

شناسایی ژن‌های عامل سرطان در شبکه ژنی با استفاده از معماری شبکه عصبی پیش‌خور

مصطفی اخوان صفار و عباسعلی رضایی

هر کدام از این روش‌ها به تعیین جهش‌های سرطانی کمک می‌نمایند؛ اما چالش اصلی این است که در بین روش‌های ارائه‌شده برای تعیین ژن‌های عامل سرطان، کدام یک از آنها نتایج دقیق‌تری را نشان می‌دهند. روش‌های پیشنهادی برای شناسایی ژن‌های عامل سرطان به دو دسته کلی روش‌های محاسباتی و شبکه‌ای قابل تقسیم است. اکثر روش‌های پیشنهادی، ژن‌های عامل سرطان را از روی تعداد جهش‌های رخ داده و بررسی داده ژنومیک شناسایی می‌کنند که جزء روش‌های محاسباتی محسوب می‌شوند. از چالش‌ها و محدودیت‌های موجود در این روش‌ها آن است که به دلیل عدم قطعیت داده‌های زیستی، دارای مقدار زیادی نویز و داده‌های پرت هستند که این باعث ایجاد نرخ مثبت کاذب بالا در نتایج می‌شود. این روش‌ها اکثراً دارای محاسبات پیچیده و هزینه بالا و از نظر کارایی دارای میانگین هارمونیک پایین هستند. از این رو ایده استفاده از روش‌های مبتنی بر شبکه یادگیری ماشین و استفاده از ساختار شبکه برای یافتن ژن‌های عامل بیماری مطرح گردید تا برخی محدودیت‌های موجود در روش‌های محاسباتی مرتفع گردد. به‌عنوان مثال روش Simon از جمله روش‌های محاسباتی است که برای بهبود شناسایی ژن‌های عامل سرطان از تخمین جهش‌های پس‌زمینه استفاده می‌کند [۶]. این روش با سه معیار این کار را انجام می‌دهد؛ تأثیر عملیاتی جهش بر پروتئین‌ها، تغییر در میزان جهش پس‌زمینه در بین تومورها و افزودگی کد ژنتیکی. تمایز قائل شدن بین جهش‌های تأثیرگذار در عملکرد پروتئین و دیگر جهش‌ها و نیز تفاوت بین تعداد جهش پس‌زمینه در نمونه‌ها و بیماران مختلف، از ویژگی‌های این روش است. روش MDPFinder یکی دیگر از روش‌های شناسایی ژن‌های عامل سرطان می‌باشد [۷] که تلاش می‌کند تا مسئله ماتریس وزن بیشینه [۸] را که برای شناسایی مسیرهای درایور جهش‌یافته طراحی شده، حل کند. برای این کار از الگوریتم ژنتیک به همراه یک رویکرد تلفیقی برای یافتن مسیر جهش‌های سرطانی و سپس پیدا کردن ژن‌های عامل سرطان استفاده می‌کند. در این روش، داده‌های بیان ژن و جهش با یکدیگر ادغام می‌شوند. یکی دیگر از روش‌های محاسباتی Oncodrive-fm است [۹] که برای غلبه بر محدودیت‌های موجود در رویکردهای سنتی مانند سختی تخمین درست میزان جهش و متکی بودن به افزایش تغییرات در آن، معیاری با نام FM bias محاسبه می‌شود. روش MeMo یکی دیگر از روش‌های محاسباتی است که به بررسی نظام‌مند ماژول‌های مسیرهای آنکوژنیک^۱ می‌پردازد [۱۰] و در آن برای تشخیص ماژول‌ها از تحلیل همبستگی و آزمون‌های آماری استفاده می‌شود. روش DriverNet یک چهارچوب محاسباتی است که در آن جهش‌های سرطانی از طریق شبکه‌های بیان miRNA شناسایی می‌شوند

چکیده: شناسایی ژن‌های آغازگر سرطان یا عامل سرطان یکی از موضوعات تحقیقاتی مهم در زمینه سرطان‌شناسی و زیست‌داده‌ورزی است. ژن‌های عامل سرطان، ژن‌هایی هستند که بعد از اینکه جهش در آنها اتفاق می‌افتد، آن جهش را از طریق برهم‌کنش‌های پروتئین-پروتئین به دیگر ژن‌ها منتقل کرده و از این طریق، باعث اختلال در عملکرد سلول و بروز بیماری و سرطان می‌شوند. تا کنون روش‌های مختلفی برای پیش‌بینی و دسته‌بندی ژن‌های عامل سرطان پیشنهاد شده که اکثراً متکی به داده‌های ژنومی و ترنسکریپتومیک هستند و از این رو میانگین هارمونیک پایینی در نتایج دارند. تحقیقات در این زمینه به‌منظور بهبود دقت نتایج ادامه دارد و از این رو روش‌های مبتنی بر شبکه و زیست‌داده‌ورزی به کمک این حوزه آمده‌اند. در این مطالعه ما رویکردی را پیشنهاد داده‌ایم که متکی به داده‌های جهش نیست و از روش‌های شبکه‌ای برای استخراج ویژگی و از شبکه عصبی سه‌لایه پیش‌خور برای دسته‌بندی ژن‌ها استفاده می‌کند. برای این منظور، ابتدا شبکه زیستی مورد نظر که شبکه تنظیم رونویسی سرطان سینه است، تشکیل و سپس ویژگی‌های مختلف هر ژن به‌صورت بردارهایی استخراج گردید. نهایتاً بردارهای به‌دست‌آمده جهت دسته‌بندی به یک شبکه عصبی پیش‌خور داده شد. نتایج به‌دست‌آمده نشان می‌دهند که استفاده از روش‌های مبتنی بر شبکه‌های عصبی چندلایه می‌تواند صحت و میانگین هارمونیک را بهبود بخشد و باعث بهبود عملکرد نسبت به سایر روش‌های محاسباتی شود.

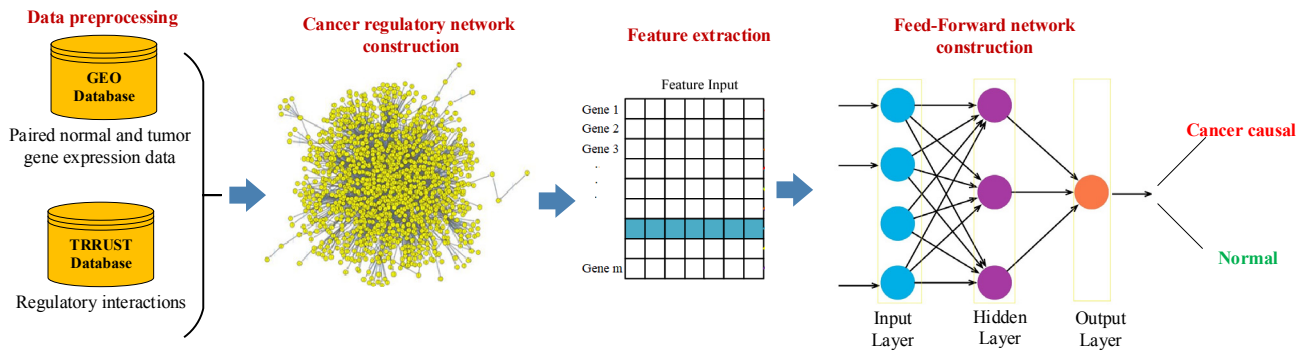
کلیدواژه: یادگیری عمیق، ژن‌های عامل سرطان، شبکه عصبی پیش‌خور، سرطان سینه.

