

## جایگاه هوش مصنوعی در علوم دارویی

مرضیه ثنائی<sup>۱</sup>، محمدابراهیم دلزنده ندامانی<sup>۲</sup>، حمید باخرد<sup>۳</sup>

## مقاله مروری

## چکیده

**مقدمه:** کشف و توسعه‌ی دارو جنبه‌های مختلف سلامت انسان را تحت تأثیر قرار می‌دهد و به طور چشمگیری بر بازار دارو تأثیر می‌گذارد. با این حال، سرمایه‌گذاری در یک داروی جدید اغلب به دلیل فرآیند طولانی و پیچیده تحقیق و توسعه‌ی دارو یک چالش پیچیده و سخت است.

**روش‌ها:** با پیشرفت تکنولوژی تجربی و ساخت‌افزار کامپیوتر، هوش مصنوعی اخیراً به عنوان ابزاری پیشرو در تجزیه و تحلیل داده‌های فراوان و با ابعاد بالا ظاهر شده است. رشد انفجاری در اندازه داده‌های زیستی مزایایی را در به کارگیری هوش مصنوعی در تمام مراحل تحقیق و توسعه‌ی دارویی فراهم می‌کند.

**یافته‌ها:** مشابه مدل یادگیری انسان، یادگیری ماشین و یادگیری عمیق می‌توانند به تدریج ویژگی‌های مختلف داده‌ها را تشخیص دهند و پارامترهای مدل خود را از طریق تکرارهای مداوم به روز کنند تا زمانی که یک مدل معتبر تشکیل شود.

**نتیجه‌گیری:** این مقاله با مروری کوتاه بر مدل‌های رایج هوش مصنوعی در زمینه‌ی کشف دارو آغاز می‌شود. سپس، کاربردهای خاص آنها را در مراحل مختلف تحقیق و توسعه دارویی، مانند کشف هدف، کشف و طراحی دارو، تحقیقات پیش بالینی و تأثیرات در بازار دارویی، به طور خلاصه مورد بحث قرار می‌دهد. در نهایت، محدودیت‌های عمده‌ی هوش مصنوعی در تحقیق و توسعه‌ی دارویی مورد بحث قرار گرفته است.

**واژگان کلیدی:** طراحی دارو؛ هوش مصنوعی؛ یادگیری ماشین؛ یادگیری عمیق

**ارجاع:** ثنائی، مرضیه، دلزنده ندامانی محمدابراهیم، باخرد حمید. جایگاه هوش مصنوعی در علوم دارویی. مجله دانشکده پزشکی اصفهان ۱۴۰۴؛ ۴۳ (۸۱۳): ۴۶۲-۴۵۱.

## مقدمه

فناوری‌ها به ندرت نقش مهمی در روند فعلی کشف داروهای جدید داشته‌اند. بنابراین، نیاز مبرمی به فناوری جدید برای پیشبرد توسعه‌ی داروها وجود دارد.

طراحی دارو به کمک رایانه (Computer-aided drug design) CADD مثالی از زیست فناوری وابسته به کامپیوتر است که بسیاری از مشکلات تولید و کشف داروهای جدید را حل نموده است (۶، ۷). با افزایش قدرت محاسباتی دستگاه‌ها، هوش مصنوعی (Artificial intelligence) AI به دلیل توانایی در یادگیری، پردازش و پیش‌بینی مقادیر انبوه اطلاعات مورد استفاده قرار گرفته است. داده‌های زیست‌پزشکی که سطح بالایی از ناهمگونی و پیچیدگی دارند، از منابع مختلف، از جمله داده‌های omics از پلتفرم‌های مختلف، داده‌های آزمایشی از آزمایشگاه‌های بیولوژیکی یا شیمیایی و داده‌های تولید

توسعه‌ی داروهای جدید برای سلامت جهانی حیاتی است، اما این فرآیند پرهزینه و زمان‌بر است، اغلب چندین سال طول می‌کشد و بیش از یک میلیارد دلار هزینه دارد (۱، ۲). بهبود در مراحل طراحی دارو می‌تواند تا حد زیادی این هزینه‌ها را کاهش داده و آن را در دسترس‌تر کند. به عبارت دیگر، زمینه‌ی فعلی تحقیق و توسعه دارو (R&D) نیازمند بهبودهای قابل توجه بهره‌وری برای کوتاه کردن چرخه‌ی زمان و هزینه توسعه دارو است (۳). فناوری‌هایی مانند فارماکولوژی شبکه، توالی‌یابی RNA (RNA-seq)، غربالگری با توان بالا (HTS) (High throughput screening)، یا غربالگری مجازی (VS) (Virtual screening) همگی کشف اهداف جدید و همچنین داروهای جدید را تا حدودی تسریع کرده‌اند (۴، ۵). با این وجود، این

۱- دکترای بیوتکنولوژی دارویی، گروه بیوتکنولوژی، دانشکده‌ی داروسازی، دانشگاه علوم پزشکی تهران، تهران، ایران

۲- کارشناس ارشد مکاترونیک، دانشکده‌ی مکانیک، دانشگاه سمنان، سمنان، ایران

۳- دانشیار بیوتکنولوژی دارویی، گروه بیوتکنولوژی، دانشکده‌ی داروسازی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

نویسنده‌ی مسؤول: حمید باخرد؛ استادیار بیوتکنولوژی دارویی، گروه بیوتکنولوژی، دانشکده‌ی داروسازی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

Email: bakheradh@pharm.mui.ac.ir

در آن را استنباط کنند و پارامترهای مدل خود را از طریق تکرارهای مداوم به روز کنند تا زمانی که یک مدل معتبر تشکیل شود.

بسته به نوع الگوریتم یادگیری مورد نیاز برای حل مسأله، مدل‌ها به سه دسته تقسیم می‌شوند: یادگیری تحت نظارت (Supervised learning)، یادگیری بدون نظارت (Unsupervised learning) و یادگیری تقویتی (Reinforcement learning). یادگیری نظارت شده یک فرایند مبتنی بر داده برچسب‌گذاری شده است که مدلی را در مورد رابطه‌ی بین ورودی و خروجی از پیش تعیین شده آن برای پیش‌بینی متغیرهای ورودی آینده آموزش می‌دهد.

در مقابل، روش‌های بدون نظارت برای شناسایی الگوها در مجموعه داده‌های بدون برچسب و کاوش در ساختارهای بالقوه‌ی یک مجموعه داده استفاده می‌شود تا امکان خوشه‌بندی داده‌ها برای تجزیه و تحلیل بیشتر فراهم شود. یادگیری تقویتی ساخت مدل را از طریق یادگیری تعاملی ثابت انجام می‌دهد، با تکیه بر مجازات برای شکست یا پاداش برای موفقیت (۱۲).

روش‌های مختلف ML و DL با موفقیت برای حل مشکلات زیست‌پزشکی استفاده شده‌اند. ML الگوریتم‌های خاصی ایجاد می‌کند که بر ویژگی‌های داده‌ها تمرکز می‌کند و آنها را به دانشی تبدیل می‌کند که ماشین‌ها می‌توانند بخوانند تا بینش‌های جدیدی را در اختیار انسان قرار دهد. الگوریتم‌های مبتنی بر ML مختلفی برای محققان وجود دارد که می‌توانند از بین آنها انتخاب کنند (۱۳). از جمله الگوریتم‌های مبتنی بر ML می‌توان به الگوریتم ساده بیز NB (Naïve Bayes)، تقویت گرادیان شدید (Extreme gradient Boosting)، SVM، XGBoost (boosting)، نزدیکترین همسایه k (K-nearest neighbor)، حداقل مربعات جزئی (Partial least squares)، PLS و رگرسیون لجستیک (Logistic regression) LR اشاره نمود. بر خلاف فرایند سنتی ML که اغلب از چندین ماژول مستقل تشکیل شده است، DL یک فرایند یادگیری و پیش‌بینی یکسره است که طی فرایند مدل‌سازی داده‌های خروجی را مستقیماً از داده‌های ورودی به دست می‌آورد و به طور مداوم مدل را بر اساس خطای بین خروجی و مقدار واقعی تنظیم و بهینه‌سازی می‌کند تا زمانی که به نتیجه مورد انتظار برسد (۱۰). یادگیری عمیق زیرمجموعه‌ای از یادگیری ماشینی است که از شبکه‌های عصبی چندلایه به نام شبکه‌های عصبی عمیق برای شبیه‌سازی قدرت تصمیم‌گیری پیچیده‌ی مغز انسان استفاده می‌کند (۱۴). یادگیری عمیق می‌تواند از اشتباهات خود درس بگیرد در حالی که یادگیری ماشین به مداخله انسانی نیاز دارد. یادگیری عمیق همچنین به داده‌های بسیار بیشتری نسبت به یادگیری ماشینی نیاز دارد که به نوبه خود نیازمند قدرت محاسباتی بسیار بیشتری است (شکل ۱) (۱۵). از جمله الگوریتم‌های رایج مبتنی

