

هوش مصنوعی؛ راهی نو برای کشف سریع و ایمن داروها

علی اصغر اسدی^{*۱}

۱ - کارشناسی رشته مهندسی تکنولوژی نرم افزار کامپیوتر، دانشگاه لیان، لیان، ایران .

(کارشناس در آمد شهرداری منطقه ۲)

چکیده

هوش مصنوعی به عنوان یکی از تحول آفرین ترین فناوری های معاصر، توانسته است مسیر کشف و توسعه دارو را به طور بنیادین دگرگون سازد. فرآیند سنتی کشف دارو همواره زمان بر، پرهزینه و با نرخ شکست بالا بوده است؛ اما ظهور الگوریتم های یادگیری ماشین و یادگیری عمیق، امکان تحلیل داده های عظیم زیستی و شیمیایی را فراهم کرده و موجب افزایش سرعت، دقت و کارایی این چرخه شده است. هدف این پژوهش، بررسی نقش هوش مصنوعی در بهبود فرآیند کشف سریع تر و ایمن تر داروها و تبیین ابعاد فنی و کاربردی آن در طراحی مولکول، پیش بینی ساختار پروتئین و اعتبارسنجی اهداف درمانی است. روش تحقیق به صورت مروری و تحلیلی بر اساس مطالعات علمی اخیر انجام شد و به بررسی رویکردهای مختلف از جمله یادگیری نظارت شده، بدون نظارت، تقویتی و مدل های مولد پرداخته شد. نتایج نشان داد که مدل های یادگیری عمیق، به ویژه شبکه های عصبی گرافی و ترانسفورمرها، توانایی پیش بینی فعالیت زیستی ترکیبات و طراحی ده نوو مولکول های دارویی را با دقت بالایی دارند. همچنین، پیشرفت هایی نظیر مدل های پیش بینی ساختار پروتئینی بر پایه یادگیری عمیق، مانند AlphaFold، موجب تسریع چشمگیر فرآیند طراحی مبتنی بر ساختار شده اند. یافته ها بیانگر آن است که کاربرد هوش مصنوعی در غربالگری مجازی، بهینه سازی چندهدفه و کشف مجدد داروها باعث کاهش هزینه و زمان توسعه، و در عین حال افزایش میزان موفقیت داروهای کاندیدا شده است. با این وجود، چالش هایی همچون کمبود داده های استاندارد، سوگیری داده ای، و نیاز به شفافیت تصمیمات مدل ها هنوز پابرجاست و باید مورد توجه قرار گیرد. در نتیجه، می توان گفت هوش مصنوعی نه تنها ابزاری کمکی، بلکه یک عنصر حیاتی در آینده کشف دارو است که با ترکیب علم داده، زیست شناسی و مهندسی نرم افزار، امکان طراحی دقیق تر، ایمن تر و مقرون به صرفه تر داروها را فراهم می سازد و افق های جدیدی را در درمان بیماری های پیچیده و نادر پیش روی محققان می گشاید.

واژگان کلیدی: هوش مصنوعی، کشف دارو، یادگیری ماشین، طراحی مولکولی، پیش بینی سمیت

مقدمه

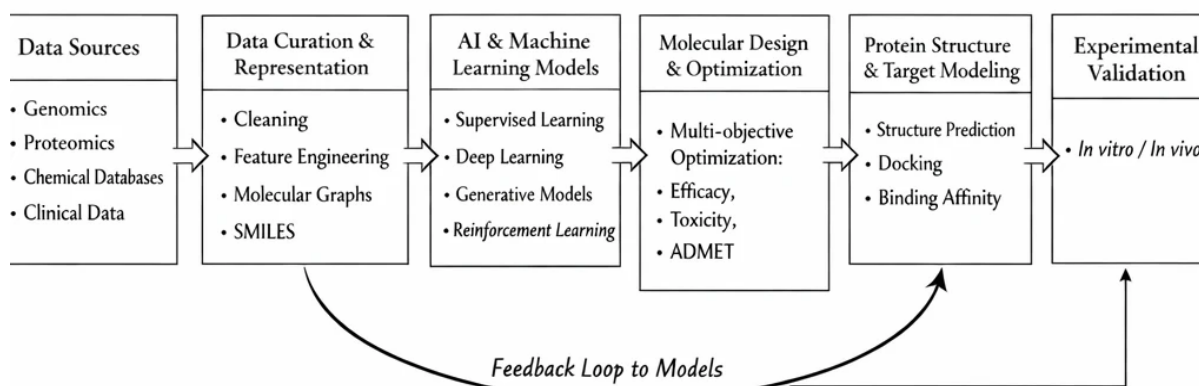
حوزه کشف و توسعه دارو (Drug Discovery and Development - DDD) همواره به عنوان یکی از پیچیده ترین و پرچالش ترین حوزه های علمی و صنعتی مطرح بوده است، که با نرخ شکست بالا و سرمایه گذاری های هنگفت همراه است. فرآیند سنتی، که شامل شناسایی اهداف، کشف ترکیبات پیشرو، بهینه سازی پیشرو، مطالعات پیش بالینی و در نهایت کارآزمایی های بالینی است، به طور متوسط بیش از یک دهه و نیم به طول می انجامد و هزینه هایی بالغ بر میلیارد ها دلار را بر سازمان های دارویی تحمیل می کند (Fleming, 2018). با توجه به افزایش شیوع بیماری های مزمن، ظهور پاتوژن های مقاوم به دارو، و نیاز مبرم به درمان های نوآورانه برای بیماری های نادر، فشارهای زیادی بر سیستم DDD وارد شده است تا کارایی خود را به شکل چشمگیری افزایش دهد. در این بستر، فناوری های محاسباتی، به ویژه هوش مصنوعی (AI) و زیرشاخه های آن مانند یادگیری ماشین (ML) و یادگیری عمیق (DL)، به عنوان یک نیروی دگرگون کننده ظهور کرده اند که پتانسیل بازتعریف کل چرخه عمر کشف دارو را دارند.

هوش مصنوعی، با قابلیت های بی نظیر خود در استخراج الگوهای پنهان از مجموعه های داده های عظیم و چندوجهی بیولوژیکی و شیمیایی (از جمله داده های ژنومیک، پروتئومیک، سوابق الکترونیکی سلامت، و پایگاه های داده ترکیبات شیمیایی)، می تواند فرآیندهای زمان بر و اغلب مبتنی بر حدس و گمان را با مدل های پیش بین و توصیه گر جایگزین سازد. این جایگزینی، امکان غربالگری سریع تر میلیون ها ترکیب بالقوه، پیش بینی دقیق تر سمیت و کارایی دارویی، و حتی طراحی ساختارهای مولکولی جدیدی که تاکنون تصور نمی شدند را فراهم می آورد. قدرت اصلی AI در این زمینه، توانایی آن در مدل سازی سیستم های زیستی با پیچیدگی بسیار بالا است؛ مسائلی که برای محاسبات سنتی مبتنی بر شبیه سازی های فیزیکی یا روش های مبتنی بر غربالگری با توان عملیاتی بالا (HTS) بسیار دشوار بوده اند (Gupta et al., 2021).

کاربردهای AI در این حوزه، از ابتدایی ترین مراحل، یعنی انتخاب اهداف مناسب برای دارو درمانی، آغاز می شود. شناسایی پروتئین ها یا مسیرهای بیوشیمیایی که در بروز یک بیماری نقش حیاتی دارند، خود یک چالش بزرگ است. الگوریتم های ML می توانند با تحلیل داده های ترانسکریپتومیکس و پروتئومیکس بیماران، اهداف درمانی جدیدی را که از طریق روش های سنتی قابل شناسایی نبودند، کشف کنند. پس از شناسایی هدف، چالش اصلی به یافتن مولکول هایی معطوف می شود که بتوانند با اتصال به آن هدف، فعالیت بیولوژیکی مورد نظر را تعدیل کنند. در این مرحله، مدل های یادگیری عمیق، به ویژه آن هایی که برای کار با نمایش های ساختاری مولکولی مانند گراف ها یا بردارهای ویژگی (Descriptors) آموزش دیده اند، قادرند خواص فیزیکوشیمیایی و فعالیت زیستی هزاران ترکیب را در کسری از ثانیه پیش بینی کنند، که این امر غربالگری مجازی (Virtual Screening) را به ابزاری قدرتمند تبدیل کرده است (Chen et al., 2023).

علاوه بر غربالگری، نوآوری های قابل توجهی در زمینه طراحی ده نوو (De Novo Drug Design) با استفاده از مدل های مولد مانند شبکه های مولد تخاصمی (GANs) و رمزگذارهای خودکار متغیر (VAEs) به دست آمده است. این مدل ها می توانند فضای شیمیایی را پیمایش کرده و مولکول هایی با مجموعه ای از ویژگی های مطلوب (مانند میل ترکیبی بالا به هدف، حلالیت مناسب، و سمیت پایین) به طور خودکار تولید کنند (Zhu, 2020). این رویکرد انقلابی، مسیر را از انتخاب از میان مجموعه های موجود به سمت خلق ترکیبات کاملاً جدید هموار می سازد. این تحول نیازمند تخصص عمیق در

مهندسی نرم افزار و یادگیری ماشین برای توسعه چارچوب‌های محاسباتی مقیاس پذیر و قابل اعتماد است تا بتوانند این ایده‌های مولکولی را به واقعیت آزمایشگاهی برسانند. در نهایت، پیشرفت‌های اخیر در مدل سازی ساختار سه بعدی پروتئین‌ها، که به دقت بی سابقه‌ای دست یافته است، تأثیر مستقیمی بر طراحی داروهای مبتنی بر ساختار (Structure-Based Drug Design) داشته است. این مدل‌ها، که اغلب بر پایه معماری‌های یادگیری عمیق پیشرفته بنا شده‌اند، می‌توانند ساختار فضایی پروتئین‌ها را تنها بر اساس توالی آمینواسیدی آن‌ها پیش‌بینی کنند (Patel & Shah, 2022). این توانایی، به محققان اجازه می‌دهد تا مدل‌های دقیق‌تری از اهداف دارویی خود داشته باشند و بهینه‌سازی اتصال مولکول‌های کوچک به این ساختارها را با دقت بیشتری انجام دهند، که این امر به کاهش ابهامات در مراحل اولیه طراحی کمک شایانی می‌کند. بنابراین، یکپارچه‌سازی موفقیت‌آمیز این ابزارهای محاسباتی با دانش بیولوژیکی و شیمیایی، سنگ بنای نسل بعدی کشف داروهای کارآمد و مقرون به صرفه خواهد بود.



