

پیش‌بینی مقاومت دارویی ویروس HIV با استفاده از ماشین‌های بردار پشتیبان

سمیرا عباسی^۱، دکتر مجید محمدبیگی^۲

چکیده

مقدمه: پژوهش حاضر با هدف بررسی عملکرد روش یادگیری ماشین‌های بردار پشتیبان برای پیش‌بینی مقاومت دارویی ویروس HIV یا (Human immunodeficiency virus) استفاده از تحلیل دنباله‌ی آمینو اسیدها انجام گرفت.

روش‌ها: در این مطالعه از ماشین‌های بردار پشتیبان استفاده شد؛ همچنین برای آموزش و آزمایش ماشین‌های بردار پشتیبان از نرم‌افزار (LIBSVM یا Library for support vector machines) استفاده گردید.

یافته‌ها: نتایج حاصل از مطالعه‌ی حاضر بر روی چهار روش بیولوژیکی که برای تفسیر نتایج تست‌های ژنوتیپی ارایه گردیده بود، بررسی شد. با مقایسه‌ی نتایج، بهترین روش بیولوژیکی برای هر دارو مشخص گردید. ماشین‌های بردار پشتیبان، به عنوان یک جداساز بسیار قوی برای تعیین مقاومت دارویی ویروس HIV با دقت ۹۸/۷۷-۸۶/۲۷ عمل می‌کند. با توجه به پارامترهای اندازه‌گیری شده، اعمال ماشین‌های بردار پشتیبان بر روی نتایج روش Abacavir, (NF) Nelfinavir, (APV) Amprenavir برای داروهای (HIVDB) HIV RT and protease sequence database, (TDF) Tenofovir disoproxil fumarate, (DDI) Didanosine, (D4T) Stavudine, (AZT) Zidovudine, (ABC) Delavirdine (DLV) و بر روی نتایج روش Agence national de recherches le SIDA (ANRS)، برای داروهای (NVP) Nevirapine و (EFV) Efavirenz, (TDF) Tenofovir disoproxil fumarate, (3TC) Lamivudine, (IDV) Indinavir بهترین نتایج را داشته است. برای داروهای (LPV) Lopinavir و (AZT) Zidovudine ماشین‌های بردار پشتیبان برای نتایج روش (REGA) Rega institute عملکرد خوبی ارایه نموده است و نیز برای داروهای (LPV) Lopinavir, (RTV) Ritonavir, (SQV) Saquinavir و (VGI) Visible genetics (VGI) عملکرد خوبی داشته است.

نتیجه‌گیری: ماشین‌های بردار پشتیبان، به عنوان یک جداساز بسیار قوی برای تعیین مقاومت دارویی ویروس HIV عمل می‌کند. می‌توان هنگام استفاده از هر دارو برای درمان، قبل از شروع درمان ابتدا مقاومت دارویی را با استفاده از روش‌های یادگیری ماشین، تعیین کرد.

واژگان کلیدی: ایدز، ماشین بردار پشتیبان، مقاومت دارویی

مقدمه

Human immunodeficiency virus (HIV) یکی از علل اصلی مرگ و میر در جهان است، این بیماری انواع خاصی از لمفوسیت‌ها را که سلول‌های کمک کننده‌ی T نامیده می‌شوند و در سیستم ایمنی بدن بسیار مهم هستند، آلوده می‌کند. بدون تعداد کافی از این سلول‌ها، سیستم ایمنی بدن قادر به دفاع در برابر عفونت‌ها و امراض نیست (۱).

مقاومت دارویی ویروس HIV یک فاکتور مهم در

شکست درمان‌های بیماری HIV است. عامل ایجاد مقاومت دارویی جهش در ژنوم HIV است که در نتیجه‌ی آن، پروتئین‌ها و ساختار ویروس عوض می‌شوند و داروهایی که در گذشته روی این ویروس اثر داشته است، نمی‌تواند آن را مهار کند. از این رو، پیش‌بینی مقاومت دارویی HIV می‌تواند در بهبود درمان مؤثرتر و بادوام‌تر بیماری اثربخش باشد.

وقتی ویروس HIV وارد بدن انسان می‌شود، در نرخ بالایی حدود میلیون‌ها در هر روز تکثیر می‌شود.

^۱ دانشجوی کارشناسی ارشد، گروه مهندسی پزشکی، دانشکده فنی و مهندسی، دانشگاه اصفهان، اصفهان، ایران.

^۲ استادیار، گروه مهندسی پزشکی، دانشکده فنی و مهندسی، دانشگاه اصفهان، اصفهان، ایران

مهار می‌کنند (۲-۳).

