مجله علمی پژوهشی دانشگاه علوم پزشکی شهید صدوقی یزد
دوره سی و یک، شماره دهم، دی ۱۴۰۲
تاریخ دریافت:۱۴۰۲/۰۱/۲۹
تاریخ پذیرش:۱۴۰۲/۰۷/۰۹

شناسایی مهارکننده جدید گیرنده دوم فاکتور رشد اندوتلیال عروقی با استفاده از مدل طبقه‌بندی ماشین‌بردار پشتیبان

نوشین عربی[۱]، محمدرضا ترابی[۲]، افشین فصیحی[۳]، فهیمه قاسمی[*۲]

مقاله پژوهشی

مقدمه: امروزه با شیوع گسترده سرطان و افزایش مرگ و میر ناشی از آن، راههای موثر برای درمان سرطان از اهمیت بالایی برخوردار است. رگ‌زایی غیرطبیعی، یکی از ویژگی‌های مشترک انواع مختلف سرطان شناخته شده است. تا کنون مهار مسیر سیگنالینگ گیرنده دوم فاکتور رشد اندوتلیال عروق، به دلیل نقش پیش رگ‌زایی آن ببسیار مورد توجه قرار گرفته است. از اینرو، یافتن مدل‌های محاسباتی قابل اطمینان برای شناسایی مهارکننده‌ها می‌تواند در کاهش زمان و هزینه موثر باشد. هدف از این مطالعه به‌کارگیری روش ماشین‌بردار پشتیبان جهت طبقه‌بندی ترکیبات در دو گروه مهارکننده و غیرمهارکننده می‌باشد.

روش بررسی: به‌منظور پیاده‌سازی مدل یادگیری ماشین، لیگاندهای مورد مطالعه در این پژوهش از پایگاه داده https://www.bindingdb.org استخراج گردید و پس از گذراندن پیش پردازش‌های لازم برخی روش‌های انتخاب ویژگی مبتنی بر فیلتر و تعبیه شده مورد استفاده قرار گرفته شد. پس از استخراج توصیفگرها از داده‌ها، با استفاده از الگوریتم انتخاب ویژگی مبتنی بر همبستگی ابعاد داده کاهش یافته است تا بدین طریق از بیش برازش مدل جلوگیری شود. برای طبقه‌بندی از مدل ماشین‌بردار پشتیبان به همراه کرنل‌های Radial Basis Function (RBF)، Polynomial، Sigmoid و Linear استفاده شده است.

نتایج: پیاده‌سازی مدل ماشین‌بردار پشتیبان با کرنل RBF به همراه روش انتخاب ویژگی مبتنی بر همبستگی صحت بالاتری به میزان ۸۲/۴٪ (P= ۰/۰۰۸) (نسبت به سایر روش‌های انتخاب ویژگی بکار گرفته شده در این مطالعه به همراه داشته است.

نتیجه‌گیری: مشاهدات بیانگر آن است که روش انتخاب ویژگی مبتنی بر همبستگی، نسبت به سایر روش‌های به‌کار گرفته شده در این مطالعه از صحت بالاتری برخوردار است.

واژه‌های کلیدی: گیرنده دوم فاکتور رشد اندوتلیال عروق، رابطه کمی ساختار فعالیت، ماشین‌بردار پشتیبان، رگ‌زایی

ارجاع: عربی نوشین، ترابی محمدرضا، فصیحی افشین، قاسمی فهیمه. **شناسایی مهارکننده جدید گیرنده دوم فاکتور رشد اندوتلیال عروقی با استفاده از مدل طبقه‌بندی ماشین‌بردار پشتیبان.** مجله علمی پژوهشی دانشگاه علوم پزشکی شهید صدوقی یزد ۱۴۰۲؛ ۳۱ (۱۰): ۱۶-۷۱۰۸.

۱- گروه بیوالکتریک، دانشکده فناوری‌های نوین علوم پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
۲- گروه بیوانفورماتیک، دانشکده فناوری‌های نوین علوم پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
۳- گروه شیمی دارویی، دانشکده داروسازی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
*(نویسنده مسئول)؛ تلفن: ۰۹۱۳۱۰۷۵۹۷۵، پست الکترونیکی: f_ghasemi@amt.mui.ac.ir، صندوق پستی: ۸۱۷۴۷۳۴۶۱

[DOI: 10.18502/ssu.v31i10.14563]

