

تحول ژنتیک نوین با هوش مصنوعی: از تحلیل داده‌های ژنومی تا پزشکی دقیق و ویرایش ژن

پرنیا پسندشانجانی^۱، زهرا عرفانی مقدم^{۲*}

۱- دبیرستان فرزنانگان ۲ دوره اول، سازمان ملی پرورش استعدادهای درخشان (سمپاد)، تهران ahani.maryam65@gmail.com

۲- دبیر پژوهش، سازمان ملی پرورش استعدادهای درخشان (سمپاد)، دکترای اصلاح نباتات- ژنتیک بیومتری

zahra.erfanimoghdam@alumni.znu.ac.ir

چکیده

رشد شتابان فناوری‌های تعیین توالی نسل جدید، فناوری‌های چندامیکس و تولید کلان داده‌های زیستی، افق‌های جدیدی را در ژنتیک نوین ایجاد کرد، اما هم‌زمان پیچیدگی تحلیل و تفسیر این داده‌ها را به یکی از مهم‌ترین چالش‌های پژوهش‌های ژنومی و کاربردهای بالینی تبدیل نمود. در این میان، هوش مصنوعی با بهره‌گیری از الگوریتم‌های یادگیری ماشین، یادگیری عمیق و مدل‌های مولد، به یکی از فناوری‌های کلیدی برای استخراج دانش از داده‌های زیستی و توسعه پزشکی دقیق تبدیل شد. این مطالعه به روش مروری انجام شد. برای گردآوری اطلاعات، مقالات منتشرشده در پایگاه‌های علمی معتبر بین‌المللی در زمینه کاربرد هوش مصنوعی در ژنتیک و ژنومیکس بررسی و تحلیل شدند. نتایج بررسی نشان داد که هوش مصنوعی در طیف گسترده‌ای از حوزه‌های ژنتیک نوین، از جمله تحلیل داده‌های تعیین توالی نسل جدید، تفسیر ژنوم، شناسایی و اولویت‌بندی واریانتهای ژنتیکی، کشف عملکرد ژن‌ها، یکپارچه‌سازی داده‌های چندامیکس، تشخیص بیماری‌های ژنتیکی، پزشکی دقیق، طراحی توالی‌های زیستی، توسعه مدل‌های بنیادین، بهینه‌سازی فناوری‌های ویرایش ژن و تسریع کشف اهداف درمانی، نقش مؤثری ایفا کرد. همچنین مشخص شد که ادغام هوش مصنوعی با داده‌های ژنومی، مولکولی و بالینی، امکان شناسایی الگوهای پیچیده زیستی، افزایش دقت پیش‌بینی‌ها و پشتیبانی از تصمیم‌گیری‌های پژوهشی و بالینی را فراهم ساخت. با این حال، چالش‌هایی نظیر کیفیت و استانداردسازی داده‌ها، قابلیت تفسیر مدل‌ها، سوگیری الگوریتمی، حفظ حریم خصوصی اطلاعات ژنتیکی و ملاحظات اخلاقی و حقوقی همچنان از مهم‌ترین موانع توسعه کاربردهای بالینی این فناوری محسوب شدند. در مجموع، این مطالعه نشان داد که هوش مصنوعی از یک ابزار محاسباتی فراتر رفته و به یکی از ارکان تحول ژنتیک نوین تبدیل شده است. توسعه مدل‌های قابل اعتماد، ایجاد زیرساخت‌های داده‌ای استاندارد، افزایش قابلیت تفسیر الگوریتم‌ها و تقویت همکاری‌های میان‌رشته‌ای، نقش تعیین‌کننده‌ای در گسترش کاربرد ایمن، مسئولانه و اثربخش این فناوری در پژوهش‌های ژنتیکی و پزشکی دقیق آینده خواهند داشت.

کلمات کلیدی: هوش مصنوعی، ژنتیک نوین، ژنومیکس، داده‌های چندامیکس، پزشکی دقیق، ویرایش ژن، CRISPR/Cas9

۱. مقدمه

در سال‌های اخیر، رشد سریع فناوری‌های تعیین توالی نسل جدید (NGS)^۱ و توسعه فناوری‌های امیکس، حجم بی‌سابقه‌ای از داده‌های زیستی را در حوزه ژنتیک و ژنومیک تولید کرده است. این داده‌ها اطلاعات ارزشمندی درباره ساختار، عملکرد و تنظیم ژن‌ها، تنوع‌های ژنتیکی و مکانیسم‌های مولکولی بیماری‌ها فراهم می‌کنند، اما حجم، پیچیدگی و ناهمگونی آن‌ها موجب شده است که روش‌های آماری و محاسباتی سنتی در بسیاری از موارد توانایی کافی برای استخراج دانش از این داده‌ها نداشته باشند [1, 2]. در چنین شرایطی، هوش مصنوعی (AI)^۲ به‌عنوان یکی از مهم‌ترین فناوری‌های تحول‌آفرین قرن حاضر، جایگاه ویژه‌ای در علوم زیستی و پزشکی یافته است. هوش مصنوعی شاخه‌ای از علوم رایانه است که با استفاده از الگوریتم‌هایی نظیر یادگیری ماشین^۳ و یادگیری عمیق^۴، امکان تحلیل داده‌های پیچیده، شناسایی الگوهای پنهان، پیش‌بینی نتایج و تصمیم‌گیری هوشمند را فراهم می‌کند [3]. پیشرفت سریع این فناوری موجب شده است که کاربردهای آن از حوزه‌هایی مانند پردازش تصویر و زبان طبیعی فراتر رفته و به بخش‌های مختلف علوم پزشکی، از جمله ژنتیک، ژنومیک و پزشکی دقیق گسترش یابد [4].

هم‌زمان، ژنتیک نوین نیز از مطالعه صرف توالی DNA فراتر رفته و با ظهور فناوری‌های چندامیکس، امکان بررسی هم‌زمان ژنوم، اپی‌ژنوم، ترنسکرپتوم، پروتئوم و سایر لایه‌های تنظیم زیستی فراهم شده است. این رویکرد جامع، اگرچه درک دقیق‌تری از مکانیسم‌های مولکولی بیماری‌ها ارائه می‌دهد، اما حجم بسیار زیاد داده‌ها و روابط پیچیده میان آن‌ها، تحلیل و تفسیر نتایج را به چالشی مهم تبدیل کرده است [2]. در این شرایط، هوش مصنوعی با توانایی پردازش داده‌های حجیم و کشف روابط غیرخطی، به یکی از ابزارهای کلیدی در تحلیل داده‌های ژنتیکی تبدیل شده است [1]. مطالعات متعدد نشان داده‌اند که الگوریتم‌های هوش مصنوعی می‌توانند در مراحل مختلف پژوهش‌های ژنتیکی نقش مؤثری ایفا کنند. این فناوری در تحلیل داده‌های توالی‌یابی، شناسایی و اولویت‌بندی واریانت‌های بیماری‌زا، پیش‌بینی عملکرد ژن‌ها، تشخیص بیماری‌های ژنتیکی، طراحی درمان‌های شخصی‌سازی شده و افزایش دقت فناوری‌های ویرایش ژن، به‌ویژه سامانه CRISPR-Cas9، کاربرد گسترده‌ای یافته است [3, 4]. همچنین توسعه مدل‌های یادگیری عمیق در سال‌های اخیر، دقت پیش‌بینی ساختار و عملکرد مولکول‌های زیستی و تحلیل شبکه‌های پیچیده ژنی را به‌طور چشمگیری افزایش داده است [2].

با وجود پیشرفت‌های چشمگیر، بهره‌گیری از هوش مصنوعی در ژنتیک با چالش‌هایی نیز همراه است. حفاظت از حریم خصوصی داده‌های ژنتیکی، سوگیری الگوریتم‌ها، قابلیت تفسیر نتایج، استانداردسازی داده‌ها و ملاحظات اخلاقی و قانونی از مهم‌ترین موانع توسعه و کاربرد گسترده این فناوری در پژوهش‌ها و خدمات بالینی محسوب می‌شوند [5]. از سوی دیگر، سرعت بالای پیشرفت فناوری‌های هوش مصنوعی موجب شده است که بسیاری از یافته‌های جدید هنوز به‌صورت یکپارچه و نظام‌مند مورد بررسی قرار نگیرند. بر این اساس، انجام یک مطالعه مروری جامع که آخرین دستاوردهای هوش مصنوعی در حوزه ژنتیک نوین را گردآوری و تحلیل کند، ضروری به نظر می‌رسد. هدف این مقاله، بررسی نقش هوش مصنوعی در تحلیل داده‌های ژنتیکی، پزشکی دقیق و فناوری‌های ویرایش ژن، معرفی مهم‌ترین کاربردها، مزایا و محدودیت‌های این

1- Next-Generation Sequencing

2- Artificial Intelligence

3- Machine Learning

4- Deep Learning

فناوری و تبیین چالش‌ها و چشم‌اندازهای آینده آن در پژوهش‌ها و کاربردهای بالینی است. انتظار می‌رود نتایج این مطالعه بتواند تصویری جامع از تعامل روزافزون هوش مصنوعی و ژنتیک نوین ارائه داده و زمینه را برای توسعه پژوهش‌های آینده و استفاده ایمن، مؤثر و مسئولانه از این فناوری در علوم زیستی و پزشکی فراهم سازد.

۲. یافته‌ها:

۲-۱. ضرورت راهبردی هوش مصنوعی در تحول ژنتیک نوین

گسترش فناوری‌های ژنومی موجب شده است که چالش اصلی پژوهش‌های ژنتیکی از تولید داده به تفسیر داده تغییر یابد. امروزه مسئله اساسی، تولید توالی‌های ژنومی یا داده‌های چندامیکس نیست، بلکه استخراج اطلاعات زیستی معنادار از مجموعه داده‌هایی است که میلیون‌ها و گاهی میلیاردها متغیر را در بر می‌گیرند. در چنین شرایطی، بسیاری از روابط میان ژن‌ها، عناصر تنظیمی، پروتئین‌ها و عوامل محیطی دارای ماهیتی غیرخطی و چندعاملی هستند و استفاده از مدل‌های آماری کلاسیک به تنهایی قادر به آشکارسازی این تعاملات پیچیده نیست. از این رو، هوش مصنوعی به عنوان یک زیرساخت محاسباتی پیشرفته، امکان تحلیل هم‌زمان داده‌های با ابعاد بالا و شناسایی الگوهای زیستی پنهان را فراهم کرده است [6, 7].

یکی از مهم‌ترین دلایل ضرورت استفاده از هوش مصنوعی، توانایی آن در ادغام داده‌های ناهمگون^۱ است. پژوهش‌های ژنتیکی امروزی تنها بر داده‌های ژنومی متکی نیستند، بلکه برای دستیابی به تصویری جامع از وضعیت زیستی هر فرد، اطلاعات حاصل از ترنسکریپتوم، اپی‌ژنوم، پروتئوم، متابولوم، میکروبیوم و سوابق بالینی نیز باید به صورت یکپارچه مورد تحلیل قرار گیرند. الگوریتم‌های یادگیری ماشین و یادگیری عمیق با تلفیق این منابع اطلاعاتی، قادر به مدل‌سازی شبکه‌های پیچیده تنظیم ژن، شناسایی برهم‌کنش‌های مولکولی و پیش‌بینی پیامدهای زیستی در سطح سامانه‌های زنده هستند. قابلیت نقش مهمی در توسعه زیست‌شناسی سامانه‌ها و پزشکی سامانه‌ای ایفا می‌کند [7, 8]. علاوه بر تحلیل داده‌ها، هوش مصنوعی به ابزاری برای تولید دانش زیستی تبدیل شده است. برخلاف روش‌های متداول که عمدتاً بر آزمون فرضیه‌های از پیش تعریف‌شده متکی هستند، مدل‌های پیشرفته هوش مصنوعی می‌توانند بدون پیش‌فرض اولیه، الگوهای ناشناخته، نشانگرهای زیستی جدید و ارتباطات عملکردی میان ژن‌ها را کشف کرده و فرضیه‌های پژوهشی جدیدی تولید کنند. این رویکرد، فرآیند کشف زیستی را از یک روش مبتنی بر مشاهده به یک فرایند داده‌محور و پیش‌بینی‌کننده تغییر داده و سرعت شناسایی اهداف درمانی و عوامل مولکولی بیماری‌ها را به طور قابل توجهی افزایش داده است [6, 9].

