

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

فاطمه شعاع‌اردبیلی<sup>۱</sup>، امین جانقربانی‌پوده<sup>۱\*</sup> و احمدفرهاد طالبی<sup>۱</sup>

<sup>۱</sup> 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 مواجهه 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 الیگونوکلوئیدهای 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 مدل سازی شده است. مجموعه داده این مقاله شامل ۹ ویژگی‌های ساختار مولکولی ۹۶ 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 داده‌گان آموزش دقت پیش‌بینی میل اتصال در شبکه عصبی GRNN برای داده‌گان آزمون و براساس معیار  $r^2$  مقدار 0.64 بدست آمده است. همین فرآیند برای شبکه عصبی MLP تکرار شد عملکرد این شبکه در پیش‌بینی میل اتصال در داده‌گان آزمون  $r^2=0.77$  حاصل شد. نتایج حاصل از این روش‌ها در مقایسه با روش رگرسیون چند خطی با مقدار  $r^2=0.66$  نشان از برتری مدل MLP در پیش‌بینی میل اتصال آپتامرها به ویروس آنفولانزا است.

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

### ۱- 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 سال ۲۰۲۰، 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 [۱]. با بهره‌گیری از روش‌های هوش مصنوعی می‌توان در تشخیص سریع و به موقع این تهدیدات موثری انجام داد. ویروس آنفولانزا اغلب دارای ساختار کروی زمخت با قطری در حدود ۰.۱-۰.۲۱ نانومتر می‌باشد اما گاهی به شکل رشته ای نیز دیده می‌شود. در ساختار مرکزی ویروس RNA تک رشته‌ای با پلارینه 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 [۲]. 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 [۳].

### ۱-۱- روش‌های تشخیص ویروس آنفولانزا

روش‌های 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 زیرگونه‌های نوع A برخوردار هستند.

این تست‌ها قابلیت ردیابی نوکلئوپروتئین‌های ویروسی را دارند و از جمله ویژگی‌های آن‌ها اختصاصیت ۹۹٪ و اعلام نتیجه در کمترین مدت زمان ممکن است. این تست‌ها را بر اساس قابلیت افتراقی‌شان به سه دسته زیر تقسیم می‌کنند:

۱- تست‌های که آنفولانزا نوع A را تشخیص می‌دهند

۲- تست‌های که دو نوع A و B را از یکدیگر تشخیص و افتراق می‌دهند

۳- 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 این کار می‌توانیم از آپتامرها بهره بگیریم.

