ریشه‌های علمی هوش مصنوعی در زیست‌فناوری: الگوریتم‌ها و روش‌های پایه

در دنیای پرشتاب امروز، هوش مصنوعی (AI) دیگر تنها یک مفهوم علمی-تخیلی نیست، بلکه به یک نیروی محرکه اصلی در پیشرفت‌های علمی و فناورانه تبدیل شده است. از خودروهای خودران گرفته تا سیستم‌های توصیه‌گر هوشمند، ردپای هوش مصنوعی در هر جنبه‌ای از زندگی مدرن به چشم می‌خورد. در این میان، زیست‌فناوری به عنوان یکی از پویاترین و تأثیرگذارترین حوزه‌ها، به طور فزاینده‌ای از قابلیت‌های هوش مصنوعی برای حل چالش‌های پیچیده و تسریع کشفیات علمی بهره می‌برد. این هم‌افزایی، افق‌های جدیدی را در درک ما از سیستم‌های بیولوژیکی، توسعه داروهای نوین، تشخیص بیماری‌ها و حتی مهندسی ژنتیک گشوده است.

هدف این گزارش، ارائه یک مرور عمیق بر ریشه‌های علمی و تکامل الگوریتم‌های هوش مصنوعی است که مسیر پیشرفت زیست‌فناوری را هموار کرده‌اند. این بررسی از مفاهیم بنیادی و الگوریتم‌های کلاسیک آغاز شده، به تحول به سمت شبکه‌های عصبی می‌پردازد و در نهایت، اهمیت روش‌های نسل جدید و نقش سخت‌افزارهای پیشرفته را در گسترش هوش مصنوعی زیستی تحلیل خواهد کرد. این سفر علمی، به درک این موضوع کمک می‌کند که چگونه مبانی ریاضی و محاسباتی هوش مصنوعی، پایه‌های انقلاب کنونی در زیست‌فناوری را بنا نهاده‌اند و چرا حرکت به سوی ترندهای پیشرفته‌تر، اجتناب‌ناپذیر است.

بخش اول: مفاهیم بنیادی هوش مصنوعی: از AI تا یادگیری عمیق

برای درک عمیق‌تر کاربردهای هوش مصنوعی در زیست‌فناوری، ابتدا باید به تعاریف و مفاهیم کلیدی این حوزه پرداخت. هوش مصنوعی، یادگیری ماشین و یادگیری عمیق، سه اصطلاح مرتبط اما با دامنه‌های متفاوت هستند که اغلب به جای یکدیگر استفاده می‌شوند.

تعاریف کلیدی: هوش مصنوعی، یادگیری ماشین، یادگیری عمیق

هوش مصنوعی (**Artificial Intelligence - AI**) یک حوزه گسترده و جامع است که به ماشین‌ها توانایی شبیه‌سازی رفتار هوشمندانه انسان را می‌دهد. هدف اصلی هوش مصنوعی، ساخت سیستم‌هایی است که بتوانند مانند انسان فکر کنند، استدلال کنند، یاد بگیرند و خود را اصلاح کنند.[1] این حوزه شامل طیف وسیعی از فناوری‌ها و رویکردها است که ماشین را قادر می‌سازد وظایف پیچیده‌ای مانند درک زبان طبیعی، تشخیص تصویر، تصمیم‌گیری و حل مسئله را انجام دهد.[1]

هوش مصنوعی با انواع داده‌های ساختاریافته، نیمه‌ساختاریافته و بدون ساختار سروکار دارد و دامنه‌ای بسیار گسترده را پوشش می‌دهد.[1]

یادگیری ماشین (**Machine Learning - ML**) زیرمجموعه‌ای از هوش مصنوعی است که به ماشین‌ها اجازه می‌دهد بدون برنامه‌نویسی صریح، از داده‌های قبلی الگوها را یاد بگیرند و عملکرد خود را بهبود بخشند.[1] در یادگیری ماشین، انسان‌ها با استفاده از داده‌ها به ماشین‌ها آموزش می‌دهند تا یک کار خاص را انجام دهند و نتایج دقیقی ارائه دهند.[1] الگوریتم‌های یادگیری ماشین الگوهای موجود در داده‌ها را تشخیص می‌دهند و پس از دریافت داده‌های جدید، نتایج را پیش‌بینی می‌کنند.[2] این الگوریتم‌ها معمولاً ساختار نسبتاً ساده‌ای دارند، مانند رگرسیون خطی یا درخت تصمیم.[2] مدل‌های یادگیری ماشین اغلب می‌توانند با مجموعه‌داده‌های کوچکتر (در حد هزاران داده) آموزش ببینند و معمولاً تفسیرپذیری بالاتری نسبت به مدل‌های پیچیده‌تر دارند.[2] با این حال، یادگیری ماشین کلاسیک بیشتر به مداخله انسانی برای مهندسی ویژگی‌ها (Feature Engineering)، یعنی تعیین ویژگی‌های مرتبط و سلسله‌مراتب داده‌ها، وابسته است.[2]

یادگیری عمیق (**Deep Learning - DL**) زیرمجموعه‌ای تخصصی از یادگیری ماشین است که از شبکه‌های عصبی مصنوعی (Artificial Neural Networks - ANN) با ساختار لایه‌ای متعدد (لایه‌های پنهان) برای پردازش داده‌های پیچیده استفاده می‌کند.[2] این شبکه‌ها از ساختار عصبی مغز انسان الهام گرفته‌اند و توانایی یادگیری بسیار بیشتری نسبت به مدل‌های استاندارد یادگیری ماشین دارند.[2] ویژگی بارز یادگیری عمیق، قابلیت استخراج خودکار ویژگی (automatic feature engineering) است؛ به این معنا که الگوریتم‌ها می‌توانند ویژگی‌های مهم را مستقیماً از داده‌های خام استخراج کرده و از خطاهای موجود در خروجی برای بهبود خود استفاده کنند، که نیاز به مداخله انسانی را به شدت کاهش می‌دهد.[2] مدل‌های یادگیری عمیق برای عملکرد صحیح به حجم بسیار بیشتری از داده‌ها (معمولاً میلیون‌ها داده) و توان محاسباتی بالا (مانند واحدهای پردازش گرافیکی یا GPU) نیاز دارند.[2] این قابلیت‌ها، یادگیری عمیق را برای کارهای پیچیده مانند تشخیص تصویر، پردازش زبان طبیعی و خودروهای خودران بسیار مناسب ساخته است.[2]

در بررسی این تعاریف، یک روند تکاملی آشکار می‌شود: از سیستم‌هایی که به صراحت برنامه‌ریزی می‌شوند (هوش مصنوعی اولیه) به سیستم‌هایی که از داده‌ها یاد می‌گیرند (یادگیری ماشین) و سپس به سیستم‌هایی که حتی ویژگی‌های مرتبط را به صورت خودکار از داده‌های خام استخراج می‌کنند (یادگیری عمیق). این پیشرفت، نشان‌دهنده حرکت به سمت خودمختاری بیشتر ماشین‌ها در درک و پردازش اطلاعات است. این تحول، به‌ویژه در مواجهه با داده‌های پیچیده و بدون ساختار

زیستی، اهمیت حیاتی دارد، زیرا آینده هوش مصنوعی کمتر به کدنویسی دستی قوانین و بیشتر به طراحی معماری‌هایی متکی خواهد بود که می‌توانند خودشان دانش را از حجم عظیمی از داده‌ها استخراج کنند. این امر به طور مستقیم بر سرعت و مقیاس‌پذیری تحقیقات در زیست‌فناوری تأثیر می‌گذارد.

همچنین، با افزایش پیچیدگی از یادگیری ماشین به یادگیری عمیق، نیاز به داده‌ها به شدت افزایش می‌یابد و تفسیرپذیری مدل‌ها کاهش می‌یابد.[2] یادگیری ماشین با داده‌های کمتر و تفسیرپذیری بیشتر کار می‌کند، در حالی که یادگیری عمیق برای دقت بالا به داده‌های فراوان و قدرت محاسباتی زیاد نیاز دارد. این یک بده‌بستان اساسی است که انتخاب روش مناسب را در کاربردهای زیست‌فناوری تعیین می‌کند. در زیست‌فناوری، که اغلب داده‌های برچسب‌دار گران و کمیاب هستند و تفسیر مدل‌ها برای تصمیم‌گیری‌های بالینی حیاتی است، این بده‌بستان اهمیت ویژه‌ای دارد. انتخاب بین یادگیری ماشین و یادگیری عمیق صرفاً بر اساس دقت نیست، بلکه بر اساس ملاحظات عملی مانند دسترسی به داده، منابع محاسباتی و نیاز به شفافیت در تصمیم‌گیری مدل است.