مقدمه

سرطان یک مشکل بهداشت عمومی در سراسر جهان است. شیوع بالای مرگ و میر ناشی از سرطان، آن را به یکی از دلایل اصلی مرگ در جهان تبدیل کرده‌است، در نتیجه نیاز مبرمی به گسترش داروهای موثر برای درمان سرطان وجود دارد. یکی از عوامل مهم در گسترش سرطان‌هایی که منجر به مرگ می‌گردد متاستاز است. در واقع در طی فرآیند متاستاز سلول‌های سرطانی از طریق عروق خونی مهاجرت نموده و به سایر بافت‌ها وارد می‌شوند و در نهایت باعث درگیر شدن بافت‌های سالم بدن می‌شوند، از این‌رو اگر از لحاظ تئوری بتوان رگ‌زایی را مهار کنیم، تومورها در اندازه کوچک باقی می‌مانند و آسیب‌رسان نخواهند شد (۱) فاکتور کلیدی و مؤثر در تکثیر و مهاجرت سلول‌های اندوتلیال، که اساس تشکیل هر رگ جدیدی است، فاکتور رشد اندوتلیال عروقی (Vascular Endothelial Growth Factor) هست. این فاکتور به‌عنوان یک محرک اولیه برای رگ‌زایی عمل می‌کند. تولید VEGF و گیرنده‌اش به‌طور مستقیم میزان رگ‌زایی در تومور را کنترل می‌کنند، با توجه به اینکه VEGF نقطه استراتژیکی در تنظیم رگ‌زایی در تومور است، هدف مهمی در وقایع درمانی محسوب می‌گردد (۲). علاوه‌براین، رگ‌زایی لازمه رشد و تداوم تومورهای جامد و متاستازهای آن‌ها است و بیماری‌های التهابی شدید به مرحله بدخیم مرتبط با رگ‌زایی، پیشرفت می‌کنند (۳). VEGF که به دامنه گیرنده خارج سلولی متصل می‌شود، باعث فعال‌شدن آنزیم تیروزین‌کیناز در دامنه گیرنده داخل‌سلولی می شود که باقی‌مانده‌های تیروزین را فسفریله می‌کند و در نتیجه چندین مسیر سیگنال‌دهی داخل‌سلولی را فعال می‌کند (۴). در سال ۲۰۱۸ در مطالعه‌ای، به تأثیرگذاری نقش VEGF و گیرنده‌های آن در رگ‌زایی فیزیولوژیک و پاتولوژیک در پیشرفت تومور و ایجاد متاستاز تأکید شد و به بررسی نقش آن در تشخیص و درمان سرطان سلول کلیوی که نوع متداول مرگ‌های ناشی از سرطان کلیه در سراسر جهان است پرداخته شد. در این مطالعه، ۳۱۰۰۰۰ ترکیب پایگاه PDB(Protein Data Bank) با استفاده از روش غربالگری مجازی مبتنی بر ساختار مورد بررسی قرار گرفتند و سرانجام ۲۳ بازدارنده بالقوه شناسایی شدند و در نهایت ترکیب SCHEMBL469307 که پتانسیل بالایی برای مهار پروتئین VEGFR2 داشت، معرفی شد (۵). در سال ۲۰۲۰، مطالعه‌ای بر روی ۲/۴ میلیون مولکول پایگاه داده Zinc با استفاده از روش غربالگری مجازی مبتنی بر لیگاند باعث کشف چهارترکیب ۳، ۷، ۱۰ و ۱۳ شد که ترکیب ۱۰ فعالیت بازدارندگی مطلوبی را با مقدار IC_{50} ، ۱۹/۳میکرومولار برای مهار VEGFR2 به نمایش گذاشت (۶). در دسترس بودن ساختار سه بعدی پروتئین‌های هدف درمانی و اکتشاف حفره محل اتصال، اساس طراحی دارویی مبتنی بر ساختار، را تشکیل می‌دهد. این رویکرد اختصاصی است و به‌طور موثر، شناسایی مولکول‌های پیشرو و بهینه‌سازی آن‌ها را تسریع می‌کند که به درک بیماری در سطح مولکولی کمک کرده است. برخی از روش‌های رایج مورد استفاده در طراحی دارو بر اساس ساختار، شامل شبیه‌سازی‌های غربالگری مجازی مبتنی بر ساختار، داکینگ مولکولی و دینامیک مولکولی است (۷). زمانی‌که ساختار سه بعدی گیرنده هدف در دسترس نباشد از طراحی دارو مبتنی بر لیگاند استفاده می‌شود. اطلاعات به‌دست آمده از مجموعه‌ای از ترکیبات فعال در برابر یک گیرنده هدف خاص را می توان در شناسایی ویژگی‌های فیزیکوشیمیایی و ساختاری مسئول فعالیت بیولوژیکی معین استفاده کرد، که بر اساس این واقعیت است که شباهت‌های ساختاری با عملکرد بیولوژیکی مشابه مطابقت دارد. برخی از تکنیک‌های رایج مورد استفاده در رویکرد غربالگری مجازی مبتنی بر لیگاند شامل مدل‌سازی فارماکوفور، کمی‌سازی رابطه ساختار- فعالیت و هوش مصنوعی است (۸). با توجه به موارد مطرح شده در بخش قبل، به‌نظر می‌رسد یافتن ترکیبات جدید جهت مهار رگ‌زایی در سرطان با عوارض جانبی کمتر و اثربخشی بالاتر می‌تواند تاسیر بسزایی در درمان داشته باشد. بدین منظور، در این مطالعه از طراحی دارو مبتنی بر لیگاند استفاده شده است تا بتوان به بهترین ترکیبات مهارکننده گیرنده VEGF دست یافت. لازم به ذکر است که در بخش طراحی دارو مبتنی بر لیگاند از مدل‌سازی Quantitative

