Derleme/*Review*

Ege Journal of Medicine / *Ege Tıp Dergisi* 2026; 65 (3):638-648

Tıbbi genetikte yapay zekânın rolü: Fırsatlar ve zorluklar

The role of artificial intelligence in medical genetics: Opportunities and challenges

Erhan Parıltay[1] Zeki Sakallı[2]

[1]İstanbul University Cerrahpaşa, Cerrahpaşa Faculty of Medicine, Department of Medical Genetics İstanbul-Türkiye

[2]Ege University, Faculty of Medicine, Department of Medical Genetics Izmir-Türkiye

ÖZ

Amaç: Tıbbi genetik, hastalıkların kalıtsal temellerini aydınlatmayı ve tanı ile tedavi süreçlerinde kullanılmak üzere gen ve gen ürünlerinin analizini hedefleyen bir bilim dalıdır. Sanger dizilemesi, PCR ve İnsan Genom Projesi gibi dönüm noktalarını izleyen teknolojik ilerlemeler, 2010'lu yıllarda yeni nesil dizileme (NGS) yöntemlerinin klinik pratiğe girmesiyle hız kazanmış; çoklu gen panelleri ve tüm ekzom dizileme gibi testler daha erişilebilir hale gelmiştir. Bu gelişmeler özellikle nadir hastalıklarda tanı süresini kısaltmış ve tanısal doğruluğu artırmıştır. Bununla birlikte NGS'nin ürettiği devasa veri yükü, varyantların konvansiyonel yöntemlerle değerlendirilmesini güçleştirmiş; biyoinformatik araçlar yaygınlaşsa da varyantların klinik olarak anlamlandırılması ve fenotiple bütünleştirilmesi aşamalarında sınırlılıklar devam etmiştir. Bu bağlamda yapay zeka (YZ) ve makine öğrenmesi, genetik verilerin hızlı, hassas ve ölçeklenebilir biçimde analiz edilmesini sağlayarak klinik genetik uzmanları ve araştırmacılar için önemli bir destek aracı olarak öne çıkmaktadır. Bu derleme, tıbbi genetikte YZ uygulamalarını ve güncel araçları özetlemeyi amaçlamaktadır.

Gereç ve Yöntem: YZ, milyonlarca varyant arasından olası patojenik adayların önceliklendirilmesi, nadir hastalıklarda etken varyantın saptanması, kompleks hastalıklarda risk tahmini ve klinik karar destek süreçlerinin hızlandırılmasında kullanılmaktadır. Protein yapı tahmininde AlphaFold, protein katlanma probleminde deneysel doğruluğa yakın performans sunarak yapısal biyolojide ve hedefe yönelik ilaç tasarımında yeni olanaklar yaratmıştır; yeni sürümlerle protein dışı biyomoleküler etkileşimlerin modellenmesi de genişlemiştir.

Bulgular: Genomik veri işleme ve varyant çağrımında DeepVariant gibi araçlar dizileme kaynaklı hataları azaltarak analiz güvenilirliğini artırırken, AlphaGenome ve AlphaMissense gibi modeller varyantların moleküler ve fonksiyonel etkilerinin yorumlanmasına katkı sağlar. SpliceAI, splicing üzerindeki etkileri öngörerek özellikle nadir hastalıkların tanısında kritik bilgi üretir. Varyant sınıflandırmasının standartlaştırılmasında ClinGen VCI gibi platformlar klinik kürasyon süreçlerini destekler. Fenotip tabanlı yaklaşımlarda Face2Gene ve GestaltMatcher gibi yüz analizi araçları dismorfoloji değerlendirmesine katkı sunarken, büyük dil modelleri genetik danışmanlıkta özetleme ve eğitim materyali üretimi gibi görevleri destekleyebilir. Son olarak, veri gizliliği, algoritmik önyargı, şeffaflık/yorumlanabilirlik ve düzenleyici gereksinimler, YZ'nin güvenli klinik entegrasyonu için temel etik ve teknik başlıklar olarak ele alınmalıdır.

Sonuç: YZ sistemleri bağımsız karar verici değil; insan uzmanlığını tamamlayan, sorumluluğu klinisyende kalan karar destek araçları olarak konumlandırıldığında tıbbi genetikte tanı, kişiselleştirilmiş tedavi ve ilaç keşfi süreçlerine anlamlı katkı sağlayacaktır.

Anahtar Sözcükler: Tıbbi genetik, yapay zekâ, yeni nesil dizileme, nadir hastalıklar, makine öğrenimi

Corresponding author: Erhan Parıltay
İstanbul University Cerrahpaşa, Cerrahpaşa Faculty of Medicine, Department of Medical Genetics İstanbul-Türkiye
E-mail: *pariltay@gmail.com*
Application date: 06.07.2026 Accepted: 29.07.2026

ABSTRACT

Aim: *Medical genetics is a discipline dedicated to elucidating the hereditary basis of disease and applying analyses of genes and their products to improve diagnostic and therapeutic decision-making. Following key milestones such as Sanger sequencing, the development of polymerase chain reaction (PCR), and the Human Genome Project, technological progress accelerated with the clinical adoption of next generation sequencing (NGS) in the 2010s. As multigene panels and whole exome sequencing became increasingly accessible, diagnostic yield improved and the time to diagnosis—particularly for rare disorders—was substantially reduced. Nevertheless, the enormous data volume generated by NGS has complicated variant interpretation using conventional approaches. Although bioinformatics pipelines have expanded rapidly, persistent challenges remain in translating genomic findings into clinically meaningful conclusions and integrating them with patient phenotypes. In this context, artificial intelligence (AI) and machine learning have emerged as powerful enabling technologies, offering rapid, accurate, and scalable analysis of genomic data to support clinical geneticists and researchers. This review summarizes major applications of AI in medical genetics and highlights contemporary tools that are shaping the field.*

Materials and Methods: *AI driven methods are increasingly used to prioritize potentially pathogenic variants among millions of candidates, identify causal variants in rare diseases, estimate risk for complex disorders, and accelerate clinical decision support workflows. In protein structure prediction, AlphaFold has achieved near experimental accuracy for many targets, catalyzing advances in structural biology and enabling structure guided drug discovery; subsequent iterations have expanded modeling capabilities beyond proteins to broader biomolecular interactions.*

Results: *In genomic data processing and variant calling, tools such as DeepVariant enhance analytical reliability by reducing sequencing and alignment artefacts. Complementary models, including AlphaGenome and AlphaMissense, support interpretation by predicting molecular and functional consequences of genetic variation. SpliceAI provides high performing predictions of splicing disruption, offering clinically actionable insights that are especially relevant to rare disease diagnostics. To promote consistency and transparency in clinical interpretation, platforms such as the ClinGen Variant Curation Interface (VCI) facilitate standardized variant classification aligned with established guidelines. Beyond sequence based analyses, phenotype driven approaches employing facial analysis—such as Face2Gene and GestaltMatcher—can assist dysmorphology assessment, while large language models may streamline genetic counseling tasks including result summarization and the creation of patient facing educational materials. Despite these advances, safe clinical implementation requires careful attention to data privacy and security, algorithmic bias, transparency and interpretability, and evolving regulatory frameworks.*

