



Derin öğrenme ile otomatik kalça kırığı tanısı

Automatic hip fracture diagnosis with deep learning

Salih Beyaz¹, Şahika Betül Yaylı², Uğur Doktor²

¹Başkent Üniversitesi, Adana Dr. Turgut Noyan Uygulama ve Araştırma Merkezi, Ortopedi ve Travmatoloji Ana Bilim Dalı, Adana
²Turkcell Teknoloji, Yapay Zeka & Dijital Analitik Çözümler, İstanbul

Röntgenin keşfinin üzerinden bir asırdan çok zaman geçmiş olmasına karşın röntgen (X-ışını) görüntüleme kırık tespitinde altın standart tanı yöntemidir. Görüntü işleme ise yapay zeka uygulamalarının en başarılı olduğu alanlardan bir tanesidir. Yaşlanan nüfusa bağlı kalça kırığı sıklığındaki artış, tanı ve sınıflandırma sistemlerine olan ilgiyi artırmaktadır. Yüksek sayıda homojen dağılımlı verinin elde edilerek, tecrübeli etiketleyiciler tarafından sınıflandırılması ve geliştirilmesi, planlanan algoritmanın gerçek hayatta başarılı olabilmesi için şarttır. Literatüre bakıldığında ise tanı koyan sistemlerin yanında sınıflandırılabilen sistemlerin de geliştirilmekte olduğu görülmektedir. Yapılan çalışmalarda sadece röntgen görüntüsü üzerinden değil, ek olarak hastaya ait kişisel sağlık kayıtlarının (radyoloji raporu, kullandığı ilaçlar vb.) tanı amaçlı kullanımına yönelik çalışmalara rastlanmaktadır. Kırık riskini değerlendiren çalışmalar ise önleyici sağlık hizmetlerinin daha etkin kullanımına yönelik gelecek vadetmektedir. Sonuç olarak yapay zeka temelli kalça kırığı tanıma sistemleri, tecrübe farklılıklarından dolayı ortaya çıkan hataların azaltılmasına katkı sunacaktır.

Anahtar sözcükler: derin öğrenme; yapay zeka; kalça kırığı; röntgen

Although more than a century has passed since the discovery of X-ray, X-ray imaging is the gold standard diagnostic method for fracture detection. Image processing is the field where artificial intelligence applications are most successful. The increase in the incidence of hip fractures due to the aging population increases the interest in diagnosis and classification systems. Obtaining a large number of homogeneously distributed data and classification by experienced experts is essential for the success of the algorithm, which is planned to be developed in real life. When we look at the literature, it is seen that besides diagnostic systems, classification systems are also being developed. In the studies conducted, there are studies not only on the X-ray image, but also on the use of the personal health records of the patient (radiology report, drugs used, etc.) for diagnostic purposes. Studies evaluating the risk of fracture, on the other hand, are promising for more effective use of preventive health services. As a result, artificial intelligence-based hip fracture recognition systems will contribute to reducing the mistake arising from experience differences.

Key words: deep learning; artificial intelligence; hip fracture; X-ray

1959 yılında Arthur Samuel^[1] yapay zekayı “programlamaya gerek olmadan bilgisayara öğrenme yeteneği verebilme kabiliyeti” olarak tanımladı. Geçen 60 yıl içinde makine öğrenmesi ile başlayan süreç derin öğrenme, insan beynini taklit eden nöral öğrenme ve evrimsel sinir ağları ile ivme kazandı. Özellikle grafik işlemcilerdeki gelişmeler ve Moore yasası uyarınca her iki yılda iki kat artan işlemci hızları, yapay zeka algoritmalarının görüntü işleme teknolojilerinde atılıma neden oldu. Birçok yenilikte olduğu gibi derin öğrenme tabanlı görüntü işleme uygulamaları önce güvenlik ve askeri alanda kullanılmaya başlamış, çokça veri kaydının tutulduğu sağlık alanında da hızla yer bulmuştur.

Wilhelm Roentgen^[2]’in 1895 yılında X ışınlarını keşfi, tıbbi görüntüleme alanında çığır açtı. Geçen sürede bilgisayarlı tomografi (BT) ve manyetik rezonans görüntüleme (MRG) gibi kas iskelet sistemi hakkında daha detaylı bilgi veren görüntüleme yöntemleri olmasına karşın röntgenin ucuz, hızlı ve kolay ulaşılabilir olmasıyla özellikle kırık tanısı koymada altın standart olma özelliğini korumaktadır. Ortopedi ve travmatoloji uzmanları, istedikleri röntgenin yorumunu yaparak tedavi planını yaparlar. Kemik kırıklarının değerlendirilmesi, sınıflandırılması ve tedavi planlarının hazırlanması çoğu zaman bir megabayt bile etmeyen bu iki boyutlu bir resim üzerinden yapılmaktadır.

İletişim / Contact: Doç. Dr. Salih Beyaz • **E-posta / E-mail:** sbeyaz@baskent.edu.tr

ORCID ID: Salih Beyaz, 0000-0002-5788-5116 • Şahika Betül Yaylı, 0000-0001-5057-8634 • Uğur Doktor, 0000-0002-0247-079X

Geliş / Received: 24 Ekim 2021 • **Kabul / Accepted:** 08 Aralık 2021

Ortalama yaşam süresinin artışına bağlı olarak yaşlanan nüfusla beraber kalça kırığı görülme sıklığı hızla artmaktadır. Otuz beş yaşından genç nüfusta kalça kırığı insidansı 2/100.000 insan yılı olurken, 85 yaş üstü kadınlarda bu oran 3032/100.000 insan yılıdır. Kalça kırıklarının %52'si 85 yaş üstü hastalarda görülürken, %90'ı 50 yaş üstü hastalarda görülmektedir.^[3] 1997 yılıyla kıyaslandığında, 2030 yılına gelindiğinde kalça kırığı görülen hasta sayısının üç katına çıkacağı tahmin edilmektedir. Bu durumun özellikle gelişmiş batı toplumları için ciddi bir halk sağlığı problemine dönüşmesi olasıdır. Kalça kırığı diğer kırık tipleri ile karşılaştırıldığında; daha ciddi bir hastalık ve eşlik eden hastalıklara bağlı ölüm oranının yüksek olması, görülme sıklığının katlanarak artacak olması ve gibi konular araştırmacıların bu konu üzerine daha çok çalışma yapmasına neden olmaktadır.^[4]

