پیش‌بینی میل اتصال آپتامر به ویروس آنفولانزا با استفاده از شبکه عصبی

فاطمه شعاع‌اردبیلی1، امین جانقربانی‌پوده1\* و احمدفرهاد طالبی1

1 گروه بیوفناوری دانشکده علوم و فناوری‌های نوین، دانشگاه سمنان، ایران

\*a.janghorbani@semnan.ac.ir

**چکیده**: ویروس آنفولانزا سالانه باعث مرگ تعداد زیادی انسان و حیوان در سراسر جهان می‌شود و به خسارات اقتصادی چشمگیری منجر می‌شود. توسعه روش‌های تشخیصی مطمئن، موثر، سریع و با حساسیت بالا، امکان کنترل همه­گیری و مواجه با خسارات را فراهم می‌کند. با توجه به پتانسیل بالای کاربرد آپتامرها در تشخیص، کنترل و حتی درمان عفونت آنفولانزا، معرفی روش­های مولکولی نوین براساس این الیگونوکلئوتیدهای متصل‌شونده به ذرات ویروسی مورد توجه قرار گرفته است. با توجه به توان بالای محاسبات هوش مصنوعی، طراحی و استخراج ویژگی‌های آپتامرهای اختصاصی ویروس آنفولانزا در مطالعه حاضر به اجرا رسیده است. در این پژوهش با استفاده از شبکه‌های عصبی میزان تمایل اتصال آپتامرها و ویروس آنفولانزا مدل‌سازی شده‌است. مجموعه داده این مقاله شامل ۹ ویژگی‌های ساختار مولکولی 96 آپتامر به عنوان توصیف‌گر و میزان تمایل اتصال آن‌ها به ویروس آنفولانزا به عنوان متغیر مستقل است. در این مقاله از دو نوع شبکه‌ عصبی MLP و GRNN برای پیش‌بینی میل اتصال آپتامرها بهره گرفته شد. برای رسیدن به این هدف ۹ ویژگی بیان شده به عنوان ورودی‌های شبکه و میزان میل اتصال به عنوان خروجی شبکه در نظر گرفته شد. بعد از آموزش شبکه‌های عصبی با استفاده از دادگان آموزش دقت پیش‌بینی میل اتصال در شبکه عصبی GRNN برای دادگان آزمون و براساس معیار r2 مقدار 0.64 بدست آمده است. همین فرآیند برای شبکه عصبی MLP تکرار شد عملکرد این شبکه در پیش بینی میل اتصال در دادگان آزمون r2=0.77 حاصل شد. نتایج حاصل از این روش‌ها در مقایسه با روش رگرسیون چند خطی با مقدار r2=0.66 نشان از برتری مدل MLP‌ در پیش‌بینی میل اتصال آپتامر‌ها به ویروس آنفولانزاست.

کلید واژه‌ها: آپتامر، ویروس آنفولانزا، میل اتصال، پیش‌بینی، هوش مصنوعی شبکه‌های عصبی، زیست‌محاسباتی

1- مقدمه

تشخیص سریع نوع، غلظت بیومولکول‌ها و انواع تعاملات آنان بحث مهمی در علوم زیستی به شمار می‌آید. دسته‌ای از این میانکنش­ها مابین سلول­های زیستی و عوامل بیماری­زا منجر به ظهور بیمار­ی­ها و عفونت­های مختلف می­شوند. در این میان ویروس آنفولانزا سالانه باعث مرگ تعداد زیادی انسان و حیوان در سراسر جهان می‌شود و به خسارات اقتصادی چشمگیری منجر می‌شود. علاوه بر این، در سال2020، تهدید جدیدی رخ داد که باعث عفونت دستگاه تنفسی در پستانداران شد و در مدت کوتاهی تبدیل به یک بیماری همه‌گیر در بین انسان‌ها و منجر به مرگ هزاران انسان شد. ویروس‌های زیادی در محیط وجود دارند که به انسان، سایر حیوانات یا هر دو در اطراف آن آلوده می‌شوند. برخی از آن‌ها به توسعه بیماری‌ها کمک می‌کنند. بیماری‌زا‌ترین سویه‌های ویروسی می‌تواند باعث بیماری‌هایی شود که منجر به آسیب دائمی به سلامتی و حتی مرگ شود ]1[. با بهره‌گیری از روش‌های هوش مصنوعی میتوان در تشخیص سریع و به موقع این تهدید اقدامات موثری انجام داد. ویروس آنفولانزا اغلب دارای ساختار کروی زمخت با قطری در حدود 0.21-0.1 نانومتر می‌باشد اما گاها به شکل رشته ای نیز دیده می‌شود. در ساختار مرکزی ویروس RNA تک رشته‌ای با پلاریته منفی به همراه پروتئین‌های حفاظت کننده وجود دارد. ویروس‌های زیادی وجود دارند که برای انسان بسیار خطرناک است. در میان آن‌ها، یکی از مهمترین آن‌ها ویروس آنفولانزا است که نه تنها به راحتی جهش می‌یابد، بلکه باعث شیوع بیشتر بیماری در انسان و حیوانات می‌شود، اما برخی از سویه‌های آن نیز باعث عفونت با میزان بقای کم می‌شود]2[. به نظر می‌رسد که کلید مقابله کارآمد با گسترش ویروس‌های خطرناک برای انسان، تشخیص سریع و صحیح است که اغلب شامل تمایز عفونت‌ها با روند مشابه در مورد ویروس آنفولانزا و SARS-CoV-2 است. در میان روش‌های استاندارد تشخیصی مورد استفاده در تشخیص عفونت‌های ویروسی، روش­های سرم­شناسی و روش‌های مولکولی بر اساس واکنش زنجیره‌ای پلیمریزاسیون غالب هستند]3[.