ضرورت هوش مصنوعی در پزشکی دقیق نیز بیش از پیش آشکار شده است. انتخاب مناسب‌ترین راهبرد درمانی برای هر بیمار مستلزم تحلیل هم‌زمان هزاران متغیر ژنتیکی، مولکولی، محیطی و بالینی است. فرایندی که از توان پردازشی روش‌های سنتی خارج است. مدل‌های هوش مصنوعی با پیش‌بینی بیماری، طبقه‌بندی واریانت‌های ژنتیکی، برآورد پاسخ به دارو و شناسایی اهداف درمانی، نقش مهمی در تصمیم‌گیری‌های بالینی ایفا می‌کنند و زمینه را برای تحقق پزشکی دقیق و

1- Data Integration

درمان‌های شخصی‌سازی شده فراهم می‌سازند [5, 6]. در مجموع، هوش مصنوعی در ژنتیک نوین تنها یک فناوری کمی برای تحلیل داده‌ها محسوب نمی‌شود، بلکه به زیرساختی راهبردی برای مدیریت، تفسیر و بهره‌برداری از داده‌های زیستی تبدیل شده است. توانایی این فناوری در ادغام داده‌های چندامیکس، کشف روابط پیچیده زیستی، تولید فرضیه‌های جدید و پشتیبانی از تصمیم‌گیری‌های بالینی، آن را به یکی از ارکان اساسی توسعه پژوهش‌های ژنتیکی و پزشکی دقیق در دهه اخیر تبدیل کرده است [7, 10].

۲-۲. نقش داده‌های ژنومی و چندامیکس در توسعه مدل‌های هوش مصنوعی

عملکرد و دقت مدل‌های هوش مصنوعی در ژنتیک نوین به‌طور مستقیم به کیفیت، تنوع و حجم داده‌هایی وابسته است که برای آموزش و اعتبارسنجی این مدل‌ها مورد استفاده قرار می‌گیرند. توسعه فناوری‌های تعیین توالی نسل جدید (NGS)، تعیین توالی تک‌سلولی^۱ و فناوری‌های چندامیکس، امکان تولید داده‌هایی را فراهم کرده است که لایه‌های مختلف اطلاعات زیستی، از جمله ژنوم، ترنسکریپتوم، اپی‌ژنوم، پروتئوم، متابولوم و میکروبیوم را به‌صورت هم‌زمان در اختیار پژوهشگران قرار می‌دهند. این مجموعه داده‌های گسترده، بستر لازم برای آموزش الگوریتم‌های یادگیری ماشین و یادگیری عمیق را فراهم کرده و موجب افزایش توان این مدل‌ها در شناسایی الگوهای زیستی پیچیده و استخراج دانش از داده‌های چندبعدی شده‌اند [6, 7]. برخلاف روش‌های تحلیلی سنتی که معمولاً هر لایه از داده‌های زیستی را به‌صورت مستقل بررسی می‌کنند، مدل‌های هوش مصنوعی قادرند اطلاعات حاصل از چندین سطح زیستی را به‌طور هم‌زمان ادغام کنند. این رویکرد که با عنوان یکپارچه‌سازی داده‌های چندامیکس^۲ شناخته می‌شود، امکان شناسایی شبکه‌های تنظیم ژنی، مسیرهای پیام‌رسانی سلولی و برهم‌کنش‌های پیچیده میان ژن‌ها، پروتئین‌ها و متابولیت‌ها را فراهم می‌سازد. در نتیجه، مدل‌های هوش مصنوعی می‌توانند تصویری جامع‌تر از سازوکارهای مولکولی بیماری‌ها ارائه داده و دقت تحلیل‌های ژنومی را به‌طور قابل توجهی افزایش دهند [8, 11]. به جدول یک مراجعه کنید.

داده‌های ژنومی همچنین نقش اساسی در توسعه مدل‌های پیش‌بینی‌کننده ایفا می‌کنند. آموزش این مدل‌ها بر روی میلیون‌ها واریانت ژنتیکی، داده‌های بیان ژن و اطلاعات فنوتیپی، امکان پیش‌بینی بیماری‌زایی واریانت‌ها، شناسایی ژن‌های مسئول بیماری، تحلیل ارتباط ژنوتیپ و فنوتیپ و تفسیر نتایج تعیین توالی را فراهم کرده است. علاوه بر این، ترکیب داده‌های ژنتیکی با اطلاعات بالینی، تصویربرداری پزشکی و پرونده‌های الکترونیک سلامت، دقت مدل‌های هوش مصنوعی را در طبقه‌بندی بیماران، پیش‌بینی سیر بیماری و انتخاب درمان‌های شخصی‌سازی شده به‌طور محسوسی افزایش داده است [6, 12].

1- Single-cell sequencing

2- Multi-omics Integration

جدول ۱- ارتباط میان لایه‌های چندامیکس، چالش‌های محاسباتی، رویکردهای هوش مصنوعی و کاربردهای زیستی- بالینی در ژنتیک نوین

لایه آمیکی	مهم‌ترین چالش تحلیلی	الگوریتم‌ها و رویکردهای متداول هوش مصنوعی	خروجی زیستی حاصل از تحلیل	کاربردهای پژوهشی و بالینی
ژنومیک (Genomics)	حجم بسیار زیاد واریانت‌ها، تفسیر واریانت‌های نادر، روابط غیرخطی ژنوتیپ-فنوتیپ	Autoencoder, CNN, Graph Neural Network, مبتنی بر یادگیری ماشین	اولویت‌بندی واریانت‌ها، پیش‌بینی بیماری‌زایی، شناسایی ژن‌های بیماری‌زا	تشخیص بیماری‌های ژنتیکی، غربالگری، پزشکی دقیق
ترنسکرپتومیک (Transcriptomics)	نویز بالا، اثرات Batch، ناهمگنی سلولی	CNN, Autoencoder, Graph Neural Network, خوشه‌بندی مبتنی بر یادگیری ماشین	شناسایی الگوهای بیان ژن، تعیین انواع سلولی، تحلیل مسیرهای زیستی	طبقه‌بندی سرطان، تعیین زیرگروه‌های بیماری، پیش‌بینی پاسخ درمان
اپی‌ژنومیک (Epigenomics)	وابستگی به بافت، تنظیم چندلایه بیان ژن، داده‌های بسیار پراکنده	Transformer, CNN, Graph Neural Network	شناسایی عناصر تنظیمی، مدل‌سازی شبکه‌های تنظیم ژنی، پیش‌بینی وضعیت اپی‌ژنتیکی	کشف بیومارکرها، اپی‌ژنتیکی، تشخیص زود هنگام بیماری
پروتئومیک (Proteomics)	پیچیدگی ساختار و برهم‌کنش‌های پروتئینی، پویایی زمانی	Deep Learning, Graph Neural Network, مدل‌های بنیادین پروتئین	پیش‌بینی ساختار و عملکرد پروتئین، تحلیل شبکه‌های برهم‌کنش	کشف اهداف دارویی، طراحی دارو، پزشکی شخصی
متابولومیک (Metabolomics)	تنوع زیاد متابولیت‌ها، وابستگی به شرایط فیزیولوژیک	یادگیری ماشین، شبکه‌های عصبی، مدل‌های چندوجهی (Multimodal AI)	شناسایی الگوهای متابولیکی، بازسازی مسیرهای متابولیکی	کشف نشانگرهای زیستی، پیش‌بینی اختلالات متابولیک
میکروبیومیکس (Microbiomics)	ناهمگنی شدید داده‌ها، تعاملات پیچیده میزبان-میکروب	Random Forest, Gradient Boosting, Graph Neural Network	مدل‌سازی ارتباط میکروبیوم با بیماری، پیش‌بینی تغییرات اکوسیستم میکروبی	پزشکی دقیق، پیش‌بینی پاسخ به درمان، تشخیص بیماری‌های التهابی

در سال‌های اخیر، پیشرفت مدل‌های یادگیری عمیق و مدل‌های بنیادین^۱ موجب شده است که کاربرد داده‌های ژنومی از تحلیل صرف فراتر رود و به پیش‌بینی ساختار و عملکرد مولکول‌های زیستی، طراحی پروتئین‌ها، شناسایی اهداف دارویی و بهینه‌سازی سامانه‌های ویرایش ژن نیز گسترش یابد. این مدل‌ها با یادگیری روابط پیچیده موجود در پایگاه‌های عظیم ژنومی، قادرند اطلاعاتی را استخراج کنند که با روش‌های آماری متداول قابل دستیابی نیست و از این طریق، فرایند کشف زیستی و توسعه پزشکی دقیق را تسریع می‌کنند [10]. با وجود این ظرفیت‌های گسترده، عملکرد مدل‌های هوش مصنوعی به کیفیت داده‌های آموزشی وابسته است. وجود خطاهای آزمایشگاهی، ناهمگونی داده‌ها، عدم توازن جمعیت‌های مورد مطالعه، تفاوت میان پلتفرم‌های تعیین توالی و فقدان استانداردهای یکپارچه می‌تواند موجب ایجاد سوگیری الگوریتمی، کاهش قابلیت تعمیم مدل‌ها و محدود شدن کاربرد بالینی آن‌ها شود. افزون بر این، حفاظت از حریم خصوصی داده‌های ژنتیکی، رعایت اصول اخلاقی، شفافیت الگوریتم‌ها و ایجاد چارچوب‌های استاندارد برای اشتراک‌گذاری داده‌ها از مهم‌ترین پیش‌نیازهای توسعه پایدار مدل‌های هوش مصنوعی در ژنتیک محسوب می‌شوند [5, 6]. در مجموع، داده‌های ژنومی و چندامیکس نه تنها ماده اولیه توسعه مدل‌های هوش مصنوعی به شمار می‌روند، بلکه کیفیت و تنوع آن‌ها عامل تعیین‌کننده در دقت، قابلیت تعمیم و ارزش بالینی این مدل‌ها است. پیش‌بینی می‌شود با گسترش بانک‌های اطلاعاتی زیستی و توسعه روش‌های پیشرفته یکپارچه‌سازی داده‌ها، نقش این داده‌ها در توسعه نسل آینده سامانه‌های هوشمند ژنتیکی و پزشکی دقیق بیش از پیش افزایش یابد.

۲-۳. روش‌ها و الگوریتم‌های هوش مصنوعی در تحلیل داده‌های ژنتیکی و ژنومیک

پیشرفت هوش مصنوعی در ژنتیک نوین تا حد زیادی مرهون توسعه الگوریتم‌هایی است که توانایی یادگیری از داده‌های زیستی پیچیده و استخراج الگوهای پنهان را دارند. بسته به نوع داده، هدف پژوهش و مسئله زیستی مورد بررسی، از طیف گسترده‌ای از روش‌های یادگیری ماشین، یادگیری عمیق و هوش مصنوعی مولد استفاده می‌شود که هر یک مزایا و محدودیت‌های خاص خود را دارند [6]. الگوریتم‌های یادگیری نظارت‌شده^۲ از پرکاربردترین روش‌ها در ژنتیک هستند و با استفاده از داده‌های برچسب‌گذاری شده برای پیش‌بینی بیماری‌زایی واریانت‌های ژنتیکی، طبقه‌بندی نمونه‌های زیستی، پیش‌بینی پاسخ به درمان و شناسایی ژن‌های مرتبط با بیماری به کار می‌روند. در مقابل، یادگیری بدون نظارت^۳ با خوشه‌بندی داده‌ها و کاهش ابعاد، امکان شناسایی زیرگروه‌های مولکولی، کشف الگوهای ناشناخته و طبقه‌بندی فنوتیپ‌های پیچیده را بدون نیاز به برچسب‌های از پیش تعیین‌شده فراهم می‌کند. همچنین، یادگیری نیمه‌نظارتی و یادگیری انتقالی^۴ در شرایطی که داده‌های برچسب‌دار محدود باشند، عملکرد مدل‌ها را بهبود می‌بخشند [6, 9].