۱- مقدمه

پیش‌بینی و دسته‌بندی ژن‌های عامل سرطان مولکولی در حوزه سرطان‌شناسی بسیار حائز اهمیت است [۱]. تا کنون تحقیقات مختلفی در خصوص یافتن ژن‌های عامل سرطان صورت گرفته [۲] تا [۴] که از طریق روش‌های مختلفی انجام می‌شوند که اکثراً از نوع محاسباتی و آماری هستند. با توجه به پژوهش‌های صورت‌گرفته، محققان بسیاری از فرایندی با عنوان توالی‌یابی ژنوم بهره می‌گیرند که در آن از طریق شبیه‌سازی جهش‌های ژنتیکی، ژن‌های مرتبط با بروز سرطان را بررسی می‌نمایند [۵]. این روش‌ها بسیار زمان‌بر و پرهزینه است و از این رو استفاده از روش‌های محاسباتی و آماری و تحلیل داده‌های ژنومیک و در ادامه استفاده از روش‌های زیست‌داده‌ورزی و شبکه‌ای پیشنهاد شده است.

این مقاله در تاریخ ۱ بهمن ماه ۱۴۰۱ دریافت و در تاریخ ۱۶ اردیبهشت ماه ۱۴۰۲ بازنگری شد.

مصطفی اخوان صفار (نویسنده مسئول)، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، تهران، (email: akhavansaffar@pnu.ac.ir).
عباسعلی رضایی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، تهران، (email: a_rezaee@pnu.ac.ir).



شکل ۱: روند کلی رویکرد پیشنهادی.

ژن‌های سرطانی می‌شوند باید هم‌زمان دو ویژگی داشته باشند؛ یکی اینکه دارای پوشش زیاد باشند و دیگری اینکه از نظر آماری، جهش رخ داده بین دو مسیر باید همبستگی بالایی داشته باشد. روش MSEA یک چهارچوب کلی و یکپارچه از طرز کار بیماری به‌جای بررسی مجزای داده ارائه می‌کند [۲۰]. این روش شامل یک خط لوله محاسباتی است که در آن از داده‌های چندبعدی مرتبط با بیماری، عملکردهای بیولوژیکی و شبکه‌های مولکولی برای بازیابی مسیرهای بیولوژیکی و نهایتاً شناسایی ژن‌های عامل سرطان استفاده می‌کند. تمام روش‌های ذکر شده متکی به داده‌های جهش و از نوع محاسباتی هستند. البته برخی روش‌ها مانند Netbox، DownRank، DriverNet، Memo و MSEA در کنار داده‌های ژنومیک و جهش از ساختار شبکه نیز کمک گرفته‌اند. برخی از روش‌ها که اخیراً پیشنهاد شده‌اند به داده‌های جهش متکی نبوده و تنها از ساختار شبکه استفاده می‌کنند که iMaxDriver-W و iMaxDriver-N از جمله این روش‌ها هستند [۲۱] که در آنها با استفاده از بیشینه‌سازی انتشار، ژن‌ها رتبه‌بندی می‌شوند. ژن‌های دارای رتبه بالاتر به‌عنوان عامل سرطان معرفی می‌گردند. روش GenHITS یکی دیگر از روش‌های شبکه‌ای است که در آن، ژن‌ها از طریق تحلیل ساختار برهم‌کنش‌های تنظیمی در شبکه تنظیم ژنی بر اساس مفهوم نفوذ در تئوری اطلاعات رتبه‌بندی می‌شوند [۲۲]. ژن‌های دارای رتبه بالاتر به‌عنوان عامل سرطان معرفی می‌گردند. روش rDDriver یکی دیگر از روش‌های شبکه‌ای است که با ترکیب شبکه عمیق و کانولوشنال به شناسایی ژن‌های تأثیرگذار در مجموعه پان‌کنسر می‌پردازد که در آن علاوه بر ژن‌های mRNA، دیگر نوع‌های RNA نیز شناسایی می‌شوند [۲۳]. در مطالعه لی و همکاران [۲۴] با استفاده از شبکه ژنی هر بیمار و یادگیری ماشین، روشی برای پیش‌بینی دقیق پیش‌آگهی سرطان پیشنهاد شده که از الگوریتم اصلاح‌شده پیچرنک برای محاسبه بردارهای ویژگی استفاده می‌کند.

با توجه اهمیت موضوع، تحقیقات در این خصوص با هدف بهبود عملکرد الگوریتم‌های پیشنهادی ادامه دارد. در این مطالعه ابتدا با استفاده از داده‌های بیان ژن و داده‌های تنظیمی و بدون نیاز به داده‌های جهش، شبکه ژنی مربوط به سرطان سینه، ایجاد و ویژگی‌های مربوط به ژن از طریق ساختار شبکه استخراج می‌شود. در ادامه با استفاده از بردارهای ویژگی به‌دست‌آمده، یک شبکه عصبی پیش‌خور به‌منظور دسته‌بندی ژن‌ها به دو دسته عادی و عامل سرطان یادگیری می‌شود. استفاده از معماری‌های شبکه عصبی چندلایه و یادگیری عمیق در شبکه تنظیم ژنی برای شناسایی ژن‌های عامل سرطان به‌تازگی مورد توجه قرار گرفته است. نمودار کلی رویکرد پیشنهادی در شکل ۱ آمده است. نتایج رویکرد پیشنهادی از نظر معیارهای کارایی با ۱۹ روش محاسباتی و شبکه‌ای پیشین مقایسه شدند.

[۱۱]. در این روش با استفاده از برهم‌کنش‌های بین ژنی، ارتباط بین انحرافات موجود در ژنوم و الگوهای رونویسی استخراج می‌گردند. روش Dendrix با ترکیب دو ویژگی پوشش یعنی پیدا کردن ژن‌ها در نمونه‌های مختلف بیماران و انحصاری بودن، یعنی جهش‌هایی که به‌ندرت در بیماران خاصی دیده می‌شود، جهش‌های سرطانی و غیرسرطانی را دسته‌بندی می‌کند [۱۲]. اغلب ژن‌های عامل سرطان از طریق جهش‌های پرتکرار در نمونه‌های مختلف تومور شناسایی می‌گردند؛ با این حال برخی جهش‌ها برای عملکرد پروتئین از دیگر جهش‌ها مهم‌تر تلقی می‌شوند. از این رو روش ActiveDriver با توجه به محل جهش و مکان عملکردی پروتئین می‌تواند مکانیسم آنها را پیش‌بینی کرده و تشخیص ژن‌های عامل سرطان را بهبود بخشد [۱۳]. روش iPAC یک رویکرد نظارت‌نشده است که در آن بر اساس آنالیز ترکیبی تعداد کپی‌ها و داده‌های بیان ژن، به‌طور نظام‌مند دنباله‌ای از آزمون‌های آماری بر روی فهرستی از ژن‌ها انجام می‌شود و از این طریق، لیست ژن‌های عامل سرطان به‌دست می‌آید [۱۴]. در روش MutsigCV با حل مسئله ناهمگونی در فرایندهای جهش و فرکانس جهش ژن‌ها تغییرات غیرعادی در ژن‌ها کشف و از این طریق، ژن‌های سرطانی شناسایی می‌شوند [۱۵]. در این روش از طریق بررسی فعالیت‌های رونویسی و مقایسه تعداد وقوع جهش در انواع سرطان و تعداد وقوع جهش در ژنوم انسان، سرطانی بودن ژن تشخیص داده می‌شود. روش OncodriveCLUST ژن‌هایی که تمایل زیادی برای خوشه‌بندی جهش در توالی پروتئین دارند، شناسایی می‌کند [۱۶] و از طریق ارزیابی جهش‌های خاموش‌کدکننده، یک مدل دسته‌بندی ژن‌ها را ایجاد می‌نماید. روش e-Driver یکی دیگر از روش‌های شناسایی ژن‌های عامل سرطان است [۱۷] که در آن، توزیع داخلی جهش‌های بین نواحی عملکردی پروتئین‌ها به‌منظور یافتن نرخ جهش در مقایسه با دیگر نواحی همان پروتئین استخراج می‌شود. در صورت مثبت بودن مشاهدات، آن ژن‌ها می‌توانند عامل بروز سرطان باشند. روش DawnRank با استفاده از داده‌های جهش و تمرکز روی ژن‌های سرطانی مخصوص هر بیمار، ژن‌های نادر و خاص هر بیمار را به‌عنوان ژن سرطانی کشف می‌کند [۱۸]. این روش از اطلاعات فردی تنها یک بیمار برای تشخیص ژن سرطانی استفاده می‌کند و در آن، ژن‌های جهش‌یافته در یک بیمار از طریق شبکه برهم‌کنش مولکولی رتبه‌بندی می‌شوند. نهایتاً ژن‌های با رتبه بالاتر به‌عنوان عامل سرطان معرفی می‌گردند. روش CoMDP که یکی دیگر از روش‌های محاسباتی است برخلاف سایر روش‌ها که بر روی ژن‌ها تمرکز دارند، ژن‌های عامل سرطان را با کمک مسیرهای مشترک جهش‌یافته شناسایی می‌کند [۱۹]. مسیرهای انتخابی که منجر به انتخاب