**1-1- روش‌های تشخیص ویروس آنفولانزا**

روش های تشخیصی: شماری از روش‌های آزمایشگاهی برای شناسایی و تشخیص آنفولانزا وجود دارد که از حساسیت و اختصاصیت‌های متفاوتی در افتراق آنفولانزا نوع A از B و C و یا افتراق زیرگونه های نوع A برخوردار هستند.

این تست ها قابلیت ردیابی نوکلوپروتئین‌های ویروسی را دارند و از جمله ویژگی‌های آن‌ها اختصاصیت 99 % و اعلام نتیجه در کمترین مدت زمان ممکنه است. این تست‌ها را بر اساس قابلیت افتراقی‌شان به سه دسته زیر تقسیم می‌کنند:

1-تست‌های که آنفولانزا نوع A را تشخیص می‌دهند

2-تست‌های که دو نوع A وB را از یکدیگر تشخیص و افتراق می‌دهند

3-تست‌های که آنفولانزا نوع A و B را تشخیص می‌دهند اما قابلیت افتراق این دو نوع را از یکدیگر ندارند.

به منظور جلوگیری از عواقب نامطلوب عفونت‌های ویروسی، بهداشتی و اقتصادی، نه تنها راه حل‌های درمانی موثر لازم است، بلکه بیش از همه روش‌های تشخیصی مطمئن، موثر و سریع با حساسیت و ویژگی بالا، که در نتیجه تشخیص، امکان درمان مناسب را فراهم می‌کند که برای این‌کار می‌توانیم از آپتامرها بهره بگیریم.

