

بهبود تخمین اثر بیولوژیکی ترکیبات مرتبط با مهار کننده‌های استیل کولین استراز و بوتیل کولین استراز با استفاده از روش غیر خطی جنگل تصادفی

فهیمة معتمدی^۱، علیرضا مهری دهنوی^۲، فهیمة قاسمی^۳

مقاله پژوهشی

چکیده

مقدمه: امروزه با توجه به افزایش جمعیت سالمندان جهان و روند رشد بیماری آلزایمر، بررسی مهار کننده‌های Acetyl cholinesterase (AChE) و Butyl cholinesterase (BChE) که تجزیه شدن آن‌ها در مغز از دلایل اصلی ایجاد این بیماری می‌باشد، امری ضروری به نظر می‌رسد. از آن جایی که سنتز و بررسی هر ترکیب جدید بسیار پرهزینه و زمان‌بر است، محققان سعی بر تخمین میزان فعالیت بیولوژیکی ترکیبات قبل از انجام فعالیت‌های آزمایشگاهی دارند. تا کنون روش‌های محاسباتی مختلفی پیشنهاد شده است که از مهم‌ترین آن‌ها می‌توان به بررسی کمی ارتباط ساختار-فعالیت اشاره نمود که بر اساس مدل‌سازی خطی و یا غیر خطی با استفاده از محاسبه‌ی توصیف‌گرهای ملکولی استوار می‌باشند. مطالعه‌ی حاضر با هدف شناسایی مدل مناسب جهت پیش‌بینی اثر بیولوژیکی ترکیبات مرتبط با مهار کننده‌های آنزیم‌های ترکیبات استیل کولین استراز و بوتیل کولین استراز با استفاده از مدل‌سازی Quantitative structure-activity relationship (QSAR) و با روش جنگل تصادفی انجام شد.

روش‌ها: در راستای پیش‌بینی فعالیت دارویی ترکیبات AChE و BChE، روش خطی حداقل مربعات جزئی و روش غیر خطی جنگل تصادفی به کار گرفته شدند. به منظور دستیابی به نتایج دقیق‌تر و قابل اعتمادتر، در هر اجرا به صورت تصادفی، ۸۰ درصد ترکیبات به عنوان نمونه‌ی آموزشی و سایر ترکیبات به عنوان نمونه‌ی آزمایشی جهت ساخت مدل و ارزیابی قدرت پیش‌بینی مدل استفاده شدند.

یافته‌ها: با اعمال مدل غیر خطی جنگل تصادفی بر روی مهار کننده‌های استیل کولین استراز و بوتیل کولین استراز، دقت ۸۹ درصد حاصل شد. در پایان، به منظور بررسی دقیق‌تر عملکرد مدل پیشنهادی، نتایج با خروجی حاصل از روش کیمینه‌ی خطای جزئی مقایسه گردید. نتایج نشان داد که روش جنگلی تصادفی غیر خطی نسبت به روش مربعات حداقل خطی، عملکرد قوی‌تری دارد.

نتیجه‌گیری: مشاهدات بیانگر این است که روش غیر خطی جنگل تصادفی، می‌تواند در تخمین اثر بیولوژیکی ترکیبات استیل کولین استراز و بوتیل کولین استراز پیشنهادی توسط پزشکان و متخصصان شیمی دارویی بسیار مؤثر باشد. بنابراین، قبل از انجام آزمایش‌های حیوانی و انسانی ترکیب پیشنهادی، فعالیت بیولوژیکی این ترکیب به طور تقریبی ۹۰ درصد برآورد می‌شود. بر اساس میزان فعالیت بیولوژیکی تخمین زده شده، می‌توان اظهار نظر کرد که «آیا ترکیب دارویی جدید با صرف هزینه و زمان، خطر تبدیل شده به یک داروی جدید را دارد یا خیر؟».