شکل ۱. چارچوب مفهومی کشف دارو مبتنی بر هوش مصنوعی (AI-Driven Drug Discovery Framework)

در شکل ۱، یک چارچوب مفهومی جامع از فرآیند کشف دارو مبتنی بر هوش مصنوعی نمایش داده شده است که جریان داده و تصمیم‌گیری را از مراحل اولیه تا اعتبارسنجی آزمایشگاهی نشان می‌دهد. این چارچوب با منابع داده‌ای متنوع شامل داده‌های ژنومیک، پروتئومیک، پایگاه‌های داده شیمیایی و اطلاعات بالینی آغاز می‌شود. سپس داده‌ها وارد مرحله پاک‌سازی، یکپارچه‌سازی و بازنمایی می‌شوند، جایی که تکنیک‌هایی مانند مهندسی ویژگی، نمایش مولکولی مبتنی بر گراف و فرمت‌های استاندارد نظیر SMILES برای آماده‌سازی داده‌ها جهت استفاده در مدل‌های یادگیری ماشین به کار گرفته می‌شود. این مرحله نقش حیاتی در افزایش دقت و قابلیت تعمیم‌پذیری مدل‌های هوشمند دارد. در ادامه، مدل‌های هوش مصنوعی و یادگیری ماشین شامل یادگیری نظارت‌شده، یادگیری عمیق، مدل‌های مولد و یادگیری تقویتی برای پیش‌بینی فعالیت زیستی، طراحی مولکول‌های جدید و بهینه‌سازی چندهدفه به کار گرفته می‌شوند. خروجی این مدل‌ها در مرحله طراحی و بهینه‌سازی مولکولی با در نظر گرفتن معیارهایی مانند اثربخشی، سمیت و ویژگی‌های ADMET پالایش می‌شود. سپس، از مدل‌سازی ساختار پروتئین و هدف دارویی برای پیش‌بینی ساختار سه بعدی، داکینگ و ارزیابی

میل اتصال استفاده می‌گردد. در نهایت، نتایج منتخب وارد مرحله اعتبارسنجی آزمایشگاهی (in vitro / in vivo) می‌شوند و بازخورد حاصل از این آزمایش‌ها به صورت یک حلقه بازخورد به مدل‌های هوش مصنوعی بازگردانده می‌شود تا بهبود مستمر عملکرد سیستم و کاهش ریسک شکست در مراحل بعدی توسعه دارو تضمین گردد.

۲. مبانی نظری و سیر تاریخی

ریشه‌های کاربرد محاسبات در شیمی دارویی به دهه‌های پیش بازمی‌گردد، اما انفجار واقعی این حوزه با ظهور قدرت محاسباتی کافی برای اجرای مدل‌های پیچیده یادگیری ماشین همراه شد. در ابتدا، رویکردهای محاسباتی اغلب متکی بر مدل‌سازی مبتنی بر اصول اولیه (First-Principles Modeling)، مانند دینامیک مولکولی (MD) و شبیه‌سازی‌های کوانتومی بودند؛ این روش‌ها اگرچه از نظر تئوری دقیق بودند، اما از نظر محاسباتی بسیار سنگین بوده و تنها امکان بررسی سیستم‌های کوچک و بازه‌های زمانی کوتاه را فراهم می‌ساختند (Sellwood et al., 2018). در دهه‌های ۱۹۸۰ و ۱۹۹۰، روش‌های شیمی‌انفورماتیک مبتنی بر استخراج ویژگی‌های کمی ساختار و فعالیت (Quantitative Structure-Activity Relationship - QSAR) با استفاده از رگرسیون‌های خطی و آماری توسعه یافتند، اما این مدل‌ها غالباً در تعمیم‌پذیری و پوشش فضای شیمیایی وسیع دچار محدودیت بودند. گذار کلیدی در اوایل قرن بیست و یکم، با ظهور یادگیری ماشین کلاسیک، نظیر ماشین‌های بردار پشتیبان (SVMs) و جنگل‌های تصادفی (Random Forests)، رخ داد. این الگوریتم‌ها قادر بودند روابط غیرخطی پیچیده‌تری را بین توصیف‌گرهای مولکولی (مانند لگاریتم ضریب توزیع یا جرم مولکولی) و فعالیت‌های بیولوژیکی مدل‌سازی کنند. این دوره شاهد استفاده گسترده از روش‌های شیمی‌انفورماتیک برای پیش‌بینی سمیت، فارماکوکینتیک (ADMET) و میل ترکیبی (Affinity) بود، که اولین گام‌های جدی در جهت جایگزینی بخشی از آزمایشات غربالگری بود (Fleming, 2018). با این حال، موفقیت این مدل‌ها به شدت وابسته به کیفیت و کمیت ویژگی‌های استخراج‌شده توسط مهندسان شیمی‌دان بود، که اغلب استخراج ویژگی‌های بهینه و کامل برای هر نوع مشکل بیولوژیکی جدید، دشوار بود.

نقطه عطف اصلی، ظهور یادگیری عمیق (Deep Learning) در دهه ۲۰۱۰ بود که امکان یادگیری خودکار سلسله‌مراتبی از ویژگی‌ها را از داده‌های خام فراهم آورد و نیاز به مهندسی دستی ویژگی‌ها را کاهش داد. شبکه‌های عصبی عمیق، به ویژه زمانی که برای پردازش داده‌های ساختاری مانند گراف‌ها (با استفاده از شبکه‌های عصبی گرافی - GNNs) یا توالی‌ها (با استفاده از RNNs و ترانسفورمرها) به کار گرفته شدند، توانایی مدل‌سازی بهتر ماهیت سه‌بعدی و تعاملی مولکول‌ها و پروتئین‌ها را نشان دادند (Jiménez-Luna et al., 2021). این پیشرفت‌ها به‌ویژه در حوزه غربالگری مجازی به نتایج خیره‌کننده‌ای منجر شدند، جایی که مدل‌های یادگیری عمیق توانستند با دقتی بالاتر از روش‌های QSAR سنتی، فعالیت ترکیبات را پیش‌بینی کنند. سیر تاریخی هوش مصنوعی در کشف دارو را می‌توان به سه مرحله اصلی تقسیم کرد: مرحله اول (تا حدود ۲۰۱۰) مبتنی بر آمار و QSAR کلاسیک بود؛ مرحله دوم (۲۰۱۰ تا حدود ۲۰۱۸) شاهد پذیرش اولیه شبکه‌های عصبی ساده‌تر و پیشرفت در مدل‌سازی مولکولی مبتنی بر توصیف‌گرهای از پیش تعریف‌شده بود؛ و مرحله سوم (از ۲۰۱۸ به بعد) که با قدرت یادگیری عمیق و معماری‌های پیشرفته مانند ترانسفورمرها و مدل‌های مولد تعریف می‌شود، متمرکز بر طراحی ده نو و مدل‌سازی ساختارهای پروتئینی پیچیده است (Deng et al., 2022). این تحولات اخیر، که از پیشرفت‌های مشابه در حوزه پردازش زبان طبیعی و بینایی ماشین الهام گرفته‌اند، امکان ایجاد مدل‌هایی را

فراهم کرده‌اند که نه تنها ویژگی‌های موجود را پیش‌بینی می‌کنند، بلکه می‌توانند ترکیبات کاملاً جدیدی را در فضای شیمیایی با اهداف خاص تولید کنند.

پیشرفت در مدل‌سازی پروتئین، که با الگوریتم‌هایی نظیر AlphaFold نشان داده شد، یکی از مهم‌ترین دستاوردهای این دوره است. توانایی پیش‌بینی ساختار سه‌بعدی پروتئین‌ها با دقتی نزدیک به روش‌های تجربی، چالش بزرگ مدل‌سازی اهداف دارویی را حل کرده است و یک زیربنای اساسی برای شبیه‌سازی‌های اتصال دارو-هدف مبتنی بر هوش مصنوعی فراهم آورده است (Patel & Shah, 2022). این پیشرفت‌ها نشان می‌دهند که هوش مصنوعی از صرفاً یک ابزار تحلیلی به یک ابزار مولد و طراحی در شیمی دارویی تبدیل شده است، که مرزهای دانش در مورد چگونگی تعامل مولکول‌ها و بیولوژی را به طرز چشمگیری جابجا می‌کند و وابستگی به غربالگری‌های آزمایشی را کاهش می‌دهد.