در سال‌های اخیر، یادگیری عمیق به یکی از مهم‌ترین رویکردهای تحلیل داده‌های ژنومی تبدیل شده است. شبکه‌های عصبی کانولوشنی (CNN)، شبکه‌های عصبی بازگشتی (RNN)، حافظه بلندکوتاه مدت (LSTM) و به‌ویژه معماری‌های مبتنی بر Transformer، قابلیت بالایی در استخراج ویژگی‌های پیچیده از توالی‌های DNA و RNA، پیش‌بینی نواحی

1- Foundation Models

2- Supervised Learning

3- Unsupervised Learning

4- Transfer Learning

تنظیمی ژن، شناسایی واریانت‌های عملکردی و مدل‌سازی تعاملات میان ژن‌ها نشان داده‌اند. این معماری‌ها بدون نیاز به استخراج دستی ویژگی‌ها، الگوهای زیستی را مستقیماً از داده‌های خام فرا می‌گیرند و دقت تحلیل‌های ژنومی را به‌طور قابل توجهی افزایش می‌دهند [13, 14]. علاوه بر این، توسعه مدل‌های بنیادین و هوش مصنوعی مولد^۱ افق‌های جدیدی را در ژنتیک گشوده است. این مدل‌ها که بر روی میلیاردها نوکلئوتید، توالی پروتئینی و داده‌های زیستی آموزش دیده‌اند، قادرند ساختار و عملکرد مولکول‌های زیستی را پیش‌بینی کرده، اثر واریانت‌های ژنتیکی را ارزیابی کنند و حتی در طراحی توالی‌های جدید DNA، RNA و پروتئین یا بهینه‌سازی سامانه‌های ویرایش ژن مشارکت داشته باشند. از این رو، نقش هوش مصنوعی از تحلیل داده‌ها فراتر رفته و به طراحی و مهندسی سامانه‌های زیستی نیز گسترش یافته است [10, 15]. در کنار این روش‌ها، پردازش زبان طبیعی (NLP)^۲ نیز جایگاه ویژه‌ای در ژنتیک یافته است. مدل‌های زبانی بزرگ با استخراج اطلاعات از مقالات علمی، گزارش‌های آزمایشگاهی، پرونده‌های الکترونیک سلامت و پایگاه‌های داده ژنتیکی، فرآیند تفسیر واریانت‌ها، شناسایی شواهد بالینی و به‌روزرسانی دانش ژنتیکی را تسهیل می‌کنند. این قابلیت، به‌ویژه در محیط‌های بالینی و پژوهشی، زمان موردنیاز برای تحلیل و تفسیر داده‌ها را به میزان قابل توجهی کاهش داده است [16].

با وجود پیشرفت‌های چشمگیر، استفاده از این الگوریتم‌ها همچنان با چالش‌هایی نظیر تفسیرپذیری محدود مدل‌های عمیق، سوگیری ناشی از داده‌های آموزشی، خطر بیش‌برازش^۳، نیاز به مجموعه داده‌های بزرگ و اعتبارسنجی مستقل نتایج همراه است. بنابراین، توسعه مدل‌های هوش مصنوعی قابل اعتماد، شفاف و قابل تفسیر، همراه با استانداردسازی روش‌های ارزیابی، یکی از مهم‌ترین اولویت‌های پژوهش‌های آینده در ژنتیک و ژنومیکس محسوب می‌شود [6, 9].

۴-۲. کاربرد هوش مصنوعی در تحلیل داده‌های تعیین توالی نسل جدید (NGS) و ژنومیکس

فناوری تعیین توالی نسل جدید (NGS) یکی از مهم‌ترین دستاوردهای ژنومیک نوین است که امکان تولید حجم عظیمی از داده‌های توالی DNA و RNA را با سرعت و دقت بالا فراهم کرده است. اگرچه این فناوری توانسته است مطالعه ژنوم، اگزوم، ترنسکرپتوم و سایر لایه‌های اطلاعات ژنتیکی را متحول سازد، اما تفسیر داده‌های حاصل از آن همچنان یکی از بزرگ‌ترین چالش‌های بیوانفورماتیک به شمار می‌رود. در این زمینه، هوش مصنوعی با خودکارسازی فرایندهای تحلیلی و افزایش دقت تفسیر داده‌ها، به یکی از اجزای کلیدی جریان کاری^۴ تحلیل داده‌های NGS تبدیل شده است [6, 9].

کاربرد هوش مصنوعی در داده‌های NGS از مرحله کنترل کیفیت داده‌ها آغاز می‌شود. الگوریتم‌های یادگیری ماشین می‌توانند خطاهای ناشی از تعیین توالی، آلودگی نمونه‌ها، توالی‌های کم‌کیفیت و داده‌های پرت را شناسایی کرده و کیفیت داده‌های ورودی را پیش از تحلیل‌های پایین‌دستی بهبود دهند. این موضوع موجب افزایش صحت مراحل بعدی، از جمله هم‌ترازی توالی‌ها^۵، شناسایی واریانت‌ها^۶ و حاشیه‌نویسی ژنومی^۷ می‌شود [9]. یکی از مهم‌ترین کاربردهای هوش مصنوعی،

1- Generative AI
 2- Natural Language Processing
 3- Overfitting
 4- Workflow
 5- Sequence Alignment
 6- Variant Calling
 7- Genome Annotation

تفسیر و اولویت‌بندی واریانت‌های ژنتیکی است. هر ژنوم انسانی میلیون‌ها واریانت ژنتیکی را در بر دارد که تنها بخش اندکی از آن‌ها دارای اهمیت بالینی هستند. مدل‌های یادگیری عمیق با تلفیق اطلاعات حاصل از توالی ژنومی، داده‌های عملکردی، حفاظت تکاملی و پایگاه‌های داده بیماری‌ها، می‌توانند احتمال بیماری‌زایی هر واریانت را پیش‌بینی کرده و واریانت‌های دارای اهمیت بالینی را با دقت بیشتری رتبه‌بندی کنند. این قابلیت، زمان موردنیاز برای تفسیر داده‌های تعیین توالی را به‌طور قابل توجهی کاهش داده و دقت تشخیص بیماری‌های ژنتیکی را افزایش داده است [6, 17].

علاوه بر تحلیل واریانت‌ها، هوش مصنوعی در تفسیر داده‌های حاصل از تعیین توالی¹ RNA، تعیین توالی تک‌سلولی² و تعیین توالی فضایی³ نیز نقش مهمی ایفا می‌کند. این الگوریتم‌ها با شناسایی الگوهای بیان ژن، تعیین انواع سلولی، بازسازی مسیرهای تمایز سلولی و تحلیل ناهمگنی بافت‌ها، امکان مطالعه دقیق‌تر مکانیسم‌های مولکولی بیماری‌ها را فراهم کرده‌اند. به‌ویژه در سرطان، بیماری‌های عصبی و بیماری‌های خودایمنی، کاربرد این فناوری‌ها موجب شناسایی زیرگروه‌های سلولی جدید و نشانگرهای زیستی بالقوه شده است [18]. یکی دیگر از روندهای نوظهور، استفاده از مدل‌های بنیادین ژنومی و معماری‌های مبتنی بر Transformer برای تحلیل مستقیم توالی‌های DNA و RNA است. این مدل‌ها که بر روی میلیاردها نوکلئوتید آموزش دیده‌اند، قادرند بدون نیاز به طراحی ویژگی‌های دستی، روابط پیچیده موجود در توالی‌های ژنتیکی را فراگیرند و در پیش‌بینی عناصر تنظیمی، اثر واریانت‌های غیرکدکننده، تعاملات کروماتینی و بیان ژن عملکرد بسیار بالایی از خود نشان دهند. چنین مدل‌هایی نسل جدید ابزارهای تحلیل ژنوم را تشکیل می‌دهند و انتظار می‌رود در آینده بخش عمده‌ای از فرایندهای بیوانفورماتیکی را خودکار کنند [13, 19].

در مجموع، کاربرد هوش مصنوعی در تحلیل داده‌های NGS تنها به افزایش سرعت پردازش محدود نمی‌شود، بلکه موجب ارتقای دقت تشخیص، کاهش خطاهای تفسیری، کشف نشانگرهای زیستی جدید و تسریع انتقال یافته‌های ژنومی به کاربردهای بالینی شده است. با توسعه پایگاه‌های داده ژنومی و ظهور مدل‌های هوش مصنوعی نسل جدید، انتظار می‌رود تحلیل داده‌های تعیین توالی به‌تدریج از یک فرایند صرفاً محاسباتی به سامانه‌ای هوشمند، خودآموز و پشتیبان تصمیم‌گیری در پزشکی دقیق تبدیل شود.

۵-۲. کاربرد هوش مصنوعی در تفسیر ژنوم، شناسایی واریانت‌ها و کشف عملکرد ژن

گسترش فناوری‌های تعیین توالی کل ژنوم (WGS)⁴ و تعیین توالی اگزوم⁵ (WES) موجب شناسایی میلیون‌ها واریانت ژنتیکی در هر فرد شده است. با این حال، تنها بخش کوچکی از این واریانت‌ها دارای پیامدهای عملکردی یا اهمیت بالینی هستند و بخش قابل توجهی از آن‌ها به‌عنوان واریانت‌های با اهمیت نامشخص (VUS)⁶ طبقه‌بندی می‌شوند. از این‌رو، یکی از مهم‌ترین چالش‌های ژنتیک نوین، تفسیر صحیح اثر واریانت‌ها و تعیین ارتباط آن‌ها با فنوتیپ و بیماری است. در سال‌های

1- RNA-Seq
 2- scRNA-seq
 3- Spatial Transcriptomics
 4- Whole Genome Sequencing
 5- Whole Exome Sequencing
 6- Variants of Uncertain Significance

اخیر، هوش مصنوعی به ابزاری کلیدی برای غلبه بر این چالش تبدیل شده و توانسته است سرعت و دقت تفسیر داده‌های ژنومی را به‌طور چشمگیری افزایش دهد [20, 21]. الگوریتم‌های یادگیری ماشین و یادگیری عمیق با بهره‌گیری از اطلاعات حاصل از توالی DNA، داده‌های حفاظت تکاملی، ویژگی‌های ساختاری پروتئین، داده‌های اپی‌ژنومی و پایگاه‌های اطلاعاتی بالینی، قادرند احتمال بیماری‌زایی واریانت‌های ژنتیکی را پیش‌بینی و آن‌ها را بر اساس اهمیت بالینی اولویت‌بندی کنند. این مدل‌ها، به‌ویژه در تحلیل واریانت‌های نادر و واریانت‌های واقع در نواحی غیرکدکننده ژنوم، عملکردی به‌مراتب دقیق‌تر از بسیاری از روش‌های مبتنی بر قواعد سنتی نشان داده‌اند و نقش مهمی در کاهش تعداد واریانت‌های با اهمیت نامشخص ایفا می‌کنند [22, 23].

علاوه بر تفسیر واریانت‌ها، هوش مصنوعی نقش فزاینده‌ای در کشف عملکرد ژن‌ها و عناصر تنظیمی ژنوم ایفا می‌کند. مدل‌های یادگیری عمیق با یادگیری مستقیم الگوهای موجود در توالی DNA، قادرند محل اتصال فاکتورهای رونویسی، عناصر تنظیمی، نواحی تقویت‌کننده^۱، پروموتورها و سایر عناصر سیس-تنظیمی را پیش‌بینی کنند. این قابلیت، درک دقیق‌تری از سازوکارهای تنظیم بیان ژن فراهم کرده و امکان شناسایی مسیرهای مولکولی مرتبط با بیماری را افزایش داده است. در نتیجه، هوش مصنوعی تنها به پیش‌بینی پیامد واریانت‌ها محدود نمی‌شود، بلکه به ابزاری برای استخراج دانش زیستی و شناسایی روابط عملکردی میان ژن‌ها و شبکه‌های تنظیمی تبدیل شده است [20, 24].

