



Yaşa Bağlı Maküla Dejenerasyonu Tespiti ve Sınıflandırmasında Otomatize ve Uzman Tasarımlı Makine Öğrenimi Modellerinin Karşılaştırmalı Analizi

Comparative Analysis of Automated vs. Expert-Designed Machine Learning Models in Age-Related Macular Degeneration Detection and Classification

© Ceren Durmaz Engin^{1,2*}, © Ufuk Beşenk^{3*}, © Denizcan Özizmiriler⁴, © Mustafa Alper Selver^{2,3}

¹İzmir Demokrasi Üniversitesi Buca Seyfi Demirsoy Eğitim ve Araştırma Hastanesi, Göz Hastalıkları Kliniği, İzmir, Türkiye

²Dokuz Eylül Üniversitesi, Sağlık Teknolojileri Geliştirici ve Hızlandırıcı Uygulama ve Araştırma Merkezi (Bioİzmir), İzmir, Türkiye

³Dokuz Eylül Üniversitesi Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Anabilim Dalı, İzmir, Türkiye

⁴Dr. Nevruz Erez Devlet Hastanesi, Göz Hastalıkları Kliniği, Iğdır, Türkiye

*Bu yazarlar çalışmaya eşit katkı sağlamıştır.

Öz

Amaç: Optik koherens tomografi (OKT) görüntülerinde yaşa bağlı maküla dejenerasyonu (YBMD) tespiti ve kuru ve yaş tip YBMD ayırt edilmesi için uzman tasarımı makine öğrenimi modelleri ile kodlama gerektirmeyen otomatik makine öğrenimi (AutoML) modellerinin etkinliğini karşılaştırmak.

Gereç ve Yöntem: Uzman modelleri, bir yapay zeka uzmanı tarafından EfficientNet V2 mimarisi kullanılarak geliştirilirken, AutoML modelleri bir oftalmolog tarafından ResNet-50 V2 ile transfer öğrenimini içeren LobeAI platformu kullanılarak oluşturulmuştur. Her iki model, normal OKT görüntülerini YBMD'den ayırt etmek ve ayrıca kuru tip YBMD ile yaş tip YBMD arasında sınıflama yapmak üzere tasarlanmıştır. Modeller, %80 eğitim ve %20 test bölünmesiyle, her tanı grubu için 500 OKT görüntüsü kullanılarak eğitilmiş ve test edilmiştir. Performans metrikleri (duyarlılık, özgüllük, doğruluk ve F1 skorları) hesaplanarak karşılaştırılmıştır.

Bulgular: Uzman tasarımı model, tüm görüntüleri sınıflamada %99,67 genel doğruluk sağlarken ve tüm ikili sınıf karşılaştırmalarında F1 skorları 0,99 veya daha yüksek olarak hesaplanmıştır. Buna karşılık, AutoML modeli %89,00 genel doğruluk sağlarken, ikili karşılaştırmalarda F1 skorları 0,86 ile 0,90 arasında değişmiştir. AutoML modelinde, kuru tip YBMD ile normal OKT'nin ayırt edilmesindeki duyarlılığın (0,85) düşük olduğu gözlenmiş olup, bu durum modelin kuru tip YBMD'yi doğru şekilde tanımakta zorluk çektiğini göstermektedir.

Sonuç: AutoML modelleri, YBMD olgularını tanımada ve sınıflamada kabul edilebilir bir performans sergilemiş olsa da, uzman tasarımı modeller önemli ölçüde daha üstün performans göstermiştir. Uzman tarafından geliştirilen modellerde kullanılan ileri sinir ağı mimarileri ve optimizasyon süreçleri, tıbbi görüntü sınıflama için yüksek hassasiyetli tanı araçlarının geliştirilmesinde uzman katkısının önemini vurgulamaktadır.

Anahtar Kelimeler: Yaşa bağlı maküla dejenerasyonu, AutoML, konvolüsyonel sinir ağları, EfficientNet V2, optik koherens tomografi

Abstract

Objectives: To compare the effectiveness of expert-designed machine learning models and code-free automated machine learning (AutoML) models in classifying optical coherence tomography (OCT) images for detecting age-related macular degeneration (AMD) and distinguishing between its dry and wet forms.

Materials and Methods: Custom models were developed by an artificial intelligence expert using the EfficientNet V2 architecture, while AutoML models were created by an ophthalmologist utilizing LobeAI with transfer learning via ResNet-50 V2. Both models were designed to differentiate normal OCT images from AMD and to also distinguish between dry and wet AMD. The models were trained and tested using an 80:20 split, with each diagnostic group containing 500 OCT images. Performance metrics, including sensitivity, specificity, accuracy, and F1 scores, were calculated and compared.

Results: The expert-designed model achieved an overall accuracy of 99.67% for classifying all images, with F1 scores of 0.99 or higher across all binary class comparisons. In contrast, the AutoML model achieved an overall accuracy of 89.00%, with F1 scores ranging from 0.86 to 0.90 in binary comparisons. Notably lower recall was observed for dry AMD vs. normal (0.85) in the AutoML model, indicating challenges in correctly identifying dry AMD.

Conclusion: While the AutoML models demonstrated acceptable performance in identifying and classifying AMD cases, the expert-designed models significantly outperformed them. The use of advanced neural network architectures and rigorous optimization in the expert-developed models underscores the continued necessity of expert involvement in the development of high-precision diagnostic tools for medical image classification.

Keywords: Age-related macular degeneration, AutoML, convolutional neural networks, EfficientNet V2, optical coherence tomography

Cite this article as: Durmaz Engin C, Beşenk U, Özizmiriler D, Selver MA. Comparative Analysis of Automated vs. Expert-Designed Machine Learning Models in Age-Related Macular Degeneration Detection and Classification. *Türk J Ophthalmol.* 2025;55:120-126

Bu çalışmanın bulguları, 2024 Türk Oftalmoloji Derneği Yaz Sempozyumu'nda serbest bildiri olarak sunulmuştur.