Conclusions: *AI systems should be positioned not as autonomous decision makers but as decision support tools that augment human expertise, with clinical responsibility remaining with qualified professionals. When integrated within robust ethical, legal, and professional standards, AI has the potential to meaningfully advance diagnosis, personalized therapeutics, and drug discovery in medical genetics*

Keywords: *Medical genetics, artificial intelligence, next generation sequencing, rare diseases, machine learning*

1.Giriş

Tıbbi genetik, hastalıkların kalıtsal temellerini anlamayı, tanı ve tedavide kullanmak üzere ilgili gen ve gen ürünlerinin analizini hedefleyen bir bilim dalıdır. Teknolojik gelişmeler, tıbbi genetikte tanı başarısını artırmış ve hastalık mekanizmalarının daha iyi anlaşılmasına katkı sağlamıştır (1). Moleküler biyolojideki önemli dönüm noktaları arasında, 1977 yılında Frederick Sanger tarafından geliştirilen dizileme yöntemiyle genetik dizilerin okunabilmesi, 1983'te Kary Mullis'in polimeraz zincir reaksiyonunu (PCR) keşfetmesiyle çok küçük DNA parçalarının çoğaltılabilmesi ve 1990'da başlatılan İnsan Genom Projesi yer almaktadır. Bu proje kapsamında insan genomunun yaklaşık %92'si

dizilenmiş ve yaklaşık 3 milyar baz çifti tanımlanmıştır. 2010'lu yıllarda yeni nesil dizileme (NGS) teknolojilerinin klinik pratiğe girmesiyle birlikte, yüzlerce geni kapsayan paneller ve tüm ekzom dizileme testleri kolaylıkla uygulanabilir hale gelmiştir (1,2). Bu gelişmeler, özellikle nadir hastalıklarda tanı süresini belirgin biçimde kısaltmış ve tanısal doğruluğu artırmıştır. Ancak NGS ile elde edilen veri miktarının katlanarak artması, bu verilerin bireysel ve konvansiyonel yöntemlerle analiz edilmesini güçleştirmiştir. Biyoinformatik analiz araçları geliştirilmiş olsa da varyantların klinik olarak anlamlandırılması ve hastanın klinik bulgularıyla bütünleştirilmesi sürecinde halen çeşitli kısıtlılıklar bulunmaktadır. Bu noktada yapay zekâ (YZ) ve makine öğrenmesi algoritmaları, genetik verilerin hızlı, hassas ve ölçeklenebilir biçimde analiz edilmesini sağlayarak tıbbi genetik uzmanları ve araştırmacılar için güçlü bir destek sunmaktadır (3,4).

Yapay zekâ, bilgisayarların çevresel verileri algılaması, dili anlaması ve üretmesi, veri analizi yapması ve karar destek mekanizmaları oluşturması gibi gelişmiş işlevleri yerine getirmesini sağlayan teknolojiler bütünüdür. Kavram ilk kez 1956 yılında Dartmouth Konferansı'nda ortaya atılmış, 1997'de IBM tarafından geliştirilen Deep Blue'nun dünya satranç şampiyonu Garry Kasparov'u yenmesiyle önemli bir dönüm noktasına ulaşmıştır. İnternetin yaygınlaşması ve büyük veriye erişimin artmasıyla birlikte, 2000'li yıllarda ses tanıma, görüntü işleme ve tercüme gibi alanlarda yapay zekâ uygulamaları öne çıkmıştır. 2010'lu yıllarda ise bu uygulamalar tıbbi görüntüleme, nesnelerin interneti ve otonom sistemler gibi alanlara yayılmıştır. 2020'li yılların başından itibaren ise üretken yapay zekâ dönemine geçilmiştir. Yapay zekânın temelini makine öğrenmesi oluşturmakta; bunun alt bileşenleri arasında derin öğrenme ve büyük dil modellerini mümkün kılan doğal dil işleme yer almaktadır. Makine öğrenmesi, bilgisayarların açıkça programlanmadan veriden öğrenerek tahmin ve sınıflandırma yapabilmesini sağlarken, derin öğrenme çok katmanlı yapay sinir ağları aracılığıyla karmaşık verilerden otomatik özellik çıkarımı gerçekleştirmektedir. Doğal dil işleme ise bilgisayarların insan dilini anlama, yorumlama ve üretme yeteneğini ifade eder; ChatGPT ve Gemini gibi büyük dil modelleri bu alanın en gelişmiş örnekleridir.

Günümüzde yapay zekâ, tıbbi genetikte hastanın ilk değerlendirmesinden tanı ve tedavi süreçlerine kadar geniş bir kullanım alanına sahiptir. Klinik bulgulara dayalı ön tanı listelerinin oluşturulması, yüz fotoğrafları aracılığıyla olası sendromların belirlenmesi ve genomik verilerin analizinde varyant sınıflandırması gibi süreçler YZ uygulamalarıyla desteklenmektedir (5). Bu gelişmeler, hızlı ve doğru tanı, belirsiz varyantların değerlendirilmesi, kişiselleştirilmiş tıp yaklaşımları ve ilaç keşfi süreçlerine doğrudan katkı sağlamıştır. Bu derlemede, tıbbi genetik pratiğine katkı sunan yapay zekâ araçlarının incelenmesi ve alandaki güncel gelişmelerin sunulması amaçlanmıştır.

Tıbbi Genetikte Yapay Zekâ Uygulamaları: Yapay zekânın tıbbi genetikteki kullanım alanları oldukça geniştir ve sürekli olarak genişlemektedir. İnsan genomunda milyonlarca genetik varyant bulunur. YZ algoritmaları, bu varyantlardan hangilerinin hastalıkla ilişkili (patojenik) olduğunu belirlemede kritik bir rol oynar (6).Özellikle nadir görülen genetik hastalıklarda, hastalık nedeni olan varyantın tespiti için YZ modelleri büyük bir doğruluk oranı sunmaktadır (7).Bunun dışında hastalık teşhisi ve risk tahmini de temel olarak kullanıldığı alanlardandır. Bu amaca yönelik olarak yapay zekâ, bireylerin genetik profillerini analiz ederek kanser, kalp hastalıkları, diyabet gibi kompleks hastalıklara yakalanma riskini tahmin edebilir, poligenik risk skorları (PRS) oluşturabilir (8). Ayrıca, belirli semptomlar ve genetik bulgular arasında bağlantı kurarak tanı koyma sürecini hızlandırabilir. Tanı süreci dışında tedavi ve ilaç geliştirme aşamaları da yapay zekânın kullanıldığı alanlardır. Örneğin, bireyin genetik yapısı ilaca vereceği tepkiyi önemli ölçüde etkileyebilir. YZ algoritmaları, bir hastanın genetik verilerine dayanarak hangi ilacın daha etkili olacağını veya hangi ilacın yan etki riskinin yüksek olduğunu öngörebilir. Bu, "doğru hastaya, doğru ilaç, doğru dozda" prensibini hayata geçiren kişiselleştirilmiş tedavilerin önünü açar. İlaç keşfi ve geliştirme aşamalarına yönelik olarak da yapay zekâ, genetik verileri kullanarak hastalıklarla ilişkili yeni biyolojik hedefler (genler, proteinler vb.) belirleyebilir. Bu sayede, yeni ilaçların keşfedilmesi ve geliştirilmesi süreci önemli ölçüde kısalabilir.