در سال‌های اخیر، توسعه مدل‌های یادگیری عمیق مبتنی بر معماری Transformer و مدل‌های بنیادین ژنومی^۲ تحول قابل‌توجهی در تفسیر ژنوم ایجاد کرده است. این مدل‌ها با آموزش بر روی میلیاردها نوکلئوتید، بدون نیاز به استخراج دستی ویژگی‌ها، قادرند اثر واریانت‌های کدکننده و غیرکدکننده، تغییرات بیان ژن و پیامدهای عملکردی جهش‌ها را پیش‌بینی کنند. همچنین، این مدل‌ها امکان تحلیل هم‌زمان داده‌های ژنومی، ترنسکریپتومی و اپی‌ژنومی را فراهم کرده و دقت پیش‌بینی روابط ژنوتیپ-فنوتیپ را بهبود بخشیده‌اند [24, 20]. کاربرد هوش مصنوعی در تفسیر ژنوم، به‌ویژه در پزشکی ژنومی، موجب تسریع فرایند تشخیص بیماری‌های نادر، شناسایی ژن‌های بیماری‌زا، کشف اهداف درمانی جدید و بهبود تصمیم‌گیری بالینی شده است. علاوه بر این، تلفیق پیش‌بینی‌های محاسباتی با داده‌های عملکردی حاصل از آزمون‌های با توان عملیاتی بالا، مانند آزمون‌های چندگانه ارزیابی اثر واریانت‌ها (MAVEs)^۳، دقت طبقه‌بندی واریانت‌ها را افزایش داده و زمینه را برای کاهش تعداد واریانت‌های با اهمیت نامشخص و توسعه پزشکی دقیق فراهم ساخته است. این روند نشان می‌دهد که آینده تفسیر ژنوم بر هم‌افزایی میان داده‌های تجربی و مدل‌های پیشرفته هوش مصنوعی استوار خواهد بود (جدول ۲).

1- Enhancers

2- Genomic Foundation Models

3- Multiplexed Assays of Variant Effect

جدول ۲- مهم ترین وظایف هوش مصنوعی در تفسیر ژنوم، شناسایی واریانت ها و کشف عملکرد ژن

کاربردهای پژوهش و بالینی	الگوریتم ها و رویکردهای متداول هوش مصنوعی	نوع داده ورودی	وظیفه ژنومی
شناسایی دقیق SNV ها، CNV ها و واریانت های ساختاری	شبکه های عصبی عمیق (CNN)، مدل های مبتنی بر یادگیری عمیق، CNN	داده های خام تعیین توالی (FASTQ)، فایل های هم ترازشده (BAM/CRAM)	فراخوانی واریانت ها ^۱
پیش بینی بیماری های ارثی، کاهش واریانت های با اهمیت نامشخص (VUS)، حمایت از تصمیم گیری بالینی	Random Forest، Deep Neural Networks، XGBoost، مدل های Ensemble	واریانت های ژنتیکی، داده های حفاظت تکاملی، پایگاه های داده بالینی، اطلاعات ساختاری پروتئین	تفسیر و اولویت بندی ^۲
شناسایی جهش های بیماری زا، تعیین اهداف درمانی	Transformer، شبکه های عصبی کانولوشنی (CNN)، مدل های بنیادین ژنومی	توالی RNA، توالی پروتئین، داده های اپی ژنومی	پیش بینی اثر عملکردی ^۳
درک مکانیسم های تنظیم ژن، کشف ژن ها و مسیرهای مولکولی جدید	Transformer، CNN، شبکه های عصبی گراف ^۴	داده های ژنوم، اپی ژنوم، ChIP-seq، RNA-seq، ATAC-seq	کشف عملکرد ژن و عناصر تنظیمی
پزشکی دقیق، پیش بینی خطر بیماری، طراحی درمان فردی	یادگیری ماشین نظارت شده، Graph Neural Networks، مدل های چندوجهی	داده های ژنتیکی، فنوتیپی، سوابق بالینی و چندآمیگس	پیش بینی ارتباط ژن ها با بیماری و فنوتیپ
شناسایی ژن ها، اگزون ها، اینترون ها و عناصر عملکردی ژنوم	یادگیری عمیق، مدل های مبتنی بر توالی، Transformer	توالی ژنوم، داده های رونویسی، پایگاه های داده ژنومی	حاشیه نویسی ژنوم ^۵
بهبود کیفیت ژنوم های مرجع، مطالعات ژنومیک مقایسه ای، بهبود تفسیر بالینی واریانت ها، افزایش قابلیت اعتماد مدل های هوش مصنوعی و توسعه پزشکی ژنومی	مدل های یادگیری نظارت شده، یادگیری انتقالی و مدل های تلفیقی	داده های حاصل از آزمون های عملکردی، از جمله آزمون های چندگانه ارزیابی اثر واریانت ها (MAVEs)، و داده های آزمایشگاهی	اعتبارسنجی پیش بینی های محاسباتی

- 1- Variant Calling
- 2- Variant Interpretation and Prioritization
- 3- Functional Effect
- 4- Graph Neural Networks
- 5- Genome Annotation

۶-۲. کاربردهای هوش مصنوعی در تشخیص بیماری‌های ژنتیکی و پزشکی دقیق

کاربرد هوش مصنوعی در پزشکی ژنومی، فراتر از تحلیل داده‌های ژنتیکی، به مرحله تصمیم‌گیری بالینی و ارائه راهکارهای درمانی شخصی‌سازی شده گسترش یافته است. در سال‌های اخیر، الگوریتم‌های یادگیری ماشین و یادگیری عمیق با تلفیق داده‌های ژنومی، فنوتیپی، تصویربرداری پزشکی، سوابق الکترونیک سلامت و سایر اطلاعات بالینی، امکان تشخیص دقیق‌تر بیماری‌های ژنتیکی و پیش‌بینی پیامدهای بالینی را فراهم کرده‌اند. این رویکرد، به‌ویژه در بیماری‌های پیچیده و چندعاملی که عوامل ژنتیکی و محیطی به‌طور هم‌زمان در بروز آن‌ها نقش دارند، موجب افزایش دقت تشخیص و کاهش زمان رسیدن به تشخیص قطعی شده است [5, 6].

یکی از مهم‌ترین کاربردهای هوش مصنوعی، تشخیص بیماری‌های نادر ژنتیکی است. بسیاری از این بیماری‌ها به دلیل تنوع بالای فنوتیپ، ناهمگونی ژنتیکی و شیوع پایین، با تأخیر یا خطای تشخیصی همراه هستند. مدل‌های هوش مصنوعی با ادغام اطلاعات حاصل از تعیین توالی ژنوم یا اگزوم، ویژگی‌های فنوتیپی بیمار و داده‌های پرونده الکترونیک سلامت، قادرند ژن‌های کاندید را اولویت‌بندی کرده و احتمال ارتباط واریانت‌های ژنتیکی با فنوتیپ مشاهده‌شده را برآورد کنند. مطالعات اخیر نشان داده‌اند که این سامانه‌ها می‌توانند فرایند تشخیص بیماری‌های نادر را تسریع کرده و نرخ تشخیص مولکولی را در مقایسه با روش‌های متداول افزایش دهند [10, 22]. در ژنومیک سرطان نیز هوش مصنوعی نقش مهمی در توسعه پزشکی دقیق ایفا می‌کند. ادغام داده‌های ژنومی، ترنسکریپتومی، پروتئومی و آسیب‌شناسی دیجیتال، امکان شناسایی زیرگروه‌های مولکولی تومورها، پیش‌بینی پیش‌آگهی بیماری، انتخاب درمان‌های هدفمند و ارزیابی احتمال پاسخ یا مقاومت به دارو را فراهم ساخته است. مدل‌های یادگیری عمیق با تحلیل هم‌زمان این داده‌ها می‌توانند اهداف درمانی جدید را شناسایی کرده و پزشکان را در انتخاب مناسب‌ترین راهبرد درمانی برای هر بیمار یاری دهند [25, 26].

یکی دیگر از محورهای مهم پزشکی دقیق، پیش‌بینی پاسخ بیماران به درمان و بهینه‌سازی انتخاب دارو است. الگوریتم‌های هوش مصنوعی با تحلیل داده‌های ژنتیکی، بیان ژن، ویژگی‌های مولکولی تومور و اطلاعات فارماکوژنومیک، قادرند احتمال اثربخشی یا بروز عوارض جانبی داروها را پیش‌بینی کنند. این قابلیت، زمینه توسعه درمان‌های فردمحور، کاهش تجویز داروهای غیرمؤثر و افزایش ایمنی درمان را فراهم کرده است. همچنین، مدل‌های یادگیری عمیق در شناسایی مکانیسم‌های مولکولی مقاومت دارویی و معرفی ترکیب‌های درمانی جایگزین، عملکرد امیدوارکننده‌ای نشان داده‌اند [6, 27]. علاوه بر تشخیص و درمان، هوش مصنوعی در توسعه سامانه‌های پشتیبان تصمیم‌گیری بالینی^۱ نیز نقش فزاینده‌ای دارد. این سامانه‌ها با یکپارچه‌سازی اطلاعات ژنتیکی، فنوتیپی و بالینی، می‌توانند در تفسیر یافته‌های ژنتیکی، انتخاب آزمون‌های تکمیلی، پیشنهاد گزینه‌های درمانی و ارزیابی خطر بیماری به پزشکان کمک کنند. در نتیجه، هوش مصنوعی از یک ابزار صرفاً تحلیلی به بستری برای پشتیبانی از تصمیم‌گیری بالینی مبتنی بر شواهد تبدیل شده است و نقش مهمی در تحقق پزشکی دقیق ایفا می‌کند [5, 6].

با وجود این پیشرفت‌ها، انتقال مدل‌های هوش مصنوعی از محیط پژوهش به کاربردهای بالینی همچنان با چالش‌هایی همراه است. قابلیت تفسیرپذیری مدل‌ها، اعتبارسنجی در جمعیت‌های متنوع، سوگیری ناشی از داده‌های آموزشی، استانداردسازی داده‌های ژنومی و بالینی، حفاظت از حریم خصوصی اطلاعات ژنتیکی و انطباق با الزامات قانونی و اخلاقی، از مهم‌ترین عوامل مؤثر بر پذیرش این فناوری در نظام‌های سلامت محسوب می‌شوند. غلبه بر این چالش‌ها، پیش‌شرط استفاده ایمن، مؤثر و پایدار از هوش مصنوعی در تشخیص بیماری‌های ژنتیکی و پزشکی دقیق خواهد بود [26, 5].

