g sterebilmektedir. Biyolojik ve tanısai bilgiyi maskeleyen bu sistemsel ve teknik farklılıklara yıđın etkisi de denilmektedir. Yıđın etkisi,  zellikle  ok merkezli  alıřmalar veya manyetik rezonans (MR) g r nt lerinin kullanıldıđı senaryolarda daha belirgin bir fakt r olabilir.   nk , manyetik rezonans g r nt lemede (MRG) piksellerin sinyal yođunluđu "Hounsfield Unit" (HU)'ten farklı olarak standardize olmayan bir řekilde belirlenir ve  reticiler arasında farklılık g sterir. Yıđın etkisini ortadan kaldırmayı ama layan   z mlere ise veri harmonizasyonu denilmektedir. Harmonizasyon g r nt  veya veri d zeyinde uygulanabilmektedir. Veri d zeyinde kullanılan etkin   z mlerden birisi olan ComBat, farklı cihaz veya merkezlerden elde edilen verileri uyumlu hale getirerek analitik sonu ların daha g venilir ve karřılařtırılabilir olmasını sađlar. G r nt  d zeyinde en  ok kullanılan yaklařım, piksel deđerlerinin kaydırıldıđı ve/veya  l eklendirildiđi bir dizi tekniđi i eren normalizasyon olup farklı g r nt   zelliklerine (uzamsal normalizasyon veya intensite normalizasyonu gibi) uygulanabilmektedir. Uzamsal normalizasyon g r nt n n uzamsal  zelliklerini (piksel boyutu, inceleme alanı-FOV, serilerin oryantasyonu vb.) deđerřtiren herhangi bir iřlemdir. Intensite normalizasyonu ile piksellerin grilik deđerleri aynı

aralığa yeniden ölçeklendirilir ve sayısal olarak büyük değerlerin etkisi dengelenmiş olur. Böylece, genellikle modelin eğitim süresi azaltılarak performansı iyileştirilmiş olur.

Görüntü yeniden örnekleme (resampling), dijital görüntülerin geometrik olarak bütün dönüştürme işlemlerinin genel adıdır. Bu, örneğin görüntü rotasyonu ile yeni veri noktalarının oluşturulması veya farklı zaman noktalarında elde edilen görüntülerin uzaysal olarak hizalanması gibi işlemleri içerir. Tıbbi görüntüler genellikle daha büyük hacimde ve üç boyutlu olduğundan, tıbbi olmayan görüntü görevlerinden daha karmaşıktır. Evrimsel sinir ağları genellikle daha küçük boyutlu (örneğin, 300 x 300 piksel) iki boyutlu görüntülerle eğitildiği için, tıbbi görüntülerin işlenmesi daha fazla hesaplama gücü gerektirebilir. Bu durumda, tıbbi görüntülerin çözünürlüğünün azaltılması veya yama tabanlı değerlendirme yapılması gerekebilir. Yama tabanlı değerlendirme, görüntü girdisinin alt bölümlere ayrıştırılmasıyla hesaplama yükünün azaltılmasını ve özel bir bölgeye (örneğin, prostat kanserini lokalize etmeyi amaçlayan bir algoritmanın yalnızca prostat bezini içeren piksellere odaklanması) dayalı sınıflandırma yapılmasını sağlar. Kantitatif görüntü analizi yöntemleri de görüntü yeniden örnekleme gerektirir. Radyomiks, doku heterojenitesini matematiksel formüllerle ifade etmeye çalıştığından, voksel boyutundan etkilenebilir. Bu nedenle, üç boyutlu radyomiks çalışmalarında voksellerin izotropik hale getirilmesi önemlidir.

Gürültü ve artefaktların giderilmesi yoluyla veri niteliğinin iyileştirilmesini içeren görüntü ön işleme ile model başarımı artırılabilir. Hareket düzeltme yordamları ile akciğer veya kardiyak bilgisayarlı tomografi (BT), dinamik görüntüleme ve difüzyon ağırlıklı görüntülemeye olduğu gibi dört boyutlu verilerle uzamsal eşleştirme yapılarak rahatlıkla çalışılabilir. Filtreleme ve hareket düzeltici yordamlar görüntülerdeki tanınabilir bilgiyi de bozabilen yöntemler olup özellikle ileriye yönelik çalışmalarda mümkün olduğunca kullanılmamalıdır. MR görüntüleri çoğunlukla manyetik alandaki inhomojeniteden kaynaklı olarak biyolojik doku özelliklerini bozan bir gürültüden olumlu yönde etkilendiğinden "bias-field" düzeltim yordamlarının kullanılması önerilmektedir.

4.2.3. Veri Gizliliğinin Sağlanması

Yapay zeka araçlarının geliştirilmesindeki temel engellerden birisi de veri gizliliğinin sağlanması ve etik problemlerin aşılmasıdır. Veri toplamaya başlamadan önce projenin yerel kişisel verilerin gizliliği mevzuatına uygun olması sağlanmalıdır. Gereklik halinde uygun gizlilik önlemlerinin alınabilmesi için veri koruma uzmanlarına danışılmalıdır. Özellikle ileriye dönük çalışmalarda hasta verilerini kullanmanın gerekçesi ve verilerin nasıl kullanılacağı konusunda hastalar bilgilendirilmeli ve aydınlatılmış onam alınmalıdır. Geriye yönelik çalışmalarda hastaların ek bir işleminden geçmesi gerekmediğinden aydınlatılmış hasta onamı gerekmeyebilir. Ancak, verilerin kullanılması için geçerli bir nedenin olup olmadığı her iki koşulda da etik kurullar tarafından mutlaka değerlendirilmelidir. Özellikle geriye yönelik çalışmalarda, hastalardan açık onam almanın olası bulunmadığı, veri paylaşımı ile ilgili risklerin en alt düzeyde olduğu ve veri sorumlularına güvenilebildiği koşullarda etik kurulların izni daha kolay olabilmektedir.

Gerçek dünyada kullanılabilir YZ tabanlı çözümlerin geliştirilmesi sürecinde özel sektör ile iş birliği yapılması gerekebilir. Yapay zeka

geliştiricilerinin PACS'a ("Picture Archiving and Communication System") doğrudan erişimi bulunmadığından verinin hazırlanarak transfer edilmesi gerekmektedir. Veri transferinden önce kimliksizleştirme yapılarak hasta gizliliğinin korunması şarttır. Kimliksizleştirme iki farklı yol ile yapılabilir: anonimleştirme ve yalancı anonimleştirme. Anonimleştirme, hasta ile ilgili bilgilerin kayıtlardan geri döndürülemez şekilde çıkarılmasını ifade eder. Bu yöntem, tıbbi verilerin paylaşımı için tercih edilen yaklaşımdır. Yalancı anonimleştirme ise, orijinal verilerin yalnızca gizli bir anahtarla ortaya çıkarılabileceği şekilde hasta bilgilerinin yapay değerler ile değiştirilmesini ifade etmektedir. Transfer edilecek hassas her türlü sağlık verisinin hem DICOM ("Digital Imaging and Communications in Medicine") meta bilgisinden hem de görüntülerden kaldırılması gereklidir. Çoğu PACS'ın dışarı aktarma sırasında sunduğu temel anonimleştirme yöntemleri yeterli olmayabilir. DICOM meta bilgisindeki tanımlanabilir bilgileri otomatik ve ücretsiz olarak kaldırmak için birden fazla araç bulunmaktadır. DICOM Library ve RSNA Clinical Trials Processor bu amaç için kullanılabilir. Ücretsiz ve kanıtlanmış iki farklı araç setidir. Anonimleştirilmiş ve meta bilgiden arındırılmış görüntüler bile, özellikle kafa görüntülemesinde, yüz özelliklerinin tespiti ile hastalarının kimlik bilgilerine ulaşmaya imkan sağlayabilir. Pydeface (<https://github.com/poldracklab/pydeface>) veya mridefacer (<https://github.com/mih/mridefacer>) gibi araçlar, bu tür yüz özelliklerinin tıbbi görüntülerden otomatik olarak kaldırılmasına yardımcı olabilir.

Veriler kimliksizleştirme sonrasında fiziksel sunucular veya buluta aktarılarak belirli bir süre depolanabilmektedir. Bulut tabanlı veri depolama güvenlik konusunda endişeleri beraberinde getirebileceği gibi pahalı ve yüksek internet hızı isteyen bir çözümdür. Ancak, veri paylaşımını kolaylaştırarak çok merkezli çalışmalarda büyük fayda sağlayabilir. Veri depolama sürecinde gizliliğin korunması için verilerin şifrelenmesi, verilere erişimin yetkili kişiler ile sınırlandırılması, veriler üzerinde gerçekleştirilen tüm hareketlerin işlem kayıtlarının güvenli olarak loglanması, verilerin bulunduğu ortamlara ait güvenlik testlerinin düzenli olarak yapılması ve veri ihlallerinin şeffaf bir şekilde raporlanması gerekmektedir. Ayrıca, yetkisiz erişimden kaynaklanan riskleri en aza indirmek için verilerin ne kadar süreyle saklanabileceği belirlenmelidir.