واژگان کلیدی: ارتباط کمی ساختار-فعالیت، استیل کولین استراز، حداقل مربعات جزئی

ارجاع: معتمدی فهیمة، مهری دهنوی علیرضا، قاسمی فهیمة. بهبود تخمین اثر بیولوژیکی ترکیبات مرتبط با مهار کننده‌های استیل کولین استراز و بوتیل کولین استراز با استفاده از روش غیر خطی جنگل تصادفی. مجله دانشکده پزشکی اصفهان ۱۳۹۸؛ ۳۷ (۵۵۸): ۱۴۰۶-۱۴۰۱

بارزترین نشانه‌های این بیماری می‌باشند (۱). در حال حاضر، به طور تقریبی ۴۶/۸ میلیون نفر در جهان به این بیماری دچار هستند و با توجه به این که افزایش سن مهم‌ترین عامل ابتلا به آلزایمر است، پیش‌بینی می‌شود با افزایش جمعیت سالمندان، تعداد افراد مبتلا به این

مقدمه

بیماری آلزایمر (Alzheimer's disease یا AD)، یک اختلال عصبی چند وجهی و شایع‌ترین نوع زوال عقل است که از دست دادن حافظه، نقص در عملکردهای شناختی و ناهنجاری‌های رفتاری از

- ۱- دانشجوی کارشناسی ارشد، کمیته‌ی تحقیقات دانشجویی، گروه بیوالکتریک، دانشکده‌ی فن‌آوری‌های نوین علوم پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران
 - ۲- استاد، گروه بیوالکتریک، دانشکده‌ی فن‌آوری‌های نوین علوم پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران
 - ۳- استادیار، گروه بیوالکتریک، دانشکده‌ی فن‌آوری‌های نوین علوم پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران
- نویسنده‌ی مسؤؤل: علیرضا مهری دهنوی؛ استاد، گروه بیوالکتریک، دانشکده‌ی فن‌آوری‌های نوین علوم پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

Email: mehri@med.mui.ac.ir

هدف اصلی این مقاله، شناسایی مدل مناسب جهت پیش بینی اثر بیولوژیکی ترکیبات مرتبط با مهار کننده های آنزیم های ترکیبات استیل کولین استراز و بوتیل کولین استراز با استفاده از مدل سازی QSAR و با روش جنگل تصادفی می باشد. برای رسیدن به این هدف، ۲۱۰ ترکیب که در سال های ۲۰۱۹-۲۰۱۸ سنتز شده بودند، استخراج شد. از طرف دیگر، با توجه به موفقیت روش کمینه ی خطای جزئی در پیش بینی اثر بیولوژیکی ترکیبات در کنار سادگی و سرعت بالای مدل در مقایسه با مدل های غیر خطی، در پایان نتایج روش پیشنهادی با این روش مورد مقایسه قرار گرفت.

آماده سازی داده ها: یکی از مسایلی که در حوزه ی طراحی دارو از اهمیت ویژه ای برخوردار است، تهیه و آماده سازی داده ها می باشد. این داده ها شامل اطلاعاتی در زمینه ی گیرنده و ملکول ها می باشد که برای رویکردهای مختلف، متفاوت است. تعداد مناسب و قابل قبولی از نمونه ها، یکی از شرایط ضروری پیاده سازی مدل QSAR می باشد. بنابراین، باید تعداد کافی از ترکیبات کولین استراز به همراه میزان فعالیت بیولوژیکی آن ها که در آزمایشگاه سنتز شده بودند، جمع آوری می شد. در این راستا، مطالعات انجام گرفته شامل ۲۱۰ ترکیب همراه با عدد IC_{50} گردآوری شد که دو دسته از مشتقات کولین استراز را در برمی گرفت (۷-۹).