2. Mutasyon ve Proteine Etkinin Analizi

Proteinler, yaşam için zorunlu olan yapı taşı moleküllerdendir. Proteinlerinin üç boyutlu yapısını amino asit dizilerinden tahmin etme sorunu, biyoinformatikte yarım yüzyılı aşkın

süredir bilinen büyük bir sorundur. Bu protein katlanma problemindeki asıl zorluklar ise dizinin 3 boyutlu yapıya geçmesinde belirleyici olan kimyasal ve fiziksel etkileşimlerin son derece karmaşık olmasından kaynaklanmaktadır (9). Bu etkileşimlerin karmaşıklığının çözülmesiyle proteinlerin yapılarının oluşumunun belirlenmesi ve işlevleri konusunda da yol gösterici olabileceğini düşünülmektedir. İşlevleri bilinen proteinlerin ise fizyolojik rollerinin belirlenmesi mümkün olacaktır. Hastalık mekanizmalarının anlaşılmasının, ilaç geliştirme süreçlerine de yardımcı olacağı düşünülmektedir. X Işını Kristalografisi gibi geleneksel yöntemler çok yüksek doğruluk oranları sağlasa da hız ve maliyet açısından sınırlı kalmışlardır.

DeepMind ekibi, geliştirdikleri AlphaFold ile makine öğrenimi algoritmalarını protein katlanma problemi gibi biyolojik sorunlarda da uygulamaya karar vermiştir. Ekip, 2018 senesinde AlphaFold'u geliştirerek Critical Assessment of protein Structure Prediction (CASP) 13'te Regular Target kategorisinde birinci olmuştur. Bu birincilik, sinir ağlarının proteinlerin yapısını tahmin etmekteki potansiyelini göstermiştir. 2020'de geliştirilen AlphaFold v2 ise CASP 14'te neredeyse deneysel doğruluk seviyesinde sonuçlar vererek protein biyolojisi alanında bir devrim yaratmıştır (10). AlphaFold v2, Evoformer adlı tamamen yeni bir sinir ağı mimarisiyle bu süreci uçtan uca tek bir modelde birleştirmiştir. Bu sayede proteinlerin evrimsel bilgisi ile yapısal bilgileri sürekli olarak birbirleriyle ilişkilendirilip rafine edilmektedir. AlphaFold v2 hem katlanma analizinde Multiple sequence alignment (MSA) hem de pair representation (çift temsili) kullanmaktadır (11).Pair representation her aminoasit çifti arasındaki olası ilişkileri kodlamaktadır. AlphaFold'un ilk versiyonu bir dizi ayrı parçayı birleştirerek çalıştıran bir sistemken AlphaFold v2 tek bir sinir ağında evrimsel bilgiyi ve yapısal bilgiyi eşzamanlı ve interaktif bir şekilde işleyen bir modeldir. Bütüncül yaklaşım sayesinde protein katlanma probleminde önemli bir başarı artışı söz konusu olmuştur. CASP14'te ortalama GDT_TS (Global Distance Test – Total Score) skoru > 90 olmuştur, bu X-ışını kristalografisine yakın çözünürlük demektir (10). Bu gelişmelerle günler haftalar süren deneysel çözüm süreçleri hızlıca hesaplanabilir bir hale inmiştir. Google ve EMBL-EBI işbirliği ile kullanıcılara fayda sağlayacak olan AlphaFold Protein Veri Tabanı (AlphaFold DB) oluşturulmuş ve kullanıcıların erişimine sunulmuştur. 2022'ye gelindiğinde sayısı 200 milyondan fazla proteinin bilgisi veri tabanında yer almaktadır (12). 2024'te AlphaFold v3 yayınlanmıştır. AlphaFold v3 ile sadece proteinler değil DNA, RNA, çeşitli ligand ve iyonlar gibi moleküller de modellemeye başlanmıştır. Bu moleküllerin birbirleriyle nasıl etkileşimde olduklarını ve nasıl kompleksler oluşturduklarını yüksek doğrulukla tahmin edebilmektedir (13). Son versiyonla birlikte biyolojik sistemleri bir bütün olarak ele alabilmektedir. Hedefe yönelik ilaç geliştirme, hastalık mekanizmaları gibi birçok alanda fayda sağlayacağı düşünülmektedir. Bu devrimsel gelişmeler 2024 Nobel Kimya Ödülü ile taçlandırılmıştır; protein yapı tahmini için Demis Hassabis ve John Jumper ödülün yarısını alırken, *de novo* hesaplamalı protein tasarımı alanındaki katkıları nedeniyle David Baker diğer yarısını almıştır.

DeepVariant, AlphaGenome ve SpliceAI gibi yapay zekâ ve makine öğrenimi algoritmalarının entegrasyonu genomik verilerin analizinde hassasiyet ve hız açısından önemli avantajlar sağlamaktadır (14,15). Bu bağlamda, DeepVariant, AlphaGenome ve SpliceAI gibi ileri düzey yapay zekâ tabanlı araçlar, tıbbi genetik pratiğinde giderek daha merkezi bir rol üstlenmektedir. Islak laboratuvar süreçlerinden sonra elde edilen büyük miktardaki verinin saflaştırılması ve sonrasındaki analiz sürecine doğru bir şekilde aktarılması gerekmektedir.

DeepVariant, Google tarafından geliştirilen ve DNA dizileme verilerini alarak, hangi genetik farklılıkların gerçek olduğunu, hangilerinin ise dizileme sırasında oluşmuş hatalar olduğunu anlamaya çalışan bir analiz aracıdır (14). İnsan genomu yaklaşık 3 milyar baz çiftinden oluşmaktadır. Yeni Nesil Dizileme yöntemlerinden bazıları bu diziyi okumak için diziyi çok sayıda küçük parçaya ayırmaktadır. Bu parçalar referans genoma göre hizalanır. Bu süreçte okuma hataları, hizalama hataları gibi sorunlar ortaya çıkmaktadır. Deepvariant bu hataları ortadan kaldırmak için geliştirilmiştir (14,16).Referans genomla uyumlu olmayan diziler işaretlenip görselleştirilir. Bu görsel, Konvolüsyonel Sinir Ağları (CNN) adı verilen bir yapay zekâ modeline aktarılır. CNN bu görselleri analiz ederek hangilerinin gerçek genetik varyantlar olduğunu, hangilerinin ise sadece dizileme hatası olduğunu yüksek doğrulukla sınıflandırır. Bu analiz sonrasında olası genetik değişikleri içeren variant calling dosyası (VCF) oluşturur. Hatalı dizileme

süreçlerini önleyen bu yaklaşım sayesinde genetik verideki ‘gürültüler’ engellenebilmektedir.