انواع یادگیری ماشین: نظارت‌شده، نظارت‌نشده، تقویتی

یادگیری ماشین به طور کلی به سه دسته اصلی تقسیم می‌شود که هر یک رویکرد متفاوتی برای یادگیری از داده‌ها دارند:

یادگیری نظارت‌شده:(Supervised Learning)

در یادگیری نظارت‌شده، مدل با استفاده از داده‌های برچسب‌دار آموزش می‌بیند؛ به این معنی که برای هر ورودی، خروجی صحیح و مورد انتظار از قبل مشخص شده است.[0] این فرآیند شبیه به یادگیری یک دانش‌آموز تحت نظارت یک معلم است که پاسخ‌های صحیح را ارائه می‌دهد.[0] هدف، یادگیری یک تابع یا نگاشت ریاضی ($y = f(x)$) است که بتواند خروجی مورد انتظار را برای ورودی‌های جدید پیش‌بینی کند[5].

فرآیند یادگیری نظارت‌شده شامل جمع‌آوری داده‌های برچسب‌دار، پیش‌پردازش آن‌ها (حذف داده‌های ناقص، استانداردسازی)، انتخاب الگوریتم مناسب و سپس آموزش مدل با استفاده از این داده‌ها است.۵ در طول آموزش، الگوریتم پارامترها یا وزن‌ها را تنظیم می‌کند تا تابع پیش‌بینی‌کننده بهترین انطباق را با خروجی‌های واقعی داشته باشد و خطای آموزش را به حداقل برساند[8].

کاربردهای یادگیری نظارت‌شده در طیف وسیعی از حوزه‌ها دیده می‌شود، از جمله طبقه‌بندی (Classification)که به پیش‌بینی خروجی‌های گسسته یا دسته‌بندی داده‌ها به کلاس‌های مشخص می‌پردازد (مانند تشخیص اسپم، تشخیص چهره، تشخیص بیماری، طبقه‌بندی ژن‌ها).۵ همچنین رگرسیون (Regression)برای پیش‌بینی مقادیر پیوسته (مانند پیش‌بینی قیمت سهام، پیش‌بینی آب و هوا، مدل‌سازی سطح بیومارکرها) [0] از دیگر کاربردهای مهم این روش است. مزایای این روش شامل دقت بالا در پیش‌بینی،

کاربردهای متنوع و توانایی یادگیری از داده‌های گذشته است.[۵] با این حال، جمع‌آوری داده‌های برچسب‌دار می‌تواند زمان‌بر و پرهزینه باشد و مدل به داده‌های آموزشی حساس است[5].

یادگیری نظارت‌نشده:(Unsupervised Learning)

در یادگیری نظارت‌نشده، مدل با داده‌های بدون برچسب آموزش می‌بیند؛ به این معنی که هیچ خروجی صحیح از پیش تعیین‌شده‌ای وجود ندارد.[۵] هدف این نوع یادگیری، کشف الگوهای پنهان، ساختارهای ذاتی یا گروه‌بندی موارد بر اساس شباهت‌ها و تفاوت‌ها در داده‌ها است.[۱۰] این روش در لحظه و به صورت بی‌درنگ انجام می‌شود و تمامی داده‌های ورودی در حین یادگیری تجزیه و تحلیل و برچسب‌گذاری می‌شوند[11].

مدل‌های یادگیری بدون ناظر برای سه وظیفه اصلی استفاده می‌شوند: خوشه‌بندی (Clustering) که به گروه‌بندی نقاط داده مشابه به خوشه‌های مختلف می‌پردازد (مانند الگوریتم K-Means)[۱۰] همبستگی (Association Rules)که روابط و الگوهای تکراری بین متغیرها در یک مجموعه داده را پیدا می‌کند (مانند اگر X آنگاه Y)[۱۰]. و کاهش ابعاد (Dimensionality Reduction) که تعداد ویژگی‌ها در داده‌های با ابعاد بالا را کاهش می‌دهد، در حالی که اطلاعات مهم حفظ می‌شود (مانند تحلیل مؤلفه اصلی یا PCA)[۳]. مزایای آن شامل توانایی یافتن الگوهای ناشناخته، سهولت در دسترسی به داده‌های بدون برچسب و کمک به مهندسی ویژگی است.۱۱ با این حال، دقت خروجی آن ممکن است کمتر باشد، زیرا برچسب‌گذاری توسط خود ماشین انجام می‌شود و تفسیر نتایج می‌تواند چالش‌برانگیز باشد10.

یادگیری تقویتی (Reinforcement Learning - RL):

یادگیری تقویتی یک رویکرد یادگیری ماشین است که در آن یک "عامل (Agent) "از طریق تعامل با یک "محیط (Environment) "و دریافت "پاداش (Reward) "یا "جریمه (Penalty)، "یاد می‌گیرد که چگونه تصمیم‌گیری کند تا "سیاست (Policy) "بهینه را برای حداکثر کردن پاداش کلی در طول زمان پیدا کند.[۱۰] این فرآیند شبیه به آموزش یک حیوان است که با انجام کارهای درست پاداش می‌گیرد.

مبانی ریاضی یادگیری تقویتی شامل نظریه احتمال (برای مدل‌سازی عدم قطعیت در محیط و تصمیم‌گیری تحت عدم قطعیت)، جبر خطی (برای نمایش فضای حالت و عمل) و تکنیک‌های بهینه‌سازی (مانند گرادیان کاهشی برای یافتن سیاست بهینه) است.[۱۶] معادله بل‌من (Bellman Equation) یک مفهوم کلیدی در یادگیری تقویتی است که به عامل کمک می‌کند با ارزیابی حالت‌ها و پاداش‌های آینده، تصمیمات بهینه بگیرد[۱]. الگوریتم Q-learning نیز یکی از الگوریتم‌های محبوب یادگیری تقویتی است که از معادله بل‌من برای تخمین مقادیر Q جفت‌های حالت-عمل) استفاده می‌کند تا عامل را در انتخاب بهترین مسیر برای حداکثر کردن پاداش‌های بلندمدت راهنمایی کند.[۱] یادگیری تقویتی در حوزه‌هایی مانند بازی‌ها (مانند AlphaGo) رباتیک، سیستم‌های خودمختار و اتوماسیون صنعتی که نیاز به تصمیم‌گیری سریع بر اساس داده‌های فراوان دارند، کاربرد دارد[18].

در مقایسه یادگیری نظارت‌شده و نظارت‌نشده، مشاهده می‌شود که یادگیری نظارت‌شده با داده‌های برچسب‌دار به دقت بالا دست می‌یابد [5]، اما جمع‌آوری این داده‌ها زمان‌بر و پرهزینه است.[5] در مقابل، یادگیری نظارت‌نشده با داده‌های بدون برچسب به کشف الگوهای پنهان می‌پردازد [11]، اما

دقت خروجی آن کمتر است و تفسیر نتایج چالش‌برانگیز است.[11] این وضعیت، توسعه روش‌های یادگیری نیمه‌نظارتی (Semi-supervised learning) را ضروری کرده است که از ترکیبی از داده‌های برچسب‌دار و بدون برچسب استفاده می‌کنند.[6] این رویکرد به‌ویژه در زیست‌فناوری، جایی که برچسب‌گذاری داده‌های حجیم (مانند تصاویر پزشکی) بسیار دشوار است، اهمیت می‌یابد.

در حالی که یادگیری نظارت‌شده و نظارت‌نشده بیشتر بر روی شناسایی الگوهای ثابت یا گروه‌بندی داده‌ها تمرکز دارند، یادگیری تقویتی به طور خاص برای سیستم‌های پویا و تصمیم‌گیری‌های متوالی در محیط‌های نامطمئن طراحی شده است.[15] سیستم‌های بیولوژیکی ذاتاً پویا و انطباق‌پذیر هستند و اغلب شامل فرآیندهای تصمیم‌گیری متوالی (مانند دوز دارو، طراحی آزمایش، یا مسیرهای تاشدگی پروتئین) می‌شوند. مبانی ریاضی یادگیری تقویتی (نظریه احتمال، جبر خطی، بهینه‌سازی، و به ویژه معادله بلمن و Q-learning) به خوبی برای مدل‌سازی چنین فرآیندهای پاداش‌محور و متوالی مناسب هستند.[1] این موضوع نشان می‌دهد که یادگیری تقویتی پتانسیل عظیمی، هرچند شاید کمتر کاوش‌شده، در کاربردهای پیشرفته زیست‌فناوری فراتر از مثال‌های فعلی دارد. این می‌تواند شامل بهینه‌سازی فرآیندهای پیچیده بیولوژیکی (مانند بهینه‌سازی بیوراکتورها)، طراحی درمان‌های انطباق‌پذیر، یا حتی شبیه‌سازی فرآیندهای تکاملی باشد، جایی که "محیط" خود سیستم بیولوژیکی و "پاداش‌ها" نتایج بیولوژیکی مطلوب هستند.