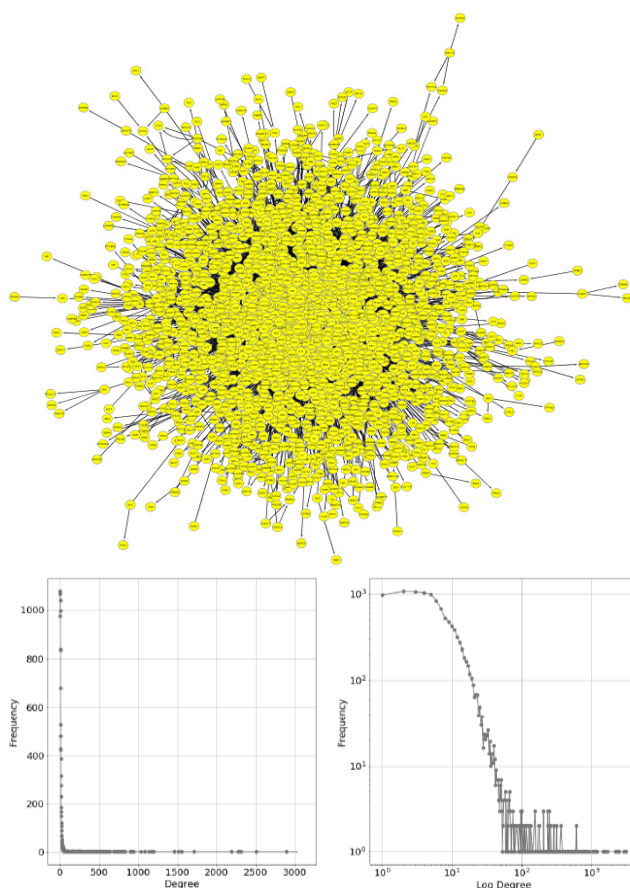
پایگاه داده GEO بارگیری گردید^۲. در این پایگاه داده، داده‌های بیان ژن هر سرطان به صورت مجزا برای بافت سرطانی و بافت سالم مجاز آن با فرمت CEL گزارش شده که این مجموعه داده‌ها برای هر سرطان توسط نرم‌افزار آنالیز تصاویر ریزآرایه^۳ ایجاد گردیده است. این فایل‌ها قبل از استفاده نیاز به پیش‌پرداز دارند. خروجی این پیش‌پردازش فایل حاوی داده‌های سطری است که در هر سطر شناسه ژن، مقدار بیان بافت نرمال و مقدار بیان بافت سرطانی مجاور آن می‌باشد. با استفاده از این مجموعه، داده‌های به دست آمده و لیست برهم‌کنش‌های تنظیمی شبکه مربوط، ساخته و برای هر شبکه، ژن مبدأ و مقصد در داده‌های بیان جستجو شد. در صورتی که هم مبدأ و هم مقصد دارای داده بیان ژن بودند یال مورد نظر، حفظ و در غیر این صورت آن برهم‌کنش از لیست نهایی حذف شد. تصویر نهایی شبکه ایجاد شده با نرم‌افزار ۳.۷.۲ Cytoscape مصورسازی گردید و در شکل ۲ آمده است. در این شکل، ساختار برهم‌کنش‌ها و نیز میانگین درجات هر ژن در شبکه برای نمایش بهتر به صورت لگاریتمی رسم شده که نشان می‌دهد شبکه از نوع توانی است.

Cytoscape یک پروژه نرم‌افزاری منبع باز برای یکپارچه‌سازی شبکه‌های برهم‌کنش مولکولی با داده‌های بیان و سایر حالات مولکولی در یک چارچوب مفهومی یکپارچه است. Cytoscape برای هر سامانه‌ای از اجزای مولکولی و برهم‌کنش‌ها قابل استفاده است، اما این نرم‌افزار، زمانی قدرتمندتر است که در ارتباط با پایگاه‌های داده بزرگ پروتئین-پروتئین، پروتئین-دنا و برهم‌کنش‌های ژنتیکی که به طور فزاینده‌ای برای انسان و موجودات در دسترس هستند، استفاده شود [۲۷].

۳- استخراج ویژگی‌ها

برای ساخت و آموزش یک مدل یادگیری نظارت شده برای دسته‌بندی ژن‌ها به بردار ویژگی برای هر ژن نیاز است. بردار ویژگی مربوط به هر ژن بر اساس معیارهای انتشار در شبکه جهت‌دار تنظیم ژنی محاسبه می‌شود. در کنار آن نیز از معیار زیستی اختلاف بیان ژن (DGE)^۴ در بافت سالم و سرطانی استفاده گردیده تا تأثیر تغییرات زیاد برخی ژن‌ها در بافت سالم و بافت سرطانی مجاور آنها در دسته‌بندی در نظر گرفته شود. این بردارهای ویژگی در مرحله بعد برای ساخت مدل مورد استفاده قرار می‌گیرند.