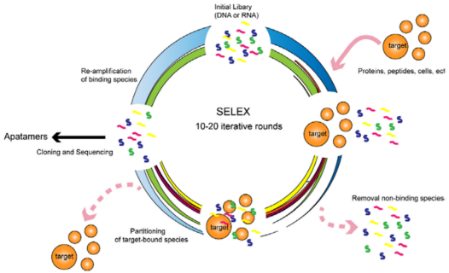
**1-2- آپتامر**

آپتامرها توالی‌های تک رشته‌ای سنتزی RNA یا DNA (اخیرا، پتیدی) هستند که به ساختارهای فضایی شکل می‌دهند و با اختصاصیت فوق العاده زیاد به اهداف معینی متصل می‌گردند. اولین بار آپتامرها در سال 1990 ارائه شدند، خصوصیات بی‌نظیر آن‌ها باعث شده است كه موثرتر از آنتی بادی‌ها عمل نمایند. آپتامرها عموما از طریق فرایند آزمایشگاهی [[1]](#footnote-1)SELEX از یک کتابخانه، غربال و انتخاب می‌شوند که می‌توانند به هر مولکول هدفی (از یون‌های معدنی کوچک تا سلول های کامل) متصل ‌شوند. آپتامرها زمانی که انتخاب شدند، طی فرایند PCR تکثیر شده تا مقادیر زیادی از آن‌ها با خلوص بالا تهیه شود. ساختار شیمیایی ساده آپتامرها آن‌ها را پذیرای اصلاحات کاربردی بیشتر با توجه به هدف‌های مختلف می‌نماید. همچنین آپتامرها بسیار پایدارتر از آنتی بادی‌ها هستند که آن‌ها را مناسب و سازگار برای شرایط دشوار (مثل بالا بودن دما) می‌نماید. مطالعات و پژوهش‌های روزافزون در زمینه کاربردهای آپتامرها همانند تشخیص و درمان در حال انجام است. احتمالاً در آینده‌ای نزدیک آپتامرها استفاده فزاینده‌ای هماهنگ با دیگر مولکول‌های درمانی پیدا کند]4[. بعلاوه کاربردهای آپتامرها مانند ابراز تشخیصی، درمانی، تصویربرداری زیستی و تحویل دارو معرفی شده است]5.[ با توجه به پتانسیل بالای آپتامرها در کنترل و مهار ویروس آنفولانزا و با توجه به اینکه روش‌های موجود در حوزه شبکه‌های عصبی مانند توانایی پیش‌بینی اتصال آپتامرهای انتخاب شده با ویژگی و میل زیاد، استخراج توالی و ویرایش ژن‌ها، بسیار کارآمد هستند. شبکه های عصبی مصنوعی، به عنوان یک الگوریتم قدرتمند، به طور وسیعی در بهره‌برداری از مقدار زیادی از اطلاعات و داده‌های تجربی از طریق یادگیری و به ویژه در تخمین کمی استفاده شده‌اند. شبکه‌های عصبی می‌توانند به طور مداوم ویژگی‌های اصلی را از طریق یک لایه پنهان استخراج کنند تا به تأثیر کاهش ابعاد برسند و کار را برای شناسایی یا طبقه‎بندی آپتامرها آسان‌تر بکنند. و با صرف زمان کوتاه می‌توان نتایج ارزشمندی بدست آورد. می‌توانیم با استفاده از ویژگی‌های استخراج شده از آپتامرها و ویروس آنفولانزا و بکارگیری آن‌ها در حوزه هوش‌مصنوعی پیشرفت‌های بسیاری در علم پزشکی حاصل کنیم و از وقوع بیماری‌های خطرناکی که امکان مرگ بیمار را دارند، جلوگیری کنیم.

**1-3- فرآیند انتخاب آپتامر تحت عنوان SELEX**

فرآیند SELEX شامل چندین چرخه انتخاب و تقویت است: الف) مجموعه‌ای از مولکول‌های RNA یا DNA نمایش داده می‌شود و آپتامرهایی که دارای قدرت اتصال به یک مولکول هدف هستند از غیر آپتامرها جدا شده و (ب) آپتامرهای نگه‌دارنده با استفاده از واکنش‌ زنجیره‌ای پلیمراز (PCR) دنباله‌ها برای دور بعدی غنی‌سازی تقویت می‌شوند. کل مراحل انتخاب بطور معمول به 15 دور انتخاب نیاز دارد و می‌تواند از چند روز تا چند ماه طول بکشد. در مرحله بعدي، اين کتابخانه تصادفي DNA یا RNA در مجاورت مستقيم مولکول هدف قرار گرفته و در ادامه، کمپلکس‌هاي ايجادشده با هدف، از اليگونوکلئوتيدهاي باند نشده (مرحله Partition) جدا مي‌شوند. اين مرحله يکي از سخت‌ترین مراحل SELEX است و بطور قابل ملاحظه‌اي بر خصوصيات اتصالي آپتامر به هدف، اثر مي‌گذارد. در ادامه اليگونوکلئوتيدهاي متصل شده به هدف، از هدف جداسازی شده (شستشو داده می‌شوند) و توسط PCR (برای (DNA SELEXو RT-PCR (برای RNA SELEX) تکثیر می‌یابند. خروجي فرآيندPCR ، dsDNA است و می‌بایست به ssDNA تبديل شود]6[. و توالي مکمل آپتامر انتخاب شده، حذف گردد. حال توالي تک‌رشته‌ی باقیمانده وارد چرخه دوم SELEX شده و به‌طور مجدد در مجاورت هدف قرار مي‌گيرد. به دليل تکرارشونده بودن اين چرخه‌ي انتخاب و تکثير اليگونوکلئوتيدها در هر مرحله، استخر توالي در هر مرحله دستخوش کاهش تنوع اليگونوکلئوتيد قرار مي‌گردد؛ در حالي که ميل ترکيبي اليگونوکلئوتيدهاي باقيمانده، رفته رفته افزايش مي‌يابد. تعداد دفعات چرخه‌ی (راند) مورد نیاز در SELEX به پارامترهاي مختلفي از جمله خصوصيات هدف و غلظت آن، طراحي کتابخانه اوليگونوکئوتيدي آغازگر، شرايط محيطي انتخاب، نسبت مولي هدف به اوليگونوکلئوتيدها و کارآيي روش جداسازی (Partitioning) وابسته است. راندهاي اضافي منحصراً با توجه به ميزان اختصاصيت اوليگونوکلئوتيدها امکان پذير است.