AlphaGenome, Google'ın DeepMind ekibi tarafından geliştirilen ve DNA'daki genetik varyantların moleküler etkilerini tahmin etmekte faydalanılan bir yapay zekâ modelidir. AlphaGenome’un amacı, AlphaFold’un protein katlanmasındaki başarılarını genomik veriye taşımak ve genlerin düzenlenmesinde rol oynadığı bilinen fakat ‘non coding’ kodlanmayan bölgeler gibi alanlara ışık tutmaktır. Genetik değişikliklerin protein bağlanması, gen ifadesi veya RNA üretimi gibi moleküler süreçleri nasıl etkilediğini ön görmeye yardımcı olur. Tahmin edilen özellikler arasında genlerin farklı hücre tipleri ve dokulardaki başlangıç ve bitiş noktaları, kesilme (splicing) bölgeleri, üretilen RNA miktarı ve ayrıca hangi DNA bazlarının erişilebilir olduğu, birbirine yakın olduğu veya belirli proteinler tarafından bağlandığı bulunur. AlphaGenome modeli 1 milyon DNA bazına kadar analiz yapar ve tahminleri tek baz çözünürlüğünde verir. Hem uzun kapsama alanı sayesinde uzaktan düzenleyen bölgelerin etkisini ön görebilmektedir hem de yüksek çözünürlük sayesinde detayları yakalayabilmektedir (17). Bazı nadir hastalıklar, splicing hatalarından kaynaklanmaktadır. AlphaGenome bu kesim noktalarının yerini ve ekspresyon düzeyini genetik diziden doğrudan modelleyebilen ilk araçtır. Varyantların fonksiyonel etkilerinin değerlendiril mesi, klinik önemi ve hastalıkla ilişkili olma potansiyelinin yorumlanması süreçleri AlphaGeno me sayesinde geliştirilmiştir.

AlphaMissense ise yine Google DeepMind tarafından geliştirilen, insan genomundaki "yanlış anlamlı varyant" olarak bilinen genetik mutasyonların hastalıklara yol açıp açmayacağını tahmin etmek için tasarlanmış bir yapay zekâ aracıdır (18). AlphaFold adlı protein yapısı tahmin modelinin bir uyarlaması olan AlphaMissense, insan ve primat popülasyon frekans veritabanları üzerinde ince ayar yapılarak geliştirilmiştir. Bu teknoloji, milyonlarca potansiyel nokta mutasyonunu analiz ederek, bu genetik değişikliklerin büyük bir kısmını "muhtemel iyi huylu" veya "muhtemel patojenik (hastalığa neden olan)" olarak sınıflandırmıştır. AlphaMissense tarafından oluşturulan bu kapsamlı katalog, bilim insanlarının ve tıp uzmanlarının genetik hastalıkları teşhis etmelerine, araştırmalarına ve potansiyel olarak daha etkili tedaviler geliştirmelerine yardımcı olma potansiyeli taşımaktadır.

SpliceAI, genetik varyantların splicing üzerindeki etkilerini tahmin etmek için tasarlanmış derin öğrenme tabanlı bir yapay zekâ aracıdır. Konvolüsyonel sinir ağları (CNN) adı verilen derin öğrenme modelleri kullanılarak geliştirilmiştir. Model, GENCODE ve GTEx gibi insan genomu veritabanlarındaki on binlerce bilinen eklenme bölgesi ve yüz binlerce hatalı eklenme bölgesi verisiyle eğitilmiştir (15).DNA’mızdaki genlerin ifadesi önce RNA’ya ardından da RNA aracılığıyla proteinleri oluşturur. Bu süreçte pre mRNA molekülünde kodlanan (ekzon) ve kodlanmayan bölgeler (intron) mevcuttur. İntronların kesilip ekzonların birleşmesiyle mRNA oluşmaktadır. Bu kesim işlemi splicing olarak adlandırılır ve bu noktadaki hatalar yanlış proteinlerin üretimine yol açabilir. SpliceAI, genetik varyantların pre-mRNA splicing üzerindeki etkilerini yüksek doğrulukla tahmin etmektedir. Bunu, bir mutasyonun yeni bir splice bölgesi yaratıp yaratmadığını veya mevcut bir splice bölgesini bozup bozmadığını analiz ederek yapar. Splice bölgelerinin değerlendirilmesi, nadir hastalıkların tanısında önemli yer tutmaktadır. Varyantın splice bölgesini etkilediğini gösterebilen bir skor, bu varyantın hastalığa neden olabileceği konusunda yol gösterici olabilmektedir. Hastalıkların moleküler mekanizmalarını anlama ve bu mekanizmalar eşliğinde tedaviye yönelik ilaç geliştirme aşamalarına SpliceAI’ın önemli katkıları olmuştur.

ClinGen Variant Curation Interface (VCI), genetik varyantların klinik önemini değerlendirmek ve sınıflandırmak için tasarlanmış, halka açık, web tabanlı bir platformdur. Ulusal Sağlık Enstitüleri (NIH) tarafından finanse edilen ClinGen Konsorsiyumu tarafından geliştirilen bu arayüz, araştırmacıların ve klinisyenlerin Amerikan Tıbbi Genetik ve Genomik Koleji (ACMG) ve Moleküler Patoloji Derneği (AMP) tarafından belirlenen kılavuzlara uygun olarak genetik varyantları sistematik bir şekilde değerlendirmelerini sağlar. Platform, klinik durum ile ilişkilendirilebilecek bilimsel kanıtların toplanması, yorumlanması ve paylaşılması süreçlerini standartlaştırarak bireysel veya grup halinde içeriğin düzenlenmesine olanak tanır. Bu sayede, genetik varyantların hastalıklarla olan ilişkisinin tutarlı bir şekilde sınıflandırılmasına yardımcı olur ve elde edilen veriler, hassas tıp ve genetik hastalıkların teşhisi alanlarında kullanılmak üzere ClinVar gibi halka açık veritabanları aracılığıyla daha geniş bilim topluluğuyla paylaşılır (19).

3. Fenotip ve Sendromlar

GestaltMatcher ve Face2Gene: Klinik genetik uygulamalarında, tanı başarısını artıran en önemli şeylerden biri dismorfolojik bulguların sistematik bir şekilde tanımlanmasıdır. Yüzdeki dismorfizm lerin belirlenmesi sendromik bozukluk ların tanısında önemli rol oynamaktadır. Fakat yüzdeki dismorfizmlerin belirlenmesinde birçok zorluk mevcuttur. Öncelikle bu fenotipik özellikler; değerlendiren doktorun kişisel bilgi birikimi ve tecrübesine bağlı olarak değişebilmektedir. Yüzdeki özellikler hastanın cinsiyet, yaş ve kökenine göre değişiklik gösterebilmektedir (20). Özellikle köken burada önemli bir yer tutmaktadır. Kökenlere göre yüz özellikleri değişebilmektedir. Örneğin, yukarı eğimli palpebral fissürler sağlıklı Asyalılarda yaygınken, diğer popülasyonlarda dismorfik olarak algılanabilir. Bu da standartlaş tırılmış ve sistematik bir değerlendirme yapmayı zorlaştırmaktadır.