رسم ترکیبات با نرم افزار Hyperchem با استفاده از نرم افزار Hyperchem ساختار دو بعدی ملکول به طور کامل ترسیم شد و طول پیوند، ساختار فضایی نمونه ها نظیر نحوه ی توزیع بار فضایی و یا جهت گیری فضایی اتم ها مانند زاویه ی پیوندی و پیچشی توسط الگوریتم هایی نظیر الگوریتم ژنتیک، گرادیان نزولی، گاس-نیوتن بهینه سازی می شوند. پس از بهینه سازی، با یکی از روش های بهینه سازی، ساختارها در قالب hin ذخیره شدند.

استخراج ویژگی های ترکیبات دارویی با نرم افزار Dragon اطلاعاتی که از ساختار دو بعدی و سه بعدی ترکیبات به دست می آید، توصیف گر (Descriptor) می گویند. به منظور دست یابی به توصیف گرهای ملکولی، لازم است ابتدا مانند قبل به ساختار دسترسی داشته باشیم؛ سپس، با استفاده از نرم افزار Dragon می توان توصیف گرها را استخراج نمود.

توصیف گرها، مقادیری هستند که به ساختار ملکول ها اختصاص دارند و برای ملکول های مختلف، متفاوت می باشند. امروزه، تعداد آن ها به بیش از ۱۰۰۰ مورد می رسد.

علاوه بر محاسبه ی ساده ترین نوع اتم ها و گروه های عاملی و شمارش اجزا، می توان تعداد زیادی از توصیف کننده های توپولوژیکی و هندسی را نیز توسط این نرم افزار محاسبه نمود. Dragon یک فایل خروجی کامل را که به سادگی توسط هر نرم افزار واکاوی همبستگی

بیماری تا سال ۲۰۵۰ سه برابر شود (۲). با توجه به این که تا کنون هیچ درمانی وجود ندارد که پیشرفت بیماری آلزایمر را متوقف یا معکوس کند، بررسی راه های جلوگیری و درمان برای این بیماری امری ضروری است.

کاهش سطح استیل کولین (ACh)، یک دلیل قابل اعتماد برای علت بیماری آلزایمر است. در حال حاضر، مؤثرترین روش برای درمان آلزایمر، افزایش سطح استیل کولین در مغز است که به طور عمده توسط کولین استراز (CHE) تجزیه می شود. بنابراین، مؤثرترین روش برای درمان آلزایمر، استفاده از مهار کننده های کولین استراز است. سیستم عصبی، دارای دو نوع کولین استراز به نام های استیل کولین استراز (AChE) و بوتیل کولین استراز (BChE) می باشد. استیل کولین استراز در مقایسه با بوتیل کولین استراز 10^{13} برابر فعال تر است و به طور تقریبی، ۸۰ درصد هیدرولیز استیل کولین در مغز را تشکیل می دهد (۳).

از سال ۲۰۰۳ تاکنون، هیچ داروی جدیدی تأیید نشده است و درمان بهبود دهنده ی بیماری (Disease-modifying treatments) یا DMTs برای بیماری آلزایمر وجود ندارد (۴). بنابراین، جستجوی داروهای درمانی مؤثر برای معالجه ی بیماری آلزایمر، یک امر مهم در شیمی دارویی (Medicinal chemistry) است. از آن جایی که چالش های توسعه ی دارو پیچیده تر شده اند و میزان زیادی از برنامه های توسعه ی داروی بیماری آلزایمر با شکست مواجه شده است، روش های مدل سازی محاسباتی مبتنی بر لیگاند و مبتنی بر ساختار (Ligand and structural-based computational modeling) در دهه های گذشته به طور گسترده مورد مطالعه قرار گرفته اند. روابط کمی ساختار- فعالیت (Quantitative Structure- Activity relationships) یا QSAR، ابزارهای مهمی برای مطالعه ی مهار کننده های AChE و BChE هستند. رابطه ی بین ساختار و فعالیت به عنوان یک مدل ریاضی تنظیم می شود و بر اساس مدل QSAR به دست آمده، میزان فعالیت بیولوژیکی (IC_{50}) ترکیبات پیشنهادی و جدید تخمین زده می شود (۵).