Yazışma Adresi/Address for Correspondence: Ceren Durmaz Engin, İzmir Demokrasi Üniversitesi Buca Seyfi Demirsoy Eğitim ve Araştırma Hastanesi, Göz Hastalıkları Kliniği; Dokuz Eylül Üniversitesi, Sağlık Teknolojileri Geliştirici ve Hızlandırıcı Uygulama ve Araştırma Merkezi (Bioİzmir), İzmir, Türkiye
E-posta: cerendurmaz@gmail.com ORCID-ID: orcid.org/0000-0001-5797-6467
Geliş Tarihi/Received: 19.12.2024 Kabul Tarihi/Accepted: 04.03.2025

DOI: 10.4274/tjo.galenos.2025.74780



Giriş

Yaşa bağlı maküla dejenerasyonu (YBMD), 55 yaş üstü bireylerde görme kaybının önemli bir nedenidir ve 2040 yılına kadar dünya çapında 288 milyon insanı etkileyeceği tahmin edilmektedir.¹ YBMD primer olarak makülayı tutarak fotoreseptörler ve retina pigment epitelinin (RPE) dejenerasyonuna yol açar.¹ Klinik olarak, YBMD iki şekilde kendini gösterir: Drusen varlığı ile karakterize kuru YBMD ve anormal kan damarı büyümesi ile ilişkili olan ve genellikle hızlı ve ciddi görme kaybına neden olan yaş veya eksüdatif YBMD.^{1,2} Optik koherens tomografi (OKT), YBMD tanısı ve izleminde kritik bir araçtır ve klinisyenlerin drusen varlığı, RPE anormallikleri ve koroidal neovasküler membranlar gibi önemli lezyonları tanımlamasına olanak sağlayan yüksek çözünürlüklü, kesitsel retina görüntüleri OKT ile elde edilebilir. OKT, anti-vasküler endotel büyüme faktörü tedavisi alan hastalarda kuru ve yaş YBMD'yi ayırt etmek, hastalık progresyonunu izlemek ve tedavi yanıtlarını değerlendirmek için özellikle değerlidir.¹

Derin öğrenme (DÖ) modelleri, özellikle ileri mühendislik ürünü evrişimli sinir ağları ("*convolutional neural networks*", CNN), tıbbi görüntülerin analizinde başarı göstermiştir. Bu DÖ modellerinin kullanıldığı birkaç çalışmada, YBMD'nin tanısı ve sınıflandırılmasında önemli başarı elde edilmiştir.^{3,4,5,6,7,8,9,10,11} Rasti ve ark.'nın³ çalışmasında olduğu gibi ilk çalışmalarda, veri setleri küçüktü ve Çok Ölçekli Evrişimsel Uzmanların Karmaşı ve AlexNet gibi daha basit modeller kullanılmış ve sınırlı veriye rağmen kayda değer sınıflandırma performansı elde edilmiştir.⁴ Veri kümelerinin boyutu ve karmaşıklığı arttıkça, ResNet ve DenseNet gibi daha karmaşık mimariler kullanılmış ve çoğunlukla doğruluk, hassasiyet ve özgüllüğü artırmak için topluluk öğrenme yaklaşımlarıyla birleştirilmiş ve çoğu zaman bu ölçütlerde %99 oranı aşılmıştır.^{7,8} Özellikle 2020'den sonra yapılan çalışmalar, longitudinal verileri işlemek için CNN ve tekrarlayan sinir ağlarını ("*recurrent neural networks*") birleştiren hibrit modellere ve özellik tanımayı geliştirmek için öz dikkat mekanizmalarının entegrasyonuna odaklanmıştır.^{9,10} [Tablo 1](#), YBMD tanısı ve sınıflandırması için DÖ modellerinin kullanıldığı önceki çalışmaları özetlemektedir. Başarılı sonuçlar elde etmelerine rağmen, inşa etmesi ciddi teknik bilgi gerektirir. Bu nedenle bu alanda teknik uzmanlığa sahip olmayan doktorlar için bir zorluk oluşturur. Yakın zamanlarda geliştirilen, kod içermeyen otomatik makine öğrenimi (AutoML) platformlarının kullanıma girmesi, tıbbi görüntü sınıflandırması için umut verici bir yaklaşım olarak ortaya çıkmış ve hekimlerin kapsamlı kodlama bilgisi olmadan modelleri geliştirebilmelerine olanak sağlamıştır.¹² AutoML sistemleri, veri ön işleme, özellik seçimi ve model optimizasyonu gibi ML ardışık işlemlerinin temel yönlerini otomatik olarak gerçekleştirir. Böylece ML'de uzman olmayanlar kişiler için başlangıçta karşılaşılan engelleri azaltır ve sağlıkta uzman kişilerin teknik karmaşıklıklardan ziyade klinik uygulamalara odaklanmasını sağlar.¹³ Farklı çalışmalarda diyabetik retinopati, retinal ven tıkanıklığı ve katarakt cerrahisi aşamaları gibi oftalmolojik hastalıklarda AutoML'nin başarısını değerlendirilmiştir.^{14,15,16}

Bildiğimiz kadarıyla, daha önce OKT görüntüleri ile YBMD'nin tanısı ve sınıflandırılmasında AutoML modellerinin performansı uzmanlar tarafından geliştirilen modellerle karşılaştırılmamıştır. Bu nedenle, bu çalışmada, hekimlerin kendi başlarına doğru ve güvenilir sınıflandırma modelleri oluşturmak için AutoML araçlarından yararlanıp yararlanamayacağını veya mühendislik alanında uzmanlığın optimal performans için çok önemli olup olmadığını araştırmak için bu iki tekniğin performansını değerlendirmeyi ve karşılaştırmayı amaçladık.

Gereç ve Yöntem

Bu çalışma Helsinki Bildirgesi ilkelerine uygun olarak yürütüldü ve Dokuz Eylül Üniversitesi Etik Kurul'dan onay alındı (protokol kodu: 2024/37-06, onay tarihi: 06.11.2024). Çalışmada açık erişimli veri kümeleri kullanıldığı, hastalarla doğrudan iletişim kurulmadığı ve hastaların tanımlanmasından kullanılabilecek verilere yer verilmediği göz önüne alındığında, hasta onamı alınması gerekmedi.