[DOI: 10.18502/ssu.v31i10.14563]

(structure-activity relationship QSAR) بهره گرفته شده است و برخی از روش‌های انتخاب ویژگی مبتنی بر فیلتر و تعبیه شده (Embedded) به منظور انتخاب بهترین ویژگی‌ها اتخاذ شده است. هم‌چنین از مدل طبقه‌بند ماشین‌بردار پشتیبان استفاده شده است زیرا روش‌های یادگیری ماشین مانند SVM معمولاً دارای قابلیت تعمیم پذیری بالا هستند، بدان معنا که مدل با توجه به داده‌های آموزشی، قادر است به درستی داده‌های جدید و ناشناخته را تشخیص دهد، SVM به خوبی در مواردی که تعداد نمونه‌ها نسبتاً کم است و ممکن است داده‌ها غیرخطی باشند، عملکرد خوبی دارد. با استفاده از توابع هسته‌ای، SVM می‌تواند ترکیبات غیرخطی را نیز مدل کند. این امکان بسیار مهم است زیرا بسیاری از مسائل واقعی دارای روابط غیرخطی هستند. معمولاً SVM مقاومت بالایی در برابر داده‌های نویزی دارد و قادر است با استفاده از همین امکان، داده‌های نویزی را از داده‌های واقعی تشخیص دهد. در برخی از حالات، مدل‌های SVM به راحتی قابل تفسیر و توجیه هستند، به این معنا که می‌توان فهمید که چه ویژگی‌هایی در تصمیم‌گیری مدل تاثیرگذار هستند هم‌چنین امکان کنترل دقیق‌تری بر فضای ویژگی‌ها را فراهم می‌کند که در مواردی که ویژگی‌های انتخاب شده برای مسئله مهم هستند می‌تواند بسیار مهم باشد.

روش‌بررسی

داده مورد نیاز در این پژوهش متشکل از گیرنده دوم فاکتور رشد اندوتلیال عروق (VEGFR2) می‌باشد، که مشتمل بر ۱۲۰۲۰ ترکیب است. برای استخراج لیگاندهای مورد مطالعه در این پژوهش از پایگاه اینترنتی https://www.bindingdb.org استفاده شده است.

پیش پردازش داده‌ها: به منظور پیش پردازش داده‌ها در ابتدا ترکیبات مورد مطالعه که به فرمت sdf می‌باشند، به صورت منحصر به فرد، متمایز می‌شوند تا مولکول‌هایی با ساختار مشابه و یکسان حذف شوند. در این مرحله به ۹۲۷۱ ترکیب منحصر به فرد دست می‌یابیم. پس از اصلاح ساختاری از ترکیبات به صورت سه بعدی استفاده می‌گردد. برای انجام این پیش پردازش‌ها از نرم افزار Openbabel بهره‌برده‌ایم.

استخراج ویژگی‌های ترکیبات با نرم افزار Dragon: کلیه ویژگی‌های پیشنهادی توسط نرم افزارDragon و با استفاده از تمام ۲۲ دسته انتخاب ویژگی آن صورت گرفته است. در این مطالعه با استفاده از نرم‌افزار۲۰۰۷ Dragon ۵.۵ تعداد ۳۲۲۴ ویژگی در اختیار قرار می‌گیرد. خروجی نرم افزارDragon یک ماتریس عددی به فرم زیر است:

رابطه (۱)

$$descriptors = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

در اینجا m تعداد مولکول‌ها و n تعداد ویژگی‌ها است.

استخراج میزان فعالیت هر لیگاند (IC_{50})

هدف این مطالعه تخمین مقادیر IC_{50} می‌باشد. با استفاده از برنامه‌نویسی پایتون میانگین IC_{50} ها مورد بررسی قرار گرفت، بدین صورت که اگر میانگین IC_{50} ها بالای یک میکرومولار باشد به عنوان غیرمهارکننده (۲۵۷۳) و اگر پایین‌تر از این آستانه باشد به عنوان مهارکننده (۵۸۹۶) در نظر گرفته شده است که این آستانه با توجه به مطالعات گذشته و سعی و خطاهای مربوط به مطالعات پیشین به‌دست آمده است. لازم به ذکر است که برخی از ترکیبات به علت نداشتن مقدار IC_{50} حذف گردیدند.