مقاومت دارویی می‌تواند با استفاده از دو تست بیولوژیکی اندازه‌گیری شود: تست‌های فنوتیپی و ژنوتیپی. تست‌های فنوتیپی درگیر تعیین مستقیم حساسیت دارویی هستند و به طور مستقیم رفتار دارویی یک نژاد از HIV را ارزیابی می‌کنند. در این روش ویروس HIV را که از یک بیمار جدا کرده‌اند کشت می‌دهند. تکثیر ویروس در کشت‌های سلولی زیر فشار انتخابی غلظت‌های مختلف و در حال افزایش داروهای ضد ویروس اندازه‌گیری و با نوع وحشی مقایسه می‌شود. تست‌های فنوتیپی معایبی دارند از جمله اینکه زمان‌بر و پرهزینه‌اند، نیاز به آزمایشگاه تخصصی دارند و از طرفی زمانی که لود (Load) ویروسی کمتر از ۲۰ درصد باشد، نمی‌توان مقاومت دارویی یک نژاد HIV را بررسی نمود.

تست‌های ژنوتیپی بر اساس آنالیز جهش‌های مرتبط با مقاومت دارویی هستند. برای انجام تست‌های ژنوتیپی ژنوم ویروس HIV گرفته شده از بیمار را استخراج و ژن‌های کد کننده‌ی آنزیم رونوشت بردار معکوس و پروتئاز را به cDNA رونویسی معکوس می‌کنند و cDNA حاصل را با PCR تکثیر می‌کنند. سپس جهش‌های موجود در این ژن‌ها (cDNA) به وسیله‌ی توالی‌یابی یا به وسیله‌ی روش‌های هیبریداسیون با نوع وحشی تعیین می‌شود. بسته به تعداد و انواع جهش‌های نشان داده شده، تست تعیین می‌کند که آیا بیمار به یک دارو یا ترکیبی از داروها مقاومت دارد یا خیر.

تست‌های ژنوتیپی هزینه‌ی زیادی ندارند و سریع انجام می‌شوند و همچنین تفسیر نتایج آن‌ها به فرایند تست کردن بستگی ندارد. بنابراین در توسعه‌ی

در طول تکثیر، کپی‌های درست و کپی‌های حاوی خطا (ویروس جهش یافته) از ویروس تولید می‌شود. جهش در ویروس HIV به دلیل نرخ بالای تکثیر ویروس و صحت پایین آنزیم رونوشت بردار معکوس (RT یا Reverse transcriptase) خیلی معمول است. در اثر تغییر در ساختار ژنتیکی HIV، توانایی یک دارو یا ترکیبی از داروهای که تکثیر ویروس را در بدن مهار می‌کنند، کاهش می‌یابد، این پدیده مقاومت دارویی نامیده می‌شود.

۱. رقابت بسیاری بین دانشمندان برای طراحی دارویی در مقابل HIV وجود دارد و ترکیب چند داروی مهار کننده تا به حال مؤثرترین درمان در مقابل این ویروس بوده است، که باعث جلوگیری طولانی مدت می‌شود. در حال حاضر، بیش از ۱۸ دارو علیه HIV موجود است؛ بیشتر آن‌ها روی دو آنزیم وابسته به ویروس تمرکز می‌کنند، یعنی پروتئاز و آنزیم رونوشت بردار معکوس. این داروها به طور کلی بر اساس هدف آنزیمی‌شان، مرحله‌ی مهار تکثیر ویروس و ترکیب شیمیایی آن‌ها به سه دسته‌ی اصلی تقسیم می‌شوند: مهار کننده‌های نوکلئوزید/ نوکلئوتید رونوشت بردار معکوس (NRTIs یا Nucleoside/Nucleotide reverse transcriptase inhibitors) که در مرحله‌ی تکثیر ویروس به وسیله‌ی بلوکه کردن امتداد بیشتر DNA پرو ویروس عمل می‌کنند، مهار کننده‌های غیر نوکلئوتیدی رونوشت بردار معکوس (NNRTIs یا Non-nucleoside reverse transcriptase inhibitors) و مهار کننده‌های پروتئاز (PIs یا Protease inhibitors) که با مرحله‌ی تجمع یافتن ذره‌ی ویروسی تداخل می‌کنند. در حقیقت، داروهای ضد HIV، آنزیم پروتئاز یا رونوشت بردار معکوس کد شده توسط ویروس را

روش‌ها

ماشین‌های بردار پشتیبان

ماشین‌های بردار پشتیبان یک روش یادگیری ماشین است که در سال ۱۹۹۲ ارایه شد. با اینکه این روش امروزه به روش‌های شناسایی الگو اضافه شده است، عملکرد خوبی در مسایل شناسایی الگو نشان داده است. به علاوه توانایی آن در عملکرد بهتر بر مبنای نمونه‌های آموزشی محدود، آن را در مسایل بیولوژیکی محاسباتی، روشی ایده‌آل ساخته است. یعنی در جاهایی که تولید داده مشکل یا هزینه‌بر است، عملکرد خوبی دارد، یکی از این مسایل تعیین مقاومت دارویی HIV است (۲).

