



جمهوری اسلامی ایران

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

گواهی پذیرش مقاله  
CERTIFICATE



مرکز آموزش ملی مهندسی  
مسابقات خیار الکتریک

شماره گواهی: ۲۵۶ / م / ۹۳

بدینوسیله گواهی میشود مقاله پژوهشگر گرامی

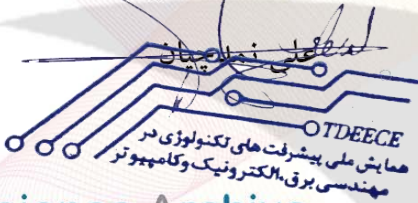
**مرضیه عبدی**

با عنوان

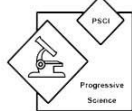
**الگوریتم جدید برای تشخیص آنتی بادی های موثر بر آنتی ژن HIV بر اساس ماشین بردار پشتیبان**

در اولین همایش ملی پیشرفت های تکنولوژی در مهندسی برق، الکترونیک و کامپیوتر که توسط مرکز آموزش عالی خیام الکتریک برگزار گردید مورد پذیرش قرار گرفته و توسط ایشان ارائه گردید. پیشرفت روزافزون ایشان را از خداوند متعال مسئلت داریم.

ریاست همایش



Science Archive



# الگوریتم جدید برای تشخیص آنتی‌بادی‌های موثر بر آنتی‌ژن HIV بر اساس ماشین بردار پشتیبان

مرضیه عبدی، مهدی سعادت‌مند طرزجان، علیرضا حق‌پرست، محمد طاهرزاده ثانی

دانشگاه فردوسی مشهد

مسئول مکاتبات: مرضیه عبدی

## چکیده

ایدز یا سندرم نقص ایمنی اکتسابی نوعی بیماری دستگاه ایمنی است که توسط ویروس نقص ایمنی (HIV) ایجاد می‌شود. این آنتی‌ژن دارای دو نوع HIV-1 و HIV-2 با ساختار ژنی/آنتی‌ژنی متفاوت می‌باشد. دفاع اختصاصی سیستم ایمنی در مقابله با آنتی‌ژن‌ها تولید آنتی‌بادی‌های اختصاصی است. آنتی‌بادی‌ها از نظر توالی اسیدآمینه‌ای (ساختار اول) با یکدیگر متفاوتند که این اختلاف ساختار باعث تفاوت در عملکرد و قدرت خنثی‌کنندگی آنها می‌شود. یکی از دلایل ضعف سیستم ایمنی در برابر آنتی‌ژن HIV، تاخیر آن در تولید آنتی‌بادی موثر و قابلیت چپش سریع ویروس می‌باشد. تزریق آنتی‌بادی‌های موثر بر HIV، در همان ساعات اولیه آلودگی، می‌تواند نقش بسزایی در کنترل و حذف عفونت داشته باشد. از این دیدگاه، تعیین ساختار آنتی‌بادی‌های موثر بر HIV نیازی ضروری است. در این مقاله، روشی جدید برای ارزیابی و تعیین آنتی‌بادی‌های موثر بر HIV با استفاده از ماشین بردار پشتیبان ارائه شده است. همچنین، ما با معرفی دو ویژگی جدید احتمال وقوع اسید آمینه در هم‌ترازی و همسایگی اسید آمینه‌ها در توالی، توانسته‌ایم عملکرد روش پیشنهادی را بهبود بخشیم. نتایج تجربی بیانگر عملکرد بسیار مناسب روش پیشنهادی (با حساسیت ۸۱/۸۰٪ و ویژگی ۹۹/۶۶٪، ارزش پیش‌بینی مثبت ۹۸/۲۹٪ و دقت ۹۶/۲۰٪) در مقایسه با دیگر الگوریتم‌های رقیب می‌باشد.

## کلمات کلیدی

آنتی‌بادی، HIV، ماشین بردار پشتیبان، ساختار اول آنتی‌بادی

## ۱- مقدمه

آنتی‌بادی‌ها، پروتئین‌های گردشی هستند که در پاسخ به آنتی‌ژن‌ها، توسط لنفوسیت‌های B تولید می‌شوند. این پروتئین‌ها که نقش واسطه‌های ایمنی را ایفا می‌کنند با اتصال به آنتی‌ژن و فعال‌سازی مکانیسم‌های اجرایی موجب حذف آن می‌گردند (ابوالعباس، ۱۳۹۱).