جدول ۱: تفاوت‌های کلیدی هوش مصنوعی، یادگیری ماشین و یادگیری عمیق

ویژگی / مفهوم	هوش مصنوعی (**AI**)	یادگیری ماشین (**ML**)	یادگیری عمیق (**DL**)
دامنه / گستره	وسیع‌ترین حوزه، شامل هر فناوری که هوش انسانی را شبیه‌سازی کند.[1]	زیرمجموعه‌ای از AI، به ماشین‌ها اجازه یادگیری از داده را می‌دهد.[1]	زیرمجموعه‌ای از ML، با استفاده از شبکه‌های عصبی چندلایه.[2]
رویکرد یادگیری	ایجاد سیستم‌های هوشمند برای انجام وظایف پیچیده.[1]	آموزش ماشین‌ها برای انجام یک کار خاص با شناسایی الگوها در داده.[1]	تحلیل داده‌ها با ساختار منطقی شبیه به مغز انسان، استخراج خودکار ویژگی.[2]

نیاز به داده	با داده‌های ساختاریافته، نیمه‌ساختاریافته و بدون ساختار.[1]	با داده‌های ساختاریافته و نیمه‌ساختاریافته، می‌تواند با داده‌های کمتر کار کند (هزاران).[1]	نیاز به حجم بسیار زیاد داده (میلیون‌ها) برای عملکرد صحیح، مناسب برای داده‌های بدون ساختار.[2]
مداخله انسانی	تمرکز بر ایجاد سیستم هوشمند.[1]	نیاز به مداخله انسانی بیشتر برای مهندسی ویژگی و انتخاب ساختار داده.[2]	نیاز به مداخله انسانی بسیار کم به دلیل مهندسی ویژگی خودکار و خودیادگیری.[2]
توان محاسباتی	متغیر، بسته به پیچیدگی وظیفه.	نسبتاً کمتر، می‌تواند روی CPUهای معمولی اجرا شود.[3]	بسیار بالا، نیاز به سخت‌افزارهای قدرتمند مانند GPU و TPU.[3]
الگوریتم‌های نمونه	سیستم‌های خبره، رباتیک، پردازش زبان طبیعی.	رگرسیون خطی، درخت تصمیم، SVM، جنگل تصادفی، K-Means.[3]	شبکه‌های عصبی کانولوشنی (CNN)، شبکه‌های عصبی بازگشتی (RNN)، ترانسفورمرها.[3]
تفسیرپذیری	متغیر، بسته به رویکرد.	عموماً تفسیرپذیرتر.[3]	معمولاً کمتر تفسیرپذیر (جعبه سیاه).[4]

این جدول یک نمای کلی ساختاریافته و مقایسه‌ای از مفاهیم اصلی هوش مصنوعی، یادگیری ماشین و یادگیری عمیق را ارائه می‌دهد. با توجه به اینکه این سه اصطلاح اغلب به اشتباه به جای یکدیگر استفاده می‌شوند، این جدول به درک تفاوت‌های ظریف و سلسله‌مراتبی آن‌ها به سرعت و وضوح کمک می‌کند. این امر پایه‌ای محکم برای درک مباحث پیچیده‌تر بعدی در گزارش فراهم می‌آورد و به عنوان یک مرجع سریع برای مخاطبان عمل می‌کند.

بخش دوم: الگوریتم‌های کلاسیک در زیست‌فناوری: پایه‌های اولیه

پیش از ظهور یادگیری عمیق و شبکه‌های عصبی پیچیده، زیست‌فناوری از الگوریتم‌های کلاسیک

یادگیری ماشین برای تحلیل داده‌های خود بهره می‌برد. این الگوریتم‌ها، اگرچه ساده‌تر به نظر می‌رسند، اما پایه‌های اولیه کاربرد هوش مصنوعی در این حوزه را بنا نهادند و مفاهیم کلیدی را معرفی کردند.

رگرسیون‌های خطی و لجستیک: مدل‌سازی بیومارکرها (قبل از ۲۰۰۵)

رگرسیون خطی: رگرسیون خطی یکی از ساده‌ترین و پرکاربردترین روش‌های آماری است که برای مدل‌سازی رابطه بین یک متغیر وابسته پیوسته و یک یا چند متغیر مستقل استفاده می‌شود[۱۹]. فرمول پایه آن برای یک متغیر مستقل به صورت $y = a + bx$ است، که در آن y متغیر وابسته x ،متغیر مستقل b ،شیب خط رگرسیون و a عرض از مبدأ است.[۲۰] هدف اصلی، یافتن خطی است که مجموع مربعات تفاوت بین مقادیر مشاهده‌شده و مقادیر پیش‌بینی‌شده را به حداقل برساند (روش حداقل مربعات)[۲۰].

در زیست‌فناوری، رگرسیون خطی اغلب برای رمزگشایی الگوهای پیچیده در مجموعه‌داده‌های بزرگ و کمی‌سازی روابط استفاده می‌شد.[۱۹] به عنوان مثال، در تحلیل داده‌های ژنومی، از آن برای پیش‌بینی نرخ جهش بر اساس نشانگرهای ژنومی شناخته‌شده ($\text{Mutation Rate} = \beta_0 + \beta_1 \times \text{SNP1} + \ldots + \varepsilon$) استفاده می‌شد. همچنین در پروتئومیکس و کشف بیومارکرها، رگرسیون خطی برای همبسته‌سازی سطوح بیان پروتئین با وضعیت بیماری ($\text{Disease Severity} = \beta_0 + \beta_1 \times \text{Protein}_1 + \ldots + \varepsilon$)کاربرد داشت.[۱۹] این مدل‌سازی‌ها به شناسایی پروتئین‌های معنی‌دار آماری و درک نقش بیولوژیکی آن‌ها کمک می‌کرد.

رگرسیون لجستیک: رگرسیون لجستیک یک الگوریتم یادگیری نظارت‌شده است که برای مسائل طبقه‌بندی، به ویژه پیش‌بینی احتمال تعلق یک ورودی به یک کلاس خاص (مانند حضور یا عدم حضور بیماری) استفاده می‌شود.[۲۲] برخلاف رگرسیون خطی که مقادیر پیوسته را پیش‌بینی می‌کند، رگرسیون لجستیک خروجی دودویی (۰ یا ۱) را هدف قرار می‌دهد.۲۲ این مدل، لگاریتم شانس (log-odds) یک پیامد را بر اساس ویژگی‌های ورودی مدل‌سازی می‌کند: $\log(\pi / (1-\pi)) = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m$ که در آن π احتمال وقوع رویداد است[۲۳]. تابع سیگموئید ($\sigma(z) = 1 / (1 + e^{-z})$) نقش کلیدی در این الگوریتم دارد، زیرا خروجی خطی مدل را به یک مقدار احتمال بین ۰ و ۱ تبدیل می‌کند[۲۲].

اگرچه منابع مستقیماً به مثال‌های پیش از ۲۰۰۵ اشاره نمی‌کنند [۲۵]، اما رگرسیون لجستیک به دلیل سادگی پیاده‌سازی و کاربرد گسترده‌اش در پزشکی، برای مدل‌سازی بیومارکرها جهت پیش‌بینی پیامدهای دودویی (مانند تشخیص بیماری، پیش‌بینی بقا) استفاده می‌شد.[۲۳] این الگوریتم امکان ترکیب چندین بیومارکر برای بهبود دقت تشخیصی را فراهم می‌آورد[26].

کاربرد رگرسیون‌های خطی و لجستیک در مدل‌سازی بیومارکرها و تحلیل داده‌های ژنومی [19] پیش از سال ۲۰۰۵، نشان‌دهنده یک تغییر مهم در زیست‌فناوری است. پیش از این، تحلیل‌های بیولوژیکی بیشتر توصیفی بودند. این مدل‌های رگرسیونی امکان کمی‌سازی روابط و پیش‌بینی پیامدها (مانند نرخ جهش یا شدت بیماری) را فراهم آوردند، که گامی اساسی به سوی علم بیولوژیکی داده‌محور و پیش‌بینی‌کننده بود. این گذار، مسیر را برای پذیرش گسترده‌تر روش‌های محاسباتی در زیست‌فناوری

هموار کرد و نشان داد که مدل‌های ریاضی می‌توانند بینش‌های عمیقی از داده‌های پیچیده بیولوژیکی استخراج کنند.