در این بخش به طور خلاصه ماشین‌های بردار پشتیبان به عنوان یک جداساز معرفی می‌شود. ابتدا ساده‌ترین حالت ماشین‌های بردار پشتیبان، با فرض جدا پذیری داده‌ها به طور خطی، در نظر گرفته می‌شود. فرض کنید $D = \{(X_i, y_i)\}_{i=1}^l$ یک مجموعه Y داده شامل l نمونه X_i با برجسب‌های $y_i \in \{1, -1\}$ از دو کلاس باشد و بخواهیم این دو کلاس را به طور خطی از هم جدا کنیم. در نگاه اول می‌توان تعداد زیادی فوق صفحه برای جدا کردن این دو مجموعه در نظر گرفت، حال این سؤال پیش می‌آید که کدام یک از این فوق صفحات جدا کننده‌ی بهینه می‌باشد و کمترین خطا را برای تقسیم‌بندی دارد. به طور حسی می‌توان گفت که مناسب‌ترین انتخاب، صفحه‌ای است که بیشترین حاشیه را بین دو کلاس ایجاد کند. بنا به تعریف، حاشیه مجموع فاصله‌ی نزدیک‌ترین نقطه از هر دو کلاس تا صفحه‌ی جدا کننده می‌باشد (شکل ۱).

حال اگر داده‌ها به طور خطی جدا پذیر نباشند، باز

روش‌های محاسباتی برای پیش‌بینی مقاومت از روش‌های ژنوتیپی استفاده می‌شود (۱-۲).

چندین روش بیولوژیکی برای تفسیر نتایج تست‌های ژنوتیپی HIV ارایه شده است. برای ارزیابی تطبیق این روش‌ها و برای تعیین دنباله‌هایی که موجب ناهماهنگی بین روش‌ها می‌شوند، Ravela و همکاران چهار روش موجود را برای دنباله‌ی آمینو اسیدهای گرفته شده از ۲۰۴۵ نفر در کالیفرنیا شمالی به کار بردند و مقاومت دارویی را به صورت S یعنی مستعد پذیرش مقاومت، R یعنی مقاوم و I یعنی میانه کلاس بندی کردند. از این تفسیرها ۴/۴ درصد به طور کامل ناهماهنگ بودند، یعنی با حداقل یک روش تعیین کننده‌ی S و بقیه تعیین کننده‌ی R . ۲۹/۲ درصد تا حدودی ناهماهنگ بودند با حداقل یک روش تعیین کننده‌ی S و بقیه I یا حداقل یک روش تعیین کننده‌ی R و بقیه I . ۶۶/۴ درصد تفاسیر به طور کامل هماهنگ بودند یعنی هر چهار روش یک تفسیر انجام دادند (۳). چهار روش بیولوژیکی اعمال شده عبارت از ANRS (Agence national de recherches le SIDA)، HIVDB (HIV RT and protease sequence)، REGA (Rega institute database)، VGI و (Visible genetics) می‌باشند که هر روش لیستی از جهش‌ها را به عنوان ورودی می‌گیرد و خروجی‌ای تولید می‌کند که داروهای ضد HIV را با توجه به اندازه‌ی فعالیت پیش‌بینی شده طبقه‌بندی می‌کند (۳).

این مقاله عملکرد ماشین‌های بردار پشتیبان را در پیش‌بینی مقاومت دارویی HIV با استفاده از نتایج حاصل از این چهار روش بررسی می‌کند و عملکرد ماشین‌های بردار پشتیبان را روی نتایج حاصل از روش‌ها برای تعیین روش برتر برای هر دارو مقایسه می‌کند.