رابطه کمی ساختار-فعالیت: پیش‌پردازش داده‌ها یکی از مهم‌ترین مراحل پردازشی در مدل‌سازی QSAR محسوب می شود. در واقع، هنگام اعمال توصیف‌گرهای مولکولی به مدل آماری، اغلب محققان با حجم وسیعی از داده‌ها مواجه می‌شوند که باتوجه به محدودیت زمانی، انتخاب بهینه‌ترین توصیف‌گرها ضروری است. به عبارت دیگر، فرآیند تشخیص ویژگی‌های مفید و حذف ویژگی‌های تکراری به‌منظور بررسی مسئله با زیر مجموعه‌ای از ویژگی‌های مرتبط، در راستای عملکرد بهتر مدل، انتخاب ویژگی نامیده می‌شود. ضرورت انتخاب ویژگی را می-توان در کاهش پیچیدگی مدل، بهبود عملکرد روش، افزایش دقت پیش‌بینی و کاهش زمان محاسبه خلاصه نمود (۹). هدف اصلی انتخاب ویژگی، انتخاب زیرمجموعه‌ای از بیشترین

[DOI: 10.18502/ssu.v31i10.14563]

ویژگی‌های همبسته از مجموعه داده‌های بزرگ است که بتواند ویژگی اصلی همه ویژگی‌های مجموعه داده را توصیف و نشان دهد. فرآیند انتخاب ویژگی با استفاده از یک مجموعه داده اصلی شروع و انتخاب ویژگی‌های ضروری به صورت مرحله به مرحله انجام می‌شود(۱۰). در این پژوهش از روش‌های انتخاب ویژگی مبتنی بر فیلتر شامل روش‌های انتخاب ویژگی مبتنی بر همبستگی (CFS)، امتیاز فیشر (FS) و اطلاعات متقابل (MI) و هم‌چنین از روش‌های انتخاب ویژگی تعبیه شده شامل روش‌های LASSO و Elastic Net استفاده شده است. برای ساخت مدل طبقه‌بند، روش ماشین‌بردار پشتیبان بکار گرفته شده است. ماشین‌بردار پشتیبانی یکی از الگوریتم‌های یادگیری ماشینی تحت نظارت است.

نتایج

ارزیابی روش اجرا: به منظور ارزیابی مدل در ابتدا داده ها به دو گروه آموزش، ۷۰٪ داده‌ها، و آزمون، ۳۰٪ داده ها، تقسیم شدند. در ادامه به منظور بررسی عملکرد مدل در حین آموزش، از روش اعتبارسنجی متقابل k-fold که یک روش استاندارد برای تخمین عملکرد یک الگوریتم یا پیکربندی یادگیری ماشین بر روی یک مجموعه داده است، بر روی داده‌های آموزش استفاده شده است. در واقع k-fold Cross-Validation، ارزیابی مدل‌های یادگیری ماشین بر روی یک مجموعه داده با استفاده از اعتبارسنجی متقابل k-fold معمول است. روش اعتبارسنجی متقابل k-fold یک مجموعه داده محدود را به k دسته یکتا تقسیم می‌کند. به هر یک از k دسته ها فرصتی داده می‌شود که به عنوان یک مجموعه آزمایش عقب نگه داشته شده استفاده شود، در حالی‌که همه دسته‌های دیگر مجموعاً به‌عنوان یک مجموعه داده آموزشی استفاده می شوند. مجموع مدل‌های k متناسب می‌شوند و در مجموعه‌های آزمون نگهدارنده k ارزیابی می‌شوند و میانگین عملکرد گزارش می‌شود. دو پارامتر n-spilit و n-repeat برابر ۵ قرار داده شده است. هنگامی‌که به بهترین مدل‌ها از نظر دقت و سایر پارامترهای ارزیابی رسیدیم از ۳۰ درصد داده‌های باقی‌مانده، داده‌های آزمون، به عنوان ارزیابی نهایی استفاده می‌شود.

برای بررسی عملکرد مدل از چهار شاخص ارزیابی: حساسیت (Sensitivity)، اختصاصی بودن (Specificity)، صحت (Accuracy) و (Matthews Correlation Coefficient) MCC که با استفاده از معادلات (۱-۴)، محاسبه شده‌اند، استفاده گردیده است (۱۱).

رابطه(۱)

$$SE = \frac{TP}{TP + FN}$$

رابطه (۲)

$$SP = \frac{TN}{TN + FP}$$

رابطه (۳)

$$Q = \frac{TP + TN}{TP + TN + FP + FN}$$

رابطه(۴)

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}}$$

مثبت واقعی (True Positive) TP ، نشان‌دهنده تعداد ترکیبات بازدارنده مناسب است که به درستی به عنوان ترکیبات بازدارنده طبقه‌بندی شده‌اند.