بر اساس تجربه و انتظاراتی که پزشک از داروی مورد نظر دارد، ابتدا یک سری ترکیبات جدید برای مهار آنزیم های AChE و BChE پیشنهاد می شود، به کمک متخصص شیمی - دارو، ترکیبات پیشنهادی به تأیید اولیه می رسند و سپس، مدل های QSAR پیش بینی کننده به بررسی مفید بودن ترکیب جدید قبل از سنتز و آزمایش در آزمایشگاه می پردازند. با کمک این رویکرد، غربالگری یک کتابخانه ی بزرگ از نامزدهای احتمالی ترکیباتی که شانس تبدیل شدن به دارو را دارند، سریع تر و راحت تر انجام می شود و زمان تحقیقات و بررسی برای کشف داروی جدید کاهش می یابد (۶).

نمونه‌ها با Bootstrap ایجاد می‌شوند و سپس، درخت تصمیم بر روی تمام این زیرمجموعه‌ها اعمال می‌گردد.

حدود دو سوم نمونه‌های کل مجموعه‌ی داده‌ی اصلی، برای آموزش و یک سوم باقی مانده، برای آزمایش در نظر گرفته شدند. سپس، روی نمونه‌های انتخاب شده برای مجموعه‌ی آموزشی، تکنیک Bootstrap اجرا گردید. این تکنیک، به تعداد درختان تعیین شده برای جنگل، زیرمجموعه‌ای از نمونه‌ها را به روش جایگزینی و به صورت تصادفی تولید می‌کند. شایان ذکر است تکنیک Bootstrap، تنها در مرحله‌ی آموزش به کار گرفته می‌شود.

مدل آموزشی جنگل تصادفی، مانند سایر مدل‌های ماشین یادگیری، متغیری کردن و بهینه‌سازی کردن را انجام می‌دهد. منظور از متغیری کردن، تعیین تعداد زیرمجموعه‌های تولید شده توسط بوت استرپ، اندازه‌ی هر زیرمجموعه از ویژگی‌ها در هر گره درخت و تعیین معیار تقسیم برای هر زیرمجموعه از ویژگی‌ها می‌باشد.

در مرحله‌ی آزمایش، از روش کیسه‌بندی استفاده می‌شود. نمونه‌ی جدید وارد می‌گردد و هر درخت، یک برچسب برای گروه یا یک مقدار برای پیش‌بینی اثر زیست‌شناختی نمونه‌ی جدید ارائه می‌کند. سپس، روش کیسه‌بندی از جمع‌آوری نتیجه‌ی تمام درختان جنگل برای تصمیم‌گیری نهایی درباره‌ی نمونه‌ی جدید استفاده می‌کند. اگر هدف، تخمین اثر زیست‌شناختی باشد، باید از میانگین تخمین تمام درختان استفاده کرد و در صورتی که هدف از به کار گرفتن الگوریتم جنگل تصادفی، انتخاب بهترین ویژگی باشد، ویژگی‌های به دست آمده از هر درخت مرتب می‌شوند و سپس، بر اساس میزان تکرار ویژگی‌ها در گره‌های بالایی، بهینه‌ترین ویژگی‌ها استخراج می‌گردند (۱۰).

مدل خطی حداقل مربعات جزئی (Partial least squares یا

PLS): روش حداقل مربعات جزئی، توسط Herman Wold در سال ۱۹۸۴ ارائه شد که هدف از آن، بیشینه‌سازی تنوع متغیرهای وابسته‌ای است که توسط متغیرهای مستقل تعریف می‌شوند. روش PLS، حساسیت کمی به حجم نمونه دارد و نیازی به طبیعی سازی داده‌ها نیست.