۲-۷. هوش مصنوعی در یکپارچه‌سازی داده‌های چندامیکس و زیست‌شناسی سامانه‌ها

اگرچه هر یک از فناوری‌های آمیکس اطلاعات ارزشمندی درباره یک لایه از سامانه‌های زیستی فراهم می‌کنند، بررسی مستقل این داده‌ها معمولاً برای تبیین کامل سازوکارهای مولکولی بیماری‌ها کافی نیست. بسیاری از فرایندهای زیستی حاصل تعامل پویا میان ژنوم، اپی‌ژنوم، ترنسکریپتوم، پروتئوم، متابولوم و عوامل محیطی هستند؛ از این‌رو، در سال‌های اخیر رویکرد یکپارچه‌سازی چندامیکس^۱ به یکی از محورهای اصلی زیست‌شناسی سامانه‌ها تبدیل شده است. در این رویکرد، هوش مصنوعی با ترکیب هم‌زمان داده‌های ناهمگن، امکان بازسازی شبکه‌های مولکولی، شناسایی مسیرهای تنظیمی و مدل‌سازی روابط پیچیده ژنوتیپ-فنوتیپ را فراهم می‌کند و تصویری جامع‌تر از عملکرد سامانه‌های زیستی ارائه می‌دهد. یکی از مهم‌ترین پیشرفت‌های اخیر، توسعه روش‌های ادغام چندوجهی^۲ است که در آن مدل‌های هوش مصنوعی داده‌های حاصل از چندین لایه آمیکس را در یک فضای ویژگی مشترک^۳ نمایش می‌دهند. این رویکرد امکان استخراج هم‌زمان اطلاعات مکمل از داده‌های ژنومی، ترنسکریپتومی، اپی‌ژنومی و پروتئومی را فراهم کرده و سبب افزایش دقت در طبقه‌بندی بیماری‌ها، کشف زیرگروه‌های مولکولی و شناسایی نشانگرهای زیستی جدید شده است. به‌ویژه، مدل‌های مبتنی بر خودرمزگذارهای تغییرپذیر (VAEs)^۴، شبکه‌های عصبی گراف (GNNs)^۵ و ترنسفورمرهای چندوجهی^۶ در سال‌های اخیر عملکرد قابل‌توجهی در ادغام داده‌های چندامیکی نشان داده‌اند.

از دیدگاه زیست‌شناسی سامانه‌ها، هوش مصنوعی امکان مدل‌سازی شبکه‌های پیچیده برهم‌کنش میان ژن‌ها، پروتئین‌ها، متابولیت‌ها و مسیرهای پیام‌رسانی سلولی را فراهم کرده است. برخلاف روش‌های آماری متداول که اغلب روابط خطی را بررسی می‌کنند، الگوریتم‌های یادگیری عمیق قادرند تعاملات غیرخطی و وابستگی‌های چندسطحی را شناسایی کنند؛ موضوعی که در مطالعه بیماری‌های پیچیده مانند سرطان، بیماری‌های عصبی، بیماری‌های خودایمنی و اختلالات متابولیک اهمیت فراوانی دارد. این مدل‌ها علاوه بر شناسایی شبکه‌های تنظیمی، می‌توانند گره‌های کلیدی^۷، مسیرهای مولکولی بحرانی و اهداف درمانی بالقوه را نیز پیش‌بینی کنند و به تولید فرضیه‌های زیستی جدید کمک نمایند. پیشرفت فناوری‌های چندامیکس تک‌سلولی^۸ نیز افق تازه‌ای در مدل‌سازی سامانه‌های زیستی ایجاد کرده است. ادغام هم‌زمان اطلاعات ژنومی،

1- Multi-omics Integration

2- Multimodal Learning

3- Shared Latent Space

4- Variational Autoencoders

5- Graph Neural Networks

6- Multimodal Transformers

7- Hub Genes

8- Single-cell Multi-omics

اپی ژنومی، رونویسی و پروتئینی در سطح یک سلول، امکان بررسی ناهمگنی سلولی، مسیرهای تمایز سلولها و تعامل میان انواع مختلف سلولی را فراهم می‌کند. مدل‌های هوش مصنوعی با تحلیل این داده‌ها قادرند مسیرهای تکوین سلولی، تغییرات ریزمحیط تومور و مکانیسم‌های پاسخ به درمان را با دقتی بی‌سابقه بازسازی کنند؛ قابلیت‌هایی که در توسعه پزشکی دقیق و طراحی درمان‌های هدفمند اهمیت روزافزونی یافته است.

با وجود این پیشرفت‌ها، یکپارچه‌سازی داده‌های چندآمیک‌های همچنان با چالش‌های محاسباتی مهمی روبه‌رو است. تفاوت در مقیاس اندازه‌گیری، ناهمگونی زیستی، وجود داده‌های مفقود، اثرات بچ^۱، ابعاد بسیار زیاد داده‌ها و محدود بودن نمونه‌های دارای داده‌های کامل، از مهم‌ترین موانع توسعه مدل‌های پایدار و قابل تعمیم محسوب می‌شوند. در سال‌های اخیر، استفاده از مدل‌های مولد، یادگیری انتقالی^۲، یادگیری خودنظارتی^۳ و روش‌های یادگیری توضیح‌پذیر^۴ به‌عنوان راهکارهایی برای افزایش پایداری، تفسیرپذیری و قابلیت استفاده بالینی این مدل‌ها مورد توجه قرار گرفته است. در مجموع، تلفیق هوش مصنوعی با داده‌های چندآمیک، زیست‌شناسی سامانه‌ها را از یک رویکرد توصیفی به یک چارچوب پیش‌بینی‌کننده و داده‌محور تبدیل کرده است. انتظار می‌رود توسعه مدل‌های چندوجهی، شبکه‌های زیستی هوشمند و مدل‌های بنیادین زیستی^۵ در سال‌های آینده امکان شبیه‌سازی دقیق‌تر سامانه‌های زیستی، کشف نشانگرهای زیستی جدید و طراحی راهبردهای درمانی کاملاً شخصی‌سازی‌شده را فراهم سازد.

۸-۲. هوش مصنوعی مولد در ژنتیک: طراحی توالی‌های زیستی، مدل‌های بنیادین و زیست‌شناسی مولد

تحول اخیر در هوش مصنوعی با ظهور هوش مصنوعی مولد^۶، افق جدیدی در پژوهش‌های ژنتیک و زیست‌شناسی مولکولی ایجاد کرده است. برخلاف الگوریتم‌های متداول یادگیری ماشین که عمدتاً به طبقه‌بندی، پیش‌بینی یا تحلیل داده‌های موجود می‌پردازند، مدل‌های مولد قادرند بر اساس الگوهای آموخته‌شده، توالی‌های زیستی جدید، ساختارهای مولکولی و حتی فرضیه‌های پژوهشی نو تولید کنند. این تغییر پارادایم، هوش مصنوعی را از ابزاری برای تحلیل داده‌ها به بستری برای طراحی زیستی^۷ و مهندسی سامانه‌های زیستی تبدیل کرده است [28, 2]. در سال‌های اخیر، مدل‌های بنیادین زیستی^۸ که بر پایه معماری ترنسفورمرها^۹ توسعه یافته‌اند، توانسته‌اند با آموزش بر میلیاردها نوکلئوتید و اسید آمینه، نمایش‌های نهفته^{۱۰} از زبان زیستی ایجاد کنند. این مدل‌ها، مشابه مدل‌های زبانی بزرگ در پردازش زبان طبیعی، قادرند ویژگی‌های تکاملی، ساختاری و عملکردی توالی‌های DNA، RNA و پروتئین را بدون نیاز به ویژگی‌های از پیش تعریف‌شده استخراج کنند. مدل‌هایی مانند Nucleotide Transformer، ESM، Evo و DNABERT نشان داده‌اند که می‌توانند در طیف وسیعی از مسائل

-
- 1- Batch Effects
 - 2- Transfer Learning
 - 3- Self-supervised Learning
 - 4- Explainable AI
 - 5- Foundation Models
 - 6- Generative Artificial Intelligence
 - 7- Biological Design
 - 8- Biological Foundation Models
 - 9- Transformers
 - 10- Latent Representations

ژنومیک، از پیش‌بینی اثر واریانت‌ها تا شناسایی عناصر تنظیمی و پیش‌بینی بیان ژن، عملکرد قابل توجهی داشته باشند، [29, 30]

یکی از مهم‌ترین کاربردهای هوش مصنوعی مولد، طراحی توالی‌های زیستی است. مدل‌های مولد با استفاده از شبکه‌های مولد تخصصی (GANs)، خودرمزگذارهای تغییرپذیر (VAEs)، مدل‌های انتشار^۱ و ترنسفورمرهای مولد، قادرند توالی‌های جدید DNA، RNA و پروتئین را با ویژگی‌های عملکردی مشخص تولید کنند. این فناوری امکان طراحی پروموتورها، تقویت‌کننده‌های رونویسی، ریبوسوئیچ‌ها، RNAهای تنظیمی و پروتئین‌های مصنوعی را فراهم کرده و زمان موردنیاز برای طراحی مولکول‌های زیستی را به‌طور چشمگیری کاهش داده است [31, 32]. در حوزه ژنتیک عملکردی، مدل‌های مولد به ابزاری قدرتمند برای پیش‌بینی اثر تغییرات ژنتیکی تبدیل شده‌اند. این مدل‌ها می‌توانند پیامد جهش‌های نوکلئوتیدی بر بیان ژن، اتصال فاکتورهای رونویسی، آرایش کروماتین و ساختار پروتئین را شبیه‌سازی کنند و از این طریق، تفسیر واریانت‌های با اهمیت نامشخص^۲ (VUS) را بهبود بخشند. همچنین، ترکیب مدل‌های بنیادین با داده‌های حاصل از آزمون‌های چندگانه ارزیابی اثر واریانت‌ها (MAVES) موجب افزایش دقت پیش‌بینی اثر عملکردی جهش‌ها و تسریع شناسایی واریانت‌های بیماری‌زا شده است [33, 17].

کاربرد دیگر هوش مصنوعی مولد، توسعه زیست‌شناسی مولد^۳ است. رویکردی که در آن مدل‌های هوشمند به‌جای تحلیل صرف سامانه‌های زیستی، به طراحی منطقی آن‌ها می‌پردازند. در این چارچوب، الگوریتم‌های هوش مصنوعی قادرند مسیرهای متابولیکی جدید، شبکه‌های تنظیم ژن، مدارهای ژنتیکی مصنوعی و حتی سامانه‌های زیستی با عملکردهای از پیش تعیین شده را پیشنهاد دهند. این قابلیت، فرصت‌های نوینی برای زیست‌شناسی ترکیبی^۴، مهندسی ژنتیک، طراحی واکسن، تولید آنزیم‌های نو ترکیب و توسعه درمان‌های مبتنی بر ژن فراهم کرده است [28, 312].

با وجود پیشرفت‌های چشمگیر، استفاده از هوش مصنوعی مولد در ژنتیک با چالش‌هایی نیز همراه است. قابلیت تفسیر محدود برخی مدل‌های بنیادین، نیاز به مجموعه داده‌های بسیار بزرگ، سوگیری داده‌های آموزشی، هزینه‌های بالای محاسباتی و ضرورت اعتبارسنجی آزمایشگاهی خروجی‌های تولیدشده، از مهم‌ترین محدودیت‌های این فناوری محسوب می‌شوند. علاوه بر این، طراحی توالی‌های زیستی جدید، پرسش‌های مهمی در زمینه ایمنی زیستی^۵، امنیت زیستی^۶، مالکیت فکری و چارچوب‌های اخلاقی ایجاد کرده است که توسعه مسئولانه این فناوری را به یکی از اولویت‌های آینده تبدیل می‌کند [34, 35]. در مجموع، هوش مصنوعی مولد در حال تغییر نقش هوش مصنوعی از یک فناوری تحلیلی به یک فناوری طراحی‌کننده در علوم زیستی است. انتظار می‌رود توسعه مدل‌های بنیادین زیستی و سامانه‌های مولد چندوجهی، در سال‌های آینده طراحی هدفمند توالی‌های زیستی، مهندسی سامانه‌های ژنتیکی و توسعه درمان‌های شخصی‌سازی شده را با سرعت و دقت بی‌سابقه‌ای امکان‌پذیر سازد.

1- Diffusion Models
 2- Variants of Uncertain Significance
 3- Generative Biology
 4- Synthetic Biology
 5- Biosafety
 6- Biosecurity

۹-۲. کاربرد هوش مصنوعی در بهینه‌سازی فناوری‌های ویرایش ژن

فناوری‌های ویرایش ژن، به‌ویژه سامانه‌های مبتنی بر CRISPR-Cas، طی یک دهه گذشته تحول چشمگیری در پژوهش‌های زیست‌پزشکی، درمان بیماری‌های ژنتیکی و مهندسی ژنوم ایجاد کرده‌اند. با وجود پیشرفت‌های گسترده، چالش‌هایی نظیر انتخاب بهینه توالی راهنمای RNA (gRNA)^۱، وقوع برش‌های خارج از هدف^۲، کارایی متغیر ویرایش، محدودیت‌های انتقال سامانه‌های ویرایشی به سلول‌های هدف و دشواری پیش‌بینی پیامدهای عملکردی ویرایش، همچنان موانع اصلی کاربرد گسترده این فناوری در محیط‌های بالینی به شمار می‌روند. در سال‌های اخیر، هوش مصنوعی به یکی از مهم‌ترین ابزارهای محاسباتی برای غلبه بر این محدودیت‌ها تبدیل شده است و نقش آن از تحلیل داده‌ها به طراحی و بهینه‌سازی سامانه‌های ویرایش ژن گسترش یافته است.