TEKNİK

Derin öğrenme, makinenin ham verilerle beslenerek örüntü tanıma için gereken örnekleri anlayabilir hale geldiği öğrenme biçimidir. Geliştirilen modelin işlevi temel olarak girdiden çıktının tahminlenmesidir. Söz konusu kalça kırığının derin öğrenme ile tespiti olduğunda algoritmayı besleyecek girdi algoritmaya sunulacak röntgen, çıktı ise bunun kırık veya değil şeklindeki yorumudur. Model eğitim sürecinde algoritma girdi ve çıktı eşleşmeleri ile beslenir; bir nevi algoritmaya hangi röntgeni nasıl yorumlayacağı öğretilir. Bu doğrultuda, modeli beslemek adına kırık (1) veya kırık değil (0) şeklinde etiketlenmiş kalça röntgenleri kullanılır.

Derin öğrenme modelinin geliştirilmesi için birincil gereksinim veridir. Çalışmalarda harcanan zamanın çok büyük kısmı veri setinin hazırlanması safhasında geçer. Veri setinin hazırlanmasında görüntülerin elde edilmesi ve etiketlenmesi aşamaları yer alır. Verinin doğru hacimde ve yeterli nitelikte olması, geliştirilecek modelin başarı oranını etkileyen en önemli faktörlerdendir.

Medikal görüntülerin depolanması için tercih edilen format Tıpta Dijital Görüntüleme ve İletişim (*Digital Imaging and Communication in Medicine, DICOM*)'dir. Bir DICOM dosyasında hem görüntülemeye ait imaj verisi hem de hastaya ve görüntülemeye dair yaş, cinsiyet ve görüntüleme tekniği gibi bilgiler bulunur. Kişisel verilerin korunması kanunu (KVKK) gereği hastaya ait ve hastanın kimliğini açığa çıkaracak verilerin depolanması ve bilimsel çalışmada kullanılması yasaktır. Bu sebeple veriler bu formatta çekilecekse mutlaka dosya içerisindeki kişisel bilgiler silinerek anonimize edilmelidir veya görüntüler sistemden, sadece röntgene ait verileri içeren, hasta bilgilerini barındırmayan .jpg ve .png gibi imaj formatlarında direkt olarak çevrilebilir. Bu şekilde

elde edilen görüntü dosyasında hastanın kişisel verileri yer almaz.

Veri toplama aşamasında normal hasta röntgenlerine ulaşmak, kırık röntgenlerine ulaşmaktan sayıca daha kolaydır. Veri seti anormal ve normal dağılımı mümkün olduğunca dengeli olmalıdır. Kırık röntgenler mümkün olduğunca tüm anomali örneklerini içermelidir. Bu sayede algoritmanın eğitim sürecinde ezberlemesinin önüne geçilecek ve bazı durumlar karşısında hatalı tahminleme eğilimlerinde bulunması engellenecektir. Veri sayısının ve çeşitliliğinin fazla olması, modelin kullanıma alındığında yeni röntgenleri yorumlamadaki başarısını artırır. Bunun için mümkün olduğunca farklı merkezde, farklı cihazlarla, farklı zamanlarda elde edilmiş görüntü kümesini çalışmaya dahil etmek önemlidir. Röntgenlerin toplanması esnasında cinsiyet ve yaş dağılımları da modelin kullanılacağı amaç doğrultusunda olabildiğince çeşitli tutulmalı, veri seti ilgili hasta nüfusun özelliklerini temsil etmelidir. Elektif şartlarda çekilen röntgenlerin yanı sıra acil şartlarında çekilmiş röntgenlerin de eğitim aşamasında sunulması; hasta duruşu ve para, fermuar gibi yabancı cisimler açısından da veri çeşitliliğini artıracaktır. Tüm bunlara dikkat edildiğinde görüntüleme protokolü, cihazlar, klinik işleyiş ve hasta karakteristikleri açısından heterojen bir veri seti elde edilir.

Veri sayısının ve çeşitliliğinin azlığı, modelde kalıpları yetersiz öğrenme ve/veya ezberleme sorununa yol açarak modelin eğitilirken ezberlediği kalite ve çeşitteki röntgenleri doğru yorumlamasına, diğer görüntülerde anlamsız çıktılar üretmesine sebep olur. Örneğin; sadece yaşlı bireylerin röntgenleri ile eğitilen bir model genç bireyler, yorumlamada aynı başarıyı göstermeyebilir. Literatürde yapay zekanın klinisyene kıyasla çok başarılı sonuçlar elde ettiğine dair çalışmaların bir kısmında, eğitimde az sayıda veri kullanıldığı ve sonuçların gerçek hayattan büyük veri setleri ile test edilemediği görülmektedir.^[5] Bu çalışmalar incelenirken modellerin ezberleme sorunuyla karşılaşmış olabileceği akılda bulundurulmalıdır.

Toplanan veriler, sonrasında etiketleme işlemine tabi tutulur. Burada uzman kişilerce güvenilir bir etiketleme yapılması esastır çünkü verilerin etiketi doğrudan modelin öğrenmesini, dolayısıyla modelin başarısını belirler. Görüntü verilerinin etiketlenmesinde, temelde iki farklı etiketleme çeşidi vardır; amaca yönelik olarak hangisinin kullanılacağı seçilir. Bunlardan birincisi ilgili bölgenin sınırlarının belirlenmesi, ikincisi ise görüntüde daha önceden belirlenmiş etiketlerin (kırık var/yok, kırık tipi gibi) seçilmesiyle yapılır. Görüntü bölüntüleme (segmentasyon) yapılacak senaryolarda birinci tip etiketleme tercih edilirken, kırık var-yok sınıflandırması yapılacağı durumlarda genellikle ikinci tip etiketlemeye başvuru-

lur. Etiketleyici sayısının birden fazla olması etiketlerin tarafsızlığı konusundaki soru işaretlerinin giderilmesini sağlar. Etiketleyicilerin değerlendirmeleri, çoğunluğun belirttiği etiket geçerli olarak belirlenir veya etiketler, etiketleyicilerin tecrübe düzeyine göre belirlenen kat-sayılaraya göre değerlendirmeye alınır. Böylece toplamda konsolide bir etiket elde edilmiş olur ve etiketlerin güvenilirliği artar.