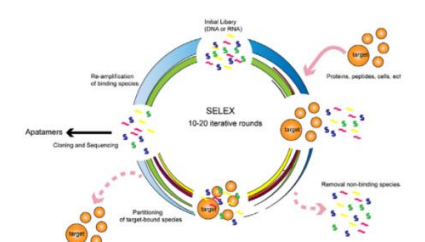
## ۲-۱- آبتامر

آبتامرها توالی‌های تک رشته‌ای سنتزی RNA یا DNA (اخیراً، پتیدی) هستند که به ساختارهای فضایی شکل می‌دهند و با اختصاصیت فوق العاده زیاد به اهداف معینی متصل می‌گردند. اولین بار آبتامرها در سال ۱۹۹۰ ارائه شدند، خصوصیات بی‌نظیر آن‌ها باعث شده است که موثرتر از آنتی‌بادی‌ها عمل نمایند. آبتامرها عموماً از طریق فرایند آزمایشگاهی SELEX<sup>۱</sup> از یک کتابخانه، غربال و انتخاب می‌شوند که می‌توانند به هر مولکول هدفی (از یون‌های معدنی کوچک تا سلول‌های کامل) متصل شوند. آبتامرها زمانی که انتخاب شدند، طی فرایند PCR تکثیر شده تا مقادیر زیادی از آن‌ها با خلوص بالا تهیه شود. ساختار شیمیایی ساده آبتامرها آن‌ها را پذیرای اصلاحات کاربردی بیشتر با توجه به هدف‌های مختلف می‌نماید. همچنین آبتامرها بسیار پایدارتر از آنتی‌بادی‌ها هستند که آن‌ها را مناسب و سازگار برای شرایط دشوار (مثل بالا بودن دما) می‌نماید. مطالعات و پژوهش‌های روزافزون در زمینه کاربردهای آبتامرها همانند تشخیص و درمان در حال انجام است. احتمالاً در آینده‌ای نزدیک آبتامرها استفاده فزاینده‌ای هماهنگ با دیگر مولکول‌های درمانی پیدا کند [۴]. علاوه بر کاربردهای آبتامرها مانند ابراز تشخیصی، درمانی، تصویربرداری زیستی و تحویل دارو معرفی شده است [۵]. با توجه به پتانسیل بالای آبتامرها در کنترل و مهار ویروس آنفولانزا و با توجه به اینکه روش‌های موجود در حوزه شبکه‌های عصبی مانند توانایی پیش‌بینی اتصال آبتامرها انتخاب شده با ویژگی و میل زیاد، استخراج توالی و ویرایش ژن‌ها، بسیار کارآمد هستند. شبکه‌های 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 آبتامرها آسان‌تر بکنند. و با صرف زمان کوتاه می‌توان نتایج ارزشمندی بدست آورد. می‌توانیم با استفاده از ویژگی‌های استخراج شده از آبتامرها و ویروس آنفولانزا و بکارگیری آن‌ها در حوزه هوش مصنوعی پیشرفت‌های بسیاری در علم پزشکی حاصل کنیم و از وقوع بیماری‌های خطرناکی که امکان مرگ بیمار را دارند، جلوگیری کنیم.

## ۳-۱- فرایند انتخاب آبتامر تحت عنوان SELEX

فرایند SELEX شامل چندین چرخه انتخاب و تقویت است: الف) مجموعه‌ای از مولکول‌های RNA یا DNA نمایش داده می‌شود و آبتامرهایی که دارای قدرت اتصال به یک مولکول هدف هستند از غیر آبتامرها جدا شده و ب) آبتامرهای نگه‌دارنده با استفاده از واکنش زنجیره‌ای پلیمرز (PCR) دنباله‌ها برای دور بعدی غنی‌سازی تقویت می‌شوند. کل مراحل انتخاب بطور معمول به ۱۵ دور انتخاب نیاز دارد و می‌تواند از چند روز تا چند ماه طول بکشد. در مرحله بعدی، این کتابخانه تصادفی DNA یا RNA در مجاورت مستقیم مولکول هدف قرار گرفته و در ادامه، کمپلکس‌های ایجادشده با هدف، از الیگونوکلوئیدهای باند نشده (مرحله Partition) جدا می‌شوند. این مرحله یکی از سخت‌ترین مراحل SELEX است و بطور قابل ملاحظه‌ای بر خصوصیات اتصال آبتامر به هدف، اثر می‌گذارد. در ادامه الیگونوکلوئیدهای متصل شده به هدف، از هدف جداسازی شده (شستشو داده می‌شوند) و توسط PCR (برای SELEX DNA) و RT-PCR (برای SELEX RNA) تکثیر می‌یابند. خروجی فرایند PCR، dsDNA است و می‌بایست به ssDNA تبدیل شود [۶]. و توالی مکمل آبتامر انتخاب شده، حذف گردد. حال توالی تک‌رشته‌ای باقی‌مانده وارد چرخه دوم SELEX شده و به‌طور مجدد در مجاورت هدف قرار می‌گیرد. به دلیل تکرار شونده بودن این چرخه انتخاب و تکثیر الیگونوکلوئیدها در هر مرحله، استخراج توالی در هر مرحله دستخوش کاهش تنوع الیگونوکلوئید قرار می‌گردد؛ در حالی که میل ترکیبی الیگونوکلوئیدهای باقی‌مانده، رفته رفته افزایش می‌یابد. تعداد دفعات چرخه‌ی (راند) مورد نیاز در SELEX به پارامترهای مختلفی از جمله خصوصیات هدف و غلظت آن، طراحی کتابخانه اولیگونوکلوئیدی آغازگر، شرایط محیطی انتخاب، نسبت مولی هدف به اولیگونوکلوئیدها و کارایی روش جداسازی (Partitioning) وابسته است. راندهای اضافی منحصراً با توجه به میزان اختصاصیت اولیگونوکلوئیدها امکان پذیر است.