Optik Koherens Tomografi Verileri

Gholami ve ark.¹⁷ tarafından 2020'de yayımlanan Optik Koherens Tomografi Görüntü Veritabanı ve Kermany ve ark.¹⁸ tarafından 2018'de yayımlanan Kaggle veri setinde yer alan maküla merkezli spektral alan OKT görüntülerini kullandık. İki deneyimli göz hekimi (C.D.E. ve D.Ö.) bu OKT görüntülerini değerlendirdi ve kuru YBMD ve yaş YBMD olarak gruplandırdı. Yaş YBMD, klinik tanı kriterlerine uygun olarak intraretinal sıvı, subretinal sıvı ve/veya subretinal hiperreflektif materyal varlığına bakılarak tanımlandı. Kuru YBMD grubuna ise drusenli ve eksüdatyon bulgusu olmayan olgular dahil edildi. Coğrafi atrofi olguları çalışmaya dahil edilmedi. Eşzamanlı başka herhangi bir retina hastalığı olan veya görüntü nedeniyle kalitesi yetersiz olan görüntüler çalışmaya dahil edilmedi. Görüntülerin tutarlı şekilde etiketlendiğini değerlendirmek için, Cohen'in kappa katsayısı kullanılarak iki göz hekiminin değerlendirmelerinin güvenilirliği ölçüldü. Cohen kappa katsayısı 0,987 idi. Bu değer OKT görüntülerinin sınıflandırılmasında mükemmel uyum olduğunu göstermektedir. Veri kümesi model eğitimi için kullanılmadan önce tutarsızlıklar fikir birliği ile çözüldü. Her gruba toplam 500 maküla merkezli OKT görüntüsü dahil edildi. Kontrol grubu olarak, aynı veri setlerinden herhangi bir retina patolojisi olmayan 500 normal spektral domain OKT görüntüsü dahil edildi. Tüm görüntüler 900x300 piksele kırpıldı ve JPEG formatına dönüştürüldü.

Modellerin Oluşturulması

Veri kümesi bölündü, her gruptaki görüntülerin %80'i modelleri eğitmek için kullanıldı ve kalan %20'si test için ayrıldı. Daha sonra, AutoML için LobeAI kullanılarak dört farklı model geliştirildi ve test edildi. Uzman tarafından modeller Python ortamında tasarlandı. Model I, yaş YBMD'yi normal görüntülerden ayırt etmek için, Model II, kuru YBMD'yi normal görüntülerden ayırt etmek için ve Model III, yaş YBMD'yi kuru YBMD'den ayırt etmek için eğitildi. Ayrıca, karmaşık bir sınıflandırma görevinde model performansını değerlendirmek için tüm patolojik ve normal OKT görüntüleri kullanılarak

Model IV eğitildi. Her iki model türü de performanslarının adil bir şekilde karşılaştırılmasını sağlamak için aynı veri kümesi kullanılarak eğitildi ve test edildi.

Otomatik Makine Öğrenimi Modelleri

Lobe yazılımı (sürüm 0.10.1130.5) resmi websayfasından (<https://www.lobe.ai/>) indirilerek kişisel bilgisayara yüklendi. Kurulundan sonra, görüntüler programa yüklendi ve üç etiket eklendi: Yaş YBMD, kuru YBMD ve normal. Lobe yazılımı, eğitim sürecinde her görüntünün otomatik olarak beş rastgele varyasyonunu oluşturur. Parlaklık, satürasyon ve kontrastı ayarlar, renk tonunu değiştirir, rotasyon, yakınlaştırma ve gürültü azaltma gibi teknikleri kullanır. Bu nedenle, başka bir veri artırma yöntemi kullanılmamıştır. Bu yazılım, mevcut görevdeki performansı artırmak için ilgili görevlerde önceden eğitilmiş modelleri kullanan ve sınırlı bir veri kümesinde bile yüksek doğruluk sağlayan bir yöntem olan transfer öğrenmeyi kullanır. ResNet-50 V2 CNN mimarisinin, Proje Ayarları menüsündeki “doğruluk için optimize et” seçeneği seçilmiştir. Görüntü sınıflandırması ve diğer görme görevlerindeki etkinliği ile bilinen 50 katmanlı bir CNN olan ResNet-50 He ve ark.¹⁹ tarafından geliştirilmiştir. Toplu normalizasyon ve ReLU aktivasyonunu evrişimli katmanlar olarak kullanan 4 set halinde gruplandırılmış 16 rezidüel bloktan oluşur. Ana konsept, girdiyi doğrudan bloğun çıkışına aktaran ve derin ağ optimizasyonuna yardımcı olan atlama bağlantısıdır. Model, global ortalamaların havuzda toplaması ve SoftMax ile tam bağlantılı bir katman ile tamamlanmaktadır. Eğitim aşamasından sonra, model performansını artırmak için uygulamanın kendi “model optimizasyonu” özelliği kullanılmış ve daha ileri iyileştirmeler yapılmıştır. AutoML modelleri için istatistiksel analizler MedCalc web sitesi (<https://www.mdcalc.com/>) kullanılarak gerçekleştirildi. Patolojik OKT görüntülerini normal görüntülerden ayırt etmek ve spesifik YBMD tipini tanımlamak için duyarlılık, özgüllük ve doğruluk hesaplandı. Ek olarak, Model IV’ün performansını daha kapsamlı değerlendirmek için Konfüzyon Matrisi Jeneratörü aracılığıyla bir karışıklık matrisi oluşturuldu.²⁰

Uzman Tarafından Tasarlanan Modeller

Uzman tarafından tasarlanan model, bilgisayar mühendisliği ve biyomedikal teknolojiler konusunda deneyimi olan ve klinik ortamlarda görüntü işleme, makine öğrenmesi ve DÖ uygulamalarında uzmanlığa sahip bir yapay zeka uzmanı (U.B.) tarafından geliştirilmiştir. Bu modelin mimarisini tasarlamaktan, ön işleme ve artırma tekniklerini uygulamaktan, hiper parametreleri optimize etmekten ve model performansını doğrulamaktan uzman sorumluydu.