منفی واقعی (True Negative) TN، نشان‌دهنده تعداد ترکیبات غیربازدارنده است که به درستی به عنوان ترکیبات غیربازدارنده طبقه‌بندی شده‌اند.

مثبت‌کاذب (False Positive)FP، نشان‌دهنده تعداد ترکیبات غیربازدارنده است که به اشتباه به عنوان ترکیبات بازدارنده طبقه‌بندی شده‌اند.

منفی‌کاذب(False Negative) FN، نشان‌دهنده تعداد ترکیبات بازدارنده است که به اشتباه در ترکیبات غیر بازدارنده طبقه‌بندی شده‌اند.

درحالیکه SE و SP به ترتیب نشان‌دهنده دقت پیش‌بینی بازدارنده‌ها و غیربازدارنده‌ها هستند، (Q) صحت پیش‌بینی کلی را برای همه ترکیبات موجود در مجموعه داده نشان می‌دهد. MCC، با مقادیر بین ۱- تا ۱، مهم‌ترین شاخص برای کیفیت طبقه‌بندی باینری است (۱۱). لازم به ذکر است که در این پژوهش موارد مثبت به ترکیبات بازدارنده و موارد منفی به ترکیبات غیر بازدارنده اشاره دارد.

[DOI: 10.18502/ssu.v31i10.14563]

روش لیگاند بنیان: به منظور تنظیمات اولیه مدل ماشین‌بردارپشتیبان و نیز انتخاب کرنل مناسب جهت پیاده‌سازی مدل، ابتدا مدل ماشین‌بردارپشتیبان با چهار کرنل مختلف آن اجرا و معیارهای ارزیابی مدل گزارش شده است. با توجه به مقادیر جدول ۱ که مدل ماشین‌بردار پشتیبان به همراه چهار کرنل مختلف آن بر روی تمامی ویژگی‌ها آماده شده است مشخص گردید که کرنل RBF برای تمام معیارهای گزارش شده دارای بالاترین صحت است بنابراین از کرنل RBF ماشین‌بردار پشتیبان به عنوان کرنل پیش‌فرض برای انجام مدل‌سازی استفاده شده است. حالی‌که کرنل RBF، جز بهترین کرنل مشخص شد از روش‌های انتخاب ویژگی مبتنی بر فیلترشامل روش‌های انتخاب ویژگی مبتنی بر همبستگی (CFS)، امتیاز فیشر، اطلاعات متقابل(MIFS)، بکار گرفته می‌شود. در روش انتخاب ویژگی مبتنی بر همبستگی(CFS)، ویژگی‌های دارای همبستگی بیشتر از ٪۹۵ به دلیل اثر یکسانی که بر روی مدل خواهند داشت حذف شده است. در سایر روش‌های انتخاب ویژگی شامل، امتیاز فیشر (FS) و اطلاعات متقابل (MIFS)، به هر ویژگی امتیازی اختصاص داده شده است که ویژگی‌هایی با امتیازهای بزرگتر از صفر و مثبت انتخاب گردیدند. با توجه به معیارهای گزارش شده در جدول ۲ مشخص می‌شود که ویژگی‌هایی که از روش انتخاب ویژگی مبتنی بر همبستگی (CFS) به‌دست آمده، نسبت به سایر روش‌های انتخاب ویژگی مبتنی بر فیلتر نتایج بهتری را دارا می‌باشد، لذا به نظر می‌رسد مدل‌سازی بر روی ویژگی‌های به دست آمده از روش انتخاب ویژگی مبتنی بر همبستگی (CFS)، نتایج بهتری را به همراه خواهد داشت. در گام بعد برای ارتقا مدل روی CFS-RBF از روش‌های مبتنی برتعبیه‌شده برای نهایی کردن مدل استفاده شده است. مدل ماشین‌بردار پشتیبان با کرنل RBF بر روی روش‌های انتخاب ویژگی مبتنی بر تعبیه شده نشان داد که روش‌های Elasctic Net و LASSO نتایج تقریبا مشابهی را دارا می‌باشند با این‌حال همان‌طور که در جدول ۳ ملاحظه می‌گردد روش Elasctic Net دارای مقدار حساسیت و صحت نسبتاً بالاتری می‌باشد.