انتشار و تحلیل مسیر آن؟؟؟ از فرایندهای مهم در شبکه‌های اجتماعی است [۲۸]. هر چیزی که بتواند به صورت فیزیکی یا مجازی بین گره‌ها جابه‌جا شود می‌تواند در این فرایند در نظر گرفته شود. برخی از آنها به صورت آشکار رخ نمی‌دهند و تنها قادریم نتایج آن را مشاهده کنیم. مثلاً در گسترش بیماری، تنها افراد را که یکی پس از دیگری بیمار می‌شوند مشاهده می‌کنیم. برای درک آنکه چگونه این انتقال صورت گرفته است باید از طریق شبکه و الگوهای موجود، جریان تبادل اطلاعات را کشف کنیم. مدل‌های مختلفی برای این منظور وجود دارد که بر اساس آنها می‌توانیم اثرگذاری موجودیتی در شبکه را ارزیابی کنیم؛ از جمله برخی الگوریتم‌ها و مرکزیت‌های شبکه. در این مطالعه در کنار ویژگی DGE- یعنی اختلاف میزان بیان ژن در بافت سالم و سرطانی- از مرکزیت‌های پیچ رنگ، هیتز، تحلیل تصادفی لینک و کتر برای محاسبه ویژگی‌های مربوط به هر ژن از طریق شبکه استفاده گردید. هر کدام از الگوریتم‌های



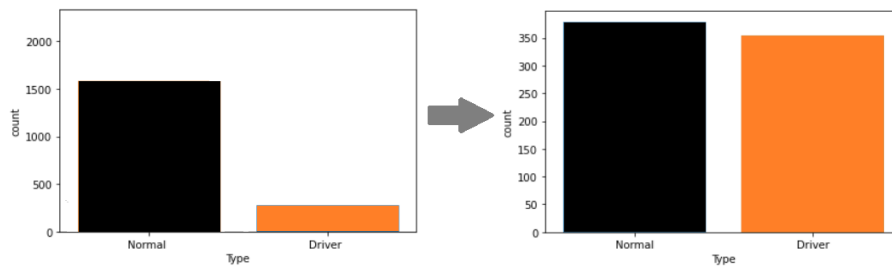
شکل ۲: شبکه ژنی سرطان سینه با ۲۴۹۹ ژن و ۷۵۴۰ برهم‌کنش.

۲- ساخت شبکه

شبکه‌های زیستی با توجه به نوع ژن‌ها و ارتباطات بین آنها در انواع مختلفی درون سلول وجود دارند. یکی از این نوع شبکه‌ها که در زمینه سرطان نیز مورد مطالعه قرار می‌گیرد شبکه تنظیم ژنی است که از داده‌های بیان ژن استخراج می‌شوند. در این شبکه جهت‌دار، گره‌ها شامل قطعات مختلف DNA مانند فاکتورهای رونویسی و RNAها می‌شوند. ارتباط بین این ژن‌ها در شبکه و ردوبدل کردن پیام بین آنها از طریق محصولات نهایی خودشان یعنی پروتئین‌ها انجام می‌شود. ژن‌ها با تأثیر مثبت یا منفی بر هم، باعث افزایش یا کاهش بیان یکدیگر می‌شوند و این تأثیرگذاری باید به گونه‌ای باشد که عملکرد سلول همیشه در وضعیت عادی باقی بماند. در این شبکه، ژن‌ها با تأثیر بر میزان بیان یکدیگر، باعث افزایش یا کاهش نرخ تولید پروتئین و در نتیجه تنظیم عملکرد سلول می‌شوند [۲۵]. ژن‌هایی که با ایجاد ناهنجاری، باعث تولید بیش از حد پروتئین و برهم‌زدن عملکرد تنظیمی سلول می‌شوند به عنوان عامل مؤثر در سرطان معرفی می‌گردند. در این مطالعه، شبکه تنظیم ژنی سرطان سینه با استفاده از لیست برهم‌کنش‌های تنظیمی و داده‌های بیان ژن، ساخته و لیست برهم‌کنش‌های تنظیمی از پایگاه داده v۲ TRRUST [۲۶] بارگیری گردید. این پایگاه داده شامل لیست برهم‌کنش‌های تنظیمی مربوط به شبکه‌های تنظیم ژنی انسان و موش است. لیست برهم‌کنش‌های مربوط به انسان شامل ۸۴۲۷ برهم‌کنش تنظیمی مربوط به ۲۸۶۷ ژن بارگیری شد. همچنین داده‌های بیان ژن سرطان سینه از

2. Download from <https://www.ncbi.nlm.nih.gov/geo>
 3. Affymetrix DNA Microarray Image
 4. Differential Gene Expression

1. Download from <https://www.gmpedia.org> in Aug. 27, 2020



شکل ۳: متوازن سازی کلاس های داده.

بیوانفورماتیک و یکی از مهم ترین پروژه ها در پایگاه داده های بیولوژیکی از حوزه زیست داده ورزی می باشد. این مجموعه برای سرعت دادن به درک محققین این حوزه از مبنای مولکولی سرطان با استفاده از فناوری های تجزیه و تحلیل ژنوم، از جمله توالی بندی ژنوم در مقیاس بزرگ تهیه شده است. مجموعه داده TCGA که حاوی بیش از دو پتابایت داده ژنومی می باشد، به صورت عمومی در دسترس است و این اطلاعات ژنومیکی به جامعه پژوهشی سرطان کمک می کند تا پیشگیری، تشخیص و درمان سرطان را بهبود بخشد.

به دلیل اینکه اکثر داده ها دارای برچسب نرمال بودند تا سرطانی، مجموعه داده ها دارای عدم تعادل کلاس ها بود. برای آنکه مدل به سمت کلاس اکثریت تمایل پیدا نکند و پیش بینی مدل جانب دارانه نباشد، مجموعه داده ها متوازن سازی شدند. برای رفع مشکل از روش کم نمونه گیری تصادفی^۳ (RUS) که یکی از روش های باز نمونه گیری در سطح داده می باشد، استفاده گردید و بدین ترتیب توازن دو کلاس ایجاد شد که در شکل ۳ آمده است.

۴-۱ تعریف معماری شبکه عصبی

برای ساخت مدل از یک شبکه سه لایه پیش خور^۴ با BatchNorm و Dropout استفاده گردید. ویژگی های به دست آمده برای هر ژن که در بخش ۳ توضیح داده شد به عنوان ورودی به شبکه عصبی وارد شدند. برای پیاده سازی شبکه عصبی مورد استفاده از زبان برنامه نویسی پایتون و کتابخانه پایتورچ^۵ برای ساخت مدل استفاده گردید. از تابع فعال سازی سیگموئید در لایه نهایی در طول آموزش استفاده نشد و به جای آن از تابع موجود در کتابخانه پایتورچ که به طور خودکار فعال سازی سیگموئید را انجام می دهد استفاده گردید.

۴-۲ تنظیم پارامترها

انتخاب پارامترهای مختلف مدل از جمله تعداد دوره ها و نرخ یادگیری از طریق انتخاب مقادیر مختلف و آزمایش مدل به منظور رسیدن به بهترین نتیجه، انجام و برای تابع هزینه از آنتروپی متقاطع^۶ استفاده شد. تابع هزینه، تفاوت نتیجه واقعی و نتیجه پیش بینی شده را مشخص می کند و با استفاده از این تابع هزینه، مقادیر وزن ها به روز رسانی می شوند و باعث می گردد که شبکه به سمت جواب های بهینه تر حرکت کند. زمانی که قصد داریم آنتروپی یک سری نمونه و بیت های متناظر آن را به حداکثر برسانیم، احتمال آماری آن نماد را در لگاریتم ضرب می کنیم تا تعداد بیت های بهینه آن را به دست آوریم. در این تابع به جای لگاریتم، مقدار