۳. داده‌ها در کشف دارو و چالش‌های کیفیت

موفقیت هر سیستم هوش مصنوعی، به‌ویژه در حوزه‌های علمی و مهندسی پیچیده‌ای مانند کشف دارو، به‌شدت به کمیت، کیفیت، و تنوع داده‌های آموزشی بستگی دارد. در حوزه داروسازی، مجموعه داده‌ها شامل ابعاد مختلفی هستند: داده‌های شیمیایی (ساختارها، ویژگی‌های فیزیکوشیمیایی، پایگاه‌های داده ترکیبات)، داده‌های بیولوژیکی (دنباله‌های ژنومی، ساختارهای پروتئینی، داده‌های بیان ژن، نتایج غربالگری فعالیت زیستی) و داده‌های بالینی (نتایج کارآزمایی‌ها و سوابق بیماران). با وجود حجم عظیم داده‌های تولیدشده در دهه‌های اخیر، دسترسی، استانداردسازی و اطمینان از کیفیت این داده‌ها همچنان بزرگ‌ترین موانع پیش روی هوش مصنوعی در این صنعت باقی مانده‌اند (Chan et al., 2019). یکی از چالش‌های اصلی، کمبود داده‌های "منفی" یا دقیقاً "عدم فعالیت" است. در بسیاری از آزمایشگاه‌ها، تنها نتایج مثبت (ترکیبات فعال) ثبت و منتشر می‌شوند، در حالی که داده‌های مربوط به ترکیبات غیرفعال که تلاش‌های زیادی برای سنتز و آزمایش آن‌ها شده است، منتشر نمی‌شوند. این سوگیری در داده‌ها (Data Bias) باعث می‌شود مدل‌های یادگیری ماشین بیش از حد به فعال بودن تمایل پیدا کنند و توانایی آن‌ها در تشخیص دقیق ترکیبات غیرفعال یا پیش‌بینی عدم سمیت کاهش یابد. علاوه بر این، داده‌های فعالیت زیستی اغلب به‌صورت نسبی (مثلاً EC50 یا IC50) گزارش می‌شوند که خود دارای خطای اندازه‌گیری قابل توجهی هستند و این نویز ذاتی در داده‌ها، یادگیری مدل‌های با دقت بالا را دشوار می‌سازد (Gupta et al., 2021).

استانداردسازی و قابلیت همکاری (Interoperability) داده‌ها نیز یک مانع ساختاری مهم است. داده‌های جمع‌آوری شده از منابع مختلف (دانشگاه‌ها، شرکت‌های دارویی، پایگاه‌های داده عمومی مانند PubChem یا ChEMBL) ممکن است از پروتکل‌های آزمایشی، فرمت‌های شیمیایی یا روش‌های اندازه‌گیری متفاوتی استفاده کنند. تبدیل این داده‌های ناهمگن به یک فرمت یکپارچه و قابل استفاده برای مدل‌های یادگیری عمیق، نیازمند تلاش مهندسی نرم‌افزار بسیار سنگینی است. استفاده از فرمت‌های استاندارد مانند SMILES برای نمایش مولکول‌ها یا PDB برای ساختارهای پروتئینی ضروری است، اما مدیریت تغییرات و تفسیرهای مختلف این فرمت‌ها همچنان نیازمند مداخله‌های دقیق است (Sellwood et al., 2018). چالش بزرگ دیگر، مربوط به داده‌های ساختاری پروتئین‌ها است. اگرچه پیشرفت‌هایی مانند مدل‌های یادگیری عمیق برای پیش‌بینی ساختار پروتئین‌ها ارائه شده‌اند (Patel & Shah, 2022)، اما هنوز برای بسیاری از اهداف

دارویی جدید، ساختار سه‌بعدی دقیق با حتی یک مدل ساختاری قابل اعتماد در دسترس نیست. در این شرایط، مدل‌های هوش مصنوعی باید بر روی داده‌های مبتنی بر توالی یا مدل‌های ساختاری کم‌دقت آموزش ببینند که منجر به کاهش اطمینان در پیش‌بینی‌های اتصال و طراحی دارو می‌شود. این امر اهمیت توسعه روش‌هایی را دوچندان می‌کند که بتوانند عدم قطعیت (Uncertainty) در پیش‌بینی‌های خود را به وضوح گزارش دهند.

همچنین، داده‌های بالینی و ایمنی (ADMET) که برای موفقیت یک دارو در مراحل نهایی حیاتی هستند، معمولاً خصوصی و بسیار محدود هستند، زیرا شرکت‌ها این اطلاعات را محرمانه نگه می‌دارند. این کمبود داده‌های ایمنی و سمیت در مراحل پایانی، باعث می‌شود مدل‌های هوش مصنوعی که آموزش خود را عمدتاً بر اساس داده‌های شیمیایی اولیه دریافت کرده‌اند، در پیش‌بینی شکست‌های ناگهانی در مراحل پیش‌بالینی یا بالینی دقت کافی نداشته باشند. برای غلبه بر این، نیاز به ایجاد کنسرسیوم‌هایی برای اشتراک‌گذاری داده‌های غیررقابتی و توسعه روش‌هایی برای یادگیری انتقالی (Transfer Learning) از حوزه‌های مشابه، ضروری است (Deng et al., 2022). در نهایت، داده‌های مورد نیاز برای مدل‌های مولد که قصد دارند ترکیبات جدیدی طراحی کنند، باید طیف وسیعی از فضای شیمیایی ممکن را پوشش دهند. اگر داده‌های آموزشی فقط ترکیبات با ویژگی‌های خاصی را در بر گیرند، مدل مولد تنها قادر به تولید مشتقاتی از آن مجموعه خواهد بود و پتانسیل واقعی نوآوری خود را از دست می‌دهد. مهندسی داده‌ها در این حوزه، فراتر از صرف جمع‌آوری صرف است و شامل تکنیک‌هایی مانند تعادل‌سازی داده‌ها، پر کردن نقاط خالی (Imputation) و اعتبارسنجی متقابل (Cross-Validation) بر روی زیرمجموعه‌های معتبر داده‌ها است تا مدل‌ها بتوانند به شیوه‌ای پایدار و قابل اعتماد عمل کنند (Jiménez-Luna et al., 2021).

۴. روش‌های هوش مصنوعی در کشف دارو

هوش مصنوعی مجموعه‌ای از ابزارها و الگوریتم‌ها را فراهم می‌آورد که فرآیند کشف دارو را در سه مرحله اصلی — شناسایی اهداف، کشف ترکیبات پیشرو، و بهینه‌سازی — به‌طور بنیادین تغییر داده است. این روش‌ها عموماً بر اساس رویکردهای یادگیری نظارت‌شده، بدون نظارت، تقویتی، و اخیراً مدل‌های مولد دسته‌بندی می‌شوند که هر کدام کاربرد خاص خود را در سلسله مراتب DDD دارند.

۴.۱. یادگیری نظارت‌شده (Supervised Learning)

یادگیری نظارت‌شده هسته اصلی بسیاری از کاربردهای اولیه هوش مصنوعی در شیمی دارویی بوده است و بر پیش‌بینی یک خروجی مشخص (مانند فعالیت یا سمیت) بر اساس ورودی‌های برچسب‌گذاری شده (مانند ساختار شیمیایی و نتایج آزمایش) متمرکز است. در این رویکرد، مدل‌ها با استفاده از مجموعه‌های داده‌ای که فعالیت زیستی آن‌ها به‌طور تجربی تأیید شده است، آموزش می‌بینند تا روابط میان ویژگی‌های مولکولی و پاسخ بیولوژیکی را مدل‌سازی کنند. شبکه‌های عصبی عمیق، به‌ویژه GNNها، در این حوزه بسیار موفق بوده‌اند، زیرا می‌توانند توپولوژی مولکولی را مستقیماً به عنوان

گراف‌های شیمیایی پردازش کنند، به جای تکیه بر توصیف‌گرهای محاسباتی سنتی. این مدل‌ها در پیش‌بینی میل ترکیبی دارو به پروتئین‌های هدف، پیش‌بینی پاسخ بیماران به داروهای خاص (Pharmacogenomics)، و ارزیابی سریع پارامترهای ADMET (جذب، توزیع، متابولیسم، دفع و سمیت) کاربرد فراوان دارند (Chen et al., 2023). موفقیت این مدل‌ها در کاهش تعداد ترکیبات نیازمند سنتز و آزمایش در مراحل اولیه، صرفه‌جویی قابل توجهی در زمان و منابع ایجاد کرده است.

۴.۲. یادگیری بدون نظارت (Unsupervised Learning)

یادگیری بدون نظارت، زمانی اهمیت می‌یابد که داده‌های برچسب‌گذاری شده کمیاب یا پرهزینه باشند. هدف اصلی در این بخش، کشف ساختارها و الگوهای ذاتی پنهان در مجموعه داده‌های بزرگ و بدون برچسب است. تکنیک‌هایی مانند خوشه‌بندی (Clustering) و کاهش ابعاد (Dimensionality Reduction) برای سازماندهی فضای شیمیایی و بیولوژیکی به کار می‌روند. در شیمی دارویی، این روش‌ها برای دسته‌بندی ترکیبات دارویی به گروه‌های ساختاری مشابه (برای شناسایی مکانیسم‌های عمل مشترک) یا خوشه‌بندی پروفایل‌های بیان ژن بیماران (برای شناسایی زیرگروه‌های بیماری) استفاده می‌شوند. علاوه بر این، روش‌های بدون نظارت مانند رمزگذارهای خودکار (Autoencoders) به عنوان ابزاری برای یادگیری یک نمایش فشرده و معنادار (Latent Representation) از ساختارهای مولکولی استفاده می‌شوند؛ این نمایش‌های فشرده می‌توانند به عنوان ورودی کارآمدتری برای مدل‌های نظارت‌شده بعدی یا به عنوان پایه‌ای برای مدل‌های مولد عمل کنند (Jiménez-Luna et al., 2021).