با این حال، هر دو رگرسیون خطی و لجستیک، مدل‌های خطی هستند.[20] در حالی که این سادگی به تفسیرپذیری کمک می‌کند، سیستم‌های بیولوژیکی ذاتاً غیرخطی و بسیار پیچیده هستند. این وابستگی اولیه به مدل‌های خطی، یک محدودیت ضمنی در توانایی مدل‌سازی دقیق تعاملات بیولوژیکی را نشان می‌دهد. ناتوانی مدل‌های خطی در ثبت کامل پیچیدگی‌های غیرخطی سیستم‌های بیولوژیکی، محرک اصلی برای جستجو و توسعه الگوریتم‌های پیشرفته‌تر بود که بتوانند الگوهای پیچیده‌تر را شناسایی و مدل‌سازی کنند.

درخت‌های تصمیم و جنگل تصادفی: طبقه‌بندی ژن‌ها (Chen et al. ۲۰۰۷)

درخت‌های تصمیم (**Decision Trees**): درخت تصمیم یک الگوریتم یادگیری نظارت‌شده است که برای مسائل طبقه‌بندی و رگرسیون استفاده می‌شود.[9] ساختار آن شبیه به یک فلوچارت است که در آن هر گره داخلی یک تست بر روی یک ویژگی را نشان می‌دهد و هر شاخه، یک نتیجه ممکن از آن تست است.[9] این الگوریتم با تقسیم‌بندی بازگشتی مجموعه داده به زیرمجموعه‌های کوچکتر، به یک گره برگ (Leaf Node) می‌رسد که طبقه‌بندی نهایی را ارائه می‌دهد.[27] معیارهایی مانند آنتروپی، بهره اطلاعات (Information Gain) و شاخص جینی (Gini Index) برای انتخاب بهترین ویژگی برای تقسیم‌بندی گره‌ها استفاده می‌شوند تا خلوص گره‌های حاصل افزایش یابد.[27] درختان تصمیم به دلیل ساختار بصری و منطق ساده‌شان، بسیار قابل تفسیر هستند.[28]

جنگل تصادفی (Random Forest): جنگل تصادفی یک روش یادگیری جمعی (Ensemble Learning) است که با ترکیب پیش‌بینی‌های چندین درخت تصمیم مستقل، دقت مدل را بهبود می‌بخشد و از بیش‌برازش (Overfitting)جلوگیری می‌کند.[9] این الگوریتم با استفاده از دو مفهوم اصلی کار می‌کند:
نمونه‌برداری بوت‌استرپ (Bootstrap Sampling): برای هر درخت تصمیم، یک زیرمجموعه تصادفی از داده‌های آموزشی با جایگذاری (با امکان تکرار داده‌ها) انتخاب می‌شود[۳۰]. این کار باعث ایجاد تنوع در مجموعه‌داده‌های آموزشی هر درخت می‌شود.
انتخاب ویژگی تصادفی (Feature Bagging): در هر مرحله تقسیم گره در هر درخت، به جای بررسی تمام ویژگی‌ها، تنها یک زیرمجموعه تصادفی از ویژگی‌ها در نظر گرفته می‌شود[۳۰]. این تصادفی‌سازی، همبستگی بین درختان را کاهش داده و تنوع در مجموعه را تضمین می‌کند.
برای طبقه‌بندی، پیش‌بینی نهایی با رأی‌گیری اکثریت از خروجی درختان انجام می‌شود؛ برای رگرسیون، میانگین پیش‌بینی‌ها گرفته می‌شود[۳۰]. جنگل تصادفی نسبت به یک درخت تصمیم تنها، بسیار قوی‌تر است، به داده‌های نویزی یا پرت حساسیت کمتری دارد و توانایی بالایی در مدیریت داده‌های با ابعاد بالا دارد31.

جنگل تصادفی به دلیل کارایی و دقت بالا، به طور گسترده در مسائل بیوانفورماتیک استفاده شده است.[29] به عنوان مثال، در مطالعه‌ای در سال ۲۰۰۷، چن و همکاران[32] یک معیار اهمیت ویژگی به نام "اهمیت عمق" (depth importance) را برای جنگل تصادفی پیشنهاد کردند که در شناسایی ژن‌های مرتبط با بیماری‌های پیچیده (مانند ژن‌های بیماری آلزایمر از روی توالی پروتئین) مؤثر بود.[29] این الگوریتم‌ها برای طبقه‌بندی ژن‌ها و انتخاب زیرمجموعه‌ای از نشانگرهای ژنتیکی مرتبط با پیش‌بینی بیماری‌های خاص مورد توجه قرار گرفتند.[32]

موفقیت جنگل تصادفی در زیست‌فناوری [29] ریشه در مفهوم "یادگیری جمعی" (Ensemble Learning) دارد. با ترکیب پیش‌بینی‌های چندین درخت تصمیم [30]، این الگوریتم توانست بر نقاط ضعف یک درخت تنها (مانند بیش‌برازش و حساسیت به نویز) غلبه کند. داده‌های بیولوژیکی ذاتاً پرنویز و متغیر هستند؛ بنابراین، رویکردی که نویز را با میانگین‌گیری از چندین مدل مستقل کاهش می‌دهد، برای این حوزه بسیار مناسب بود. این موضوع نشان‌دهنده یک اصل مهم در تحلیل داده‌های بیولوژیکی است: تجمیع اطلاعات از منابع متعدد یا مدل‌های کمی متفاوت، می‌تواند به نتایج بسیار قوی‌تر و قابل اعتمادتر از تکیه بر یک مدل واحد منجر شود. این ایده، اساس بسیاری از روش‌های پیشرفته‌تر در بیوانفورماتیک، از جمله ادغام داده‌های چند-اومیکس، را تشکیل می‌دهد.

درختان تصمیم به خودی خود تفسیرپذیر هستند [28]، و جنگل تصادفی نیز ابزارهایی برای تحلیل اهمیت ویژگی(Feature Importance) فراهم می‌کند.[30] پیشنهاد معیار "اهمیت عمق" توسط چن و همکاران [32] برای شناسایی ژن‌های خطر، نشان‌دهنده تمایل قوی جامعه بیوانفورماتیک در آن زمان به درک چرایی پیش‌بینی‌ها، نه فقط انجام پیش‌بینی، بود. این موضوع در تضاد با ماهیت "جعبه سیاه" مدل‌های یادگیری عمیق‌تر قرار می‌گیرد که بعدها ظهور کردند. این زمینه تاریخی بر ارزش پایدار مدل‌های هوش مصنوعی قابل تفسیر در زیست‌شناسی تأکید می‌کند، به ویژه در تحقیقات بالینی و ترجمه‌ای که درک مکانیسم‌های بیولوژیکی زیربنایی برای کشف دارو و پزشکی شخصی‌سازی‌شده حیاتی است.

خوشه‌بندی K-Means: گروه‌بندی بیان ژن (۲۰۰۸)

الگوریتم **K-Means** یک الگوریتم یادگیری نظارت‌نشده و یکی از محبوب‌ترین روش‌های خوشه‌بندی داده‌ها است.[33] هدف آن گروه‌بندی نقاط داده بدون برچسب به k خوشه مجزا بر اساس شباهت ویژگی‌ها است.[33] این الگوریتم به صورت تکراری عمل می‌کند: ابتدا k نقطه تصادفی را به عنوان مرکز اولیه خوشه‌ها (Centroids) انتخاب می‌کند. سپس هر نقطه داده به نزدیک‌ترین مرکز خوشه (معمولاً با استفاده از فاصله اقلیدسی) اختصاص داده می‌شود. در مرحله بعد، مراکز خوشه‌ها به

عنوان میانگین تمام نقاط داده اختصاص‌یافته به آن خوشه، مجدداً محاسبه می‌شوند. این فرآیند تا زمانی که موقعیت مراکز خوشه‌ها همگرا شوند یا حداکثر تعداد تکرارها انجام شود، تکرار می‌شود.[33] هدف اصلی K-Means، به حداقل رساندن مجموع مربعات فاصله بین نقاط داده و مراکز خوشه‌های اختصاص‌یافته به آن‌ها (WCSS یا Inertia) است، که نشان‌دهنده فشردگی و شباهت نقاط درون یک خوشه است.[33]

K-Means به طور گسترده‌ای برای خوشه‌بندی داده‌های بیان ژن (Gene Expression - GE) استفاده شده است.[35] در مطالعات پیشگامانه RNA-seq در سال ۲۰۰۸ (مانند Marioni et al.,, ۲۰۰۸ و Sultan et al.,, ۲۰۰۸)، از خوشه‌بندی برای بررسی پروفایل‌های بیان ژن و شناسایی گروه‌هایی از ژن‌ها با الگوهای بیان مشابه استفاده شد.[35] این امر به محققان کمک می‌کرد تا ژن‌هایی را که احتمالاً از نظر عملکردی مرتبط یا با هم تنظیم می‌شوند، کشف کنند.[35]

اگرچه K-Means از نظر محاسباتی کارآمد است [36]، اما چالش‌هایی نیز دارد. این الگوریتم نیاز به تعیین قبلی تعداد خوشه‌ها (k) دارد که انتخاب آن می‌تواند دشوار باشد.[13] همچنین، به انتخاب اولیه مراکز خوشه‌ها حساس است و می‌تواند به نتایج ناسازگار منجر شود.[33] با این حال، سادگی و کارایی آن باعث شد که به ابزاری محبوب در تحلیل داده‌های بیولوژیکی، به ویژه در دورانی که حجم داده‌ها رو به افزایش بود، تبدیل شود.