کلیدی ابتدا بر اساس داده های زیستی متناسب سازی شده اند. روابط کلی هر معیار به صورت زیر است

$$PageRank(p_i, t+1) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{a_{ji} \cdot PageRank(p_j, t)}{L(p_j)} \quad (1)$$

$$a_p^{(k)} = \sum_{q: e_{pq} \in E} h_q^{(k-1)} \quad (2)$$

$$h_p^{(k)} = \sum_{q: e_{pq} \in E} a_q^{(k)} \text{ for } k = 1, 2, 3, \dots \quad (3)$$

$$Katz_i = \alpha \sum_{j=1}^n A_{ij} \cdot Katz_j + \beta \quad (4)$$

$$P_A(i_a, j_a) = \sum_{\substack{k_h \rightarrow i_a \\ k_h \in G_b}} \frac{1}{in_deg(i_a)} \times \frac{1}{out_deg(k_h)} \quad (5)$$

$$P_H(i_h, j_h) = \sum_{\substack{i_h \rightarrow k_a \\ k_h \in G_b}} \sum_{\substack{j_h \rightarrow k_a \\ k_a \in G_b}} \frac{1}{in_deg(i_h)} \times \frac{1}{out_deg(k_a)} \quad (6)$$

که در روابط فوق $L(P_j)$ تعداد لینک های خروجی از p_j ، N تعداد ژن ها، $M(p_i)$ مجموعه ژن هایی که به p_i متصل هستند، A ماتریس مجاورت شبکه ژنی، d فاکتور دمپینگ بین ۰ و ۱ (معمولاً ۰٫۸۵)، $PageRank(p_i, t+1)$ رتبه ژن p_i ، $Katz_i$ رتبه کتر ژن i ، α ضریب میرایی، $in_deg(i_a)$ درجه ورودی ژن i ، $out_deg(k_a)$ درجه خروجی ژن k و $P_A(i_a, j_a)$ احتمال انتقال از i به j است.

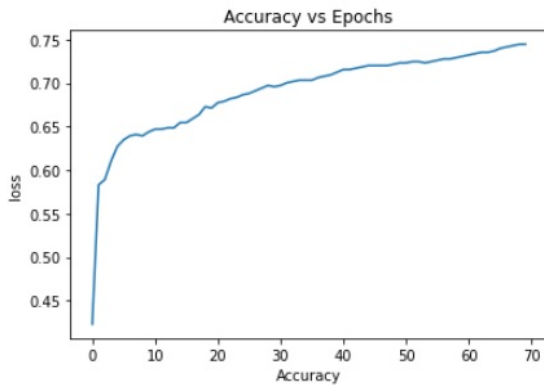
۴-۳ ساخت مدل و تنظیم پارامترها

قبل از ساخت مدل نیاز است تا داده ها برچسب گذاری شوند و برای برچسب زدن داده ها به پایگاه داده استاندارد و تأیید شده ژن های عامل سرطان نیاز است. برای این منظور از مجموعه ژن های سرطانی استاندارد و تأیید شده^۱ CGC استفاده گردید. مجموعه ژن های مربوط به سرطان سینه با نام TCGA-BRCA از طریق تارنما^۲ در دسترس است. ژن های مربوط به بارگیری و سپس ژن های مشخص شده توسط CGC به عنوان ژن های نهایی پالایه شدند و سایر اطلاعات ژن ها از مجموعه داده ارزیابی حذف گردیدند. نهایتاً برای سرطان سینه، ۵۷۲ ژن عامل سرطان استخراج شد که این مجموعه ژن ها با ۱ به معنی عامل سرطان و مابقی با ۰ به معنی عادی برچسب گذاری شدند.

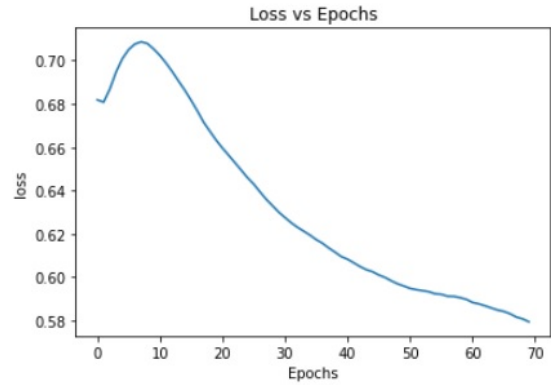
اطلس ژنوم سرطان (TCGA) پروژه های است برای فهرست نویسی جهش های ژنتیکی عامل سرطان با استفاده از توالی یابی ژنوم و

3. Random Under Sampling
4. Feed-Forward
5. PyTorch
6. Cross Entropy

1. Cancer Gene Census
2. <https://cancer.sanger.ac.uk/census>



شکل ۵: نمودار میزان افزایش دقت در هر تکرار.



شکل ۴: نمودار میزان سرعت کاهش هزینه در هر تکرار.

جدول ۱: ماتریس اغتشاش و بخش‌های مختلف آن.

	مقادیر واقعی	
	مثبت	منفی
مقادیر	درست مثبت	نادرست مثبت خطای نوع یک
پیش‌بینی شده	منفی	نادرست منفی خطای نوع دو

میزان دقت و هزینه در هر دوره برای بررسی تغییر مدل در طول آموزش ترسیم شد. برای محاسبه نسبت یادگیری، میزان سرعت کاهش هزینه در هر تکرار محاسبه گردید که در شکل‌های ۴ و ۵ آمده است. نسبت یادگیری باید با دقت زیادی انتخاب شود تا به اندازه‌ای زیاد نباشد که حالت بهینه را رد کنیم و به اندازه‌ای کم نباشد که یادگیری شبکه، سال‌ها زمان ببرد. همان طور که مشاهده می‌شود هزینه‌ها در هر تکرار به‌طور قابل توجهی کاهش می‌یابد. شکل‌های ۴ و ۵ نشان می‌دهند که مدل ایجادشده به‌خوبی یادگیری شده است.

۳-۴ تحلیل مدل

۵- معیارهای ارزیابی

رویکرد پیشنهادی با ۱۹ روش محاسباتی و مبتنی بر شبکه که در بخش ادبیات موضوع توضیح داده شد، مقایسه گردید. اکثر معیارهای ارزیابی با استفاده از ماتریس اغتشاش حاصل از اجرای مدل، محاسبه شده‌اند. استفاده از ماتریس اغتشاش در مسائل دسته‌بندی بانظارت معمول است. بر اساس این ماتریس مقادیر می‌توانند در یکی از دسته‌های منفی صحیح^۳، مثبت صحیح^۴، منفی کاذب^۵ و مثبت کاذب^۶ قرار گیرند. در مسائل درمانی و زمانی که دسته‌بندی به‌صورت بیمار و سالم است، دسته‌بندی درست نمونه‌ها در هر دسته، اهمیت بالایی دارد. مثبت صحیح یعنی مدل نمونه را در دسته مثبت طبقه‌بندی کرده و نمونه هم مثبت است. مثبت کاذب یعنی مدل نمونه را در دسته مثبت طبقه‌بندی کرده اما نمونه منفی است. منفی صحیح یعنی مدل نمونه را در دسته منفی طبقه‌بندی کرده و نمونه هم منفی است و منفی کاذب یعنی مدل نمونه را در دسته منفی طبقه‌بندی کرده اما نمونه مثبت است. بخش‌های مختلف ماتریس اغتشاش در جدول ۱ نشان داده شده است.