Face2Gene, klinik ve genetik tanı alanında kullanılan bir yapay zekâ uygulamasıdır. Temel amaç, yapay zekâ modellerini kullanarak nadir hastalıklara sahip bireylerin tanı süresini kısaltmak ve tanıya yardımcı olmaktır. Face2Gene, derin öğrenme ve bilgisayar görüşü teknolojilerini kullanarak çalışmaktadır (21). Hastanın frontal yüz fotoğrafının çekilmesi ve uygulamaya eklenmesiyle yapay zekânın çekilen fotoğraftaki birçok noktayı analiz etme süreci başlamaktadır. Gözler, burun, ağız, kulaklar, çene hattı gibi bölgelerdeki mesafeleri, boyutları ve şekilleri ölçen uygulama, bu özellikleri veriye dönüştürmektedir. Bu veriler algoritma sayesinde daha önceden bilinen yüzlerce nadir hastalığın ve binlerce hastanın olduğu kendi veri tabanıyla karşılaştırmaktadır. Bu veri tabanlarında daha önceden standardize edilmiş HPO (The Human Phenotype Ontology) terimlerini içeren doğrulanmış bilgiler bulunmaktadır. Sonuç olarak hastanın yüz özelliklerini, veri tabanındaki bilinen sendromlarla eşleştirir ve benzerlik olasılığına göre bir liste oluşturur. Bu, klinisyene, göz önünde bulundurması gereken potansiyel sendromlar için bir fenotipik önceliklendirme sunmaktadır. Ön tanılar eşliğinde, doğru genetik test yöntemlerinin seçilmesiyle nadir hastalıkların tanı süreçlerini kısaltmaktadır.

GestaltMatcher, Face2Gene'e benzer bir misyona sahip olmakla birlikte daha nadir hastalıkların araştırılmasını hedefleyen bir araştırma platformudur (22).Farklı etnik kökenlere sahip olan hastaları kapsayarak FAIR (findable, accessible, interoperable, reusable - bulunabilir, erişilebilir, birlikte çalışabilir, yeniden kullanılabilir) prensiplerine uygun olarak tasarlanmış ilk veri tabanıdır. 6 Nisan 2024 itibarıyla, GMDB-FAIR veri setinde 8.346 hastaya ait 10.980 portre görüntüsü bulunmaktadır. Bu veriler ile 581 genetik hastalık kapsanmaktadır. Hastalar arasında 2.224 bilimsel yayından derlenen vakalar da yer almaktadır. Hastaların fotoğraflarının yanı sıra genetik bilgisi ve tanımlanmış HPO bilgileri de yer almaktadır. Özellikle çalışmada Avrupa ve Kuzey Amerika dışı hastaların da temsili göz önünde bulundurulmuştur. Uygulama, yüz görünümünden elde ettiği verilerin sonucunda olası sendromların bir listesini vermektedir. Avrupalı olmayan hastaların bilgilerinin sisteme eklenmesi ile uygulamanın tanı başarı performansında artış gözlenmiştir. Özellikle ilk 5'teki doğruluk oranı %11,29 artmış olmakla birlikte Avrupalı hastaların tanısındaki performansında azalma olmamıştır. Artan verilerin analiziyle ilerleyen yıllarda tanı başarısını daha da artıracağı düşünülmekte ve yeni nadir hastalıkların yüz görünümlerinin tanımlanması mümkün olacaktır (23).

Büyük dil modellerinin (LLM'ler) genetik danışmanlık alanındaki kullanımı, önemli fırsatlar sunarken aynı zamanda dikkatli bir biçimde ele alınması gereken etik ve teknik zorlukları da beraberinde getirmektedir. LLM'ler, genetik test sonuçlarının özetlenmesi, hasta raporları ve bilimsel literatürden ilgili bilgilerin çıkarılması, kişiye özel eğitim materyallerinin hazırlanması ve sık sorulan soruların yanıtlanması gibi görevleri otomatikleştirerek genetik danışmanların iş akışını önemli ölçüde hızlandırabilir (24,25). Bu teknoloji, özellikle genetik danışman sayısının yetersiz olduğu bölgelerde hizmetlere erişimi artırma ve bekleme sürelerini kısaltma potansiyeli taşımaktadır (24,26).

Bununla birlikte, hasta verilerinin gizliliğinin korunması, modellerin yanlış veya "halüsinasyon" olarak adlandırılan uydurma bilgiler üretme riski ve mevcut genetik veri tabanlarındaki etnik köken temelli yanlılıkların algoritmalara yansıması gibi önemli sorunlar mevcuttur (27,28). Bu nedenle, LLM'lerin bağımsız bir danışman olarak değil, insan uzmanlığını destekleyen ve verimliliği artıran yardımcı araçlar olarak konumlandırılması kritik öneme sahiptir (26). Nitekim empati kurma, karmaşık aile dinamiklerini anlama ve etik açıdan hassas durumlarda rehberlik etme gibi insani

beceriler, genetik danışmanlığın vazgeçilmez bileşenleri olmaya devam etmektedir (25,26).

DNABERT, genomik dizileri bir dil gibi ele almak amacıyla geliştirilmiş ve Google'ın BERT modelinden uyarlanmış bir yapay zekâ modelidir (29,30). Modelin temel hedefi, DNA dizilim lerindeki nükleotidler arasındaki anlamsal ve bağlamsal ilişkileri yakalayarak DNA'nın "dilini" çözümlemektir (29). DNABERT, DNA dizisini *k-mer* adı verilen, birbirleriyle örtüşen küçük parçalara ayırarak çalışır (31). Ardından, insan genomu gibi büyük ve etiketsiz veri setleri üzerinde ön eğitime tabi tutulur ve bu süreçte dizideki maskelenmiş nükleotidleri tahmin etmeyi öğrenir. Bu ön eğitim sayesinde model, DNA'nın yapısal kurallarını ve tekrar eden kalıplarını derinlemesine kavrama yeteneği kazanır.

Ön eğitim tamamlandıktan sonra DNABERT, daha küçük ve özel veri setleri kullanılarak kolayca "ince ayar" yapılabilir ve çeşitli genomik görevlerde yüksek doğrulukla kullanılabilir (29). Bu görevler arasında gen ekspresyonunu başlatan promotör bölgelerinin tanımlanması, gen birleştirme (splicing) bölgelerinin tahmini ve transkripsiyon faktörü bağlanma bölgelerinin belirlenmesi gibi kritik analizler yer alır. Modelin en dikkat çekici avantajlarından biri, yalnızca eğitildiği organizmada (örneğin insan) değil, farklı türlerin genomik verilerinde de başarıyla uygulanabilmesidir. Ayrıca DNABERT, belirli bir fonksiyon için önemli nükleotidleri görselleştirerek yüksek düzeyde yorumlanabilirlik sağlar (29). DNABERT'in başarısının ardından, daha verimli tokenizasyon yöntemleri kullanan ve performansı artırılmış DNABERT-2 gibi gelişmiş versiyonları da geliştirilmiştir (32).