آنتی‌بادی‌ها نقش بسیار مهمی در کندشدن روند بسیاری از بیماری‌ها از جمله بیماری ایدز دارند، اما به خاطر ساختار متغیر ویروس (و تولید نسبی در هر هنگام آنتی‌بادی) قادر به حذف کامل آن نخواهند بود. بنابراین شناخت ساختار آنها و تحلیل آن، ما را در درمان HIV و تولید داروهای موثر کمک می‌کند.

## ۱-۱- ساختار آنتی‌بادی

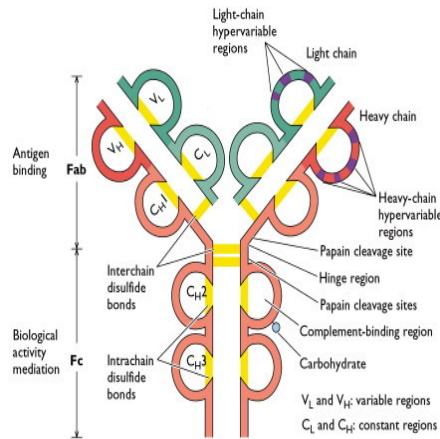
آنتی‌بادی‌ها مولکول‌های گلیکوپروتئینی هستند که دارای ساختاری قرینه شامل دو زنجیره سبک و دو زنجیره سنگین یکسان می‌باشند (دارای دو جایگاه اتصال به آنتی‌ژن). هر دو نوع زنجیره، از تعداد زیادی اسیدآمینه تشکیل شده است که هر اسیدآمینه می‌تواند یکی از بیست نوع اسیدآمینه موجود باشد (ابوالعباس، ۱۳۹۱).

زنجیره‌های سبک و سنگین، هر دو در انتهای آمینی دارای نواحی متغیر (V) و در انتهای کربوکسیلی خود دارای نواحی ثابت (C) بوده که به ترتیب در شناسایی آنتی‌ژن و فعالیت‌های اجرایی نقش دارند. این قطعه‌های متغیر، ناحیه اتصال به آنتی‌ژن را تشکیل می‌دهند که هر کدام از هفت ناحیه FR1، CDR1، FR2، CDR2، FR3، CDR3، FR4 تشکیل شده‌اند (شکل ۱ را ببینید).

همه مولکول‌های آنتی‌بادی، اساس ساختمانی یکسانی دارند؛ اما براساس محل اتصال به آنتی‌ژن دارای تنوع قابل ملاحظه‌ای می‌باشند که این باعث عملکرد متفاوت آنها می‌شود (ابوالعباس، ۱۳۹۱). با توجه به نقش نواحی متغیر در شناسایی آنتی‌ژن، در این تحقیق نیز تنها از نواحی متغیر زنجیره سبک و سنگین در ارزیابی ساختار موثر آنتی‌بادی استفاده می‌شود.

## ۲-۱- مروری بر تحقیقات گذشته

یکی از راههای شناسایی ساختارهای موثر آنتی‌بادی در اتصال به یک آنتی‌ژن خاص، ایجاد تغییرات در توالی آنتی‌بادی و تست آنتی‌بادی جدید با روش‌های معمول آزمایشگاهی است. از آنجا که این روش بسیار پرهزینه و وقت‌گیر می‌باشد؛ استفاده از روش‌های هوشمند می‌تواند برای ما راه‌گشا باشد.