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



شکل ۱- شمایند فرایند SELEX برای کتابخانه‌های RNA و DNA [۶]

## ۴-۱- کاربرد هوش مصنوعی در طراحی آبتامرها

در مطالعه‌ی توان اتصال آبتامرها به پروتئین هم‌گلوئین ویروس آنفولانزا در سال ۲۰۱۴، شش خصوصیت فضایی و فیزیکوشیمیایی آبتامرها برای این نوع اتصال اختصاصی به کمک محاسبات هوش مصنوعی مهم تشخیص داده شدند. در واقع محاسبات انجام شده توسط کامپیوتر، آبتامرهایی را طراحی و پیشنهاد کردند که تا ۲۰ برابر مؤثرتر به ذرات ویروسی متصل می‌شدند [۷]. در مطالعات مشابهی تحقیقات بسیاری روی روابط ساختار-فعالیت کمی<sup>۲</sup> و روابط ساختار-ویژگی کمی انجام و گزارش شده است. با این حال مقالات کمی در مورد QSARs<sup>۳</sup> برای آبتامرها منتشر شده است که علت آن هم دشواری بدست آوردن توصیف‌گرها<sup>۴</sup> دو بعدی و سه بعدی مستقیماً از آبتامرها می‌باشد. در مطالعه‌ای با استفاده از توالی آمینواسیدهایی که از توالی آبتامر و ارتقاء مدل‌های QSAR که مربوط به تمایل اتصال آبتامرها به ویروس آنفولانزا هستند، به توصیف روش بدست آوردن توصیف‌گرهای مولکولی پرداخته شده است و از شبکه عصبی رگرسیون عمومی بر مبنای تخمین پنجره‌های Parzen برای ساخت مدل QSAR با اعمال شش توصیف‌گر مولکولی استفاده شده است [۸].

1 Systematic Evolution of Ligands by Exponential enrichment

2 QSARs

3 Quantitative structure-activity relationship

4 descriptors

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 دست‌یافت [۹].

برای بررسی بهتر و کارآمدتر نحوه اتصال و کیفیت اتصال آیتامرها و ویروس‌ها می‌توان از روش استخراج ویژگی استفاده نمود تا با در دسترس داشتن ویژگی‌های مهمی که بیشترین تاثیر را در اتصال آیتامر-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 آیتامر به ویروس آنفولانزا پیش‌بینی شوند.

## ۲- روش‌ها و ابزارها

### ۲-۱- مجموعه داده

در این پژوهش از مجموعه داده مقاله‌ای سال ۲۰۱۴ که در رابطه با پیش‌بینی میل اتصال آیتامر به ویروس آنفولانزا بر اساس روش‌های رگرسیون آماری منتشر شده است بهره گرفته شد [۷]. در مجموعه داده این مقاله شامل ۹۹ توالی آیتامر به همراه تمایل اتصال به ویروس آنفولانزا نمایش داده شده است. که این مجموعه از آیتامرها را به دو مجموعه آموزش (۷۲ آیتامر) و مجموعه آزمون (۲۴ آیتامر) تقسیم کرده‌اند. در این مجموعه داده تعداد ۹ ویژگی یا توصیف‌گر براساس ویژگی‌های ساختار مولکولی که بیشترین تاثیر را در بررسی میزان اتصال آیتامر و ویروس آنفولانزا را دارند، ارائه شده است. (جدول ۱) که به عنوان ورودی‌های مدل پیش‌بینی میل اتصال مورد استفاده قرار گرفته است.