Etiketlerin hazırlanmasında bir diğer yöntem ise elektronik hasta kayıtlarından ve görüntü saklama ve iletişim sistemleri (*Picture Archiving and Communication Systems, PACS*)’nde kaydedilen klinik açıklamalardan doğal dil işleme teknikleriyle veri madenciliği yapılmasıdır. Bu yöntem kimi durumlarda kullanışlı olabilir fakat veriyi elde etmeye yarayacak algoritmaların geliştirilmesi aşamasını da ek olarak sürece ilave eder. Bu şekilde elde edilen etiketlerin doğruluğu, sadece veri madenciliğinde kullanılan yöntemlerin başarısına bağlı değildir ve aynı zamanda tutulan kayıtların güvenilirliği ile de ilişkilidir. Elektronik hasta kayıtlarında özensiz, yanlış ve yetersiz kayıtlara sıkça rastlanır. Oakden-Rayner, etiketleri veri madenciliğiyle çıkarılmış olan Chest X-ray 14 veri setinde etiketlerin içlerinde hatalar olduğunu gösteren bir çalışma paylaşmıştır.^[6]

Etiketleme aşaması tamamlandıktan sonra elde edilen veriler üç gruba ayrılır: eğitim, doğrulama ve test. Eğitim veri seti model eğitilirken kalıpları öğrenmesi için kullanılır. Doğrulama veri seti her döngünün sonunda kullanılır ve modelin o noktaya kadar geldiği başarıyı ölçmeyi sağlar. Böylece model parametreleri burdan gelen sonuçlara göre değiştirilir ve yeni döngü eğitimi başlar. Döngü sayısı, modelin eğitim kümesindeki bütün görüntüleri kaç kez elden geçireceğini belirler ve her bir döngü sonunda doğrulama kümesi ile model ağırlıklarının ne şekilde güncellenmesi gerektiği belirlenir. Döngü sayısı eğitim başlamadan önce belirleneceği gibi, dinamik olarak da belirlenebilir.

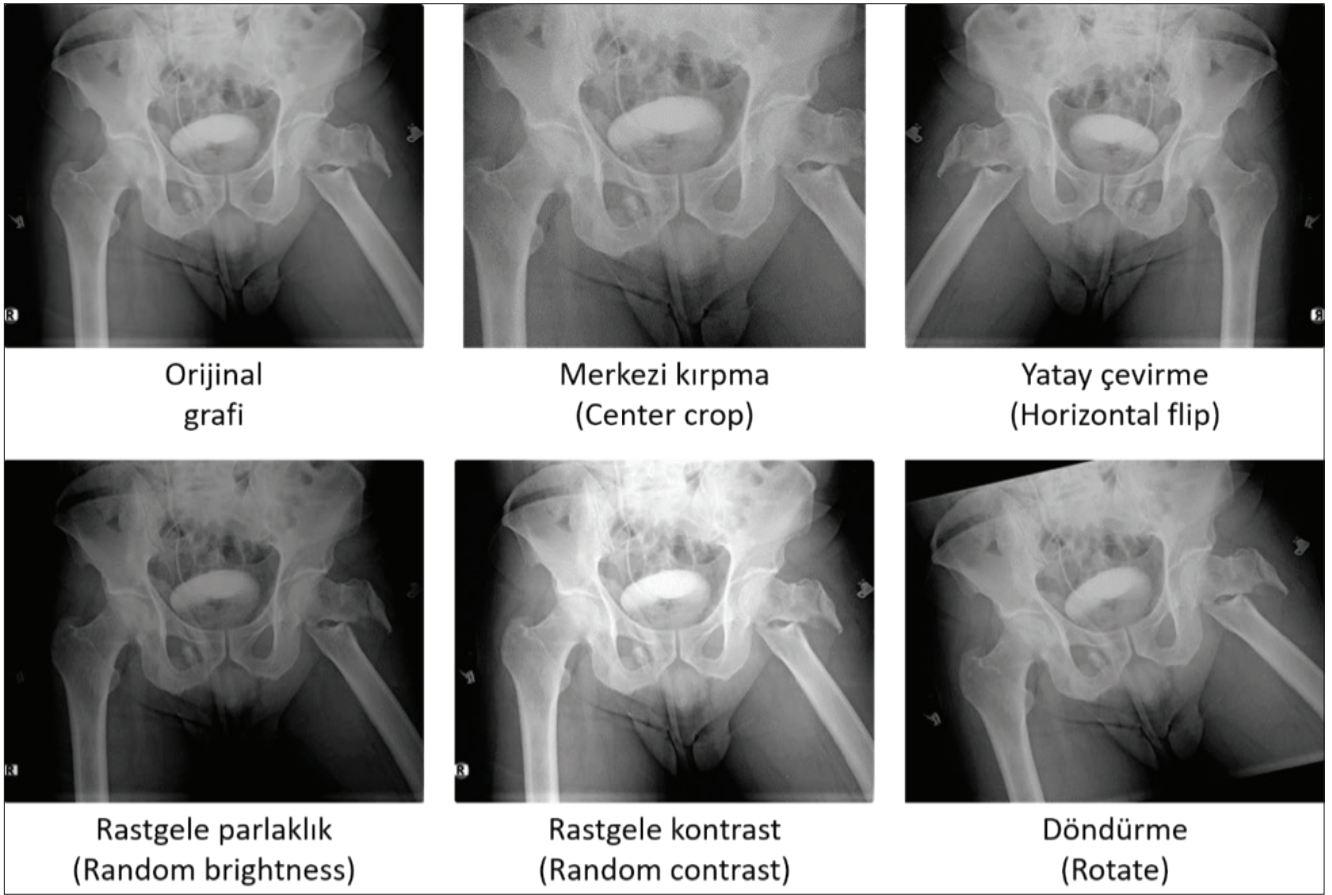
Derin öğrenme model eğitiminde evrimsel sinir ağlarından yararlanır. Evrimsel sinir ağları sıralı olarak düzenlenen ve çok sayıda doğrusal olmayan işlemlerin yapıldığı katmanlardan oluşur. Her katmanın temsili bir sonraki katmandan beslenir ve katmanlar arasında ilerledikçe görüntüdeki noktalar ve çizgiler gibi temel özelliklerden nihai hedefe, yani kırık var/yok ayrımına doğru öğrenme sağlanır. Modelin bu öğrenme sürecinin işleyişi tam olarak izah edilemediği için bu bölüm kara kutu olarak adlandırılır. Ancak bazı yöntemler kullanılarak modelin tahminlemelerini görüntü üzerindeki hangi örüntülere göre yaptığı yorumlanabilir.