Ülkemizde de 6698 sayılı Kişisel Verilerin Korunması Kanunu (KVKK) yayımlanmış ve böylece kişisel verilerin korunması, kaydedilmesi, işlenmesi ve paylaşılması gibi faaliyetler düzenlenmiştir. Buna ek olarak 21 Haziran 2019 tarihinde Resmî Gazetede yayımlanarak yürürlüğe giren "Kişisel Sağlık Verileri Hakkında Yönetmelik" ile Sağlık Bakanlığı sağlık verilerinin işlenmesini, erişimini, gizlenmesini, düzeltilmesini, imha edilmesini, aktarılmasını ve güvenliğini kapsayan genel çerçeveyi belirlemiştir. Yönetmeliğin 16. maddesine göre veri sorumlusu tarafından anonim hale getirilmesi koşuluyla kişisel sağlık verileri ile bilimsel çalışma yapılabileceği belirtilmiştir. Kişisel sağlık verilerinin aktarılması ise yönetmeliğin 15. maddesi ile düzenlenmiştir. Buna göre sağlık verileri ancak kamu sağlığının korunması, koruyucu hekimlik, tıbbî teşhis, tedavi ve bakım hizmetlerinin yürütülmesi, sağlık hizmetleri ile finansmanının planlanması ve yönetimi amacıyla, sır saklama yükümlülüğü altında bulunan kişiler veya yetkili kurum ve kuruluşlar tarafından ilgilinin açık rızası aranmaksızın aktarılabilir. Diğer bir deyişle, sağlık verilerinin gizliliğini düzenleyen yönetmelikler, tamamen anonimleştirilmiş verilerin

işlenmesi için açık hasta rızası gerekmediğini ve bu şekilde verilerin kötüye kullanılmasının teorik olarak engellenebileceğini belirtmektedir. KVKK'nın 12. maddesinde veri güvenliğine ilişkin yükümlülükler tanımlanmıştır. Buna göre, veri sorumlusu, kişisel verilerin güvenliğini sağlamak ve hukuka aykırı erişim ve işlemeyi önlemek için gerekli tüm teknik ve idari tedbirleri almak ve bu yükümlülüğü veri işleyenlerle birlikte müştereken yerine getirmek zorundadır. Bu sebeple, bilimsel araştırma projelerinde araştırmacıların verilerin saklanması, erişimi, paylaşımı ve yok edilmesi ile ilgili güvenlik ve gizlilik önlemlerini alarak bunları bir veri yönetim planında kayıt altına alması gerekebilir.

Veri gizliliği ile ilgili endişelerin ve mevzuat kısıtlamalarının üstesinden gelebilmek için ortaya çıkmış olan birleşik öğrenme (federated learning) ile veriler hastane dışına aktarılmaz, bunun yerine YZ modeli hastanelere gönderilerek burada eğitilmesi sağlanır. Doğası gereği geliştiricilerin verilere kör olması nedeniyle hata ayıklamasının zor olması model performansının beklenin altında kalmasına yol açabilir.

4.2.4. Görüntü Etiketleme ve Anotasyon

Denetimli makine öğrenmesi, "etiket" verisiyle tıbbi görüntü veya klinik değişkenlerin eşleştirilmesi için yordamlar oluşturmayı içerir. Etkin bir model geliştirilmesi etiketli verilerin sayısına ve kalitesine bağlıdır. Etiketli veriler üzerinde eğitildikten sonra, algoritma öğrendiği kalıplara göre yeni bir görüntünün etiketini tahmin etmeye çalışır. Referans standardı, modelin öğrenmesi gereken bilgiyi temsil etmektedir. Modelin amacına göre referans standartları değişkenlik gösterebilir. Örneğin, lokalizasyon görevi için sınırlayıcı kutular (bounding boxes), segmentasyon görevi için uzmanlar tarafından oluşturulmuş piksel tabanlı maskeler, sınıflandırma görevi için kategori etiketleri içeren görüntüler, izlem görevi için ölçüm işaretlemeleri veya prognoz görevi için klinik sonuçlar (sağ kalım, metastaz, progresyon vb.) gibi.

Modelin amacına bağlı olarak etiketler radyoloji raporlarından, uzman incelemelerinden, klinik veya patolojik verilerden kaynaklanabilir. Yüksek kapasiteli sinir ağlarının eğitilmesi için büyük veri kümelerinin etiketlenmesi gerekliliği ve etiketleme görevinin emek yoğun doğası, alternatif çözümlerin gerekliliğini ortaya koymaktadır. Bunlar arasında YZ destekli otomatik etiketleme araçları (örneğin, otomatik segmentasyon araçları veya radyoloji raporlarından otomatik etiket çıkarılması gibi), kitle kaynağı ile etiketleme ve zayıf denetimli öğrenme (tamamlanmamış etiketler veya düşük kaliteli etiketler ile eğitim) bulunmaktadır. Yapay zeka destekli otomatik etiketleme yöntemleri, büyük veri kümelerini hızlı bir şekilde etiketlemeye olanak tanırken, her senaryoya uygun otomatik etiketleme çözümleri bulunmayabilir ve etiket kalitesinin kontrolü için insan faktörü gerekir. Otomatik etiketleme yöntemlerinden biri de doğal dil işleme veya tekrarlayan sinir ağları ile radyoloji raporlarından bilgi çıkarımıdır. Yapılandırılmamış radyoloji raporlarından çıkarılan bilgiler insan kaynaklı (yazım yanlışları, yorumlama hataları vb.) veya YZ kaynaklı (dil kullanım stiline bireysel farklılıklar vb.) hatalar içerebilir. Radyoloji raporlarının %2-20'sinin gösterilebilir hatalar içerdiği tahmin edilmektedir. Raporlardan daha kaliteli veri elde etmek ve verileri daha kolay erişilebilir hale getirmek için yapılandırılmış raporlama bir çözüm olabilir. Etiketleme görevinin daha fazla sayıda insana paylaştırılması bireysel etiketleme yükünü azaltabilir, ancak bu durumda etiketleyiciler arasında uyum ve tutarlılık sorunları ortaya çıkabilir. Günümüzde otomatik hizmetleri desteklemek

için büyük miktarlardaki verinin incelenmesi ve etiketlenmesi için uzman olmayan kişiler kullanılabilmektedir. Son zamanlarda yapılan çalışmalarda bazı medikal görüntülerin etiketlenmesinde uzman olmayan gözlemcilerin kullanılabileceği raporlanmıştır. Bu yöntemler, etiketleme sürecini kolaylaştırır da genellikle elde edilen verilerin düşük kaliteli oluşu önemli bir kısıtlılıktır. Ancak, geniş veri kümeleri kullanılarak eğitilen modellerin düşük kaliteli verilerle bile iyi sonuçlar gösterebileceği bilinmektedir. Verilerin eksik veya güvenilmez olduğu bu tip durumlardaki model eğitimine zayıf denetimli öğrenme denir. Eksik etiketlenmiş verilerle başa çıkmak için kullanılabilecek iki temel zayıf denetimli öğrenme yaklaşımı aktif öğrenme ve yarı denetimli öğrenmedir. Aktif öğrenme, uzmanlar tarafından veri havuzundan en bilgilendirici örneklerin yinelenmeli olarak etiketleme görevi için seçilmesi ve minimum insan etiketleme çabasıyla modelin eğitilmesini amaçlar.

Referans standardının belirlenmesi, bir çalışmanın güvenilirliği ve sonuçlarının doğruluğu açısından kritik bir öneme sahiptir. Etiketleme sürecinde, raporlardan çıkarılan zayıf etiketler yerine mümkün olduğunca uzman radyologların yeniden değerlendirme ve etiketlemesine öncelik verilmelidir. Birden fazla etiketleyici varsa etiketlerin güvenilirliği, gözlemciler arası uyum (örneğin Cohen'in kappa istatistiği veya intra-class correlation-ICC) gibi yöntemlerle ölçülmelidir. Etiketleme öncesinde standart bir etiketleme kılavuzu belirlenmesi (örneğin, hangi bulguların "pozitif" veya "negatif" sayılacağı gibi), gözlemcilerin bu konuda eğitilmesi, küçük bir test grubunda deneme etiketlemesi yapılması tutarlılığı artıracaktır. Düzenli aralıklarla rastgele seçilen vakalar üzerinden çapraz kontrol yapılarak etiketleme kalitesi ve uyumu izlenebilir. Manuel etiketleme yönteminin kullanıldığı durumlarda, etiketleme sürecinin detayları açıkça raporlanmalıdır. Bu, etiketleme yapan kişinin deneyimi, zor vakaların ele alınmasındaki strateji ve önyargıyı minimize etmek için alınan önlemleri içermelidir. Ayrıca, manuel etiketleme sürecinde gözlemciler arası değişkenliğin nasıl ele alındığı da raporlanmalıdır.