جدول ۱- معرفی توصیف‌گرهای مولکولی استفاده شده در [۷]

شماره ردیف	توضیح توصیف‌گر	محدوده تغییرات توصیف‌گرها
۱	تعداد کل بازها	میانگین ۵۵.۴۱، بیشینه ۶۹، کمینه ۳۰، انحراف معیار ۹.۱۶
۲	تعداد انطباقات محاسبه شده	میانگین ۱۰.۸، بیشینه ۴، کمینه ۰، انحراف معیار ۰.۵۹
۳	طول توالی‌های C جفت نشده که با سایر توالی‌های کمتر از ۸ نوکلئوتیدی قطع نشده باشد	میانگین ۱۷.۱۷، بیشینه ۵۵، کمینه ۰، انحراف معیار ۱۵.۷۸
۴	مجموع نوکلئوتیدهای جفت نشده C با طول سه یا بیشتر	میانگین ۱۷.۰۴، بیشینه ۵۵، کمینه ۰، انحراف معیار ۱۶.۱۰
۵	مجموع تمام نوکلئوتیدهای جفت نشده	میانگین ۳۳.۲۴، بیشینه ۵۵، کمینه ۴، انحراف معیار ۹.۰۳
۶	اندازه بزرگترین حلقه	میانگین ۲۲.۲۱، بیشینه ۳۹، کمینه ۰، انحراف معیار ۱۲.۵۸
۷	نسبت بین نوکلئوتیدهای جفت نشده به کل	میانگین ۰.۶۱۳، بیشینه ۱، کمینه ۰.۰۵۹، انحراف معیار ۰.۱۷۰
۸	نسبت بین بزرگترین حلقه به کل	میانگین ۰.۴۰۹، بیشینه ۰.۷۷۸، کمینه ۰، انحراف معیار ۰.۲۴۰
۹	نسبت بین طول توالی‌های C جفت نشده به کل	میانگین ۰.۳۲۵، بیشینه ۱.۰۰۰، کمینه ۰، انحراف معیار ۰.۳۰۱

### ۲-۲- شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی، به عنوان یک الگوریتم قدرتمند، به طور وسیعی در رگرسیون غیرخطی و مدل‌سازی ارتباط بین متغیرهای توصیف‌گر و متغیر پاسخ بهره گرفته شده است. در این پژوهش به منظور مدل‌سازی ارتباط بین ویژگی‌های آیتامر با میل اتصال آن‌ها به ویروس آنفولانزا از شبکه‌های GRNN<sup>۵</sup> و MLP<sup>۶</sup> بهره گرفته شده است.

#### 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<sup>۷</sup>، 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 یک نورون الگو فاصله اقلیدسی نمونه تست را از نقطه مرکزی نورون 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<sup>۸</sup> از دو نوع گره تشکیل شده است: گره‌های مخرج<sup>۹</sup> و گره‌های صورت<sup>۱۰</sup> 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  $\sigma$  (یعنی پارامتر صاف یا ضریب عرض) تابع گاوسی (که D نشان دهنده یک تابع اسکالر است) باید تنظیم شود. برای مقدار  $\sigma$  بزرگتر، نمایش احتمالی نقطه نمونه ارزیابی شده برای دامنه وسیع‌تری از X امکان‌پذیر است. برای مقدار  $\sigma$  کوچک، نمایش به محدوده باریکی از X محدود می‌شود.