شده توسط شرکت‌های داروسازی به دست می‌آیند. از هوش مصنوعی می‌توان برای یادگیری الگوهای بالقوه در این حجم عظیم از داده‌های زیست‌پزشکی استفاده کرد و در نتیجه فرصت‌ها و چالش‌های جدیدی را برای صنایع داروسازی به ارمغان آورد. به دلیل توانایی بالای AI در استفاده از داده‌های زیست‌پزشکی مرتبط برای درک سیستم‌های پیچیده بیولوژیکی و فضا‌های واکنش شیمیایی، هوش مصنوعی تأثیری انقلابی بر تمام مراحل تحقیق و توسعه دارو از جمله تحقیق بر روی پروتئین‌ها و مولکول‌های کوچک، و همچنین کمک به طراحی آزمایشات بالینی و نظارت پس از فروش داشته است. در حال حاضر برخی شرکت‌های داروسازی این الگوریتم‌ها را در فرآیندهای طراحی داروی خود ادغام می‌کنند (۸). داروهای توسعه یافته با این روش‌ها، مانند DSP-1181، EXS21546 و DSP-0038، به آزمایش‌های بالینی رسیده‌اند (۹)، که اثربخشی AI در تولید عوامل درمانی مناسب را اثبات می‌کند.

تکنیک‌های هوش مصنوعی در این زمینه عمدتاً شامل یادگیری ماشینی (Machine learning) ML و یادگیری عمیق (Deep learning) DL می‌شوند. هر دو رویکر ML و DL حاوی الگوریتم‌هایی هستند که در کشف و اعتبارسنجی هدف، کشف و طراحی دارو، و تحقیقات دارویی پیش‌بالینی نقش دارند، جایی که از آنها برای تجزیه و تحلیل ویژگی‌های داده‌های مختلف استفاده می‌شود. بنابراین، در این بررسی، مروری بر برخی از جنبه‌های استفاده از هوش مصنوعی در علوم دارویی خواهیم داشت (۱۰).

### مفاهیم اولیه هوش مصنوعی

هوش مصنوعی اولین بار در کنفرانس دارتموث در سال ۱۹۵۶ پیشنهاد شد و به عنوان الگوریتمی تعریف شد که به ماشین‌ها توانایی استدلال و انجام عملکردها را می‌دهد. از ماشین‌های ادراکی گرفته تا ماشین‌های برداری (Support vector machines) SVM و شبکه‌های عصبی مصنوعی (Artificial neural networks) ANN، توسعه هوش مصنوعی فراز و نشیب‌های زیادی را پشت سر گذاشته است و در حال حاضر به لطف پشتیبانی سخت‌افزاری که اکنون در دسترس است، شکوفا شده است. هر دو روش ML و DL تحت رده‌ی هوش مصنوعی قرار می‌گیرند. به طور دقیق‌تر، DL را می‌توان در دسته ML قرار داد (۱۰).

برخلاف محاسبات برنامه‌نویسی سنتی، ML و DL می‌توانند الگوهای بالقوه را از داده‌های ورودی، بدون برنامه‌نویسی صریح بیاموزند. آنها با فرمت داده‌های ورودی محدود نمی‌شوند و می‌توانند شامل متن، تصاویر، صدا و همه انواع داده‌هایی که می‌توانند رمزگذاری شوند باشند (۱۱). مشابه مدل یادگیری انسان، ML و DL می‌توانند به تدریج ویژگی‌های مختلف داده‌ها را تشخیص دهند، الگوهای موجود

بنابراین، رویکردهای ML و DL را می‌توان برای یادگیری دانش بالقوه از مجموعه داده‌های omics در مقیاس بزرگ مورد استفاده قرار داد، که می‌تواند به کشف ژن‌ها یا مسیرهای حیاتی برای فرایندهای بیولوژیکی کمک کند (۲۰). برای مثال، در مطالعه‌ای توسط Jeon و همکاران از یک الگوریتم SVM با هسته تابع پایه شعاعی (Radial basis function) RBF برای ساخت سه مدل جهت پیش‌بینی اهداف بالقوه خاص سرطان سینه (BrCa)، سرطان پانکراس (PaCa) و سرطان تخمدان (OvCa) استفاده شد (۲۱). اساسی بودن ژن، بیان ژن، تنوع تعداد کپی DNA، جهش سوماتیک و توپولوژی شبکه (Protein-Protein interactions) PPI ویژگی‌های اصلی ورودی بودند و SVM قادر بود به طور عمیق ارتباط و تفاوت بین این ویژگی‌ها را برای تشخیص اهداف دارویی بالقوه از پروتئین‌های غیرهدف بررسی کند. با استفاده از مدل آموزش‌دیده برای پیش‌بینی ۱۵۶۶۳ پروتئین انسانی و امتیازدهی به نتایج پیش‌بینی، در مجموع ۱۲۲ هدف برای انواع سرطان شناسایی شد (۶۹ مورد از آن‌ها با ۱۱۶ هدف شناخته‌شده که به‌شدت تأیید شده بودند مطابقت داشت). علاوه بر این، تعداد زیادی از اهداف بالقوه خاص BrCa، PaCa، OvCa شناسایی شدند.

بر همین اساس پلتفرم‌هایی نیز طراحی شده که در شناسایی کاندیدهای هدف بالقوه نقش دارند. برای مثال، PharmMapper یک سرویس اینترنتی است که ۵۲۴۳۱ مدل فارماکوفور را از TargetBank، DrugBank، BindingDB و پایگاه داده هدف دارویی بالقوه (PDTD (Potential drug target database) جمع‌آوری کرده و از آنها برای شناسایی کاندیدهای هدف بالقوه برای مولکول‌های کوچک دارویی با استفاده از یک رویکرد نقشه برداری سریع فارماکوفور استفاده می‌کند (۲۲).

ChemMapper یکی دیگر از سرویس‌های اینترنتی است که هدف آن پیش‌بینی اثرات پلی فارماکولوژی، اهداف پروتئینی بالقوه و حالت‌های عمل (Modes of action) مولکول‌های کوچک بر اساس محاسبات شباهت سه بعدی است و به این منظور از یک پایگاه داده حاوی ۴۳۵۰۰۰۰ ساختار شیمیایی با فعالیت‌های زیستی و تفسیر مولکول‌های هدف مرتبط با آنها بهره می‌برد (۲۳). پلتفرم iDrug نیز یک ابزار آنلاین کاربرپسند و کارآمد برای طراحی دارو به کمک رایانه است که بر اساس فارماکوفور و جستجوی شباهت مولکولی سه بعدی عمل می‌کند و امکان تشخیص سایت‌های اتصال و پیش‌بینی هدف دارو را به صورت تعاملی از طریق یک رابط یکپارچه فراهم می‌کند (۲۴).

ML در هدایت فرایند یادگیری داده‌های مولتی اومیکس بسیار مفید بوده است، اما می‌تواند توسط داده‌های مولتی اومیکس بزرگتر و مشکلات پیچیده‌تر درهم شکسته و مغلوب شود (۲۵). با این حال، DL می‌تواند مقادیر بسیار بیشتری از داده‌های مولتی اومیکس را مدیریت کند

بر DL می‌توان به شبکه‌های عصبی کانولوشنال (Convolutional Neural Networks (CNN)، شبکه‌های عصبی مکرر (Recurrent Neural Networks (RNN)، پرسپترون‌های چند لایه (Multilayer Perceptrons (MLPs) و شبکه‌های متخاصم مولد (Generative Adversarial Networks (GAN) اشاره نمود (۱۱).

### کاربردهای هوش مصنوعی در علوم دارویی

#### شناسایی و اعتبارسنجی هدف

به طور کلی دو روش مرسوم برای کشف داروهای جدید (کلاس اول) وجود دارد: کشف داروی فنوتیپی (Phenotypic drug Target-based) PDD و کشف داروی مبتنی بر هدف (Target-based drug discovery) TDD (۱۶). PDD با استفاده از تکنیک‌های اولیه بیولوژیکی مانند میکروسکوپی، تصویربرداری و تکنیک‌های سلولی، غربالگری مجموعه‌ای از ترکیبات یا آنتی‌بادی‌ها را با ساختن مدل حیوانی یا آزمایشی که بسیار مرتبط با بیماری است، انجام می‌دهد. سپس، پاسخ سلول‌ها یا حیوانات آزمایشی به این ترکیبات مشاهده می‌شود تا مولکول‌هایی با سطح خاصی از کارایی برای اصلاح و بهینه‌سازی بیشتر ساختاری شناسایی شوند (۱۷).