4.3. Modelleme ve Geçerlilik Testi (Validasyon)

Model, belirli kalıpları tanımak amacıyla eğitilen bir program veya algoritmadır. Bu algoritmalar giriş verilerinden öngörü, saptama, sınıflandırma, bölütleme veya öneri gibi çıktılar üretir. Görüntü analizinde kullanılan modelleme yaklaşımları genellikle iki gruba ayrılır: geleneksel makine öğrenimi (önceden belirlenmiş özelliklere ihtiyaç duyar) ve derin öğrenme (özellikleri verilerden otomatik olarak öğrenir). Modelleme aşamasında kullanılacak en uygun yöntemin belirlenmesi; veri tipine (örneğin metin, görüntü, kategorik veya sayısal), problem türüne (örneğin sınıflandırma, regresyon veya sağ kalım analizi), erişilebilir bilgi işlem kaynaklarına (CPU veya GPU), problemin karmaşıklığına ve modelin doğruluk ile yorumlanabilirliği arasında tercih edilen dengeye bağlıdır. Geleneksel makine öğrenimi yöntemlerinde, görüntülerden önceden tanımlanmış ve formüle edilmiş özellikler (örneğin radyomiks doku özellikleri) çıkarılır ve bu özellikler kullanılarak çeşitli makine öğrenimi algoritmalarıyla modeller oluşturulur. Bu yaklaşım, küçük veri kümelerinde genellikle daha kullanışlı ve yorumlanabilir sonuçlar sunar. Ancak, görüntü analizi ve metin tanıma gibi karmaşık problemler için performans açısından sınırlı olabilir.

Geleneksel makine öğreniminde sıklıkla kullanılan algoritmalara örnek olarak destek vektör makinesi (SVM), rastgele orman (RF), k-en yakın komşu (kNN) ve Naive Bayes gösterilebilir. Her problem için en iyi sonucu verecek tek ve evrensel bir algoritma bulunmamaktadır. Bu nedenle, farklı algoritmaların ayrı ayrı eğitilerek karşılaştırılması ve her probleme özgü en uygun algoritmanın belirlenmesi önerilir.

Derin öğrenme ise çok katmanlı sinir ağlarına dayanan bir makine öğrenimi alt alanıdır. Bu yöntem, karmaşık özellikleri eğitim süreci sırasında ham verilerden otomatik olarak öğrenir ve böylece özelliklerin manuel olarak tasarlanması gerekliliğini ortadan kaldırır. Bununla birlikte derin öğrenmenin dezavantajları da bulunmaktadır: yüksek miktarda etiketli veriye ihtiyaç duyması, eğitim sürecinin uzun ve maliyetli olması ve "kara kutu" yapısı nedeniyle yorumlanabilirliğinin düşük olması bunlardan başlıcalarıdır. Günümüzde farklı problemlerin çözümü için özelleştirilmiş birçok derin öğrenme mimarisi geliştirilmiştir. Bu mimariler arasında en önemlileri evrimsel sinir ağları (CNN), tekrarlayan sinir ağları (RNN), dönüştürücü (transformer) ve üretici modellerdir (generative model). Örneğin, CNN özellikle görüntü analizi alanında en sık kullanılan derin öğrenme mimarisidir. Sonuç olarak, sınırlı ve iyi yapılandırılmış veri kümelerinde veya hızlı ve yorumlanabilir sonuçlar gerektiğinde geleneksel makine öğrenimi yöntemleri tercih edilebilir. Buna karşın, geniş ve karmaşık veri kümeleri ile uğraşırken, özellikle görüntü analizi veya metin tanıma gibi problemlerde derin öğrenme yöntemleri daha uygundur.

Tıbbi görüntüleme alanında üretilen YZ modelleri için, makine öğrenimi ve tıp literatüründe kullanılan "validasyon" terimleri arasında bir karışıklık mevcuttur. Makine öğrenimi penceresinden bakıldığında, validasyon en iyi modeli seçmeyi ve onu ayarlamayı ifade eder. Ancak tıp açısından bakıldığında, validasyon genellikle, makine öğrenimindeki "test seti"ne benzer bir şekilde, bir modelin performansını görünmeyen veriler üzerinde doğrulama sürecini ifade eder. Bu farklılık karışıklığa neden olabilir, bu yüzden tıbbi bağlamda bazen "doğrulama seti" yerine "geliştirme seti" ya da "geçerlilik test seti" terimleri kullanılır.

Veriler ideal olarak üç kümeye ayrılır; eğitim, doğrulama (geliştirme) ve test. Optimum ayırma oranı her problem için değişmekte olup tek bir çözüm yoktur. Temel kural olarak veriyi %80 eğitim, %10 doğrulama ve %10 test olarak parçalamak yaygın bir uygulamadır. Bununla birlikte, bazı kaynaklarda doğrulama (geliştirme) kümesinin kullanılmadığı görülmektedir. Bu durumda eğitim:test oranının genellikle 60:40 ile 90:10 arasında seçilmektedir. Doğrulama kümesi en iyi sonucu elde etmek için optimal hiperparametreleri belirlemek ve en iyi modeli seçmek amacıyla kullanılır. Modelin genellenebilirliğini ölçmek için test seti yalnızca bir kez kullanılır. Daha küçük veri kümelerinde bölünmüş doğrulama şeması eğitim setinde yeterli çeşitlilik sağlanamaması ve test setindeki performans ölçümünün yanıltıcı olabilmesi nedeniyle güçlü modellerin üretilmesini engelleyebilir. Bu gibi durumlarda daha fazla veri elde etmek mümkün değilse çapraz validasyon kullanılabilir. Bu yöntemde veri kümesi k eşit katmana bölündükten sonra her eğitim için algoritma neredeyse tüm katmanlar üzerinde eğitilir ve dışarıda bırakılan katmanda test edilir. Nihai performans ölçülen k adet performansın ortalaması olarak kaydedilir.

Modelin potansiyelini ve genellenebilirliğini doğru bir şekilde değerlendirmek için harici bir veri setinde bağımsız doğrulama, dahili

doğrulamaya tercih edilmelidir. Tıp literatüründe, harici doğrulama genellikle bir modelin güvenliğini, güvenilirliğini ve genellenebilirliğini kesin olarak değerlendirmek için nihai test olarak kabul edilir. Ancak, literatürde harici doğrulama yapılmış çalışma sayısının %6 civarında olduğu görülmektedir. İdeal olan, dış doğrulamada kullanılan verilerin modelin nihai kullanımının amaçlandığı popülasyonları yansıtmasıdır. Ancak, bu veriler sıklıkla uygunluk ve erişilebilirlik temelinde seçilmektedir. Dış doğrulamanın kural olarak farklı kurumlardan bağımsız araştırmacılar tarafından gerçekleştirilmesi gerekir. Prospektif doğrulama oldukça nadir olsa da literatür tarafından tercih edilir, çünkü modellerin gerçek kullanılabilirliği hakkında daha iyi bir fikir verebilir.

Teknik başarımlar geçerlilik testi, modelin başarımının eğitim ve test kümelerinde çeşitli kantitatif metrikler kullanılarak ölçülmesidir. Performans doğrulama metrikleri modelin amacına göre değişebilir. Örneğin, sınıflandırma problemleri için ayırt etme receiver operating curve (ROC) altındaki alan, duyarlılık, özgüllük, pozitif ve negatif tahmin değerleri, kalibrasyon ve karar eğrisi analizi; tespit etme (saptama) problemleri için intersection over union (IoU) ve/veya mean average precision (mAP); segmentasyon problemleri için IoU ve/veya Dice skoru kullanılabilir. Tüm başarımlar ölçütleri hem eğitim hem de test kümeleri için ayrı ayrı raporlanmalıdır. Bu, modelin aşırı uyum durumunu değerlendirmede bilgilendirici olduğu için önemlidir. Ayrıca, modelin amacına uygun birden fazla performans ölçütünün skorlarının güven aralıkları ile birlikte raporlanması ve hatalı sınıflandırılan vakaların hata analizlerinin yapılması önemlidir. Modelin başarımının mevcut en iyi radyolojik uygulamalar veya alternatif YZ algoritmaları ile karşılaştırılması, performans değerlendirme sürecinde gerekli bir yöntemdir.

