

طراحی مهارکننده‌های جدید $GSK-3\beta$ مبتنی بر هوش مصنوعی و مدل‌سازی مولکولی

محمد عظیمی، سید احمد عبادی*

مرکز تحقیقات گیاهان دارویی و فرآورده‌های طبیعی، دانشگاه علوم پزشکی همدان، همدان، ایران

تاریخ وصول: ۱۴۰۴/۰۳/۱۸ تاریخ پذیرش: ۱۴۰۴/۰۵/۲۶

چکیده

زمینه و هدف: با توجه به نقش کلیدی $GSK-3\beta$ در پاتوژنز بیماری پارکینسون و چالش‌های موجود در طراحی مهارکننده‌های اختصاصی، لذا هدف از این مطالعه تعیین و طراحی مهارکننده‌های جدید $GSK-3\beta$ مبتنی بر هوش مصنوعی و مدل‌سازی مولکولی بود.

روش بررسی: این یک مطالعه توصیفی - تحلیلی از نوع محاسباتی است که در سال ۱۴۰۴ انجام شد، در این پژوهش، از یک چارچوب نوآورانه مبتنی بر هوش مصنوعی جهت طراحی و ارزیابی مهارکننده‌های جدید آنزیم $GSK-3\beta$ با پتانسیل درمان بیماری پارکینسون استفاده شد. ابتدا با بهره‌گیری از مدل‌های زبانی شیمیایی (CLMS)، مجموعه‌ای از ترکیبات جدید به صورت از آغاز طراحی شد. ترکیبات جدید بر اساس معیارهای داروهمانندی، تنوع ساختاری، ویژگی‌های فیزیکوشیمیایی پالایش و ارزیابی شدند. به منظور عمومیت دادن به مدل یادگیری ماشین از داده‌افزایی ۲۰ و افزایش تنوع در طراحی ساختارهای جدید از نمونه‌برداری دمایی ۰/۸ استفاده گردید. داده‌های جمع‌آوری با استفاده از تحلیل‌های توصیفی و مقایسه‌ای مبتنی بر شاخص‌های محاسباتی تجزیه و تحلیل شدند.

یافته‌ها: مدل با نرخ ۷۳/۱ درصد در طراحی مهارکننده‌های جدید موفق عمل کرد. از این بین ۷۸ ترکیب با قدرت مهارتی کمتر از ۲۰ میکرومولار برای انجام داکینگ مولکولی در جایگاه فعال آنزیم $GSK-3\beta$ انتخاب شدند. نتایج داکینگ نشان دادند که ترکیب A با داشتن هسته ایزوکیلینولی، با انرژی پیشبینی شده ۵/۱۶ - و ترکیب B با داشتن هسته کینولینی، با انرژی پیشبینی شده ۵/۲۵ - بهترین کاندیدها از نظر انرژی اتصال بودند. همچنین ارزیابی ویژگی‌های داروهمانندی دو ترکیب منتخب، حاکی از انطباق کامل آن‌ها با قوانین فارماکوکینتیکی و معیارهای جذب خوراکی بود.

نتیجه‌گیری: رویکرد تلفیقی استفاده از هوش مصنوعی و مدل‌سازی مولکولی می‌تواند به عنوان چارچوبی کارآمد در طراحی مهارکننده‌های اختصاصی و مؤثر در حوزه درمان بیماری پارکینسون مورد استفاده قرار گیرد.

واژه‌های کلیدی: هوش مصنوعی، مدل زبانی شیمیایی، $GSK-3\beta$ ، بیماری پارکینسون، داکینگ مولکولی

* نویسنده مسئول: سید احمد عبادی، همدان، دانشگاه علوم پزشکی همدان، مرکز تحقیقات گیاهان دارویی و فرآورده‌های طبیعی

Email: a.ebadi@umsha.ac.ir

"نشریه علمی پژوهشی ارمغان دانش وابسته به دانشگاه علوم پزشکی یاسوج، یک نشریه با دسترسی آزاد است و تمامی مقالات منتشر شده در این نشریه به صورت دسترسی آزاد منتشر می‌شوند."

مقدمه

در دهه‌های اخیر، استفاده از روش‌های محاسباتی و مدل‌های هوشمند در کشف و طراحی دارو به عنوان یکی از حوزه‌های پیشرو در علوم زیستی و دارویی مورد توجه گسترده قرار گرفته است (۱ و ۲). با افزایش پیچیدگی فرآیند توسعه داروهای جدید، نیاز به رویکردهایی که بتوانند هزینه، زمان و نرخ بالای شکست در مراحل اولیه طراحی و غربالگری را کاهش دهند، بیش از پیش احساس می‌شود. در این میان، فناوری‌های مبتنی بر هوش مصنوعی (AI) و یادگیری ماشین (ML) به ویژه در زمینه تولید و ارزیابی ترکیبات شیمیایی جدید، جایگاهی کلیدی یافته‌اند. یکی از این نوآوری‌ها، استفاده از مدل‌های زبانی شیمیایی (Chemical Language Models یا CLMs) است که با بهره‌گیری از بازنمایی متنی ساختارهای مولکولی مانند SMILES، امکان تولید ترکیبات جدید با ویژگی‌های مطلوب زیستی و شیمیایی را فراهم می‌سازد (۳-۷).

مدل‌های زبانی مولد، هم‌چون شبکه‌های عصبی بازگشتی (RNNs) و ترنسفورمرها، با یادگیری از پایگاه‌های داده بزرگ حاوی اطلاعات ساختاری و بیولوژیکی ترکیبات، توانایی طراحی از آغاز ترکیبات هدفمند را پیدا کرده‌اند. این مدل‌ها می‌توانند به صورت خودکار و بر اساس ویژگی‌های تعریف شده، مولکول‌هایی با فعالیت زیستی بالا، انتخاب‌پذیری مناسب و خصوصیات فارماکوکینتیکی قابل قبول طراحی کنند. هنگامی که این مدل‌های نوین با

روش‌های کلاسیک محاسباتی مانند داکینگ مولکولی ادغام شوند، بستر توانمندی برای شناسایی، بهینه‌سازی و اعتبارسنجی مهارکننده‌های بالقوه فراهم می‌گردد. این ادغام به‌ویژه در مواردی که هدف دارویی دارای پیچیدگی ساختاری یا عملکردی است، از جمله آنزیم‌های درگیر در بیماری‌های دستگاه اعصاب مرکزی، اثربخشی بیشتری دارد (۸-۱۰).

بیماری پارکینسون یکی از شایع‌ترین اختلالات عصبی تحلیل‌برنده (neurodegenerative) است که عمدتاً در افراد میانسال و سالمند رخ می‌دهد. این بیماری به دلیل تخریب تدریجی سلول‌های عصبی در ناحیه‌ای از مغز به نام جسم سیاه (Substantia Nigra) به وجود می‌آید، این سلول‌ها مسئول تولید دوپامین هستند، که نقشی حیاتی در تنظیم حرکات بدن ایفا می‌کند. کاهش سطح دوپامین منجر به بروز علائم حرکتی از جمله لرزش در حال استراحت، کندی حرکت (برادی‌کینزیا)، سفتی عضلات و اختلال در تعادل و راه رفتن می‌شود. علاوه بر علائم حرکتی، بیماران ممکن است دچار اختلالات غیرحرکتی مانند افسردگی، اختلال خواب، کاهش حافظه و زوال شناختی شوند. علت دقیق پارکینسون هنوز به‌طور کامل شناخته نشده، اما عوامل ژنتیکی، محیطی و آسیب‌های اکسیداتیو از جمله دلایل محتمل هستند. با توجه به ماهیت پیش‌رونده بیماری، درمان‌های موجود بیشتر جنبه کنترلی دارند و بر کاهش علائم تمرکز دارند. در سال‌های اخیر، تمرکز بر

مسیرهای مولکولی بیماری، از جمله نقش آنزیم گلیکوژن سنتاز کیناز- β ($GSK-3\beta$)^(۱)، مسیرهای نوینی را برای طراحی داروهای هدفمند با کمک هوش مصنوعی و روش‌های محاسباتی فراهم کرده است (۱۱-۱۴).

یکی از این اهداف دارویی مهم، $GSK-3\beta$ است که به عنوان یک کیناز سرین / ترئونین در طیف گسترده‌ای از فرآیندهای سلولی نظیر تنظیم چرخه سلولی، آپوپتوز، رشد نورونی و انتقال سیگنال ایفای نقش می‌کند (۱۵). بررسی‌های متعددی نشان داده‌اند که افزایش فعالیت $GSK-3\beta$ با پیشرفت بیماری‌های نورودژنراتیو، به ویژه بیماری پارکینسون، ارتباط نزدیکی دارد. این بیماری مزمن عصبی که با تحلیل تدریجی نورون‌های دوپامین‌ساز در ناحیه جسم سیاه مغز همراه است، موجب بروز علائمی مانند؛ لرزش، سختی عضلات، برادی‌کینزی (کندی حرکات) و ناپایداری در راه رفتن می‌شود. مهار هدفمند $GSK-3\beta$ با هدف بازگرداندن تعادل مسیرهای سیگنال‌دهی مختل‌شده و جلوگیری از مرگ سلول‌های عصبی، به عنوان یک استراتژی درمانی نویدبخش در بیماری پارکینسون مطرح شده است (۱۷-۱۵ و ۱۲).

پروتئین $GSK-3\beta$ ساختاری نسبتاً محافظت شده و چندبخشی دارد. این پروتئین با وزن مولکولی حدود ۴۶ کیلو دالتون است که در انسان از ۴۳۳ اسید آمینه تشکیل شده و به صورت چند دومینی سازمان یافته است. دومین N-ترمینال (اسیدهای آمینه ۱ تا ۱۳۵) دارای ساختار بتا-برلی متشکل از ۷ رشته β است. دومین مرکزی با ساختار α -مارپیچی (اسیدهای

آمینه ۱۳۹ تا ۳۴۲) محل فعالیت کینازی آنزیم را دربر می‌گیرد و جایگاه اتصال ATP در ناحیه بین دومین N-ترمینال و دومین مرکزی قرار دارد. دومین C-ترمینال (اسیدهای آمینه ۳۴۳ تا ۴۳۳) از ساختار مارپیچ و حلقه تشکیل شده و با دومین مرکزی تعامل دارد، اما خارج از تاخوردگی کلاسیک کینازهای سرین / ترئونین قرار می‌گیرد (۱۹ و ۱۸، ۱۱). در بخش N-ترمینال، اسید آمینه تیروزین ($Tyr216$) وجود دارد که فسفریلاسیون آن برای فعال شدن کامل آنزیم ضروری است، در حالی که فسفریلاسیون اسید آمینه $Ser9$ در ناحیه N-ترمینال باعث مهار عملکرد آن می‌شود. ساختار $GSK-3\beta$ به گونه‌ای است که حتی در غیاب سوبسترا، در وضعیت فعال قرار دارد، که آن را از بسیاری از کینازهای دیگر متمایز می‌سازد. این ویژگی ساختاری، همراه با جایگاه‌های اتصال مشخص، $GSK-3\beta$ را به هدفی مناسب برای طراحی مهارکننده‌های انتخابی با استفاده از روش‌های مدلسازی مولکولی و طراحی دارو تبدیل کرده است (۲۰-۲۲).