رھیافت PLS به عنوان روش‌های کوچک‌سازی (Shrinkage methods) رفتار می‌کند. این الگوریتم، به جای بازتولید ماتریس کوواریانس تجربی بر بیشترین واریانس تبیین شده‌ی متغیرهای وابسته به وسیله‌ی متغیرهای مستقل، تمرکز دارد و برای پیش‌بینی به کار می‌رود و به نظریه‌ی قوی نیاز ندارد. این تکنیک، امکان بررسی روابط متغیرهای پنهان و متغیرهای قابل مشاهده را به صورت هم‌زمان فراهم می‌سازد (۱۱).

قابل به کارگیری است، فراهم می‌سازد تا به کمک آن، بتوان مطالعات QSAR را انجام داد. خروجی Dragon آن به فرم زیر می‌باشد:

$$descriptors = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{bmatrix} \quad (1-2)$$

در این جا m تعداد ملکول‌ها و n تعداد توصیف‌گرها می‌باشد. در پایان، دو دسته تارگت به صورت ماتریس عددی آماده شد که اطلاعات این دو دسته داده در جدول ۱ قابل مشاهده است.

جدول ۱. داده‌های کولین استراز

نام تارگت	نام اختصاری تارگت	تعداد توصیف‌گرها	تعداد ترکیبات
Acetyl choline esterase	AChE	۹۲۸	۲۱۰
Butyl choline esterase	BChE	۸۱۷	۱۰۴

روش‌ها

ابتدا لازم است تعریف علمی از واژگان کلیدی که در مطالعه به آن‌ها اشاره می‌شود، ارائه گردد:

طراحی محاسباتی دارو: روشی برای طراحی و ساخت داروهای با بازده و تأثیر بیشتر که در آن داروها می‌توانند هم‌زمان چندین پروتئین را هدف قرار دهند.

رابطه‌ی کمی ساختار- فعالیت: مطالعه‌ی رابطه‌ی کمی ساختار- فعالیت (QSAR)، یکی از روش‌های کمومتریکی است که مدل‌های ریاضی را ایجاد می‌کند تا بتواند رابطه‌ی معنی‌دار آماری بین ساختار و عملکرد توصیف‌گر را به دست آورد. هدف اصلی تحقیقات QSAR، پیش‌بینی خواص فیزیکی، بیولوژیکی و شیمیایی پیچیده‌ی مواد شیمیایی از طریق ساختارهای ملکولی آن‌ها می‌باشد.

اثر زیست‌شناختی: غلظتی از اسانس و عصاره که منجر به مهار ۵۰ درصد آنزیم می‌شود (IC₅₀).

مدل غیر خطی جنگل تصادفی: جنگل تصادفی، یک الگوریتم مبتنی بر درخت تصمیم است که در بر گیرنده‌ی سه تکنیک نمونه‌برداری به روش جایگزینی از مجموعه‌ی داده‌ی اصلی، انتخاب تصادفی زیرمجموعه از ویژگی‌ها و ترکیب نتایج کلیه‌ی درختان است. نمونه‌برداری از کل نمونه‌های موجود توسط روش Bootstrap و ترکیب نتایج درختان با استفاده از روش کیسه‌ای بر روی داده‌های استیل کولین استراز انجام شد.

یکی از ویژگی‌های بارز جنگل تصادفی، ساختار موازی آن است. منظور از ساختار موازی، به کار گرفتن چندین درخت تصمیم به صورت هم‌زمان است که برای این منظور، زیرمجموعه‌هایی از

بیماری، از مهم ترین درمان های بیماری آلزایمر هستند. با استفاده از داروهای مؤثر، می توان مانع پیشرفت بیماری شد و شدت اختلال حافظه و مشکلات رفتاری بیمار را کاهش داد. رویکردهای درمانی کلی، متکی به استفاده از مهار کننده های استیل کولین استراز است که می تواند عملکرد عصب کولینرژیک مغز را در برخی بیماران حفظ کند. با این وجود، هنوز هم استفاده از روش های سنتی تجربی یا داروسازی، مشکلات زیاد و هزینه های سنگینی را در بر می گیرد.