Klinik geçerlilik çalışmaları, modellerin entegrasyonunu hızlandırabilecek potansiyele sahip olabilir, ancak henüz bu tür çalışmalar için standartlar belirlenmemiştir. Bununla birlikte, radyologların bu standartların oluşturulması ve uygulanmasında aktif bir rol oynamaları son derece önemlidir. Klinik doğrulamada, modelin doğruluğunun yanı sıra YZ kullanımıyla artan patoloji belirleme oranı, raporlama süresindeki hızlanma, hasta sağkalımındaki değişim veya maliyet etkinlik analizleri yapılabilir. Modelin entegrasyonu ile birlikte, hasta bakım kalitesinde veya hastane operasyonlarındaki potansiyel değişimler analiz edilebilir. Modelin hatalı tahminleri (yanlış pozitif ve yanlış negatif) olgu bazında incelenerek olası hata senaryoları belirlenebilir.

Gerçek dünyaya uyarlanabilir ve genellenebilir etkin bir modelin oluşturulmasının önündeki en büyük engellerden biri aşırı uyumdur (overfitting). Bu durumun diğer bir nedeni, örneklem hatası içeren ve çeşitliliği düşük bir veri kümesi ile eğitim yapılmasıdır. Model, eğitim kümesindeki özelliklere fazla uyum sağladığında, gürültüye veya veri içindeki rastgele değişikliklere aşırı duyarlı hale gelir ve yeni veri kümelerine genellenebilir olmaz. Bu da model performansında düşüşe neden olabilir. Aşırı uyumun anlaşılmasında ayarlama seti oldukça önemlidir. Model, eğitim setinde doğrulama setine kıyasla iyi performans gösteriyorsa, muhtemelen eğitim verilerine aşırı uyum sağlamıştır. Aşırı uyumu azaltmak için modelin karmaşıklığını azaltmak veya daha fazla veriyle eğitim yapmak gibi adımlar atılmalıdır.

Eğitim veri kümelerinin aksine, doğrulama veri kümelerinin merkezi olarak tutulması ve tarafsız üçüncü kişiler tarafından model validasyonu yapılması, modellerin performans sonuçlarına duyulan

güveni artırabileceği gibi klinik entegrasyonunu da kolaylaştırabilir. Yapay zeka modellerinin gelişimine yön verme, merkezi ve kapsayıcı bir validasyon kümesi sağlayarak test aşamasını güvenilir ve etkin kılma, aynı amaçtaki modeller arasından en iyi olanlarının seçilmesiyle radyologlar ve düzenleyici kurumların güvenini artırmada düzenlenen yarışma (challenge) etkinlikleri faydalı olabilir.

Doğrulama veri kümelerinin referans standartlarının eğitim için kullanılan veri kümelerinden daha yüksek standartlara sahip olması önemlidir. Radyologlar, doğrulama veri kümelerinin mümkün olduğunca hedef popülasyona yakın olmasını sağlamakla kalmayıp algoritma doğrulaması için kullanılan ölçütlerin tanımlanmasında da kritik bir rol oynamalıdır.

4.3.1. Başarım (Performans) Değerlendirmesi

Modelin sonuç başarısını değerlendirirken, sadece görev başarısı değil, aynı zamanda şeffaflık (açıklanabilirlik), klinik fayda, güvenlik ve istikrar da dikkate alınmalıdır. Bununla birlikte, modelin hataları analiz edilmeli ve modelin kısıtlılıkları ile potansiyel gelişim fırsatları da araştırılmalıdır.

4.3.1.1. Öngörü Başarımı Değerlendirmesi

Basit ikili sınıflandırma problemlerinde, tanısal araçların veya yordamların başarımı, duyarlılık, özgüllük, pozitif ve negatif tahmin değeri gibi ölçütlere dayalı olarak değerlendirilebilir. Değişen eşikler ile çoklu duyarlılık ve özgüllük çiftleri üretildiğinden karşılaştırmada kullanılacak tekil bir performans metriği ihtiyacı doğar. Receiver Operating Curve (ROC) analizi, bu senaryoda modelin ayırt etme başarımını etkili bir şekilde değerlendirebilen önemli bir ölçüttür. ROC eğrisinin en yaygın özet ölçüsü, eğri altındaki alan olarak bilinen C istatistiği veya AUC'dir. AUC, hastalığı olan bir hastanın modelden sağlıklı bir hastaya göre daha yüksek bir risk puanı alma olasılığını yansıtır. AUC, 0 ile 1 arasında değişir ve tanı testinin ayırt edicilik performansını belirler. Ancak, 0.5'lik bir AUC değeri rastgele tahmin etmekle eşdeğer olduğundan pratik olarak alt sınır olarak kabul edilmektedir. Sınıf dengesizliği durumunda AUC tek başına doğru sonuç vermeyebileceğinden doğruluk, duyarlılık, özgüllük, F1 ölçütü ve Matthews korelasyon katsayısı gibi diğer performans ölçütlerinin de raporlanması önemlidir. Özellikle sınıflandırma görevlerinde, modellerin yetkinlikleri hakkında daha ayrıntılı bilgiler sağlayabilecek konfüzyon matrislerinin değerlendirilmesi önemlidir. Performans ölçütleri klinik senaryo ve modelin amacına göre verilerin özelliklerine göre dikkatle seçilmelidir. Klinisyenler ile birlikte net ve ölçülebilir başarı kriterleri ve performans eşikleri önceden belirlenmelidir. Çok kategorili sınıflandırmalarda (örneğin, birden fazla patolojik tanı durumu), her bir sınıf için ayrı ayrı performans istatistikleri raporlanmalıdır. Bu istatistiklerin ortalaması alınabilir veya her bir sınıfın test setindeki yaygınlığına göre ağırlıklandırılabilir. Ayrıca, belirli sınıfların sıklıkla karıştırılıp karıştırılmadığını değerlendirebilmek için hata matrisleri mutlaka verilmelidir.

Regresyon görevlerinde sürekli ve sayısal bağımlı bir değişken tahmin edilmeye çalışılır. Regresyon modelleri için R2, ortalama kare hatası, kök ortalama kare hatası, kök ortalama kare logaritmik hatası ve ortalama mutlak hata gibi ölçütler kullanılır.

Obje segmentasyonu başarısını değerlendirmek için çok çeşitli metrikler bulunmaktadır. Ancak, bu amaçla genellikle örtüşme temelli ölçütler yaygın olarak kullanılmaktadır. Objeye tespiti ve segmentasyonu, model tarafından tahmin edilerek işaretlenen bir alanın, genellikle bir radyolog tarafından belirlenmiş olan referans alan ile ne kadar iyi eşleştiğini tanımlayan metrikleri gerektirir. Kesişim (intersection), referans alan ile modelin tahmin ettiği alan arasındaki örtüşme iken birleşme (union), tahmin edilen alan ve referans alan tarafından kapsanan toplam alandır. Birleşme üzerinden kesişim (Intersection over Union, IoU) ve Dice skoru, segmentasyon görevlerindeki model performansını ölçmek için yaygın olarak kullanılan iki metriktir. Ancak, her iki ölçüm yönteminin bazı kısıtlılıkları ve tuzakları da bulunmaktadır. Örneğin, küçük yapıların segmentasyon hatalarına fazla hassas olma, hatalı referans segmentasyonların varlığında gereğinden düşük skorların elde edilmesi, tahmini segmentasyonların şekline duyarlı olmama ve fazla ya da eksik segmentasyonlardan eşit şekilde etkilenmeme gibi dezavantajlar, bunlar arasında en bilinenleridir. Objeye tespiti görevinde IoU için bir eşik değeri belirlenerek tespitin doğru olup olmadığı sınıflandırma problemlerinde olduğu gibi incelenebilir.

4.3.1.2. Kalibrasyon Başarımı Değerlendirmesi

ROC eğrileri yaygın olarak kullanılmasına rağmen, bazı kısıtlılıkları bulunmaktadır. Temel olarak, AUC bir sıralama ölçütüdür ve modelin hastaları ne kadar iyi ayırdığını gösterir. Fakat, olasılık tahminlerinin doğruluğu hakkında bilgi vermez. Bu nedenle, tahmin modeli değerlendirilmesinde kalibrasyon da oldukça önemlidir. Kalibrasyon, tahmin edilen olasılık değerleri ile gerçek olasılıklar arasındaki benzerliği değerlendiren bir yöntemdir.

Kalibrasyon eğrisinde, x eksenini tahmini olasılıklar iken y eksenini gerçekleşen olasılıklardır. Tahmini olasılıklar model tarafından verilen sayısal çıktılardan elde edilir. Ancak, gerçekleşen olasılıklar gözlemlenebilir değildir (örneğin, örnekler ya hastalık pozitif ya da negatiftir). Bu nedenle, model tahminlerinin gerçeğe ne kadar uyumlu olduğunu tespit edebilmek için benzer tahmini olasılık hesaplanan hastalar gruplandırılarak (tipik olarak tahmin değerlerinin ondalık dilimlerine göre) x eksenini düşük olasılık grubundan yüksek olasılık grubuna doğru yerleştirilir. Her gruptaki ortalama tahmin değeri x koordinatı, gerçek olasılık (yani, gruptaki hasta olan örneklerin toplam örneklerle oranı) y koordinatı olarak kullanılarak elde edilen noktalardan bir grafik çizilir. Mükemmel kalibrasyon, 45 derece çizgisi ile paraleldir. Hosmer-Lemeshow uyum iyiliği testi gibi istatistiksel testler, görsel değerlendirmeyi tamamlar. Anlamlı olmayan sonuçlar (örneğin, $p > .05$), iyi bir kalibrasyonu önerir, ancak yetersiz veri boyutu yalancı iyi kalibrasyon sonuçlarına sebep olabilir.