Tan ve Le²¹ tarafından geliştirilen EfficientNet modelleri, ResNet veya VGG gibi mimariyle kıyasla saniyede daha az parametre ve kayan nokta işlemi yaparak yüksek performans sergilemeyi amaçlamaktadır. Model derinliğini, genişliğini ve çözünürlüğünün en iyi şekilde dengeleyen bir yöntem olan “bileşik ölçeklendirme” yöntemini geliştirmişlerdir. Derinlik katman sayısını, genişlik katman başına kanal sayısını ve çözünürlük

yüklenen görüntü boyutunu ifade eder ve EfficientNet’in farklı boyutlarda verimli bir şekilde ölçeklenmesine olanak tanır.

Modellerimiz için, OKT görüntüleri ile eğitim ve değerlendirme süreci, veri seti istatistiklerinin hesaplanmasıyla başlamıştır. Ortalama (μ) ve standart sapma (σ) eğitim setinden türetilmiştir. Daha sonra, bu istatistiklere dayalı olarak eğitim, doğrulama ve test setleri için tanımlanan dönüşümlerle veri ön işleme ve artırma gerçekleştirilmiştir. Veri kümesi daha sonra eğitim ve doğrulama alt kümelerine bölünmüş ve DataLoaders buna göre oluşturulmuştur. Özel bir EfficientNetV2 mimarisine dayanan model sınıflandırma görevi için başlatıldı ve He başlatma kullanıldı. Kayıp fonksiyonu olarak, etiket yumuşatma çapraz entropi kullanıldı. Hiperparametre optimizasyonu, Optuna kullanılarak gerçekleştirildi. Optuna, hiperparametre alanını sistematik olarak araştırarak optimum konfigürasyonları belirlemek ve manuel ayar ihtiyacını en aza indirmek için tasarlanmış otomatik bir yazılım çerçevesidir.²² Optuna’da, öğrenme oranı (α), bırakma oranı (p) ve ağırlık azalması (λ) için arama alanları önce tanımlandı, daha sonra doğrulama hassasiyetini en yüksek düzeye çıkarılarak optimal hiperparametreler ($\theta^* = \{\alpha^*, p^*, \lambda^*\}$) belirlendi. Eğitim aşamasında, optimizör optimal öğrenme oranı (α^*) ve ağırlık azalması (λ^*) ile başlatıldı. Aşırı uyum gösterme riskini azaltmak için en iyi bırakma oranı (p^*) modele dahil edildi. Model, birçok kez eğitildi; her eğitimde, model parametreleri güncellenerek yapılan eğitim aşamasını doğrulama setindeki performansı değerlendirmek için doğrulama aşaması izledi. Doğrulama hassasiyetinde bir iyileşme gözlemlendiğinde model kontrol noktası kaydedildi. Eğitimin tamamlanmasının ardından, nihai model (M^*) için istatistiksel metrikler doğrudan Python ortamında hesaplandı. Duyarlılık, özgüllük, hassasiyet, hatırlama ve F1 skorları scikit-learn kütüphanesi ile modelin tahminleri temel alınarak hesaplandı. Ayrıca, sınıflandırma performansını görselleştirmek için Python dilinde programatik olarak bir karışıklık matrisi oluşturuldu. Uzman model parametreleri ve eğitim ayarları [Ek Tablo 1](#)’de özetlenmiştir.

Bulgular

Lobe uygulaması kullanılarak geliştirilen ML modelleri, normal ile yaş veya kuru YBMD OKT görüntüleri arasında ayırım yapabilmemiş ve sırasıyla %92,00 ve %90,00 duyarlılık ve %94,00 ve %91,00 özgüllük elde etmiştir. Buna karşılık, yaş ve kuru YBMD’yi normal OKT görüntülerinden ayırt etmek için uzman tarafından tasarlanan modellerde, sırasıyla %100,00 ve %99,00 duyarlılık ve %99,00 ve %100,00 özgüllük sağlanmıştır.

Yaş ve kuru YBMD’yi birbirinden ayırmak için tasarlanan AutoML Model III, normal OKT görüntülerinden yaş (doğruluk: %93,00) ve kuru YBMD’yi (doğruluk: %90,50) ayrı ayrı tanıyan modellere kıyasla daha düşük performans göstererek %86,00 doğruluk göstermiştir. Buna karşılık, uzman tarafından tasarlanan Model III, mükemmele yakın performansa göstererek %99,50 doğruluk sergilemiştir. Tüm modellerin performans metrikleri [Tablo 2](#)’de özetlenmiştir.

Normal OKT görüntülerine kıyasla tüm patolojik görüntüleri tanıyan AutoML Model IV, 0,88 ağırlıklı F1 skoru ile %89,00 doğruluk gösterdi. Buna karşılık, uzman tarafından tasarlanan

Model IV modeli %99,67 doğruluk ve 0,97 ağırlıklı F1 puanı elde etti. Hem AutoML hem de uzman tarafından tasarlanan Model IV için karışıklık matrisleri [Şekil 1](#)'de gösterilmiştir.