یکی از مهم‌ترین کاربردهای هوش مصنوعی، طراحی و بهینه‌سازی gRNA است. مدل‌های یادگیری عمیق با آموزش بر میلیون‌ها توالی حاصل از آزمایش‌های ویرایش ژن، قادرند کارایی برش در جایگاه هدف^۳ و احتمال ایجاد برش‌های خارج از هدف را به‌طور هم‌زمان پیش‌بینی کنند. این مدل‌ها با در نظر گرفتن ویژگی‌هایی مانند ترکیب نوکلئوتیدی، ساختار ثانویه RNA، دسترسی کروماتینی، وضعیت اپی‌ژنتیکی و ویژگی‌های توالی ژنوم، راهنماهایی با بیشترین اختصاصیت و کمترین احتمال خطا پیشنهاد می‌کنند. در نتیجه، هوش مصنوعی نه تنها دقت طراحی gRNA را افزایش می‌دهد، بلکه تعداد آزمایش‌های تجربی مورد نیاز برای غربالگری را نیز به‌طور قابل‌توجهی کاهش می‌دهد. کاربرد مهم دیگر، پیش‌بینی و کاهش اثرات خارج از هدف است که یکی از چالش‌های اصلی ایمنی در درمان‌های مبتنی بر CRISPR محسوب می‌شود. الگوریتم‌های یادگیری ماشین با ادغام اطلاعات توالی، ویژگی‌های اپی‌ژنومی، ساختار سه‌بعدی کروماتین و داده‌های تجربی حاصل از روش‌هایی مانند GUIDE-seq و CIRCLE-seq، قادرند جایگاه‌های بالقوه برش ناخواسته را پیش‌بینی کرده و احتمال وقوع آن‌ها را رتبه‌بندی کنند. این قابلیت، انتخاب سامانه‌های ویرایشی ایمن‌تر را تسهیل کرده و نقش مهمی در توسعه کاربردهای بالینی فناوری‌های ویرایش ژن دارد.

پیشرفت‌های اخیر نشان داده‌اند که هوش مصنوعی تنها به بهینه‌سازی سامانه CRISPR-Cas9 محدود نیست. بلکه در توسعه نسل‌های جدید فناوری‌های ویرایش ژن نیز نقش اساسی ایفا می‌کند. مدل‌های یادگیری عمیق برای پیش‌بینی کارایی ویرایش باز^۴، ویرایش نخست^۵ و سامانه‌های نوظهور CRISPR طراحی شده‌اند و قادرند با تحلیل ویژگی‌های توالی، احتمال موفقیت ویرایش، ایجاد محصولات جانبی و پیامدهای عملکردی هر تغییر ژنتیکی را پیش‌بینی کنند. این رویکرد امکان انتخاب دقیق‌ترین راهبرد ویرایش را برای هر هدف ژنتیکی فراهم می‌سازد. علاوه بر طراحی gRNA، هوش مصنوعی در مهندسی پروتئین‌های Cas نیز نقش فزاینده‌ای یافته است. ترکیب مدل‌های بنیادین پروتئینی، پیش‌بینی ساختار مبتنی بر یادگیری عمیق و طراحی مولد، امکان شناسایی یا طراحی ویرایشگرهای جدید با اختصاصیت بیشتر، اندازه کوچک‌تر، کارایی بالاتر و ایمنی بهتر را فراهم کرده است. این رویکردها مسیر توسعه نوکلئازهای نسل جدید، ویرایشگرهای باز، ویرایشگرهای

1- Guide RNA
 2- Off-target Effects
 3- On-target Efficiency
 4- Base Editing
 5- Prime Editing

نخست و سامانه‌های نوین انتقال ژن را تسریع کرده و فاصله میان طراحی محاسباتی و اعتبارسنجی آزمایشگاهی را کاهش داده‌اند. کاربرد نوظهور دیگر، استفاده از هوش مصنوعی برای پیش‌بینی پیامدهای عملکردی ویرایش ژن است. مدل‌های محاسباتی می‌توانند پیش از انجام آزمایش، اثر تغییرات ژنتیکی بر بیان ژن، شبکه‌های تنظیمی، مسیرهای پیام‌رسانی سلولی و فنوتیپ سلول را شبیه‌سازی کنند. این قابلیت، انتخاب اهداف درمانی مناسب، طراحی آزمایش‌های دقیق‌تر و کاهش هزینه و زمان توسعه درمان‌های ژنی را امکان‌پذیر ساخته است. در سال‌های اخیر، توسعه مفهوم سلول مجازی^۱ مبتنی بر هوش مصنوعی، چشم‌انداز جدیدی برای ارزیابی پیامدهای ویرایش ژن پیش از اجرای آزمایش‌های زیستی ایجاد کرده است.

با وجود پیشرفت‌های قابل توجه، کاربرد هوش مصنوعی در ویرایش ژن همچنان با چالش‌هایی روبه‌رو است. وابستگی مدل‌ها به کیفیت و تنوع داده‌های آموزشی، محدود بودن داده‌های تجربی برای برخی سامانه‌های نوظهور، قابلیت تفسیر پایین برخی مدل‌های یادگیری عمیق و ضرورت اعتبارسنجی نتایج در شرایط آزمایشگاهی و بالینی، از مهم‌ترین محدودیت‌های موجود هستند. علاوه بر این، استفاده از سامانه‌های هوشمند در طراحی ویرایش‌های ژنومی، ضرورت تدوین چارچوب‌های دقیق اخلاقی، ایمنی زیستی و نظارت قانونی را بیش از پیش آشکار ساخته است. در مجموع، هوش مصنوعی در حال تغییر جایگاه فناوری‌های ویرایش ژن از سامانه‌هایی مبتنی بر آزمون و خطا به سامانه‌هایی پیش‌بینی‌پذیر، دقیق و داده‌محور است. انتظار می‌رود تلفیق هوش مصنوعی با فناوری‌های نوین ویرایش ژن، طراحی درمان‌های ژنی شخصی‌سازی‌شده، کاهش خطرات خارج از هدف و توسعه نسل آینده ابزارهای مهندسی ژنوم را با سرعت و دقت بیشتری امکان‌پذیر سازد.

۱۰-۲. چالش‌های فنی، ملاحظات اخلاقی و الزامات حقوقی کاربرد هوش مصنوعی در ژنتیک

با وجود پیشرفت‌های چشمگیر هوش مصنوعی در تحلیل داده‌های ژنومی، پزشکی دقیق و ویرایش ژن، انتقال این فناوری از محیط‌های پژوهشی به کاربردهای بالینی همچنان با چالش‌های متعددی روبه‌رو است. این چالش‌ها تنها به محدودیت‌های محاسباتی محدود نمی‌شوند، بلکه ابعاد اخلاقی، حقوقی و اجتماعی نیز دارند و می‌توانند بر قابلیت اعتماد، پذیرش و استقرار ایمن سامانه‌های هوش مصنوعی در ژنتیک تأثیر بگذارند. از این‌رو، توسعه مسئولانه این فناوری مستلزم ایجاد تعادل میان نوآوری، ایمنی، شفافیت و حفاظت از حقوق افراد است [36, 37].

یکی از مهم‌ترین چالش‌های فنی، کیفیت و نمایندگی داده‌های آموزشی است. عملکرد مدل‌های هوش مصنوعی به‌طور مستقیم به کیفیت، تنوع و کامل بودن داده‌های ژنومی وابسته است. بسیاری از پایگاه‌های داده ژنتیکی موجود، عمدتاً بر جمعیت‌های اروپایی متمرکز هستند و نمایندگی محدودی از سایر گروه‌های قومی دارند؛ موضوعی که می‌تواند موجب کاهش دقت مدل‌ها در جمعیت‌های کمتر مطالعه‌شده و افزایش سوگیری الگوریتمی شود. علاوه بر این، تفاوت در روش‌های تعیین توالی، استانداردهای پردازش داده، اثرات بچ و ناهمگونی داده‌های چندمرکزی، قابلیت تعمیم‌پذیری مدل‌ها را کاهش می‌دهد و اعتبار نتایج را در محیط‌های بالینی تحت تأثیر قرار می‌دهد [37, 38]. چالش مهم دیگر، قابلیت تفسیرپذیری^۲ مدل‌های هوش مصنوعی است. بسیاری از الگوریتم‌های یادگیری عمیق، به‌ویژه مدل‌های بنیادین و شبکه‌های عصبی عمیق، اگرچه از

1- Virtual Cell
 2- Explainability

دقت بالایی برخوردارند، اما فرآیند تصمیم‌گیری آن‌ها برای پژوهشگران و پزشکان شفاف نیست. در ژنتیک بالینی، که تصمیم‌ها می‌توانند پیامدهای درمانی یا ارثی مهمی داشته باشند، اتکای صرف به مدل‌های جعبه سیاه می‌تواند اعتماد متخصصان و بیماران را کاهش دهد. از این‌رو، توسعه روش‌های هوش مصنوعی توضیح‌پذیر^۱ (XAI) و چارچوب‌های اعتبارسنجی مستقل، به یکی از اولویت‌های اصلی این حوزه تبدیل شده است [39, 36].

از منظر اخلاقی، حفاظت از حریم خصوصی و امنیت داده‌های ژنتیکی یکی از مهم‌ترین دغدغه‌ها محسوب می‌شود. برخلاف بسیاری از داده‌های پزشکی، اطلاعات ژنتیکی ماهیتی منحصربه‌فرد، پایدار و قابل انتساب به اعضای خانواده دارند؛ بنابراین، افشای این داده‌ها می‌تواند پیامدهایی فراتر از فرد، از جمله تبعیض ژنتیکی، نقض محرمانگی خانوادگی و سوءاستفاده‌های احتمالی در حوزه بیمه یا اشتغال به همراه داشته باشد. به همین دلیل، روش‌هایی مانند یادگیری فدرال^۲، رمزنگاری هم‌ریخت^۳، محاسبات امن چندطرفه و حفظ حریم خصوصی تفاضلی^۴ به‌عنوان راهکارهای نوین برای کاهش مخاطرات امنیتی مورد توجه قرار گرفته‌اند [40, 37]. مسئله دیگر، عدالت الگوریتمی و برابری در دسترسی به فناوری‌های هوش مصنوعی است. اگر داده‌های آموزشی یا طراحی الگوریتم‌ها نماینده همه جمعیت‌ها نباشند، سامانه‌های هوش مصنوعی ممکن است برای برخی گروه‌های قومی، جنسیتی یا جغرافیایی عملکرد ضعیف‌تری داشته باشند و در نتیجه، نابرابری‌های موجود در خدمات سلامت را تشدید کنند. از این‌رو، بسیاری از سازمان‌های بین‌المللی بر توسعه الگوریتم‌های منصفانه، ارزیابی مستمر سوگیری و استفاده از مجموعه داده‌های متنوع تأکید کرده‌اند تا عدالت در تصمیم‌گیری‌های ژنتیکی حفظ شود [41, 38].