4.3.1.3. Klinik Fayda Analizi

Yapay zekanın tahmin odaklı ilerleyişi, klinik etki ve sonuçlarının göz ardı edilmesine neden olmuştur. Ancak, yüksek tahmin başarısına sahip bir model her zaman klinik açıdan faydalı olmayabilir. Çünkü klinik fayda, modelin ayırt etme başarımının yanı sıra kalibrasyon performansı ile ilgilidir. Örneğin, bir testin risk tahmin skorlarını ona bölsek bile, en uygun eşik değeri de buna göre değişeceğinden testin sınıflandırma performansı değişmez. Ancak, hastanın kanser olma olasılığı %30'dan %3'e düşerse, hastanın veya doktorun biyopsi kararı etkilenir ve agresif bir kanserin kaçırılma riski ortaya çıkar. Bu noktada karar eğrisi analizi

(KEA), tanı araçlarının hem ayırt etme hem de kalibrasyon performansını dikkate alarak testlerin klinik faydasını değerlendiren bir yöntem olarak değerlidir. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) rehberi tarafından da kullanımı önerilmektedir.

Karar eğrisi, çeşitli klinik stratejilerin farklı risk eşiklerine göre net faydalarını grafiksel olarak gösteren bir araçtır. Yatay ekseninde önceden belirlenen olası risk eşikleri bulunurken, düşey ekseninde net fayda yer alır. Karar eğrisinin altındaki alan (KEA), tahmin modelinin klinik faydasının bir ölçüsüdür. Her KEA'da, iki olası seçenek yaklaşımı gösteren biri yatay diğeri çapraz en az iki referans çizgisi vardır. Noktalı yatay çizgi, hiçbir hastanın tedavi edilmediği bir stratejinin net faydasını gösterir. Hiçbir hastayı tedavi etmemenin net faydası her zaman sıfırdır. Diyagonal kesikli çizgi ise tüm hastaların tedavi edildiği bir klinik stratejinin net faydasını göstermektedir. Hastalık prevelansının altındaki risk eşikleri için, "bütün hastaları tedavi etmek", "hiçbir hastayı tedavi etmemek" seçeneğinden daha yüksek bir net faydaya sahiptir. Prevelans üzerindeki risk eşikleri için ise bunun tersi geçerlidir ve "bütün hastaları tedavi etmek" yaklaşımı negatif bir net fayda oluşturacaktır. Bu iki çizgi, mümkün olan en uç iki stratejiyi temsil eder. Gerçekte, başka herhangi bir klinik strateji belirli hastaların tedavi edilmesini ve diğerlerinin tedavi edilmemesini içerecektir. Bir modelin belirli bir eşik değerinde klinik açıdan faydalı olarak kabul edilebilmesi için "tümünü tedavi et" ya da "hiçbirini tedavi etme" seçeneğinden daha yüksek net faydaya sahip olması gerekir.

4.3.2. Açıklanabilirlik ve Yorumlanabilirlik

Açıklanabilir YZ, sorumlu YZ uygulamalarının temel bir parçasıdır. Yorumlanabilirlik (interpretability) ve açıklanabilirlik (explainability) terimleri, makine öğrenmesi kaynaklarında karıştırılan iki kavramdır. Yorumlanabilirlik, modelin iç işleyişini ve çıktıları doğrudan (ilave bir analiz tekniğine gerek kalmadan) anlayabilmemizi ifade eder. Açıklanabilirlik ise özellikle karmaşık veya "kara kutu" nitelikteki modellerin hangi sebeple belli bir karara ulaştığını gösterebilmek için LIME, SHAP ve Grad-CAM gibi ek araç veya yöntemler kullanma gerekliliğini vurgular. Dolayısıyla, yorumlanabilirlik modelleme yönteminin içkin bir özelliği iken, açıklanabilirlik genellikle model eğitildikten sonra geliştirilen ek çözümleri kapsar. Açıklanabilir YZ, insan karar verme sürecini desteklemek, YZ sistemleri ve insanlar arasındaki şeffaflığı artırmak, beklenmedik davranışlar ortaya çıktığında hata ayıklamayı mümkün kılmak, yasal gereklilikleri karşılamak için denetimi kolaylaştırmak, YZ'ye güveni uygun seviyelere çekerek klinik entegrasyonu kolaylaştırmak ve genelleme yeteneğini doğrulamak gibi çeşitli amaçlara hizmet eder.

Literatürde önerilen çok sayıda açıklanabilir YZ tekniği bulunmaktadır. Açıklanabilir YZ tekniklerini sınıflandırmak için üç farklı unsur temel alınmaktadır: (i) yerel ve genel, (ii) modele özgü ve modelden bağımsız, (iii) model öncesi (doğal) ve model sonrası.

Doğal açıklanabilirlik, basit yapıları nedeniyle yorumlanabilir olarak kabul edilen makine öğrenimi modellerini (karar ağaçları, doğrusal regresyon vb.) ifade eder. Model sonrası açıklanabilirlik, modelin sonradan uygulanan çeşitli yöntemlerle (permütasyon özellik önemi, bütünleşmiş gradyan vb.) açıklanabilir hale getirilmesini ifade eder. Model bağımsız yöntemler, farklı YZ modelleri ile birlikte

çalışabilen esnek tekniklerdir, oysa modele özgü yöntemler, adından da anlaşılacağı gibi yalnızca belirli modellerle birlikte çalışabilirler. Genel açıklanabilirlik yöntemleri, modelin tahminlerine dayanak sağlayan genel kalıpları anlamak için tüm veri kümesini analiz eder. Bu yöntem, verilerdeki hangi örüntülerin modelin tahminleri için önemli olduğuna dair açıklamalar sağlar. Öte yandan, yerel açıklanabilirlik yöntemleri model tarafından yapılan tek bir tahmini açıklamaya odaklanır. Yani, modelin neden tek bir çıktıyı o şekilde verdiğine dair iç görü sağlamayı amaçlar.

Modellerin açıklanabilirliği üç farklı yöntem ile iyileştirilebilir; özellik, metin ve örnek tabanlı. Özellik tabanlı yöntemler, hangi girdinin model sonucu üzerinde en büyük etkiye sahip olduğunu özgün görüntü üzerine işaretleyebilirler. Metin tabanlı yöntemde, modelin kararını açıklamak için semantik tanımlamalar (örneğin, mamografide obje saptama ve sınıflandırma görevi bulunan bir algoritma saptanan lezyonun malignansi olasılığının yanında buna etki eden spiküler kenar özelliği veya yüksek dansiteden bahsedebilir) kullanılır. Ek olarak, etki fonksiyonu yöntemleri, bir YZ modelinin belirli bir tahmin sonucunun hangi eğitim örneklerinden etkilendiğini analiz etmeye yarar.

Görselleştirme yöntemleri, kullanıcıların sisteme olan güvenini artırmanın yanı sıra sistemin performansını etkileyebilecek hastalıkla ilişkili olmayan karıştırıcı faktörlerin (örneğin, pnömotoraks tanısının göğüs tüpü yardımıyla konması veya portabl göğüs filmlerinde pnömoni prevalansının artması gibi) ortaya çıkarılmasına da yardımcı olabilir. İlgili haritaları sık kullanılan görselleştirme yöntemleri olup görüntüleri doğru sınıflandırmada önemli olan pikselleri işaretleyerek ısı haritaları oluştururlar. Ancak, bu haritaların bazı riskleri ve kısıtlılıkları bulunmaktadır. Bu yöntemler, öncelikle modelin genel davranışını analiz etmekten ziyade yalnızca bireysel tahminleri açıklamada kullanılan yerel açıklayıcılardan birisidir. Bunun yanında, bir görüntüdeki vurgulanmış bir alanın gerçekte neyi ifade ettiğinin anlaşılması bazen zor olabilir.

Özetle, basit modelleme metodolojileri (lineer regresyon, karar ağaçları vb.) doğal olarak yorumlanabilir olduklarından ek bir yöntemle ihtiyaç duymazlar. Ancak, derin öğrenme gibi karmaşık modellerin karar süreçlerinin anlaşılması için ek araçlar kullanılması şarttır.