5 General regression neural network

6 Multilayer perceptron

7 pattern

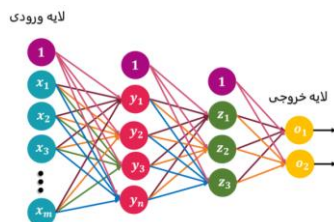
8 summation

9 denominator

10 numerator

## MLP

شبکه عصبی خود را تغییر داده و از شبکه عصبی MLP استفاده کردیم. پرسپترون چند لایه یک شبکه عصبی مصنوعی پیشرو<sup>۱۱</sup> است که مجموعه‌ای از خروجی‌ها را از مجموعه ورودی‌ها تولید می‌کند. در یک شبکه عصبی پرسپترون چند لایه، حداقل سه لایه وجود خواهند داشت که این لایه‌ها عبارتند از یک لایه ورودی<sup>۱۲</sup>، یک لایه پنهان<sup>۱۳</sup> و یک لایه خروجی<sup>۱۴</sup>. آموزش شبکه عصبی پرسپترون چند لایه توسط الگوریتم یادگیری پس انتشار خطا<sup>۱۵</sup> انجام می‌شود [۱۰]. ساختار پرسپترون چند لایه در شکل ۲ نشان داده شده است:



شکل ۲- شماتیک شبکه عصبی MLP

به این شبکه ورودی‌های یکسان با شبکه عصبی GRNN اعمال گردید. شبکه عصبی MLP دولایه پنهان با تعداد نورون ۸ در لایه اول و تعداد نورون ۱۰ 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.

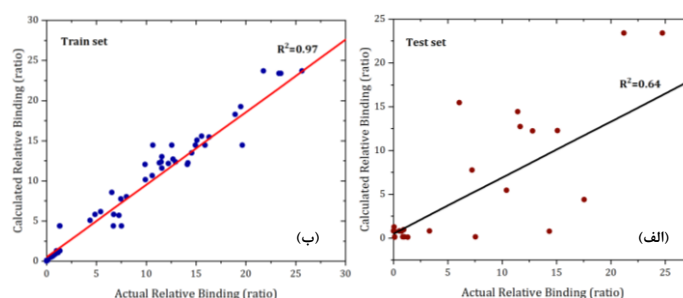
## ۳- نتایج آزمایشگاهی و تحلیل آن

در این پژوهش با استفاده از ویژگی‌ها و میل اتصال ۹۶ آبتامر به ارائه شبکه عصبی مناسبی جهت پیش‌بینی میزان میل اتصال آبتامر-ویروس آنفولانزا پرداختیم. از این مجموعه، ۷۲ آبتامر به عنوان آموزش و ۲۴ آبتامر به عنوان آزمون استفاده شده است. ویژگی‌های هر آبتامر توصیف‌گرهایی هستند که نشان دهنده خواص فیزیکی-شیمیایی 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  $r^2$  به دست آمده از این برازش، مدنظر قرار گرفت. برای مدل GRNN مقدار  $r^2=0.64$  بدست آمد. در مرحله بعد با تغییر شبکه عصبی به MLP و تکرار همین مراحل مقدار  $r^2=0.77$  حاصل شد. شکل ۳ و ۴ نمودار میل اتصال آبتامر بر حسب مقادیر واقعی آن برای دو شبکه‌ی GRNN و MLP و برای دادگان آموزش و آزمون رسم شده است. علاوه بر این خط برازش شده  $y=x$  به این دادگان به همراه مقدار  $r^2$  نمایش داده شده است.

در مطالعه مشابه با بکارگیری روش MLR برای پیش‌بینی میل اتصال آبتامر بر اساس همین توصیف‌گرها مقدار  $r^2=0.66$  گزارش شده است [۷]. 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 آبتامر به ویروس آنفولانزا را پیش‌بینی کند. جدول ۱ نتایج حاصل از روش‌های بکار گرفته شده در این مقاله را با روش‌های مطالعات دیگر مقایسه می‌کند.