طراحی دارویی با کمک رایانه (Computer-aided drug design یا CADD)، یک روش قدرتمند در تولید دارو برای پیش گیری یا درمان بیماری آلزایمر است. پزشک بر اساس نقص ها و عوارضی که داروهای فعلی روی بیماران آلزایمر دارند، ترکیبات جدید با عملکرد بهتری پیشنهاد می دهد. سپس، متخصص شیمی - دارو، این ترکیبات را تأیید یا رد می کند. اگر ترکیبات پیشنهادی پزشک مورد تأیید قرار بگیرند، به کمک روش طراحی دارو با کامپیوتر، بدون صرف زمان و هزینه های سنگین آزمایشگاهی می توان تخمین زد که ترکیب جدید تا چه اندازه می تواند در درمان بیماری آلزایمر موفق باشد.

همان گونه که در بخش آماده سازی داده اشاره شد، این داده ها از مقالات ۲۰۱۸-۲۰۱۹ گردآوری شده اند و تاکنون در مدل سازی QSAR استفاده نشدند. بنابراین، امکان مقایسه ی عملکرد مدل های پیشنهادی با مدل های گذشته بر روی این دسته از ترکیبات کولین استراز وجود نداشت. دقت نامناسب در خروجی مدل ها و زمان بر بودن اجرای مدل ها، همواره دو چالش اساسی در مدل های اجرا شده بر روی ترکیبات دارویی مختلف در سال های گذشته بوده است. روش جنگل تصادفی، با دقت حدود ۹۰ درصد و زمان اجرای کوتاه، توانست بر این دو چالش غلبه کند.

نتیجه گیری نهایی این که اختلاف زیاد بین همبستگی حاصل شده از مدل جنگل تصادفی و حداقل مربعات جزئی، نشان دهنده ی قدرت پیش بینی خوب جنگل تصادفی می باشد. از آن جایی که رابطه ی بین توصیف گر ها و فعالیت ترکیبات یک رابطه ی غیر خطی است، به نظر می رسد استفاده از روش های غیر خطی نظیر جنگل تصادفی کارآمد تر است و مدل مناسب تر و دقیق تری جهت تخمین اثر زیست شناختی ترکیبات را فراهم می کند. بنابراین، از بین ترکیبات زیاد پیشنهادی توسط داروسازان، می توان با روش جنگل تصادفی با دقتی نزدیک به ۹۰ درصد میزان فعالیت زیست شناختی را پیش بینی کرد و اگر این عدد تخمین زده شده مناسب بود، به عنوان کاندیدای داروی جدید در نظر گرفته می شود. با تخمین IC_{50} ترکیبات پیشنهادی، می توان در همان گام اول بدون صرف هزینه و زمان زیاد، یک غربالگری با دقت بالا انجام داد و ترکیباتی که شانس تبدیل شدن به دارو ندارند، حذف گردند.

یافته ها

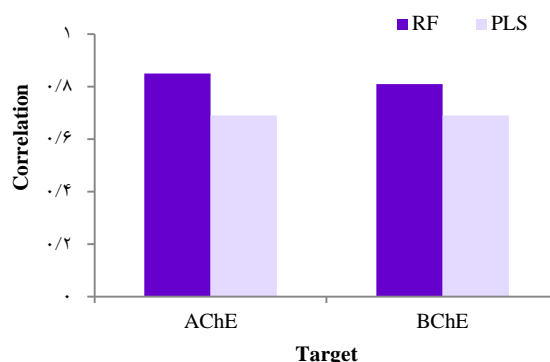
دو مدل پیشنهادی بر روی دو تارگت AChE و BChE پیاده شد و IC_{50} نمونه های آزمایش توسط این دو مدل تخمین زده شد. ضریب همبستگی (Correlation) بین فعالیت زیست شناختی تخمین زده شده توسط مدل ها و فعالیت زیست شناختی به دست آمده در آزمایشگاه و حاصل از سنتز برای هر ترکیب توسط تابع CORREL در نرم افزار Excel محاسبه گردید. به خاطر تصادفی انتخاب شدن مجموعه ی آموزش و آزمایش، هر مدل ۱۰ بار برای ۱۰ مجموعه ی آموزشی و آزمایشی تصادفی و متفاوت اجرا شد و در نهایت، میانگین این ۱۰ بار به عنوان خروجی مدل در نظر گرفته شد. این نتایج در جدول ۲ قابل مشاهده است.