با این حال، طراحی و شناسایی مهارکننده‌های انتخابی و مؤثر برای $GSK-3\beta$ همچنان چالش‌برانگیز است. این چالش‌ها شامل تمایز میان دو ایزوform $GSK-3\alpha$ و $GSK-3\beta$ ، شباهت بالای جایگاه‌های اتصال و الزامات سخت‌گیرانه در خصوص خواص زیستی و فارماکوکینتیکی ترکیبات می‌شود. بنابراین، بهره‌گیری از یک چارچوب طراحی پیشرفته و چندمرحله‌ای که بتواند مولکول‌های مؤثر را به صورت هدفمند تولید،

1-Glycogen Synthase Kinase-3beta ($GSK-3\beta$)

غربالگری و ارزیابی نماید، امری ضروری است (۲۴ و ۲۳).

در این پژوهش، رویکردی تلفیقی مبتنی بر مدل‌های زبانی شیمیایی، ابزارهای شیمی - انفورماتیک، اتصال مولکولی و شبیه‌سازی داکینگ مولکولی برای طراحی، ارزیابی و بهینه‌سازی مهارکننده‌های بالقوه GSK-3 β با هدف درمان بیماری پارکینسون ارایه شده است. ابتدا مجموعه‌ای از ترکیبات شیمیایی جدید با استفاده از مدل‌های CLM تولید شد. این ترکیبات بر اساس ویژگی‌های ساختاری و پتانسیل زیستی مورد ارزیابی اولیه قرار گرفتند و سپس تحت فرآیند غربالگری مجازی از طریق داکینگ مولکولی با جایگاه فعال GSK-3 β بررسی شدند. لذا هدف از این مطالعه تعیین و طراحی مهارکننده‌های جدید GSK-3 β مبتنی بر هوش مصنوعی و مدل‌سازی مولکولی بود.

روش بررسی

این یک مطالعه توصیفی-تحلیلی از نوع محاسباتی است که در سال ۱۴۰۴ انجام شد، داده‌های مربوط به مهارکننده‌های آنزیم GSK-3 β از پایگاه داده ChEMBL (<https://www.ebi.ac.uk/chembl>) استخراج گردید. تنها ترکیباتی انتخاب شدند که مهارکننده آنزیم GSK-3 β انسانی بودند و دارای داده‌های فعالیت زیستی به صورت کمی (نظیر IC₅₀) بودند. شناسه‌های منحصر به فرد ترکیبات و اطلاعات زیستی مرتبط با آن‌ها، از جمله رشته‌های SMILES، جمع‌آوری شد. این داده‌ها به عنوان پایه‌ای برای آموزش مدل‌های یادگیری ماشین استفاده شدند.

برای ترکیباتی که بیش از یک مقدار IC₅₀ در منابع مختلف برای آن‌ها گزارش شده بود، میانگین این مقادیر به عنوان مقدار نهایی در نظر گرفته شد و انحراف معیار آن‌ها برای ارزیابی همخوانی داده‌ها محاسبه گردید. سپس ترکیباتی که دارای مقادیر خارج از محدوده سه برابر انحراف معیار از میانگین بودند، به عنوان داده پرت در نظر گرفته شده و حذف شدند تا از ورود نویز آماری به مدل جلوگیری شود.

در مرحله‌ی بعد، رشته‌های SMILES با استفاده از ابزار (RDKit (۲۵) (این اسم برنامه هست معادل فارسی ندارد) مورد پردازش قرار گرفتند. تمامی نمک‌ها، ایزومرهای تکراری و ساختارهای طولانی‌تر از ۹۰ کاراکتر حذف شدند. پس از این مرحله، مجموعه‌ای از ترکیبات منحصر به فرد و آماده برای استفاده در مدل‌سازی باقی ماند.

به منظور تبدیل رشته‌های SMILES به فرمتی قابل پردازش به وسیله به وسیله مدل‌های یادگیری ماشین، دو روش مختلف بازنمایی مورد استفاده قرار گرفت. در روش نخست، از کدگذاری یک‌بارز (One-Hot Encoding) برای نمایش کاراکترهای SMILES استفاده شد، که در آن هر کاراکتر به صورت برداری دودویی با طول برابر تعداد کل کاراکترهای منحصر به فرد نمایش داده شد. در روش دوم، از بازنمایی تعبیه‌ای (Embedding) استفاده گردید که در آن هر کاراکتر به یک بردار عددی قابل یادگیری تبدیل شد. این روش، با یادگیری روابط بین توکن‌ها در طول آموزش، به مدل اجازه می‌دهد تا درک عمیق‌تری از ساختارهای شیمیایی به دست آورد (۲۶).

برای افزایش حجم و تنوع داده‌های آموزشی و بهبود توان تعمیم مدل، از روش افزایش داده (Augmentation) مبتنی بر بازنویسی تصادفی رشته‌های SMILES استفاده شد. به این ترتیب، برای هر ترکیب چندین نمایش غیرتکراری از همان ساختار شیمیایی با ترتیب متفاوت توکن‌ها تولید گردید. این بازنویسی‌ها در سطوح مختلف شامل: ۰، ۲، ۵، ۱۰ و ۲۰ نسخه رشته SMILES برای هر ترکیب انجام شد. این تکنیک باعث افزایش تنوع آماری در داده‌ها و کاهش احتمال بیش‌برازش شد (۲۷).

برای تولید ترکیبات جدید مشابه مهارکننده‌های $GSK-3\beta$ ، از یک مدل زبان‌محور در سطح کاراکتر (Character-level Language Model) استفاده شد. این مدل با استفاده از زبان برنامه‌نویسی Python (نسخه ۳.۱۲) و کتابخانه‌های Keras و TensorFlow پیاده‌سازی گردید. معماری مدل شامل یک لایه تعبیه‌گر برای تبدیل توکن‌ها به بردارهای عددی قابل یادگیری، دو لایه بازگشتی GRU به ترتیب با ۵۱۲ و ۲۵۶ واحد برای مدل‌سازی وابستگی‌های ترتیبی و در نهایت یک لایه چگال برای تولید خروجی بود. هم‌چنین از لایه‌های نرمال‌سازی دسته‌ای (Batch Normalization) برای پایدارسازی روند آموزش استفاده شد (۲۸).

مدل با تابع زیان آنتروپی متقاطع (Categorical Crossentropy) و بهینه‌ساز آدام (Adam) با نرخ یادگیری اولیه ۰/۰۰۱ آموزش داده شد. داده‌ها به نسبت ۸۰ درصد برای آموزش و ۲۰ درصد برای آزمون تقسیم شدند. مدل به مدت ۱۰۰ دوره آموزش دید و پارامترهای آن در هر دوره ذخیره شدند.

برای تحلیل دقیق‌تر فعالیت زیستی ترکیبات تولیدشده، یک مدل طبقه‌بند نیز توسعه یافت. در این مدل از معماری مشابه مدل مولد استفاده گردید، با این تفاوت که خروجی آن شامل ۳ دسته‌بندی مربوط به قدرت مهارتی ترکیبات بود. دسته‌بندی ترکیبات بر اساس مقادیر IC_{50} آن‌ها صورت گرفت و شامل ۳ دسته بندی: قدرت مهارتی کمتر از ۲۰ میکرومولار (دسته بندی ۰)، ۲۰ تا ۵۰ میکرومولار (دسته بندی ۱) و بیشتر از ۵۰ میکرومولار (دسته بندی ۲) بود (۲۹).

کدگذاری ورودی در این مدل نیز با استفاده از تعبیه (Embedding) انجام گرفت و خروجی مدل با استفاده از تابع softmax و آنتروپی متقاطع چند دسته‌بندی، آموزش داده شد. این مدل امکان غربالگری سریع ترکیبات تولیدی بر اساس فعالیت پیش‌بینی شده آن‌ها را فراهم می‌کرد.

برای افزایش تنوع مولکول‌های تولیدشده به وسیله مدل CLM، از تکنیک نمونه‌برداری کنترل‌شده با دما استفاده شد (۳۰). در این روش، توزیع احتمالاتی خروجی مدل (که به وسیله تابع softmax تولید می‌شود) با یک پارامتر دما تنظیم شد. دماهای پایین (مانند ۰/۲ یا ۰/۴) منجر به تولید مولکول‌هایی با صحت ساختاری بالا و شباهت بیشتر به داده‌های اولیه می‌شوند، در حالی که دماهای بالاتر (۰/۶ تا ۰/۸) امکان تولید ساختارهای جدیدتر و متفاوت‌تر را فراهم می‌آورند. از این‌رو، انتخاب دمای بهینه در فرآیند نمونه‌برداری، تعادلی میان اعتبار شیمیایی (validity) و نوآوری ساختاری (novelty) ایجاد می‌کند که برای طراحی ترکیبات دارویی جدید از اهمیت

معماری شبکه و پارامترهای مدل انجام شد. بهترین نتیجه حاصل از آموزش مدل پایه با استفاده از دو لایه GRU با تعداد واحدهای ۵۱۲ و ۲۵۶ به همراه بردار تعبیه‌گر با اندازه ۱۰۰، دقت قابل قبولی را نشان دادند (جدول ۱). این مقادیر بیانگر توان مدل در یادگیری ساختارهای شیمیایی پیچیده و تعمیم مناسب به داده‌های جدید هستند.