Yapay zekâ (YZ), geleneksel olarak on yıllar süren ve milyarlarca dolarlık yatırım gerektiren ilaç geliştirme sürecini köklü biçimde dönüştürmektedir (33,34). Bu dönüşümün ilk ve en kritik aşaması, hastalıklarla ilişkili doğru biyolojik hedeflerin (genellikle genler veya proteinler) tanımlanmasıdır (35). YZ algoritmaları, genomik ve proteomik veriler, klinik kayıtlar ve milyonlarca bilimsel yayın gibi geniş ve karmaşık veri kaynaklarını sistematik olarak analiz ederek, insan araştırmacıların gözden kaçırabileceği biyolojik desenleri ve ilişkileri ortaya çıkarmaktadır (33).Bu sayede hastalıkların moleküler mekanizmalarının daha hızlı aydınlatılması ve yeni, daha etkili ilaç hedeflerinin tespit edilmesi mümkün olmaktadır (35).Böylece araştırma süreçleri en başından doğru hedefe yönlendirilerek sonraki aşamalardaki başarısızlık olasılığı önemli ölçüde azaltılmaktadır (33).

Hedef belirleme sürecini takiben ikinci temel aşama, bu hedefi etkileyebilecek aday ilaç moleküllerinin tasarlanması ve keşfidir. Geleneksel yaklaşımlar binlerce kimyasal bileşiğin laboratuvar ortamında taranmasını gerektirirken, YZ bu süreci büyük ölçüde dijitalleştirmektedir (36). Özellikle üretken yapay zekâ (Generative AI) modelleri, hedefe yüksek bağlanma potansiyeli ve düşük toksisite gibi arzu edilen özelliklere sahip tamamen yeni moleküller tasarlayabilmektedir (37). Google DeepMind'ın AlphaFold modeli gibi araçlar sayesinde protein yapı tahmini yüksek doğrulukla yapılabilmekte; bu yapısal bilgiler, yapay zekânın hedef biyomoleküllere optimum şekilde bağlanacak özgün moleküller tasarlamasını mümkün kılmaktadır. Bu "akıllı ilaç tasarımı" yaklaşımı, potansiyel adayların keşif süresini aylar yerine haftalar mertebesine indirerek süreç verimliliğini dramatik biçimde artırmaktadır (34).

YZ, ayrıca ilaç geliştirmenin en maliyetli ve en yüksek başarısızlık oranına sahip aşaması olan klinik denemelerde de kritik rol oynamaktadır. Modeller, aday moleküllerin potansiyel toksisite ve yan etki profillerini klinik aşamadan önce öngörerek başarısız olma ihtimali yüksek moleküllerin elenmesini sağlar (35). Bunun yanı sıra, elektronik sağlık kayıtları ve genetik verileri analiz ederek klinik çalışmalar için en uygun hasta popülasyonunu belirlemede önemli katkı sunar. Böylece, daha küçük ölçekli, daha kısa süreli ve daha yüksek başarı şansına sahip klinik denemelerin tasarlanması mümkün olur.

Sonuç olarak, yapay zekâ teknolojileri ilaç geliştirme sürecini hızlandırarak maliyeti düşürmekte ve doğruluğu artırmaktadır. Bu ilerlemeler, kişiselleştirilmiş tedavilerin ve yeni nesil ilaçların klinik uygulamaya geçişini hızlandırarak hastalara daha kısa sürede ulaşmasını sağlayan stratejik bir dönüşüm yaratmaktadır (34).

4. Zorluklar ve Etik Hususlar

Bu heyecan verici gelişmelere karşın, yapay zekânın tıbbi genetik alanında etkin ve güvenilir biçimde kullanılabilmesi için dikkate alınması gereken önemli sınırlılıklar ve etik sorumluluklar bulunmaktadır (38,39). Teknolojik ilerleme hızının yüksek olması, mevcut klinik altyapı, etik

çerçevelerin ve hukuki regülasyonların aynı hızda uyum sağlamasını zorlaştırmakta; bu nedenle yapay zekâ uygulamalarının kontrollü, bilinçli ve güvenli biçimde hayata geçirilmesi gerekmektedir.

Yapay zekâ temelli sistemlerin kullanımına ilişkin temel zorluklardan biri veri gizliliği ve güvenliğidir (40). Genetik veriler, kişilerin sağlık geçmişi, aile bilgisi ve kalıtsal hastalık risklerine ilişkin önemli ve hassas bilgiler içermektedir. Bu nedenle bu verilerin güvenli biçimde saklanması, işlenmesi ve yetkisiz erişimlere karşı korunması zorunludur. Veri sızıntılarını engellemek için, doğru şifreleme tekniklerini uygulamak ve güvenli veri paylaşım protokolleri geliştirmek, yapay zekâ tabanlı genetik analizlerin etik kullanımının temel koşulları arasında yer almaktadır.

Bir diğer önemli konu algoritmik ön yargı (bias) meselesidir. Yapay zekâ sistemlerinin geliştirilmesinde kullanılan eğitim verilerinin çeşitliliği ve temsil gücü, algoritmaların doğruluğunda belirleyici rol oynamaktadır. Eğitim veri setleri belirli etnik grupları, coğrafi bölgeleri veya genetik varyasyonları yeterince temsil etmiyorsa, model sonuçları bu az temsil edilen gruplar için hatalı veya eksik olabilir (41).Günümüzde kullanılan yapay zekâ araçlarının başarısının azalmasında da önemli bir sebeptir. Bu durum tıbbi karar süreçlerinde eşitsizlikleri artırma riskini de barındırmaktadır. Dolayısıyla, modellerin farklı popülasyonları kapsayan dengeli ve yüksek kalitede veri setleri ile eğitilmesi klinik güvenilirlik açısından kritik öneme sahiptir.

Ayrıca, şeffaflık ve yorumlanabilirlik ilkeleri de tıbbi uygulamalar açısından merkezi bir önem taşımaktadır. Bazı yapay zekâ modelleri “kara kutu” niteliği taşıyarak karar süreçlerinin iç dinamiklerini açıkça ortaya koymamaktadır. Bu dinamiklerin bilinmesiyle klinik uygulamalardaki tanı ve tedavi süreçlerindeki başarı oranlarını artıracağı düşünülmektedir. Klinik uygulamalarda kullanılan sistemlerin, hekimlerin karar mekanizmalarına entegre edilebilmesi ve gerektiğinde klinik muhakeme ile doğrulanabilmesi için sonuçların açık, anlaşılır ve yorumlanabilir olması gereklidir. Bu gereklilik, hasta güvenliği ve klinik sorumluluk açısından da önem arz etmektedir.