جدول ۲. نتایج حاصل از اجرای مدل های خطی و غیر خطی

Target	Random forest	PLS
AChE	0.893 ± 0.026	0.714 ± 0.137
BChE	0.855 ± 0.054	0.717 ± 0.132

PLS: Partial least squares; AChE: Acetyl choline esterase; BChE: Butyl choline esterase

شکل ۱، نشان دهنده ی دقت عملکرد دو مدل در راستای تخمین IC_{50} می باشد. با توجه به مقدار همبستگی بالا بین مقدار IC_{50} تخمین زده شده توسط مدل و IC_{50} واقعی ترکیبات مورد بررسی، می توان ترکیب پیشنهادی جدید را ترسیم کرد و به عنوان نمونه ی آزمایش به مدل داد و مقدار فعالیت زیست شناختی را تخمین زد. سپس، با توجه به محدوده ی IC_{50} مورد نظر، فرد داروساز یا شیمی دارو، می توان آن ترکیب پیشنهادی را برای راه یابی به مرحله ی آزمایشگاهی تأیید یا رد کرد.



شکل ۱. عملکرد دو مدل پیشنهادی در راستای تخمین IC_{50}

بحث

تاکنون داروی قطعی برای درمان بیماری آلزایمر کشف نشده است، اما درمان اختلالات رفتاری و کشف داروهای کاهنده ی سیر پیشرفت

معاونت پژوهشی این دانشگاه به دلیل حمایت مالی از این مطالعه سپاسگزاری می‌گردد.

تشکر و قدردانی

این مقاله برگرفته از پایان‌نامه‌ی دوره‌ی کارشناسی ارشد به شماره‌ی ۳۹۷۲۰۵ در دانشگاه علوم پزشکی اصفهان می‌باشد. بدین وسیله از

References

1. Zeng H, Wu X. Alzheimer's disease drug development based on Computer-Aided Drug Design. *Eur J Med Chem* 2016; 121: 851-63.
2. Cheng ZQ, Zhu KK, Zhang J, Song JL, Muehlmann LA, Jiang CS, et al. Molecular-docking-guided design and synthesis of new IAA-tacrine hybrids as multifunctional AChE/BChE inhibitors. *Bioorg Chem* 2019; 83: 277-88.
3. Yan A, Wang K. Quantitative structure and bioactivity relationship study on human acetylcholinesterase inhibitors. *Bioorg Med Chem Lett* 2012; 22(9): 3336-42.
4. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimers Dement (NY)* 2019; 5: 272-93.
5. Velazquez-Libera JL, Caballero J, Toropova AP, Toropov AA. Estimation of 2D autocorrelation descriptors and 2D Monte Carlo descriptors as a tool to build up predictive models for acetylcholinesterase (AChE) inhibitory activity. *Chemometr Intell Lab Syst* 2019; 184: 14-21.
6. Hooda N, Bawa S, Rana PS. B2FSE framework for high dimensional imbalanced data: A case study for drug toxicity prediction. *Neurocomputing* 2018; 276: 31-41.
7. Okten S, Ekiz M, Kocyigit UM, Tutar A, Celik İ, Akkurt M, et al. Synthesis, characterization, crystal structures, theoretical calculations and biological evaluations of novel substituted tacrine derivatives as cholinesterase and carbonic anhydrase enzymes inhibitors. *J Mol Struct* 2019; 1175: 906-15.
8. Bicer A, Taslimi P, Yakali G, Gulcin I, Serdar GM, Turgut CG. Synthesis, characterization, crystal structure of novel bis-thiomethylcyclohexanone derivatives and their inhibitory properties against some metabolic enzymes. *Bioorg Chem* 2019; 82: 393-404.
9. Rezaei M, Bayrak C, Taslimi P, Gulcin I, Menzek A. The first synthesis and antioxidant and anticholinergic activities of 1-(4,5-dihydroxybenzyl)pyrrolidin-2-one derivative bromophenols including natural products. *Turk J Chem* 2018; 42: 808-25.
10. Kensert A, Alvarsson J, Norinder U, Spjuth O. Evaluating parameters for ligand-based modeling with random forest on sparse data sets. *J Cheminform* 2018; 10(1): 49.
11. Ranjan P, Athar M, Jha PC, Krishna KV. Probing the opportunities for designing anthelmintic leads by sub-structural topology-based QSAR modelling. *Mol Divers* 2018; 22(3): 669-83.