به‌منظور بهبود کیفیت و تنوع ترکیبات تولیدشده، تأثیر تعداد نسخه‌های افزایش داده و دمای نمونه‌برداری بر سه شاخص کلیدی شامل درصد SMILES‌های معتبر، درصد ترکیبات یکتا و درصد ترکیبات جدید نسبت به کتابخانه اولیه مورد بررسی قرار گرفت (جدول ۲). بر اساس نتایج حاصل، مشاهده شد که در دمای پایین (دمای برابر با ۰/۲) و بدون بازنویسی، گرچه درصد ساختارهای معتبر قابل قبول بود (۶۷ درصد)، اما تنوع ساختاری وجود نداشت و تنها ۵/۶ درصد ترکیبات طراحی شده یکتا بودند. در مقابل، با افزایش تدریجی تعداد بازنویسی‌ها (تا ۲۰ نسخه) و افزایش دما از ۰/۲ تا دمای ۱ تعادل مطلوب‌تری بین درصد ترکیبات معتبر، یکتا و جدید طراحی شده به وسیله مدل حاصل شد. این نتایج نشان می‌دهد که عملکرد مدل CLM به شدت تحت تأثیر تنظیم هم‌زمان پارامترهای دما و میزان افزایش داده قرار دارد. در نهایت، افزایش داده ۲۰، دمای ۰/۸ با ۸۲ درصد ترکیب معتبر به عنوان بهترین مدل در نظر گرفته شد.

برای بررسی تنوع ساختاری ترکیبات تولیدشده، خروجی پنج مدل مولد افزایش داده (افزایش داده ۱ تا ۵) به ترتیب با دماهای نمونه‌برداری ۰/۲،

بالایی برخوردار است (۳۱). دماهای مورد استفاده در این مطالعه شامل مقادیر: ۰/۲، ۰/۴، ۰/۶، ۰/۸ و ۱/۰ بودند.

ساختار سه‌بعدی آنزیم GSK-3 β از پایگاه داده PDB با شناسه 1H8F دریافت شد. برای آماده‌سازی لیگاندها، ابتدا ساختار سه‌بعدی آن‌ها از رشته‌های SMILES با استفاده از OpenBabel تولید شد (۳۲) و سپس با روش نیمه‌تجربی AM1 در نرم‌افزار ORCA بهینه‌سازی هندسی شدند (۳۳).

داکینگ در دو مرحله انجام شد، در مرحله نخست، از نرم‌افزار AutoDock Vina برای جستجوی حالت‌های اتصال و رتبه‌بندی اولیه استفاده گردید (۳۴). سپس، برای افزایش دقت، ۹ کانفورمر پیش‌بینی شده بار دیگر با استفاده از تابع امتیازدهی AutoDock 4 مورد ارزیابی قرار گرفتند (۳۵). مختصات مرکز جعبه جست و جو در مختصات $x=40/333$ ، $y=7/527$ ، $z=37/21$ با پارامتر فاصله ۱ آنگسترومی تعیین شد که ابعاد آن $20 \times 20 \times 20$ آنگستروم بود.

برای نمایش برهم‌کنش‌های دو بعدی لیگاند-پروتئین از نرم‌افزار (۳۶) LigPlot+ استفاده گردید. همچنین برای بررسی و تصویرسازی سه‌بعدی ساختارهای داک‌شده و برهم‌کنش‌ها، از UCSF Chimera نسخه ۱.۱۵ استفاده شد که امکان تحلیل دقیق فضایی و نمایش تعاملی را فراهم نمود.

یافته‌ها

برای ارزیابی کارایی مدل مولد زبان شیمیایی (CLM) در تولید ترکیبات جدید مهارکننده GSK-3 β ، مجموعه‌ای از آزمایش‌های بهینه‌سازی

۰/۴، ۰/۶، ۰/۸ و ۱/۰ با ترکیبات مهارکننده $GSK-3\beta$ موجود در پایگاه داده *CHEMBL* مقایسه شد (شکل ۱). تحلیل فضای شیمیایی مهارکننده‌های تولید شده با استفاده از محاسبه اثرانگشت مورگان و نگاشت دوبعدی *UMAP* نشان داد که اگرچه توزیع داده‌ها متفاوت است، اما ترکیبات جدید ناحیه‌ای مشابه با فضای شیمیایی کتابخانه *CHEMBL* را پوشش می‌دهند. بر همین اساس می‌توان نتیجه گرفت که مدل به خوبی ویژگی‌های فیزیکوشیمیایی مهارکننده آنزیم *GSK* را آموخته است و علاوه بر آن توانایی طراحی ترکیباتی را دارد که با کتابخانه اولیه همپوشانی ندارد. این موضوع نشان دهنده پتانسیل مدل در طراحی ساختارهای نوآورانه با ویژگی‌های فیزیکوشیمیایی متفاوت و درعین حال منطقی است (شکل ۱).

برای ارزیابی ویژگی‌های فیزیکوشیمیایی ترکیبات طراحی شده به وسیله مدل مولد آموزش داده شده برای آنزیم $GSK-3\beta$ ، مجموعه‌ای از ویژگی‌های فیزیکوشیمیایی کلیدی در چارچوب قانون پنج‌گانه لپینسکی (*Lipinski's Rule of Five*) مورد ارزیابی قرار گرفت که شاخصی رایج در پیش‌بینی قابلیت جذب خوراکی ترکیبات دارویی است. مطابق این قانون، ترکیب‌هایی که دارای وزن مولکولی کمتر از ۵۰۰ دالتون، $LogP$ کمتر از ۵، حداکثر ۵ دهنده پیوند هیدروژنی (*HBD*) و حداکثر ۱۰ پذیرنده پیوند هیدروژنی (*HBA*) باشند، احتمال جذب مناسب خوراکی بالاتری دارند. همچنین مساحت قطبی سطحی (*TPSA*) کمتر از ۱۴۰ آنگستروم مربع که عاملی تعیین کننده برای عبور از غشای سلولی و جذب خوراکی است، تعداد پیوندهای قابل چرخش (*RTB*) کمتر از ۱۰ که

برای فراهم زیستی خوراکی دارو ها حیاتی می‌باشد و در نهایت درصد کربن‌های با هیبریداسیون درصد کربن‌ها با هیبریداسیون SP^3 برای این ترکیبات محاسبه شد. نتایج حاصل در شکل ۲ ارایه شده‌اند که مقایسه‌ای میان ترکیبات تولید شده به وسیله مدل و ترکیبات موجود در کتابخانه *ChEMBL* (به‌عنوان مرجع) را به تصویر می‌کشد.

تحلیل نمودار نشان می‌دهد که ترکیبات تولیدی مدل در محدوده استاندارد و قابل قبول ویژگی‌های دارو همانندی قرار دارند. مقدار FSP^3 نیز که شاخصی از اشباع بودن ساختار است، در ترکیبات تولیدی مقادیر متنوعی را شامل می‌شود و نشان می‌دهد که مدل مولد توانسته ترکیباتی بالقوه زیست‌فعال‌تر نسبت به ترکیبات مسطح طراحی کند. در بخش دیگری از تحلیل، با استفاده از الگوریتم کاهش ابعاد *UMAP*، ویژگی‌های فیزیکوشیمیایی ترکیبات به فضای دو بعدی نگاشته شد تا امکان مقایسه تصویری میان ترکیبات طراحی‌شده و ترکیبات *ChEMBL* فراهم شود (شکل ۳). این نگاشت نشان می‌دهد که ترکیبات تولیدشده، اگرچه بخشی از فضای شیمیایی مشترکی با ترکیبات *ChEMBL* دارند، اما در برخی نواحی نیز به صورت خوشه‌های متمایز ظاهر شده‌اند که نمایانگر نوآوری شیمیایی مدل مولد است. این تفاوت فضایی می‌تواند نشان‌دهنده توانایی مدل در کشف ساختارهایی باشد که در کتابخانه‌های سنتی کمتر مشاهده شده‌اند. بر اساس این یافته‌ها می‌توان نتیجه گرفت که مدل مولد نه تنها قادر به تولید ترکیباتی با ویژگی‌های دارویی مناسب بوده، بلکه ساختارهایی نوآورانه و غیرتکراری نیز ارایه داده است. این امر

با استفاده از نرم‌افزار *AutoDock Vina* و امتیازدهی مجدد با *AutoDock 4* انجام شد. ترکیبات منتخب حاصل از مدل مولد و مرحله یادگیری انتقالی با دقت بالا در جایگاه فعال آنزیم جای‌گذاری شدند و برهم‌کنش آن‌ها با اسید آمینه‌های کلیدی مورد بررسی قرار گرفت. جایگاه فعال آنزیم $GSK-3\beta$ از اسید آمینه‌های *Val135*, *Asp133*, *Lys85*, *Ile62*, *Pro136*, *Arg141* و *Asp200* تشکیل شده است (۳۸). بررسی ساختار کریستالی پروتئین $GSK-3\beta$ نشان‌دهنده وجود یک ناحیه حفاظت‌شده (*conserved region*) برای اتصال *ATP* می‌باشد. این ناحیه به واسطه حضور موتیف‌های اختصاصی نظیر حلقه‌ی غنی از گلیسین (*G-loop*) و ناحیه‌ی لولا (*hinge region*) به‌عنوان جایگاه اصلی اتصال لیگاندها و تنظیم فعالیت آنزیمی شناخته می‌شود. حلقه غنی از گلیسین شامل اسید آمینه‌های *Ile62* تا *Gly65* است. این حلقه بخشی از لوب N-ترمینال است و با تشکیل یک داربست (*scaffold*) انعطاف‌پذیر که با گروه‌های فسفات *ATP* تعامل دارد، نقش مهمی در اتصال و کاتالیز *ATP* ایفا می‌کند. این انعطاف‌پذیری برای نقش کاتالیتیکی $GSK-3\beta$ مهم است، زیرا حلقه غنی از گلیسین به قرار گیری مناسب فسفات‌های *ATP* برای انتقال کارآمد فسفوریل کمک می‌کند (۱۹). ناحیه لولا، ستون فقرات ساختاری برای قرارگیری *ATP* از طریق اسید آمینه‌های *Val135* و *Asp133* می‌باشد. این اسید آمینه‌ها در یک الگوی پیوند هیدروژنی که برای جهت‌گیری حلقه آدنین *ATP* حیاتی است، شرکت می‌کنند. به طور خاص، تشکیل برهم‌کنش هیدروژنی با *Val135*، جهت‌گیری مسطح باز پورین را در داخل حفره اتصال تثبیت می‌کند (۳۹).