آن‌ها غیر صفر می‌باشد، بردار پشتیبان می‌گویند که این بردارهای پشتیبان روی مرز بین دو کلاس قرار می‌گیرند. در عمل استفاده از جداساز خطی جهت جدا کردن داده‌های غیر خطی، سبب کاهش چشم‌گیر کارایی می‌شود، بنابراین بهتر است از جداساز غیر خطی استفاده شود، این کار به راحتی با تصویر کردن داده‌ها به یک فضای ویژگی با ابعاد بالا امکان‌پذیر می‌باشد، به طوری که:

$$X \in R^d \Rightarrow Z(X) = (\phi_1(X), \dots, \phi_n(X)) \in R^n$$

حال می‌توان روابط مربوط به جداساز خطی را در این فضای جدید نوشت. در نتیجه، تابع تصمیم‌گیری مربوط به این حالت به فرم زیر تبدیل می‌شود:

$$f(x) = \text{sign}\left(\sum_{i=1}^l \lambda_i y_i Z^T(X)Z(X_i) + b\right)$$

یک نکته‌ی کلیدی درباره‌ی ماشین‌های بردار پشتیبان این است که برای محاسبه‌ی توابع تصمیم‌گیری، تنها مقداری که باید محاسبه شود حاصل ضرب نقطه‌ای $Z^T(X)Z(X_i)$ می‌باشد. برای راحتی کار، تابع کرنل K معرفی می‌شود:

$$Z^T(X)Z(X_i) = \sum_{i=1}^{\infty} \alpha_i \phi_i(X)\phi_i(Y) = K(X, Y)$$

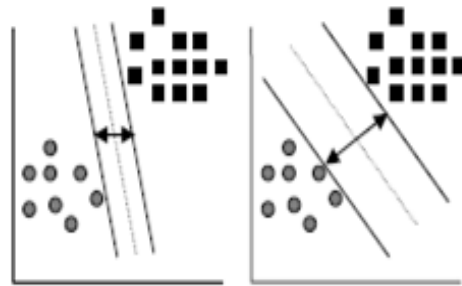
که $\{\alpha_i\}_{i=1}^{\infty}$ و $\{\phi_i\}_{i=1}^{\infty}$ ها به ترتیب یک سری اعداد و توابع حقیقی می‌باشند. به این ترتیب، تابع تصمیم‌گیری به فرم زیر تبدیل می‌شود:

$$(3) \quad f(x) = \text{sign}\left(\sum_{i=1}^l \lambda_i y_i K(X, X_i) + b\right)$$

باها از حل معادله‌ای مشابه معادله‌ی ۲ به دست می‌آیند با این تفاوت که $D_{ij} = y_i y_j K(X_i, X_j)$ در نظر گرفته می‌شود (۴).

در این مقاله برای آموزش ماشین‌های بردار پشتیبان از کرنل RBF استفاده شده است، همچنین برای آموزش و آزمایش ماشین‌های بردار پشتیبان از نرم‌افزار

هم می‌توان مانند حالت قبل فوق صفحه‌ای را تعیین کرد که دو کلاس را به طور خطی با بیشترین حاشیه و کمترین خطا (نمونه‌هایی که به ناچار غلط طبقه‌بندی می‌شوند) از هم جدا کند. تعادل بین حاشیه و خطای نمونه‌های غلط طبقه‌بندی شده را می‌توان توسط مقدار مثبت C که از قبل تعیین می‌شود، کنترل کرد. در این حالت می‌توان نشان داد که تابع تصمیم‌گیری $f(x)$ به فرم زیر بیان می‌شود:



شکل ۱. مرزهای جداکننده‌ی دو کلاس (۴).

$$(1) \quad f(x) = \text{sign}\left(\sum_{i=1}^l \lambda_i y_i X^T X_i + b\right)$$

که با ضرایب لاگرانژ حاصل از حل معادله‌ی زیر می‌باشند:

$$\text{Minimize: } w(A) = -A^T 1 + \frac{1}{2} A^T D A$$

$$(2) \text{ subject to: } A^T y = 0$$

$$A - C_1 \leq 0$$

$$-A \leq 0$$

$$D_{ij} = y_i y_j X_i^T X_j \text{ و } (1)_i = 1, (\Lambda)_i = \lambda_i$$

با حل معادله‌ی بالا یک سری ضرایب لاگرانژ غیر صفر به دست می‌آید که متناظر با تعدادی از داده‌های آموزشی هستند. البته تعدادشان نسبت به تعداد کل داده‌های موجود در پایگاه داده‌ی آموزشی کم می‌باشد. بقیه‌ی ضرایب هم که متناسب با دیگر داده‌ها هستند، صفر می‌شوند. به داده‌هایی که ضرایب لاگرانژ متناظر با

تست شد و داده‌ها به صورت تصادفی به داده‌های استفاده شده برای آموزش و آزمایش تقسیم شدند.