Improving the Biological Activity Prediction of Acetylcholinesterase and Butyl Cholinesterase Inhibitors Using Nonlinear Random Forest Algorithm

Fahimeh Motamedi¹, Alireza Mehridehnavi², Fahimeh Ghasemi³

Original Article

Abstract

Background: Due to the growing population of the elderly and the increasing trend of Alzheimer's disease, evaluation of acetylcholinesterase (AChE) and butyl cholinesterase (BChE) inhibitors, as major causes of Alzheimer's disease, is essential. Since the synthesis and investigation of each new compound is very costly and time-consuming, computational modeling techniques have been used to estimate biological activity. Up to now, various computational methods have been proposed which one of the major approaches, quantitative structure activity relationship, is based on the linear and non-linear methods using calculating the independent molecular descriptors. This study aimed to improve the biological activity prediction of AChE and BChE inhibitors using nonlinear random forest algorithm.

Methods: In order to predict the biological activity of AChE and BChE compounds, linear partial least squares and nonlinear random forest algorithms were used. To obtain more accurate and reliable results, 80% of the compounds were randomly used as a training sample, and the rest as a test sample, to construct the model and evaluate the predictive power of the model.

Findings: By applying nonlinear random forest model on AChE and BChE inhibitors, the accuracy of 89% was achieved. Finally, in order to examine more accurately the performance of the proposed model, the results were compared with the results obtained from the minimum partial error method, and the nonlinear random forest method had stronger performance than linear least squares method.

Conclusion: The observations indicated that the nonlinear random forest method could be very effective in predicting the biological activity of AChE and BChE compounds proposed by physicians and pharmaceutical chemists. Therefore, before animal and human testing of the proposed compound, the biological activity of the compound was estimated approximately to be 90%. Based on the estimated biological activity, it can be argued that a new drug combination, at the expense of time and money, has the potential of becoming a new drug or not.

Keywords: Quantitative structure-activity relationship, Acetylcholinesterase, Least squares

Citation: Motamedi F, Mehridehnavi A, Ghasemi F. **Improving the Biological Activity Prediction of Acetylcholinesterase and Butyl Cholinesterase Inhibitors Using Nonlinear Random Forest Algorithm.** J Isfahan Med Sch 2020; 37(558): 1401-6.

1- MSc Student, Student Research Committee, Department of Biomedical Engineering, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

2- Professor, Department of Biomedical Engineering, School of Advanced Technologies in Medicine AND Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

3- Assistant Professor, Department of Biomedical Engineering, School of Advanced Technologies in Medicine AND Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

Corresponding Author: Alireza Mehridehnavi; Professor, Department of Biomedical Engineering, School of Advanced Technologies in Medicine AND Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran; Email: mehri@med.mui.ac.ir