ظرفیت بالای روش‌های یادگیری عمیق را برای طراحی ترکیبات جدید در حوزه کشف دارو به خوبی نشان می‌دهد.

نتایج مدل دسته‌بندی مهارکننده‌های آنزیم $GSK-3\beta$ در جدول ۳ گزارش شده است. از ۷۳۱ ترکیب جدید و منحصر به فرد بهترین مدل (افزایش داده ۲۰ و دمای ۰/۸) برای پیش‌بینی دسته‌بندی مهارکننده‌های تولید شده به وسیله مدل مولد استفاده شد (جدول ۳). مدل دسته‌بندی توسعه‌یافته بر پایه همان معماری *CLM*، با خروجی سه دسته بندی، با قدرت مهارتی کمتر از ۲۰ میکرومولار (دسته بندی ۰) بین ۲۰ تا ۵۰ میکرومولار (دسته بندی ۱) و بیشتر از ۵۰ میکرومولار (دسته بندی ۲) قادر به پیش‌بینی فعالیت زیستی ترکیبات تولیدی شد. بر اساس نتایج آموزش، با افزایش تعداد داده‌ها، دقت مدل بهبود یافت و در نهایت با استفاده از افزایش داده ۲۰، دقت آموزش و اعتبارسنجی به ترتیب به ۸۲ و ۸۵ درصد رسید. پیش‌بینی مدل دسته‌بندی روی ۷۳۱ ترکیب تولید شده به وسیله مدل مولد نشان داد که ۱۰/۶۷ درصد ترکیبات (۷۸ ترکیب) در دسته بندی ۱ (با IC_{50} کمتر از ۲۰ میکرومولار) قرار گرفتند که نشان دهنده پتانسیل ترکیبات تولیدی در مهار مؤثر آنزیم $GSK-3\beta$ است (جدول ۴). این یافته مؤید آن است که مدل مولد نه تنها ساختارهای معتبر و متنوع تولید می‌کند، بلکه ترکیبات حاصل پتانسیل بالایی برای فعالیت زیستی نیز دارند.

به منظور ارزیابی توانایی ترکیبات طراحی شده در اتصال مؤثر به آنزیم $GSK-3\beta$ ، داکینگ مولکولی برای ۷۸ ترکیب انتخاب شده از مراحل قبلی

نتایج حاصل از تحلیل حالت‌های اتصال (*poses*) نشان داد که تمامی لیگندهای تولیدشده قادر به برقراری پیوندهای هیدروژنی، برهم‌کنش‌های هیدروفوبیک و نیروهای واندروالس با اسید آمینه‌های کلیدی جایگاه فعال آنزیم هستند که نشان‌دهنده جهت‌گیری صحیح مولکول‌ها در داخل پاکت فعال و قابلیت اتصال اختصاصی آن‌ها می‌باشد. نتایج داکینگ مولکولی پیشبینی کرد که انرژی اتصال پیشبینی شده برای ۷۸ ترکیب منتخب در بازه ی ۴/۱۶- تا ۵/۲۵- قرار می‌گیرد. از این بین، ترکیب A با داشتن هسته ایزوکیلینولی، با انرژی پیشبینی شده ۵/۱۶- و ترکیب B با داشتن هسته کیلولینی، با انرژی پیشبینی شده ۵/۲۵- بهترین کاندیداها از نظر انرژی اتصال بودند (شکل ۴). در مجموع، نتایج داکینگ مولکولی حاکی از آن است که ۸۰ درصد از ترکیبات تولیدی به وسیله مدل از نظر برهم‌کنش‌های مولکولی قابلیت ایجاد پیوند هیدروژنی با اسید آمینه‌های کلیدی آنزیم *GSK-3β* را دارند. بررسی تعاملات لیگندهای منتخب با اجزای ساختاری کلیدی جایگاه فعال آنزیم *GSK-3β*، نشان‌دهنده نقش کلیدی دو ناحیه محافظت‌شده یعنی ناحیه لولا و حلقه غنی از گلیسین در پایداری کمپلکس لیگاند- پروتئین می‌باشد (شکل ۵). برقراری پیوند هیدروژنی با *Val135* که یکی از مؤلفه‌های اصلی اتصال *ATP* در جایگاه فعال آنزیم *GSK-3β* است، در هر دو ترکیب A و B مشاهده شد و بیانگر تقلید موفقیت‌آمیز از الگوی اتصال طبیعی *ATP* به وسیله این ترکیبات می‌باشد. از سوی دیگر، حضور پیوندهای واندروالس و هیدروفوبیک بین اسید آمینه *Ile62* در حلقه غنی از گلیسین و لیگندها نیز تأییدکننده

درگیری این لوپ انعطاف‌پذیر در پایدارسازی کمپلکس نهایی است. با توجه به آنکه حلقه غنی از گلیسین در جهت‌دهی فسفات‌های *ATP* برای فعالیت آنزیمی نقشی کلیدی دارد، برهم‌کنش مؤثر لیگندها با این ناحیه می‌تواند بیانگر توان بالقوه آن‌ها در ممانعت از فعالیت کینازی *GSK-3β* باشد. این یافته‌ها پایه‌ای منطقی برای انتخاب ترکیبات منتخب جهت ورود به مرحله بررسی ویژگی‌های فیزیکی- شیمیایی ترکیبات و دارو همانند در مرحله بعدی فراهم می‌سازد.

برای ارزیابی داروهمانندی ترکیبات با پایین‌ترین انرژی پیشبینی شده داکینگ مولکولی در مرحله قبل، ویژگی‌های فیزیکی-شیمیایی با استفاده از ابزار آنالین *SwissADME* محاسبه شد (۴۰). هر دو ترکیب از نظر تمام پارامترهای لپینسکی در محدوده مجاز قرار گرفتند و هیچ تخطی از این قانون نشان ندادند (جدول ۵). با توجه به این‌که ترکیبات با *PSA* کمتر از ۱۴۰ آنگستروم مربع معمولاً توانایی عبور از غشای سلولی و جذب خوراکی مناسبی دارند، هر دو ترکیب در این زمینه نیز واجد شرایط هستند. همچنین تعداد پیوندهای قابل چرخش (*RBC*) برای هر دو ترکیب کمتر از ۱۰ بود که به زیست‌دسترسی خوراکی آنها کمک می‌کند. لازم به ذکر است هر دو ترکیب *FCsp³* صفر دارند که نشان از مسطح و ریجید بودن ساختار است. در مجموع، ارزیابی ویژگی‌های فیزیکی-شیمیایی نشان می‌دهد که هر دو ترکیب از نظر ویژگی‌های دارو همانندی دارای شاخص‌های استاندارد جذب خوراکی و پتانسیل توسعه به‌عنوان مهارکننده‌های بالقوه محسوب

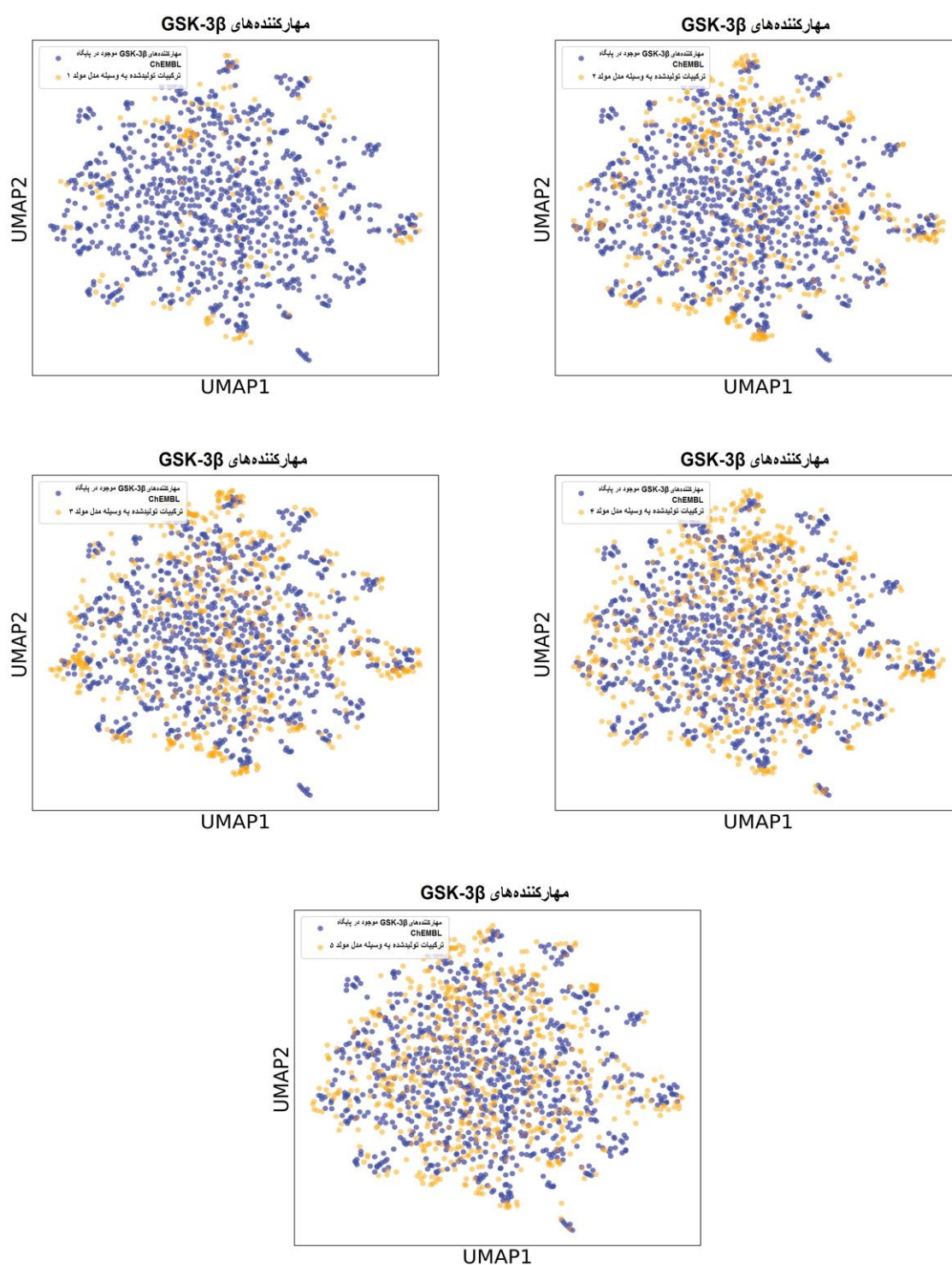
می‌شوند.