Model, tamamen eldeki verilerle sıfırdan eğitilebileceği gibi *transfer learning* yaklaşımı da kullanılabilir.

Bu yöntemde direkt çözülmek istenen problemle ilgili olması gerekmeksizin daha önceden milyonlarca veri ile eğitilmiş modeller kullanılabilir. Bu modellerin yalnızca bazı katmanları kalça verileri ile eğitmek üzere yeniden dizayn edilebilir ve model eğitim süreci sıfır noktadan (rastgele belirlenmiş ağırlıklar) başlamak yerine bazı temel, orta ve üst düzey örüntüleri başarılı olarak çıkartabildiği noktadan başlayabilir çünkü daha önce bahsettiğimiz gibi model eğitim aşamasında ilk katmanlarda öğrenilen özellikler, çözülmek istenen problemden bağımsız, görüntüye ait temel bileşenlerdir. Model genellikle son katmanlara yaklaştıkça ilgili probleme özgül komplike örüntüleri yakalamaya odaklanır.

Eğitim aşamasında verilerin daha da çeşitlendirilmesi için mevcut verilerden yeni röntgenler türetilir. Büyütme veya artırma (*augmentation*) adı verilen bu aşamada yapılan işlemler görüntülerin simetriğini alma, döndürme, kırpma, görüntü ekleme, kontrast ve parlaklık değerlerini değiştirme ve daha ileri transformasyon tekniklerdir (Şekil 1). Bu türetme işlemleri model eğitimi esnasında, belleğe görüntüler kaydedilmeden yapılabileceği gibi veriler modele sunulmadan önce transforme edilen hallerini kaydedip sonra modele sunma şeklinde de ilerlenebilir. İşlemci kaynağı yeterli ise özellikle büyük veri setlerinde ilk yöntem tercih edilir. Böylece her döngüde yeni veriler türetilir ve eğitim bu türetilen veriler ile yapılır. Toplam çeşitlilik transformasyon için tercih edilen tekniklerin kombinasyonlarını içerir. Küçük veri seti ve yetersiz işlemci durumlarda ikinci yöntem tercih edilebilir. Bu yöntemde görüntüler bellekte kaydedildiği için bellek ihtiyacı vardır. Her iki yöntem için de kullanılan verinin versiyonlanması, tekrar edilebilir eğitim süreçleri için kritiktir. Artırma teknikleri ile veri seti yüzlerce kat büyütülebilir fakat tamamen bağımsız, tekil örneklerden oluşan aynı sayıda veri içeren veri seti ile karşılaştırıldığında genellikle daha niteliksizdir. Çünkü her yeni veri yeni bir bilgi barındırır ve büyütme/artırma ile kırık paterininin veri setinde olmayan bir versiyonu elde edilemez, bunu sağlamak ancak bu verilerin veri setinde olmasıyla mümkündür.

Eğitim süreci bittiğinde elde edilen modelin başarısını test etmek için ayrılan test verilerinden yararlanır, böylece modelin daha önce hiç görmediği verileri nasıl yorumladığı gözlemlenir. Modelin başarısı bu test veri setinden elde edilen verilerle değerlendirilir, modelin ezberleme veya yetersiz öğrenme yapıp yapmadığı gözlemlenir. Değerlendirme için doğruluk, kesinlik, duyarlılık ve F1-puanı gibi metriklerden yararlanır. Değerlendirme yapılırken doğruluk değerinin veri setinin kırık-kırık değil dağılımından da etkilenebileceği göz önünde bulundurulmalıdır. Örneğin; %95’i normal, %5’i kırık veriden oluşan bir veri setinde tüm röntgenlere normal yoru-



Şekil 1. Aynı röntgenin çeşitli büyütme/artırma teknikleri ile dönüşüme uğratılmış versiyonları.

mu yapan bir modelin doğruluk değeri %95 olacaktır fakat bu değer gerçek başarıyı işaret edemez. Kesinlik ve duyarlılık değerlerinin harmonik ortalaması olan F1-puanı bu gibi durumlarda çok daha güvenilirdir. Hata matrisi, yalancı pozitif/negatif değerlerin dağılımını daha rahat incelemek için kullanılır. ROC eğrisi (*receiver operator characteristics curve*) ve buradan elde edilen AUC'de (*Area Under the Curve*) model başarısını karşılaştırmaya yardımcı olur (Şekil 2).

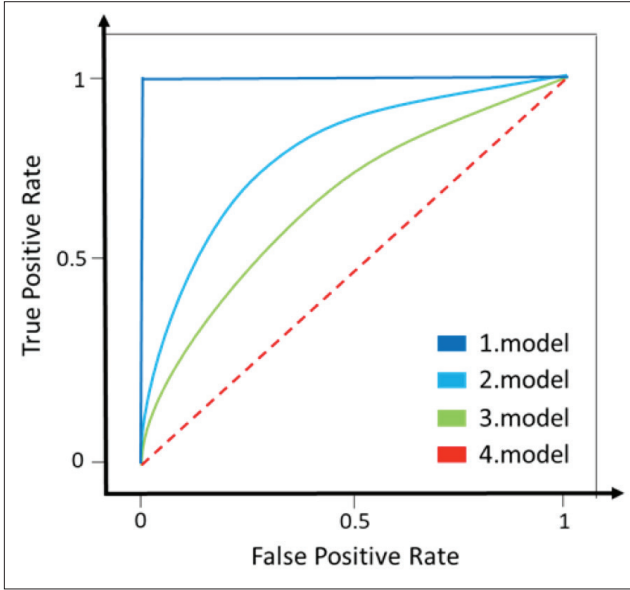
Modeli değerlendirmek için kullanılacak bir yöntem de Grad-CAM (*Gradient-weighted Class Activation Mapping*)'dir (Şekil 3). Çıkarılan ısı haritası, modelin röntgenin hangi bölümlerinin nihai kararda belirleyici olduğunu gösterir. Modelin doğru sınıflandırma yapıp yapmadığını gözlemlenmek ve yanlış sınıflandırılan görüntülerin analizini yapmak amacıyla bu yönetime başvurulur.