جدول۱: پیاده‌سازی مدل روی کلیه ویژگی‌ها برای انتخاب بهترین کرنل ماشین‌بردار پشتیبان

نوع کرنل	صحت (٪)	ضریب همبستگی متیو (٪)	حساسیت (٪)	اختصاصی (٪)
RBF	۸۱/ ۸(P=۰/۰۰۸)	۵۷/۷(P=۰/۰۴۶)	۹۲/۳(P=۰/۰۱)	۸۲/۴(P=۰/۰۰۶)
Linear	۷۷/۷ (P=۰/۰۱۲)	۴۹/۷ (P=۰/۰۶۵)	۸۳/۷ (P=۰/۰۱۳)	۸۲/۸ (P=۰/۰۱)
Polynomial	۷۵/۸ (P=۰/۰۰۸)	۴۲/۴ (P=۰/۰۶۳)	۹۶ (P=۰/۰۰۴)	۷۴/۷ (P=۰/۰۰۶)
Sigmoid	۶۳/۲(P=۰/۰۰۹)	۱۶/۷ (P=۰/۰۵۳)	۷۳/۴(P=۰/۰۱۷)	۷۱/۹ (P=۰/۰۰۶)

جدول۲: پیاده‌سازی مدل ماشین‌بردار پشتیبان با کرنل RBF بر روی برخی از روش‌های انتخاب ویژگی مبتنی بر فیلتر

روش های انتخاب ویژگی مبتنی بر فیلتر	تعداد ویژگی‌ها	صحت (٪)	ضریب همبستگی متیو (٪)	حساسیت (٪)	اختصاصی (٪)
کرنل RBF بدون انتخاب ویژگی	۳۲۲۴	۸۱/۸ (P=۰/۰۰۸)	۵۷/۷ (P=۰/۰۴۶)	۹۲/۳ (P=۰/۰۱)	۸۲/۴(P=۰/۰۰۶)
انتخاب ویژگی مبتنی بر همبستگی	۱۴۳۱	۸۲/۴ (P=۰/۰۰۸)	۵۹/۱ (P=۰/۰۴۶)	۹۲/۲ (P=۰/۰۰۶)	۸۳/۱ (P=۰/۰۰۷)
اطلاعات متقابل	۱۸۵۹	۸۱/۵ (P=۰/۰۱۱)	۵۷ (P=۰/۰۶۳)	۹۱/۹ (P=۰/۰۱۶)	۸۲/۳ (P=۰/۰۰۸)
امتیاز فیشر	۱۱۱۷	۸۱/۲ (P=۰/۰۱)	۵۶/۲ (P=۰/۰۶۲)	۹۲/۹ (P=۰/۰۱۲)	۸۱/۸ (P=۰/۰۰۷)

[DOI: 10.18502/ssu.v31i10.14563]

جدول۳: پیاده‌سازی مدل ماشین‌بردار پشتیبان با کرنل RBF برروی روش‌های انتخاب ویژگی مبتنی بر تعبیه شده

روش انتخاب ویژگی مبتنی بر تعبیه	تعداد ویژگی‌ها	صحت(٪)	ضریب همبستگی متیو(٪)	حساسیت(٪)	اختصاصی(٪)
Elastic- Net	۲۰۰	۸۱/۶ (P=۰/۰۰۶)	۵۴ (P=۰/۰۴۳)	۹۳/۳(P=۰/۰۰۴)	۸۲/۵ (P=۰/۰۰۵)
Shrinkage and Least Absolute (LASSO) Selection Operator	۷۸	۸۰/۸ (P=۰/۰۰۶)	۵۲ (P=۰/۰۳۹)	۹۲/۴ (P=۰/۰۰۷)	۸۲/۲ (P=۰/۰۰۳)

بحث

همانگونه که در بخش‌های قبل به آن اشاره شد، هدف اصلی این مطالعه شناسایی ترکیبات جدید به عنوان مهارکننده رگ‌زایی در سرطان می‌باشد. برای این منظور روش لیگاند بنیان بر مبنای روش‌های یادگیری ماشین پیشنهاد شد. یکی از مشکلات مهم و رایج در این روش‌ها، تعداد زیاد توصیف‌گر در مقابل تعداد کم ترکیب است که غالبا منجر به مشکل بیش برازش در هنگام آموزش مدل می‌شود. برای جلوگیری از این مشکل روش استخراج و کاهش ویژگی پیشنهاد می‌شود. از آنجایی‌که روش طبقه‌بند ماشین‌بردار پشتیبان، SVM، یکی از روش‌های موفق در حوزه طراحی دارو بوده است، لذا در این مطالعه بر آن شدیم تا با به‌کارگیری روش‌های کاهش و استخراج بهینه ترین ویژگی‌ها به بهترین مدل برای شناسایی ترکیبات ضد رگ‌زایی دست پیدا کنیم. برای رسیدن به این هدف، ابتدا، تلاش بر آن شد تا بهترین مدل SVM شناسایی شود. برای به‌دست آوردن بهترین کرنل چهار هسته رایج به نام‌های RBF، Linear، Polynomial و Sigmoid مورد مطالعه قرار گرفت. در این میان کرنل RBF بهترین نتایج را داشت. در ادامه به بررسی روش‌های کاهش ویژگی مبتنی بر فیلترشامل روش‌های انتخاب ویژگی مبتنی بر همبستگی (CFS)، امتیاز فیشر، اطلاعات متقابل (MIFS) پرداخته شد. که در میان آن‌ها روش انتخاب ویژگی مبتنی بر همبستگی (CFS) بهترین نتیجه را به همراه داشت. در پایان برای ارتقا مدل، روش‌های مبتنی برتعبیه‌شده برای نهایی کردن مدل استفاده شد. در پایان روش Elastic- Net با دقت ۸۱/۶٪ بهترین روش پیشنهاد شد.