جدول ۱: نتایج بهینه‌سازی تعداد واحدهای Gru در مدل مولد مهارکننده‌های GSK-3 β

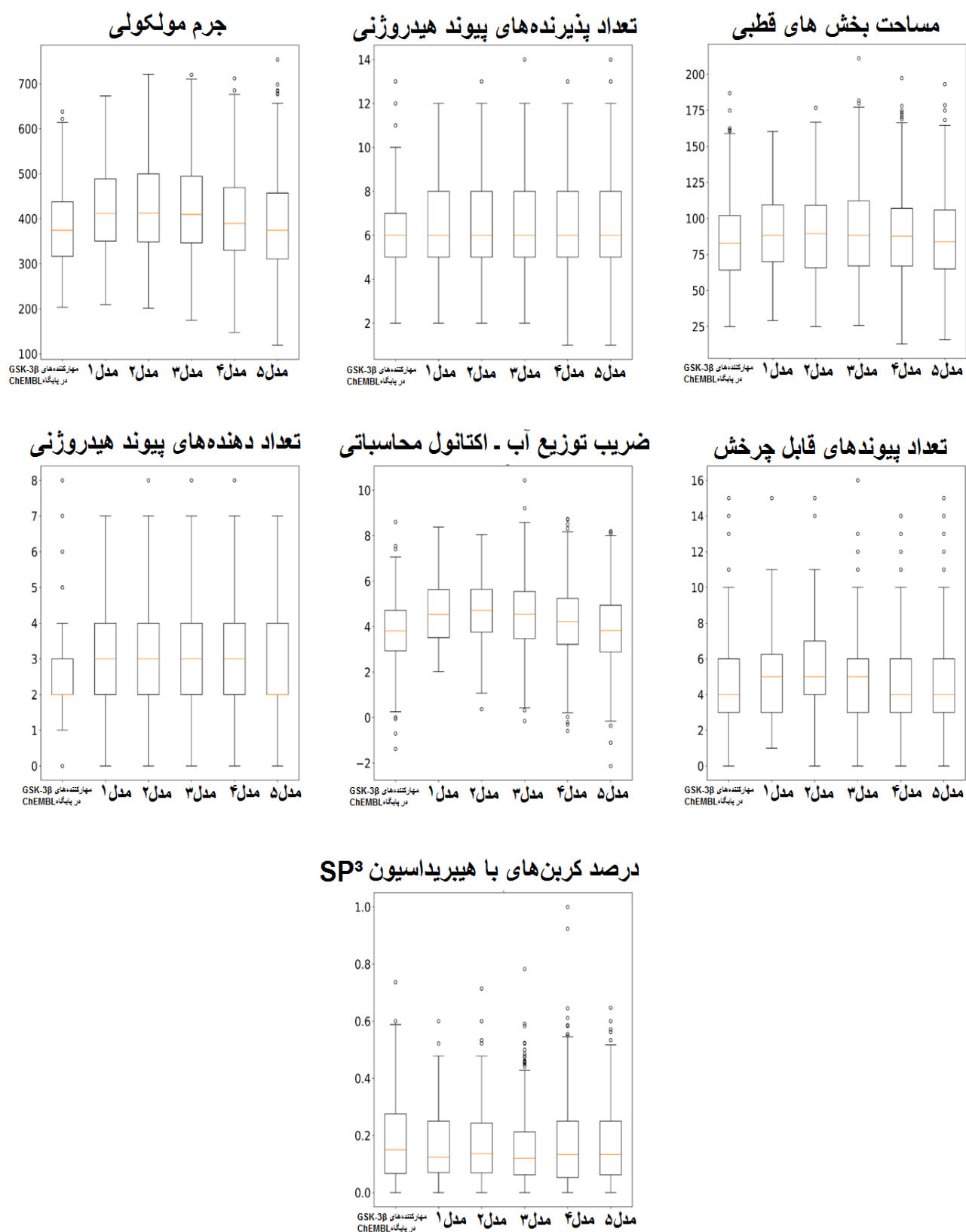
دقت آزمون (درصد)	دقت آموزش (درصد)	Embedding	Gru Units	دوره
۸۲	۸۵	۱۰۰	۲۵۶-۵۱۲	۱۰۰

جدول ۲: بهینه‌سازی تعداد افزایش داده در مدل مولد مهارکننده‌های GSK-3 β . معتبر: درصد SMILES‌های تولید شده بدون خطا، یکتا: درصد SMILES‌های تولید شده غیرتکراری و جدید: درصد SMILES‌های غیرتکراری و جدید در مقایسه با کتابخانه اولیه

افزایش داده	دما (درصد)	معتبر (درصد)	یکتا (درصد)	جدید (درصد)
	۰/۲	۶۷	۵/۶	۷۶/۳
	۰/۴	۵۲	۳۱/۴	۸۷/۲
۰	۰/۶	۳۴	۶۶/۱۷	۸۸/۸۸
	۰/۸	۲۲	۹۰/۷	۹۲/۱۹
	۱/۰	۱۳	۹۹/۲۴	۹۵/۴۵
	۰/۲	۸۹	۱۰/۶۹	۸۷/۵
	۰/۴	۷۰	۴۳/۵	۹۳/۵۴
۲	۰/۶	۵۶	۸۰/۶۳	۹۲/۲۹
	۰/۸	۳۷	۹۷/۶۲	۹۵/۴
	۱/۰	۲۷	۱۰۰	۹۸/۱۸
	۰/۲	۸۶	۲۴/۹۴	۸۸/۸۳
	۰/۴	۸۰	۶۶/۷۰	۹۲/۰۶
۵	۰/۶	۷۱	۸۹/۹۱	۹۲/۰۵
	۰/۸	۶۴	۹۵/۶۳	۹۵/۱۰
	۱/۰	۵۲	۹۹/۶۱	۹۵/۷۶
	۰/۲	۹۵	۱۸	۸۷/۱۳
	۰/۴	۸۸	۶۲/۹۳	۸۸/۸۶
۱۰	۰/۶	۸۲	۸۶/۱۴	۸۹/۸۴
	۰/۸	۷۱	۹۵/۲۵	۹۲/۹۷
	۱/۰	۶۲	۹۸/۲۴	۹۵/۹۳
	۰/۲	۹۸	۱۹/۴۳	۸۱/۲۵
	۰/۴	۹۴	۵۵/۸۴	۸۶/۰۳
۲۰	۰/۶	۹۰	۸۱/۶۵	۹۱/۳۳
	۰/۸	۸۲	۹۵/۳۶	۹۳/۴۷
	۱/۰	۶۸	۹۸/۸۳	۹۶/۳۲

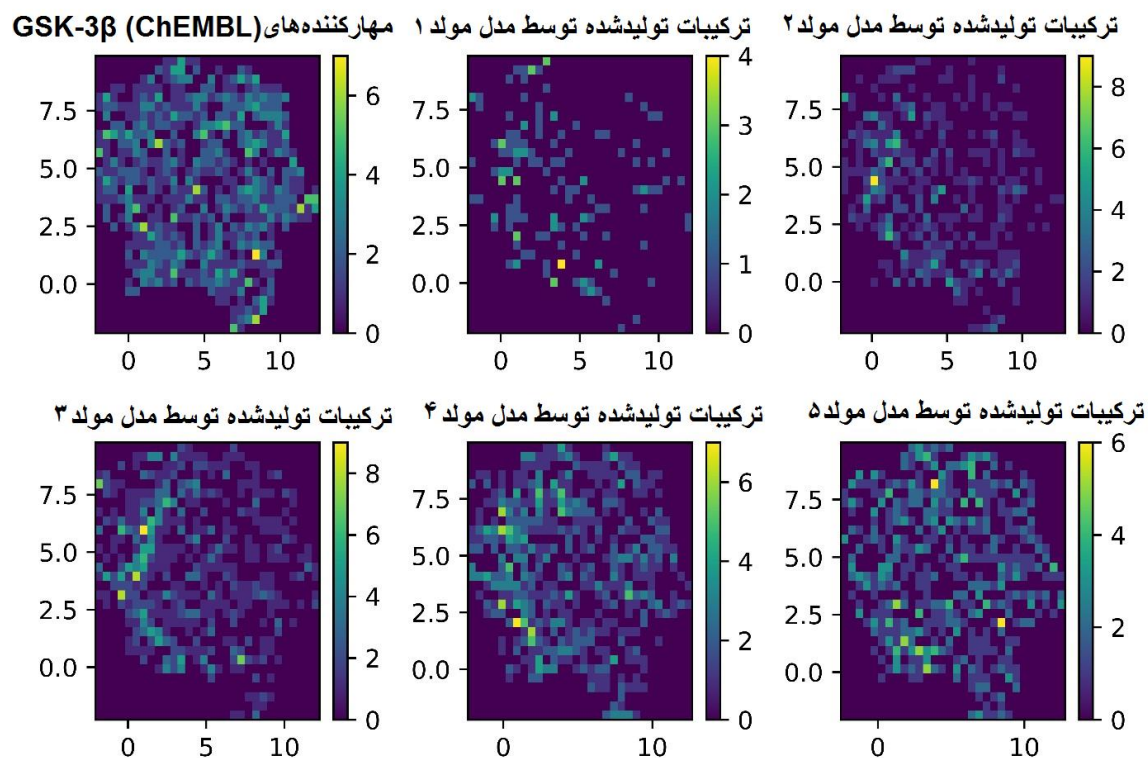


شکل ۱: مقایسه فضای شیمیایی مهارکننده‌های آنزیم *GSK-3β* در دو کتابخانه *ChEMBL* و تولید شده به وسیله مدل مولد. فضای شیمیایی با محاسبه اثر انگشت مورگان به دست آمده است.