با توسعه زیست‌شناسی مولکولی و تکنیک‌های مختلف توالی‌یابی، TDD به تدریج جایگزین PDD به عنوان الگوی اصلی کشف دارو شده است. TDD بر مفهوم «یک ژن، یک دارو، یک بیماری» متمرکز است. این رویکرد متکی بر یک هدف بسیار مرتبط با بیماری است که می‌تواند یک آنزیم، پروتئین یا سایر محصولات ژنی باشد، همراه با طراحی دقیق مولکول کوچک داروی اختصاصی هدف که می‌تواند منجر به تغییر در روند پیشرفت بیماری شود (۱۸). بنابراین، کشف هدف اولین مرحله حیاتی در فرایند توسعه دارو است. فرایند کشف هدف شامل تحقیقات چند وجهی، از جمله مطالعه ژن‌های مرتبط با بیماری، مسیرهای سیگنالینگ، برهمکنش‌های پروتئینی، و برهمکنش‌های پروتئین-مولکول‌های کوچک (دارو) است.

اهداف بالقوه مولکول‌هایی هستند که با یک بیماری خاص مرتبط هستند و کمترین میزان ارتباط ممکن را با سایر بیماری‌ها دارند. بیماری‌های پیچیده‌ای مانند سرطان، قلبی عروقی و ایمنی اغلب توسط چندین ژن کلیدی یا مسیرهای سیگنالی تنظیم می‌شوند، بنابراین اغلب لازم است که ارتباط بین چندین مولکول و بیماری را بررسی کنیم. لذا پردازش و تجزیه و تحلیل داده‌های omics (ژنومیکس، ترانس کریپتومیکس، پروتئومیکس، متابولومیکس) در تسهیل درک محققان از سیستم‌ها و فرایندهای بیولوژیکی پیچیده کارگشا خواهد بود (۱۹). جدول ۱ پایگاه‌های omics مختلفی را نشان می‌دهد که مورد استفاده محققان قرار می‌گیرند (۱۰). با این حال، پردازش و تجزیه و تحلیل این داده‌های omics پیچیده و با ابعاد بالا، بسیار چالش‌برانگیز است.

داده‌ها در نظر می‌گیرد که به طور ضمنی رابطه بین ویژگی‌ها و نتایج تجربی را یاد می‌گیرد و امتیاز نهایی را می‌دهد. برای مثال یک روش VS موسوم به ID-Score، نه کلاس از توصیفگرهای ویژگی شامل برهمکنش واندروللس، برهمکنش پیوند هیدروژنی، برهمکنش الکترواستاتیک، برهمکنش سیستم p، برهمکنش پیوند فلز-لیگاند، اثر تجزیه، اثر از دست دادن آنتروپیک، تطبیق شکل و تطبیق ویژگی سطح را به عنوان ویژگی داده در نظر گرفته و از ۲۲۷۸ ترکیب به عنوان مجموعه آموزشی استفاده کرد و از الگوریتم رگرسیون بردار پشتیبان (SVR) برای بررسی میزان تمایل اتصال مولکول‌های کوچک به پروتئین‌ها استفاده کرد. نتایج نشان داد که ID-Score می‌تواند لیگاندهای مشابه ساختاری را به درستی تشخیص دهد که استفاده از آن را به عنوان ابزاری قدرتمند برای ارزیابی میل ترکیبی دارو-پروتئین مبتنی بر ساختار نشان می‌دهد (۳۴).

در مطالعه‌های دیگر، از CNN برای امتیازدهی لیگاندهای پروتئینی استفاده شد. برخلاف روش‌های سنتی، CNNها به اندازه‌ای قدرتمند هستند که نمایش‌های سه‌بعدی تعاملات پروتئین-لیگاند را به عنوان ورودی بپذیرند. در طول آموزش مدل، CNN ویژگی‌های کلیدی مؤثر بر اتصال را از نمایش سه‌بعدی می‌آموزد که برای تعیین وضعیت اتصالات درست/نادرست استفاده می‌شود (۳۵). (جدول ۱)

#### روش‌های مبتنی بر لیگاند

در مقایسه با رویکرد SBVS که توسط اطلاعات ساختاری پروتئین هدف محدود می‌شود، روش‌های غربالگری مبتنی بر لیگاند LBVS (Ligand-based VS) می‌توانند از داده‌های زیست‌فعال لیگاند شناخته شده استفاده کنند و پایگاه داده بزرگی از ترکیبات را برای کشف ترکیبات پیشرو بالقوه غربال کنند. بنابراین، VS مبتنی بر هوش مصنوعی تمایل به حمایت از LBVS دارد. نقطه شروع LBVS این فرض است که ترکیبات ساختاری مشابه، فعالیت‌های بیولوژیکی مشابهی دارند. بنابراین، روش‌های هوش مصنوعی مورد استفاده در این زمینه شامل مدل‌های رگرسیونی برای پیش‌بینی فعالیت و مدل‌های طبقه‌بندی بر اساس شباهت ترکیبی است (۱۰). مدل‌های مبتنی بر رابطه کمی ساختار-فعالیت که به اختصار (Quantitative structure-activity relationship) QSAR نامیده می‌شود به طور گسترده در LBVS استفاده می‌شود زیرا از مدل‌های ریاضی برای ارتباط ساختارهای مولکولی با فعالیت‌های کمی بیولوژیکی استفاده می‌کند. یک مدل QSAR فرض می‌کند که ساختار یک ترکیب ویژگی‌های فیزیکوشیمیایی و فعالیت بیولوژیکی آن را تعیین می‌کند. بنابراین، می‌توان روابط کمی بین ساختار یک ترکیب و خواص فیزیکوشیمیایی، فعالیت بیولوژیکی، اثرات سم‌شناسی و غیره ایجاد کرد. فرایند تجزیه و تحلیل QSAR معمولاً شامل تهیه مجموعه

و ارتباطات عمیق‌تری را کشف کند. اکثر این مدل‌های کشف هدف از مدل‌های سرتاسری (End-to-end models) برای کشف پروتئین‌های با قابلیت دارو استفاده می‌کنند (۲۶). DL همچنین می‌تواند نقش کلیدی در چندین مرحله خاص از فرایند کشف هدف از جمله پیش‌بینی بیان ژن و انواع پیرایش یک ژن انجام دهد و این همان نقش یک ژن-یک آنزیم-یک بیماری را معرفی می‌کند. برای مثال، SpliceAI یک ابزار مبتنی بر DL است که برای پیش‌بینی انواع پیرایش از توالی رونوشت pre-mRNA به کار می‌رود (۲۷). (Single-cell variational Inference) scVI دیگر ابزار مبتنی بر DL است که برای پیش‌بینی و تجزیه و تحلیل احتمالات بیان ژن در سلول‌های منفرد از داده‌های بدست آمده از ترانسکریپتومیکس استفاده می‌کند (۲۸).

#### شناسایی برهمکنش‌های دارو-هدف

شناسایی برهمکنش‌های دارو-هدف (Drug-target interactions) DTIs می‌تواند به تحقیقات در زمینه‌ی کشف دارو کمک کند. DTIs تازه کشف شده را می‌توان برای یافتن اهداف جدیدی که با داروهای موجود برهمکنش دارند یا برای کشف ترکیبات جدیدی که با یک هدف مرتبط با بیماری تعامل دارند، استفاده کرد. بنابراین، نتایج تحقیقات در مورد DTIsها به طور گسترده در زمینه‌های کشف ترکیبات پیشرو (Lead compounds)، کشف هدف جدید، بازکاربردیی دارو و پیش‌بینی عوارض جانبی دارو استفاده می‌شود (۲۹). وقوع یک DTI تحت تأثیر عوامل متعددی است و مربوط به ویژگی‌های چند بعدی است که ساختار و خواص دارو و هدف را نشان می‌دهد. چنین مشکلاتی به عنوان مشکلات رگرسیون در نظر گرفته می‌شود. پیش‌بینی DTIsها معمولاً به عنوان یک مشکل طبقه‌بندی باینری در نظر گرفته می‌شود. به طور کلی، روش‌های پیش‌بینی DTI به سه رویکرد اصلی تقسیم می‌شوند: روش‌های مبتنی بر لیگاند، روش‌های مبتنی بر ساختار و روش‌های کم‌ژنومیک (۱۰).