با توجه به زمانبر و هزینه‌بر بودن فرآیند SELEX به نظر می‌رسد ارزیابی بکارگیری روش‌های محاسباتی برای پیش‌بینی میل اتصال آپتامر‌ها بر اساس ویژگی‌های مختلف این توالی‌ها بتواند کمک شایانی به شناسایی آپتامر‌های مناسب و با پتانسیل بالا در اتصال به ویروس آنفولانزا نماید. هوش مصنوعی ازجمله ابزارهای محاسباتی است که کاربردهای فراوانی در زمینه پزشکی و حوزه بیوانفورماتیک دارد و می‌تواند کاربردهای فراوانی در بررسی و طبقه‌بندی آپتامرها و انتخاب مناسب آپتامرها داشته باشد.



**شکل 1**- شماتیک فرآیند SELEX برای کتابخانه‌های RNA و DNA ]6[

**1-4- کاردبرد هوش مصنوعی در طراحی آپتامرها**

در مطالعه‌ی توان اتصالی آپتامرها به پروتئین هماگلوتینین ویروس آنفولانزا در سال 2014، شش خصوصیت فضایی و فیزیکوشیمیایی آپتامرها برای این نوع اتصال اختصاصی به کمک محاسبات هوش‌مصنوعی مهم تشخیص داده شدند. در واقع محاسبات انجام شده توسط کامپیوتر، آپتامرهایی را طراحی و پیشنهاد کردند که تا 20 برابر مؤثرتر به ذرات ویروسی متصل می­شدند]7[. در مطالعات مشابهی تحقیقات بسیاری روی روابط ساختار-فعالیت کمّی [[2]](#footnote-2)و روابط ساختار-ویژگی‌کمّی انجام و گزارش شده است. با این حال مقالات کمی در مورد QSARs[[3]](#footnote-3) برای آپتامرها منتشر شده است که علت آن هم دشواری بدست آوردن توصیف‌گرها[[4]](#footnote-4) دو بعدی و سه بعدی مستقیما از آپتامرها می‌باشد. در مطالعه‌ای با استفاده از توالی آمینواسیدهایی که از توالی آپتامر و ارتقاء مدل‌هایQSAR که مربوط به تمایل اتصال آپتامرها به ویروس آنفولانزا هستند، به توصیف روش بدست آوردن توصیف‌گرهای مولکولی پرداخته شده است و از شبکه‌عصبی رگرسیون عمومی بر مبنای تخمین پنجره های Parzen برای ساخت مدل QSAR با اعمال شش توصیف‌گر مولکولی استفاده شده است]8[.

در مطالعه‌ای دیگر با توجه به همه‌گیری فصلی ویروس‌های آنفولانزا A که تهدید بزرگی برای سلامت انسان است و بنابراین زیرگروه‌بندی سریع ویروس‌های آنفولانزای A اهمیت فزاینده‌ای دارد. روش‌های سنتی بیوشیمیایی برای زیرگونه سازی این ویروس‌ها گران و زمانبر هستند. تکنیک‌های مختلف توالی‌یابی و روش‌های یادگیری عمیق فرصتی را برای تجزیه و تحلیل و کسب اطلاعات بیشتر از آن‌ها فراهم می‌کند. این مقاله یک مدل یادگیری مجموعه‌ای مبتنی بر شبکه عصبی پیچیده عمیق را برای تشخیص دقیق همه زیرگروه ویروس‌های آنفولانزای A پیشنهاد می‌کند. آزمایشات نشان می‌دهد که روش پیشنهادی می‌تواند به عملکرد پیشرفته‌ای برای زیرگونه‌سازی ویروس‌های آنفولانزای A و تشخیص زیرگروه‌های جدید آن با توجه به داده‌های توالی دست یابد]9[.

برای بررسی بهتر و کارآمدتر نحوه اتصال و کیفیت اتصال آپتامرها و ویروس‌ها می‌توان از روش استخراج ویژگی استفاده نمود تا با در دسترس داشتن ویژگی‌های مهمی که بیشترین تاثیر را در اتصال آپتامر-ویروس دارند با افزایش دقت و کاهش زمان، کیفیت نتایج نیز افزایش یابد. لذا در این مطالعه سعی شده است با استفاده از شبکه‌های رگرسیون تعمیم‌یافته و شبکه عصبی پرسپترون چند لایه بر اساس ویژگی‌های مولکولی میل اتصال آپتامر به ویروس آنفولانزا پیش‌بینی شوند.