برای ارزیابی عملکرد روش، اندازه‌گیری‌های مختلف می‌توان انجام داد. در این مقاله برای این منظور چهار پارامتر دقت (Accuracy)، حساسیت (Sensitivity)، اختصاصی بودن (Specificity) و ضرایب همبستگی Matthews (MCC) اندازه‌گیری و به صورت زیر تعریف شده‌اند:

$$(۴) \text{ accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$(۵) \text{ sensitivity} = \frac{TP}{TP + FN}$$

$$(۶) \text{ specificity} = \frac{TN}{FP + TN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(FN + FP)(TN + FP)(TP + FP)(TP + FN)}} \quad (۷)$$

که TP، FN، FP و TN به ترتیب تعداد مثبت درست، مثبت نادرست، منفی درست و منفی نادرست می‌باشد.

عملکرد مدل با بررسی این پارامترها، بررسی شد و نتایج با استفاده از ارزیابی ۱۰ مرحله‌ای میانگین‌گیری گردیدند.

یافته‌ها

به طور معمول برای پیش‌بینی مقاومت دارویی ویروس HIV از نتایج روش بیولوژیکی HIVDB استفاده می‌شود. در این مقاله روش ماشین‌های بردار پشتیبان برای پیش‌بینی مقاومت دارویی HIV با استفاده از نتایج حاصل از چهار روش بیولوژیکی موجود، به کار گرفته و نتایج با هم مقایسه شده‌اند و در مورد هر دارو روش بهتر مشخص شده است. این نتایج در

LIBSVM استفاده شده است که در سایت <http://www.csie.ntu.edu.tw/~cjlin/libsvm> موجود می‌باشد.

مقاومت دارویی به عنوان یک مسأله‌ی شناسایی الگو
در این بخش راجع به اینکه چگونه مسأله‌ی مقاومت دارویی HIV می‌تواند به عنوان یک مسأله‌ی شناسایی الگو در نظر گرفته شود، بحث می‌شود.

بعضی از جهش‌ها باعث مقاومت می‌شوند و بعضی نه، در تست‌های ژنوتیپی محل‌های خاصی در طول دنباله‌ی ژنتیکی برای وجود آمینو اسیدهای خاصی چک می‌شوند و بسته به نتیجه، حساسیت نسبی به دارو اندازه‌گیری می‌شود. حال اگر مسأله از دیدگاه محاسباتی و شناسایی الگو بررسی شود، هر دنباله از آمینو اسیدها را می‌توان با یک بردار نمایش داد و هر دنباله به صورت S یا R نسبت به داروی خاصی برچسب زده می‌شود. حال اگر مجموعه‌ای از داده‌های ژنوتیپی موجود باشد، آیا ماشین بردار پشتیبان می‌تواند آموزش داده شود و مقاومت دارویی یک دنباله‌ی جدید را پیش‌بینی کند؟

مقاومت دارویی HIV می‌تواند S (Susceptible)، R (Resistance) و یا I (Intermediate) باشد اما تعداد اندکی از الگوها جزء دسته‌ی I طبقه‌بندی شده‌اند، بنابراین می‌توان آن‌ها را نیز R در نظر گرفت و داده‌ها را به عنوان دو کلاس S و R برچسب زد.

داده‌های استفاده شده در این مقاله، داده‌های استفاده شده توسط Ravela و همکاران (۳) هستند که شامل دنباله‌ی آمینو اسیدهای گرفته شده از ۲۰۴۵ نفر می‌باشند و در پایگاه داده‌ی <http://hivdb.stanford.edu/> موجود می‌باشند. عملکرد مدل با به کار بردن ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد باقی مانده برای آزمایش،