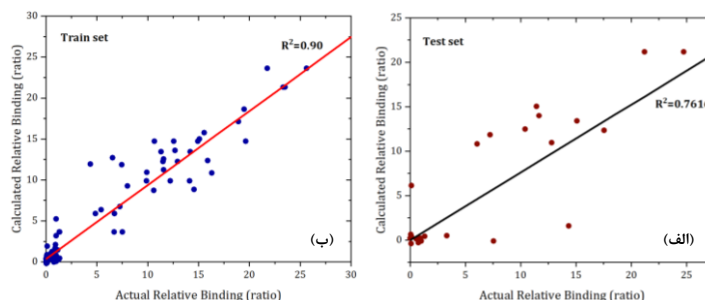
جدول ۱- مقایسه نتایج حاصل از این مطالعه با مطالعات پیشین

روش	دقت پیش‌بینی ( $R^2$ )
رگرسیون چند خطی [۷]	۰/۶۶
شبکه عصبی GRNN	۰/۶۴
شبکه عصبی MLP	۰/۷۷



شکل ۳- نمودار binding score محاسبه شده توسط شبکه عصبی GRNN به مقدار واقعی آن برای (الف) دادگان تست (ب) دادگان آموزش

<sup>11</sup> feedforward  
<sup>12</sup> Input Layer  
<sup>13</sup> Hidden Layer  
<sup>14</sup> Output layer  
<sup>15</sup> backpropagation



شکل ۴- نمودار میل اتصال محاسبه شده توسط شبکه عصبی MLP برحسب مقدار واقعی آن برای (الف) دادگان تست (ب) دادگان آموزش

#### ۴- جمع‌بندی

گفتنی است، تا به امروز دستاوردهای بسیار خوبی در مورد انتخاب، اصلاحات و کاربرد آبتامرها حاصل شده است. با این حال، تعداد کمی از محصولات مبتنی بر آبتامر با موفقیت در استفاده بالینی و صنعتی وارد شده‌اند. علاوه بر این، هنوز هم به دست آوردن شیوهی کارآمد و کم هزینه‌ای برای تجزیه و تحلیل آبتامرها چالشی بزرگ می‌باشد. بنابراین، بررسی کامل کمبود فعلی و دستیابی به فناوری و روش مناسب در تحقق این امر بسیار مهم است. با این وجود، برای ورود بهتر و سریع‌تر آبتامرها به این حوزه ها باید گام‌هایی در جهت ایجاد روش‌های سریع‌تر و ارزان‌تر برای بررسی اتصال آبتامر-ویروس انجام شود. برای تحقق این امر می‌توانیم از روش‌های هوش مصنوعی و شبکه‌های عصبی استفاده نماییم که در این پژوهش با استفاده از شبکه‌های عصبی میزان میل اتصال آبتامر-ویروس آنفولانزا بررسی شده است. ورودی این شبکه تعداد ۹۶ آبتامر به همراه میزان میل اتصال آن‌ها به ویروس آنفولانزا می‌باشد. این مجموعه از آبتامرها به دو مجموعه تست (۷۲ آبتامر) و مجموعه آموزش (۲۴ آبتامر) تقسیم شده‌اند. در این مجموعه داده تعداد ۹ ویژگی یا توصیف‌گر که بیشترین تاثیر را در بررسی میزان اتصال آبتامر و ویروس آنفولانزا دارند، مورد استفاده قرار گرفته‌اند و در شبکه عصبی GRNN مقدار  $R^2=0.64$  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 آبتامر به ویروس آنفولانزا را پیش‌بینی کند و به  $R^2=0.77$  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-آبتامر بهره بگیریم.

#### ۵- مراجع

- [۱] 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.
- [۲] The Latest Achievements in the Construction of Influenza Virus Detection Aptasensors Ewelina Węedrowska 1,\* , Tomasz Wandtke 1 , El 'zbieta Piskorska 2 and Piotr Kopi 'nski 1,\*
- [۳] Dziabowska, K.; Czarczyk, E.; Nidzworski, D. Detection Methods of Human and Animal Influenza Virus-Current Trends. Biosensors 2018, 8, 94. [CrossRef]
- [۴] 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.
- [۵] 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.
- [۶] 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
- [۷] 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.
- [۸] 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.
- [۹] 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.
- [۱۰] 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 Shooa<sup>1</sup>, \*Amin Janghorbani<sup>1</sup>, Ahmad Farhad Talebi<sup>1</sup>

<sup>1</sup> 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  $r^2=0.66$  and MLP with  $r^2=0.77$ . 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,