شکل ۲: ویژگی‌های فیزیکوشیمیایی مهارکننده‌های تولیدشده به وسیله بهترین مدل مولد آنزیم $GSK-3\beta$ در مقایسه با کتابخانه ترکیبات

نگاشت دوبعدی ویژگی‌های قانون پنج‌گانه لیپینسکی



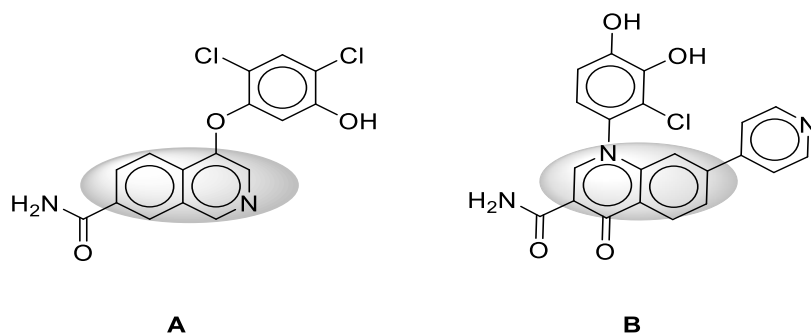
شکل ۳: مقایسه ویژگی‌های فیزیکوشیمیایی مهارکننده‌های آنزیم GSK-3 β در دو کتابخانه ChEMBL و تولید شده به وسیله مدل مولد. ویژگی‌های فیزیکوشیمیایی با استفاده از روش UMAP به دو بعد کاهش داده شده‌اند.

جدول ۳: نتایج آموزش مدل طبقه‌بندی جهت دسته‌بندی مهارکننده‌های آنزیم GSK-3 β در کتابخانه ChEMBL. بهترین مدل به صورت پررنگ نشان داده شده است.

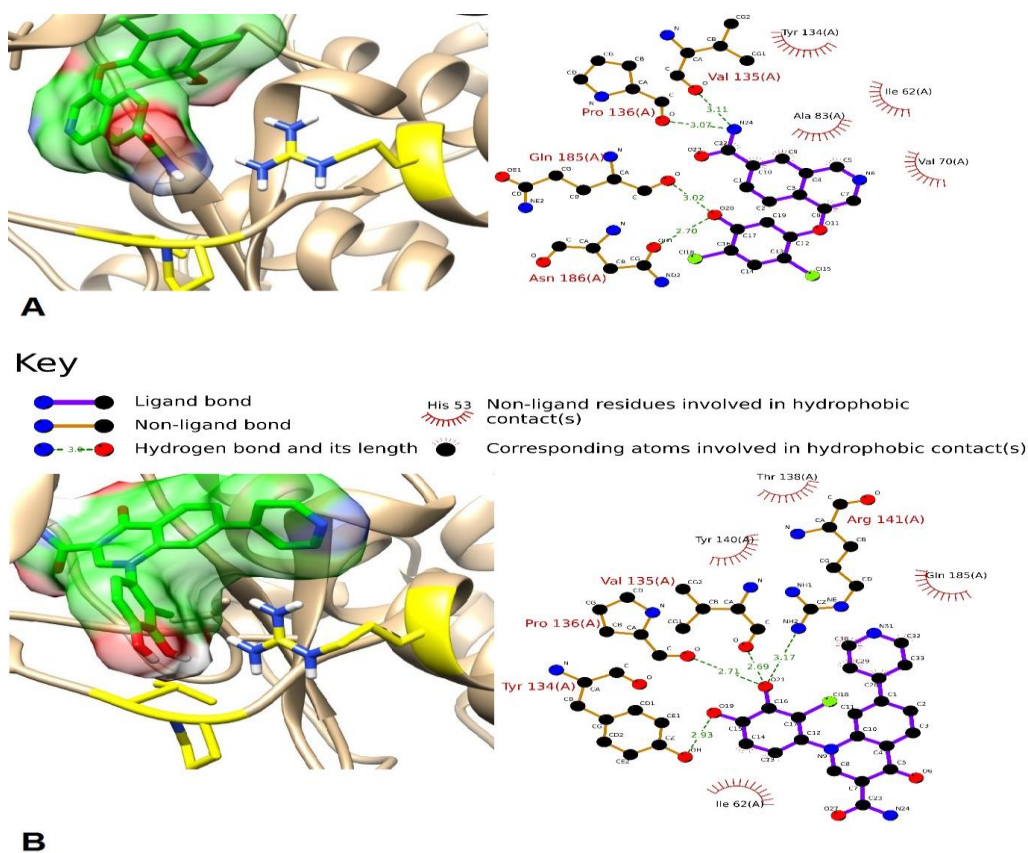
دوره	Gru Units	Embedding	افزایش داده	دقت آموزش	اعتبارسنجی دقت
۱۰۰	۵۱۲-۲۵۶	۱۰۰	۰	۰/۷۵	۰/۶۸
۱۰۰	۵۱۲-۲۵۶	۱۰۰	۲	۰/۷۰	۰/۶۹
۱۰۰	۵۱۲-۲۵۶	۱۰۰	۵	۰/۷۵	۰/۷۳
۱۰۰	۵۱۲-۲۵۶	۱۰۰	۱۰	۰/۸۳	۰/۸۰
۱۰۰	۵۱۲-۲۵۶	۱۰۰	۲۰	۰/۸۵	۰/۸۲

جدول ۴: درصد ترکیبات پیش‌بینی شده در هر دسته بندی از بهترین مدل مولد GSK-3 β (۷۳۱ ترکیب). دسته‌بندی بین سه دسته با قدرت مهارتی کمتر از ۲۰ میکرومولار (دسته بندی ۰) بین ۲۰ تا ۵۰ میکرومولار (دسته بندی ۱) و بیشتر از ۵۰ میکرومولار (دسته بندی ۲)

درصد	دسته بندی
۸۷/۵۵	۰
۱۰/۶۷	۱
۱/۷۸	۲



شکل ۴: ساختار ترکیبات آنتی‌کولین استراز منتخب طراحی شده توسط مدل آموزش دیده



شکل ۵: نحوه قرار گیری دو بعدی و سه بعدی دو لیگاند منتخب از نظر انرژی اتصال به وسیله داکینگ مولکولی

جدول ۵: ویژگی‌های فیزیکوشیمیایی مهارکننده‌های $GSK-3\beta$ منتخب پیش‌بینی شده به وسیله مدل

شماره ترکیب	فرمول شیمیایی	HBD	HBA	MW	Log P	RBC	PSA	FCsp ³
A	C ₁₆ H ₁₀ Cl ₂ N ₂ O ₃	۲	۴	۳۴۹/۱۷	۴/۱۴	۳	۸۵/۴۴	۰
B	C ₂₁ H ₁₄ ClN ₃ O ₄	۳	۵	۴۰۷/۸۱	۳/۲۲	۳	۱۱۸/۴۴	۰

Abbreviations: MW, molecular weight; HBA, number of H-bond acceptors; HBD, number of H-bond donors; Log P, the octanol-water partition coefficient; RBC, rotatable bond count; PSA, polar surface area; FCsp³, Fraction Csp³. Calculated by SwissADME

بحث

رویکردهای محاسباتی به عنوان ابزارهای ضروری در شناسایی موفقیت‌ها و سرنخ‌ها در پژوهش‌های دارویی ظهور کرده‌اند. ابزارهای یادگیری ماشین و هوش مصنوعی، از جمله مدل‌های مولد و پیش‌بینی کننده، با شناسایی نامزدهای دارویی با اختصاصیت بالا و کاهش عوارض جانبی ناخواسته، دقت و سرعت کشف دارو را افزایش می‌دهند. چنین روش‌هایی نه تنها توسعه دارو را ساده می‌کنند، بلکه هزینه‌ها و اتکا به آزمایش‌های تجربی پرهزینه و زمان‌بر را نیز کاهش می‌دهند (۴ و ۵). مدل‌های زبان شیمیایی مولد رویکردی پیشگامانه در کشف دارو هستند که از بازنمایی‌های متنی ساختارهای شیمیایی برای ایجاد ترکیبات جدید با خواص خاص برای اهداف درمانی استفاده می‌کنند. CLM ‌ها با استفاده از الگوهایی مانند شبکه‌های عصبی بازگشتی، ترانسفورماتورها، طراحی *de novo* ترکیبات را با

یادگیری از مجموعه داده‌های شیمیایی گسترده امکان‌پذیر می‌کنند (۹ و ۸). این قابلیت به ویژه در هدف قرار دادن پروتئین‌های درمانی مانند گلیکوژن سنتاز کیناز ۳ که در بیماری پارکینسون نقش دارد، ارزشمند است. نکته قابل توجه این است که ادغام CLM ‌ها در روش‌های *in silico* مانند داکینگ مولکولی، کارایی بهینه‌سازی ترکیب را افزایش می‌دهد (۱۰). لذا هدف از این مطالعه تعیین و طراحی مهارکننده‌های جدید $GSK-3\beta$ مبتنی بر هوش مصنوعی و مدل‌سازی مولکولی بود.

در این مطالعه، با به‌کارگیری مدل‌های زبانی شیمیایی (CLM) و یک چارچوب ترکیبی شامل افزایش داده و نمونه‌برداری دمایی، مجموعاً ۷۳۱ ترکیب جدید و منحصر به فرد طراحی شد. پس از غربالگری زیستی اولیه با استفاده از مدل دسته‌بندی، ۷۸ ترکیب (۱۰ درصد ترکیبات طراحی شده) از بین آن‌ها به عنوان مهارکننده‌های بالقوه و قوی با IC_{50} کمتر از ۲۰

شباهت با ترکیبات مرجع و هم از نظر بهبود ویژگی‌های فارماکوکینتیکی، عملکرد قابل قبولی دارند. این مطالعه نقش کلیدی یادگیری عمیق را در توسعه ترکیبات نوظهور و هدایت‌شده برای کاربردهای دارویی برجسته می‌سازد (۱۰).

در پژوهشی دیگر کاربرد مدل‌های مولد مبتنی بر یادگیری عمیق در طراحی دارو به صورت گسترده مورد بررسی قرار گرفته است. در این مطالعه، مدل‌هایی از جمله *RNN*، *LSTM*، *VAE* و *GAN* معرفی شده‌اند که می‌توانند با دریافت توالی‌های *SMILES* به عنوان ورودی، ساختارهای شیمیایی جدیدی با پتانسیل دارویی تولید کنند. این مدل‌ها با تکیه بر یادگیری داده‌محور قادرند ترکیباتی متنوع، معتبر و نوآورانه را طراحی کرده و آن‌ها را به سمت اهداف زیستی خاص هدایت نمایند. یافته‌های این مطالعه نشان می‌دهد که استفاده از مدل‌های مولد می‌تواند روند طراحی دارو را تسریع کرده و آن را به سمت تولید ترکیبات هدفمند و اثربخش سوق دهد (۴۲).