نتیجه‌گیری

امروزه نیاز به درمان‌های جدید برای بیماری‌های ناشی از سرطان در یک بازه زمانی کوتاه‌تر ضروری است. کشف و توسعه داروی سنتی شامل چندین مرحله برای کشف یک داروی جدید و کسب تاییدیه بازاریابی است. بنابراین کشف راهکارهای جدید برای کاهش بازه زمانی کشف دارو ضروری است. طراحی محاسباتی دارو که بر دو روش مبتنی بر لیگاند و مبتنی بر ساختار استوار است در جهت کاهش هزینه‌ها و زمان روشی موثر است. در این مطالعه گیرنده دوم فاکتور رشد اندوتلیال عروق که جزء گیرنده های مهم رگ‌زایی است و در بسیاری از سرطان‌ها نقش موثر ایفا می‌کند، به کمک مدل‌سازی QSAR که یکی از روش‌های مبتنی بر لیگاند است مورد بررسی قرار گرفت با استفاده از مدل ماشین‌بردار پشتیبان و به‌کارگیری روش‌های انتخاب ویژگی مبتنی بر فیلتر، نظیر روش انتخاب ویژگی مبتنی بر همبستگی، اطلاعات متقابل، امتیاز فیشر و برخی از روش‌های انتخاب ویژگی نظیر روش‌های Lasso، Elastic Net معیار های ارزیابی مدل نظیر صحت، ضریب همبستگی متیو، حساسیت و اختصاصی بودن گزارش گردید که از این میان، روش انتخاب ویژگی مبتنی بر همبستگی بالاترین صحت را به خود اختصاص داد و برتری این روش نسبت به سایر روش‌های به کارگرفته شده نشان داده شد. صحت کارهای انجام شده را می‌توان با روش‌های آزمایشگاهی هم مورد ارزیابی قرار داد که متاسفانه به علت هزینه بالای روش‌های آزمایشگاهی و زمان بر بودن فرایندها این مرحله صورت نپذیرفته است که از جمله محدودیت‌ها و کاستی‌های این پژوهش می‌باشد.

[DOI: 10.18502/ssu.v31i10.14563]

سپاس‌گزاری

این مقاله برگرفته از پایان‌نامه دوره کارشناسی ارشد به شماره ۳۴۰۰۵۹۴ در دانشگاه علوم پزشکی اصفهان می‌باشد. بدین وسیله از معاونت پژوهشی این دانشگاه به دلیل حمایت مالی از این مطالعه سپاسگزاری می‌گردد.

حامی مالی: معاونت آموزشی/ دانشگاه علوم پزشکی اصفهان

تعارض در منافع: وجود ندارند.

ملاحظات اخلاقی

کارگروه/ کمیته اخلاق در پژوهش معاونت تحقیقات و فناوری- دانشگاه علوم پزشکی اصفهان مصوب ۱۴۰۰/۰۸/۱۶ (کد اخلاق IR.MUI.RESEARCH.REC.1400.322)

مشارکت نویسندگان

نوشین عربی: تحقیق و بررسی، روش شناسی، نوشتن – پیش نویس اصلی

محمدرضا ترابی: نظارت بر داده‌ها، استفاده از نرم‌افزار، اعتبار سنجی، ویرایش متن

افشین فصیحی: تحلیل، نوشتن–ویرایش متن

فهیمه قاسمی: مفهوم سازی، نظارت، مدیریت پروژه، نوشتن – بررسی و ویرایش، جذب بودجه و همه نویسندگان در تدوین، ویرایش اولیه و نهایی مقاله و پاسخگویی به سوالات مرتبط با مقاله سهیم هستند.

References:

1-Li WW, Li VW, Hutnik M, Chiou AS. ***Tumor Angiogenesis as a Target for Dietary Cancer Prevention.*** J Oncol 2021 2012: 879623.

2-Farzaneh Behelgardi M, Zahri S, Gholami Shahvir Z, Mashayekhi F, Mirzanejad L, Asghari SM. ***Targeting Signaling Pathways of VEGFR1 and VEGFR2 as a Potential Target in the Treatment of Breast Cancer.*** Mol Biol Rep 2020; 47(3): 2061-71.

3-Melincovici CS, Boşca AB, Şuşman S, Mărginean M, Mihu C, Istrate M, et al. ***Vascular Endothelial Growth Factor (VEGF)-Key Factor in Normal and Pathological Angiogenesis.*** Rom J Morphol Embryol 2018; 59(2): 455-67.