#### روش‌های مبتنی بر ساختار

VS مبتنی بر ساختار (Structure-based VS) SBVS لیگاندهای بالقوه را بر اساس ترکیب سه بعدی پروتئین انتخاب می‌کند و توانایی لیگاند برای اتصال به پروتئین را بر اساس دانش وارد شده از روش‌های بیوفیزیکی امتیاز می‌دهد و در نتیجه رتبه‌بندی نامزدهای دارو را به دست می‌آورد (۳۰). پیش از این، شبیه‌سازی با استفاده از نرم‌افزارهای مختلف Docking رویکرد غالب بود (۳۱). برای مثال سرور HADDOCK یکی از برنامه‌های محبوب بررسی میانکنش‌های گیرنده-لیگاند است که در بسیاری مطالعات مورد استفاده قرار گرفته است (۳۲، ۳۳). ساخت تابع امتیازدهی، محدودیت اولیه نتایج شبیه‌سازی بود که باید فاکتورهای زیادی را به عنوان پارامتر در نظر می‌گرفت. هوش مصنوعی این فاکتورهای زیاد را به عنوان ویژگی‌های

جدول ۱. پایگاه داده‌های omics

معرفی	پایگاه	امیکس
یک مرورگر ژنومی که تفسیر ژن، محاسبات تراز چندگانه، پیش‌بینی عملکرد تنظیمی و جمع‌آوری داده‌های بیماری را انجام می‌دهد. حاوی اطلاعاتی در مورد هم‌ترازی بین ژنومی، توالی‌های مختلف، فنوتیپ‌ها، پروفایل بیان، اطلاعات تنظیمی، وارته‌ها، مناطق تکراری و حفاظت شده ژنی.	Ensembl UCSC	ژنومیکس
داده‌های ژنومیک عملکردی از جمله جهش، تعداد کپی، mRNA و بیان پروتئین را جمع‌آوری می‌کند. پایگاه داده پروتئومیک‌ها، از جمله شناسایی پروتئین و پپتید و اصلاحات پس از ترجمه پایگاه داده پروتئینی حاوی توالی پروتئین، اطلاعات عملکردی و نمایه‌سازی مقالات است.	TCGA PRIDE UniProt HPA	پروتئومیکس
پایگاه داده حاوی میلیون‌ها تصویر با وضوح بالا است و توزیع فضایی پروتئین‌ها را در ۴۶ رده سلولی مختلف انسانی، ۲۰ نوع سرطان مختلف و ۴۴ بافت طبیعی انسان نشان می‌دهد.	HPM GEO EA	ترانسکریپتومیکس
یک منبع تعاملی است که مقدار زیادی از داده‌های توالی یابی پپتیدی را از نقشه پیش‌نویس پروژه پروتئوم انسانی ترکیب می‌کند. مخزنی از داده‌های بیان ژن با توان بالا	HMDB	
داده‌های بیان ژن پایه را در گونه‌های مختلف جمع‌آوری کرده و شامل مطالعات افتراقی است که تغییرات بیان را بین دو شرایط مختلف گزارش می‌کند.	KEGG Pathway MetaCyc Reactome	متابولومیکس
یک پایگاه داده که اطلاعات جامعی در مورد متابولیت‌های مولکولی کوچک که در بدن انسان کشف شده و به طور تجربی تأیید شده‌اند ارائه می‌دهد.		
مجموعه‌ای از نقشه‌های مسیرهای متابولیک است که به صورت دستی ترسیم شده همراه با ادغام اطلاعات مسیر متابولیک، ژن و پروتئین. شامل مسیرهای متابولیسم اولیه و ثانویه و همچنین متابولیت‌ها، واکنش‌ها، آنزیم‌ها و ژن‌هایی است که همراه با آنها هستند. یک پایگاه داده مسیر است که ابزارهای بیوانفورماتیک کاربر پسند را برای تجسم، تفسیر و تجزیه و تحلیل دانش مسیر فراهم می‌کند.		

HPA: human protein atlas; HPM: human proteome map; HMDB: human metabolome database; KEGG: Kyoto Encyclopedia of Genes and Genomes

می‌کند و این اطلاعات را به منابع متعدد اطلاعات زیست‌پزشکی برای پیش‌بینی بهتر DTIها متصل می‌کند. منابع پایگاه داده در دسترس عموم حاوی حجم زیادی از داده‌های زیست‌پزشکی ساختاریافته و بدون ساختار هستند تا دسترسی به اطلاعات را پشتیبانی کنند (جدول ۲ برخی پایگاه‌های داده برای تحقیقات DTI را معرفی می‌کند). ML و DL می‌توانند جزئیات عملکردی مرتبط را استخراج کنند و نويز را از این داده‌های نااهمگن بزرگ کاهش دهند تا اهداف پروتئینی جدید را به طور دقیق و کارآمد کشف کنند. استفاده از رویکرد ML برای پیش‌بینی DTI بسیار راحت است، که معمولاً فقط به دستیابی به SMILE (Simplified molecular-input lineentry system) (نوعی سیستم نمایش ساختاری که مولکول‌های کوچک را به شکل رشته نشان می‌دهد) مولکول‌های کوچک و توالی پروتئین‌های هدف نیاز دارد. این دنباله‌ها از طریق قوانین مختلف به بردارهای ویژگی تبدیل می‌شوند و بعداً به عنوان ورودی یک مدل برای پیش‌بینی طبقه‌بندی نهایی آنها استفاده می‌شوند (۹). برای مثال، مولکول‌های کوچکی که با استفاده از اثر انگشت MACCS (اثر انگشت ساختاری ۲ بعدی که برای اندازه‌گیری شباهت مولکولی استفاده می‌شوند) مشخص می‌شوند، با بردارهای پروتئینی که توسط توصیفگرهای (Composition, transition, and distribution) CTD مشخص می‌شوند، به هم متصل شده و به عنوان ورودی SVM برای پیش‌بینی DTIs استفاده شدند (۴۰).

داده‌های اولیه، محاسبه و انتخاب توصیفگرهای مولکولی، ایجاد مدل‌های مربوطه، و ارزیابی و اعتبارسنجی نتایج مدل است. NB، RF (Random forest) و SVM الگوریتم‌های رایج در LBVS و QSAR هستند (۳۶).

برای مثال، AbdulHameed و همکاران پایگاه داده‌ای با نزدیک به ۲۰۰۰ ترکیب را با استفاده از یک مدل مبتنی بر QSAR با الگوریتم NB برای تعیین آبگریزی فعال‌کننده‌های گیرنده X پرگنن (Pregnane X receptor) PXR (receptor) غربال کردند (۳۷). در مطالعه‌ای دیگر از روش pQSAR 2.0 مبتنی بر الگوریتم RF برای پیش‌بینی مقادیر نیمه حداکثر غلظت بازدارندگی PXR (IC50) استفاده شد (۳۸). روش‌های QSAR تنها ابزار مورد استفاده برای LBVS نیستند. برای مثال، Li و همکاران از چندین روش ML برای ساخت مدل‌های طبقه‌بندی برای انتخاب آگونیست‌های گیرنده X کبد (LXR) استفاده کردند (۳۹). از آنجایی که روش‌های DL بیشتری به تدریج برای مطالعات مرتبط با QSAR مورد استفاده قرار گرفته‌اند، محققان دریافته‌اند که DL در یادگیری تک وظیفه‌ای و چند وظیفه‌ای بهتر از ML عمل می‌کند. با افزایش سریع تعداد ترکیبات شناخته شده در سال‌های اخیر، معماری DL برای پردازش مجموعه داده‌های ترکیبی بزرگ مناسب‌تر است (۹).

### روش‌های کم‌ژنومیک

رویکرد کم‌ژنومیک از اطلاعات مربوط به دارو و هدف استفاده

نانوذره است، کشف کردند. این روش همچنین ویژگی‌های ساختاری مولکولی مهمی را نشان داد که به خودآرایی و تشکیل نانوذرات اجازه می‌دهد. نتایج نشان داد که با کمک سولفات ایندوسیانین، این داروها به نانوذرات با راندمان بارگذاری ۹۰ درصد مونتاژ شدند (۴۳). الگوی انتشار یک دارو نیز برای درمان بیماری بسیار مهم است. تولید داروهایی که در پاسخ به تفاوت سیگنال‌های فیزیولوژیکی (از جمله pH، گونه‌های ردوکس فعال، آنزیم‌ها، گلوکز، یون‌های مختلف، آدنوزین تری فسفات (ATP) و اکسیژن) اندام‌ها و بافت‌های مختلف منتشر می‌شوند، می‌تواند اثربخشی دارو را افزایش دهد، از عوارض سمی و جانبی ناشی از اهداف غیراختصاصی جلوگیری کند و به درمان ایمن و دقیق دست یابد. هوش مصنوعی می‌تواند ارزیابی حالت رهاسازی دارو را تسهیل کند و می‌تواند بازخوردی را برای فرمولاسیون حامل‌های دارو از طریق ML ارائه دهد (۴۴).