5. Klinik Uygulamalarda Uygulanması Beklenen İlkeler

5.1. Uygulama Seçerken

Yapay zeka uygulamalarını radyoloji pratiği için seçerken “hasta yararı” ve gereksinim belirleyici unsur olmalıdır. Bunun için bir gereksinim analizi yapılarak klinik ortamda karşılaşılan problemler ve bu problemlere yönelik “gereklilik ve en uygun YZ çözümleri saptanmalıdır. Seçilecek uygulamalar klinikte hasta yönetimini anlamlı biçimde etkileyecek ve iş akışını kolaylaştırıp üretkenliği artıracak şekilde olmalıdır. Ayrıca uygulamanın; “uzaktan erişim”, “lisans satın alım”, “kullandığın kadar öde modeli” gibi özellikleri kurum yöneticileri ve uygulayıcı radyologlar tarafından tartışılmalı ve bedel/etkinlik gözetilmelidir.

Seçilecek YZ uygulamasının geçerliliği ve etkinliği, nitelikli bilimsel araştırmalarla desteklenmiş olmalıdır. Bu araştırmalar, klinik doğruluk, hassasiyet ve güvenilirlik açısından makul ve sağlam kanıtları sunmalıdır. Uygulamalar kullanılırken güvenlik, mahremiyet ve etkinlik açısından genel ve yerel standartlar karşılanmalıdır. Seçilecek uygulama CE, FDA ya da yerel benzeri sağlık otoriteleri tarafından onaylanmış olmalıdır. Bu onaylar yardımı ile uygulamanın klinik kullanımı için gerekli yasal ve etik standartları karşıladığı garanti edilmelidir. Seçilecek YZ uygulamaları klinik pratikte entegrasyonu

hızlandıracak biçimde kolay kullanılabilir olmalıdır. Uygulama mevcut sistemlerle sorunsuz entegre olabilmeli ve klinik iş akışını bozmayacak ve kolaylaştıracak şekilde çalışmalıdır. Bu nedenle mümkünse bilgi sistemlerine ve PACS' a entegre olması tercih nedeni olmalıdır. Bu şekilde güvenli ve etkili bir şekilde kullanılması kolaylaşacaktır. Sağlık personelinin uygulamayı hızlı ve etkili bir şekilde benimsemesini sağlayacak tarzda kullanıcı dostu arayüzler bulunmalıdır.

5.2. Uygulama Kullanımında

Klinikte YZ uygulamalarının etkin ve güvenli kullanımı için bütün önlemler alınmalıdır. Bu amaçla YZ uygulamalarının potansiyel zayıflıkları, güvenlik ölçütleri bilinmeli ve bunlar ilgili taraflarla yeterince paylaşılmalıdır.

Kullanıcıların temel bilgi sistemleri ve bilişim konuları yanı sıra YZ uygulamaları ve seçilecek YZ uygulaması özelinde yeterli eğitimi almış olması sağlanmalıdır. Bu eğitim programları uygulamanın etkili ve tüm işlevlerini kullanabilecek şekilde olmalı ve bu şekilde uygulama doğru ve etkili kullanılmalıdır.

Hasta verilerinin gizliliğinin ve güvenliğinin sağlanması garanti altına alınmalıdır. Veri güvenliği açısından hasta bilgilerine yetkisiz erişimlerden korunmalıdır. Standart protokoller ile etik ve yasal zorunluluklara uyumu garanti altına alınmalıdır. Klinik uygulamalarda tutarlılık ve güvenlik sağlanmalıdır.

Yapay zeka uygulamalarının performansı sürekli izlenmesi ve değerlendirilmesi için düzenekler oluşturulmalı ve bu şekilde olası hataların ve iyileştirme alanlarının belirlenmesi sağlanmalıdır. Bu sürekli değerlendirme, uygulamanın güvenilirliğini ve etkinliğini artırma amaçlanmalıdır.

Kullanıcıların geri bildirimlerini iletebilecekleri mekanizmalar oluşturulmalıdır. Bu şekilde uygulamanın sürekli olarak geliştirilmesi amaçlanmalı ve kullanıcı deneyimleri ile karşılaşılan sorunlar raporlanarak bu amaçla kullanılmalıdır.

5.3. Uygulamalar Hakkında Hastaların Bilgilendirilmesi

Yapay zeka uygulamalarının kullanımı hakkında hastaların bilgilendirilmesi öncelikle hasta hakları ve etik değerler gözetilmek üzere yapılmalıdır. Bu amaçla; hastalara uygulamaların kullanılacağı belirtilmeli ve kullanımları konusunda açık ve bilgilendirilmiş onam alınmalıdır.

Bilgilendirilmiş onam, şeffaf bir biçimde hastaların bilinçli karar vermelerini sağlayarak uygulamanın nasıl çalıştığı, ne amaçla kullanıldığı ve hangi verilerin toplandığını açıkça içermelidir. Böylece hastaların uygulama hakkında tam bir anlayışa sahip olması sağlanmalıdır.

Yapay zeka kullanımlarında avantaj-dezavantajları ve zayıf-güçlü noktaları hakkında hastaların uygun bilgi sahibi olacağından emin olunmalıdır. Yapay zeka kullanımının etkileri abartılarak ticari bir tanıtıma dönüşmemelidir.

Uygulamanın olası riskleri ve potansiyel faydaları hastalara açıklanmalıdır. Bu açıklamalarla hastaların kolaylıkla avantajları ve dezavantajları hakkında bilgi sahibi olmaları için bütün önlemler alınmalıdır.

Hastaların hangi verilerinin nasıl kullanılacağı ve veri gizliliği hakları konusunda bilgilendirilmesi sağlanmalıdır. Verilerin nasıl korunduğu ve hangi amaçlarla kullanıldığı belirtilmelidir. Hastaların uygulama ile ilgili sorularının tatmin edici biçimde cevaplanacağı koşullar sağlanmalıdır.

Uygulama hakkında hastaların yazılı ya da sözlü bilgilendirilmesinde hastaların kültürel düzeyleri gözetilmeli ve her hastaya (ya da sorumlu yakınına) ulaşacak çözümler sağlanmalıdır.

5.4. Uygulamaların Güvenirlik Düzeyleri

Kullanılacak uygulamaların geliştirildikten sonra gerçek veri setleri ile klinik ortamda nasıl başarımlar gösterdiği konusunda tercihen birden fazla geçerlilik testlerinin yapılmış olması klinik ortamda güvenilirliklerini artırır. Bu konudaki yayımlanmış bilimsel yayınların yanı sıra uygulayıcıların kendi klinik ortamlarında yapacakları çalışmalar önemlidir. Uygulamaların düzenli kalite kontrol testlerine tabi tutularak performanslarının sürekli olarak yüksek düzeyde kalması sağlanmalıdır.

Uygulamaların genel performans ve doğruluk ölçütlerinin yanı sıra yanıt sürelerinin izlenmesi önemlidir. Yanıt sürelerinin spesifik-klinik beklentiyi karşılayacak düzeyde kısa olmalıdır. Uygulamaların değerlendirilmesinin bağımsız kurumlar tarafından yapılması önemlidir. Uygulamaların en son teknolojik yeniliklere uyumlu olacak şekilde sürekli olarak düzenli aralıklarla güncellenmesi ve iyileştirilmesi sağlanmalıdır.

5.5. Sürekli İyileştirme ve Kalite Denetimi

Tüm sistemlerde olduğu gibi YZ algoritmalarının sürekli izlemi ve iyileştirilmesi gereklidir. Bunun için kurumsal düzeyde çalışma gruplarının oluşturulması ve saptanacak parametreler ile YZ araçlarının izlenmesi ve performanslarının ölçülmesi önerilir. Saptanan sorunların düzeltici eylemlerle iyileştirmeye yardımcı olması beklenir.

5.6. Geri Ödeme

İş akışlarına eklenecek ve görüntü ile tanı kalitesini artırıcı YZ araçlarının kullanılması halinde bir geri ödemedan yararlanması sağlanmalıdır. Bu otorite kurumlarca yapılacak düzenlemelerle adil olarak gerçekleştirilmelidir.

5.7. Sorumluluklar

Yapay zeka araçlarının kullanımında doğacak sorunlardan kimin sorumlu olacağı yasal düzenlemelerle tanımlanmalıdır. Halihazırda ulusal düzeyde bir düzenleme yapılmadığından, uygulamaların geçerli olan yasalar doğrultusunda yürütülmesi gereklidir. Radyoloji özelinde değerlendirme ve raporun tek sorumlusu radyoloji uzmanıdır ve YZ araçlarından yararlanılmış ise bunun raporda belirtilmesi uygun olacaktır.