2- روش‌ها و ابزارها

**2-1- مجموعه داده**

در این پژوهش از مجموعه داده مقاله‌ای سال 2014 که در رابطه با پیش‌بینی میل اتصال آپتامر به ویروس آنفولانزا بر اساس روش‌های رگرسیون آماری منتشر شده است بهره گرفته شد ]7[. در مجموعه داده این مقاله شامل 99 توالی آپتامر به همراه تمایل اتصال به ویروس آنفولانزا نمایش داده شده است. که این مجموعه از آپتامرها را به دو مجموعه آموزش (72 آپتامر) و مجموعه آزمون (24آپتامر) تقسیم کرده‌اند. در این مجموعه داده تعداد 9 ویژگی یا توصیف­گر براساس ویژگی‌های ساختار مولکولی که بیشترین تاثیر را در بررسی میزان اتصال آپتامر و ویروس آنفولانزا را دارند، ارائه شده است. (جدول ۱) که به عنوان ورودی‌های مدل پیش‌بینی میل اتصال مورد استفاده قرار گرفته است.

**جدول 1**- معرفی توصیف­گرهای مولکولی استفاده شده در ]7[

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| شماره ردیف | توضیح توصیف‌گر | محدوده تغييرات توصیف‌گر‌ها | | | | |
|  |  | میانگین | بیشینه | کمینه | انحراف معيار |
| ۱ | تعداد کل بازها | 55.41 | 69 | 30 | 9.16 |
| ۲ | تعداد انطباقات محاسبه شده | 1.08 | 4 | 0 | 0.59 |
| ۳ | طول توالی‌های C جفت نشده که با سایر توالی‌های کمتر از 8 نوکلئوتیدی قطع نشده باشد | 17.17 | 55 | 0 | 15.78 |
| ۴ | مجموع نوکلئوتیدهای جفت نشده C با طول سه یا بیشتر | 17.04 | 55 | 0 | 16.10 |
| ۵ | مجموع تمام نوکلئوتیدهای جفت نشده | 33.24 | 55 | 4 | 9.03 |
| ۶ | اندازه بزرگترین حلقه | 22.21 | 39 | 0 | 12.58 |
| ۷ | نسبت بین نوکلئوتیدهای جفت نشده به کل | 0.613 | 1 | 0.059 | 0.170 |
| ۸ | نسبت بین بزرگترین حلقه به کل | 0.409 | 0.778 | 0 | 0.240 |
| ۹ | نسبت بین طول توالی‌هایC جفت نشده به کل | 0.325 | 1.000 | 0 | 0.301 |

**2-2- شبکه عصبی مصنوعی**

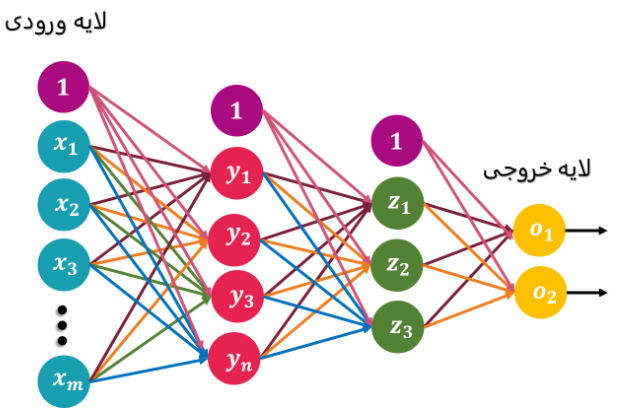
شبکه‌های عصبی مصنوعی، به عنوان یک الگوریتم قدرتمند، به طور وسیعی در رگرسيون غير‌خطي و مدل‌سازي ارتباط بين متغيرهاي توصيف‌گر و متغير پاسخ بهره‌ گرفته شده است. در این پژوهش به منظور مدل‌سازی ارتباط بین ویژگی‌های آپتامر با میل اتصال آن‌ها به ویروس آنفولانزا از شبکه‌های GRNN[[5]](#footnote-5) و MLP[[6]](#footnote-6) بهره گرفته شده‌است.