۴.۳. یادگیری تقویتی (Reinforcement Learning - RL)

یادگیری تقویتی یک پارادایم جذاب برای طراحی دارو است زیرا فرآیند کشف را شبیه به یک بازی یا فرآیند تصمیم‌گیری متوالی مدل‌سازی می‌کند، جایی که یک عامل (Agent) در یک محیط (فضای شیمیایی) اقداماتی (مانند تغییر یک گروه عاملی) انجام می‌دهد و برای رسیدن به یک هدف مشخص (مانند حداکثر کردن میل ترکیبی)، پاداش دریافت می‌کند. RL به‌ویژه در بهینه‌سازی مولکول‌ها در فضای شیمیایی با محدودیت‌های چندگانه بسیار مؤثر است (Zhu, 2020). به جای صرفاً پیش‌بینی یک ویژگی، عامل RL یاد می‌گیرد که دنباله‌ای از تغییرات مولکولی را ایجاد کند که منجر به تولید یک ترکیب نهایی با مجموعه‌ای از خواص بهینه (مثلاً پایداری بالا و اندازه مناسب) شود. این رویکرد نیازمند تعریف دقیق تابع پاداش است که معمولاً ترکیبی از امتیازات پیش‌بینی‌شده توسط مدل‌های نظارت‌شده (مانند کارایی و سمیت) است.

۴.۴. مدل‌های مولد (Generative Models)

مدل‌های مولد، که پیشرفته‌ترین بخش کاربرد هوش مصنوعی در DDD محسوب می‌شوند، به‌طور فعال به دنبال خلق مولکول‌های جدید هستند، نه فقط ارزیابی مولکول‌های موجود. مدل‌هایی مانند VAEs, GANs و به ویژه مدل‌های مبتنی بر ترانسفورمر (که اغلب برای توالی‌های SMILES یا گراف‌ها استفاده می‌شوند) می‌توانند بر اساس داده‌های آموزشی، توزیع شیمیایی را فرا گرفته و مولکول‌هایی با ویژگی‌های از پیش تعیین‌شده تولید کنند (Zhu, 2020). این مدل‌ها به شیمی‌دانان اجازه می‌دهند تا به جای غربالگری یک کتابخانه عظیم، مستقیماً به سمت طراحی ترکیبات "مطلوب" حرکت کنند. مدل‌های Diffusion، که اخیراً در گرافیک محبوب شده‌اند، نیز در حال تطبیق برای تولید ساختارهای مولکولی سه‌بعدی با حفظ ویژگی‌های فیزیکی شیمیایی مورد نظر هستند، و پتانسیل عظیمی برای غلبه بر محدودیت‌های روش‌های قدیمی‌تر در طراحی ده نو دارند.

۴.۵. شیمی انفورماتیک پیشرفته و یادگیری عمیق

این بخش از هوش مصنوعی بر توسعه نحوه نمایش مولکول‌ها و تعاملات آن‌ها برای ورود به شبکه‌های عصبی تمرکز دارد. شبکه‌های عصبی گرافی (GNNs) در این زمینه پیشگام هستند، زیرا نمایش مولکول به صورت یک گراف (اتم‌ها به عنوان گره‌ها و پیوندها به عنوان یال‌ها) اجازه می‌دهد تا اطلاعات هندسی و توپولوژیکی به‌طور مستقیم در محاسبات لحاظ شوند (Jiménez-Luna et al., 2021). همچنین، استفاده از نمایش‌های مبتنی بر بردار فضایی (Embeddings) برای توالی‌های پروتئینی و مولکولی، الهام گرفته از مدل‌های پردازش زبان طبیعی (NLP)، امکان یادگیری روابط عمیق‌تر میان اجزای بیولوژیکی را فراهم کرده است. در این زمینه، رویکردهای ترکیبی که دانش شیمیایی را با یادگیری عمیق ادغام می‌کنند، نتایج قوی‌تری ارائه می‌دهند، به‌ویژه زمانی که از مدل‌های فیزیکی-محاسباتی برای تولید ویژگی‌های غنی‌تر استفاده شود (Patel & Shah, 2022).

۵. پیش‌بینی ساختار و عملکرد پروتئین‌ها

پروتئین‌ها هدف اصلی بیش از ۹۰ درصد داروهای موجود هستند، و درک دقیق ساختار سه‌بعدی و دینامیک عملکردی آن‌ها برای طراحی داروهای مؤثر حیاتی است. روش‌های سنتی تعیین ساختار، مانند بلورنگاری اشعه ایکس یا کرایو-ام (Cryo-EM)، بسیار پرهزینه، زمان‌بر و اغلب برای پروتئین‌های خاصی (مانند پروتئین‌های غشایی) غیرممکن هستند. هوش مصنوعی، با تکیه بر توانایی‌های یادگیری عمیق در مدل‌سازی روابط پیچیده در داده‌های بیولوژیکی، این محدودیت‌ها را به چالش کشیده است (Patel & Shah, 2022). موفقیت‌های اخیر در پیش‌بینی ساختار پروتئین، عمدتاً توسط مدل‌های مبتنی بر ترانسفورمر و شبکه‌های عصبی که توالی‌های آمینواسیدی را به عنوان ورودی می‌گیرند، رقم خورده است. این مدل‌ها با تحلیل ماتریس‌های توالی‌های هم‌تراز (Multiple Sequence Alignment - MSA) و الگوهای تکاملی، قادرند فاصله میان جفت‌های آمینواسیدی دور از هم در ساختار سه‌بعدی را با دقت فوق‌العاده‌ای تخمین بزنند. این تخمین‌های فاصله، سپس به عنوان محدودیت‌هایی برای یک فرآیند ساختاردهی استفاده می‌شوند که می‌تواند مدل سه‌بعدی نهایی را با دقتی رقابت‌پذیر با روش‌های تجربی تولید کند (Patel & Shah, 2022). این تحول به این معنی

است که برای هر پروتئینی که توالی آن شناخته شده باشد، یک مدل ساختاری با کیفیت بالا به طور معمول ظرف چند دقیقه در دسترس است، که این امر چرخه کشف دارو را به طور چشمگیری کوتاه می کند.

علاوه بر ساختار ثابت، درک عملکرد پروتئین ها و تغییر شکل آن ها در حین تعامل با لیگاندها یا تغییرات محیطی، اهمیت مضاعفی دارد. مدل های یادگیری عمیق اکنون در حال گسترش برای پیش بینی دینامیک پروتئین ها و همچنین پیش بینی نقاط اتصال آلوستریک (Allosteric Sites) هستند. پیش بینی سایت های آلوستریک، که نقاطی غیر از محل فعال هستند و اتصال دارو به آن ها می تواند عملکرد پروتئین را تنظیم کند، برای طراحی داروهای نسل جدید که نیاز به کنترل دقیق تری بر مسیرهای بیولوژیکی دارند، ضروری است (Deng et al., 2022). مدل های یادگیری عمیق می توانند با تحلیل ارتعاشات ساختاری موجود در داده های دینامیک مولکولی یا مدل های ساختاری پیش بینی شده، مناطقی را که به تغییرات محلی حساس هستند، شناسایی کنند. کاربرد مستقیم این مدل های ساختاری در کشف دارو، تقویت شبیه سازی های داکینگ (Docking Simulation) است. در حالی که داکینگ سنتی اغلب بر اساس ساختارهای ثابت پیش بینی شده یا تجربی انجام می شد، استفاده از مدل های ساختاری بسیار دقیق تر، دقت امتیازدهی ترکیب (Scoring Function) را بهبود می بخشد. همچنین، مدل های هوش مصنوعی می توانند نحوه واکنش ساختار پروتئین به حضور یک لیگاند را به طور پویا مدل سازی کنند، که فرآیندی است که شبیه سازی های دینامیک مولکولی سنتی برای انجام آن به زمان های محاسباتی بسیار طولانی نیاز دارند. این قابلیت پیش بینی تعامل، به مهندسان نرم افزار اجازه می دهد تا الگوریتم هایی برای تولید ساختارهای پروتئین-لیگاند طراحی کنند که انرژی آزاد اتصال را با کارایی بالاتری تخمین بزنند.

پیش بینی محل های اتصال دارو در سطح سلولی نیز از دیگر حوزه های پیشرفته است. پروتئین ها در محیط سلولی در یک "محیط زیستی" پیچیده وجود دارند و این محیط بر شکل و عملکرد آن ها تأثیر می گذارد. مدل های هوش مصنوعی قادرند با ادغام داده های ژنومی و پروتئومیکی، پیش بینی کنند که کدام پروتئین ها در یک مسیر خاص در یک نوع سلول خاص دخیل هستند و ساختار آن ها تحت چه شرایطی تغییر می کند، که این امر به انتخاب اهداف دارویی با اختصاصیت بافتی بالاتر کمک می کند (Yang et al., 2019). این سطح از تجزیه و تحلیل نیازمند تلفیق داده های چندگانه مقیاس از داده های ساختاری گرفته تا داده های عملکردی سلولی است، که صرفاً با توانایی های محاسباتی پیشرفته قابل دستیابی است. در نهایت، استفاده از مدل های مولد برای طراحی لیگاندها به طور همزمان با ساختار پروتئینی هدف، یک رویکرد نوین است. به جای طراحی لیگاند برای یک ساختار ثابت، مدل های یادگیری عمیق می توانند همزمان ساختار لیگاند و ساختار پروتئین هدف را طوری بهینه کنند که بهترین اتصال ممکن را ایجاد کنند، که این امر به طراحی داروهایی با میل ترکیبی بهینه کمک می کند. این سطح از همکاری بین طراحی لیگاند و طراحی پروتئین، که توسط هوش مصنوعی تسهیل می شود، چشم انداز کاملاً جدیدی را برای مقابله با اهداف دارویی چالش برانگیز، مانند برهم کنش های پروتئین-پروتئین، ترسیم می کند (Parvatikar et al., 2023).