جدول ۲: خلاصه‌ای از الگوریتم‌های کلاسیک و کاربردهای آن‌ها در زیست‌فناوری

الگوریتم	نوع یادگیری	کاربرد اصلی در زیست‌فناوری	سال/دوره زمانی مرتبط	مبانی ریاضیاتی کلیدی
رگرسیون خطی	نظارت‌شده (رگرسیون)	مدل‌سازی داده‌های بیومارکرها، پیش‌بینی نرخ جهش ژنی، شدت بیماری [19]	قبل از ۲۰۰۵	معادله خطی (y=a+bx)، روش حداقل مربعات [20]
رگرسیون لجستیک	نظارت‌شده	مدل‌سازی بیومارکرها برای	قبل از ۲۰۰۵	تابع سیگموئید، لگاریتم شانس

	(طبقه‌بندی)	پیش‌بینی خروجی‌های دودویی (حضور/عدم حضور بیماری) [23]		log-odds) [22])
درخت تصمیم	نظارت‌شده (طبقه‌بندی/رگرسیون)	طبقه‌بندی ژن‌ها [29]	Chen et al. 2007	آنتروپی، بهره اطلاعات، شاخص جینی [27]
جنگل تصادفی	نظارت‌شده (طبقه‌بندی/رگرسیون)	طبقه‌بندی ژن‌ها، شناسایی ژن‌های مرتبط با بیماری [29]	Chen et al. 2007	نمونه‌برداری بوت‌استرپ، انتخاب ویژگی تصادفی، یادگیری جمعی [30]
خوشه‌بندی **K-Means**	نظارت‌نشده (خوشه‌بندی)	گروه‌بندی بیان ژن، شناسایی ژن‌های با الگوهای بیان مشابه [35]	۲۰۰۸	فاصله اقلیدسی، به حداقل رساندن مجموع مربعات درون خوشه‌ای WCSS) [33])

این جدول به صورت فشرده، الگوریتم‌های کلاسیک معرفی‌شده، نوع یادگیری آن‌ها، کاربردهای مشخص در زیست‌فناوری در دوره زمانی مربوطه، و مبانی ریاضیاتی اصلی هر یک را خلاصه می‌کند. این اطلاعات به شنوندگان کمک می‌کند تا یک دید کلی از ابزارهای محاسباتی اولیه در این حوزه به دست آورند.

بخش سوم: تحول به سمت شبکه‌های عصبی: از پرسپترون تا آلفافولد

با افزایش پیچیدگی داده‌های بیولوژیکی و نیاز به مدل‌سازی روابط غیرخطی، جامعه علمی به تدریج به سمت شبکه‌های عصبی مصنوعی متمایل شد. این تحول، از مدل‌های ساده مانند پرسپترون آغاز شد و به اوج خود در پروژه‌های پیشرفته یادگیری عمیق مانند AlphaFold رسید.

پرسپترون: گام‌های اولیه در دهه ۱۹۸۰

پرسپترون (**Perceptron**) به عنوان اولین مدل شبکه عصبی مصنوعی، توسط فرانک روزنبلات در سال ۱۹۵۸ معرفی شد و از نورون‌های بیولوژیکی الهام گرفته بود.[38] این مدل یک الگوریتم یادگیری نظارت‌شده برای طبقه‌بندی‌کننده‌های دودویی است که ورودی‌های عددی را به یکی از دو دسته (مثبت یا منفی) نگاشت می‌کند.[40] مبنای ریاضی پرسپترون شامل یک جمع وزنی از ورودی‌ها به همراه یک بایاس ($z = w_1x_1 + w_2x_2 + ... + w_nx_n + b$) است که سپس از یک تابع فعال‌سازی آستانه‌ای (معمولاً تابع پله‌ای) عبور می‌کند تا خروجی دودویی (۰ یا ۱) تولید شود.[41] وزن‌ها و بایاس در طول آموزش تنظیم می‌شوند تا خطا بین خروجی پیش‌بینی‌شده و خروجی واقعی به حداقل برسد.[41]

با این حال، پرسپترون تک‌لایه دارای محدودیت‌های قابل توجهی بود؛ مهم‌ترین آن عدم توانایی در حل مسائل غیرخطی مانند تابع XOR بود.[39] این محدودیت‌ها منجر به دوره‌ای از رکود در تحقیقات شبکه‌های عصبی شد که به "زمستان هوش مصنوعی" معروف است. اما در دهه ۱۹۸۰، تحقیقات در زمینه شبکه‌های عصبی با کشف الگوریتم پس‌انتشار (Backpropagation) و توسعه پرسپترون‌های چندلایه (Multi-Layer Perceptrons - MLP) احیا شد.[39] این پیشرفت، مسیر را برای مدل‌های پیچیده‌تر و قدرتمندتر هموار کرد.

شبکه‌های چندلایه (MLP): الگوشناسی بیان ژنی در ۲۰۱۲

شبکه‌های چندلایه (Multi-Layer Perceptrons - MLP) اولین شبکه عصبی مصنوعی با ساختار کامل محسوب می‌شوند و توانایی حل مسائل غیرخطی را دارند.[۴۰] یک MLP شامل یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی است که هر لایه به طور کامل به لایه بعدی متصل است.[۲۴] هر گره در لایه‌های پنهان از یک تابع فعال‌سازی غیرخطی (مانند سیگموئید ReLU یا Tanh) استفاده می‌کند که قابلیت مدل‌سازی روابط پیچیده را به شبکه می‌دهد[24].

آموزش MLP از طریق فرآیندی به نام پس‌انتشار (Backpropagation) انجام می‌شود.[۲۴] در این فرآیند، ابتدا داده‌ها از لایه ورودی به سمت لایه خروجی حرکت می‌کنند (انتشار رو به جلو). سپس، خطای بین خروجی پیش‌بینی‌شده و خروجی واقعی محاسبه می‌شود. این خطا سپس به صورت معکوس از لایه خروجی به سمت لایه‌های پنهان و ورودی منتشر می‌شود و گرادیان‌های تابع زیان نسبت به وزن‌ها و بایاس‌ها محاسبه می‌گردد. در نهایت، وزن‌ها و بایاس‌ها با استفاده از الگوریتم بهینه‌سازی گرادیان کاهشی (Gradient Descent)تنظیم می‌شوند تا تابع زیان به حداقل برسد.۲۴ این فرآیند تکراری به شبکه اجازه می‌دهد تا الگوهای پیچیده را یاد بگیرد.

در سال ۲۰۱۲، محققان دانشگاه میسوری یک مدل یادگیری عمیق مبتنی بر MLP و Stacked Denoising Auto-encoder (MLP-SAE) را برای پیش‌بینی بیان ژن از روی ژنوتیپ‌ها پیشنهاد کردند.[46] این مدل با استفاده از یک خودرمزگذار کاهنده نویز انباشته برای استخراج ویژگی‌های مفید و بهره‌گیری از MLP برای پس‌انتشار، توانست عملکرد بهتری نسبت به روش‌های سنتی مانند

رگرسیون Lasso و جنگل تصادفی ارائه دهد.[46] این کاربرد، نشان‌دهنده پتانسیل بالای یادگیری عمیق در تحلیل داده‌های ژنومی و الگوشناسی بیان ژنی بود.[46]

پروژه آلفافولد: پیش‌بینی ساختار پروتئین‌ها با یادگیری عمیق (DeepMind، ۲۰۲۰)

پروژه **AlphaFold** که توسط DeepMind توسعه یافته است، در سال ۲۰۲۰ به عنوان یک نقطه عطف بی‌سابقه در زیست‌شناسی ساختاری شناخته شد.[47] این پروژه، مسئله "تاشدگی پروتئین" را که یکی از چالش‌های بزرگ و حل‌نشده در زیست‌شناسی برای بیش از ۵۰ سال بود، با دقت و سرعت بی‌سابقه‌ای حل کرد.[47] پروتئین‌ها، مولکول‌های حیاتی هستند که تقریباً تمامی فرآیندهای بیولوژیکی را در موجودات زنده کنترل می‌کنند. عملکرد یک پروتئین به شدت به ساختار سه‌بعدی پیچیده‌اش بستگی دارد، اما تعیین این ساختار به صورت تجربی سال‌ها زمان و صدها هزار دلار هزینه نیاز داشت.[47]