**GRNN**

شبکه عصبی رگرسیون عمومی (GRNN) ابزاری بسیار مفید برای پیاده‌سازی پیش‌بینی‌های غیرخطی است. الگوی GRNN دارای مزایای مختلفی در مقایسه با سایر شبکه‌های عصبی مانند MLP است. GRNN نیازی به تعریف تعداد لایه‌های پنهان از قبل و آموزش تکراری ندارد. علاوه بر این از GRNN می‌توان برای تعداد نسبتاً کمی نمونه استفاده کرد. ساختار متداول یک GRNN شامل لایه‌های ورودی، الگو[[7]](#footnote-7)، جمع و خروجی است. نورون‌های ورودی برابر با تعداد متغیرهای ورودی در مساله است. تعداد نورون‌های موجود در لایه الگو با تعداد نمونه‌های آموزش برابر است. یک نورون الگو فاصله اقلیدسی نمونه تست را از نقطه مرکزی نرون محاسبه می‌کند و سپس یک تابع کرنل RBF را بر روی آن اعمال و مقدار بدست آمده به نورون‌های لایه جمع فرستاده می‌شود. لایه جمع [[8]](#footnote-8) از دو نوع گره تشکیل شده است: گره‌های مخرج[[9]](#footnote-9) و گره‌های صورت[[10]](#footnote-10) به این خاطر که در فرایند محاسبات این شبکه عصبی در نهایت به خروجی، دو مقدار سیگما می‌رسد که باید تقسیم بر هم شوند. گره صورت مقادیر وزن ضرب شده با مقدار هدف را برای هر نورون لایه الگو جمع می‌کند و گره دوم یا مخرج دارای سیگمای مقادیر هدف برای هر یک از نورون لایه الگو است. واحد خروجی مقادیر برآورد مورد نظر را ارائه می‌دهد. برای GRNN تنها ضریب پخش σ (یعنی پارامتر صاف یا ضریب عرض) تابع گاوسی (که D نشان دهنده یک تابع اسکالر است) باید تنظیم شود. برای مقدار σ بزرگتر، نمایش احتمالی نقطه نمونه ارزیابی شده برای دامنه وسیع‌تری از X امکان‌پذیر است. برای مقدار σ کوچک، نمایش به محدوده باریکی از X محدود می‌شود.

**MLP**

شبکه عصبی خود را تغییر داده و از شبکه عصبی MLP استفاده کردیم. پرسپترون چند لایه یک شبکه عصبی مصنوعی پیشرو[[11]](#footnote-11) است که مجموعه‌ای از خروجی‌ها را از مجموعه ورودی‌ها تولید می‌کند. در یک شبکه‌عصبی پرسپترون چند لایه، حداقل سه لایه وجود خواهند داشت که این لایه‌ها عبارتند از یک لایه ورودی[[12]](#footnote-12)، یک لایه نهان[[13]](#footnote-13)و یک لایه خروجی[[14]](#footnote-14). آموزش شبکه عصبی پرسپترون چند لایه توسط الگوریتم یادگیری پس انتشار خطا[[15]](#footnote-15) انجام می‌شود]10[. ساختار پرسپترون چند لایه در شکل 2 نشان داده شده‌است:



**شکل2**- شماتیک شبکه عصبی MLP

به این شبکه ورودی‌های یکسان با شبکه عصبی GRNN اعمال گردید. شبکه عصبیMLP دولایه پنهان با تعداد نورون 8 در لایه اول و تعداد نورون 10 در لایه دوم در نظر گرفته شد. فرآیند آموزش شبکه در epoch=300 انجام شد که در آن ۸۵ درصد دادگان به عنوان داده آموزش و ۱۵ درصد دادگان برای داده اعتبار سنجی به صورت تصادفی در نظر گرفته شدند. پس از آموزش شبکه دادگان آزمون به عنوان ورودی به شبکه اعمال و خروجی آن با استفاده از شبکه عصبی آموزش دیده به دست آمد.

1. نتایج آزمایشگاهی و تحلیل آن‌

در این پژوهش با استفاده از ویژگی‌ها و میل اتصال 96 آپتامر به ارائه شبکه عصبی مناسبی جهت پیش‌بینی میزان میل اتصال آپتامر-ویروس آنفولانزا پرداختیم. از این مجموعه، 72 آپتامر به عنوان آموزش و 24 آپتامر به عنوان آزمون استفاده شده است. ویژگی‌های هر آپتامر توصیف‌گرهایی هستند که نشان دهنده خواص فیزیکوشیمیایی آن‌ها می‌باشند. در ابتدا شبکه عصبی GRNN مورد بررسی قرار داده شد. بدین منظور ابتدا خروجی GRNN بدست آمد و در صفحه‌ای برحسب مقادیر آزمایشگاهی ترسیم و خط y=x بر روی این نقاط برازش شد. برای ارزیابی عملکرد این مدل مقدار r2 به دست آمده از این برازش، مدنظر قرار گرفت. برای مدل GRNN مقدار r2=0.64 بدست آمد. در مرحله بعد با تغییر شبکه عصبی به MLP و تکرار همین مراحل مقدارr2=0.77 حاصل شد. شکل 3 و 4 نمودار میل اتصال آپتامر بر حسب مقادیر واقعی آن برای دو شبکه‌ی GRNN و MLP‌ وبرای دادگان آموزش و آزمون رسم شده است. علاوه بر این خط برازش شده y=x به این دادگان به همراه مقدار r2 نمایش داده شده است.