شکل ۱: ساختار آنتی‌بادی

یکی از روش‌های متداول برای تعیین ساختار موثر آنتی‌بادی‌ها بهره‌گیری از ویژگی‌های فیزیکی - شیمیایی اسیدهای آمینه و فراوانی آنها در توالی می‌باشد. به عنوان مثال محمودی و همکارانش (۱۳۸۲، ۱۳۸۳) با بهره‌بردن از دو ویژگی آبدوستی و PH ایزوالکتریک، به دسته‌بندی اتوانتی‌بادی‌ها (در بیماری لوپوس) با استفاده از شبکه عصبی پرسپترون چند لایه پرداختند. این روش توانست به حساسیت و ویژگی ۷۱٪ دست یابد. در ارتباط با تعیین پاراتوپ‌ها<sup>۳</sup> نیز چندین وب‌سرور وجود دارد که براساس شباهت‌های ساختاری توالی ورودی با توالی‌های شناخته شده، به صورت آنلاین به پیش‌بینی نواحی اتصال می‌پردازد. به عنوان مثال، وب‌سرور Paratome (Kunik et al, 2012)، با همترازسازی ساختاری پروتئین‌ها و استفاده از شباهت‌های ساختاری به تعیین پاراتوپ می‌پردازد. همچنین وب‌سرور proABC (Olimpieri et al, 2013) تنها با استفاده از توالی و با بهره‌گیری از الگوریتم جنگل تصادفی به تعیین نواحی اتصال می‌پردازد. این وب‌سرور قادر به پیش‌بینی احتمال اتصال هر اسیدآمینه بر اساس نوع تعامل (تعاملات با تشکیل پیوند هیدروژنی<sup>۴</sup>، تعاملات آب‌گریز<sup>۵</sup>، تعاملات بدون اتصال<sup>۶</sup>) و نوع اتم‌های آن (زنجیره اصلی، زنجیره جانبی و یا هردو) نیز می‌باشد. دو وب‌سرور مذکور از همترازسازی ساختاری یا همترازسازی توالی برای پیش‌بینی بهره می‌برند. این در حالی است که به دلیل جهش‌های غیرهمبسته در کمپلکس‌های آنتی‌بادی - آنتی‌ژن (و در نتیجه، عدم دسترسی به اطلاعات تکاملی) استفاده از همترازی چندان قابل قبول نیست. لذا، در سال ۲۰۱۳، کرافچیک و همکارانش وب‌سرور Antibody i-Patch را معرفی نمودند. در این وب‌سرور بجای همترازی، از ساختار همولوژی آنتی‌بادی‌ها برای تعیین احتمال اتصال اسیدآمینه‌ها به آنتی‌ژن استفاده شده است.

علاوه بر روش‌هایی برای تعیین پاراتوپ، وب‌سرورهای متعددی هم در ارتباط با تعیین ساختار آنتی‌ژن‌ها (به صورت خاص تعیین اپی‌توپ‌های<sup>۷</sup> آن) ارائه شده است. به عنوان مثال می‌توان از وب‌سرورهای BcePred (Saha & Raghava, 2004) (پیش‌بینی براساس خصوصیات فیزیکی - شیمیایی اسیدآمینه مانند آبدوستی، انعطاف‌پذیری، میزان قطبی بودن، سطح در معرض و میزان دسترسی به حلال)، CEP (Kulkuranikale et al, 2005) (پیش‌بینی براساس میزان دسترس‌پذیری حلال و توزیع فضایی اسیدهای آمینه)، DiscoTope (Andersen et al, 2008) (پیش‌بینی براساس ساختار و نمرات گرایش اسیدهای آمینه)، ELLIPRO (Ponomarenko et al, 2008) (پیش‌بینی براساس ساختار فضایی)، EPITOPIA (Rubinstein et al, 2009) (پیش‌بینی براساس خصوصیات فیزیکی - شیمیایی و ویژگی‌های هندسی ساختار فضایی و با استفاده از طبقه‌بند بیز ساده)، EPCES (Liang et al, 2009) و EPSVR (Liang et al, 2010) (پیش‌بینی براساس شش خصوصیت باقی‌مانده شامل تعداد تماس، نمره حفاظت، نوع ساختار دوم، انرژی زنجیره جانبی، میزان تمایل به اپی‌توپ بودن و میزان مسطح بودن آن)، SEPPA (Qi et al, 2009)، EPMeta (Liang et al, 2010) و Bpredictor (Zhang et al, 2011) نام برد.

در اکثر روش‌های پیش‌بینی ساختار، آموزش طبقه‌بند با انواع مختلفی از آنتی‌بادی‌ها و آنتی‌ژن‌ها صورت می‌گیرد که این موجب پیچیده شدن فضای مدل‌سازی خواهد شد. این در حالی است که می‌توان تنها ساختار آنتی‌بادی‌های قابل تعامل با یک آنتی‌ژن خاص (یا آنتی‌ژن‌هایی با ساختار مشابه) مد نظر قرار گیرد. به این ترتیب، با ساده‌سازی فضای داده‌ها می‌توان به نتایج دقیق‌تری دست یافت. این ایده، اولین بار، در روش محمودی و همکارانش (۱۳۸۲، ۱۳۸۳)، برای ارزیابی ساختار اتوانتی‌بادی‌ها با استفاده از خواص فیزیکی - شیمیایی پیشنهاد شده است.

در این مقاله، از ماشین بردار پشتیبان<sup>۸</sup> (SVM) برای ارزیابی ساختار موثر آنتی‌بادی‌های قابل تعامل با ویروس HIV استفاده شده است. به این ترتیب، با توجه به ساده شدن فضای داده‌ها، می‌توان انتظار داشت که نتایج ارزیابی‌های مدل دقت بهتری در مقایسه با روش‌های رقیب داشته باشد. در واقع هدف ما ارزیابی عملکرد آنتی‌بادی‌ها در برابر آنتی‌ژن HIV، تنها با استفاده از ساختار اول آنها (یعنی توالی اسیدآمینه‌ها) است. برای این منظور، برای کدگذاری توالی اسیدهای آمینه در ساختار آنتی‌بادی، از سه ویژگی مختلف شامل آبدوستی، احتمال رخداد در هم‌ترازی و همسایگی بهره گرفته شده است که دو ویژگی آخر برای اولین بار در این مقاله معرفی گردیده‌اند.