آروس - پوس و همکاران در پژوهشی تجربی به بررسی تأثیر استفاده از نمایش‌های تصادفی‌شده‌ی *SMILES* بر عملکرد مدل‌های زبانی شیمیایی پرداختند. در این مطالعه، مدل تولیدی مبتنی بر معماری *LSTM* با داده‌هایی شامل نسخه‌های مختلفی از *SMILES* های یکسان، اما با ترتیب متفاوت، آموزش داده شد. مقایسه عملکرد این مدل با مدلی که از *SMILES* کانیکال استفاده می‌کرد، نشان داد که تنوع ترتیبی

میکرومولار آنزیم *GSK-3 β* شناسایی شدند. داکینگ مولکولی ترکیبات منتخب مرحله قبل با ساختار سه‌بعدی *GSK-3 β* نشان داد که ترکیب‌های *A* و *B* به ترتیب با انرژی اتصال ۵/۱۶- و ۵/۲۵- کیلوکالری بر مول، بهترین انرژی‌های اتصال را دارند. همچنین داکینگ مولکولی نشان داد که هر دوی آن‌ها توانایی برقراری پیوندهای هیدروژنی با اسیدآمین‌های کلیدی *Val135* و *Pro136* جایگاه فعال را دارند. تحلیل فیزیکوشیمیایی این دو ترکیب نشان داد که هر دو در چارچوب قانون پنج‌گانه لیپینسکی قرار دارند و هیچ‌گونه تخلفی در ویژگی‌های جذب خوراکی ندارند. قابل ذکر است که با وجود داشتن ساختارهای مسطح (*FCsp³*) برابر با صفر، این ترکیبات پتانسیل بالایی برای توسعه به عنوان داروی خوراکی دارند.

در سال‌های اخیر، ظهور فناوری‌های مبتنی بر هوش مصنوعی (*AI*) چشم‌انداز جدیدی در کشف دارو فراهم کرده است. به طور خاص، *CLMs* توانسته‌اند تحول چشم‌گیری در طراحی ترکیبات دارویی ایجاد کنند.

اشنایدر و همکاران با تمرکز بر قابلیت‌های مدل‌های یادگیری عمیق در حوزه طراحی دارو، مطالعه‌ای را ارایه کردند که در آن از توالی‌های *SMILES* و معماری‌های مولد پیشرفته مانند مدل‌های ترنسفورمر بهره گرفته شد. یافته‌های این تحقیق نشان داد که چنین مدل‌هایی می‌توانند ساختارهای شیمیایی جدیدی با پتانسیل دارویی تولید کنند که هم از نظر

داده‌های ورودی می‌تواند به‌طور چشم‌گیری موجب افزایش نرخ تولید ترکیبات معتبر، منحصربه‌فرد و ساختاراً نوآورانه شود. این یافته‌ها نقش مهم روش‌های *data augmentation* مبتنی بر بازنویسی تصادفی را در افزایش توانایی مدل‌های یادگیری عمیق برای طراحی ترکیبات جدید تقویت کرده و کاربرد آن‌ها را در توسعه مدل‌های پیشرفته‌تر در حوزه شیمی دارویی برجسته می‌سازد (۴۳).

کنگ و همکاران در مطالعه‌ای کاربردی به بررسی پتانسیل مدل‌های مولد مبتنی بر نمایش *SMILES* در طراحی مولکول‌های دارویی جدید پرداختند. در این پژوهش، از معماری‌های یادگیری عمیق بر پایه *RNN* برای آموزش مدل بر روی مجموعه‌ای از توالی‌های *SMILES* استفاده شد تا ساختارهایی تولید شوند که علاوه بر اعتبار شیمیایی، از نظر زیستی نیز دارای فعالیت بالقوه باشند. نتایج مطالعه حاکی از آن بود که مدل توسعه‌یافته توانایی تولید ترکیباتی با تنوع ساختاری بالا، نحوه نگارش معتبر و خلاقیت ساختاری قابل‌توجه را دارد. همچنین، روند غربالگری و ارزیابی بیولوژیکی پس از تولید، به انتخاب ترکیبات بهینه برای مراحل بعدی کمک کرد. این تحقیق بر کارآمدی مدل‌های زبانی مبتنی بر *SMILES* در فرآیند طراحی *de novo* مولکول‌های دارویی تأکید دارد و نقش آن‌ها را به‌عنوان ابزاری نوین در شیمی محاسباتی تثبیت می‌کند (۴۴).

وانگ و همکاران در پژوهشی نوآورانه، رویکردی ترکیبی برای بهبود فرآیند تولید *de novo* ترکیبات دارویی ارائه کردند. در این مطالعه، از معماری *CycleGAN* به‌عنوان چارچوب مولد استفاده شد و به‌منظور افزایش دقت و کیفیت خروجی، لایه‌های *LSTM* و مکانیزم توجه (Attention Mechanism) در ساختار مدل تعبیه گردید. هدف اصلی، تولید مولکول‌هایی با صحت ساختاری بالا، تنوع مناسب و پتانسیل دارویی بود. نتایج نشان داد که افزودن *LSTM* به مدل موجب یادگیری بهتر وابستگی‌های ترتیبی *SMILES* شده و استفاده از توجه، به مدل امکان تمرکز بر ویژگی‌های مهم مولکول‌ها را می‌دهد. این تلفیق منجر به بهبود چشم‌گیر در اعتبار، نوآوری و کنترل‌پذیری مولکول‌های تولیدی شد. این مطالعه نشان می‌دهد که ادغام مؤلفه‌های دنباله‌محور و توجهی در چارچوب‌های مولد می‌تواند کیفیت طراحی ترکیبات جدید را در شیمی محاسباتی به‌طور معنی‌داری ارتقا دهد (۴۵).

در مطالعه‌ای که به یوهانسن و همکاران منتشر شد، از یک مدل زبانی شیمیایی تقویت شده با قطعات شیمیایی برای طراحی *de novo* آگونیست‌های گیرنده *Nurr1* در شرایط داده‌های کم استفاده شد. نتایج حاکی از آن بود که حتی با داده‌های بسیار محدود نیز می‌توان ترکیباتی با فعالیت زیستی بالا تولید کرد. این رویکرد ضمن حفظ نوآوری ساختاری، ابزار ارزشمندی برای توسعه دارو در اهداف یتیم

فراهم می‌آورد و نقش الگوریتم‌های مولد را در شناسایی *scaffold* های جدید برجسته می‌سازد (۴۶).

در مطالعه‌ای به وسیله مایکل و همکاران در رویکردی نوین برای طراحی ترکیبات دارویی با استفاده از مدل‌های زبانی شیمیایی پیشنهاد شد. آن‌ها از داده‌های *SMILES* برای آموزش مدلی بهره گرفتند که قادر به یادگیری الگوهای پیچیده موجود در ساختارهای مولکولی است و در عین حال از اطلاعات زیستی برای هدایت بهتر فرآیند تولید استفاده می‌کند. مدل طراحی شده نه تنها توانایی تولید ترکیبات منحصر به فرد و ساختاراً معتبر را داشت، بلکه قابلیت پیش‌بینی فعالیت زیستی آن‌ها نیز به شکل قابل توجهی بهبود یافت. این مطالعه نشان می‌دهد که استفاده از *CLM* ها در کنار داده‌های بیولوژیکی می‌تواند فرآیند کشف داروهای جدید را از نظر سرعت و دقت ارتقاء دهد (۴۷).

اگرچه یافته‌های این پژوهش نشان‌دهنده توان بالای چارچوب طراحی مبتنی بر هوش مصنوعی در شناسایی و تولید مهارکننده‌های جدید آنزیم *GSK-3 β* است، با این حال تمامی مراحل طراحی، غربالگری و ارزیابی ترکیبات به صورت *in silico* و بر پایه‌ی مدل‌سازی‌های محاسباتی انجام شده است و هنوز تأیید آزمایشگاهی برای ترکیبات منتخب صورت نگرفته است. لذا پیشنهاد می‌شود در گام‌های آینده، ترکیبات منتخب از نظر سمیت، زیست‌فراهمی و پایداری در محیط‌های زیستی مختلف مورد بررسی

تجربی قرار گیرند. همچنین گسترش این چارچوب به سایر اهداف دارویی و بهبود مدل‌های یادگیری ماشین می‌تواند دقت و کارایی فرآیند کشف دارو را افزایش دهد.

نتیجه‌گیری

در مجموع، این پژوهش نشان داد که استفاده از رویکردهای مبتنی بر هوش مصنوعی در کنار شبیه‌سازی‌های محاسباتی می‌تواند مسیر کشف مهارکننده‌های جدید و اختصاصی برای *GSK-3 β* را تسهیل نماید و افق‌های تازه‌ای در درمان بیماری‌های نورودژنراتیو مانند پارکینسون بگشاید.

تقدیر و تشکر

نویسندگان این مقاله مراتب تشکر و قدردانی خود را از معاونت تحقیقات و فناوری دانشگاه علوم پزشکی همدان اعلام می‌دارند.

تعارض منافع

نویسندگان مقاله هیچ‌گونه تعارض مناعفی در این مطالعه نداشتند.

حمایت مالی

این تحقیق با حمایت مالی معاونت تحقیقات و فناوری دانشگاه علوم پزشکی همدان به انجام رسیده است.

ملاحظات اخلاقی

این مقاله برگرفته از طرح تحقیقاتی "طراحی مهارکننده‌های جدید GSK-3 β مبتنی بر هوش مصنوعی و مدل‌سازی مولکولی" از دانشگاه علوم پزشکی همدان، با کد اخلاق IR.UMSHA.REC.1402.770 می‌باشد.

مشارکت نویسندگان

محمد عظیمی انجام مدل‌سازی مولکولی و نگارش اولیه مقاله را بر عهده داشت. سید احمد عبادی طراحی بخش هوش مصنوعی و نظارت بر کل پروژه را انجام داد و در بازبینی و اصلاح مقاله مشارکت داشت.