TARTIŞMA

Sato ve ark., tek taraflı proksimal femur kırığı olan 4851 hastanın 5242 AP Pelvis röntgenini ikiye bölerek toplam 10.484 hastanın röntgenini çalışmaya almışlardır.^[7] Makine

öğrenmesi ve evrimsel sinir ağlarını kombine ettikleri çalışmada *framework* olarak *Pytorch1.3* ve *Fast ai* kullanmışlardır. *EfficientNet-B4* önceden eğitilmiş *ImageNet* modelini uygulamışlardır. Çalışmalarında güvenilirlik, kesinlik, duyarlılık, F1 ve AUC değerleri sırasıyla 0,961; 0,952; 0,969; 0,961 ve 0,99 olarak bulmuşlar ve *Grad-Cam* görüntüleri ile algoritmanın doğru yere odaklanarak tanı koyduğunu saptamışlardır.^[7]

Ortopedi ve travmatoloji uzmanları kalça kırığı tanısı koyarken sadece radyolojik görüntüden yararlanamazlar. Hastaya ait bazı özellikler (yaş, cinsiyet, düşme hikayesi gibi bilgiler) doktorun tanı koymasında yardımcıdır. Bagledey^[8] ve ark. 17.587 kalça röntgeni ile eğittikleri derin öğrenme modeline ek olarak NLP metodu yardımıyla hastaya ait beş değişkeni [yaş, cinsiyet, vücut kitle indeksi (*body mass index*, BMI), ağrı, düşme hikayesi] ve 14 hastane işlemini de (çekimi yapan teknisyen, hastanın başvuru saati, radyoloji raporundaki kırık ifadesi, hasta sıra işlemleri vb.) kayıt altına almıştır. Sonuçta elde ettikleri bilgisayar yardımıyla kırık tanıma sisteminde sadece görüntülerin yorumlanması ile elde edilen başarının (AUC= 0,78), hasta verilerini ve hastane sürecinin



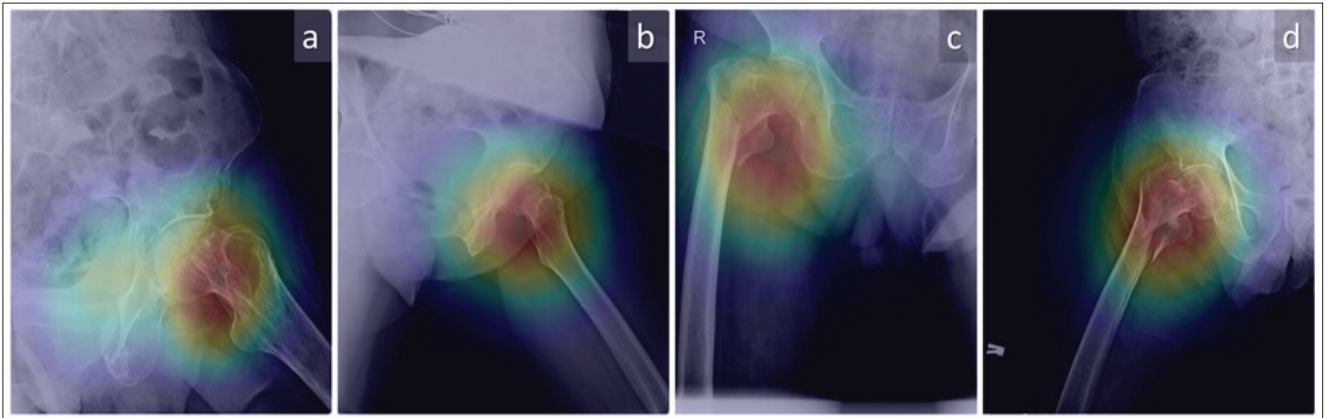
Şekil 2. ROC eğrisi bir olasılık eğrisidir. Modelin *false positive* ve *true positive* arasındaki dengesini gösterir. ROC eğrisi ne kadar sol üst köşeye yakınsa modelin yorumlama kapasitesi o kadar iyidir. Birinci çizgi mükemmel yorumlayan modele aittir. Bu model, tüm yorumlamalarını *true positive* olarak yapmaktadır. Dördüncü çizgi ise random yorumlayan modele aittir. Yani model rastgele sonuç üretiyor; ürettiği sonuçlarda doğru ve yanlış yüzdesi eşit demektir. Birinci çizgiden dördüncü çizgiye doğru gidildikçe modellerin başarısı düşmektedir. AUC ise eğrinin altında kalan alanı temsil eder. Birinci modelin AUC değeri 1; dördüncü modelinki ise 0,5'tir. AUC azaldıkça modelin güvenilirliği azalır.

işlenmesiyle elde edilen başarıyla (AUC= 0,86) kombine ettiklerinde sonuçların daha iyi (AUC= 0,91) olduğu sonucuna ulaşmışlardır. Bu çalışma kalça kırığı tanısı için grafi ile birlikte hasta ve hastane verilerinin kullanılmasının başarı yüzdesini artıracaklarını ispatlamaktadır.^[8]

Engels ve ark., kalça kırığı geçirme ihtimalini ön görme-yeye yönelik bir çalışma planlamışlardır.^[9] Çalışmalarında