فهرست ژن‌های عامل سرطان پیش‌بینی‌شده با روش‌های NetBox، MDPFinder، Dendrix، OncodriveFM، DriverNet، Simon، e-driver، MutSigCV، oncodriveCLUST، iPac، ActiveDriver، dawnRank، CoMDP و MSEA از پایگاه داده v۲ DriverDB به‌دست آمدند [۲۹]. این پایگاه داده حاوی ژن‌های عامل سرطان پیش‌بینی‌شده توسط ۱۵ ابزار رایج است که در آن برای پیش‌بینی ژن‌های بیماری توسط هر روش از مجموعه داده‌های استاندارد TCGA-BRCA

نتیجه‌گرفته‌شده از شبکه عصبی را در آن ضرب می‌کنیم. اگر مقدار درستی نباشد باعث بالارفتن مقدار آنتروپی متقاطع می‌شود و ما متوجه خطای شبکه می‌شویم. همچنین هرچه این اختلاف بین مقدار ارائه‌شده توسط شبکه عصبی و مقدار واقعی بیشتر باشد، شبکه به‌صورت لگاریتمی رشد می‌کند؛ یعنی برای مقادیر کوچک، کوچک‌تر و برای مقادیر بزرگ، بزرگ‌تر می‌شود و خطا مشخص می‌گردد. این تابع حتی در حالتی که مشتق گرہ‌ها نزدیک به صفر است نیز تأثیر خطا در وزن‌ها را ایجاد و به‌طور کلی مانند اسمش یعنی آنتروپی عمل می‌کند. این تابع برای ارزیابی شبکه‌هایی که گرہ‌های پایانی آن به مفهوم یا دسته خاصی اشاره دارند و فایز شدن یک گرہ به‌منزله احتمال صحت آن مفهوم یا دسته است، می‌تواند انتخاب مناسبی باشد

$$Binary\ Cross - entropy = -(P(X) \cdot \log q(X) + (1 - P(X)) \cdot \log (1 - q(X))) \quad (7)$$

که در رابطه فوق اگر هدف ۰ باشد بخش اول رابطه و در صورتی که هدف ۱ باشد بخش دوم رابطه در نظر گرفته نمی‌شود. برای پیاده‌سازی تابع هزینه نیز از تابع BCELoss کتابخانه پایتورچ که تابع هزینه آنتروپی متقاطع استاندارد برای استفاده در مسائل طبقه‌بندی دودویی است استفاده گردید. برای شبکه مورد مطالعه، نرخ یادگیری برابر ۰٫۰۱، تعداد تکثیر رفت‌وبرگشت^۱ برابر ۷۰۰ و همچنین پارامتر BATCH_SIZE برابر ۶۴ در نظر گرفته شد. هرچند استفاده از دوره‌های بیشتر منجر به دقت بیشتر در شبکه می‌شود، اما زمان یادگیری شبکه افزایش می‌یابد. به‌علاوه اگر تعداد دوره‌ها بیش از حد زیاد باشد، ممکن است که هضم آن برای شبکه سنگین باشد و شبکه دچار بیش‌برازش^۲ گردد. برای ساخت مدل، ابتدا داده‌ها به مجموعه‌های آموزش و آزمون تقسیم شدند و برای این منظور ۳۳ درصد از داده‌ها برای قرارگرفتن در مجموعه آزمون انتخاب گردیدند. سپس برای فهم ماشین، ژن‌های عادی با ۰ و عامل سرطان با ۱ برچسب زده شدند.

3. True Negative
4. True Positive
5. False Negative
6. False Positive

1. Epoch
2. Over-Fit

استفاده از معیار حساسیت به‌تنهایی برای ارزیابی عملکرد سیستم درست نیست و باید در کنار معیار صحت مورد استفاده قرار گیرد. صحت در حقیقت درستی پیش‌بینی را نشان می‌دهد و معادل تعداد ژن‌های عامل سرطانی است که به‌عنوان عامل سرطان تشخیص داده می‌شوند. صحت، نسبت تعداد پیش‌بینی‌های صحیح انجام‌شده برای نمونه‌های یک کلاس خاص به تعداد کل پیش‌بینی‌ها برای نمونه‌های همان کلاس خاص را ارزیابی می‌کند. مقدار بالا برای معیار دقت، بیانگر تعداد کم داده‌هایی است که به اشتباه در کلاس خاص دسته‌بندی شده‌اند. این معیار، بیشتر زمانی استفاده می‌شود که مدل دسته‌بندی، تعلق یک نمونه به یک کلاس خاص را پیش‌بینی کرده باشد؛ یعنی زمانی که مثبت کاذب دارای ارزش بالایی باشد. در هنگام ارزیابی عملکرد یک مدل دسته‌بندی متن، بهتر است که از این معیار در کنار معیار حساسیت استفاده شود

$$Precision = \frac{TP}{TP + FP} \quad (۹)$$

صحت و حساسیت به‌تنهایی برای نشان‌دادن عملکرد سیستم کافی نیستند. میانگین ریاضی هم برای محاسبه این دو معیار مناسب نیست و به همین جهت از میانگین هارمونیک استفاده می‌شود. این معیار، پارامترهای صحت و حساسیت را با هم ترکیب می‌کند تا مشخص شود مدل تا چه حد عملکرد خوبی از خود نشان می‌دهد. به این معیار، میانگین هارمونیک دو معیار صحت و حساسیت نیز گفته می‌شود. این معیار، نسبت به معیار صحت، تصویر دقیق‌تری از نحوه عملکرد مدل پیش‌بینی برای تمامی کلاس‌های موجود در داده‌ها نشان می‌دهد و در بهترین حالت، یک و در بدترین حالت صفر است

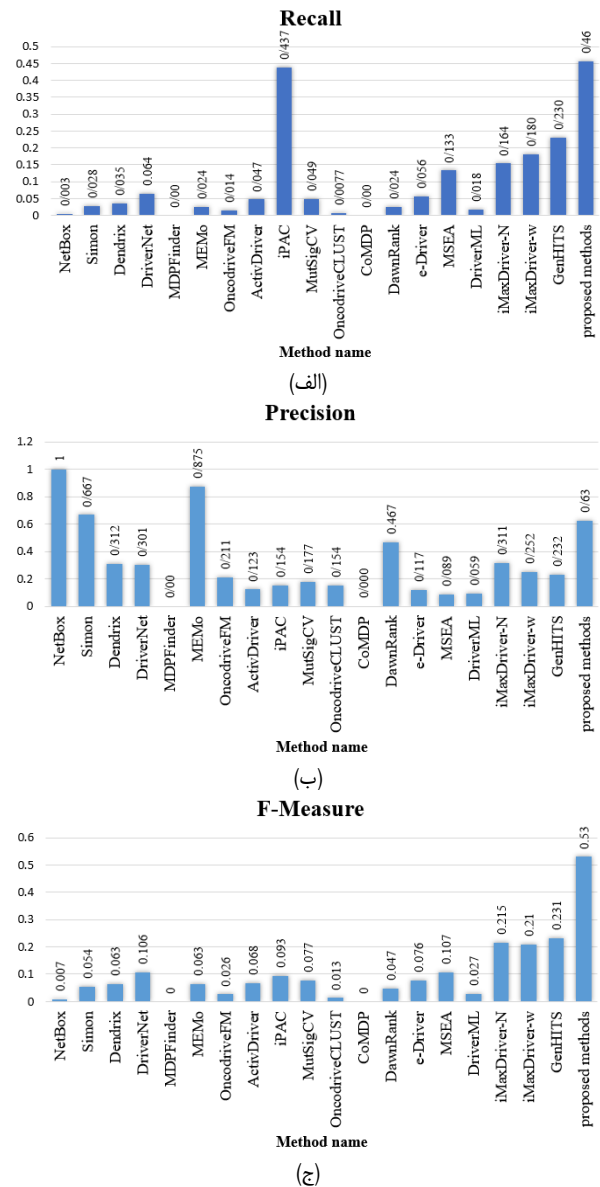
$$F_measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (۱۰)$$