۶. طراحی مولکول و بهینه سازی چندهدفه

طراحی مولکول های دارویی مؤثر نه تنها نیازمند دستیابی به میل ترکیبی بالا به هدف است، بلکه باید مجموعه ای از خواص فیزیکوشیمیایی مانند جذب، حلالیت، پایداری متابولیک و سمیت را نیز به طور همزمان برآورده سازد. این مسئله

یک مسئله بهینه‌سازی چندهدفه (Multi-Objective Optimization) در فضای شیمیایی وسیع است که در آن بهبود یک ویژگی اغلب منجر به تضعیف ویژگی‌های دیگر می‌شود (مانند افزایش میل ترکیبی ممکن است حلالیت را کاهش دهد). هوش مصنوعی، به‌ویژه با استفاده از مدل‌های مولد و یادگیری تقویتی، ابزارهای قدرتمندی برای پیمایش این فضای پیچیده ارائه داده است.

مدل‌های مولد مانند VAEs و GANs نقش کلیدی در طراحی ده نوو ایفا می‌کنند. در این رویکرد، مدل‌ها بر روی توزیع مولکول‌های فعال شناخته‌شده آموزش می‌بینند و سپس می‌توانند با هدایت فضای نهفته (Latent Space)، مولکول‌هایی با خواص ترکیبی مورد نظر تولید کنند. برای مثال، با ترکیب دو نقطه در فضای نهفته که نشان‌دهنده مولکول‌های با فعالیت و حلالیت متفاوت هستند، می‌توان مولکول‌های جدیدی در میانه این دو نقطه تولید کرد که تعادلی مطلوب از هر دو ویژگی را ارائه دهند (Zhu, 2020). این قابلیت نیازمند نگاشت دقیق فضای نهفته به خواص شیمیایی قابل اندازه‌گیری است، که اغلب از طریق شبکه‌های عصبی پیش‌بینی‌کننده که به‌عنوان توابع ارزیابی عمل می‌کنند، انجام می‌شود. یادگیری تقویتی (RL) روشی قدرتمند برای دستیابی به این بهینه‌سازی چندهدفه است. در چارچوب RL، عامل به‌طور متوالی اتم‌ها یا گروه‌های عاملی را به یک ساختار مولکولی موجود اضافه یا اصلاح می‌کند. تابع پاداش در این حالت به‌طور مصنوعی طراحی می‌شود تا ترکیبی از امتیازات پیش‌بینی‌شده توسط مدل‌های نظارت‌شده برای اهداف مختلف باشد. برای مثال، پاداش می‌تواند تابعی از میل ترکیبی (ضریب مثبت بالا) و سمیت (ضریب منفی بالا) باشد. عامل RL یاد می‌گیرد که دنباله‌ای از تغییرات را انتخاب کند که ماکزیمم پاداش تجمعی را به دست آورد، که معادل بهینه‌سازی همزمان همه اهداف است (Zhu, 2020). این روش به‌ویژه برای ساخت داروهای پیچیده با ساختارهای خاص (مانند پپتیدها یا ماکروسیکل‌ها) مفید است، جایی که قوانین شیمی سنتی ممکن است پیچیده باشند.

رویکردهای مبتنی بر شیمی‌انفورماتیک پیشرفته نیز در این زمینه حیاتی هستند. استفاده از شبکه‌های عصبی گرافی (GNNs) برای مدل‌سازی ساختارهای مولکولی به مدل‌ها اجازه می‌دهد تا تأثیر تغییرات کوچک در اتصالات اتمی را بر خواص کلی مولکول با دقت بیشتری ارزیابی کنند (Jiménez-Luna et al., 2021). این دقت در ارزیابی محلی، به‌ویژه هنگام بهینه‌سازی ویژگی‌هایی که به ساختار دقیق وابسته هستند (مانند برهم‌کنش‌های فضایی در محل اتصال)، بسیار ارزشمند است. مهندسان نرم‌افزار در حال توسعه روش‌هایی برای ادغام مستقیم اطلاعات هندسی سه‌بعدی از مدل‌های پیش‌بینی ساختار پروتئین (بخش قبلی) با مدل‌های طراحی مولکول هستند تا اطمینان حاصل شود که مولکول‌های طراحی‌شده نه تنها از نظر شیمیایی امکان‌پذیرند، بلکه به‌طور فضایی نیز با هدف سازگارند.

بهینه‌سازی پارامترهای سینتیک و فارماکوکینتیک (PK) یکی دیگر از ابعاد کلیدی است که هوش مصنوعی آن را متحول کرده است. به‌جای آزمایش کورکورانه در مراحل پیش‌بالینی برای تعیین جذب و متابولیسم، مدل‌های ML می‌توانند این خواص را بر اساس ساختار پیش‌بینی کنند و بهینه‌سازی را به سمت ترکیباتی هدایت کنند که احتمال موفقیت بالاتری در جذب از روده و رسیدن به بافت هدف دارند (Patel & Shah, 2022). این کاهش وابستگی به آزمایش‌های در ویوو اولیه، ریسک‌های مالی و اخلاقی مرتبط با تست داروهای ناکارآمد بر روی مدل‌های حیوانی را به شدت کاهش می‌دهد. چالش باقی‌مانده در این حوزه، اعتبارسنجی است. ترکیبات طراحی‌شده توسط هوش مصنوعی اغلب نوآورانه هستند و ممکن است از مرزهای فضای شیمیایی شناخته‌شده فراتر روند. بنابراین، سنتز و آزمایش فیزیکی این ترکیبات برای تأیید پیش‌بینی‌های مدل ضروری است، که مستلزم یک لوله‌عملیاتی قوی و خودکار برای سنتز و غربالگری است (Chan et al., 2023).

(al., 2019). موفقیت نهایی در طراحی مولکول، مستلزم یکپارچگی بدون درز میان ابزارهای محاسباتی پیشرفته و توانایی‌های سنتز شیمیایی با توان عملیاتی بالا است که توسط مهندسان نرم‌افزار پشتیبانی می‌شود.

۷. اکتشاف هدف و اعتبارسنجی

یکی از پرخطرترین مراحل در کشف دارو، انتخاب هدف مولکولی مناسب است؛ هدفی که نقش محوری در پاتوفیزیولوژی بیماری داشته باشد و پتانسیل درمانی قابل قبولی را ارائه دهد. روش‌های سنتی اغلب بر روی اهداف "داروشناسی‌شده" (Druggable) و شناخته‌شده تمرکز داشتند، اما هوش مصنوعی در حال باز کردن مسیرهایی برای کشف اهداف جدید و کم‌تر شناخته‌شده است که ممکن است در بیماری‌های پیچیده‌ای مانند سرطان‌های مقاوم یا بیماری‌های نورودژنراتیو کلید حل مشکل باشند (Yang et al., 2019). روش‌های یادگیری عمیق، به‌ویژه زمانی که داده‌های چندگانه اومیکس (Omics) از بیماران مختلف با موفقیت ترکیب شوند، می‌توانند ارتباطات پنهانی میان ژن‌ها، مسیرهای سیگنالینگ و پیامدهای بیماری را آشکار سازند. مدل‌های هوش مصنوعی می‌توانند با تحلیل شبکه‌های تنظیم ژنی و پروتئینی، مسیری را شناسایی کنند که در زیرگروه‌های خاصی از بیماران بیش از حد فعال است و در عین حال، این مسیر از نظر ساختاری برای مهار توسط یک مولکول کوچک مناسب باشد. این فرآیند، به‌طور موثری، غربالگری اهداف را از یک جستجوی مبتنی بر حدس و گمان به یک فرآیند مبتنی بر شواهد بیولوژیکی جمعی تبدیل می‌کند (Deng et al., 2022).

اعتبارسنجی اهداف (Target Validation) مرحله‌ای است که برای اطمینان از اینکه مهار یا فعال‌سازی هدف انتخابی، تأثیر درمانی مطلوب و قابل قبولی در مدل‌های در ویوو و بالینی خواهد داشت، انجام می‌شود. هوش مصنوعی در این مرحله نیز نقش دارد. با استفاده از یادگیری ماشین بر روی داده‌های مرتبط با سمیت و اثرات جانبی مربوط به اهداف مشابه (Target Family)، مدل‌ها می‌توانند احتمال موفقیت یک هدف جدید را پیش‌بینی کنند. این شامل تحلیل همبستگی بین مکانیسم عمل داروهای موجود که بر روی اهداف هم‌خانواده عمل می‌کنند و پروفایل‌های عوارض جانبی آن‌ها است (Gupta et al., 2021). علاوه بر این، مدل‌های یادگیری ماشینی می‌توانند برای پیش‌بینی اینکه آیا مهار هدف منجر به فعال‌سازی مکانیسم‌های جبرانی (Compensatory Mechanisms) در سلول خواهد شد یا خیر، استفاده شوند. بسیاری از اهداف دارویی در ابتدا امیدوارکننده به نظر می‌رسند، اما بدن با فعال کردن مسیرهای جایگزین، به سرعت مقاومت ایجاد می‌کند. مدل‌های پیش‌بین هوش مصنوعی می‌توانند با تحلیل داده‌های دینامیکی سیستم بیولوژیکی، این مسیرهای جبرانی را پیش‌بینی کرده و طراحان دارو را وادار کنند تا استراتژی‌های ترکیبی دارو درمانی (Combination Therapy) را از ابتدا در نظر بگیرند (Yang et al., 2019).