65 yaş üstü 288.086 bireyin yaş, cinsiyet, daha önce geçirilmiş kırık öyküsü, kullandığı ilaçlar gibi değişkenleri kaydederek geliştirdikleri ve süper öğrenci olarak adlandırdıkları algoritmayı kendi veri setleri ile test etmiştir. Eldeki veri ile geliştirilen algoritmalarda bilinen klasik yolu (%80 eğitim, %20 test) kullanmışlardır. Hastaların %3'ünün takipte kalça kırığı geçirdiğini, ilaca bağlı risk faktörleri içerisinde ilk sırayı %19 proton pompa inhibitörleri alırken onu sırasıyla %14,9 ile antidepresanlar, %12,8 ile anti-diyabetikler, %11,6 ile oftalmik ajanlar ve %10,3 ile tıkaçıcı havayolu hastalıkları için kullanılan ilaçlarının geldiğini saptamışlardır. Geliştirdikleri algoritmanın kalça kırığı öngörüsü için kullanılabilmesi için umut ortaya koyduğunu bildirmişlerdir.^[9]

Kırığın saptanmasının yanı sıra planlamasında ve tedavinin planlanmasında da önemli rol oynar. Mutasa ve ark. femur boyun kırığının saptanmasının yanı sıra Garden sınıflandırma sistemi üzerinde çalışmışlardır.^[10] Yaptıkları çalışmada 550 hastanın, 127 Garden I/II, 610 Garden III/IV ve 326 normal kalça röntgenini kullanmışlardır. Veri eğitim setini artırmak için çekişmeli üretici ağlar (*Generative Adversarial Networks, GAN*) tekniği ve ameliyat öncesi çekilmiş kalça BT rekonstrüksiyonlarından faydalanmışlardır. Elde ettikleri gerçek ve oluşturulmuş 9063 röntgen ile derin sinir ağı tasarladılar. Test verilerinin sonucunda kırık saptama (AUC= 0,92) güvenilirliği 0,923, duyarlılık değerini 0,91 özgüllük değerini ise 0,93 olarak yayınladılar. Üç tahminli yani Garden I/II, Garden III/IV ve kırık yok olarak sınıflandırdıklarında ise (AUC= 0,96) güvenilirliğin 0,86, kesinlik değerinin 0,79, duyarlılık değerinin ise 0,90 olduğunu bildirmişlerdir. Hiçbir veri seti çoğaltma tekniği kullanmadan kırık varlığını saptamada AUC değeri 0,80 olarak bulunurken, sadece BT rekonstrüksiyonlarının eklenmesiyle bu değer 0,91, sadece GAN tekniği kullanılmasında ise 0,87'ye çıktığını



Şekil 3.a-d. Kalça röntgeni için eğitilen bir modelin dört farklı kırık röntgenini değerlendirirken odaklandığı alanlar *Grad-CAM* metodu ile görselleştirilmiştir. Görüntüler incelendiğinde modelin karar vermede odaklandığı alanın kırık olan alanla örtüştüğü, modelin dikkate alınabilecek bir değerlendirme yaptığı anlaşılmaktadır.

Tablo 1. Lee ve ark.'nın yaptıkları çalışmada elde ettikleri değerler^[12]

Model	Ölçüm	2 Sınıf	3 Sınıf	7 Sınıf
Temel Ağ (<i>Inception V3</i>)	Toplam güvenilirlik	0,793	0,7313	0,6608
	Ortalama F1 puanı	0,792	0,717	0,458
M1	Toplam güvenilirlik	0,8502	0,7974	0,696
	Ortalama F1 puanı	0,845	0,791	0,493
M2	Toplam güvenilirlik	0,8678	0,8238	0,7489
	Ortalama F1 puanı	0,867	0,817	0,501

ortaya koymuşlardır. Bu çalışmanın sonucunda elde edilen kısıtlı veri setlerinin artırılmasında BT rekonstrüksiyonları ve GAN tekniği kullanılmasının daha doğru daha yüksek güvenilirliğe sahip algoritma tanımlanmasında kullanılabilirliğini göstermişlerdir.^[10] Bu çalışmada bir diğer önemli noktada sınıflama çalışmalarında birbirine yakın sınıfların birleştirilerek bir grup oluşturulmasıdır (Garden I/II ve III/IV gibi). Çünkü tedavinin planlanmasında asılan kriter kırığın deplasmanıdır. Garden I/II kırıklarda tedavi yöntemi daha çok osteosentez olurken, Garden III/IV kırıklarda tedavi daha çok artroplasti yönündedir.

Kalça ağrısı ile acil servise başvuran hastaya çekilen AP pelvis röntgeni oldukça komplike bir görüntüdür. Bu görüntü içerisinde proksimal femur, pelvik kemikler, sakrum ve lomber vertebra kemikleri de yer almaktadır. Acil servise düşme sonrası yürüyememe şikayetiyle başvuran hastanın röntgeninde ortopedi ve travmatoloji uzmanının öncelikle odaklandığı lokalizasyon proksimal femurdur. Bu yaklaşımı göz önünde bulunduran Jimenez-Sanchez ve ark. günlük pratiğe benzer şekilde ön-arka pelvis röntgeninde proksimal femuru işaretlemeyi önermişlerdir.^[11] Bu yöntem ROI (*Region of interest*) olarak adlandırılmaktadır. Algoritmanın görüntünün tamamından ziyade istenilen noktaya odaklanmasını amaçlamaktadır. Araştırmacılar bu tezlerini ispat etmek için 1347 kalça kırığı röntgeninde proksimal femuru diktörtgen şeklinde işaretleyerek eğittikleri algoritma ile test datalarında %9'luk bir iyileşme bulduklarını yayınlamışlardır.^[11] Bu yaklaşımın diğer medikal görüntüleme tekniklerinde de uygulanabileceğini belirtmişlerdir. Röntgende aynı anda her noktayı değerlendirmek yerine kırık olasılığı daha yüksek olan alanı değerlendirmek zamandan ve doğruluk oranında kazandırabilir.