در ادامه مقاله، بخش ۲ به معرفی پایگاه داده مورد استفاده برای آموزش و روش‌های مختلف الگوریتم پیشنهادی اختصاص یافته است. در بخش ۳، روش‌های ارزیابی عملکرد SVM معرفی گردیده است. همچنین، نتایج تجربی بدست آمده مورد بحث و بررسی قرار گرفته‌اند. در نهایت، بخش ۴ به نتیجه‌گیری می‌پردازد.

## ۲- مواد و روش‌ها

روش پیشنهادی شامل سه بخش پایگاه داده، کدگذاری و طبقه‌بندی می‌باشد که در ادامه، به تفصیل شرح داده خواهد شد.

### ۲-۱- پایگاه داده

هر دو پایگاه داده‌ی زنجیره‌ی سنگین (به اختصار، پایگاه داده سنگین) و زنجیره‌ی سبک (به اختصار، پایگاه داده سبک) از بانک اطلاعاتی Abysis<sup>۱</sup> استخراج شده‌اند. پایگاه داده سنگین شامل ۱۲۱۰ آنتی‌بادی موثر بر HIV (مجموعه داده +) و ۵۰۵۱ آنتی‌بادی غیرموثر بر HIV (مجموعه داده -) می‌باشد. به بیان دقیق‌تر، مجموعه داده منفی از ۸۴۲ آنتی‌بادی موثر بر بیماری‌های غیر ویروسی و ۴۲۰۹ آنتی‌بادی موثر بر بیماری‌های ویروسی تشکیل شده است. همچنین پایگاه داده سبک شامل ۱۹۹ آنتی‌بادی موثر بر HIV (مجموعه داده +) و ۲۰۰۷ آنتی‌بادی غیرموثر بر HIV (مجموعه داده -) می‌باشد. مجموعه منفی (-) در این پایگاه داده نیز از جمع‌آوری ۳۲۲ آنتی‌بادی موثر بر بیماری‌های غیر ویروسی و ۱۶۸۵ آنتی‌بادی موثر بر بیماری‌های ویروسی بدست آمده است.

با توجه به تعداد کم آنتی‌بادی‌هایی که دارای هر دو زنجیره سبک و سنگین هستند در پایگاه داده، آموزش و تست SVM برای هر زنجیره به صورت جداگانه انجام می‌شود. برای مشاهده جزئیات بیشتر درباره ترکیب پایگاه‌های داده سنگین و سبک و تعداد نمونه‌های آموزشی و تست (به تفکیک بیماری) به جدول شماره ۱ مراجعه گردد.

### ۲-۲- کدگذاری توالی

قبل از کدکردن زنجیره‌های سبک و سنگین آنتی‌بادی‌ها، باید تمامی زنجیره‌های سبک یا سنگین (به صورت جداگانه) هم‌طول شوند. برای این منظور، ابتدا هفت ناحیه مربوط به قطعه‌های متغیر (بخش ۱-۱ را ببینید) در هر زنجیره مشخص شده و سپس، با اضافه کردن صفر، نواحی مشابه در همه توالی‌ها هم‌طول می‌گردند. به این ترتیب، طول زنجیره‌های سبک و سنگین به ترتیب به ۱۳۵ و ۱۶۹ افزایش خواهد یافت. در این مقاله، برای کدگذاری توالی اسیدآمینها در هر زنجیره از ۳ ویژگی مختلف شامل آبدوستی، احتمال رخداد در هم‌ترازی و همسایگی استفاده شده است. از جمله ویژگی‌های فیزیکی و شیمیایی اسیدهای آمینه می‌توان به آبدوستی، PH، ایزوالکتریک، میزان قطبش، مساحت سطح و میزان یونی بودن اشاره کرد که بهترین تفکیک‌پذیری با ویژگی آبدوستی بدست آمد. بنابراین، در این مقاله (از میان همه ویژگی‌های فیزیکی - شیمیایی)، تنها از آبدوستی برای کدگذاری توالی استفاده می‌گردد. برای محاسبه ویژگی آبدوستی (برای هر اسیدآمین) از روش ارائه شده توسط محمودی و همکارانش (۱۳۸۳، ۱۳۸۲) استفاده شد.