Tablo 1. 2017 ile 2024 yılları arasında OKT görüntüleri ile YBMD sınıflandırması üzerine yapılan çalışmaların özeti, veri kümelerini, modelleri, performans ölçütleri ve temel bulguları detaylı şekilde verilmiştir

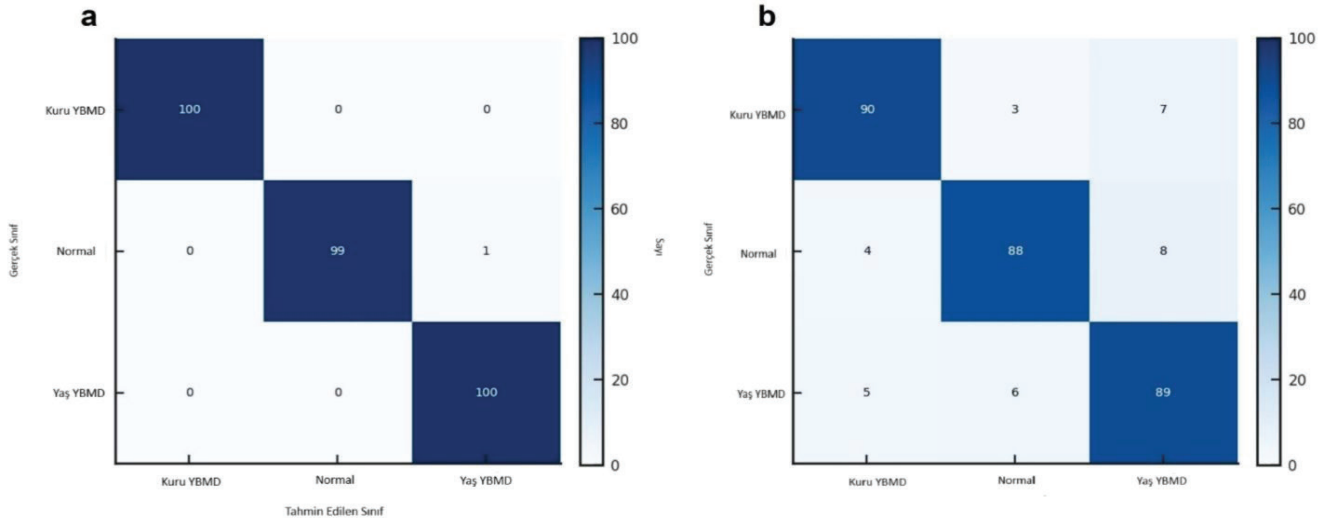
Yıl	Çalışma	Veri kümesi	Model	Performans ölçütleri	Açıklamalar
2017	[3]	148 OKT (50 normal, 48 kuru YBMD, 50 DME), 45 genel erişim	Çok ölçekli evrimsel uzman karması (ÇÖEUK)	EAA: 0,998; Hassasiyet: %98,86; Hatırlama: %99,36; F1-skoru: %99,34	Minimum ön işleme ve ÇÖEUK, geleneksel modellerden daha iyi performans göstermiştir
2018	[4]	83,484 OKT görüntüsü (sağlıklı, kuru YBMD, yaş YBMD, DMÖ)	AlexNet (tam eğitilmiş)	Doğruluk: %97,1; Duyarlılık: %99,6; Özgüllük: %98,4	AlexNet, YBMD sınıflandırması için transfer öğrenmeden daha iyi performans göstermiştir
2019	[5]	83,484 OKT görüntüsü (sağlıklı, kuru YBMD, yaş YBMD, DMÖ)	ResNet (18 katman), AlexNet	Doğruluk: %99,5; Duyarlılık: %98,0; Özgüllük: %100	ResNet, kuru ve yaş YBMD tanısında AlexNet'ten daha iyi performans göstermiştir
2019	[6]	185 normal OKT, 535 sıvı içeren YBMD, 514 sıvı içermeyen YBMD	VGG16 (transfer öğrenme)	EAA: 0,999; Doğruluk: %99,2; EAA: 0,992; Doğruluk: %95,1	Normal ile karşılaştırmak için iki aşamalı transfer öğrenme modeli Sıvı içeren YBMD ile sıvı içermeyen YBMD karşılaştırması
2020	[7]	281 eğitim hastası, 69 test hastası (longitudinal OKT)	DenseNet + RNN	EAA: 0,85 (az tedavi); EAA: 0,81 (çok tedavi); R2: 0,22	Tedavi ihtiyacını tahmin etmek için uçtan uca DÖ modeli, geleneksel modellerden daha iyi performans göstermiştir
2020	[8]	108,309 eğitim görüntüsü, 1,000 test görüntüsü	3 ResNet152 modelinde oluşan grup	Doğruluk: %98,9; Duyarlılık: %98,9; Özgüllük: %99,6	CNN grubu, ResNet152 modellerini birleştirerek daha iyi sınıflandırma yapmıştır
2022	[9]	4927 OKT görüntüsü (neovasküler YBMD, PKV, yaş olmayan YBMD)	Yığılı Autoencoder- VGG16	F1: %86,81; Doğruluk: %88,28; Hassasiyet: %86,34; Hatırlama: %87,28	Kontrastlı öğrenme ve kendine-dikkat VGG16, YBMD alt tip sınıflandırmasını iyileştirmiştir
2023	[10]	94 hastanın OKT görüntüleri, 20,482 B-taraması	ResNet + Rastgele Orman + RNN	Rastgele orman: Doğruluk: %95; RNN: Doğruluk: %71	Sıralı veri tahmini için RNN ve tedavi tahmini için özellik ekstraksiyonu
2024	[11]	167 hastanın 1285 OKT B-taraması	Açıklanabilir yapay zeka (AI) tabanlı sistem	Doğruluk: %90,82; Kappa: %89,10	Çok aşamalı AI sistemi, YBMD tanısında retina uzmanlarını taklit etmiştir

OKT: Optik koherens tomografi, YBMD: Yaşa bağlı maküla dejenerasyonu, DMÖ: Diyabetik maküla ödemi, PKV: Polipoidal koroidal vaskülopati, EAA: Eğri altındaki alan, CNN: Evrimsel sinir ağı, RNN: Tekrarlayan sinir ağı