6. Yararlanılan Kaynaklar

- 1) *Britannica Sözlük*. Adres: <https://www.britannica.com>
- 2) Hsu HH, Ko KH, Chou YC, ve ark. Performance and reading time of lung nodule identification on multidetector CT with or without an artificial intelligence-powered computer-aided detection system. *Clin Radiol*. 2021;76(8):626.
- 3) Larsen M, Hoff SR, Auensen S. AI risk score on screening mammograms preceding breast cancer diagnosis. 2023;(20):19–24.
- 4) Dreizin D, Staziaki P V., Khatri GD, ve ark. Artificial intelligence CAD tools in trauma imaging: a scoping review from the American Society of Emergency Radiology (ASER) AI/ML Expert Panel. *Emerg Radiol [Internet]*. 2023;30(3):251–65.
- 5) Moawad AW, Fuentes DT, Elbanan MG, ve ark. Artificial intelligence in diagnostic radiology: where do we stand, challenges, and opportunities. *J Comput Assist Tomogr*. 2022;46(1):78–90.
- 6) Galbusera F, Cina A. Image annotation and curation in radiology: an overview for machine learning practitioners. *Eur Radiol Exp*. 2024 Feb 6;8(1):11.
- 7) *Türk Dil Kurumu Sözlüğü*. Adres: <https://sozluk.gov.tr/>
- 8) Eklund A, Dufort P, Forsberg D, LaConte SM. Medical image processing on the GPU - Past, present and future. *Med Image Anal [Internet]*. 2013;17(8):1073–94.
- 9) Walston SL, Seki H, Takita H, Mitsuyama Y, Sato S, Hagiwara A, Ito R, Hanaoka S, Miki Y, Ueda D. Data set terminology of deep learning in medicine: a historical review and recommendation. *Jpn J Radiol*. 2024 Jun 10.
- 10) Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture analysis of imaging: what radiologists need to know. *AJR Am J Roentgenol*. 2019 Mar;212(3):520-528.
- 11) Litjens G, Kooi T, Bejnordi BE, ve ark. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42(December 2012):60–88.

- 12) Chae A, Yao MS, Sagreiya H, ve ark. Strategies for implementing machine learning algorithms in the clinical practice of radiology. *Radiology*. 2024;310(1).
- 13) Park SH, Choi J, Byeon JS. Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence. *Korean J Radiol*. 2021;22(3):442–53.
- 14) Moezzi SAR, Ghaedi A, Rahmiani M, Mousavi SZ, Sami A. Application of deep learning in generating structured radiology reports: a transformer-based technique. *J Digit Imaging [Internet]*. 2023;36(1):80–90.
- 15) Akinci D'Antonoli T, Stanzione A, Bluethgen C, ve ark. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol*. 2024 Mar 6;30(2):80–90.
- 16) Özkan M, Kar G. Türkçe dilinde yazılan bilimsel metinlerin derin öğrenme tekniği uygulanarak çoklu sınıflandırılması. *Mühendislik Bilim ve Tasarım Derg*. 2022;10(2):504–19.
- 17) Karataş E, Karaca U. Yapay zeka tarafından meydana getirilen fikri ürünlerin 5846 sayılı fikir ve sanat eserleri kanununa göre korunması. *Maltepe Üniversitesi Hukuk Fakültesi Derg*. 2022;(1):17–50.
- 18) Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications. *Radiology*. 2024;310(1).
- 19) Baheti AD, Thakur MH, Jankharia B. Informed consent in diagnostic radiology practice: Where do we stand?. *Indian J Radiol Imaging* 2017;27:517–20.
- 20) Ueda D, Kakinuma T, Fujita S, ve ark. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol [Internet]*. 2024;42(1):3–15.
- 21) Park HJ. Patient perspectives on informed consent for medical AI: A web-based experiment. *Digit Heal*. 2024;10.
- 22) Rowell C, Sebro R. Who will get paid for artificial intelligence in medicine? *Radiol Artif Intell*. 2022 Aug 3;4(5):e220054.
- 23) Yapay Zeka Avrupa Komisyonu Yasası. Adres: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-act-different-rules-for-different-risk-levels-0>
- 24) YZ Avrupa Komisyonu Kararı. Adres: https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en
- 25) Amerika YZ Yasaları Yönerge Taslağı. Adres: <https://www.whitehouse.gov/ostp/ai-bill-of-rights>
- 26) Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: Practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. *J Med Imaging Radiat Oncol*. 2024 Feb;68(1):7–26.
- 27) Cambridge Sözlük Adres: <https://dictionary.cambridge.org/dictionary/english/regulator>
- 28) Najjar R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics (Basel)*. 2023 Aug 25;13(17):2760.
- 29) Graziani M, Dutkiewicz L, Calvaresi D, Pereira J. A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences [Internet]. C. 56, *Artificial Intelligence Review*. Springer Netherlands; 2023. 3473–3504 s.
- 30) Mehta N, Pandit A. International Journal of Medical Informatics Concurrence of big data analytics and healthcare : A systematic review. *Int J Med Inform [Internet]*. 2018;114(January):57–65.
- 31) Chen R, Ch I, Hatabu H, Valtchinov VI, Siegelman J. Detection of unwarranted CT radiation exposure from patient and imaging protocol meta-data using regularized regression. *Eur J Radiol Open [Internet]*. 2019;6(April):206–11.
- 32) Nobel JM, Kok EM, Robben SGF. Redefining the structure of structured reporting in radiology. *Insights Imaging*. 2020 Feb 4;11(1):10.
- 33) Datta S, Godfrey-Stovall J, Roberts K. RadLex Normalization in Radiology Reports. *AMIA . Annu Symp proceedings AMIA Symp*. 2020;2020:338–47.
- 34) Şenol U, Aktaş A, Saka O. Radyoloji Bilgi Sistemi. *Akad Bilişim'07 - IX Akad Bilişim Konf Bildir 31 Ocak - 2 Şubat 2007 Dumlupınar Üniversitesi, Kütahya*. 2007;431–3.
- 35) Dijital Hastane Adres: <https://dijitalhastane.saglik.gov.tr/TR-4877/dicom-digital-imaging-and-communications-in-medicine---tipta-dijital-goruntuleme-ve-iletisim.html>
- 36) Subaşıoğlu, F. (2024). "Elektronik sağlık kayıtlarında etik", *Elektrik Sağlık Kayıtları (11-25)*, Nobel Akademik Yayıncılık, Ankara
- 37) Tıp Bilişimi Derneği HL7 Özet Bilgi Dökümanı. Adres: <https://turkmia.net/file/HL7-ozet-bilgi.pdf>
- 38) Tang A, Tam R, Cadrin-ch A, ve ark. Canadian Association of Radiologists white paper on artificial intelligence in radiology. 2018;69:120–35.
- 39) Sutherland J, Belec J, Sheikh A, ve ark. Applying modern virtual and augmented reality technologies to medical images and models. *J Digit Imaging*. 2019 Feb;32(1):38–53.
- 40) Gupta S, Johnson EM, Peacock JG, ve ark. Radiology, Mobile Devices, and Internet of Things (IoT). *J Digit Imaging*. 2020;33(3):735–46.
- 41) Türkiye Bilişim Derneği Bilişim Terimleri Sözlüğü. Adres: <http://bilisimde.ozenliturkce.org.tr/onerilen-tum-terimler-ingilizce-turkce>
- 42) Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO.
- 43) Rowell C, Sebro R. Who will get paid for artificial intelligence in medicine? *Radiol Artif Intell*. 2022 Aug 3;4(5):e220054.