AlphaFold با استفاده از الگوریتم‌های یادگیری عمیق، قادر است ساختار سه‌بعدی پروتئین‌ها را تنها از روی توالی اسید آمینه آن‌ها، در عرض چند دقیقه و با دقتی چشمگیر پیش‌بینی کند.[47] این سیستم بر روی پایگاه داده پروتئین (Protein Data Bank - PDB) که شامل بیش از ۲۰۰ هزار ساختار پروتئینی تعیین‌شده تجربی است، آموزش دید.[49] موفقیت AlphaFold به ایجاد پایگاه داده‌ای شامل بیش از ۲۰۰ میلیون ساختار پروتئینی منجر شده است که تقریباً تمامی پروتئین‌های شناخته‌شده در علم را شامل می‌شود.[47] این دستاورد، میلیون‌ها دلار و صدها میلیون سال زمان تحقیق را برای دانشمندان در سراسر جهان صرفه‌جویی کرده و درک ما از تعاملات مولکولی حیات را به طرز چشمگیری افزایش داده است.[47]

موفقیت AlphaFold نشان داد که یادگیری عمیق می‌تواند از یک کنجکاوی نظری به یک ابزار انقلابی در زیست‌شناسی تبدیل شود. این پروژه، پتانسیل تحول‌آفرین هوش مصنوعی را در حل مسائل بنیادی بیولوژیکی که پیش از این غیرقابل حل به نظر می‌رسیدند، به وضوح اثبات کرد. این دستاورد، نه تنها درک ما از بیماری‌ها را عمیق‌تر کرده، بلکه به تسریع کشف دارو و مهندسی پروتئین‌های جدید نیز کمک شایانی کرده است.[47]

با این حال، AlphaFold یک چالش مهم را نیز برجسته می‌کند: معضل "جعبه سیاه" در کاربردهای با تأثیر بالا. در حالی که AlphaFold با دقت بی‌سابقه‌ای ساختار پروتئین‌ها را پیش‌بینی می‌کند، اما لزوماً مکانیسم تاشدگی پروتئین یا پویایی‌های آن را توضیح نمی‌دهد.[49] این مدل یک ساختار واحد را پیش‌بینی می‌کند و ممکن است نتواند تغییرات ساختاری پویا یا اثرات جهش‌های نقطه‌ای

را به طور کامل نشان دهد.[49] این موضوع اهمیت درک محدودیت‌های مدل‌های یادگیری عمیق و نیاز به ترکیب آن‌ها با دانش مکانیکی و تجربی را نشان می‌دهد، به ویژه در حوزه‌هایی مانند طراحی دارو که درک دقیق مکانیسم‌های مولکولی حیاتی است.

بخش چهارم: اهمیت روش‌های نسل جدید یادگیری عمیق و نقش سخت‌افزار

ظهور یادگیری عمیق و توانایی‌های بی‌نظیر آن در پردازش داده‌های پیچیده، به تنهایی کافی نبود. پیشرفت‌های همزمان در سخت‌افزار و زیرساخت‌های محاسباتی، نقش حیاتی در گسترش و موفقیت این روش‌ها در زیست‌فناوری ایفا کردند.

استخراج خودکار ویژگی: مزیت کلیدی یادگیری عمیق

یکی از مهم‌ترین مزایای روش‌های نسل جدید یادگیری عمیق نسبت به الگوریتم‌های یادگیری ماشین سنتی، قابلیت استخراج خودکار ویژگی(Automatic Feature Extraction) است.[2] در یادگیری ماشین سنتی، مهندسان داده باید به صورت دستی ویژگی‌های مرتبط را از داده‌های خام استخراج کنند. این فرآیند، که به آن مهندسی ویژگی (Feature Engineering) گفته می‌شود، زمان‌بر، پرهزینه و نیازمند تخصص عمیق در هر دو حوزه (داده و دامنه کاربرد) است.[52] برای مثال، در تشخیص تصویر، یک مهندس باید به صورت دستی ویژگی‌هایی مانند لبه‌ها، بافت‌ها یا اشکال را برای طبقه‌بندی استخراج کند.[52]

در مقابل، شبکه‌های عصبی عمیق، به ویژه شبکه‌های عصبی کانولوشنی (CNNs) برای تصاویر، قادرند به صورت خودکار ویژگی‌های سلسله‌مراتبی را مستقیماً از داده‌های خام (مانند مقادیر پیکسل) استخراج کنند.[51] لایه‌های اولیه شبکه ممکن است الگوهای اساسی مانند لبه‌ها و بافت‌ها را تشخیص دهند، در حالی که لایه‌های عمیق‌تر این الگوها را ترکیب کرده و ویژگی‌های سطح بالاتر مانند اشکال یا اجزای یک شی را شناسایی می‌کنند.[52] این قابلیت، نیاز به مداخله انسانی را به شدت کاهش می‌دهد و به مدل اجازه می‌دهد تا الگوهای پیچیده‌تر و ظریف‌تری را که ممکن است از دید انسان پنهان بمانند، کشف کند.[51]

مزایای استخراج خودکار ویژگی در زیست‌فناوری بسیار چشمگیر است. داده‌های بیولوژیکی، مانند تصاویر پزشکی (MRI، CT)، داده‌های ژنومیکس و پروتئومیکس (اومیکس)، و داده‌های میکروسکوپی، ذاتاً پیچیده، پرنویز و دارای ابعاد بالا هستند.[4] استخراج ویژگی‌های معنی‌دار از این داده‌ها به صورت دستی تقریباً غیرممکن است. یادگیری عمیق با توانایی استخراج خودکار ویژگی، این چالش را برطرف کرده و به محققان اجازه می‌دهد تا بر روی مدل‌سازی و تفسیر نتایج تمرکز

کنند، نه بر روی فرآیند خسته‌کننده مهندسی ویژگی. این امر منجر به کاهش هزینه‌های محاسباتی، بهبود عملکرد مدل، جلوگیری از بیش‌برازش و امکان استفاده از یادگیری انتقالی (Transfer Learning) شده است.[52]

نقش رایانش ابری و پیشرفت GPU/TPU در میانه دهه ۲۰۱۰

با وجود مزایای فراوان یادگیری عمیق، این الگوریتم‌ها به دلیل ساختار پیچیده و نیاز به آموزش بر روی حجم عظیمی از داده‌ها، به توان محاسباتی بسیار بالایی نیاز دارند.[18] آموزش یک مدل یادگیری عمیق با استفاده از پردازنده‌های مرکزی (CPU) سنتی، ممکن است ماه‌ها یا حتی سال‌ها به طول انجامد.[18] این نیاز مبرم به قدرت محاسباتی، با ظهور و پیشرفت واحدهای پردازش گرافیکی (**GPUs**) و واحدهای پردازش تانسوری (**TPUs**) در میانه دهه ۲۰۱۰ برطرف شد.[18]

GPUها که در ابتدا برای رندرینگ گرافیکی در بازی‌ها طراحی شده بودند، به دلیل معماری موازی خود که قادر به انجام هزاران محاسبه کوچک به صورت همزمان است، برای بارهای کاری هوش مصنوعی و یادگیری عمیق بسیار مناسب تشخیص داده شدند.[18] شرکت‌هایی مانند NVIDIA با توسعه پلتفرم‌هایی مانند CUDA، امکان برنامه‌نویسی GPUها را برای محاسبات عمومی (GPGPU) فراهم آوردند.[55] موفقیت مدل‌هایی مانند AlexNet در چالش ImageNet در سال ۲۰۱۲، که از GPUها برای کاهش زمان آموزش و بهبود دقت طبقه‌بندی تصویر استفاده کرد، نقطه عطفی در پذیرش گسترده GPUها در تحقیقات هوش مصنوعی بود.[18] پس از آن، NVIDIA و AMD با معرفی ویژگی‌های خاص هوش مصنوعی مانند Tensor Cores، مرزهای فناوری GPU را بیشتر گسترش دادند.TPU[18]ها، که توسط گوگل به طور خاص برای بارهای کاری یادگیری عمیق بهینه‌سازی شده‌اند، نیز توان محاسباتی بی‌سابقه‌ای را ارائه کردند.[18]

همزمان با پیشرفت سخت‌افزار، ظهور و گسترش رایانش ابری (**Cloud Computing**) در میانه دهه ۲۰۱۰، دسترسی به این منابع محاسباتی قدرتمند را دموکراتیزه کرد.[18] شرکت‌هایی مانند Google Cloud، Amazon Web Services (AWS) و Microsoft Azure، GPUها و TPUها را به صورت سرویس‌های قابل دسترسی از طریق اینترنت ارائه دادند. این امر نیاز به سرمایه‌گذاری‌های کلان اولیه در سخت‌افزار گران‌قیمت را از بین برد و به سازمان‌های کوچکتر، استارتاپ‌ها و محققان انفرادی در دانشگاه‌ها اجازه داد تا به قدرت محاسباتی لازم برای تحقیقات پیشرفته هوش مصنوعی دسترسی پیدا کنند.[18]