جدول ۲. نتایج ماشین‌های بردار پشتیبان برای نتایج روش ANRS

دارو	دقت	حساسیت	اختصاصی بودن	MCC
APV	۹۵/۱۵	۸۲/۷۹	۹۷/۳۹	۸۱/۶۰
IDV	۹۸/۶۸	۹۷/۶۲	۹۹/۵۲	۹۷/۲۵
LPV	۹۴/۹۰	۷۹/۷۲	۹۷/۳۰	۷۸/۴۸
NFV	۹۶/۸۱	۹۷/۹۳	۹۵/۷۶	۹۳/۵۷
RTV	۹۶/۲۳	۹۵/۶۱	۹۶/۵۷	۹۱/۵۷
SQV	۹۶/۵۲	۹۳/۹۴	۹۷/۹۸	۹۲/۴۳
3TC	۹۸/۷۷	۹۷/۶۴	۹۹/۹۰	۹۷/۵۳
ABC	۹۱/۷۲	۸۴/۶۸	۹۵/۰۰	۸۰/۲۹
AZT	۹۲/۷۰	۸۹/۹۵	۹۴/۹۸	۸۵/۰۵
D4T	۹۴/۱۷	۹۵/۴۲	۹۲/۹۹	۸۷/۹۶
DDI	۹۲/۵۰	۸۴/۲۲	۹۶/۶۵	۸۲/۸۴
TDF	۹۳/۵۸	۷۸/۱۱	۹۸/۹۱	۸۲/۸۱
DLV	۸۹/۶۶	۷۴/۴۲	۹۶/۱۱	۷۴/۵۵
EFV	۸۹/۴۱	۷۳/۷۲	۹۶/۵۲	۷۴/۸۵
NVP	۸۹/۰۷	۷۲/۸۶	۹۶/۶۲	۷۴/۲۶

ANRS: Agence national de recherches le SIDA

APV: Amprenavir

LPV: Lopinavir

RTV: Ritonavir

3TC: Lamivudine

AZT: Zidovudine

DDI: Didanosine

TDF: Tenofovir disoproxil fumarate

EFV: Efavirenz

IDV: Indinavir

NFV: Nelfinavir

SQV: Saquinavir

ABC: Abacavir

D4T: Stavudine

DLV: Delavirdine

NVP: Nevirapine

جدول ۳. نتایج ماشین‌های بردار پشتیبان برای نتایج روش REGA

دارو	دقت	حساسیت	اختصاصی بودن	MCC
APV	۹۳/۶۳	۹۳/۲۹	۹۳/۹۶	۸۷/۱۰
IDV	۹۱/۷۶	۸۹/۶۱	۹۵/۰۳	۸۳/۴۵
LPV	۹۶/۶۷	۹۱/۹۶	۹۷/۹۰	۸۹/۹۸
NFV	۹۲/۷۹	۹۴/۴۸	۸۹/۳۴	۸۳/۶۳
RTV	۹۱/۷۶	۸۹/۶۱	۹۵/۰۳	۸۳/۴۵
SQV	۹۴/۱۷	۹۵/۰۵	۹۳/۰۹	۸۸/۱۶
3TC	۹۲/۵۵	۸۶/۹۰	۹۹/۸۹	۸۵/۷۷
ABC	۹۶/۶۷	۹۷/۷۳	۹۳/۹۲	۹۲/۱۲
AZT	۹۵/۵۹	۹۲/۱۵	۹۸/۶۵	۹۱/۱۹
D4T	۹۴/۳۶	۹۰/۲۹	۹۷/۱۳	۸۷/۹۹
DDI	۹۶/۶۲	۹۶/۷۵	۹۶/۴۲	۹۲/۶۷
TDF	۹۳/۱۹	۸۶/۱۸	۹۶/۸۰	۸۴/۵۱
DLV	۸۶/۹۶	۶۹/۲۱	۹۶/۴۰	۷۰/۸۶
EFV	۸۶/۸۶	۶۸/۴۸	۹۶/۸۹	۷۱/۰۶
NVP	۸۶/۲۷	۶۷/۴۴	۹۶/۸۶	۷۰/۱۱

REGA: Rega institute

IDV: Indinavir

NFV: Nelfinavir

SQV: Saquinavir

ABC: Abacavir

D4T: Stavudine

TDF: Tenofovir disoproxil fumarate

EFV: Efavirenz

APV: Amprenavir

LPV: Lopinavir

RTV: Ritonavir

3TC: Lamivudine

AZT: Zidovudine

DDI: Didanosine

DLV: Delavirdine

NVP: Nevirapine

جدول‌های ۱ تا ۴ آورده شده است و در مورد هر دارو بهترین نتیجه پررنگ‌تر از بقیه نشان داده شده است. به طور کلی ماشین‌های بردار پشتیبان به عنوان یک جداساز بسیار قوی برای پیش‌بینی مقاومت دارویی HIV عمل کرده است و دقت این روش برای نتایج مربوط به روش‌های گفته شده بین ۹۸/۷۷-۸۶/۲۷ درصد می‌باشد.