### تحقیقات دارویی پیش‌بالینی

به منظور جلوگیری از شکست آزمایشات بالینی که مرحله‌ای پرهزینه در فرایند تحقیق و توسعه دارو محسوب می‌شود، مطالعات پیش‌بالینی ضرورت دارد. این مطالعات بر بررسی ویژگی‌های فارماکولوژی غیربالینی، فارماکوکینتیک و سم‌شناسی متمرکز است. خواص فیزیوکوشیمیایی یک دارو و خواص (Absorption, ADMET (distribution, metabolism, excretion, and toxicity (جذب-توزیع-متابولیسم-دفع-سمیت) آن برای مطالعات فارماکوکینتیک و سم‌شناسی ضروری است (۴۵، ۴۶). ویژگی‌های ADMET یک داروی کاندید به ویژه داروهایی که مصرف خوراکی دارند، می‌تواند مستقیماً تحت تأثیر ویژگی‌های فیزیوکوشیمیایی آن باشد و تأثیر مهمی بر موفقیت یک دارو در ورود به بازار خواهد داشت (۴۷). به عنوان مثال، ثلث یونیزاسیون (pKa)، که پارامتر

استفاده از نمایش تک ویژگی به طور کامل مولکول‌ها یا پروتئین‌های کوچک را مشخص نمی‌کند، بنابراین برخی از مطالعات از توصیف‌گرهای متعدد برای توصیف مولکول‌ها و پروتئین‌های کوچک استفاده کرده‌اند و این ویژگی‌ها را به‌عنوان بردار ورودی‌ها برای پیش‌بینی DTI‌ها ادغام کرده‌اند. در چنین مواردی DL بهتر از ML عمل کرده و طبقه‌بندی مدل را تا حد معینی بهبود می‌بخشد (۹). (جدول ۲)

### انتشار دارو

با پیشرفت در روش‌های جدید کشف دارو، نیاز به گسترش سیستم‌های پیشرفته تحویل دارو نیز به سرعت افزایش یافته است. یک سیستم تحویل دارو را می‌توان به عنوان یک «گاری» (یعنی حامل) تجسم کرد که «کالاها» (یعنی مواد درمانی) را به مقصد مناسب منتقل می‌کند. در مقایسه با حامل‌های دارویی معمولی، نانوحامل‌ها می‌توانند حلالیت دارو را بهبود بخشند و اثرات نامطلوب حل‌کننده‌های معمولی را کاهش دهند. علاوه بر محافظت از دارو در برابر زوال، نانوحامل‌ها می‌توانند در هدف‌گیری عملکردی دارو نقش ایفا کنند (۴۱). با این وجود، بسته به داروها، مواد کمکی (Excipients) و شرایط واکنش از جمله دما، زمان و سرعت هم زدن، تهیه‌ی یک نانوحامل مناسب بسیار پیچیده است. آزمایش‌ها به تنهایی نمی‌توانند همه این پارامترها را بررسی کنند. علاوه بر تعیین هدف مولکولی دارو و فعالیت بیولوژیکی، هوش مصنوعی می‌تواند شرایط بهینه تشکیل نانوحامل را به دقت پیش‌بینی کند (۴۲).

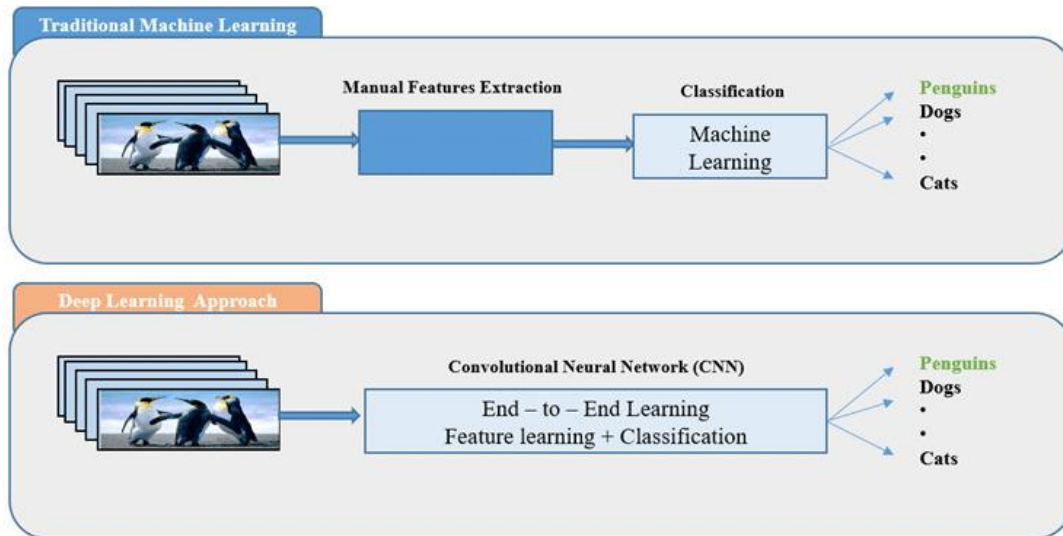
برای مثال، در مطالعه‌ای با استفاده از محاسبات کمی پیش‌بینی مونتاژ ساختار- نانوذره (Quantitative structure-nanoparticle assembly prediction) QSNAP، دو توصیف‌گر مولکولی را برای پیش‌بینی اینکه کدام فرم داروی ایندوسیانین قادر به تشکیل فرم

جدول ۲. برخی پایگاه‌های داده برای تحقیقات DTI

پایگاه داده	معرفی
ChEMBL	کتابخانه‌ای است که به صورت دستی از ترکیبات فعال زیستی با خواص شبیه دارو تشکیل شده است. داده‌های مربوط به مواد شیمیایی، فعالیت بیولوژیکی و ژنومیک یکپارچه شده است.
Pharos	پایگاه داده‌ای است که یک پایگاه دانش جامع و یکپارچه برای ژنوم با قابلیت دارو فراهم می‌کند تا بخشی از ژنوم را که به خوبی توصیف نشده است، روشن کند.
GtoPdb	یک پایگاه داده از فعل و انفعالات لیگاند-فعالیت-هدف است که حاوی داده‌های کمی در مورد اهداف دارویی و داروهای آزمایشی و تجویزی است که بر آنها تأثیر می‌گذارد.
DockCov2	یک پایگاه داده دارویی با ۳۵۴۸ ترکیب است که میل پیوندی پروتئین‌های مرتبط با SARS-CoV-2 و ۶۷ پروتئین انسانی را پیش‌بینی می‌کند.
SIDER	شامل اطلاعاتی در مورد داروهای تأیید شده و ADRها، فرکانس عوارض جانبی دارو و دسته‌های عوارض جانبی، و روابط دارو-هدف است.
PDTD	پایگاه داده پروتئینی برای شناسایی هدف است که شامل ۱۱۸۶ پروتئین است و ۸۳۱ هدف دارویی شناخته شده یا بالقوه را پوشش می‌دهد. همچنین تفسیر پروتئین و ساختارهای سایت فعال، بیماری‌های مرتبط، عملکردهای بیولوژیکی و مسیرهای تنظیم‌کننده (سیگنال) مرتبط را ارائه می‌دهد.
PDID	حاوی تعداد زیادی از فعل و انفعالات بالقوه و بومی پروتئین-دارو در پروتئوم ساختاری انسان است.

GtoPdb: guide to pharmacology database; PDID: protein-drug interaction database; SIDER: side-effect resource; ADR: adverse drug reaction.