معیار دقت^۴ به این معناست که مدل به چه میزان توانسته که خروجی را درست پیش‌بینی کند و بر اساس آن می‌توان فهمید که آیا مدل، درست آموزش دیده یا خیر و کارایی مدل در چه حدی است. این معیار تنها اطلاعات کلی را درباره کارایی مدل ارائه می‌دهد و چیزی در خصوص اطلاعات جزئی بازگو نمی‌کند

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (۱۱)$$

۶- نتایج

در این مطالعه، ابتدا با استفاده از داده‌های برهم‌کنش تنظیمی و داده‌های بیان ژن و پس از پیش‌پردازش‌های لازم، شبکه تنظیم ژنی سرطان سینه ساخته شد. سپس ویژگی‌های مبتنی بر انتشار مختلف که نشان‌دهنده میزان تأثیر یک ژن بر ژن‌های هدف است از طریق شبکه سرطان سینه ساخته‌شده محاسبه گردیدند و ویژگی‌های به‌دست‌آمده برای طبقه‌بندی ژن‌ها به یک شبکه عصبی پیش‌خور وارد شدند. پس از تنظیم پارامترها و اجرای مدل خروجی به‌صورت فهرستی حاوی ژن‌های عامل سرطان و ژن‌های عادی به دست آمد. نتایج به‌دست‌آمده از رویکرد پیشنهادی و ۱۹ روش دیگر با استفاده از معیارهای اشاره‌شده در بخش قبل مورد ارزیابی قرار گرفتند. همان‌طور که در شکل ۶ مشاهده می‌شود روش پیشنهادی با مقدار فراخوانی برابر ۰،۴۶، بالاترین مقدار فراخوانی را در بین همه روش‌های شبکه‌ای و محاسباتی مورد مقایسه دارد و بعد از



شکل ۶: مقایسه معیارهای کارایی روش پیشنهادی و دیگر روش‌های پیشین.

استفاده گردیده و برای سایر روش‌ها از مقالات منتشرشده مربوط گرفته شده است. برای اعتبارسنجی روش پیشنهادی و دیگر روش‌ها از مجموعه ژن‌های استاندارد و تأییدشده CGC که در بخش قبل توضیح داده شد استفاده گردید. برای بررسی میزان کارایی الگوریتم پیشنهادی و مقایسه آن با دیگر روش‌ها از معیارهای فراخوانی^۱ یا حساسیت، صحت^۲ و میانگین هارمونیک^۳ که در ارزیابی روش‌های دسته‌بندی مرسوم می‌باشد استفاده شده است. معیار فراخوانی نسبت تعداد ژن‌هایی را که به‌صورت درست به‌عنوان عامل سرطان تشخیص داده شده‌اند به تعداد کل ژن‌های عامل سرطان نشان می‌دهد. این معیار بیان‌کننده نسبت تعداد ژن‌های درست دسته‌بندی‌شده در یک کلاس خاص به تعداد کل ژن‌هایی است که باید در همان کلاس خاص دسته‌بندی شوند

$$Recall = \frac{TP}{TP + FN} \quad (۸)$$

1. Recall
2. Precision
3. F-Measure

- [6] A. Youn and R. Simon, "Identifying cancer driver genes in tumor genome sequencing studies," *Bioinformatics*, vol. 27, no. 2, pp. 175-181, Jan. 2011.
- [7] J. Zhao, S. Zhang, L. Y. Wu, and X. S. Zhang, "Efficient methods for identifying mutated driver pathways in cancer," *Bioinformatics*, vol. 28, no. 22, pp. 2940-2947, 15 Nov. 2012.
- [8] F. Vandin, E. Upfal, and B. J. Raphael, "De novo discovery of mutated driver pathways in cancer," *Genome Research*, vol. 22, no. 2, pp. 375-385, Feb. 2012.
- [9] A. Gonzalez-Perez and N. Lopez-Bigas, "Functional impact bias reveals cancer drivers," *Nucleic Acids Research*, vol. 40, no. 21, Article ID: e169, Nov. 2012.
- [10] G. Ciriello, E. Cerami, C. Sander, and N. Schultz, "Mutual exclusivity analysis identifies oncogenic network modules," *Genome Research*, vol. 22, no. 2, pp. 398-406, Feb. 2012.
- [11] A. Bashashati, G. Haffari, J. Ding, G. Ha, K. Lui, J. Rosner, D. G. Huntsman, C. Caldas, S. A. Aparicio, and S. P. Shah, "DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer," *Genome Biology*, vol. 13, no. 12, pp. 1-4, Dec. 2012.
- [12] F. Vandin, E. Upfal, and B. J. Raphael, "De novo discovery of mutated driver pathways in cancer," *Genome Research*, vol. 22, no. 2, pp. 375-385, Feb. 2012.
- [13] J. Reimand, O. Wagih, and G. D. Bader, "The mutational landscape of phosphorylation signaling in cancer," *Scientific Reports*, vol. 3, no. 1, Article ID: 2651, Oct. 2013.
- [14] M. R. Aure, et al., "Identifying in-trans process associated genes in breast cancer by integrated analysis of copy number and expression data," *PLoS One*, vol. 8, no. 1, Article ID: e53014, Jan. 2013.
- [15] M. S. Lawrence, et al., "Mutational heterogeneity in cancer and the search for new cancer-associated genes," *Nature*, vol. 499, pp. 214-218, 2013.
- [16] D. Tamborero, A. Gonzalez-Perez, and N. Lopez-Bigas, "OncodriveCLUS: exploiting the positional clustering of somatic mutations to identify cancer genes," *Bioinformatics*, vol. 29, no. 18, pp. 2238-2244, Sept. 2013.
- [17] E. Porta-Pardo and A. Godzik, "E-Driver: a novel method to identify protein regions driving cancer," *Bioinformatics*, vol. 30, no. 21, pp. 3109-3114, Nov. 2014.
- [18] J. P. Hou and J. Ma, "DawnRank: discovering personalized driver genes in cancer," *Genome Medicine*, vol. 6, Article ID: 56, 16 pp., Jul. 2014.
- [19] J. Zhang, L. Y. Wu, X. S. Zhang, and S. Zhang, "Discovery of co-occurring driver pathways in cancer," *BMC Bioinformatics*, vol. 15, Article ID: 271, 14 pp., 2014.
- [20] D. Arneson, A. Bhattacharya, L. Shu, V. P. Mäkinen, and X. Yang, "Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration," *BMC Genomics*, vol. 17, Article ID: 772, 9 pp., 2016.
- [21] M. Rahimi, B. Teimourpour, and S. A. Marashi, "Cancer driver gene discovery in transcriptional regulatory networks using influence maximization approach," *Computers in Biology and Medicine*, vol. 114, Article ID: 103362, Nov. 2019.
- [22] M. Akhavan-Safar, B. Teimourpour, and M. Kargari, "GenHITS: a network science approach to driver gene detection in human regulatory network using gene's influence evaluation," *J. of Biomedical Informatics*, vol. 114, Article ID: 103661, Feb. 2021.
- [23] Y. Lu, Y. Wang, N. Sheng, H. Wang, Y. Fu, and Y. Tian, "RDDriver: a novel method based on multi-layer heterogeneous transcriptional regulation network for identifying pancreatic cancer biomarker," in *Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine, BIBM'22*, pp. 497-502, Las Vegas, NV, USA, 6-8 Dec. 2022.
- [24] S. Lee, H. Jung, J. Park, and J. Ahn, "Accurate prediction of cancer prognosis by exploiting patient-specific cancer driver genes," *International J. of Molecular Sciences*, vol. 24, no. 7, Article ID: 6445, Apr. 2023.
- [25] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, "A census of human transcription factors: function, expression and evolution," *Nature Reviews Genetics*, vol. 10, no. 4, pp. 252-263, Apr. 2009.
- [26] H. Han, et al., "TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions," *Nucleic Acids Research*, vol. 46-D1, pp. 380-386, Jan. 2018.
- [27] S. M. Cheng, V. Karyotis, P. Y. Chen, K. C. Chen, and S. Papavassiliou, "Diffusion models for information dissemination dynamics in wireless complex communication networks," *J. of Complex Systems*, vol. 2013, Article ID: 972352, 13 pp., 2013.
- [28] I. F. Chung, C. Y. Chen, S. C. Su, C. Y. Li, K. J. Wu, H. W. Wang, and W. C. Cheng, "DriverDBv2: a database for human cancer driver