برای محاسبه ویژگی (جدید) احتمال رخداد در هم‌ترازی، ابتدا در مجموعه آنتی‌بادی‌های موثر بر HIV (مجموعه مثبت پایگاه داده)، احتمال رخداد هر یک از اسیدآمین‌های را در موقعیت‌های مختلف زنجیره بدست می‌آوریم. سپس، برای کدگذاری یک زنجیره، به هر اسید آمینه در توالی، احتمال وقوع اسید آمینه مذکور در آن موقعیت (هم‌ترازی) را نسبت می‌دهیم.

در نهایت، در ویژگی همسایگی از یک پنجره لغزان (با طول  $n$ ) برای کدگذاری الگوی مجاورت اسیدآمین‌ها در توالی استفاده شده است. به این ترتیب که مرکز پنجره بر روی اولین اسیدآمین در توالی قرار گرفته و عددی مشخص به الگوی همسایگی درون پنجره نسبت داده می‌شود. سپس، پنجره به سمت جلو لغزانده می‌شود تا مرکز آن بر روی اسیدآمین بعدی قرار گیرد. فرآیند فوق تا انتهای زنجیره تکرار می‌شود تا ویژگی همسایگی برای همه اسیدآمین‌های توالی محاسبه گردد.

### ۲-۳- طبقه‌بندی با ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری بر مبنای تئوری یادگیری آماری Vapnik می‌باشد. مزیت این روش هوشمند بر شبکه‌های عصبی، بهره‌گیری از اصول کمینه‌سازی ریسک ساختاری<sup>۱۱</sup> (SRM) است (Dibike et al, 2001). از مهمترین برتری‌های ماشین بردار پشتیبان در مقایسه با شبکه‌های عصبی می‌توان به آموزش آسان و قابلیت تعمیم‌پذیری بهتر اشاره نمود. بعلاوه، در شرایطی که تعداد نمونه‌های مجموعه آموزشی کم باشد یا تعداد نمونه‌های دو کلاس بسیار متفاوت باشند، ماشین بردار پشتیبان عملکرد بهتری خواهد داشت. لذا، در این مقاله برای تعیین آنتی‌بادی‌های موثر در درمان HIV، از ماشین بردار پشتیبان استفاده می‌شود. انتظار می‌رود که ماشین بردار پشتیبان با آموزش توالی (اسیدآمین‌های) آنتی‌بادی‌های موثر بر HIV (کلاس +) و آنتی‌بادی‌های غیرموثر بر HIV (کلاس -) قادر به ارزیابی تاثیر سایر آنتی‌بادی‌ها بر HIV باشد. به این ترتیب که خروجی SVM باید بیانگر میزان تعلق آنتی‌بادی موردنظر به هر کدام از کلاس‌های + و - باشد. بعلاوه، با توجه به حضور نواحی متغیر (موثر بر HIV) هم در زنجیره سبک و هم در زنجیره سنگین توالی، برای هر زنجیره یک طبقه‌بندی جداگانه در نظر گرفته می‌شود.

در این مقاله، از کرنل RBF با  $\sigma=40$  برای ماشین‌های بردار پشتیبان هر دو زنجیره استفاده شده است. همچنین، در SVM‌های زنجیره‌های سبک و سنگین به ترتیب مقادیر بهینه  $C=1$  و  $C=300$  انتخاب شده‌اند. ورودی طبقه‌بندها نیز می‌تواند هر یک از بردارهای ویژگی معرفی شده در بخش ۲-۱ یا ترکیبی از آنها باشد (در بخش بعد در این‌باره بیشتر بحث خواهد شد).

### ۳- نتایج و بحث

برای ارزیابی عملکرد روش پیشنهادی از معیارهای ارزیابی حساسیت (sensitivity)، ویژگی (specificity)، ارزش پیش‌بینی مثبت (PPV)، دقت (accuracy) استفاده گردیده است (Ponomarenko et al, 2007).

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

$$PPV = \frac{TP}{TP + FP} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

که TP تعداد مثبت حقیقی (تعداد آنتی‌بادی‌های کلاس مثبت که به درستی پیش‌بینی شده‌اند)، FP تعداد مثبت کاذب (تعداد آنتی‌بادی‌های کلاس منفی که در کلاس مثبت دسته‌بندی شده‌اند)، TN تعداد منفی حقیقی (تعداد آنتی‌بادی‌های کلاس منفی که به درستی پیش‌بینی شده‌اند) و FN تعداد منفی کاذب (تعداد آنتی‌بادی‌های کلاس مثبت که در کلاس منفی دسته‌بندی شده‌اند) است.