Son olarak, düzenleyici çerçevenin oluşturulması yapay zekânın sağlık alanında güvenli ve standartlara uygun kullanımının en kritik boyutlarından biridir. Yapay zekâya dayalı tanı ve tedavi destek sistemlerinin klinik pratiğe entegre edilebilmesi için ulusal ve uluslararası düzeyde açık regülasyonlar, validasyon kriterleri ve sertifikasyon süreçlerinin geliştirilmesi gerekmektedir (38). Bu düzenlemeler, hem hasta güvenliğini garanti altına almak hem de teknolojinin etik ve bilimsel temellere uygun şekilde ilerlemesini sağlamak açısından vazgeçilmez öneme sahiptir.

5. Sonuç

Yapay zekâ teknolojilerinin tıbbi genetik alanına entegrasyonu, yalnızca mevcut uygulamaları iyileştirmekle kalmayıp, genomik verilerin analizinden kişiselleştirilmiş tedavi planlarının oluşturulmasına kadar geniş bir yelpazede paradigma değişimi yaratmaktadır (39).Bu kapsamda, YZ destekli sistemler; genetik varyantların önceliklendirilmesi, genomik veri entegrasyonu, klinik karar destek sistemleri ve halk sağlığı ölçekli genetik risk modellemeleri gibi alanlarda devrim niteliğinde katkılar sunma potansiyeline sahiptir (Tablo-1).

Tablo-1.Tıbbi Genetikte kullanılan örnek yapay zekâ araçları ve temel işlevleri

Kullanım Alanı	İlgili Örnek Araç/Model	Temel İşlev
Mutasyon ve Proteine Etkinin Analizi	AlphaFold	Protein 3B yapı tahmini; DNA/RNA/ligand etkileşim modellemesi
	DeepVariant	Dizileme/hizalama hatalarının ayıklanması, variant calling
	AlphaGenome	Varyantların gen ifadesi ve splicing üzerindeki moleküler etkisinin tahmini
	AlphaMissense	Yanlış anlamlı (missense) varyantların patojenite sınıflandırması
	SpliceAI	Splicing bölgesi bozulmalarının tahmini
Klinik Varyant Yorumlama	ClinGen VCI	ACMG/AMP kriterlerine göre standardize varyant kürasyonu ve sınıflandırma

Fenotip ve Sendromlar	Face2Gene	Yüz fotoğrafından dismorfik özellik çıkarımı, sendrom eşleştirme
	GestaltMatcher	Yüz fotoğrafından dismorfik özellik çıkarımı, sendrom eşleştirme
Genomik Dil Modelleme	DNABERT / Nucleotide Transformer	DNA dizisinin "dil" gibi modellenmesi; promotör/splicing/TF bağlanma tahmini
Genetik Danışmanlık	Büyük Dil Modelleri (LLM: ChatGPT, Claude, Gemini vb.)	Sonuç özetleme, hasta bilgilendirme materyali üretimi
İlaç Geliştirme	(Genel YZ/ML yaklaşımları)	Hedef belirleme, aday molekül tasarımı, klinik deneme risk tahmini

Ancak, bu teknolojinin vaat edilen potansiyelinden tam olarak yararlanılabilmesi için veri gizliliği, algoritmik önyargı (bias) ve şeffaflık-yorumlanabilirlik gibi teknik ve etik zorlukların dikkatle ele alınması gerekmektedir. Uluslararası düzeyde geliştirilmiş rehberler ve çerçeveler, bu alanlarda yol gösterici olmaktadır. Örneğin, Global Alliance for Genomics and Health (GA4GH) "Framework for Responsible Sharing of Genomic and Health-related Data" başlığı altında şeffaflık, hesap verebilirlik, veri kalitesi ve güvenliği ilkelerini vurgulamaktadır (42). Özellikle yapay zekâ destekli otomatize araştırma altyapıları, ar-ge süreçlerini hızlandırsa da üretken yapay zekânın halüsinasyona eğilimli olması üretilen verinin bilimsel ve fonksiyonel olarak yeniden değerlendirilmesi gerekliliği göz ardı edilmemelidir.

YZ destekli sistemler, insan uzmanlığının yerini alacak bir alternatif değil; aksine klinik genetik uzmanlarının ve araştırmacıların yeteneklerini tamamlayan, karar alma süreçlerini güçlendiren bir destek aracı olarak konumlandırılmalıdır (39). Klinik uygulamalarda nihai sorumluluk hala genetik uzmanlarına ait olmalı ve YZ önerileri, etik, yasal ve profesyonel çerçeveler içinde değerlendirilmelidir.

Sonuç olarak, yapay zekâ tıbbi genetik alanında yeni bir çağın kapısını aralamaktadır. Sürekli gelişen algoritmalar, artan veri kapasitesi, ulusal ve uluslararası düzeyde standartlaşma çabaları, asistan araçların klinik entegrasyonuyla birlikte sağlık sistemlerinin daha bütüncül, hızlı, ekonomik ve erişilebilir bir yapıya dönüşmesini mümkün kılmaktadır. Ancak bu dönüşümün sürdürülebilir ve güvenli biçimde gerçekleşebilmesi için teknoloji, etik ilkeler ve klinik uzmanlık arasındaki dengeli ilişkinin korunması elzemdir.

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding: No funding or financial support was received for this study.

Acknowledgments: None.

Kaynaklar

1. Lohmann K, Klein C. Next generation sequencing and the future of genetic diagnosis. Neurotherapeutics. 2014 Oct;11(4):699–707. doi:10.1007/s13311-014-0288-8
2. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. New England Journal of Medicine. 2013 Oct 17;369(16):1502–11. doi:10.1056/NEJMoa1306555
3. Eraslan G, Avsec Ž, Gagneur J, Theis FJ. Deep learning: new computational modelling techniques for genomics. Nat Rev Genet. 2019 Jul 10;20(7):389–403. doi:10.1038/s41576-019-0122-6
4. Routhier E, Mozziconacci J. Genomics enters the deep learning era. PeerJ. 2022 Jun 24;10:e13613. doi:10.7717/peerj.13613
5. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019 Jan 7;25(1):60–4. doi:10.1038/s41591-018-0279-0
6. De La Vega FM, Chowdhury S, Moore B, Frise E, McCarthy J, Hernandez EJ, et al. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med. 2021 Oct 14;13(1):153. doi:10.1186/s13073-021-00965-0
7. Germain DP, Gruson D, Malcles M, Garcelon N. Applying artificial intelligence to rare diseases: a literature review highlighting lessons from Fabry disease. Orphanet J Rare Dis. 2025 Apr 17;20(1):186. doi:10.1186/s13023-025-03655-x