جدول ۱. نتایج ماشین‌های بردار پشتیبان برای نتایج روش

HIVDB				
دارو	دقت	حساسیت	اختصاصی بودن	MCC
APV	۹۸/۵۸	۹۷/۶۷	۹۹/۱۹	۹۷/۰۴
IDV	۹۷/۲۵	۹۴/۴۰	۹۹/۲۷	۹۴/۳۴
LPV	۹۶/۳۲	۹۵/۱۰	۹۷/۱۷	۹۲/۲۱
NFV	۹۸/۷۳	۹۷/۵۲	۱۰۰	۹۷/۴۴
RTV	۹۷/۳۰	۹۴/۴۰	۹۹/۴۰	۹۴/۴۵
SQV	۹۷/۸۴	۹۶/۳۴	۹۸/۹۹	۹۵/۴۶
3TC	۹۶/۳۲	۹۲/۹۸	۱۰۰	۹۲/۷۸
ABC	۹۷/۴۵	۹۷/۳۱	۹۸/۱۶	۹۳/۸۴
AZT	۹۵/۵۹	۹۲/۵۴	۹۸/۸۲	۹۱/۲۱
D4T	۹۵/۲۹	۹۳/۳۰	۹۷/۲۰	۹۰/۴۳
DDI	۹۶/۹۶	۹۷/۲۱	۹۶/۰۸	۹۲/۷۹
TDF	۹۳/۴۳	۹۳/۴۷	۸۵/۹۹	۸۵/۹۹
DLV	۸۹/۷۵	۷۳/۰۱	۹۷/۲۳	۷۵/۵۰
EFV	۸۸/۸۷	۷۱/۴۱	۹۷/۲۴	۷۴/۲۱
NVP	۸۸/۶۳	۷۰/۸۷	۹۷/۲۳	۷۳/۷۶

HIVDB: HIV database

IDV: Indinavir

NFV: Nelfinavir

SQV: Saquinavir

ABC: Abacavir

D4T: Stavudine

TDF: Tenofovir disoproxil fumarate

EFV: Efavirenz

APV: Amprenavir

LPV: Lopinavir

RTV: Ritonavir

3TC: Lamivudine

AZT: Zidovudine

DDI: Didanosine

DLV: Delavirdine

NVP: Nevirapine

اعمال ماشین‌های بردار پشتیبان بر روی نتایج روش HIVDB برای داروهای APV، NFV، ABC، AZT، D4T، DDI، TDF و DLV بهترین نتایج را با توجه به پارامترهای اندازه‌گیری شده داشته است. روش،

عملکرد خوبی داشته است و برای داروهای IDV, LPV, RTV, SQV و DDI ماشین‌های بردار پشتیبان برای روش VGI عملکرد خوبی نشان داده است.

بحث

اغلب برای پیش‌بینی مقاومت دارویی ویروس HIV از نتایج روش بیولوژیکی HIVDB استفاده می‌شود، در این مقاله ماشین‌های بردار پشتیبان به عنوان یک جداساز قوی برای پیش‌بینی مقاومت دارویی HIV بررسی شد. این روش روی نتایج حاصل از چهار روش بیولوژیکی موجود آزمایش شد و در مورد هر دارو، روشی که نتایج بهتری با توجه به پارامترهای ذکر شده داشته است، تعیین گردید؛ تا هنگام استفاده از هر دارو برای درمان، ابتدا برای تعیین مقاومت دارویی با استفاده از روش‌های یادگیری ماشین، از نتایج روش بیولوژیکی بهتر برای آموزش روش یادگیری ماشین استفاده شود.

همچنین ماشین‌های بردار پشتیبان پیش‌بینی مقاومت دارویی را با دقتی بین ۹۸/۷۷-۸۶/۲۷ درصد، حساسیت ۹۸/۵۸-۹۷/۴۴ درصد، اختصاصی بودن ۱۰۰-۸۹/۳۴ درصد و MCC ۹۷/۵۳-۷۰/۱۱ درصد برای چهار روش بیولوژیکی موجود انجام داده است.

جدول ۴. نتایج ماشین‌های بردار پشتیبان برای نتایج روش VGI

دارو	دقت	حساسیت	اختصاصی بودن	MCC
APV	۹۸/۰۹	۹۶/۲۶	۹۹/۲۵	۹۵/۸۳
IDV	۹۸/۶۸	۹۷/۶۲	۹۹/۵۲	۹۷/۲۵
LPV	۹۶/۶۲	۹۶/۱۰	۹۶/۷۳	۹۲/۰۲
NFV	۹۸/۳۸	۹۸/۷۷	۹۸/۱۷	۹۶/۷۳
RTV	۹۸/۶۸	۹۷/۶۲	۹۹/۵۲	۹۷/۲۵
SQV	۹۸/۷۳	۹۸/۵۸	۹۸/۹۳	۹۷/۲۵
3TC	۹۷/۲۵	۹۴/۸۲	۹۹/۹۰	۹۴/۵۴
ABC	۹۳/۰۹	۸۸/۸۱	۹۷/۸۱	۸۶/۲۸
AZT	۹۴/۵۶	۹۲/۱۶	۹۷/۳۶	۸۸/۹۴
D4T	۹۲/۲۵	۸۸/۹۸	۹۵/۰۸	۸۴/۲۰
DDI	۹۶/۹۱	۹۷/۸۸	۹۵/۰۷	۹۳/۰۱
TDF	۹۲/۹۹	۸۵/۲۳	۹۷/۷۳	۸۵/۰۷
DLV	۸۷/۲۵	۶۹/۵۵	۹۶/۵۰	۷۱/۳۳
EFV	۸۶/۷۲	۶۸/۱۴	۹۶/۹۶	۷۰/۸۸
NVP	۸۶/۷۲	۶۸/۱۴	۹۶/۹۶	۷۰/۸۸