برای تعیین بهترین ترکیب بردار ویژگی، طبقه‌بندهای زنجیره‌های سبک و سنگین (به اختصار، طبقه‌بندهای سبک و سنگین) با بردارهای ویژگی مختلف آموزش داده شدند. نتایج ارزیابی‌ها در جدول ۲ گزارش شده است. همانطور که مشاهده می‌شود، در مجموع، طبقه‌بند سنگین در مقایسه با طبقه‌بند سبک عملکرد بهتری داشته است. بنابراین، می‌توان نتیجه گرفت که که نواحی متغیر زنجیره سنگین در مقایسه با زنجیره سبک، نقش بیشتری در تعامل با HIV ایفا می‌کنند.

در نگاهی دقیقتر به نتایج طبقه‌بند سبک مشاهده می‌شود که ارزش پیش‌بینی مثبت دارای مقادیر نسبتاً پایینی می‌باشد که بیانگر گرایش بیشتر طبقه‌بند به سمت کلاس مثبت (در مقایسه با کلاس منفی) است. همچنین، خطای کمتر این طبقه‌بند در پیش‌بینی نمونه‌های کلاس منفی باعث افزایش مقدار معیار ویژگی در مقایسه با حساسیت شده است. در نهایت، بهترین نتایج برای طبقه‌بند سبک با استفاده از ترکیب هر سه ویژگی آبدوستی، احتمال و همسایگی با حساسیت ۷۵/۸۱٪، ویژگی ۹۳/۷۳٪، ارزش پیش‌بینی مثبت ۵۴/۶۵٪ و دقت ۹۲/۱۰٪ بدست آمده است.

در طبقه‌بند سنگین، مقدار معیار ویژگی بیش از ۹۹٪ می‌باشد که بیانگر خطای ناچیز طبقه‌بند در پیش‌بینی مجموعه منفی است. همچنین در این طبقه‌بند، معیار ارزش پیش‌بینی مثبت نسبت به حساسیت به مقادیر بزرگتری دست یافته است. لذا استنباط می‌شود که پیش‌بینی‌های طبقه‌بند گرایش به سمت نمونه‌های منفی دارد (واضح است که خطای کلاس مثبت به خاطر پیچیده بودن توالی‌های آن بیشتر بوده است). در نهایت، طبقه‌بند سنگین نیز با بهره‌گیری از هر سه ویژگی آبدوستی، احتمال و همسایگی بهترین نتایج را با حساسیت ۸۱/۸۰٪، ویژگی ۹۹/۶۶٪، ارزش پیش‌بینی مثبت ۹۸/۲۹٪ و دقت ۹۶/۲۰٪ ارائه داده است. واضح است که استفاده از ویژگی‌های جدید احتمال و همسایگی در روش پیشنهادی در کنار ویژگی آبدوستی موجب بهبود چشمگیر نتایج شده است. به عبارت دیگر، ویژگی‌های پیشنهادی احتمال (در هم‌ترازی) و همسایگی توانسته‌اند اطلاعات موثری از توالی اسیدهای آمینه در زنجیره‌های سبک و سنگین در اختیار طبقه‌بند قرار دهند.

در جدول ۳، عملکرد روش پیشنهادی با نتایج وب‌سرورها و همچنین روش ارائه شده توسط محمودی و همکارانش (۱۳۸۳، ۱۳۸۲) مقایسه شده است. همانطور که مشاهده می‌شود، روش پیشنهادی (برای ارزیابی عملکرد آنتی‌بادی‌ها در برابر آنتی‌ژن HIV) عملکرد بسیار بهتری در مقایسه با دیگر روش‌های رقیب داشته است. شاید بتوان گفت که یک دلیل عمده برای این اختلاف فاحش را در تخصصی بودن روش پیشنهادی (برای شناسایی آنتی‌بادی‌های موثر بر آنتی‌ژن HIV) و عمومی بودن الگوریتم‌های رقیب (برای شناسایی سایت‌های تعامل آنتی‌بادی - آنتی‌ژن) دانست. علاوه بر روش پیشنهادی در مقایسه با روش هوشمند و تخصصی ارائه شده برای پیش‌بینی آنتی‌بادی‌های موثر بر بیماری لوپوس (محمودی و همکارانش ۱۳۸۲، ۱۳۸۳) نیز از دقت بالاتری برخوردار است.