توسعه مدل‌های مبتنی بر داده برای پیش‌بینی ارتباط هدف-بیماری، نیازمند زیرساخت‌های داده‌ای بسیار قوی است که بتوانند داده‌های ژنومی، بالینی، و ساختاری را به‌طور همزمان در یک چارچوب تحلیلی واحد ادغام کنند. این کار مستلزم استفاده از تکنیک‌های پیچیده یکپارچه‌سازی داده‌ها در مهندسی نرم‌افزار است تا اطمینان حاصل شود که الگوهای استخراج شده از یک نوع داده، به‌طور معناداری با سایر انواع داده‌ها مرتبط می‌شوند (Chan et al., 2019). در نهایت، مدل‌سازی اثرات دارویی در بافت‌های مختلف (Tissue Specificity) یک نیاز حیاتی برای جلوگیری از عوارض جانبی سیستمیک است. هوش مصنوعی می‌تواند با تحلیل داده‌های بیان ژن در بافت‌های مختلف بدن، پیش‌بینی کند که آیا

یک دارو که بر یک هدف در کبد اثر می‌گذارد، آیا ممکن است اثرات ناخواسته‌ای بر روی قلب یا کلیه داشته باشد، حتی اگر هدف اصلی در آن بافت‌ها فعال نباشد، بلکه به واسطه برهم‌کنش‌های مسیر سیگنالینگ فرعی دخیل باشد. این توانایی پیش‌بینی اختصاصیت بافتی، اعتبار هدف را به‌طور چشمگیری افزایش می‌دهد و ریسک شکست در مراحل بالاتر توسعه را کاهش می‌دهد.

۸. کشف مجدد دارو با هوش مصنوعی

کشف مجدد دارو (Drug Repurposing یا Repositioning) به معنای یافتن کاربردهای جدید برای داروهایی است که قبلاً تأیید شده‌اند یا در مراحل توسعه اولیه شکست خورده‌اند، اما برای اهداف درمانی دیگر پتانسیل دارند. این رویکرد به‌طور ذاتی ریسک کمتری دارد و زمان توسعه کوتاه‌تری را طلب می‌کند، زیرا داده‌های ایمنی و سمیت آن‌ها برای استفاده در انسان‌ها از پیش موجود است. هوش مصنوعی به دلیل توانایی بی‌نظیرش در یافتن ارتباطات غیرمستقیم و شبکه‌ای بین داروها و بیماری‌ها، به یک ابزار محوری برای این حوزه تبدیل شده است (Sellwood et al., 2018). یکی از رایج‌ترین روش‌های مورد استفاده در کشف مجدد، استفاده از مدل‌های مبتنی بر شبکه‌های دانش (Knowledge Graphs) است که ارتباطات بین داروها، اهداف مولکولی، بیماری‌ها، مسیرهای بیوشیمیایی و اثرات جانبی را به صورت گره‌ها و یال‌ها مدل‌سازی می‌کنند. الگوریتم‌های یادگیری عمیق، به‌ویژه GNNs، می‌توانند این گراف‌های عظیم را پیمایش کرده و مسیرهای کوتاهی را که نشان‌دهنده یک ارتباط بیولوژیکی محتمل بین یک داروی موجود و یک بیماری جدید است، شناسایی کنند (Jiménez-Luna et al., 2021). برای مثال، اگر داروی A بر پروتئین P1 اثر بگذارد و مهار P1 شناخته شده است که بر مسیر B تأثیر می‌گذارد، و اگر بیماری جدید X با اختلال در مسیر B مرتبط باشد، هوش مصنوعی این ارتباط غیرمستقیم را استنتاج می‌کند.

رویکرد دیگر، مبتنی بر شباهت پروفایل است. این روش بر این فرض استوار است که داروهایی که پروفایل‌های بیان ژن مشابهی را در سلول‌های بیمار ایجاد می‌کنند، احتمالاً بر روی یک مکانیسم بیولوژیکی واحد تأثیر می‌گذارند، صرف نظر از اینکه هدف اولیه آن‌ها چه بوده است (Sellwood et al., 2018). با استفاده از مدل‌های یادگیری عمیق برای استخراج ویژگی‌های معنایی (Semantic Features) از داده‌های ترانسکریپتومیکس ناشی از درمان با یک دارو، می‌توان یک "اثر انگشت" (Signature) بیولوژیکی ایجاد کرد. سپس، این اثر انگشت با اثر انگشت بیماری‌های مختلف مقایسه می‌شود تا داروهایی که می‌توانند اثر انگشت بیماری را معکوس کنند، شناسایی شوند. مدل‌های مولد نیز می‌توانند در این زمینه نقش ایفا کنند، اگرچه کاربرد مستقیم آن‌ها کمتر رایج است. مدل‌های مولد می‌توانند برای بازطراحی ساختارهای شیمیایی داروهای قدیمی به منظور بهبود خواص ADMET آن‌ها (بدون تغییر هدف اصلی) استفاده شوند، در حالی که از پتانسیل ایمنی آن‌ها بهره برده می‌شود. این کار به "بهینه‌سازی مجدد" دارو کمک می‌کند و باعث می‌شود داروهایی که در گذشته به دلیل سمیت یا فارماکوکینتیک ضعیف کنار گذاشته شده بودند، دوباره قابل استفاده شوند (Zhu, 2020).

چالش اصلی در کشف مجدد، اعتبارسنجی سریع و کارآمد فرضیه‌های تولیدشده توسط هوش مصنوعی است. از آنجا که زمان در این فرآیند حیاتی است، ادغام سیستم‌های هوش مصنوعی با آزمایشگاه‌های کوچک و سریع برای اجرای غربالگری‌های اولیه بر روی داروها و بیماری‌های هدف، اهمیت مضاعفی پیدا می‌کند. توسعه لوله‌عملیاتی خودکار که

بتواند فرضیه‌های هوش مصنوعی را به سرعت به مدل‌های سلولی و حیوانی کوچک‌تر منتقل کند، کلید موفقیت این استراتژی است (Fleming, 2018).

۹. لوله‌عملیاتی داده‌محور (Data-Driven Pipeline)

انتقال کشف دارو از یک فرآیند مبتنی بر آزمایش‌های فیزیکی پراکنده به یک فرآیند یکپارچه، خودکار و داده‌محور، مستلزم توسعه لوله‌عملیاتی (Pipelines) پیچیده مهندسی نرم‌افزار است که هوش مصنوعی را در مرکز خود قرار می‌دهد. یک لوله‌عملیاتی مؤثر کشف دارو مبتنی بر هوش مصنوعی باید توانایی مدیریت داده‌های حجیم، اجرای مدل‌های یادگیری عمیق در مقیاس بزرگ، و بازخورد مداوم نتایج آزمایشگاهی به سیستم را داشته باشد (Parvatikar et al., 2023). این لوله‌عملیاتی از مراحل متعددی تشکیل شده است که همگی به‌طور خودکار یا نیمه‌خودکار اجرا می‌شوند. مرحله اول شامل جمع‌آوری، پاک‌سازی و استانداردسازی داده‌ها از منابع مختلف بیولوژیکی و شیمیایی است؛ این مرحله که مدیریت داده‌های بزرگ را طلب می‌کند، زیربنای همه تحلیل‌های بعدی است. مرحله دوم، اجرای مدل‌های پیش‌بین هوش مصنوعی برای غربالگری مجازی و رتبه‌بندی ترکیبات بالقوه است. مدل‌هایی مانند GNNها برای پیش‌بینی میل ترکیبی و مدل‌های پیش‌بین ADMET در اینجا به کار گرفته می‌شوند (Chen et al., 2023).

مرحله کلیدی در این لوله‌عملیاتی، حلقه‌های بازخورد (Feedback Loops) هستند. داده‌های حاصل از هر دور آزمایشگاهی (مثلاً نتایج غربالگری ترکیبات پیش‌بینی‌شده)، باید به سرعت به سیستم بازگردانده شوند تا مدل‌های یادگیری ماشینی بتوانند خود را تنظیم و کالیبره کنند. این تنظیم مستمر، که اساس یادگیری تقویتی و یادگیری فعال (Active Learning) است، تضمین می‌کند که مدل در هر تکرار به سمت بهینه‌سازی ویژگی‌های مطلوب حرکت کند و از نقاط کور فضای شیمیایی دوری گزیند (Gupta et al., 2021). برای طراحی مولکول‌های جدید، لوله‌عملیاتی باید شامل مازول‌های مدل‌سازی مولد باشد که ترکیباتی را تولید کرده و سپس از طریق یک مازول ارزیابی سریع (که از مدل‌های پیش‌بین استفاده می‌کند)، رتبه‌بندی می‌شوند. ترکیبات با بالاترین امتیازدهی برای سنتز فیزیکی انتخاب می‌شوند. در این مرحله، یک مازول سنتز خودکار یا رباتیک می‌تواند به کار گرفته شود تا فرآیند آزمایش را تسریع بخشد (Parvatikar et al., 2023).