شکل ۱. مقایسه‌ی بین روش یادگیری ماشین و یادگیری عمیق (۱۵)

می‌شوند. به عنوان مثال، یک مدل شبکه عصبی پس‌انتشار پیش‌خور کلاسیک (Backpropagation neural network) BPNN و یک رویکرد اعتبارسنجی متقاطع مکرر (Repeated double cross-validation) rdCV برای تخمین نفوذ سد خونی مغزی ترکیب شدند (۵۴). DL اجازه می‌دهد تا یک مدل با استفاده از مجموعه داده‌های بزرگتر آموزش داده شود. لذا این اطمینان حاصل می‌شود که طیف وسیع‌تری از ترکیبات نسبت به آنچه در ML امکان پذیر است، پوشش داده شده است. مطالعات، پتانسیل مدل‌های چندوظیفه‌ای را برای پیش‌بینی ویژگی‌های ADMET نشان داده است. بر این اساس، نرم‌افزارهای کاربرپسند و سرورهای تحت وب برای پیش‌بینی ویژگی‌های ADMET مولکول‌ها توسعه یافته‌اند. در این میان ADMETlab 2.0 به طور گسترده مورد تحسین قرار گرفته است. ADMETlab 2.0 بر اساس یک چارچوب توجه به نمودار چند وظیفه‌ای (Multi-task graph attention) MGA است و می‌تواند چندین ویژگی ADMET داروها را پیش‌بینی کند (در مجموع ۸۸ پارامتر شامل ۲۳ ویژگی ADMET، ۲۷ نقطه‌ی پایانی سمیت و ۸ ساختار مرتبط با سمیت دارو را پیش‌بینی می‌کند) (۵۳). بیشتر داده‌های مورد استفاده در این نرم‌افزار از داده‌های زیست‌فعالیتی موجود در پایگاه‌های داده عمومی، مقالات مرتبط و نرم‌افزار پیش‌بینی سمیت Toxicity Estimation Software Tool (TEST) مشتق شده‌اند. بر این اساس، برخی از ویژگی‌های پیش‌بینی شده توسط ADMETlab 2.0 در مقایسه با نتایج ابزارهای مشابه منحصر به فرد هستند.

#### طراحی کارآزمایی بالینی

مرحله‌ی کارآزمایی بالینی بیشتر وقت و سرمایه‌گذاری را در طول تحقیق و توسعه دارویی مصرف می‌کند. نرخ بالای شکست کارآزمایی‌های

اساسی زیربنای خواصی چون ضریب توزیع اکتانول-آب (logD) و حلالیت است، بر حلالیت آبی یک مولکول تأثیر می‌گذارد که به نوبه خود می‌تواند بر روش فرمولاسیون دارو تأثیر بگذارد (۴۸).

علاوه بر این، ADMET ترکیبات تحت شرایط pH مختلف عمیقاً تحت تأثیر وضعیت بار ترکیبات است. با این حال، ویژگی‌های فیزیکوشیمیایی به راحتی مستقیماً اندازه‌گیری نمی‌شوند (۴۹). برای پرداختن به موضوع پیش‌بینی همزمان چندین ویژگی فیزیکوشیمیایی مولکول‌های کوچک، محققان بر یادگیری و مشخصه‌یابی ویژگی‌های مولکولی تمرکز کرده‌اند. مثال‌ها شامل یادگیری و نمایش ویژگی‌های مولکولی بر اساس مدل GNN (Graph neural network) (۵۰)، ترکیب رویکردهای نمایش مولکولی سستی با ویژگی‌های آموخته‌شده توسط شبکه‌های عصبی پیام‌رسان (Message-passing neural networks) MPNNs (۵۱) و شکلی از نمایش گرافیکی طراحی مولکولی بر اساس اثر انگشت دایره‌ای اتصال توسعه‌یافته (Extended-connectivity) ECFPs (circular fingerprints) است (۵۲).

با توجه به اینکه صدها ترکیب در انتظار ارزیابی خواص ADMET خود در مرحله‌ی اولیه کشف دارو هستند، اعتبارسنجی هر یک از آنها از طریق مطالعات گسترده حیوانی زمان‌بر و پرهزینه است. بنابراین، استفاده از هوش مصنوعی برای پیش‌بینی سریع و دقیق خواص ADMET داروها به طور گسترده مورد استفاده قرار گرفته است. QSAR و مدل‌های کمی ساختار-ویژگی (QSPR) نقش‌های محوری در پیش‌بینی ADMET مولکول‌های کوچک دارند (۵۳).

بسیاری از روش‌های ML، در ترکیب با مدل‌های QSAR یا QSPR، در پیش‌بینی ADMET به خوبی عمل کرده‌اند. رویکردهای DL نیز به طور گسترده برای پیش‌بینی ویژگی‌های ADMET اعمال

در گروهی از بیماران دیابتی کمک گرفتند (۶۱). جدای از کاربرد داده‌های ساختاریافته، داده‌های بدون ساختار مستخرج از مجموعه‌های زیست‌پزشکی و بالینی را می‌توان برای تشخیص و طبقه‌بندی تداخل دارو-دارو (DDI (Drug-drug interaction)) و پیش‌بینی ADR با کمک روش‌های پردازش زبان طبیعی (Natural language processing) استفاده کرد (۶۲).

#### محدودیت‌های AI

اگرچه در این بررسی کاربردهای هوش مصنوعی در فرایند تحقیق و توسعه دارو توضیح داده شد اما در حال حاضر، هوش مصنوعی واقعاً سیستم دارویی سنتی را از بین نبرده است و بسیاری از فرآیندهای تحقیقاتی هنوز در انتظار «بهینه‌سازی» توسط هوش مصنوعی هستند و دلیل آن، محدودیت‌هایی است که در کاربرد هوش مصنوعی برای کشف دارو وجود دارد. اولین و مهم‌ترین چالش پیش‌رو، محدودیت داده است. داده‌های باکیفیت و دقیق گاهی اوقات می‌توانند مدل‌های ساده را قادر به عملکرد بهتر از مدل‌های پیچیده کنند. در واقع، توسعه الگوریتم‌های هوش مصنوعی را نمی‌توان از وجود داده جدا کرد. ساخت الگوریتم‌های هوش مصنوعی به شدت به داده‌های با کیفیت بالا و کافی متکی است. به دلیل محدودیت‌های فناوری کنونی، دستیابی به داده‌های با کیفیت بالا برای سیستم‌های بیولوژیکی پیچیده مشکل است و پردازش این داده‌ها به داده‌های استاندارد با اطمینان بالا هزینه‌بر است (۶۳).

برای مثال، برخی از داده‌ها توسط سنجش آزمایشگاهی به دست آمده است. با این حال، به دلیل عدم درک کامل از پاسخ در ارگانسیم، داده‌های آزمایشگاهی اغلب به طور قابل توجهی با داده‌های واقعی in vivo متفاوت است. بنابراین، نتایج پیش‌بینی مدل‌های آموزش دیده با داده‌های به‌دست‌آمده از آزمایش‌های آزمایشگاهی اغلب قانع‌کننده نیستند. عدم تعادل داده‌ها نیز یک مشکل عمده در آموزش مدل است. همانطور که قبلاً ذکر شد، مجموعه داده‌های مثبت به راحتی در زمینه داروسازی در دسترس هستند، اما مجموعه داده‌های منفی اغلب به طور دقیق شناسایی نمی‌شوند زیرا داده‌های ناموفق اغلب در دسترس عموم قرار نمی‌گیرند. علاوه بر مشکل کیفیت و تعادل داده‌ها، برخی از انواع داده‌ها عموماً در دسترس محققان نیستند. برای مثال، داده‌های بالینی شامل حریم خصوصی بیمار است و معمولاً در دسترس عموم نیستند (۶۴). علاوه بر محدودیت‌های داده، روش‌های AI فاقد قابلیت تفسیر هستند. برای مثال، هنگامی که یک مدل DL نتیجه جدیدی به دست می‌دهد که با تحقیقات قبلی در تضاد است، عدم تفسیرپذیری نتیجه را غیرقابل قبول می‌کند. DL فقط می‌تواند ورودی را بپذیرد و خروجی پیش‌بینی شده را ارائه دهد اما نمی‌تواند توضیحات کافی برای چگونگی استخراج این خروجی ارائه دهد. به عنوان مثال، برای

بالینی (حدود ۹۰ درصد از نامزدهای دارویی در آزمایش‌های بالینی حذف می‌شوند)، آن را به سخت‌ترین مرحله در توسعه داروی جدید تبدیل می‌کند، که در آن هر آزمایش بالینی ناموفق تقریباً ۰/۸ تا ۱/۴ میلیارد دلار هزینه دارد (۵۴). اگرچه هوش مصنوعی نمی‌تواند مستقیماً برای پیش‌بینی نتایج کارآزمایی بالینی داوطلبان دارو استفاده شود، می‌توان از آن برای کمک به طراحی اصولی‌تر آزمایش‌های بالینی به منظور ارثه‌ی پاسخ واقعی‌تر به نتایج کارآزمایی بالینی دارو استفاده کرد. در حال حاضر برای غلبه بر این کاستی‌ها، چندین رویکرد مبتنی بر هوش مصنوعی برای کمک به انتخاب و غربالگری بیمار و افزایش نظارت بر بیمار، در دسترس هستند (۵۵). در موضوع انتخاب بیمار، هوش مصنوعی می‌تواند برای بررسی ارتباط بیومارکرهای بیمار با نشانه‌های خارجی برای پیش‌بینی پاسخ احتمالی درمانی بیماران استفاده شود، که می‌تواند به غربالگری بیماران با موفقیت بالینی بالا کمک کند. علاوه بر این، فنوتیپ الکترونیکی می‌تواند ناهمگونی جمعیت بیمار را کاهش دهد و از طریق غنی‌سازی پیش‌آگهی یا پیش‌بینی به انتخاب بیمار کمک کند (۵۶).