REFERENCES

1. Lin X, Li X, Lin XJM. A review on applications of computational methods in drug screening and design. *Molecules* 2020; 25(6): 1375.
2. Brogi S. Computational approaches for drug discovery. *Molecules* 2019; 24(17): 3061.
3. Naithani U, Guleria V. Integrative computational approaches for discovery and evaluation of lead compound for drug design. *Front Drug Discov* 2024; 4: 1362456.
4. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature*. 2023; 616(7958): 673–85.
5. Agamah FE, Mazandu GK, Hassan R, Bope CD, Thomford NE, Ghansah A, et al. Computational/in silico methods in drug target and lead prediction. *Brief Bioinform* 2020; 21(5): 1663–75.
6. Han R, Yoon H, Kim G, Lee H, Lee Y. Revolutionizing medicinal chemistry: the application of artificial intelligence (AI) in early drug discovery. *Pharmaceuticals(Basel)* 2023; 16(9): 1259.
7. Deng J, Yang Z, Ojima I, Samaras D, Wang F. Artificial intelligence in drug discovery: applications and techniques. *Brief Bioinform* 2022; 23(1): bbab430.
8. Wu K, Xia Y, Deng P, Liu R, Zhang Y, Guo H, et al. TamGen: drug design with target-aware molecule generation through a chemical language model. *Nat Commun* 2024; 15(1): 9360.
9. Bian Y, Xie XQ. Generative chemistry: drug discovery with deep learning generative models. *J Mol Model* 2021; 27: 1–18.
10. Moret M, Pachon Angona I, Cotos L, Yan S, Atz K, Brunner C, et al. Leveraging molecular structure and bioactivity with chemical language models for de novo drug design. *Nat Commun* 2023; 14(1): 114.
11. Dajani R, Fraser E, Roe SM, Yeo M, Good VM, Thompson V, et al. Structural basis for recruitment of glycogen synthase kinase β to the axin—APC scaffold complex. *EMBO J* 2003; 11(2): 495–507.
12. Golpich M, Amini E, Hemmati F, Ibrahim NM, Rahmani B, Mohamed Z, et al. Glycogen synthase kinase-3 beta ($\text{GSK-3}\beta$) signaling: implications for Parkinson's disease. *Pharmacol Res* 2015; 97: 16–26.
13. Yu H, Xiong M, Zhang Z. The role of glycogen synthase kinase 3 beta in neurodegenerative diseases. *Front Mol Neurosci* 2023; 16: 1209703.
14. Li J, Ma S, Chen J, Hu K, Li Y, Zhang Z, et al. $\text{GSK-3}\beta$ contributes to Parkinsonian dopaminergic neuron death: Evidence from conditional knockout mice and tideglusib. *Front Mol Neurosci* 2020; 13: 81.
15. Li DW, Liu ZQ, Chen W, Yao M, Li GR. Association of glycogen synthase kinase- β with Parkinson's disease. *Mol Med Rep* 2014; 9(6): 2043–50.
16. Morales-García JA, Susín C, Alonso-Gil S, Pérez DI, Palomo V, Pérez C, et al. Glycogen synthase kinase-3 inhibitors as potent therapeutic agents for the treatment of Parkinson disease. *ACS Chem Neurosci* 2013; 4(2): 350–60.
17. Park SH, Xu Y, Park YS, Seo JT, Gye MC. Glycogen synthase kinase-3 isoform variants and their inhibitory phosphorylation in human testes and spermatozoa. *World J Mens Health* 2022; 41(1): 215.
18. Jacobs KM, Bhawe SR, Ferraro DJ, Jaboin JJ, Hallahan DE, Thotala D. $\text{GSK-3}\beta$: a bifunctional role in cell death pathways. *Int J Cell Biol* 2012; 2012: 930710.
19. Stamos JL, Chu MLH, Enos MD, Shah N, Weis WI. Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *ELife* 2014; 3: e01998.
20. Beurel E, Grieco SF, Jope RS. Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. *Pharmacol Ther* 2015; 148: 114–31.
21. Hardt SE, Sadoshima J. Glycogen synthase kinase- β : a novel regulator of cardiac hypertrophy and development. *Circ Res* 2002; 90(10): 1055–63.
22. Luo J. Glycogen synthase kinase β ($\text{GSK3}\beta$) in tumorigenesis and cancer chemotherapy. *Cancer Lett* 2009; 273(2): 194–200.
23. Asl FSS, Malverdi N, Mojahedian F, Baziyaar P, Nabi-Afjadi M. Discovery of effective $\text{GSK-3}\beta$ inhibitors as therapeutic potential against Alzheimer's disease: a computational drug design insight. *Int J Biol Macromol* 2025; 306: 141273.
24. Ganai SA, Mohan S, Padder SA. Exploring novel and potent glycogen synthase kinase- β inhibitors through systematic drug designing approach. *Sci Rep* 2025; 15(1): 4118.
25. Landrum G. *RDKit: Open-source cheminformatics*. Zenodo; 2006. Available from: <https://www.rdkit.org>

26. Skinnider MA. Invalid SMILES are beneficial rather than detrimental to chemical language models. *Nat Mach Intell* 2024; 6(4): 437-48.
27. Bjerrum EJ. SMILES enumeration as data augmentation for neural network modeling of molecules. *arXiv preprint arXiv:1703.07076*. 2017.
28. Segler MH, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci* 2018; 4(1): 120-31.
29. Sheridan RP. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J Chem Inf Model* 2013; 53(4): 783-90.
30. Thomas M, O'Boyle NM, Bender A, De Graaf C. MolScore: a scoring, evaluation and benchmarking framework for generative models in de novo drug design. *J Cheminform* 2024; 16(1): 64.
31. Cafiero M. Variable-temperature token sampling in decoder-GPT molecule-generation can produce more robust and potent virtual screening libraries. *Phys Chem Chem Phys* 2025; 27(14): 18462-74.
32. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform* 2011; 3: 1-14.
33. Neese F, Wennmohs F, Becker U, Riplinger C. The ORCA quantum chemistry program package. *J Chem Phys* 2020; 152(22): 224108;
34. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010; 31(2): 455-61.
35. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009; 30(16): 2785-91.
36. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng Des Sel* 1995; 8(2): 127-34.
37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004; 25(13): 1605-12.
38. Dajani R, Fraser E, Roe SM, Young N, Good V, Dale TC, et al. Crystal structure of glycogen synthase kinase 3 β : structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* 2001; 105(6): 721-32.
39. Shri SR, Manandhar S, Nayak Y, Pai KSR. Role of GSK-3 β inhibitors: new promises and opportunities for Alzheimer's disease. *Adv Pharm Bull* 2023; 13(4): 688.
40. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 2017; 7(1): 42717.
41. Naithani U, Guleria V. Integrative computational approaches for discovery and evaluation of lead compounds for drug design. *Front Drug Discov* 2024; 4: 1362456.
42. Bian Y, Xie XQ. Generative chemistry: drug discovery with deep learning generative models. *J Mol Model* 2021; 27(1): 1-18.
43. Arús-Pous J, Johansson SV, Prykhodko O, Bjerrum EJ, Tyrchan C, Reymond JL, et al. Randomized SMILES strings improve the quality of molecular generative models. *J Cheminform* 2019; 11(1): 1-13.
44. Kong W, Hu Y, Zhang J, Tan Q. Application of SMILES-based molecular generative model in new drug design. *Front Pharmacol* 2022; 13: 1046524.
45. Wang F, Feng X, Guo X, Xu L, Xie L, Chang S. Improving de novo molecule generation by embedding LSTM and attention mechanism in CycleGAN. *Front Genet* 2021; 12: 709500.
46. Ballarotto M, Willems S, Stiller T, Nawa F, Marschner JA, Grisoni F, et al. De novo design of Nurr1 agonists via fragment-augmented generative deep learning in low-data regime. *J Med Chem* 2023; 66(12): 8170-7.
47. Moret M, Pachon Angona I, Cotos L, Yan S, Atz K, Brunner C, et al. Leveraging molecular structure and bioactivity with chemical language models for de novo drug design. *Nat Commun* 2023; 14(1): 114.

Design of Novel GSK-3 β Inhibitors Based on Artificial Intelligence and Molecular Modeling

Azimi M, Ebadi SA*

Medicinal Plants and Natural Products Research Center, Hamadan University of Medical Sciences, Hamadan, Iran.

Received Date: 08 Jun 2025 Accepted Date: 17 Aug 2025

Abstract

Background & aim: Given the key role of GSK-3 β in the pathogenesis of Parkinson's disease and the challenges in designing specific inhibitors, the aim of the present study was to identify and design new GSK-3 β inhibitors based on artificial intelligence and molecular modeling.

Methods: The present descriptive–analytical computational (in silico) study was conducted in 2025. An AI-driven workflow was applied to design and evaluate novel GSK-3 β inhibitors with potential therapeutic activity against Parkinson's disease. Using CLMs, a set of de novo compounds was generated and subsequently filtered based on drug-likeness, structural diversity, and physicochemical properties. Data augmentation (20-fold) and temperature sampling ($T = 0.8$) were applied to enhance generalizability and molecular diversity. The collected data were analyzed using descriptive and comparative computational analyses based on model performance and physicochemical indices

Results: The model achieved a success rate of 73.1% in generating valid inhibitors. Among these, 78 compounds with predicted inhibitory potency below 20 μ M were selected for molecular docking at the active site of GSK-3 β . Docking analysis revealed that Compound A, featuring an isoquinoline scaffold, achieved a predicted binding energy of -16.5 kcal/mol, while Compound B, containing a quinoline scaffold, achieved -25.5 kcal/mol. Both compounds demonstrated favorable pharmacokinetic properties and full compliance with drug-likeness rules, including oral absorption criteria.

Conclusion: The integration of artificial intelligence with molecular modeling provides an efficient strategy for the design of selective and potent GSK-3 β inhibitors, offering a promising pathway for the development of novel therapeutics for Parkinson's disease.

Keywords: Artificial intelligence, Chemical language model, GSK-3 β , Parkinson's disease, Molecular docking.

Corresponding Author: Ebadi SA, Medicinal Plants and Natural Products Research Center, Hamadan University of Medical Sciences, Hamadan, Iran.
Email: ahmadebadie@gmail.com, a.ebadi@umsha.ac.ir

Please cite this article as follows: Azimi M, Ebadi SA. Design of Novel GSK-3 β Inhibitors Based on Artificial Intelligence and Molecular Modeling. Armaghane-danesh 2025; 30(3): 388-409.

The scientific research journal Armaghan Danesh, affiliated with Yasuj University of Medical Sciences, is an open-access publication. All articles published in this journal