در مطالعه مشابه با بکارگیری روش MLR برای پیش‌بینی میل اتصال آپتامر بر اساس همین توصیف‌گر‌ها مقدار r2=0.66 گزارش شده است]7[. با توجه به اینکه شبکه عصبی MLP قابلیت بازنمایی تابع غیرخطی بیانگر ارتباط بین ورودی و خروجی را داراست، توانسته است به صورت دقیق‌تری نسبت به دو روش دیگر مقدار میل اتصال آپتامر به ویروس آنفولانزا را پیش‌بینی کند. جدول ۱ نتایج حاصل از روش‌های بکارگرفته شده در این مقاله را با روش‌های مطالعات دیگر مقایسه می‌کند.

**جدول1**- مقابسه نتایج حاصل از این مطالعه با مطالعات پیشین

|  |  |
| --- | --- |
| روش | دقت پیش‌بینی(R2) |
| رگرسیون چند خطی ]7[ | **۶۶/۰** |
| شبکه‌عصبیGRNN | **۶۴/۰** |
| شبکه عصبیMLP | **۷۷/۰** |



**شکل3**- نمودار binding score محاسبه شده توسط شبکه عصبی GRNN به مقدار واقعی آن برای (الف) دادگان تست (ب) دادگان آموزش



**شکل4**- نمودار میل اتصال محاسبه شده توسط شبکه عصبی MLP برحسب مقدار واقعی آن برای (الف) دادگان تست (ب) دادگان آموزش

1. **جمع‌بندی**

گفتنی است، تا به امروز دستاوردهای بسیار خوبی در مورد انتخاب، اصلاحات و کاربرد آپتامرها حاصل شده است. با این حال، تعداد کمی از محصولات مبتنی بر آپتامر با موفقیت در استفاده بالینی و صنعتی وارد شده‌اند. علاوه بر این، هنوز هم به دست آوردن شیوه‌ی کارامد و کم هزینه‌ای برای تجزیه و تحلیل آپتامرها چالشی بزرگ می‌باشد. بنابراین، بررسی کامل کمبود فعلی و دستیابی به فناوری و روش مناسب در تحقق این امر بسیار مهم است. با اين وجود، براي ورود بهتر و سريع‌تر آپتامرها به اين حوزه ها بايد گام‌هايي در جهت ايجاد روش‌هاي سريع‌تر و ارزان‌تر براي بررسی اتصال آپتامر-ویروس انجام شود. برای تحقق این امر می‌توانیم از روش‌های هوش‌مصنوعی و شبکه‌های عصبی استفاده نماییم که در این پژوهش با استفاده از شبکه‌های عصبی میزان میل اتصال آپتامرها-ویروس‌آنفولانزا بررسی شده است. ورودی این شبکه تعداد 96 آپتامر به همراه میزان میل اتصال آن‌ها به ویروس آنفولانزا می‌باشد.این مجموعه از آپتامرها به دو مجموعه تست (72 آپتامر) و مجموعه آموزش (24آپتامر) تقسیم شده‌اند. در این مجموعه داده تعداد 9 ویژگی یا توصیف­گر که بیشترین تاثیر را در بررسی میزان اتصال آپتامر و ویروس آنفولانزا دارند، مورد استفاده قرار گرفته‌اند و در شبکه عصبی GRNN مقدار r2=0.64 بدست آمده است. برای بهبود نتایج و ارتقاء مدل از شبکه عصبی پرسپترون چندلایه استفاده شده که قابلیت بازنمایی تابع غیرخطی بیانگر ارتباط بین ورودی و خروجی را داراست، توانسته است به صورت دقیق‌تری نسبت به دو روش دیگر مقدار میل اتصال آپتامر به ویروس آنفولانزا را پیش‌بینی کند و به r2=0.77 دست یابیم. علاوه بر این برای بدست آوردن نتایج بهتر می‌توانیم از استخراج ویژگی‌های دیگری که کیفیت این پیش‌بینی را بالا می‌برند استفاده و از آن‌ها برای بررسی میزان اتصال ویروس-آپتامر بهره بگیریم .