Tablo 2. AutoML ve uzman tarafından tasarlanmış modellerin temel metrikleri

	Duyarlılık (%95 GA)		Özgüllük (%95 GA)		Hassasiyet		Hatırlama		F1 skoru	
	AutoML	Uzman	AutoML	Uzman	AutoML	Uzman	AutoML	Uzman	AutoML	Uzman
Model I: Yaş YBMD ile normal	92,00 (84,84 ila 96,48)	100,00 (96,38 ila 100,00)	94,00 (87,40 ila 97,77)	99,00 (94,55 ila 99,97)	93,88 (87,36 ila 97,14)	99,00 (94,55 ila 99,82)	92,00 (85,00 ila 95,89)	99,00 (94,45 ila 99,82)	92,93	99,00
Model II: Kuru YBMD ile normal	90,00 (82,38 ila 95,10)	99,00 (94,55 ila 99,97)	91,00 (83,60 ila 95,80)	100,00 (96,38 ila 100,00)	90,90 (83,65 ila 93,12)	100,00 (96,30 ila 100,00)	90,00 (82,56 ila 94,47)	99,00 (95,80 ila 100,00)	90,45	99,50
Model III: Yaş YBMD ile kuru YBMD	85,00 (76,40 ila 91,35)	99,00 (94,55 ila 99,97)	87,00 (78,80 ila 92,89)	100,00 (96,38 ila 100,00)	86,73 (76,72 ila 90,69)	100,00 (94,55 ila 100,00)	85,00 (75,80 ila 92,24)	99,00 (94,55 ila 99,82)	85,86	99,50

YBMD: Yaşa bağlı maküla dejenerasyonu, GA: Güven aralığı, AutoML: Otomatik makine öğrenimi



Şekil 1. Model IV'ün tüm yaşa bağlı maküla dejenerasyonu alt tiplerini ve normal kontrollerden oluşan karışıklık matrisleri. a) Uzman tarafından tasarlanan model. b) Otomatik makine öğrenimi modeli

YBMD: Yaşa bağlı maküla dejenerasyonu

Tartışma

EfficientNet V2'nin kullanıldığı uzman tarafından tasarlanan modeller ile ResNet-50 V2 ile transfer öğrenmeyi kullanan AutoML modellerinin karşılaştırılmalı analizi, performanstaki önemli farklılıkları ortaya koymuş ve uzman mühendislerin özel tıbbi görüntüleme görevlerindeki kritik rolünün altını çizmiştir. Uzman tarafından geliştirilen model, tüm OKT veritabanında YBMD'yi tespit etmek gibi en karmaşık görevde, tüm sınıflarda 0,99 veya daha yüksek F1 puanları ile %99,67'lik olağanüstü bir doğruluk oranı elde etmiştir. Bu yüksek performans seviyesi, modelin yalnızca pozitif tahminlerinde doğru olmadığını, aynı zamanda OKT görüntülerinde neredeyse tüm ilgili YBMD'leri belirlemede de etkili olduğunu göstermektedir. Minimum yanlış sınıflandırma oranları, hastalık tanısı ve tedavi planlaması için hassas sınıflandırmanın çok önemli olduğu klinik uygulamalarda modelin güvenilirliğini güçlendirmektedir. Buna karşılık, AutoML modeli, ikili sınıflandırmalarda 0,8725 ile 0,9045 arasında değişen F1 puanları ile daha düşük bir oran olan %89,00'lık bir genel doğruluk göstermiştir. YBMD'nin hatalı tanısı veya tanıda geç kalınması progresif görme kaybına yol açabileceğinden ve tedavi sonuçlarını etkileyebileceğinden, bu yetersizliğin önemli klinik sonuçları olabilir.

Uzman tarafından geliştirilen modelin daha üstün performans göstermesi birkaç temel faktöre bağlanabilir. Öncelikle, EfficientNet V2'nin gelişmiş mimarisi, ağ derinliğini, genişliğini ve çözünürlüğünü eşzamanlı olarak ölçeklendiren ve OKT görüntülerinde karmaşık desenleri daha etkili bir şekilde yakalamasını sağlayan bir bileşik ölçeklendirme yöntemi kullanmaktadır.²¹ Bu özellik, YBMD sınıfları arasındaki küçük farklılıkları ayırt etmek için çok önemlidir. EfficientNet V2'de ayrıca kanal dikkat mekanizmaları da bulunmaktadır ve bu sayede model daha az alakalı olan kanalları baskılayarak en

bilgilendirici kanallara seçici olarak odaklanır. Bu stratejik dikkat, model doğruluğunun artmasına katkıda bulunarak özellik temsilini iyileştirir. Çalışmamıza benzer şekilde Kansal ve ark.²³ tarafından yakın zamanda yapılan bir çalışmada, bu iki CNN algoritması olan ResNet-50 ve EfficientNetB0'ın koronavirüs hastalığı-2019 (COVID-19), bakteriyel ve viral pnömoni ve normal olgulardan oluşan akciğer röntgeni veri setleri üzerindeki performansı değerlendirilmiştir. Sonuçlara göre EfficientNetB0 performansının daha üstün olduğu görülmüş ve 0,98 eğitim doğruluğu ve 0,99 test doğruluğu elde edilmiştir. Buna karşılık, ResNet-50 modeli ile eğitim doğruluğu 0,83 ve test doğruluğu ise 0,96 bulunmuştur. Bu sonuçlar EfficientNetB0'ın bu bağlamda daha etkin olduğuna işaret etmektedir. De La Fuente ve ark.²⁴ tarafından yapılan bir çalışmada, sentetik görüntülerle desteklenmiş özofagogastroduodenoskopi görüntülerinden oluşan küçük, dengesiz bir veri seti üç kategoride sınıflandırılmış ve iki DÖ modelinin sınıflandırma performansı değerlendirilmiştir. Sonuçlar, EfficientNet V2 ile elde edilen genel doğruluk oranının %92,19 ve genel F1 skorunun ise 91,03 olduğunu göstermiştir. Buna karşılık, ResNet-50 algoritması ile elde edilen genel doğruluk oranı %89,49 ve genel F1 skoru 88,74 olmuştur. Bu bulgular, görüntü sınıflandırma görevlerinde EfficientNet'in daha üstün bir performans gösterdiğine işaret etmektedir.