در کنار چالش‌های فنی و اخلاقی، چارچوب‌های حقوقی و نظارتی نیز نقش تعیین‌کننده‌ای در استقرار ایمن هوش مصنوعی در ژنتیک دارند. در سال‌های اخیر، تصویب مقرراتی مانند قانون هوش مصنوعی اتحادیه اروپا (EU AI Act) و دستورالعمل‌های سازمان‌های بین‌المللی، نشان‌دهنده حرکت به سوی نظام‌های قانون‌گذاری مبتنی بر ارزیابی ریسک است. بر اساس این رویکرد، سامانه‌های هوش مصنوعی مورد استفاده در تشخیص بیماری، تصمیم‌گیری درمانی و تحلیل داده‌های ژنتیکی در گروه سامانه‌های پرخطر قرار می‌گیرند و باید از نظر ایمنی، قابلیت ردیابی، شفافیت، کیفیت داده، نظارت انسانی و پایش پس از استقرار به‌طور مستمر ارزیابی شوند [42, 43]. از سوی دیگر، توسعه مدل‌های بنیادین و هوش مصنوعی مولد، پرسش‌های جدیدی درباره مالکیت فکری، مسئولیت حقوقی و اعتبار علمی خروجی‌های تولیدشده ایجاد کرده است. هنوز مشخص نیست در صورت بروز خطا در پیش‌بینی یک مدل هوش مصنوعی یا پیشنهاد یک راهبرد درمانی نادرست، مسئولیت حقوقی بر عهده توسعه‌دهنده الگوریتم، ارائه‌دهنده خدمات سلامت یا پزشک معالج خواهد بود. همچنین، استفاده از مدل‌های مولد در طراحی توالی‌های زیستی، ضرورت تدوین استانداردهای جدید در زمینه ایمنی زیستی، امنیت زیستی و کنترل سوءاستفاده‌های احتمالی را بیش از پیش آشکار ساخته است [34, 44]. در مجموع، آینده کاربرد هوش مصنوعی در ژنتیک تنها به توسعه الگوریتم‌های دقیق‌تر وابسته نیست، بلکه مستلزم ایجاد زیرساخت‌های داده‌ای استاندارد، افزایش شفافیت مدل‌ها، تدوین قوانین منسجم، رعایت اصول اخلاق زیستی و تقویت همکاری میان متخصصان ژنتیک، علوم رایانه، حقوق،

1- Explainable AI
 2- Federated Learning
 3- Homomorphic Encryption
 4- Differential Privacy



اخلاق پزشکی و سیاست‌گذاران است. تنها در چنین چارچوبی می‌توان از ظرفیت‌های گسترده هوش مصنوعی برای توسعه پزشکی دقیق، درمان‌های ژنی و پژوهش‌های ژنومیک به شکلی ایمن، عادلانه و مسئولانه بهره‌برداری کرد.

۱۱-۲. چشم‌اندازهای آینده و فناوری‌های نوظهور در همگرایی هوش مصنوعی و ژنتیک

همگرایی هوش مصنوعی و ژنتیک در حال ورود به مرحله‌ای فراتر از توسعه ابزارهای تحلیلی است و به سمت ایجاد سامانه‌های هوشمندی حرکت می‌کند که قادرند داده‌های زیستی را در مقیاس چندلایه تفسیر کرده، فرضیه‌های پژوهشی تولید کنند و حتی بخشی از فرایند طراحی و اجرای آزمایش‌های زیستی را به‌صورت خودکار انجام دهند. در این چشم‌انداز، نقش هوش مصنوعی از یک ابزار کمکی برای تحلیل داده‌ها به یک زیرساخت شناختی برای تولید دانش زیستی تغییر خواهد کرد؛ تغییری که می‌تواند شیوه انجام پژوهش‌های ژنومی، تشخیص بیماری‌ها و توسعه درمان‌های شخصی‌سازی شده را به‌طور اساسی متحول سازد [36, 7].

یکی از مهم‌ترین روندهای آینده، توسعه مدل‌های بنیادین زیستی است. این مدل‌ها با آموزش بر روی میلیاردها نوکلئوتید DNA، توالی‌های RNA، ساختارهای پروتئینی و داده‌های چندامیکس، قادر خواهند بود نمایش‌های^۱ جامعی از سامانه‌های زیستی ایجاد کنند که در طیف گسترده‌ای از وظایف ژنومی قابل استفاده باشند. برخلاف مدل‌های سنتی که برای یک کاربرد مشخص طراحی می‌شوند، مدل‌های بنیادین قابلیت انتقال دانش^۲ و تنظیم مجدد^۳ برای مسائل مختلف از جمله پیش‌بینی اثر واریانت‌ها، شناسایی عناصر تنظیمی، طراحی RNA و تفسیر داده‌های ژنومی را فراهم می‌کنند و انتظار می‌رود به هسته اصلی سامانه‌های هوشمند زیستی در سال‌های آینده تبدیل شوند [13, 34]. روند مهم دیگر، توسعه دوقلوهای دیجیتال زیستی^۴ است. در این رویکرد، با ادغام داده‌های ژنومی، ترنسکریپتومی، پروتئومی، داده‌های تصویربرداری، سوابق الکترونیک سلامت و اطلاعات محیطی، یک مدل دیجیتال پویا از هر فرد ایجاد می‌شود که می‌تواند پاسخ او به داروها، خطر بروز بیماری، پیشرفت اختلالات ژنتیکی و نتایج مداخلات درمانی را پیش‌بینی کند. چنین سامانه‌هایی زمینه‌ساز تحول در پزشکی پیش‌بینانه و تصمیم‌گیری فردمحور خواهند بود و امکان ارزیابی مجازی سناریوهای درمانی را پیش از اجرای بالینی فراهم خواهند کرد [36, 45]. یکی دیگر از فناوری‌های نوظهور، هوش مصنوعی عامل‌محور^۵ است که از مدل‌های زبانی بزرگ فراتر رفته و قادر به برنامه‌ریزی، استدلال و اجرای خودکار زنجیره‌ای از وظایف پژوهشی است. انتظار می‌رود این سامانه‌ها بتوانند با جستجوی خودکار منابع علمی، استخراج داده‌های ژنومی، طراحی تحلیل‌های بیوانفورماتیکی، انتخاب مدل‌های مناسب و تولید گزارش‌های پژوهشی، بخش قابل توجهی از فرایند تحقیقات ژنتیک را تسریع کنند. در آینده، این عامل‌های هوشمند می‌توانند به‌عنوان دستیاران پژوهشی تخصصی در کنار متخصصان ژنتیک و زیست‌شناسی مولکولی فعالیت کنند [46].

1- Representations
2- Transfer Learning
3- Fine-tuning
4- Digital Twins
5- Agentic AI

در حوزه پژوهش‌های آزمایشگاهی نیز مفهوم آزمایشگاه‌های خودران^۱ به سرعت در حال توسعه است. در این سامانه‌ها، الگوریتم‌های هوش مصنوعی با اتصال به تجهیزات رباتیک، سامانه‌های تصویربرداری و ابزارهای توالی‌یابی، چرخه طراحی فرضیه، اجرای آزمایش، تحلیل نتایج و اصلاح فرضیه را به صورت خودکار انجام می‌دهند. چنین رویکردی می‌تواند زمان مورد نیاز برای کشف ژن‌های عملکردی، طراحی سامانه‌های ویرایش ژن و توسعه درمان‌های ژنی را به طور چشمگیری کاهش دهد و بازده پژوهش‌های زیستی را افزایش دهد [47, 48]. از سوی دیگر، افزایش نگرانی‌ها درباره محرمانگی اطلاعات ژنتیکی موجب شده است که توسعه هوش مصنوعی حفظ‌کننده حریم خصوصی^۲ به یکی از محورهای اصلی تحقیقات آینده تبدیل شود. روش‌هایی مانند یادگیری فدرال، یادگیری توزیع‌شده، رمزنگاری هم‌ریخت و محاسبات امن چندطرفه این امکان را فراهم می‌کنند که مدل‌های هوش مصنوعی بدون انتقال مستقیم داده‌های ژنومی آموزش ببینند. این فناوری‌ها علاوه بر کاهش خطر افشای اطلاعات حساس، همکاری میان مراکز پژوهشی و درمانی را نیز تسهیل کرده و امکان ایجاد مدل‌های جهانی با حفظ محرمانگی داده‌ها را فراهم می‌سازند [37, 49].

چشم‌انداز آینده این حوزه همچنین با گسترش هوش مصنوعی چندوجهی^۳ همراه خواهد بود. این سامانه‌ها قادر خواهند بود داده‌های ژنومی، تصاویر پاتولوژی، تصاویر پزشکی، اطلاعات مولکولی، پرونده الکترونیک سلامت و حتی متون علمی را به صورت هم‌زمان تحلیل کنند و از طریق ایجاد نمایش‌های مشترک، تصویری جامع از وضعیت زیستی هر بیمار ارائه دهند. انتظار می‌رود چنین مدل‌هایی نقش مهمی در تشخیص زودهنگام بیماری‌های پیچیده، انتخاب درمان‌های هدفمند و پیش‌بینی دقیق‌تر پیامدهای بالینی ایفا کنند [47]. در کنار این پیشرفت‌های فناورانه، آینده همگرایی هوش مصنوعی و ژنتیک وابسته به توسعه زیرساخت‌های داده‌ای استاندارد، افزایش قابلیت تفسیر مدل‌ها، اعتبارسنجی چندمرکزی، آموزش نیروی انسانی متخصص و تدوین چارچوب‌های حقوقی و اخلاقی هماهنگ خواهد بود. همکاری میان متخصصان ژنتیک، علوم داده، زیست‌فناوری، اخلاق زیستی و سیاست‌گذاران، نقش تعیین‌کننده‌ای در بهره‌برداری ایمن و مسئولانه از این فناوری‌ها خواهد داشت. در مجموع، انتظار می‌رود طی دهه آینده، هوش مصنوعی از یک ابزار تحلیلی به یک شریک شناختی در پژوهش، تشخیص و درمان بیماری‌های ژنتیکی تبدیل شود و زمینه شکل‌گیری نسل جدیدی از پزشکی دقیق، زیست‌شناسی سامانه‌ها و درمان‌های ژنی هوشمند را فراهم سازد [36, 37].

۳. نتیجه‌گیری

پیشرفت هم‌زمان فناوری‌های ژنومیک، تعیین توالی نسل جدید، فناوری‌های چندامیکس و هوش مصنوعی، چشم‌انداز پژوهش‌ها و کاربردهای بالینی ژنتیک را به طور بنیادین دگرگون کرده است. یافته‌های این مطالعه مروری نشان می‌دهد که هوش مصنوعی، به‌ویژه روش‌های یادگیری ماشین، یادگیری عمیق و مدل‌های بنیادین زیستی، با فراهم کردن امکان تحلیل داده‌های زیستی پیچیده و چندبعدی، نقش تعیین‌کننده‌ای در ارتقای دقت، سرعت و مقیاس تحلیل‌های ژنتیکی ایفا می‌کند.

1- Self-driving Laboratories
 2- Privacy-Preserving AI
 3- Multimodal AI

این فناوری نه تنها تفسیر داده‌های ژنومی و شناسایی واریانت‌های بیماری‌زا را تسهیل کرده، بلکه در کشف عملکرد ژن‌ها، یکپارچه‌سازی داده‌های چندآمیگس، توسعه پزشکی دقیق، طراحی مولکول‌های زیستی، بهینه‌سازی فناوری‌های ویرایش ژن و تسریع فرایندهای کشف دارو نیز به یکی از ارکان اصلی پژوهش‌های زیست‌پزشکی تبدیل شده است.