8. Rezayi S, R Niakan Kalhori S, Saeedi S. Effectiveness of Artificial Intelligence for Personalized Medicine in Neoplasms: A Systematic Review. Biomed Res Int. 2022 Apr 7;2022(1):7842566. doi:10.1155/2022/7842566
9. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. Science (1979). 2012 Nov 23;338(6110):1042–6. doi:10.1126/SCIENCE.1219021/
10. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Applying and improving AlphaFold at CASP14. Proteins. 2021 Dec;89(12):1711–21. doi:10.1002/prot.26257
11. Bouatta N, Sorger P, AlQuraishi M. Protein structure prediction by AlphaFold2: are attention and symmetries all you need? Acta Crystallogr D Struct Biol. 2021 Aug 1;77(Pt 8):982–91. doi:10.1107/S2059798321007531
12. Fleming J, Magana P, Nair S, Tsenkov M, Bertoni D, Pidruchna I, et al. AlphaFold Protein Structure Database and 3D-Beacons: New Data and Capabilities. J Mol Biol. 2025 Aug;437(15):168967. doi:10.1016/j.jmb.2025.168967
13. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024 Jun 13;630(8016):493–500. doi:10.1038/s41586-024-07487-w
14. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018 Oct 24;36(10):983–7. doi:10.1038/nbt.4235
15. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019 Jan;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015
16. Abdelwahab O, Belzile F, Torkamaneh D. Performance analysis of conventional and AI-based variant callers using short and long reads. BMC Bioinformatics. 2023 Dec 14;24(1):472. doi:10.1186/s12859-023-05596-3
17. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021 Oct 4;18(10):1196–203. doi:10.1038/s41592-021-01252-x
18. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science (1979). 2023 Sep 22;381(6664):eadg7492. doi:10.1126/science.adg7492
19. Preston CG, Wright MW, Madhavrao R, Harrison SM, Goldstein JL, Luo X, et al. ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines. Genome Med. 2022 Dec 18;14(1):6. doi:10.1186/s13073-021-01004-8
20. Solomon BD, Adam MP, Fong CT, Girisha KM, Hall JG, Hurst ACE, et al. Perspectives on the future of dysmorphology. Am J Med Genet A. 2023 Mar 1;191(3):659-671. doi:10.1002/AJMG.A.63060
21. Rudy, N. L., A. Gomes, T. Tkemaladze, et al. 2025. " Artificial Intelligence Software Changes Rare Disease Testing Strategy in Real Time: An International Case Series Using Face2Gene ." American Journal of Medical Genetics Part A 197, no. 11: e64171. https://doi.org/10.1002/ajmg.a.64171.
22. Hsieh TC, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet. 2022 Mar 10;54(3):349–57. doi:10.1038/s41588-021-01010-x
23. Lesmann H, Moosa S, Pantel T, Rosnev S, Hustinx A, Javanmardi B, et al. P196: GestaltMatcher Database: A FAIR database for medical imaging data of rare diseases. Genetics in Medicine Open. 2023;1(1):100225. doi:10.1016/j.gimo.2023.100225
24. Wong K. How Artificial Intelligence will help Genetic Counselors [Internet]. 2023 [cited 2025 Oct 7]. Available from: https://perspectives.nsgc.org/Article/how-artificial-intelligence-will-help-genetic-counselors
25. Meekins-Doherty L, Dive L, McEwen A, Sexton A. Generative AI and the profession of genetic counseling. J Genet Couns. 2025 Apr 20;34(2). doi:10.1002/jgc4.2009
26. Burgess M. AI in Genetic Counselling [Internet]. 2024 [cited 2025 Oct 7]. Available from: https://www.trakgene.com/2024/07/26/ai-in-genetic-counselling/
27. Kostick-Quenet K. Ethical, legal, and social implications of generative AI (GenAI) in healthcare. ELSIhub Collections. 2024 Feb 8. doi:10.25936/w5ry-bq46
28. Fareed M, Fatima M, Uddin J, Ahmed A, Sattar MA. A systematic review of ethical considerations of large language models in healthcare and medicine. Front Digit Health. 2025 Sep 11;7. doi:10.3389/fdgth.2025.1653631
29. Ji Y, Zhou Z, Liu H, Davuluri R V. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. Bioinformatics. 2021 Aug 9;37(15):2112–20. doi:10.1093/bioinformatics/btab083

30. Kassab D. DNABERT, a linguistic approach for sequential predictions within Biology and Health [Bachelor's Thesis]. [Buinen]: University of Groningen; 2024.
31. Sanabria M, Hirsch J, Poetsch AR. Distinguishing word identity and sequence context in DNA language models. BMC Bioinformatics. 2024 Sep 13;25(1):301. doi:10.1186/s12859-024-05869-5
32. Zhou Z, Ji Y, Li W, Dutta P, Davuluri R, Liu H. DNABERT-2: Efficient Foundation Model and Benchmark For Multi-Species Genome. arXiv [Preprint]. 2023 Jun 23 [revised 2024 Mar 18]. Available from: arXiv:2306.15006.
33. Blanco-González A, Cabezón A, Seco-González A, Conde-Torres D, Antelo-Riveiro P, Piñeiro Á, et al. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. Pharmaceuticals. 2023 Jun 18;16(6):891. doi:10.3390/ph16060891
34. Bai F, Li S, Li H. AI enhances drug discovery and development. Natl Sci Rev. 2024 Feb 1;11(3). doi:10.1093/nsr/nwad303
35. Kim H, Kim E, Lee I, Bae B, Park M, Nam H. Artificial Intelligence in Drug Discovery: A Comprehensive Review of Data-driven and Machine Learning Approaches. Biotechnology and Bioprocess Engineering. 2020 Dec 7;25(6):895–930. doi:10.1007/s12257-020-0049-y
36. Zeng X, Wang F, Luo Y, Kang S gu, Tang J, Lightstone FC, et al. Deep generative molecular design reshapes drug discovery. Cell Rep Med. 2022 Dec;3(12):100794. doi:10.1016/j.xcrm.2022.100794
37. Gangwal A, Ansari A, Ahmad I, Azad AK, Kumarasamy V, Subramaniyan V, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. Front Pharmacol. 2024 Feb 7;15. doi:10.3389/fphar.2024.1331062
38. Boudi AL, Boudi M, Chan C, Boudi FB. Ethical Challenges of Artificial Intelligence in Medicine. Cureus. 2024 Nov 26. doi:10.7759/cureus.74495
39. Johnson KB, Wei W, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision Medicine, AI, and the Future of Personalized Health Care. Clin Transl Sci. 2021 Jan 12;14(1):86–93. doi:10.1111/cts.12884
40. Zhou J, Huang C, Gao X. Patient privacy in AI-driven omics methods. Trends in Genetics. 2024 May;40(5):383–6. doi:10.1016/j.tig.2024.03.004
41. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: Implications for clinical decision-making. PLOS Digital Health. 2024 Nov 7;3(11):e0000651. doi:10.1371/journal.pdig.0000651
42. Rehm HL, Page AJH, Smith L, Adams JB, Alterovitz G, Babb LJ, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genomics. 2021 Nov;1(2):100029. doi:10.1016/j.xgen.2021.100029