این هم‌افزایی بین توانایی استخراج خودکار ویژگی در یادگیری عمیق و دسترسی بی‌سابقه به قدرت

محاسباتی از طریق GPU/TPU و رایانش ابری، تأثیر عمیقی بر زیست‌فناوری گذاشت. انفجار داده‌های بیولوژیکی حاصل از فناوری‌های توالی‌یابی نسل جدید (NGS)، تصویربرداری با توان بالا و طیف‌سنجی جرمی، نیاز به روش‌های تحلیلی جدیدی را ایجاد کرده بود.GPU [55]ها و رایانش ابری امکان پردازش و تحلیل این حجم عظیم از داده‌ها را با سرعتی بی‌سابقه فراهم آوردند. به عنوان مثال، تحلیل توالی‌یابی ژنوم کامل انسان که پیش از این به زمان زیادی نیاز داشت، اکنون در عرض چند ثانیه با استفاده از GPUها قابل انجام است.[55] این پیشرفت‌ها، زمینه را برای انقلاب داده‌محور در علوم زیستی فراهم کرد و کاربردهای هوش مصنوعی را در بیوانفورماتیک، از جمله هم‌ترازی توالی‌ها، تشخیص واریانت‌ها، شبیه‌سازی دینامیک مولکولی و طراحی دارو به شدت تسریع بخشید.[55]

این تحولات نشان می‌دهد که سخت‌افزار به عنوان توانمندساز خاموش انقلاب هوش مصنوعی در زیست‌شناسی عمل کرده است. رشد نمایی در حجم داده‌های بیولوژیکی با رشد نمایی در قدرت محاسباتی مواجه شد، که این هم‌زمانی، امکان پیشرفت‌های چشمگیر را فراهم آورد. این پیشرفت‌ها منجر به دموکراتیزه شدن قابلیت‌های تحقیقاتی پیشرفته شد. رایانش ابری و سخت‌افزارهای تخصصی، دسترسی به هوش مصنوعی پیشرفته را برای طیف وسیع‌تری از جامعه تحقیقاتی فراهم کردند، که به نوبه خود نوآوری و کشفیات را در سراسر زیست‌فناوری تسریع بخشید.

نتیجه‌گیری: چشم‌انداز هوش مصنوعی در زیست‌فناوری

در این گزارش، مسیر تکاملی هوش مصنوعی در زیست‌فناوری از مبانی نظری تا کاربردهای پیشرفته را بررسی شد. از الگوریتم‌های کلاسیک مانند رگرسیون‌های خطی و لجستیک که پایه‌های مدل‌سازی بیومارکرها را بنا نهادند، تا درختان تصمیم و جنگل‌های تصادفی که طبقه‌بندی ژن‌ها را متحول کردند، هر گام نشان‌دهنده تلاش برای استخراج دانش از داده‌های بیولوژیکی بود. سپس، با ظهور شبکه‌های عصبی، از پرسپترون‌های اولیه تا شبکه‌های چندلایه که الگوشناسی بیان ژنی را بهبود بخشیدند، شاهد گذار به مدل‌های پیچیده‌تر و قدرتمندتر بودیم. اوج این تحول، در پروژه AlphaFold به وضوح نمایان شد که با استفاده از یادگیری عمیق، مسئله تاشدگی پروتئین را حل کرد و افق‌های جدیدی را در زیست‌شناسی ساختاری گشود. قابلیت استخراج خودکار ویژگی در یادگیری عمیق، همراه با پیشرفت‌های سخت‌افزاری چشمگیر در GPU و TPU و دسترسی به رایانش ابری، نقش محوری در تسریع این انقلاب ایفا کرده‌اند. درک این ریشه‌های علمی و تکامل الگوریتم‌ها، برای هر علاقه‌مند به زیست‌فناوری ضروری است تا بتواند پتانسیل کامل این هم‌افزایی را درک کند.

نگاهی به آینده نشان می‌دهد که هوش مصنوعی در زیست‌فناوری تنها در ابتدای راه است. ترندهای آتی شامل استفاده گسترده از هوش مصنوعی مولد (**Generative AI**) برای طراحی داروهای جدید،

مهندسی پروتئین‌ها با عملکردهای خاص و حتی سنتز داده‌های اومیکس مصنوعی خواهد بود.[57] یادگیری تقویتی نیز پتانسیل بالایی در بهینه‌سازی فرآیندهای بیولوژیکی پیچیده و طراحی درمان‌های انطباق‌پذیر دارد. هوش مصنوعی به طور فزاینده‌ای در پزشکی شخصی‌سازی‌شده، از تشخیص زودهنگام بیماری‌ها تا انتخاب درمان‌های هدفمند بر اساس پروفایل ژنتیکی فردی، نقش ایفا خواهد کرد. این تحولات نه تنها سرعت کشفیات علمی را افزایش می‌دهند، بلکه به ارائه راهکارهای درمانی مؤثرتر و دقیق‌تر برای چالش‌های بهداشتی جهانی منجر خواهند شد.

این حوزه در حال تکامل، فرصت‌های بی‌نظیری را برای نوآوری و تأثیرگذاری فراهم می‌کند. از اساتید و دانشجویان گرفته تا علاقه‌مندان به زیست‌فناوری، هر یک می‌توانند نقشی در شکل‌دهی به آینده این علم ایفا کنند. برای همراهی با این انقلاب و مشارکت فعال در آن، ضروری است که دانش خود را در این زمینه تعمیق بخشید، مهارت‌های محاسباتی را تقویت کنید و همواره به دنبال کشف مرزهای جدید باشید. آینده زیست‌فناوری، با هوش مصنوعی گره خورده است و مشارکت شما می‌تواند این آینده را روشن‌تر و پربارتر سازد.

منابع مورداستناد

1. زمان دسترسی: ژوئن Quera, - تفاوت هوش مصنوعی و یادگیری ماشین چیست؟ - کوئرابلاگ 10, 2025، https://quera.org/blog/the-differnce-between-ai-and-machine-learning/
2. تفاوت یادگیری ماشین و یادگیری عمیق چیست؟ آیا این دو یک چیز ...، زمان دسترسی: ژوئن 10, 2025، https://cafetadris.com/blog/%D8%AA%D9%81%D8%A7%D9%88%D8%AA-%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D9%85%D8%A7%D8%B4%DB%8C%D9%86-%D9%88-%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D8%B9%D9%85%DB%8C%D9%82/
3. Difference between AI ML vs DL | A Comparative Analysis - NAV IT Consulting, زمان دسترسی: ژوئن 10, 2025، https://nav-it.com/machine-learning-artificial-intelligence-deep-learning/
4. Deep learning in bioinformatics - PMC - PubMed Central, زمان دسترسی: ژوئن 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC11045206/
5. یادگیری با ناظر (Supervised Learning) چیست ؟ - مزایا و معایب, زمان دسترسی: ژوئن 10, 2025، https://avalai.ir/blog/what-is-supervised-learning/
6. درسمن، شتاب‌دهنده شما برای ورود به بازار کار برنامه نویسی, زمان دسترسی: ژوئن 10, 2025، https://darsman.com/blog/ai/what-is-supervised-learning/

7. Supervised Machine Learning - GeeksforGeeks, زمان دسترسی: ژوئن 10, 2025، https://www.geeksforgeeks.org/supervised-machine-learning/
8. What is Supervised Learning? - Definition and Explaination | C3 AI, زمان دسترسی: ژوئن 10, 2025، https://c3.ai/introduction-what-is-machine-learning/supervised-learning/
9. Classification vs Regression in Machine Learning - GeeksforGeeks, زمان دسترسی: ژوئن 10, 2025، https://www.geeksforgeeks.org/ml-classification-vs-regression/
10. Supervised and Unsupervised learning - GeeksforGeeks, زمان دسترسی: ژوئن 10, 2025، https://www.geeksforgeeks.org/supervised-unsupervised-learning/
11. چیست؟ | کافه‌تدریس, زمان دسترسی: ژوئن (Unsupervised Learning) یادگیری بدون ناظر 10, 2025، https://cafetadris.com/blog/%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D8%A8%D8%AF%D9%88%D9%86-%D9%86%D8%A7%D8%B8%D8%B1/
12. What is Unsupervised Learning? - GeeksforGeeks, زمان دسترسی: ژوئن 10, 2025، https://www.geeksforgeeks.org/unsupervised-learning/
13. یادگیری نظارت نشده چیست؟ – توضیح کامل به زبان ساده – فرادرس - مجله, زمان دسترسی: ژوئن 10, 2025، https://blog.faradars.org/%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D9%86%D8%B8%D8%A7%D8%B1%D8%AA-%D9%86%D8%B4%D8%AF%D9%87-%DA%86%DB%8C%D8%B3%D8%AA/
14. Unsupervised learning: clustering and dimensionality reduction | Advanced R Programming Class Notes | Fiveable, زمان دسترسی: ژوئن 10, 2025، https://library.fiveable.me/introduction-to-advanced-programming-in-r/unit-7/unsupervised-learning-clustering-dimensionality-reduction/study-guide/oFpp6lL1BZiYCXBn
15. Q-Learning Explained: Learn Reinforcement Learning Basics - Simplilearn.com, زمان دسترسی: ژوئن 10, 2025، https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-q-learning
16. The Math Behind RL - Number Analytics, زمان دسترسی: ژوئن 10, 2025، https://www.numberanalytics.com/blog/math-behind-reinforcement-learning
17. Understanding the Bellman Equation in Reinforcement Learning - DataCamp, زمان دسترسی: ژوئن 10, 2025، https://www.datacamp.com/tutorial/bellman-equation-reinforcement-learning
18. Accelerating Artificial Intelligence: The Role of GPUs in Deep Learning and Computational Advancements - EAST PUBLICATION & TECHNOLOGY, زمان دسترسی: ژوئن 10, 2025،