1. **مراجع**

[1] Yadav, M.P.; Singh, R.K.; Malik, Y. Epidemiological Perspective in Managing Viral Diseases in Animals. In Recent Advances in Animal Virology; Malik, Y., Yadav, S.R.M., Eds.; Springer: Singapore, 2019.

[2] The Latest Achievements in the Construction of Influenza Virus Detection Aptasensors Ewelina W˛edrowska 1,\*, Tomasz Wandtke 1 , El ˙zbieta Piskorska 2 and Piotr Kopi ´nski 1,\*

[3] Dziabowska, K.; Czaczyk, E.; Nidzworski, D. Detection Methods of Human and Animal Influenza Virus-Current Trends. Biosensors 2018, 8, 94. [CrossRef]

[4] Reinholt, Sarah J., and Harold G. Craighead. "Microfluidic device for aptamer-based cancer cell capture and genetic mutation detection." Analytical chemistry 90.4 (2018): 2601-2608.

[5] Ouyang, Changhe, et al. "Precision-guided missile-like DNA nanostructure containing warhead and guidance control for aptamer-based targeted drug delivery into cancer cells in vitro and in vivo." Journal of the American Chemical Society 142.3 (2020): 1265-1277.

[6] Blank M, Weinschenk T, Priemer M, Schluesener H. Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels: Selective targeting of endothelial regulatory protein pigpen. J Biol Chem. 2001; 276(19): 16464-8

[7] Musafia, Boaz, Rony Oren-Banaroya, and Silvia Noiman. "Designing anti-influenza aptamers: novel quantitative structure activity relationship approach gives insights into aptamer–virus interaction." PLoS One 9.5 (2014): e97696.

[8] Yu, X., et al. "Prediction of the binding affinity of aptamers against the influenza virus." SAR and QSAR in Environmental Research 30.1 (2019): 51-62.

[9] Wang, Yu, et al. "Rapid Detection and Prediction of Influenza A Subtype using Deep Convolutional Neural Network based Ensemble Learning." Proceedings of the 2020 10th International Conference on Bioscience, Biochemistry and Bioinformatics. 2020.

[10] Wisesty, Untari Novia, Febryanti Sthevanie, and Rita Rismala. "Momentum Backpropagation Optimization for Cancer Detection Based on DNA Microarray Data." International Journal of Artificial Intelligence Research 4.2 (2020): 127-134.

Prediction of specific binding affinity of aptamers to the influenza virus using Neural Network

Fatemeh Shoaa1, \*Amin Janghorbani1, Ahmad Farhad Talebi1  
  
1 Faculty of New Sciences and Technologies, Seman University, Semnan, Iran

\*a.janghorbani@semnan.ac.ir

*Abstract*— The influenza virus kills many people and animals worldwide each year and causes significant economic and social damages. Developing reliable, easy-to-use, high-sensitive diagnostic methods will help us to control epidemics and their side effects. In this vein, the proposing of new molecular methods based on aptamers has been considered due to their high potential in the diagnosis, control, and even treatment of influenza infection. Thanks to artificial intelligence, now, it is easy to evaluate the effectiveness of aptamers in the detection of influenza viruses. This study applied neural networks to predict aptamers’ binding score to the influenza viruses using nine descriptors. In this study, GRNN model predictions matched the experimental binding scores with r2=0.66 and MLP with r2=0.77. These results indicate that MLP neural network outperforms the multilinear regression method applied in the previous study on the same dataset.

Keywords— Aptamer, Influenza Virus, Binding Affinity, Prediction, Artificial Intelligence, Neural Networks,

1. Systematic Evolution of Ligands by Exponential enrichment [↑](#footnote-ref-1)
2. QSPRs3 [↑](#footnote-ref-2)
3. Quantitative structure-activity relationship [↑](#footnote-ref-3)
4. descriptors [↑](#footnote-ref-4)
5. [General regression neural network](https://en.wikipedia.org/wiki/General_regression_neural_network) [↑](#footnote-ref-5)
6. [Multilayer perceptron](https://en.wikipedia.org/wiki/Multilayer_perceptron) [↑](#footnote-ref-6)
7. pattern [↑](#footnote-ref-7)
8. summation [↑](#footnote-ref-8)
9. denominator [↑](#footnote-ref-9)
10. numerator [↑](#footnote-ref-10)
11. feedforward [↑](#footnote-ref-11)
12. Input Layer [↑](#footnote-ref-12)
13. Hidden Layer [↑](#footnote-ref-13)
14. Output layer [↑](#footnote-ref-14)
15. backpropagation [↑](#footnote-ref-15)