Gelişmiş ülkelerin bir çoğunda röntgen görüntüleri radyologlar tarafından değerlendirilmektedir. Klasik derin öğrenme yöntemlerinin yanı sıra meta-öğrenme (*meta-learning*) tekniğinin kullanıldığı bir çalışmayı Lee ve ark. yürütmüştür.^[12] Bu çalışmada kırığın saptanması ve sınıf-

landırılması amacıyla, radyolojik görüntülere ek olarak eğitim aşamasında radyoloji raporlarını da girdi olarak alan iki farklı derin öğrenme mimarisi kullanmışlardır. M1 ve M2 olarak adlandırdıkları mimarilerin her birinin bir kodlayıcı-kod çözücü (*encoder-decoder*) mimarisi vardır. Amaç, kodlayıcı ile röntgeni sıkıştırdıktan sonra kod çözücü aşamasında bu bilgiye metinden gelen bilgileri de ekleyerek sınıflandırma sonucuna gitmektir. Çalışmaya 459'u radyoloji raporu bulunan röntgen olmak üzere toplamda 786 pelvis röntgenini dahil etmişlerdir. Algoritmanın eğitimi için raporu olan 459 röntgen, doğrulama için 100 ve test için ise 227 röntgen kullanmışlardır. Raporlarda İngilizce olmayan ve yazım hataları bulunduran kısımlar bir ortopedi ve travmatoloji uzmanı tarafından çıkarıldıktan sonra *Global Vectors for Word Representation* (GloVe) yöntemiyle radyoloji raporundaki kelimelerin vektörel temsilleri elde edilmiştir ve eğitimde girdi olarak kullanılmıştır. Sonuçta iki mimaride eğitilen modeller ve raporlar olmaksızın eğitilen, iki mimaride de kod çözücü için kullanılan *GoogLeNet-inception V3* modeli (bazal model) sonuçlarını karşılaştırmışlardır. Çalışma sonuçları üç düzeyde değerlendirmiştir (Tablo 1). Bunlardan birincisi kırık-normal değerlendirmesidir. İkinci değerlendirme düzeyinde ise normal röntgenlere karşılık kırığın saptandığı yerin boyunda (A) veya trokanterik (B) bölgede olmasına göre değişkenlik gösteriyordu. Subtrokanterik kırıkların (C) sayısı yetersiz olduğu için çalışmaya dahil edilmemiştir. Üçüncü değerlendirme düzeyinde ise normal röntgenlerin yanında AO/OTA sınıflandırması baz alınarak altı alt grubu (A1, A2, A3, B1, B2, B3) tanımlaması beklenmiştir. Bu çalışmada algoritmanın kırığı tanımanın yanı sıra sınıflandırması test edilmiştir.^[12] Geliştirdikleri M2 mimarinin istenilen alana odaklandığını Grad-CAM görüntüleri ile göstermişlerdir. Sonuç olarak Lee ve ark. görüntü değerlendirmesine radyoloji raporlarından destek alabilen algoritmaların kırığı tanıma ve sınıflandırma konusunda gelecek vadettiğini göstermişlerdir.^[12]

Ülkemizden çıkan konu ile ilgili ilk yayının 2016 yılında Bayram ve Çakıroğlu^[13] tarafından yapıldığı görülmekte-

dir. Geliştirmiş oldukları ve femur shaft (cisim) kırığı sınıflandırma sistemi (*Diaphyseal Femur Fracture Classifier System, DIFFRACT*) adını verdikleri algortima ile röntgen görüntülerinden femur diyafiz kırıklarını, AO Müller kırık sınıflandırma sistemi üzerinde çalışmışlardır. 0,899 başarı oranı ile kırık sınıflandırması yapabilen sistemlerinin günlük pratikte doktorlara yardımcı olabileceğini belirtmişlerdir. Beyaz ve ark. ise femur boyun kırığının tanısı amacıyla derin öğrenme tekniklerini, genetik algoritma ile kombine ettiği bir çalışma yayınlamışlardır.^[14] Altmış beş hastanın 234 ön-arka pelvis röntgenini büyütme/ artırma yöntemleri kullanarak veri setini 2106'ya çıkarmışlardır. Kesinlik ve duyarlılık değerlerini sırayla 0,83 ve 0,73 olarak bulurken eğitimde genetik algoritmalar ile yaptıkları çalışmada, derin öğrenme tekniklerine genetik algoritmaların eklenmesiyle oranlarda %1,6'lık bir iyileşme saptamışlardır.^[14] Yine aynı çalışmada görüntülerin farklı çözünürlükte test edilmesi sonucunda daha düşük çözünürlükte daha başarılı sonuçlar elde edildiğini belirtmişlerdir.

Açıcı ve ark.'nın yaptıkları tekniğe yönelik çalışmada röntgen görüntülerindeki femur boyun kırıklarını tespit etmek için kanonik makine öğrenme ve derin öğrenme yaklaşımlarını karşılaştırmışlardır.^[15] En yakın komşu (*K-Nearest Neighbors, kNN*), NB (*Naïve Bayes*), Destek Vektör Makineleri (*Support vector machine, SVM*) ve Rastgele Orman (*Random Forest, RF*) olmak üzere dört kanonik makine öğrenme yöntemi ile CNN, uzun kısa süreli hafıza (*Long short-term memory, LSTM*) ve iki yönlü uzun kısa süreli hafıza (*Bidirectional Long Short Term Memory, BiLSTM*) olmak üzere üç derin öğrenme yöntemi, dengeli ve dengesiz veri kümeleri üzerinde uygulanmıştır. Dengesiz veri kümesinde 149 adet femur boyun kırığı içeren görüntü mevcutken 85 adet görüntüde kırık bulunmamaktadır. Veri çoğaltma yöntemleri kullanılarak dengesiz veri kümesinin boyutu, 1341 adet femur boyun kırığı içeren görüntü ve 765 adet kırık içermeyen görüntü olmak üzere toplamda 2106 görüntüye çıkarılmıştır. Dengeli veri kümesinde ise toplamda 2682 adet görüntü bulunmaktadır. Kanonik sınıflandırıcıları beslemek için görüntülerden; geleneksel, yerel ikili örüntüler (*Local Binary Patterns, LBP*) ve yönlü gradyanların histogramı (*Histogram of Oriented Gradient, HOG*) olmak üzere üç farklı öznitelik grubu manuel olarak elde edilmiştir. Derin öğrenme mimarilerinde ise öznitelikler, evrimsel katmanlardan otomatik olarak çıkarılmıştır. CNN mimarisinde evrimsel katmanlarda kullanılacak evrimsel filtre boyutunu ve oluşturulacak öznitelik haritası sayısını optimize etmek için Genetik Algoritma (GA) ve Parçacık Sürüsü Optimizasyonu (PSO) üst-sezgisel optimizasyon algoritmaları kullanılmıştır. Farklı deneysel