4-Koch S, Tugues S, Li X, Gualandi L, Claesson-Welsh L. ***Signal Transduction by Vascular Endothelial Growth Factor Receptors.*** Biochem J 2011; 437(2): 169-83.

5-Sharma K, Patidar K, Ali MA, Patil P, Goud H, Hussain T, et al. ***Structure-Based Virtual Screening for the Identification of High Affinity Compounds as Potent VEGFR2 Inhibitors for the Treatment of Renal Cell Carcinoma.*** Curr Top Med Chem 2018; 18(25): 2174-85.

6-Selvam C, Mock CD, Mathew OP, Ranganna K, Thilagavathi R. ***Discovery of Vascular Endothelial Growth Factor Receptor-2 (VEGFR-2) Inhibitors by Ligand-Based Virtual High Throughput Screening.*** Mol Inform 2020; 39(7): e1900150.

7-Yu W, MacKerell AD Jr. ***Computer-Aided drug Design Methods.*** Methods Mol Biol 2017; 1520: 85-106.

8-Gurung AB, Ali MA, Lee J, Farah MA, Al-Anazi KM. ***An Updated Review of Computer-Aided drug Design and its Application to COVID-19.*** Biomed Res Int 2021; 2021: 8853056.

[DOI: 10.18502/ssu.v31i10.14563]

9-Masoomi Sefiddashti F, Asadpour S, Haddadi H, Ghanavati Nasab S. ***QSAR Analysis of Pyrimidine Derivatives as VEGFR-2 Receptor Inhibitors to Inhibit Cancer Using Multiple Linear Regression and Artificial Neural Network.*** Res Pharm Sci 2021; 16(6): 596-611.

10-Sheikhi S, Kheirabadi MT, Author C, Bazzazi A, Prof A. ***A Novel Scheme for Improving Accuracy of KNN Classification Algorithm Based on the New Weighting Technique And Stepwise Feature Selection.*** J Inf Technol Manag 2020; 12(4): 90-104.

11-Kang D, Pang X, Lian W, Xu L, Wang J, Jia H, et al. ***Discovery of VEGFR2 Inhibitors by Integrating Naïve Bayesian Classification, Molecular Docking and drug Screening Approaches.*** RSC Adv 8(10): 5286-5297.

[DOI: 10.18502/ssu.v31i10.14563]

Journal of Shahid Sadoughi University of Medical Sciences
Vol 31| NO 10 | Jan 2024
Received: 18 Apr 2023
Accepted: 1 Oct 2023

Finding New VEGFR2 Inhibitors Using Support Vector Machine Classification Model

Nooshin Arabi[1], Mohammad Reza Torabi [2], Afshin Fassihi[3], Fahimeh Ghasemi[†2]

Original Article

Introduction: In our current era, the prevalence of cancer and its associated mortality rates have become a pressing concern. As such, finding effective methods for treating cancer has become a matter of significant importance. Abnormal angiogenesis is one of the common characteristics of different types of cancer. So far, the inhibition of vascular endothelial growth factor receptor 2 signaling pathway has received much attention due to its pro-angiogenic role. Therefore, finding reliable computational models to identify inhibitors can be effective in reducing time and cost. The purpose of this study was to use the support vector machine method to classify compounds into two inhibitory and non-inhibitory groups.

Methods: In order to implement the machine learning model, the ligands studied in this research were extracted from the https://www.bindingdb.org database and after passing the necessary pre-processing, some filter-based and embedded feature selection methods were used. After extracting the descriptors from the data, using the feature selection algorithm based on correlation, the dimensions of the data have been reduced in order to avoid overfitting the model. The classification task utilized a support vector machine model, employing various kernels such as Radial Basis Function (RBF), Polynomial, Sigmoid, and Linear.

Results: The implementation of the support vector machine model with the RBF kernel along with the feature selection method based on correlation has resulted in a higher accuracy of 82.4% (P=0.008) compared to other feature selection methods used in this study.

Conclusion: Observations indicate that the correlation-based feature selection method is more accurate than other methods used in this study.

Keywords: Vascular Endothelial Growth Factor Receptor II, Quantitative Structure-Activity Relationship, Support Vector Machine, Angiogenesis.

Citation: Arabi N, Torabi M.R, Fassihi A, Ghasemi F. **Finding New VEGFR2 Inhibitors Using Support Vector Machine Classification Model** J Shahid Sadoughi Uni Med Sci 2023; 31(10): 7108-16.

[1]Department of Bioelectric, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.
[2]Department of Bioinformatics, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.
[3]Department of Medicinal Chemistry, Faculty of Pharmacy, Isfahan University of Medical Sciences, Isfahan, Iran.
***Corresponding author: Tel:** 09131075975**, email:** f_ghasemi@amt.mui.ac

[DOI: 10.18502/ssu.v31i10.14563]