نظارت بر بیمار در آزمایشات بالینی نیز یک فرآیند حیاتی است. با ترکیب فناوری پوشیدنی مانند ساعت‌های هوشمند، هوش مصنوعی می‌تواند به خودکار سازی و شخصی سازی نظارت بر بیمار در زمان واقعی کمک کند، در نتیجه حجم کاری بیمار را کاهش می‌دهد و مشکلات پایبندی به دارو را بهبود می‌بخشد. داده‌های دقیق پایبندی به دارو می‌تواند نتایج آزمایش‌های بالینی را بهتر منعکس کند. AiCure یک پلتفرم جدید هوش مصنوعی است که برای اندازه‌گیری پایبندی به دارو استفاده می‌شود. علاوه بر این، هوش مصنوعی برای بهینه‌سازی دوز برای کاهش عوارض جانبی، بهبود ایمنی پروتکل‌های آزمایشی و کاهش پیش فرض‌های بیمار به دلیل نگرانی‌های ایمنی استفاده شده است (۵۷).

#### نظارت پس از فروش

پس از ورود موفقیت آمیز یک دارو به بازار، بررسی طولانی مدت برای نظارت بیشتر و ارزیابی ایمنی دارو الزامی است. استخراج پرونده الکترونیک سلامت (EHR (Electronic health record)) یک منبع داده مهم برای کاربردهای هوش مصنوعی در نظارت پس از فروش است که در آن استفاده از داده‌های ساختاریافته می‌تواند فرایند پیش پردازش داده‌ها را ساده کند. روش‌های موجود مورد استفاده در EHR شامل مدل سری موارد خودکنترلی (SCCS (Self-control case series)) (۵۸)، روش‌های همگروهی و مورد-کنترلی (۵۹) و الگوریتم‌های کشف الگوی زمانی است (۶۰). مدل خودکنترلی کانولوشنل ConvSCCS (Convolutional SCCS) یک مدل مقیاس پذیر برای پیش‌بینی ویژگی‌های طولی با استفاده از SCCS است. Morel و همکاران از این مدل محاسباتی در تشخیص واکنش‌های نامطلوب دارویی (ADRs (Adverse drug reactions))



صنعت داروسازی به سرعت در حال تکامل است و کاربرد هوش مصنوعی در جنبه‌های مختلف این صنعت از کشف دارو تا مدیریت زنجیره‌ی تأمین، روز به روز در حال افزایش است. داروسازی محاسباتی که توسط هوش مصنوعی و داده‌های بزرگ تسهیل می‌شود، با ارائه‌ی رویکردهای کارآمدتر، مقرون به صرفه و مبتنی بر داده، توسعه و تحویل دارو را متحول می‌کند. با این حال با توجه به محدودیت داده‌های بیولوژیک، عدم دسترسی به تمامی داده‌ها و عدم تفسیرپذیری نتایج که با نتایج بدست آمده از روش‌های سنتی توسعه‌ی دارو مغایرت دارند، کاربرد هوش مصنوعی در علوم دارویی با چالش رو به رو شده است. لذا تحقیقات بیشتری برای استفاده از این تکنولوژی جدید در حوزه‌ی علوم دارویی مورد نیاز است.

### تشکر و قدردانی

بدین وسیله از تمام کسانی که در انجام این پژوهش ما را یاری کردند سپاسگزاریم.

حاشیه نویسی عملکرد پروتئین، اگرچه روش‌های DL می‌توانند (GOA (Gene ontology annotation) یک پروتئین خاص را پیش‌بینی کنند (۶۵)، فرایند محاسباتی شناخته شده نیست و اکثر پیش‌بینی‌ها زمانی که دقت قابل اعتماد نباشد پذیرفته نمی‌شوند. حتی از نظر روش‌های نمایش داده‌ها، هیچ استاندارد یکسانی در مورد اینکه کدام روش نمایش برای کدام مطالعه مناسب‌تر است و کدام روش‌های نمایش منجر به از دست رفتن اطلاعات می‌شود، ایجاد نشده است.

### یافته‌ها

مشابه مدل یادگیری انسان، یادگیری ماشین و یادگیری عمیق می‌توانند به تدریج ویژگی‌های مختلف داده‌ها را تشخیص دهند و پارامترهای مدل خود را از طریق تکرارهای مداوم به روز کنند تا زمانی که یک مدل معتبر تشکیل شود..

### نتیجه‌گیری

### References

- DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016; 47: 20-33.
- Clark DE, editor. *Evolutionary algorithms in molecular design*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 1999.
- Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010; 9(3): 203-14.
- Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008; 4(11): 682-90.
- Shoichet BK. Virtual screening of chemical libraries. *Nature* 2004; 432(7019): 862-5.
- Khorasgani RG, Sanaei M, Beigi MM. Computer-aided vaccine design: a brief report [in Persian]. *Tehran Univ Med J* 2012; 70(6): 383-8.
- Aghamollaei H, Mousavi Gargari SL, Ghanei M, Rasaei MJ, Amani J, Bakherad H, et al. Structure prediction, expression, and antigenicity of c-terminal of GRP78. *Biotechnol Appl Biochem* 2017; 64(1): 117-25.
- Mak KK, Balijepalli MK, Pichika MR. Success stories of AI in drug discovery-where do things stand?. *Expert Opin Drug Discov* 2022; 17(1): 79-92.
- Wills T. AI drug discovery: assessing the first AI-designed drug candidates to go into human clinical trials. [online 2022]. Available from: <https://www.cas.org/resources/cas-insights/ai-drug-discovery-assessing-the-first-ai-designed-drug-candidates-to-go-into-human-clinical-trials>
- Lu M, Yin J, Zhu Q, Lin G, Mou M, Liu F, et al. Artificial intelligence in pharmaceutical sciences. *Engineering* 2023; 27: 37-69.
- Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. *Anesthesiology* 2020; 132(2): 379-94.
- Rifaioglu AS, Atas H, Martin MJ, Cetin-Atalay R, Atalay V, Doğan T. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Brief Bioinform* 2019; 20(5): 1878-912.
- Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, Cedrón F, Novoa FJ, Carballal A, et al. A review on machine learning approaches and trends in drug discovery. *Comput Struct Biotechnol J* 2021; 19: 4538-58.
- Bengio Y. "Learning Deep Architectures for AI", *Foundations and Trends® in Machine Learning* 2009; 2(1): 1-127.
- Ahmed IT, Der CS, Jamil N, Mohamed MA. Improve of contrast-distorted image quality assessment based on convolutional neural networks. *International Journal of Electrical & Computer Engineering* (2088-8708). 2019; 9(6): 5604-14.
- Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov* 2011; 10(7): 507-19
- Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nat Rev Drug Discov* 2022; 21(12): 899-914.
- Li YH, Li XX, Hong JJ, Wang YX, Fu JB, Yang H, et al. Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs. *Brief Bioinform* 2020; 21(2): 649-62.
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; 45(D1): D353-D361.

20. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol Adv* 2021; 49: 107739.
21. Jeon J, Nim S, Teyra J, Datti A, Wrana JL, Sidhu SS, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Med* 2014; 6(7): 57.
22. Wang X, Shen Y, Wang S, Li S, Zhang W, Liu X, Lai L, Pei J, Li H. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res* 2017; 45(W1): W356-60.
23. Gong J, Cai C, Liu X, Ku X, Jiang H, Gao D, et al. ChemMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. *Bioinformatics* 2013; 29(14): 1827-9.
24. Wang X, Chen H, Yang F, Gong J, Li S, Pei J, et al. i Drug: a web-accessible and interactive drug discovery and design platform. *J Cheminform* 2014; 6: 28.
25. Pabon NA, Xia Y, Estabrooks SK, Ye Z, Herbrand AK, Süß E, et al. Predicting protein targets for drug-like compounds using transcriptomics. *PLoS Computational Biology* 2018; 14(12): e1006651.
26. Zhong F, Wu X, Yang R, Li X, Wang D, Fu Z, et al. Drug target inference by mining transcriptional data using a novel graph convolutional network framework. *Protein Cell* 2022; 13(4): 281-301.
27. Jaganathan K, Panagiotopoulou SK, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. *Cell* 2019; 176(3): 535-548.e24.
28. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018; 15(12): 1053-8.
29. Keiser MJ, Setola V, Irwin JJ, Lagner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. *Nature* 2009; 462(7270): 175-81.
30. Ballante F, Kooistra AJ, Kampen S, de Graaf C, Carlsson J, Schulte G. Structure-based virtual screening for ligands of G protein-coupled receptors: what can molecular docking do for you?. *Pharmacol Rev* 2021; 73(4): 1698-736.
31. Khalilzadeh M, Saberi S, Noori G, Vaziri M, Sepehri S, Bakherad H, et al. Synthesis, biological assessment, and computational investigations of nifedipine and monastrol analogues as anti-leishmanial major and anti-microbial agents. *Mol Divers* 2023; 27(6): 2555-75.
32. Bakherad H, Farahmand M, Setayesh N, Ebrahim-Habibi A. Engineering an anti-granulocyte colony stimulating factor receptor nanobody for improved affinity. *Life Sci* 2020; 257: 118052.
33. Rezaei Adriani R, Mousavi Gargari SL, Bakherad H, Amani J. In silico designing and optimization of anti-epidermal growth factor receptor scaffolds by complementary-determining regions-grafting technique. *Quantitative Biology* 2024; 12(3): 301-12.
34. Li GB, Yang LL, Wang WJ, Li LL, Yang SY. ID-Score: a new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *J Chem Inf Model* 2013; 53(3): 592-600.
35. Xie QQ, Zhong L, Pan YL, Wang XY, Zhou JP, Di-Wu L, Huang Q, Wang YL, Yang LL, Xie HZ, Yang SY. Combined SVM-based and docking-based virtual screening for retrieving novel inhibitors of c-Met. *Eur J Med Chem* 2011; 46(9): 3675-80.
36. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design-a review. *Curr Top Med Chem* 2010; 10(1): 95-115.
37. AbdulHameed MD, Ippolito DL, Wallqvist A. Predicting rat and human pregnane X receptor activators using Bayesian classification models. *Chem Res Toxicol* 2016; 29(10): 1729-40.
38. Martin EJ, Polyakov VR, Tian L, Perez RC. Profile-QSAR 2.0: kinase virtual screening accuracy comparable to four-concentration IC50s for realistically novel compounds. *J Chem Inf Model* 2017; 57(8): 2077-88.
39. Li Y, Wang L, Liu Z, Li C, Xu J, Gu Q, Xu J. Predicting selective liver X receptor  $\beta$  agonists using multiple machine learning methods. *Mol Biosyst* 2015; 11(5): 1241-50.
40. Cao DS, Liu S, Xu QS, Lu HM, Huang JH, Hu QN, et al. Large-scale prediction of drug-target interactions using protein sequences and drug topological structures. *Anal Chim Acta* 2012; 752: 1-0.
41. Mitchell MJ, Billingsley MM, Haley RM, Wechsler ME, Peppas NA, Langer R. Engineering precision nanoparticles for drug delivery. *Nat Rev Drug Discov* 2021; 20(2): 101-24.
42. Vora LK, Gholap AD, Jetha K, Thakur RR, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics* 2023; 15(7): 1916.
43. Shamay Y, Shah J, Işık M, Mizrahi A, Leibold J, Tschaharganeh DF, et al. Quantitative self-assembly prediction yields targeted nanomedicines. *Nature materials* 2018; 17(4): 361-8.
44. Lu Y, Aimetti AA, Langer R, Gu Z. Bioresponsive materials. *Nature Reviews Materials* 2016; 2(1): 1-7.
45. Meanwell NA. Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. *Chem Res Toxicol* 2011; 24(9): 1420-56.
46. Tamardash LK, Bakherad M, Bakherad H, Jalali F, Mozafari Z, Keivanloo A. Synthesis, and molecular docking studies of novel 1, 2, 3-triazoles-linked pyrazole carboxamides as significant anti-microbial and anti-cancer agents. *Results in Chemistry* 2024; 11: 101842.
47. Manallack DT, Prankerd RJ, Yuriev E, Oprea TI, Chalmers DK. The significance of acid/base properties in drug discovery. *Chem Soc Rev* 2013; 42(2): 485-96.
48. Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem* 2019; 63(16): 8749-60.
49. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, Guzman-Perez A, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model* 2019; 59(8): 3370-88.

50. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010; 50(5): 742-54.
51. Tao L, Zhang P, Qin C, Chen SY, Zhang C, Chen Z, et al. Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. *Adv Drug Deliv Rev* 2015; 86: 83-100.
52. Radchenko EV, Dyabina AS, Palyulin VA. Towards deep neural network models for the prediction of the blood-brain barrier permeability for diverse organic compounds. *Molecules* 2020; 25(24): 5901.
53. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic acids research* 2021; 49(W1): W5-14.
54. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol* 2014; 32(1): 40-51.
55. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci* 2019; 40(8): 577-91.
56. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in electronic phenotyping: from rule-based definitions to machine learning models. *Annu Rev Biomed Data Sci* 2018; 1(1): 53-68.
57. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, Hanina A. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. *JMIR Mhealth Uhealth* 2017; 5(2): e18.
58. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995;51(1): 228-35.
59. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, Reich CG, Schuemie MJ, Madigan D. A comparison of the empirical performance of methods for a risk identification system. *Drug Saf* 2013; 36(Suppl 1): 143-58.
60. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Disc* 2010; 20: 361-87.
61. Morel M, Bacry E, Gaïffas S, Guilloux A, Leroy F. ConvSCCS: convolutional self-controlled case series model for lagged adverse event detection. *Biostatistics* 2020; 21(4): 758-74.
62. Mower J, Subramanian D, Cohen T. Learning predictive models of drug side-effect relationships from distributed representations of literature-derived semantic predications. *J Am Med Inform Assoc* 2018; 25(10): 1339-50.
63. Zhang C, Mou M, Zhou Y, Zhang W, Lian X, Shi S, et al. Biological activities of drug inactive ingredients. *Brief Bioinform* 2022; 23(5): bbac160.
64. Khan B, Fatima H, Qureshi A, Kumar S, Hanan A, Hussain J, et al. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomed Mater Devices* 2023; 1(2): 731-8.
65. Xia W, Zheng L, Fang J, Li F, Zhou Y, Zeng Z, Zhang B, Li Z, Li H, Zhu F. PFMulDL: a novel strategy enabling multi-class and multi-label protein function annotation by integrating diverse deep learning methods. *Comput Biol Med* 2022; 145: 105465.

## The Role of Artificial Intelligence in Pharmaceutical Sciences

Marzieh Sanaei<sup>1</sup>, Mohammadebrahim Delzende Nedamani<sup>2</sup>, Hamid Bakherad<sup>3</sup>

### Review Article

#### Abstract

**Background:** Drug discovery and development affects human health and the drug market. However, investing in a new drug is often a complex and difficult challenge due to the long and complex drug research and development process.

**Methods:** With the advancement of experimental technology and computer hardware, artificial intelligence has recently emerged as a tool for analyzing abundant and high-dimensional data. Explosive growth in the size of biological data provides advantages in applying artificial intelligence in all stages of pharmaceutical research and development.

**Findings:** Similar to human learning models, machine learning and deep learning can gradually recognize different features of data, and update their model parameters through continuous iterations until a valid model is formed.

**Conclusion:** This article begins with a brief overview of common AI models in drug discovery. Then, it briefly discusses their specific applications in different drug research and development phases, such as target discovery, drug discovery and design, preclinical research, and effects on the drug market. Finally, major limitations of artificial intelligence in pharmaceutical research and development are discussed.

**Keywords:** Drug design, Artificial intelligence, Deep learning, Machine learning

**Citation:** Sanaei M, Delzende Nedamani M, Bakherad H. **The Role of Artificial Intelligence in Pharmaceutical Sciences.** J Isfahan Med Sch 2025; 43(813): 451-62.

1- PhD. Pharmaceutical Biotechnology, School of Pharmacy, Tehran University of Medical Sciences, Tehran, Iran

2- MSc of Mechatronics, School of Mechanics, Semnan University, Semnan, Iran

3-Associate Professor of Biotechnology, School of Pharmacy, Isfahan University of Medical Sciences, Isfahan, Iran

**Corresponding Author:** Hamid Bakherad, Assistant professor, School of Pharmacy, Isfahan University of Medical Sciences, Isfahan, Iran; Email: bakheradh@pharm.mui.ac.ir