از منظر مهندسی نرم‌افزار، ساخت چنین لوله‌عملیاتی نیازمند زیرساخت‌های ابری مقیاس‌پذیر (Scalable Cloud Infrastructure) برای اجرای مدل‌های یادگیری عمیق است که نیازمند GPU هستند. مدیریت مدل‌ها، نسخه‌بندی داده‌ها، و اطمینان از تکرارپذیری نتایج (Reproducibility) نیازمند استفاده از چارچوب‌هایی مانند MLflow یا Kubeflow است تا تغییرات در مدل‌ها و داده‌ها به‌طور سیستمی ردیابی شوند (Chan et al., 2019). یک لوله‌عملیاتی موفق، علم شیمی، زیست‌شناسی و علوم کامپیوتر را در یک جریان کار واحد ادغام می‌کند، که این امر به‌طور مستقیم به کاهش زمان و هزینه کلی توسعه دارو منجر می‌شود.

۱۰. مطالعات موردی و ارزیابی و اعتبارسنجی

اثر بخشی هوش مصنوعی در کشف دارو تنها در تئوری نیست، بلکه در چندین مطالعه موردی برجسته در سال‌های اخیر به اثبات رسیده است. یکی از مشهورترین مثال‌ها، استفاده از مدل‌های یادگیری عمیق برای کشف ترکیبات با فعالیت ضدباکتریایی در برابر سویه‌های مقاوم به دارو است. در یک مطالعه موردی خاص، محققان با استفاده از یک مدل یادگیری عمیق آموزش‌دیده بر روی داده‌های فعالیت زیستی گسترده، توانستند صدها هزار ترکیب را در عرض چند روز غربال کنند و کاندیداهایی را شناسایی کنند که در برابر برخی از سخت‌ترین سویه‌های باکتریایی (مانند *Acinetobacter baumannii*) فعال بودند (Deng et al., 2022). این کاندیداها پس از سنتز و آزمایش در مراحل بعدی، کارایی بالایی نشان دادند و نشان دادند که هوش مصنوعی می‌تواند به‌طور مؤثر از محدودیت‌های کتابخانه‌های شیمیایی موجود فراتر رود.

مورد دیگر، موفقیت‌های چشمگیر در پیش‌بینی ساختار پروتئین است. معرفی سیستم‌هایی مانند AlphaFold، اگرچه عمدتاً یک دستاورد در بیوانفورماتیک است، تأثیر مستقیمی بر داکینگ دارویی مبتنی بر هوش مصنوعی گذاشته است (Patel & Shah, 2022). در مطالعات مرتبط، تیم‌هایی از شرکت‌های دارویی از این مدل‌های ساختاری پیش‌بینی‌شده برای طراحی لیگاندها استفاده کردند و با موفقیت کاندیداهایی را شناسایی کردند که پیش‌بینی می‌شد به بهترین شکل با ساختار دقیق پروتئین تعامل داشته باشند، که این امر در داکینگ‌های سنتی بر اساس مدل‌های ساختاری قدیمی‌تر، دستیابی به آن دشوار بود. اعتبارسنجی مدل‌ها در این حوزه حیاتی است و فراتر از اندازه‌گیری سادگی معیارهایی مانند دقت (Accuracy) یا AUC (Area Under the Curve) است. در کشف دارو، اعتبارسنجی باید شامل توانایی مدل در تعمیم‌پذیری (Generalization) به مولکول‌هایی باشد که از نظر ساختاری یا فعالیت زیستی از داده‌های آموزشی فاصله دارند؛ یعنی توانایی مدل در اکتشاف (Exploration) در فضای شیمیایی (Zhu, 2020). روش‌های اعتبارسنجی استاندارد شامل استفاده از مجموعه‌های آزمایشی مستقل و همچنین اعتبارسنجی خارجی (External Validation) بر روی داده‌های جمع‌آوری‌شده توسط آزمایشگاه‌های کاملاً مجزا است.

یکی از مهم‌ترین چالش‌ها در اعتبارسنجی، تفکیک بین "یادگیری مکانیکی" (Memorization) و "یادگیری واقعی" (True Learning) است. مدل‌هایی که فقط داده‌های آموزشی را حفظ کرده‌اند، در مواجهه با ترکیبات جدید شکست می‌خورند. برای مقابله با این امر، تکنیک‌هایی مانند یادگیری فعال (Active Learning) مورد استفاده قرار می‌گیرند؛ در این روش، مدل به‌طور هوشمندانه ترکیبات بعدی را که باید آزمایش شوند، انتخاب می‌کند تا بیشترین اطلاعات جدید را کسب کند و دانش خود را افزایش دهد، به‌جای اینکه صرفاً بر روی داده‌های پرتکرار تمرکز کند (Gupta et al., 2021). اعتبارسنجی باید همچنین شامل اندازه‌گیری میزان نوآوری باشد. ترکیبات تولیدشده توسط مدل‌های مولد (Generative Models) باید از نظر شیمیایی جدید بوده و قابل سنتز باشند. بنابراین، ارزیابی‌های اضافی مانند امتیازهای سنتزپذیری (Synthesizability Scores) که اغلب با استفاده از مدل‌های ML/RL دیگر پیش‌بینی می‌شوند، باید در فرآیند اعتبارسنجی یکپارچه شوند (Parvatikar et al., 2023). در نهایت، مطالعات موردی موفق اغلب نشان می‌دهند که مدل‌های هوش مصنوعی بهترین عملکرد را زمانی دارند که به‌جای جایگزینی کامل شیمی‌دانان، به عنوان ابزاری برای هدایت آزمایش‌ها و کاهش فضای جستجو عمل کنند.

۱۱. مسائل اخلاقی و شفافیت

همان‌طور که نقش هوش مصنوعی در تصمیم‌گیری‌های حساس پزشکی و دارویی پررنگ‌تر می‌شود، مسائل اخلاقی مرتبط با شفافیت، سوگیری و مسئولیت‌پذیری بیش از پیش اهمیت می‌یابند. در زمینه کشف دارو، فقدان شفافیت در مدل‌های یادگیری عمیق (مسئله "جعبه سیاه") می‌تواند مانع پذیرش آن‌ها در محیط‌های نظارتی و بالینی شود، زیرا شیمی‌دانان و رگولاتورها نیاز دارند تا بفهمند چرا یک ترکیب خاص به‌عنوان کاندیدای برتر انتخاب شده است (Fleming, 2018).

شفافیت (Explainability یا XAI) برای اطمینان از اعتبار علمی ضروری است. اگر یک مدل پیش‌بینی کند که یک مولکول بسیار سمی خواهد بود، تیم تحقیق باید بتواند مکانیسم بیولوژیکی یا ویژگی شیمیایی خاصی را که مدل بر اساس آن این تصمیم را گرفته است، شناسایی کند. این امر به‌ویژه در مورد مدل‌های GNN و ترانسفورمرها که الگوهای پیچیده‌ای را در نمایش‌های مولکولی می‌آموزند، چالش‌برانگیز است. تلاش‌ها برای توسعه روش‌های XAI، مانند استفاده از نقشه‌های توجه (Attention Maps) در مدل‌های ترانسفورمر برای برجسته کردن بخش‌های مهم مولکول یا پروتئین، در حال افزایش است تا شیمی‌دانان بتوانند نتایج مدل را به‌طور شهودی تفسیر کنند (Chen et al., 2023).

مسئله سوگیری داده‌ای (Data Bias) یک نگرانی اخلاقی جدی دیگر است. اگر مجموعه داده‌های آموزشی عمدتاً بر روی جمعیت‌های قومی خاص یا بیماری‌هایی که مورد توجه تاریخی بوده‌اند، جمع‌آوری شده باشند، مدل‌های هوش مصنوعی ممکن است داروهایی را تولید یا اولویت‌بندی کنند که برای زیرمجموعه‌های خاصی از بیماران کارایی کمتری دارند یا حتی مضر باشند. این امر نابرابری‌های موجود در دسترسی به سلامت را تشدید می‌کند. مقابله با این سوگیری نیازمند تلاش‌های فعال در مهندسی داده برای ایجاد مجموعه‌های داده‌ای است که از نظر جمعیتی و بیولوژیکی متنوع‌تر باشند (Gupta et al., 2021). مسئولیت‌پذیری نیز یک مفهوم اخلاقی پیچیده است: اگر یک داروی طراحی‌شده توسط هوش مصنوعی در کارآزمایی‌های بالینی شکست بخورد یا منجر به عوارض جانبی پیش‌بینی‌نشده شود، مسئولیت حقوقی و اخلاقی بر عهده کیست؟ توسعه‌دهنده الگوریتم، شیمی‌دان یا شرکت دارویی؟ این موضوع نیازمند تعریف چارچوب‌های قانونی جدیدی است که نقش ابزارهای خودکار در فرآیند تصمیم‌گیری نهایی را مشخص کند. به‌طور کلی، اجماع بر این است که تصمیم‌گیری نهایی همیشه باید توسط یک متخصص انسانی صورت پذیرد، اما این انسان نیازمند درک کافی از عملکرد سیستم هوش مصنوعی است (Fleming, 2018).

در نهایت، پتانسیل سوءاستفاده از فناوری‌های مولد برای طراحی عوامل زیان‌آور (Dual-Use Potential) نیز مطرح است. مدل‌هایی که قادر به طراحی مولکول‌های فعال با ویژگی‌های مطلوب هستند، می‌توانند توسط بازیگران بدخواه برای طراحی سلاح‌های شیمیایی یا سموم استفاده شوند. این امر نیاز به ایجاد پروتکل‌های امنیتی قوی و فیلترهای اخلاقی در خود مدل‌ها برای جلوگیری از تولید ساختارهایی که پتانسیل آسیب‌زایی بالایی دارند (مانند سموم شناخته‌شده) را ضروری می‌سازد (Zhu, 2020).