Özelleştirilmiş ön işleme ve veri artırma, bir modelin ML çalışmalarında genelleme yapma yeteneğini önemli ölçüde artırır. Lobe uygulaması her görüntünün beş rastgele varyasyonunu oluştururken, uzman modeller daha hedefli ve çeşitli artırma tekniklerinden yararlanabilir. Ayrıca çalışmamızda Optuna aracılığıyla öğrenme oranı, bırakma oranı ve ağırlık azalması gibi parametreler doğrulama hassasiyetini en üst düzeye çıkarmak için titizlikle ayarlanmıştır. Bu da uzman tarafından geliştirilen modellerde genelleme ve model performansı iyileştirilmiştir.²² Bu bulgu, akciğer bilgisayarlı tomografi görüntülerinden

COVID-19 pnömonisi tanısında CNN modelinin duyarlılığını 0,94'ten 0,97'ye ve doğruluğunu 0,87'den 0,88'e çıkardığını gösteren Lacerda ve ark.²⁵ çalışmasında olduğu gibi Optuna yazılım çerçevesi ile hiperparametre optimizasyonunu yapılmasını destekler niteliktedir. Buna karşılık, LobeAI model optimizasyonu için belirli hiperparametrelerde ince ayar yapılmasına izin vermeden standart ayarlar yapılmasına izin vermektedir. Bu da karmaşık sınıflandırma görevlerinde performansı sınırlamış olabilir.

Hem AutoML hem de uzman tarafından geliştirilen modellerin yaş YBMD'yi normal olgulardan ayırmada kuru YBMD'yi normal olgulardan ayırmaya kıyasla daha başarılı olduğunu gözlemledik. Bu bulgu literatürle uyumludur ve yaş YBMD'nin fenotipik özelliklerinin (koroidal neovaskülarizasyon, intraretinal kistler ve subretinal sıvı) daha belirgin ve güçlü olduğunu göstermektedir. Bu nedenle görüntüleme tabanlı modellerle sınıflandırılması daha kolaydır.²⁶ Klinik açıdan, yaş YBMD sınıflandırmasının yüksek doğruluk ile yapılıyor olması bir avantajdır, çünkü bu alt tipin erken ve doğru tanısı, ciddi görme kaybını önleyebilen zamanında müdahale için çok önemlidir. Ayrıca, beklediğimiz gibi, yaş YBMD'yi kuru tipten ayırarak sınıflandırma başarısı, normal ile herhangi bir tip YBMD'yi birbirinden ayırmaya kıyasla daha düşük olmuştur. Bu, üst üste binen özelliklerden ve kuru ve yaş YBMD arasındaki çok küçük farklılıklardan kaynaklanabilir. Bu nedenle otomatik bir sistemin iki alt türü karıştırmadan güvenilir bir şekilde ayırt etmesi zor olabilir.²

Tüm patolojik görüntüler ile normal OKT görüntülerini birbirinden ayırarak sınıflandıran AutoML Model IV'nin doğruluk düzeyi %89,00 ve ağırlıklı F1 skoru 0,88 bulunmuştur. Bu değerler patolojik ve normalin birbirinden ayırımı için makul değerlerken, uzman tarafından tasarlanan model yaklaşık mükemmel sonuç vermiş ve doğruluk oranı %99,67 ve ağırlıklı F1 skoru 0,97 bulunmuştur. AutoML modelinin ResNet-50 V2 ve genel ön işleme adımlarına dayalı bir model olması, tıbbi görüntü analizi gibi özel alanlardaki etkinliğini sınırlayabilir. ResNet-50 V2, güçlü olsa da, OKT görüntülerinde ince detayları EfficientNetV2 kadar verimli bir şekilde yakalayamayabilir. Ayrıca, AutoML platformları genellikle bazı görevler için en uygun olmayabilecek varsayılan ayarlar kullanmaktadır, bu da mimarinin suboptimal olmasına ve hiperparametre ayarlamasının yetersiz kalmasına yol açar. Bu bulgular, uzman mühendisler tarafından geliştirilen modellerin, özellikle yüksek hassasiyet gerektiren karmaşık görevlerde, yalnızca AutoML platformları aracılığıyla üretilenler modellerden daha iyi performans gösterebileceğini düşündürmektedir.

Çalışmanın Kısıtlılıkları

Bu umut verici sonuçlara rağmen, bazı kısıtlılıklar olduğu vurgulanmalıdır. Veri kümesi, bu çalışma için yeterli olsa da, daha geniş hasta popülasyonlarında bulunan değişkenliği tam olarak yansıtmayabilir. Gelecekteki araştırmalar, modellerin genellenabilirliğini geliştirmek için daha büyük ve daha çeşitli veri kümeleri ile yapılmalıdır. Ayrıca, Google Vertex AI, Microsoft Azure AutoML ve Amazon SageMaker gibi daha yeni

AutoML platformları, kodlama deneyimi olmayan kullanıcıların parametreleri optimize etmelerini sağlayarak performansı artırılabilir.²⁷ Bununla birlikte, bu platformlarda modelleme için deneme limitleri vardır ve limit aşıldığında ödeme yapılması gerekir.

Sonuç

Bu çalışma, OKT görüntüleri ile YBMD'nin sınıflandırılması için AutoML ve uzman tarafından geliştirilen ML modellerinin bir karşılaştırmasını sunmaktadır. Bulgularımız, Lobe AI AutoML platformunun klinisyenlerin ML modelleri geliştirmesi için uygun bir yol olduğunu ancak uzman girdilerinin özel görevlerde performansı optimize etmek için çok önemli olduğunu göstermektedir. Uzman tarafından tasarlanan EfficientNet V2 modeli doğruluk ve hassasiyet açısından üstündü. Bu, gelişmiş mimarilerin, özelleştirilmiş veri artırmanın ve hiperparametrelerde ince ayarlar ile optimizasyon yapılmasının değerini göstermektedir. AutoML'nin kolay erişilebilirliği ile uzman mühendisliği bir araya getirmek, klinisyenler için kullanımın kolay olmasını sağlarken model performansını daha da artırabilir. Mühendisler ve sağlık uzmanları arasında yapılacak işbirlikleri, hem etkili hem de klinik olarak kullanılabilir yapay zeka çözümleri geliştirmek ve nihayetinde hasta bakımının iyileştirilmesine katkıda bulunmak için gereklidir.