VGI: Visible Genetics Inc

APV: Amprenavir

LPV: Lopinavir

RTV: Ritonavir

3TC: Lamivudine

AZT: Zidovudine

DDI: Didanosine

DLV: Delavirdine

NVP: Nevirapine

IDV: Indinavir

NFV: Nelfinavir

SQV: Saquinavir

ABC: Abacavir

D4T: Stavudine

TDF: Tenofovir disoproxil fumarate

EFV: Efavirenz

ماشین‌های بردار پشتیبان برای داروهای IDV, TDF 3TC EFV و NVP بهترین نتایج را بر روی روش ANRS داشته است. برای داروهای LPV و AZT ماشین‌های بردار پشتیبان برای نتایج روش REGA

References

- Bonet I, Garcia MM, Salazar S, Sanchez R, Saey Y, Van de Peer Y, et al. Predicting human immunodeficiency virus (HIV) drug resistance using recurrent neural networks. Proceedings of the 10th International Electronic Conference on Synthetic Organic Chemistry; 2006 Nov 1-30; Basel, Switzerland.
- Tesfamichael S, Hazelhurst S. Support vector machine prediction of HIV-1 drug resistance using the viral nucleotide patterns. Transactions of the Royal Society of South Africa 2009; 64(1): 62-72.
- Ravela J, Betts B, Brun-Vezinet F, Vandamme A, Descamps D, Laethem K, et al. HIV-1 protease and reverse transcriptase mutation patterns responsible for discordances between genotypic drug resistance interpretation algorithms. J Acquir Immune Defic Syndr 2003; 33(1): 8-14.
- Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge, UK: Cambridge University Press; 2000.

Predicting Human Immunodeficiency Virus Drug Resistance Using Support Vector Machines

Samira Abbasi¹, Majid Mohammadbeigi PhD²

Abstract

Background: The purpose of this study was to investigate the performance of support vector machines (SVMs) in predicting human immunodeficiency virus (HIV) drug resistance using amino acids sequence analysis.

Methods: In order to evaluate the performance of SVMs in predicting HIV drug resistance, SVMs were trained and tested using LibSVM software. Biological methods developed by HIV database (HIVDB), Visible Genetics Inc. (VGI) and the French National Agency for AIDS Research (ANRS) and REGA algorithm were employed to interpret the results of HIV genotypic resistance tests. The results were compared and the most efficient method for each drug was finally determined.

Findings: With an accuracy of 86.27- 98.77, SVMs are highly successful classifiers for predicting HIV drug resistance. Using SVMs for HIVDB results has the best performance for amprenavir (APV), nelfinavir (NFV), abacavir (ABC), zidovudine (AZT), stavudine (D4T), didanosine (DDI), tenofovir disoproxil fumarate (TDF), and delavirdine (DLV). Using SVMs for the results of ANRS method has the best performance for indinavir (IDV), lamivudine (3TC), TDF, efavirenz (EFV), and nevirapine (NVP). Using SVMs for REGA results has the best performance for lopinavir (LPV) and AZT. Finally, using SVMs for the results of VGI method has the best performance for IDV, LPV, ritonavir (RTV), saquinavir (SQV), and DDI.

Conclusion: SVMs are highly successful classifiers for predicting HIV drug resistance. Before starting treatment with each drug, researchers can determine HIV drug resistance with machine learning methods.

Keywords: Human immunodeficiency virus, Support vector machine, Drug resistance

¹ MSc Student, Department of Biomedical Engineering, School of Engineering, University of Isfahan, Isfahan, Iran

² Assistant Professor, Department of Biomedical Engineering, School of Engineering, University of Isfahan, Isfahan, Iran

Corresponding Author: Samira Abbasi, Email: samira.abbasi@gmail.com