- 44) YZ Avrupa Komisyonu Yasası. Adres: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-act-different-rules-for-different-risk-levels-0>
- 45) YZ Avrupa Komisyonu Kararı. Adres: https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en
- 46) Amerika YZ Yasaları Yönerge Taslağı. Adres: <https://www.whitehouse.gov/ostp/ai-bill-of-rights>
- 47) Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: Practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. *J Med Imaging Radiat Oncol*. 2024 Feb;68(1):7-26.
- 48) Ueda D, Kakinuma T, Fujita S, ve ark. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol [Internet]*. 2024;42(1):3–15.
- 49) Tang A, Tam R, Cadrin-ch A, ve ark. Canadian Association of Radiologists white paper on artificial intelligence in radiology. 2018;69:120–35.
- 50) Yükseköğretim kurumları bilimsel araştırma ve yayın faaliyetlerinde üretken yapay zekâ kullanımına dair etik rehber. Adres: <https://www.yok.gov.tr/Documents/2024/yapay-zeka-kullanimina-dair-etik-rehber.pdf>
- 51) Koçak B. Key concepts, common pitfalls, and best practices in artificial intelligence and machine learning: focus on radiomics. *Diagn Interv Radiol*. 2022;28(5):450-462.
- 52) Define-AI Use Case Directory | American College of Radiology. Accessed March 28, 2024. <https://www.acrdsi.org/DSI-Services/Define-AI>
- 53) Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577.
- 54) Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295(1):4-15.
- 55) Sun C, Shrivastava A, Singh S, Gupta A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *Proceedings of the IEEE International Conference on Computer Vision*. 2017;2017-October:843-852.
- 56) Ranschaert ER, Morozov S, Algra PR. Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks.; 2019.
- 57) Shur JD, Doran SJ, Kumar S, et al. Radiomics in Oncology: A Practical Guide. *RadioGraphics*. 2021;41(6):1717-1732.
- 58) Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging*. 2020;20(1):33.
- 59) Alwosheel A, van Cranenburgh S, Chorus CG. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*. 2018;28:167-182.
- 60) Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? Published online November 19, 2015.
- 61) Jain AK, Chandrasekaran B. Dimensionality and sample size considerations in pattern recognition practice. In: ; 1982:835-855.
- 62) Baum EB, Haussler D. What Size Net Gives Valid Generalization? *Neural Comput*. 1989;1(1):151-160.
- 63) Haykin S. *Neural Networks and Learning Machines*. Vol 3.; 2008.
- 64) Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging*. 2013;26(6):1045-1057.
- 65) ProstateNET | Prostate Imaging Archive. Accessed March 28, 2024. <https://prostatenet.eu/>
- 66) Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *J Cogn Neurosci*. 2007;19(9):1498-1507.
- 67) Candemir S, Nguyen X V., Folio LR, Prevedello LM. Training Strategies for Radiology Deep Learning Models in Data-limited Scenarios. *Radiol Artif Intell*. 2021;3(6).
- 68) Cheng PM, Montagnon E, Yamashita R, et al. Deep learning: An update for radiologists. *Radiographics*. 2021;41(5):1427-1445.
- 69) Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9(4):611-629.
- 70) Galbusera F, Cina A. Image annotation and curation in radiology: an overview for machine learning practitioners. *Eur Radiol Exp*. 2024;8(1).
- 71) McMahan B, Ramage D. Federated Learning: Collaborative Machine Learning without Centralized Training Data. Google AI Blog. Published April 6, 2017. Accessed March 28, 2024. <https://blog.research.google/2017/04/federated-learning-collaborative.html>
- 72) Matheny M, Israni ST, Whicher D, Ahmed M. Artificial intelligence in health care: The hope, the hype, the promise, the peril. In: *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril* (2019). ; 2023.
- 73) de Hond AAH, Leeuwenberg AM, Hooft L, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med*. 2022;5(1).
- 74) Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ Digit Med*. 2022;5(1).
- 75) Diaz O, Kushibar K, Osuala R, et al. Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools. *Physica Medica*. 2021;83:25-37.
- 76) Castiglioni I, Rundo L, Codari M, et al. AI applications to medical images: From machine learning to deep learning. *Phys Med*. 2021;83:9-24.

- 77) Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging*. 2004;22(1):81-91.
- 78) Montagnon E, Cerny M, Cadrin-Chênevert A, et al. Deep learning workflow in radiology: a primer. *Insights Imaging*. 2020;11(1):22.
- 79) Aryanto KYE, Oudkerk M, van Ooijen PMA. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy. *Eur Radiol*. 2015;25(12):3685-3695.
- 80) Library D. DICOM Library - Anonymize, Share, View DICOM files ONLINE. Accessed April 28, 2024. <https://www.dicomlibrary.com/>
- 81) RSNA. CTP-The RSNA Clinical Trial Processor. Accessed April 28, 2024. https://mircwiki.rsna.org/index.php?title=MIRC_CTP
- 82) Kişisel Verilerin Korunması Kanunu. 2016;57(29677). Accessed April 27, 2024. <https://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.pdf>
- 83) Levy A, Agrawal M, Satyanarayan A, Sontag D. Assessing the impact of automated suggestions on decision making: Domain experts mediate model errors but take less initiative. In: *Conference on Human Factors in Computing Systems - Proceedings*. ; 2021.
- 84) Lakhani P, Kim W, Langlotz CP. Automated extraction of critical test values and communications from unstructured radiology reports: An analysis of 9.3 million reports from 1990 to 2011. *Radiology*. 2012;265(3).
- 85) Chen MC, Ball RL, Yang L, et al. Deep learning to classify radiology free-text reports. *Radiology*. 2018;286(3).
- 86) Brady A, Laoie RO, McCarthy P, McDermott R. Discrepancy and error in radiology: Concepts, causes and consequences. *Ulster Medical Journal*. 2012;81(1).
- 87) Heim E, Roß T, Seitel A, et al. Large-scale medical image annotation with crowd-powered algorithms. *Journal of Medical Imaging*. 2018;5(03).
- 88) Mehta P, Sandfort V, Gheysens D, Braeckvelt GJ, Berte J, Summers RM. Segmenting the kidney on CT scans via crowdsourcing. In: *Proceedings - International Symposium on Biomedical Imaging*. Vol 2019-April. ; 2019.
- 89) Wahid KA, Fuentes D. Weak Supervision, Strong Results: Achieving High Performance in Intracranial Hemorrhage Detection with Fewer Annotation Labels. *Radiol Artif Intell*. 2024;6(1).
- 90) Zhou ZH. A brief introduction to weakly supervised learning. *Natl Sci Rev*. 2018;5(1):44-53. doi:10.1093/nsr/nwx106
- 91) Boehringer AS, Sanaat A, Arabi H, Zaidi H. An active learning approach to train a deep learning algorithm for tumor segmentation from brain MR images. *Insights Imaging*. 2023;14(1).
- 92) McCague C, Ramlee S, Reinius M, et al. Introduction to radiomics for a clinical audience. *Clin Radiol*. 2023;78(2):83-98.
- 93) Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. *Korean J Radiol*. 2019;20(3).
- 94) Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1).
- 95) Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58(5).
- 96) Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40(19).
- 97) Norgeot B, Quer G, Beaulieu-Jones BK, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med*. 2020;26(9).
- 98) Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. *Insights Imaging*. 2024;15(1):16.
- 99) Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives. *Korean J Radiol*. 2019;20(7):1124.
- 100) Kocak B, Baessler B, Bakas S, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMI. *Insights Imaging*. 2023;14(1). doi:10.1186/s13244-023-01415-8
- 101) Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol Artif Intell*. 2020;2(2).
- 102) Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology*. 2018;286(3):800-809.
- 103) Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models: The C statistic. *JAMA - Journal of the American Medical Association*. 2015;314(10).
- 104) Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed.; 2019. <http://www.springer.com/series/2848>
- 105) Kocak B, Kus EA, Kilickesmez O. How to read and review papers on machine learning and artificial intelligence in radiology: a survival guide to key methodological concepts. *Eur Radiol*. 2021;31(4):1819-1830.
- 106) Han K, Song K, Choi BW. How to develop, validate, and compare clinical prediction models involving radiological parameters: Study design and statistical methods. *Korean J Radiol*. 2016;17(3):339-350.
- 107) Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med Imaging*. 2015;15(1).

- 108) Reinke A, Tizabi MD, Sudre CH, et al. Common Limitations of Image Processing Metrics: A Picture Story. Published online April 12, 2021.
- 109) Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: ACM International Conference Proceeding Series. Vol 148. ; 2006.
- 110) Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022;12(1).
- 111) Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes. 2022;15(1):210.
- 112) Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3(1):18.
- 113) Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 2015;162(1).
- 114) Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol. 2018;74(6).
- 115) Piovani D, Sokou R, Tsantes AG, Vitello AS, Bonovas S. Optimizing Clinical Decision Making with Decision Curve Analysis: Insights for Clinical Investigators. Healthcare (Switzerland). 2023;11(16).
- 116) Kurdziolek M. Explaining the hard to explain: An overview of Explainable AI (XAI) for UX. Published online 2022. Accessed April 13, 2024. https://www.youtube.com/watch?v=JzK_SBhakUQ
- 117) Reyes M, Meier R, Pereira S, et al. On the interpretability of artificial intelligence in radiology: Challenges and opportunities. Radiol Artif Intell. 2020;2(3).
- 118) de Vries BM, Zwezerijnen GJC, Burchell GL, van Velden FHP, Menke-van der Houven van Oordt CW, Boellaard R. Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review. Front Med (Lausanne). 2023;10.
- 119) Neri E, Aghakhanyan G, Zerunian M, et al. Explainable AI in radiology: a white paper of the Italian Society of Medical and Interventional Radiology. Radiologia Medica. 2023;128(6).
- 120) Groen AM, Kraan R, Amirkhan SF, Daams JG, Maas M. A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: Limited use of explainable AI? Eur J Radiol. 2022;157.

7. Hazırlayanlar

Bu rehber Mart 2024- Mart 2025 tarihleri arasında, Dr.Oğuz Dicle Başkanlığında, TRD Görüntüleme Bilişimi Çalışma Grubu üyeleri Dr.Mustafa Özmen, Dr.Utku Şenol, Dr.Fırat Atak, Dr. Nur Hürsoy ve Dr.Sinem Gezer'in katkılarıyla hazırlanmıştır.