kurulumlar için elde edilen sonuçlara göre, derin öğrenme yöntemleri kanonik makine öğrenme yöntemlerine üstün gelmiştir. GA ve PSO algoritmalarının CNN mimarisine dahil edilmesi başarımlar ölçütlerine göre elde edilen sonuçlarda iyileştirme sağlamıştır. Kappa başarımlar ölçütü üzerinden en yüksek değerlere, dengesiz veri kümesinde iki katmanlı LSTM mimarisi ile ulaşılırken, dengeli veri kümesinde ise nüfus büyüklüğü ve iterasyon/jenerasyon sayısı 10 olan PSO algoritmasını kullanan CNN mimarisi ile ulaşılmıştır.^[15]

SONUÇ

Derin öğrenme uygulamaları radyolojik görüntülerden ve hasta verilerinden kalça kırığı tanıma ve sınıflandırma konusunda teorik anlamda oldukça başarılı sonuçlar elde etmektedir. Bu çalışmaların temel amacı gelecekte hastaya ait bireysel değişkenlerin de analiz edilerek, her hasta için en iyi tedavi planının yapılabilmesine olanak sağlamaktadır. Ortopedi ve travmatoloji uzmanlarının kendi içerisinde bile yaklaşım farklılıkları olduğu düşünüldüğünde bu elbette kolay olmayacaktır. Ancak sınırsız bellek kapasiteleri ve saniyede trilyonlarca işlem yapma kapasitesine sahip bilgisayarın sürekli gelişmesi, 20 yıl önce bilim kurgu filmlerine konu olabilecek projelerin yakın gelecekte hayatımıza entegre olacağı görülmektedir. Öncelikle sağlık profesyonellerine destek amacıyla geliştirilen sistemler, günlük pratikte tecrübe farklılıklarından kaynaklanan sorunların ortadan kaldırılmasına büyük katkı sunacaktır.

Makalede kullanılan kavramların açıklamaları:

True positive (TP): Gerçek pozitif

True negative (TN): Gerçek negatif

False positive (FP): Yalancı pozitif

False negative (FN): Yalancı negatif

Confusion Matrix: TP, TN, FP, FN değerlerini barındıran matrisi görselleştirmesi

Accuracy: Doğruluk

Recall: Kesinlik

Precision: Duyarlılık

F1-puanı: Kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır. Testin güvenilirliğini gösterir.

Gradient-weighted Class Activation Mapping (Grad-CAM): Eğitim tamamlandıktan ve parametreler sabitlendikten sonra modelin odak alanlarını görselleştirme için oluşturulan ısı haritası biçimidir.

Convolutional Neural Network (CNN): Evrimsel Sinir Ağı

Receiving Operating Characteristic (ROC) eğrisi: Makine öğrenmesi ile dengesiz veri setlerinin performans ölçü birimi

Area Under the Curve (AUC): ROC eğrisi altında kalan alan

Region of Interest (ROI): İlgi alanı, resim içerisindeki odaklanılan bölge

Convolutional Neural Network (CNN): Evrimsel sinir ağıları

Digital Imaging and Communication in Medicine (DICOM): Tıbbi görüntülerin saklanmasında kullanılan format

Natural Language Processing (NLP): Doğal dil işleme

K-Nearest Neighbors (kNN): En yakın komşu

KAYNAKLAR

- Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev 1959;3:210-29. [Crossref](#)
- Röntgen WC. Ueber eine neue art von strahlen. Annalen Der Physik 1898;300:12-7. [Crossref](#)
- Dhanwal DK, Dennison EM, Harvey NC, Cooper C. Epidemiology of hip fracture: worldwide geographic variation. Indian J Orthop 2011;45:15-22. [Crossref](#)
- Beyaz S, Yaylı SB. Ortopedi ve travmatolojide yapay zeka. JAIHS 2021;1:12-5. [Crossref](#)
- Urakawa T, Tanaka Y, Goto S. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiology 2019;48:239-44. [Crossref](#)
- Oakden-Rayner L. Exploring the Chest X-ray 14 dataset problems. 2017 <https://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/>
- Sato Y, Takegami Y, Asamoto T, Ono Y, Hidetoshi T, Goto R, et al. Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. BMC Musculoskeletal Disord 2021;22:1-10. [Crossref](#)
- Badgeley M, Zech JR, Rayner LO, Glicksberg BS, Liu M, Gale W, McConnell MV, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. NPJ Digit Med 2019;2:1-10. [Crossref](#)
- Engels A, Reber KC, Lindlbauer I, Rapp K, Büchele G, Klenk J. Osteoporotic hip fracture prediction from risk factors available in administrative claims data – a machine learning approach. PLoS One 2020;15:1-14. [Crossref](#)
- Mutasa S, Varada S, Goel A, Wong TT, Rasiej MJ. Advanced deep learning techniques applied to automated femoral neck fracture detection and classification. J Digit Imaging 2020;33:1209-17. [Crossref](#)
- Jimenez-Sanchez A, Kazi A, Albarqouni S, Kirchoff S, Sträter A, Biberthaler PD, et al. Weakly-supervised localization and classification of proximal femur fractures, 2018. arXiv:1809.10692.
- Lee C, Jang J, Lee S. Classification of femur fracture in pelvic X-ray images using meta-learned deep neural network. Sci Rep 2020;10:13694. [Crossref](#)
- Bayram F, Cakiroğlu M. Diffract: Diaphyseal femur fracture classifier system. Biocybernetics and Biomedical Engineering 2016;36:157-71. [Crossref](#)
- Beyaz S, Açıcı K, Sümer E. Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. Jt Dis Relat Surg, 2020;31:175-83. [Crossref](#)
- Açıcı K, Sümer E, Beyaz S. Comparison of different machine learning approaches to detect femoral neck fractures in X-ray images. Health Technol 2021;11:643-53. [Crossref](#)