جدول ۱: ترکیب داده‌ها در پایگاه‌های داده سبک و سنگین به تفکیک نوع بیماری و تعداد نمونه‌های مورد استفاده در آموزش و تست طبقه‌بندها

مجموعه		مجموعه داده منفی		مجموعه مثبت	
بیماری‌ها	ایدز	بیماری‌هایی با ویروس مشابه با HIV	بیماری‌هایی با ویروس غیر مشابه HIV	بیماری‌های غیر ویروسی	بیماری‌های غیر ویروسی
تعداد: زنجیره -	مجموعه آموزش	۷۷۱	۴۰۶	۲۰۸	۲۰۸
های سبک	مجموعه تست	۳۱۵	۱۹۳	۱۱۴	۱۱۴
تعداد: زنجیره -	مجموعه آموزش	۲۳۴۰	۳۸۰	۵۶۸	۵۶۸
های سنگین	مجموعه تست	۱۲۵۶	۲۰۳	۳۰۴	۳۰۴

جدول ۱: ارزیابی عملکرد طبقه‌بندهای سبک و سنگین پیشنهادی، به تفکیک ویژگی، در پیش‌بینی آنتی‌بادی‌های موثر بر HIV.

بهترین نتایج با فونت ضخیم مشخص شده‌اند.

روش‌ها	ویژگی‌های بکاربرده شده برای کدگذاری اطلاعات	حساسیت	ویژگی	ارزش پیش‌بینی مثبت	دقت
طبقه‌بند ۱	همسایگی	٪ ۸۳/۶۹	٪ ۹۹/۳۸	٪ ۹۶/۹۹	٪ ۹۶/۳۴
طبقه‌بند ۲	آبدوستی (محمودی و همکارانش ۱۳۸۲، ۱۳۸۳)	٪ ۸۳/۹۲	٪ ۹۹/۵۵	٪ ۹۷/۸۰	٪ ۹۶/۵۲
طبقه‌بند ۳	آبدوستی + همسایگی	٪ ۸۰/۶۴	۹۲/۶۰	٪ ۵۲/۰۸	٪ ۹۱/۵۲
طبقه‌بند ۴	احتمال	٪ ۸۴/۱۶	٪ ۹۸/۸۶	٪ ۹۴/۶۸	٪ ۹۶/۰۲
طبقه‌بند ۵	احتمال + همسایگی	٪ ۸۲/۹۸	٪ ۹۹/۲۱	٪ ۹۶/۱۶	٪ ۹۶/۰۶
طبقه‌بند ۶	احتمال + آبدوستی	٪ ۷۰/۹۷	٪ ۹۱/۶۴	٪ ۴۵/۸۳	٪ ۸۹/۷۷
طبقه‌بند ۷	احتمال + همسایگی + آبدوستی	٪ ۸۱/۸۰	٪ ۹۹/۶۶	٪ ۹۸/۲۹	٪ ۹۶/۲۰
		٪ ۷۵/۸۱	٪ ۹۳/۷۳	٪ ۵۴/۶۵	٪ ۹۲/۱۰

جدول ۳: مقایسه عملکرد روش پیشنهادی با روش طاهرزاده و همکاران [۲-۳] و وب سرورهای Promate [۴]، proABC [۵] و Antibody i-Patch [۶] برحسب معیارهای ارزیابی حساسیت، ویژگی، ارزش پیش‌بینی مثبت و دقت (بهترین نتایج با فونت ضخیم مشخص شده‌اند).

روش‌ها	حساسیت	ویژگی	ارزش پیش‌بینی مثبت	دقت
روش پیشنهادی	٪ ۸۱/۸۰	٪ ۹۹/۴۳	٪ ۹۷/۱۹	٪ ۹۶/۰۲
اتوانتی‌بادی‌ها (محمودی و همکارانش ۱۳۸۲، ۱۳۸۳)	٪ ۷۱	٪ ۷۱	-	-
Promate (Kunik et al, 2012)	٪ ۱۰۰	٪ ۱۱	٪ ۳۶	٪ ۴۱
proABC (Olimpieri et al, 2013)	٪ ۸۰	٪ ۸۲	٪ ۶۹	٪ ۸۱
Antibody i-Patch (Krawczyk, et al, 2013)	٪ ۹۴	-	٪ ۴۰	-