۱۲. تنظیم‌گری، روندهای آینده، محدودیت‌ها و ریسک‌ها

ورود گسترده هوش مصنوعی به کشف دارو، سازمان‌های نظارتی مانند سازمان غذا و داروی آمریکا (FDA) را با چالش‌هایی جدید در زمینه تأیید و اعتبارسنجی مواجه ساخته است. سازمان‌های تنظیم‌گر نیازمند چارچوب‌هایی هستند که بتوانند اثربخشی و ایمنی مدل‌های AI را که در مراحل میانی و پایانی توسعه دارو دخیل بوده‌اند، ارزیابی کنند. تمرکز اصلی در این زمینه، بر روی اعتبارسنجی داده‌ها، پایداری الگوریتم و مستندسازی فرآیند استدلال مدل (Explainability) متمرکز است (Deng et al., 2022). روندهای آینده در این حوزه نشان‌دهنده حرکت به سمت مدل‌های ترکیبی و چندمقیاسی است. انتظار می‌رود که مدل‌های مبتنی بر هوش مصنوعی، مدل‌های شبیه‌سازی فیزیکی-شیمیایی سنتی (مانند دینامیک مولکولی) را کاملاً جایگزین نکنند، بلکه آن‌ها را تکمیل کنند. به این ترتیب، مدل‌های یادگیری ماشین برای شناسایی سریع نقاط مناسب در فضای محاسباتی به کار می‌روند و سپس شبیه‌سازی‌های فیزیکی دقیق‌تر تنها بر روی کاندیداهای بسیار منتخب انجام می‌شود تا دقت نهایی تضمین شود. این رویکرد تلفیقی، کارایی محاسباتی را به حداکثر می‌رساند (Parvatikar et al., 2023).

یکی از روندهای مهم، ادغام هوش مصنوعی با بیولوژی مصنوعی و مهندسی پروتئین است. مدل‌های مولد نه تنها مولکول‌های کوچک، بلکه توالی‌های جدید پروتئینی و آنزیمی را نیز طراحی خواهند کرد که می‌توانند به عنوان دارو یا ابزارهای تشخیصی عمل کنند. این امر، حوزه "طراحی داروهای بیولوژیک" (Biologics Design) را متحول خواهد ساخت (Yang et al., 2019). همچنین، هوش مصنوعی در حال ورود به مرحله نهایی توسعه دارو، یعنی پیش‌بینی نتایج کارآزمایی‌های بالینی و بهینه‌سازی جمعیت‌های بیماران برای ورود به کارآزمایی‌ها (Patient Stratification) است تا احتمال موفقیت در فاز سوم افزایش یابد (Jiménez-Luna et al., 2021). با این حال، محدودیت‌ها و ریسک‌ها همچنان پابرجا هستند. بزرگ‌ترین محدودیت، همان‌طور که در بخش داده‌ها ذکر شد، وابستگی شدید به داده‌های با کیفیت و کافی است. شکست مدل‌ها در مواجهه با داده‌های دنیای واقعی که از پروتکل‌های آزمایشی متفاوتی تولید شده‌اند، همچنان یک ریسک بزرگ است (Chan et al., 2019). ریسک دیگر، وابستگی بیش از حد به مدل‌های پیش‌بین است که منجر به ایجاد "بحران تکرارپذیری" می‌شود، اگر فرضیات اساسی مدل به مرور زمان یا با تغییر جمعیت‌های مورد مطالعه تغییر کند.

ریسک عملیاتی در زمینه مهندسی نرم‌افزار نیز مطرح است. توسعه و نگهداری لوله‌عملیاتی‌های یادگیری عمیق که مستلزم به‌روزرسانی‌های مکرر و کالیبراسیون مجدد مدل‌ها با توجه به داده‌های جدید هستند، نیازمند تخصص مهندسی نرم‌افزار سطح بالایی است که اغلب در محیط‌های سنتی زیست‌پزشکی کمبود آن احساس می‌شود (Parvatikar et al., 2023). این امر ضرورت آموزش متخصصانی در رشته‌های بین‌رشته‌ای مانند بیوانفورماتیک محاسباتی و مهندسی تکنولوژی نرم‌افزار را دوچندان می‌کند تا بتوانند این زیرساخت‌های پیچیده را مدیریت کنند.

۱۳. نتیجه‌گیری نهایی

انقلاب هوش مصنوعی، یک نقطه عطف غیرقابل بازگشت در حوزه کشف و توسعه دارو رقم زده است. این فناوری، با قدرت بی‌نظیر خود در تحلیل داده‌های چندوجهی و مدل‌سازی روابط پیچیده بیولوژیکی، نویدبخش کاهش چشمگیر زمان و هزینه‌های مرتبط با آوردن داروهای جدید به بازار است. از شناسایی اهداف دارویی جدید و اعتبارسنجی آن‌ها بر

اساس شواهد بیولوژیکی جمعی (Yang et al., 2019)، تا طراحی ده نوو مولکول‌های با ویژگی‌های چندگانه بهینه با استفاده از مدل‌های مولد و یادگیری تقویتی (Zhu, 2020)، هوش مصنوعی به‌طور مداوم کارایی فرآیند را در مراحل مختلف بهبود بخشیده است.

موفقیت‌های اخیر در پیش‌بینی ساختارهای پروتئینی با دقت بالا (Patel & Shah, 2022) و ادغام این مدل‌های ساختاری در فرآیندهای غربالگری مجازی و داکینگ، یک زیربنای محکم برای طراحی مبتنی بر ساختار فراهم کرده است. در عین حال، لوله‌عملیاتی‌های داده‌محور که امکان یادگیری فعال و بازخورد سریع را فراهم می‌آورند (Parvatikar et al., 2023)، تضمین می‌کنند که دانش به صورت تجمعی و کارآمد در طول چرخه‌های توسعه انباشته شود. با این حال، پذیرش کامل پتانسیل هوش مصنوعی نیازمند غلبه بر چالش‌های اساسی در حوزه داده، اخلاق و تنظیم‌گری است. کیفیت، دسترسی و استانداردسازی داده‌ها همچنان محدودکننده اصلی باقی می‌مانند، و مسائل مربوط به شفافیت مدل‌ها (XAI) و سوگیری‌های نهفته در آن‌ها، نیازمند توجه ویژه مهندسان نرم‌افزار و متخصصان اخلاق است (Chen, Fleming, 2018; et al., 2023).

در مجموع، هوش مصنوعی دیگر یک ابزار کمکی نیست، بلکه یک جزء ضروری در استراتژی نوآوری دارویی آینده است. برای رشته مهندسی تکنولوژی نرم‌افزار کامپیوتر، این حوزه فرصتی بی‌نظیر برای توسعه الگوریتم‌های پیشرفته، زیرساخت‌های محاسباتی مقیاس‌پذیر و چارچوب‌های اعتبارسنجی قوی فراهم می‌آورد که نه تنها کارایی صنعت داروسازی را افزایش می‌دهد، بلکه به طور مستقیم در ارتقاء سلامت جهانی و کشف درمان‌های سریع‌تر برای بیماری‌های لاعلاج سهم خواهد بود. مسیر پیش رو، مسیری ترکیبی است که در آن قدرت محاسباتی هوش مصنوعی با دانش عمیق بیولوژیکی و شیمیایی ترکیب می‌شود.

منابع

- Chen, W., Liu, X., Zhang, S., & Chen, S. (2023). Artificial intelligence for drug discovery: Resources, methods, and applications. *Molecular therapy Nucleic acids*, 31, 691-702.
- Jiménez-Luna, J., Grisoni, F., Weskamp, N., & Schneider, G. (2021). Artificial intelligence in drug discovery: recent advances and future perspectives. *Expert opinion on drug discovery*, 16(9), 949-959.
- Chan, H. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends in pharmacological sciences*, 40(8), 592-604.
- Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, 25(3), 1315-1360.
- Sellwood, M. A., Ahmed, M., Segler, M. H., & Brown, N. (2018). Artificial intelligence in drug discovery. *Future medicinal chemistry*, 10(17), 2025-2028.
- Zhu, H. (2020). Big data and artificial intelligence modeling for drug discovery. *Annual review of pharmacology and toxicology*, 60(1), 573-589.
- Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature*, 557(7706), S55-S55.
- Deng, J., Yang, Z., Ojima, I., Samaras, D., & Wang, F. (2022). Artificial intelligence in drug discovery: applications and techniques. *Briefings in Bioinformatics*, 23(1).

- Patel, V., & Shah, M. (2022). Artificial intelligence and machine learning in drug discovery and development. *Intelligent Medicine*, 2(3), 134-140.
- Álvarez-Machancoses, Ó., & Fernández-Martínez, J. L. (2019). Using artificial intelligence methods to speed up drug discovery. *Expert opinion on drug discovery*, 14(8), 769-777.
- Yang, X., Wang, Y., Byrne, R., Schneider, G., & Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. *Chemical reviews*, 119(18), 10520-10594.
- Parvatikar, P. P., Patil, S., Khaparkhantikar, K., Patil, S., Singh, P. K., Sahana, R., ... & Raghu, A. V. (2023). Artificial intelligence: Machine learning approach for screening large database and drug discovery. *Antiviral Research*, 220, 105740.