مرور شواهد موجود نشان می‌دهد که تحول کنونی صرفاً ناشی از افزایش توان پردازشی رایانه‌ها نیست، بلکه حاصل تغییر پارادایم در شیوه استخراج دانش از داده‌های زیستی است. در این رویکرد، الگوریتم‌های هوش مصنوعی از سطح ابزارهای تحلیلی فراتر رفته و به سامانه‌هایی برای شناسایی الگوهای پیچیده، تولید فرضیه‌های زیستی، پیش‌بینی پیامدهای مولکولی و پشتیبانی از تصمیم‌گیری‌های پژوهشی و بالینی تبدیل شده‌اند. این تحول، زمینه را برای شناخت سازوکار بیماری‌ها و توسعه درمان‌های هدفمند گشوده است. با وجود این ظرفیت‌های چشمگیر، بهره‌برداری گسترده از هوش مصنوعی در ژنتیک مستلزم رفع چالش‌های مهمی از جمله کیفیت و استانداردسازی داده‌ها، قابلیت تفسیر مدل‌ها، سوگیری الگوریتمی، حفظ حریم خصوصی اطلاعات ژنتیکی، اعتبارسنجی بالینی و تدوین چارچوب‌های اخلاقی و حقوقی شفاف است. موفقیت آینده این حوزه، بیش از هر چیز، به ایجاد زیرساخت‌های داده‌ای قابل اعتماد، توسعه مدل‌های هوش مصنوعی قابل تفسیر، همکاری میان‌رشته‌ای متخصصان ژنتیک، علوم داده، زیست‌فناوری و پزشکی، و تدوین استانداردهای بین‌المللی برای استفاده ایمن و مسئولانه از این فناوری وابسته خواهد بود.

پیام اصلی مقاله آن است که آینده ژنتیک نوین نه در جایگزینی متخصصان با هوش مصنوعی، بلکه در هم‌افزایی دانش زیستی و توان محاسباتی سامانه‌های هوشمند شکل خواهد گرفت. هوش مصنوعی زمانی بیشترین ارزش علمی و بالینی را ایجاد خواهد کرد که در کنار دانش تخصصی پژوهشگران و پزشکان، به‌عنوان ابزاری برای ارتقای کیفیت تصمیم‌گیری، تسریع اکتشافات علمی و توسعه پزشکی دقیق به‌کار گرفته شود. از این رو، سرمایه‌گذاری در توسعه الگوریتم‌های قابل اعتماد، ایجاد پایگاه‌های داده استاندارد، گسترش همکاری‌های بین‌رشته‌ای و آموزش نیروی انسانی متخصص، از مهم‌ترین الزامات تحقق نسل آینده پژوهش‌ها و کاربردهای بالینی در ژنتیک خواهد بود.

مراجع

1. Saremi, A., Abbasi, B., Karimi-MansoorAbad, E., and Ashourian, Y. (2023). Artificial Intelligence in Medical Genetics: Enhancing Genetic Counseling and Patient Care through Advanced Data Analysis. *Sarem Journal of Medical research*, 8(3), 181-193.
2. Ghosh, S., Zhao, X., Alim, M., Brudno, M., and Bhat, M. (2025). Artificial intelligence applied to 'omics data in liver disease: Towards a personalised approach for diagnosis, prognosis and treatment. *Gut*, 74(2), 295-311.
3. Vilhekar, R. S., and Rawekar, A. (2024). Artificial intelligence in genetics. *Cureus*, 16(1).
4. Chen, Y. M., Hsiao, T. H., Lin, C. H., and Fann, Y. C. (2025). Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence. *Journal of Biomedical Science*, 32(1), 16.



5. Duong, D., and Solomon, B. D. (2025). Artificial intelligence in clinical genetics. *European Journal of Human Genetics*, 33(3), 281-288.
6. Maqsood, K., Hagrass, H., and Zabet, N. R. (2024). An overview of artificial intelligence in the field of genomics. *Discover Artificial Intelligence*, 4(1), 9.
7. Li, Y., Stanojevic, S., and Garmire, L. X. (2022). Emerging artificial intelligence applications in spatial transcriptomics analysis. *Computational and Structural Biotechnology Journal*, 20, 2895-2908.
8. Reddy, S. S. P., Francis, D. L., Thirumoorthi, H., Manohar, B., Krishnan, S. A., and Chopra, S. S. (2023). Artificial intelligence in the genomics era: a blessing or a curse. *Journal of Regenerative Biology and Medicine*, 5(3), 228-235.
9. Goecks, J., Jalili, V., Heiser, L. M., and Gray, J. W. (2020). How machine learning will transform biomedicine. *Cell*, 181(1), 92-101.
10. Thomson, T., Li, G., Strilchuk, A., Cui, H., Wang, B., and Li, B. (2026). Harnessing artificial intelligence to advance CRISPR-based genome editing technologies. *Nature Reviews Genetics*, 27, 212-230.
11. Tanaka, M. (2024). From Serendipity to Precision: Integrating AI, Multi-Omics, and Human Models for Personalized Neuropsychiatric Care.
12. Aradhya, S., Facio, F. M., Metz, H., Manders, T., Colavin, A., Kobayashi, Y., ... and Nussbaum, R. L. (2023, September). Applications of artificial intelligence in clinical laboratory genomics. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 193, No. 3, p. e32057). Hoboken, USA: John Wiley and Sons, Inc..
13. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., ... and Kelley, D. R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nature methods*, 18(10), 1196-1203.
14. Olawade, D. B., Kade, A., Egbon, E., Usman, S. O., Fapohunda, O., Ijiwade, J., and Ogbonna, C. E. (2025). Bioinformatics and artificial intelligence in genomic data analysis: current advances and future directions. *Molecular Genetics and Genomics*, 300(1), 111.
15. ZENG, W. Y., ZHENG, S. Y., ZHAO, X. Q., WANG, Z., and ZHANG, Y. J. (2026). Advances in generative biology and bio-inspired artificial intelligence. *Chinese Bulletin of Life Sciences*, 38(2), 248-267.
16. Zhou, J., Li, H., Chen, S., Chen, Z., Han, Z., and Gao, X. (2025). Large language models in biomedicine and healthcare. *npj Artificial Intelligence*, 1(1), 44/.
17. Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J. K., Brock, K., ... and Marks, D. S. (2021). Disease variant prediction with deep generative models of evolutionary data. *Nature*, 599(7883), 91-95.
18. Heumos, L., Schaar, A. C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., ... and Theis, F. J. (2023). Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, 24(8), 550-572.
19. Ji, J., Chen, B., Lou, H., Hong, D., Zhang, B., Pan, X., ... and Yang, Y. (2024). Aligner: Efficient alignment by learning to correct. *Advances in Neural Information Processing Systems*, 37, 90853-90890.
20. Wong, A. K., Sealfon, R. S., Theesfeld, C. L., and Troyanskaya, O. G. (2021). Decoding disease: from genomes to networks to phenotypes. *Nature Reviews Genetics*, 22(12), 774-790.
21. Critical Assessment of Genome Interpretation Consortium. (2024). CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods. *Genome Biology*, 25, 53.



22. Riccio, C., Jansen, M. L., Guo, L., and Ziegler, A. (2024). Variant effect predictors: A systematic review and practical guide. *Human Genetics*, 143(5), 625–634.
23. Jain, S., Bakolitsa, C., Steven, E. B., Radivojac, P., Moul, J., Repo, S., ... and Kim, S. (2024). Critical assessment of genome interpretation consortium. CAGI, the critical assessment of genome interpretation, establishes progress and pprospects for computational genetic variant interpretation methods. *Genome Biology*, 25(1), 1-46.
24. Novakovsky, G., Dexter, N., Libbrecht, M. W., Wasserman, W. W., and Mostafavi, S. (2023). Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Reviews Genetics*, 24, 125–137.
25. Lin, P. C., Tsai, Y. S., Yeh, Y. M., and Shen, M. R. (2022). Cutting-edge AI technologies meet precision medicine to improve cancer care. *Biomolecules*, 12(8), 1133.
26. Donadio, D., and Terry, S. F. (2023). The application of artificial intelligence in the diagnosis of cancer and rare genetic diseases. *Genetic Testing and Molecular Biomarkers*, 27(7), 203-204.
27. Dlamini, Z., Skepu, A., Kim, N., Mkhabele, M., Khanyile, R., Molefi, T., ... and Hull, R. (2022). AI and precision oncology in clinical cancer genomics: From prevention to targeted cancer therapies-an outcomes based patient care. *Informatics in Medicine Unlocked*, 31, 100965.
28. van Breugel, B., Liu, T., Oglic, D., and van der Schaar, M. (2024). Synthetic data in biomedicine via generative artificial intelligence. *Nature Reviews Bioengineering*, 2(12), 991-1004.
29. Dalla-Torre, H., Gonzalez, L., Mendoza-Revilla, J., Lopez Carranza, N., Grzywaczewski, A. H., Oteri, F., ... and Pierrot, T. (2025). Nucleotide transformer: building and evaluating robust foundation models for human genomics. *Nature Methods*, 22(2), 287-297.
30. Hayes, T., Rao, R., Akin, H., Sofroniew, N. J., Oktay, D., Lin, Z., ... and Rives, A. (2025). Simulating 500 million years of evolution with a language model. *Science*, 387(6736), 850-858.
31. Tang, X., Dai, H., Knight, E., Wu, F., Li, Y., Li, T., and Gerstein, M. (2024). A survey of generative AI for de novo drug design: new frontiers in molecule and protein generation. *Briefings in Bioinformatics*, 25(4), bbae338.
32. Zhang, J., Yang, Y., Si, J., Chen, D., Dong, C., and Han, Z. (2025). Artificial intelligence in the discovery and modification of biological elements in medicinal plants. *Medicinal Plant Biology*, 4(1).
33. Livesey, B. J., and Marsh, J. A. (2023). Updated benchmarking of variant effect predictors using deep mutational scanning. *Molecular systems biology*, 19(8), MSB202211474.
34. Notin, P., Rollins, N., Gal, Y., Sander, C., and Marks, D. (2024). Machine learning for functional protein design. *Nature biotechnology*, 42(2), 216-228.
35. Niu, K., Wang, Y., Zou, L., and Huang, J. (2026). When Biology Meets Medicine: A Perspective on Foundation Models. *Advanced Intelligent Discovery*, e202500245.
36. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1), 44-56.
37. Nogel, M. (2025). Sustainable and equal integration of genomics into healthcare. *JOURNAL OF INFRASTRUCUTRE, POLICY AND DEVELOPMENT*, 9(1), No-10090.
38. Ghassemi, M., Oakden-Rayner, L., and Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The lancet digital health*, 3(11), e745-e750.
39. González-Alday, R., García-Cuesta, E., Kulikowski, C. A., and Maojo, V. (2023). A scoping review on the progress, applicability, and future of explainable artificial intelligence in medicine. *Applied Sciences*, 13(19), 10778.
40. Torkzadehmahani, R., Nasirigerdeh, R., Blumenthal, D. B., Kacprowski, T., List, M., Matschinske, J., ... and Baumbach, J. (2022). Privacy-preserving artificial intelligence techniques in biomedicine. *Methods of information in medicine*, 61(S 01), e12-e27.



41. World Health Organization. (2024). Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. World Health Organization.
42. Lognoul, M. (2025). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act–AI Act). *Revue du Droit des Technologies de l'information*, (3-4), 145-189.
43. Food, U. S. (2025). Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions: Guidance for Industry and Food and Drug Administration Staff. Guidance Document. Silver Spring, MD: US Food and Drug Administration.
44. Rauf, S., Shahid, A., Faizan, M., Khalid, M. N., and Amjad, I. (2024). CRISPER/CAS: a potential tool for genomes editing.
45. Ammar, M., and Al-Raweshidy, H. (2023). Integration of Digital Twins and Artificial Intelligence for Classifying Cardiac Ischemia. *Journal of Artificial Intelligence*, 5, 195.
46. Gridach, M., Nanavati, J., Abidine, K. Z. E., Mendes, L., and Mack, C. (2025). Agentic ai for scientific discovery: A survey of progress, challenges, and future directions. *arXiv preprint arXiv:2503.08979*.
47. Rapp, J. T., Bremer, B. J., and Romero, P. A. (2024). Self-driving laboratories to autonomously navigate the protein fitness landscape. *Nature chemical engineering*, 1(1), 97-107.
48. Li, B., Saini, A. K., Hernandez, J. G., and Moore, J. H. (2026). Agentic AI and the rise of in silico team science in biomedical research. *Nature Biotechnology*, 44(5), 711-725.
49. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... and Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, 3(1), 119.