#### ۴- نتیجه‌گیری

در این مقاله روشی جدید برای تعیین ساختار آنتی‌بادی‌های موثر بر آنتی‌ژن HIV ارائه شد. ما به جای آموزش طبقه‌بند با انواع مختلفی از ترکیب‌های آنتی‌بادی - آنتی‌ژن، تنها ساختار آنتی‌بادی‌های قابل تعامل با آنتی‌ژن مورد نظر (یا چند آنتی‌ژن با ساختار مشابه) را به طبقه‌بند آموزش دادیم. به این ترتیب، با توجه به ساده شدن فضای مدل‌سازی، می‌توان انتظار داشت که دقت پیش‌بینی‌های طبقه‌بند افزایش یابد. همچنین، ما در آموزش طبقه‌بندهای سبک و سنگین توانستیم با موفقیت از دو ویژگی جدید احتمال (در هم‌ترازی) و همسایگی در کنار ویژگی آبدوستی بهره‌برده و معیارهای ارزیابی روش پیشنهادی را بهبود بخشیم. علاوه بر این، نشان داده شد که زنجیره سنگین در مقایسه با زنجیره سبک، نقش بیشتری در تعامل با آنتی‌ژن HIV ایفا می‌کند. نتایج تجربی بیانگر عملکرد بسیار بهتر روش پیشنهادی در مقایسه با دیگر الگوریتم‌های رقیب برای پیش‌بینی آنتی‌بادی‌های موثر بر HIV بوده است.

#### منابع مورد استفاده:

- [۱] عباس، ابول. ک. ۲۰۱۲؛ ایمونولوژی سلولی و مولکولی، ترجمه حسینی، ر. فرید، انتشارات ابن سینا و مجد دانش ویرایش هفتم
- [۲] طاهرزاده ثانی، م، میرصالحی، م، خادمی، م، محمودی، م، نقیسی سیستانی، م.، عازمی، الف، رملستان ۸۲؛ تعیین ساختارهای موثر در اتصال آنتی‌بادی‌ها به DNA با استفاده از شبکه مصنوعی، مجله علوم پزشکی ایران، جلد ۶، شماره ۴

[۳] طاهرزاده ثانی، م، میرصالحی، م، خادمی، م، محمودی، م، نقیبی سیستانی، م، عازمی، الف، پاییز ۸۳؛ کاربرد الگوریتم ژنتیک و پرسپترون چندلایه در پیشگویی ساختارهای موثر در اتصال آنتی‌بادی، انتشارات امیرکبیر، سال پانزدهم

- [4] Kunik, V. et al 2012. Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. *Nucleic Acids Res.* 40:521-524,
- [5] Olimpieri. P. P. et al 2013. Prediction of site-specific interactions in antibody-antigen complexes: the proABC method and server. *Bioinformatics.* 29:18:2285–2291
- [6] Krawczyk. K. et al 2013. Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking. *Protein Engineering. Design & Selection.* 26:10:621–629
- [7] Saha S, Raghava GPS. BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. *Computer Science 2004 3239:4:197-204*
- [8] Kulkuranikale U, Bhoale S, Kolaskar AS. CEP: a conformational epitope prediction server. *Nucleic Acids Research* 2005; 33:168-171
- [9] Andersen. PH, Nilsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Science* 2008; 15:11:2558-2567
- [10] Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 2008; 9:514
- [11] Rubinstein. ND, Mayrose I, Martz E, Pupko T, Epiptopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics* 2009;10:287
- [12] Liang S, Zheng D, Zhang C, Zacharias M. Prediction of antigenic epitopes on protein surfaces by consensus scoring. *BMC Bioinformatics* 2009; 10:302
- [13] Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010;11:381
- [14] Sun. J, Wu D, Xu T, Wang X, Xu X, Tao L, Li YX, Cao ZW. SEPPA: a computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Research* 2009;37:612-616,
- [15] Zhang W, Xiong Y, Zhao M, Zou H, Ye X, Liu J. Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. *BMC Bioinformatics* 2011;12:341
- [16] Dibike. Y. B, Velikov. S, Solomatine. D, Abbot. M. B 2001. Model induction with support vector machines-introduction and applications. *Journal of Computing in Civil Engineering.* 15:3:208-216
- [17] Ponomarenko. J. V, Bourne. P. E 2007. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Structural Biology.* 7:64-83

## زیر نویس ها

<sup>1</sup> Variable regions

<sup>2</sup> Constant regions

<sup>4</sup> Hydrogen bond

<sup>5</sup> Hydrophobic

<sup>6</sup> Non-bonded interactions

<sup>8</sup> Support vector machine (SVM)

<sup>9</sup><http://www.bioinf.org.uk/abysis/>

<sup>10</sup>Structural risk minimization

<sup>۲</sup> نواحی از آنتی‌بادی که به آنتی‌ژن متصل و باعث خنثی سازی آن می‌شود.

<sup>۷</sup> نواحی از آنتی‌ژن که توسط آنتی‌بادی شناسایی می‌شود.