ژن، آن، iPac با فراخوانی ۰/۴۳ در رتبه دوم قرار دارد. از نظر معیار حساسیت نیز بعد از روش‌های محاسباتی Memo، Netbox و Simon در رتبه بعد قرار دارند. همچنین روش پیشنهادی از نظر معیار F-measure که میانگین هارمونیک دو معیار قبل است از همه روش‌های پیشین عملکرد بالاتری دارد. هرچند برخی روش‌های صحت بالاتری را داشتند اما از میانگین هارمونیک مناسبی برخوردار نیستند. مقایسه معیارهای ارزیابی نشان می‌دهد که روش پیشنهادی عملکرد مناسب‌تری نسبت به سایر روش‌ها دارد. علاوه بر این، روش پیشنهادی با دقت برابر ۰/۶۶ بالاترین دقت را در بین همه روش‌ها به خود اختصاص داد و MSEA با ۰/۲۶ و GenHITS در رتبه‌های بعدی قرار گرفتند.

۷- نتیجه‌گیری

روش‌های مختلفی برای شناسایی ژن‌های عامل سرطان پیشنهاد شده که اکثر آنها متکی به داده‌های جهش بوده و دارای مقادیر کاذب زیاد در نتایج هستند. در این مطالعه، یک رویکرد مبتنی بر شبکه عصبی چندلایه برای شناسایی ژن‌های عامل سرطان سینه در شبکه تنظیم ژنی انسان پیشنهاد شد و در آن ابتدا با استفاده از داده‌های بیان ژن و برهم‌کنش‌های ژنی، شبکه تنظیم ژنی سرطان سینه ساخته شد. سپس داده‌ها برای آموزش مدل برچسب‌گذاری شدند. با توجه به اینکه داده‌ها نامتوازن بودند، با استفاده از روش کم نمونه‌گیری تصادفی، متوازن‌سازی شدند. کلیه مراحل با استفاده از زبان پایتون و کتابخانه پایتورچ پیاده‌سازی گردیده است. سپس بردارهای ویژگی بر اساس اختلاف بیان هر ژن در بافت سالم و سرطانی و همچنین معیارهای انتشار که بر اساس داده‌های زیستی متناسب‌سازی شده بودند، به دست آمد. بردارهای به دست آمده به یک شبکه سه‌لایه پیش‌خور داده شد و نهایتاً پس از آموزش مدل، ژن‌ها به دو دسته عامل سرطان و نرمال دسته‌بندی گردیدند. نتایج به دست آمده نشان می‌دهند که رویکرد پیشنهادی توانسته از نظر معیارهای کارایی بهتر از سایر روش‌های محاسباتی و شبکه‌ای پیشین عمل کند؛ به طوری که توانست به ترتیب ۵۶/۴۱ درصد میانگین هارمونیک و ۵ درصد فراخوانی را بهبود بخشد. همچنین روش پیشنهادی بالاترین مقدار دقت را در بین همه روش‌ها به خود اختصاص داده است. نتایج حاصل از این مطالعه نشان می‌دهند که استفاده از رویکردهای مبتنی بر شبکه‌های عصبی و یادگیری عمیق برای شناسایی ژن‌های عامل سرطان می‌تواند باعث بهبود کارایی روش‌های موجود شده و در مطالعات آینده، بیشتر مورد توجه قرار گیرد.

مراجع

- [1] M. H. Bailey, et al., "Comprehensive characterization of cancer driver genes and mutations," *Cell*, vol. 173, no. 2, pp. 371-385, Apr. 2018.
- [2] L. Ding, et al., "Somatic mutations affect key pathways in lung adenocarcinoma," *Nature*, vol. 455, no. 7216, pp. 1069-1075, Oct. 2008.
- [3] L. Mularoni, R. Sabarinathan, J. Deu-Pons, A. Gonzalez-Perez, and N. Lopez-Bigas, "OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations," *Genome Biology*, vol. 17, Article ID: 128, 13 pp, Dec. 2016.
- [4] J. Reimand, O. Wagih, and G. D. Bader, "The mutational landscape of phosphorylation signaling in cancer," *Scientific Reports*, vol. 3, no. 1, Article ID: 2651, 9 pp., 2013.
- [5] M. Helmy, M. Awad, and K. A. Mosa, "Limited resources of genome sequencing in developing countries: challenges and solutions," *Applied & Translational Genomics*, vol. 9, pp. 15-19, Jun. 2016.

عباسعلی رضایی در سال ۱۳۷۳ مدرک کارشناسی خود را در رشته مهندسی کامپیوتر گرایش سخت افزار از دانشگاه شهید بهشتی تهران و در سال ۱۳۸۰ مدرک کارشناسی ارشد خود را در رشته کامپیوتر گرایش معماری سیستم‌های کامپیوتری از دانشگاه اصفهان و در سال ۱۳۹۲ مدرک دکتری خود را در رشته کامپیوتر گرایش سخت افزار از دانشگاه علوم و تحقیقات تهران دریافت نمود. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: سیستم‌های توزیع شده، شبکه‌های کامپیوتری، شبکه‌های حسگر بی‌سیم، شبکه‌های نسل جدید و محاسبات نرم.

gene research," *Nucleic Acids Research*, vol. 44-D1, pp. 975-979, Jan. 2016.

- [29] P. Shannon, *et al.*, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498-2504, Nov. 2003.

مصطفی اخوان صفار مدرک کارشناسی ارشد خود را در سال ۱۳۸۸ در رشته مهندسی فناوری اطلاعات از دانشگاه علم و صنعت ایران و نیز مدرک دکتری خود را در سال ۱۴۰۰ در رشته مهندسی فناوری اطلاعات - مدیریت سیستم‌های اطلاعاتی از دانشگاه تربیت مدرس دریافت نمود. نام‌برده هم‌اکنون بعنوان استادیار گروه مهندسی فناوری اطلاعات، در دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه پیام نور مشغول به کار است. زمینه‌های علمی مورد علاقه ایشان متنوع بوده و شامل موضوعاتی مانند بایوانفورماتیک، یادگیری ماشین، داده‌کاوی، سیستم‌های اطلاعاتی و تحلیل شبکه‌های اجتماعی می‌باشد.