Etik

Etik Kurul Onayı: Bu çalışma Helsinki Bildirgesi ilkelerine uygun olarak yürütüldü ve Dokuz Eylül Üniversitesi Etik Kurul'dan onay alındı (protokol kodu: 2024/37-06, onay tarihi: 06.11.2024).

Hasta Onayı: Çalışmada açık erişimli veri kümeleri kullanıldığı, hastalarla doğrudan iletişim kurulmadığı ve hastaların tanımlanmasından kullanılabilecek verilere yer verilmediği göz önüne alındığında, hasta onamı alınması gerekmedi.

Beyan

Yazarlık Katkıları

Cerrahi ve Medikal Uygulama: C.D.E., Konsept: C.D.E., M.A.S., Dizayn: C.D.E., D.Ö., M.A.S., Veri Toplama veya İşleme: D.Ö., U.B., Analiz veya Yorumlama: C.D.E., U.B., Literatür Arama: C.D.E., U.B., Yazan: C.D.E., U.B.

Çıkar Çatışması: Yazarlar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemiştir.

Finansal Destek: Çalışmamız için hiçbir kurum ya da kişiden finansal destek alınmamıştır.

Kaynaklar

- Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. *Lancet*. 2018;392:1147-1159.
- Fleckenstein M, Keenan TDL, Guymer RH, Chakravarthy U, Schmitz-Valckenberg S, Klaver CC, Wong WT, Chew EY. Age-related macular degeneration. *Nat Rev Dis Primers*. 2021;7:31.
- Rasti R, Rabbani H, Mehrdehnavi A, Hajizadeh F. Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Trans Med Imaging*. 2017;37:1024-1034.

4. Kaymak S, Serener A. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. *IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. 2018:265-269.
5. Serener A, Serte S. Dry and wet age-related macular degeneration classification using oct images and deep learning. *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*. 2019:1-4.
6. An G, Yokota H, Motozawa N, Takagi S, Mandai M, Kitahata S, Hirami Y, Takahashi M, Kurimoto Y, Akiba M. Deep learning classification models built with two-step transfer learning for age related macular degeneration diagnosis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019:2049-2052.
7. Romo-Bucheli D, Erfurth US, Bogunovic H. End-to-end deep learning model for predicting treatment requirements in neovascular AMD from longitudinal retinal OCT imaging. *IEEE J Biomed Health Inform*. 2020;24:3456-3465.
8. Kim J, Tran L. Ensemble learning based on convolutional neural networks for the classification of retinal diseases from optical coherence tomography images. *Annual IEEE Symposium on Computer-Based Medical Systems*. 2020:532-537.
9. Haihong E, Ding J, Yuan L. SAE-wAMD: a self-attention enhanced convolution neural network for fine-grained classification of wet age-related macular degeneration using OCT Images. *Image Processing, Computer Vision and Machine Learning (ICICML), International Conference on*. 2022:619-627.
10. Bucur AI, Muntean GA, Marginean A, Nicoara SD. Predicting necessary treatment plans for patients with Age-Related Macular Degeneration using characteristics derived from optical coherence tomography B-scans. *IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. 2023:301-306.
11. Elsharkawy M, Sharafelddeen A, Khalifa F, Soliman A, Elnakib A, Ghazal M, Sewelam A, Thanos A, Sandhu HS, El-Baz A. A clinically explainable AI-based grading system for age-related macular degeneration using optical coherence tomography. *IEEE J Biomed Health Inform*. 2024;PP.
12. He X, Zhao K, Chu X. AutoML: a survey of the state-of-the-art. *Knowl Based Syst*. 2021;212:106622.
13. Hutter F, Kotthoff L, Vanschoren J. *Automated machine learning: methods, systems, challenges*. Springer Nature. 2019.
14. Jacoba CMP, Doan D, Salongcay RP, Aquino LAC, Silva JPY, Salva CMG, Zhang D, Alog GP, Zhang K, Locaylocay KLRB, Saunar AV, Ashraf M, Sun JK, Peto T, Aiello LP, Silva PS. Performance of automated machine learning for diabetic retinopathy image classification from multi-field handheld retinal images. *Ophthalmol Retina*. 2023;7:703-712.
15. Antaki F, Coussa RG, Kahwati G, Hammamji K, Sebag M, Duval R. Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. *Br J Ophthalmol*. 2023;107:90-95.
16. Touma S, Antaki F, Duval R. Development of a code-free machine learning model for the classification of cataract surgery phases. *Sci Rep*. 2022;12:2398.
17. Gholami P, Roy P, Parthasarathy MK, Lakshminarayanan V. OCTID: Optical coherence tomography image database. *Comput Electric Eng*. 2020;81:106532.
18. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172:1122-1131.
19. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. *Computer Vision*. 2016:630-645.
20. Damiano Perri. Confusion Matrix Generator. Accessed October 14, 2024. <https://www.damianoperri.it/public/confusionMatrix/>
21. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. *PMLR*. 2021:10096-10106.
22. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. *The 25th ACM SIGKDD International Conference*. 2019:2623-2631.
23. Kansal K, Chandra TB, Singh A. ResNet-50 vs. EfficientNet-B0: multi-centric classification of various lung abnormalities using deep learning. *Procedia Comput Sci*. 2024;235:70-80.
24. De La Fuente N, Majó M, Luzko I, Córdova H, Fernández-Esparrach G, Bernal J. Enhancing image classification in small and unbalanced datasets through synthetic data augmentation. *Clinical Image-Based Procedures*. 2024:11-21.
25. Lacerda P, Barros B, Albuquerque C, Conci A. Hyperparameter optimization for COVID-19 pneumonia diagnosis based on chest CT. *Sensors (Basel)*. 2021;21:2174.
26. Wu M, Lu Y, Hong X, Zhang J, Zheng B, Zhu S, Chen N, Zhu Z, Yang W. Classification of dry and wet macular degeneration based on the ConvNeXT model. *Front Comput Neurosci*. 2022;16:1079155.
27. Raghavendran KR, Elragal A. Low-code machine learning platforms: a fastlane to digitalization. *Informatics*. 2023;10:50.