https://eastpublication.com/index.php/eje/article/download/34/13
19. 7 Breakthrough Linear Regression Uses in Modern Biotechnology, زمان دسترسی: ژوئن 10, 2025، https://www.numberanalytics.com/blog/breakthrough-linear-regression-biotech
20. Linear Regression Formula - GeeksforGeeks, ژوئن زمان دسترسی: 10, 2025، https://www.geeksforgeeks.org/linear-regression-formula/
21. Linear Regression in Machine Learning - Analytics Vidhya, ژوئن زمان دسترسی: 10, 2025، https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/
22. Logistic Regression in Machine Learning - GeeksforGeeks, ژوئن زمان دسترسی: 10, 2025، https://www.geeksforgeeks.org/understanding-logistic-regression/
23. Understanding logistic regression analysis - PMC - PubMed Central, زمان دسترسی: ژوئن 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC3936971/
24. Multi-Layer Perceptron Learning in Tensorflow - GeeksforGeeks, زمان دسترسی: ژوئن 10, 2025، https://www.geeksforgeeks.org/multi-layer-perceptron-learning-in-tensorflow/
25. Biomarker discovery process at binomial decision point (2BDP): Analytical pipeline to construct biomarker panel - PMC, ژوئن زمان دسترسی: 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC10562676/
26. Combining Multiple Biomarker Models in Logistic Regression - CiteSeerX, زمان ژوئن دسترسی: 10, 2025، https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0d5c99ccc41b905b9467a3ee286295ded36ecfb4
27. Decision Tree Algorithm, Explained - KDnuggets, ژوئن زمان دسترسی: 10, 2025، https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html
28. Exploring the Core Principles of Decision Tree in Machine Learning - Certisured, ژوئن زمان دسترسی: 10, 2025، https://certisured.com/blogs/exploring-the-core-principles-of-decision-tree-in-machine-learning/
29. k-Skip-n-Gram-RF: A Random Forest Based Method for Alzheimer's Disease Protein Identification - Frontiers, ژوئن زمان دسترسی: 10, 2025، https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00033/full
30. Comprehensive Guide to Random Forest - Siddhant Academy, زمان دسترسی: ژوئن 10, 2025، https://siddhantbhattarai.hashnode.dev/comprehensive-guide-to-random-forest
31. A Beginner's Guide to Random Forests and Their Effective Use - Number Analytics, ژوئن زمان دسترسی: 10, 2025، https://www.numberanalytics.com/blog/beginners-guide-random-forests-

effective-use
32. Random Forest for Bioinformatics - CMU School of Computer Science, زمان دسترسی: ژوئن 10, 2025، https://www.cs.cmu.edu/~qyj/papersA08/11-rfbook.pdf
33. What is k-means clustering? | IBM, زمان دسترسی: ژوئن 10, 2025، https://www.ibm.com/think/topics/k-means-clustering
34. Understanding K-Means Clustering for Smart Data Segmentation - Number Analytics, زمان دسترسی: ژوئن 10, 2025، https://www.numberanalytics.com/blog/understanding-k-means-clustering-for-smart-data-segmentation
35. Model-based clustering for RNA-seq data | Bioinformatics - Oxford Academic, زمان دسترسی: ژوئن 10, 2025، https://academic.oup.com/bioinformatics/article/30/2/197/217752
36. Assisted clustering of gene expression data using regulatory data from partially overlapping sets of individuals, زمان دسترسی: ژوئن 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC9734806/
37. Clustering of Gene Expression Data Based on Shape Similarity - PMC - PubMed Central, زمان دسترسی: ژوئن 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC3171421/
38. Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era - MDPI, زمان دسترسی: ژوئن 10, 2025، https://www.mdpi.com/2227-7390/10/24/4730
39. The History of the Perceptron: From Inception to Modern AI - Nomidl, زمان دسترسی: ژوئن 10, 2025، https://nomidl.hashnode.dev/the-history-of-the-perceptron-from-inception-to-modern-ai
40. Perceptron - Wikipedia, زمان دسترسی: ژوئن 10, 2025، https://en.wikipedia.org/wiki/Perceptron
41. What is perceptron and why it is important to understand neural ..., زمان دسترسی: ژوئن 10, 2025، https://data-intelligence.hashnode.dev/what-is-perceptron-and-why-it-is-important-to-understand-neural-networks
42. The Perceptron - A Guided Tutorial Through Its History and Implementation In Python, زمان دسترسی: ژوئن 10, 2025، https://pabloinsente.github.io/the-perceptron
43. Delving into Deep Learning | American Scientist, زمان دسترسی: ژوئن 10, 2025، https://www.americanscientist.org/article/delving-into-deep-learning
44. Multi-layered perceptron and its applications in biotechnology - ResearchGate, زمان دسترسی: ژوئن 10, 2025، https://www.researchgate.net/publication/376705126_Multi-layered_perceptron_and_its_applications_in_biotechnology

45. Multi-Layer Perceptron Explained: A Beginner's Guide - QuarkML, زمان دسترسی: ژوئن 10, 2025، https://www.quarkml.com/2023/01/multi-layer-perceptron-a-complete-overview.html
46. A Predictive Model of Gene Expression Using a Deep Learning ..., زمان دسترسی: ژوئن 10, 2025، https://calla.rnet.missouri.edu/cheng/dn_gene_expression.pdf
47. AlphaFold - Google DeepMind, ژوئن زمان دسترسی: 10, 2025، https://deepmind.google/science/alphafold/
48. About - AlphaFold Protein Structure Database, ژوئن زمان دسترسی: 10, 2025، https://alphafold.ebi.ac.uk/about
49. AlphaFold and protein folding: Not dead yet! The frontier is conformational ensembles - PMC, ژوئن زمان دسترسی: 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC11892350/
50. What is AlphaFold? - EMBL-EBI, ژوئن زمان دسترسی: 10, 2025، https://www.ebi.ac.uk/training/online/courses/alphafold/an-introductory-guide-to-its-strengths-and-limitations/what-is-alphafold/
51. Deep Learning- and Expert Knowledge-Based Feature Extraction and Performance Evaluation in Breast Histopathology Images - MDPI, زمان دسترسی: ژوئن 10, 2025، https://www.mdpi.com/2072-6694/15/12/3075
52. What is the importance of feature extraction in deep learning? - Milvus, زمان ژوئن دسترسی: 10, 2025، https://milvus.io/ai-quick-reference/what-is-the-importance-of-feature-extraction-in-deep-learning
53. What is Feature Extraction? - GeeksforGeeks, ژوئن زمان دسترسی: 10, 2025، https://www.geeksforgeeks.org/what-is-feature-extraction/
54. The 2010s – Smartphones, data centers and the dawn of AI - Simon-Kucher, ژوئن زمان دسترسی: 10, 2025، https://www.simon-kucher.com/en/insights/2010s-smartphones-data-centers-and-dawn-ai
55. From GPUs to AI and quantum: three waves of acceleration in bioinformatics, ژوئن زمان دسترسی: 10, 2025، https://www.researchgate.net/publication/380052451_From_GPUs_to_AI_and_quantum_three_waves_of_acceleration_in_bioinformatics
56. Graphics processing units in bioinformatics, computational biology and systems biology - PMC - PubMed Central, ژوئن زمان دسترسی: 10, 2025، https://pmc.ncbi.nlm.nih.gov/articles/PMC5862309/
57. Benefits and Drawbacks of Generative AI in Biotech - SciSure, ژوئن زمان دسترسی: 10, 2025، https://www.scisure.com/blog/benefits-drawbacks-of-generative-ai-in-biotech