



دانشگاه شهید چمران اهواز

دانشکده مهندسی

شماره پایان نامه: ۹۳۱۴۲۳۶

پایان نامه کارشناسی ارشد گروه مهندسی کامپیوتر

گرایش هوش مصنوعی

عنوان:

تشخیص سرطان از مجموعه داده‌های ریزآرایه با به‌کارگیری مدل‌های ترکیبی

استاد راهنما:

دکتر علی‌رضا عصاره

استاد مشاور:

دکتر بیتا شادگار

نگارنده:

محمد مروت

تیر ماه ۱۳۹۳

باسمه تعالی

دانشگاه شهید چمران اهواز

دانشکده مهندسی

(نتیجه ارزشیابی پایان نامه کارشناسی ارشد)

پایان نامه آقای محمد مروت دانشجوی رشته: مهندسی کامپیوتر گرایش: هوش مصنوعی

دانشکده مهندسی به شماره دانشجویی ۹۱۱۴۲۰۴

با عنوان:

تشخیص سرطان از مجموعه داده‌های ریزآرایه با به‌کارگیری مدل‌های ترکیبی

جهت اخذ مدرک: کارشناسی ارشد در تاریخ: ۹ تیرماه ۱۳۹۳ توسط هیأت داوران مورد ارزشیابی قرار گرفت و با درجه عالی تصویب گردید.

امضاء	رتبه علمی	اعضای هیأت داوران:
.....	دانشیار	استاد راهنما: دکتر علی‌رضا عصاره
.....	استادیار	استاد مشاور: دکتر بیتا شادگار
.....	استادیار	استاد داور: دکتر کریم انصاری اصل
.....	استادیار	استاد داور: دکتر احسان نامجو
.....	دانشیار	نماینده تحصیلات تکمیلی: دکتر مرتضی بهبهانی نژاد
.....	استادیار	مدیر گروه: دکتر مرجان نادران طحان
.....	استادیار	معاون پژوهشی و تحصیلات تکمیلی دانشکده: دکتر علی حقیقی
.....	استاد	مدیر تحصیلات تکمیلی دانشگاه: دکتر عبدالرحمن راسخ

تقدیم

این مجموعه را...

به مادرم

آن که آفتاب مهرش در آستانه قلبم، پنهان پارچاست

و هرگز غروب نخواهد کرد،

و با بسبب بردستان پدرم

به پدرم که نمی دانم از بزرگی اش بگویم یا مردانگی،

سخاوت، معرفت، مهربانی و...

و به تندیس های امید، برادر و خواهر عزیزم

تقدیم می نمایم.

اگر پزندگان پرواز را، اگر ابرها کریستن را و اگر انسان با محبت را فراموش کنند... من هرگز شمارا فراموش نخواهم کرد.

پاسکزاری

پاس و ستایش خداوند را سزا است که عشق به آموختن را در انسان به ودیعه نهاد، عشقی که رسالت انگیز است، رسالتی انبیاگونه. سگکزاریم که در این راه یاریان نمود و از اومی خواهیم که ما را در جست به کارگیری آموخته‌هایمان در راه خیرمد فرماید.

اکنون که این پژوهش به زیور چاپ آراسته میگردد بر خود لازم می‌دانم والاترین مراتب پاس خویش را به محضر استادان فرزانه‌ای که در طول تحصیل بمواره افتخار شاکردی آنها را داشته و اینجانب را مریون الطاف و عنایات خالصانه خود قرار داده اند اعلام میدارم. از این رو شایسته است از اساتید فرزانه جناب آقای دکتر علی رضا عصاره و سرکار خانم دکتر میتا شاکار که بارها بهایی‌های سودمند و تکررات عالمانه خود در تمامی مراحل این تحقیق اینجانب را یاری نموده اند صمیمانه تشکر و قدردانی نمایم.

از لطف و محبت دوستان عزیز و مهربانم و تمام کسانی که به نحوی در انجام این تحقیق مرلاری نموده اند

کمال تشکر و قدردانی را می‌نمایم.

فهرست مطالب

الف.....	فرم ارزشیابی
ب.....	تقدیم
پ.....	سپاسگزاری
ت.....	فهرست مطالب
ج.....	فهرست شکل‌ها و نمودارها
خ.....	فهرست جدول‌ها
د.....	چکیده پایان‌نامه به زبان فارسی

فصل اول: مقدمه

۱	
۲.....	۱-۱ بیولوژیکی مولکولی
۲.....	۲-۱ بیولوژی سرطان
۴.....	۳-۱ فناوری ریزآرایه
۵.....	۱-۳-۱ بیان ژنی و فناوری ریزآرایه
۷.....	۴-۳-۱ کاربردهای ریزآرایه دی-ان-ای
۸.....	۴-۱ اهداف تحقیق
۸.....	۵-۱ ساختار پایان‌نامه
۹.....	۶-۱ جمع‌بندی

فصل دوم: پیشینه‌ی تحقیق

۱۰	
۱۰.....	۱-۲ انتخاب ژن
۱۳.....	۲-۲ خوشه‌بندی ژن
۱۴.....	۳-۲ دسته‌بندی نمونه‌ها
۱۶.....	۴-۲ جمع‌بندی

۱۷	فصل سوم: مبانی تحقیق
۱۷	۱-۳ مشخصات مجموعه داده‌ها
۲۱	۲-۳ مراحل تجزیه و تحلیل داده‌های ریزآرایه
۲۱	۱-۲-۳ مرحله پیش پردازش داده‌ها
۲۲	۱-۲-۳ جایگزینی مقادیر گم شده
۲۲	۲-۱-۲-۳ استانداردسازی مجموعه داده‌ها
۲۳	۲-۲-۳ مرحله انتخاب ویژگی
۲۶	۱-۲-۲-۳ انواع روش‌های انتخاب ویژگی براساس قطعیت یا عدم قطعیت در انتخاب
۲۶	۲-۲-۲-۳ انواع روش‌های انتخاب ویژگی براساس رتبه‌بندی ویژگی یا کاهش ویژگی‌ها
۲۷	۳-۲-۲-۳ انواع روش‌های انتخاب ویژگی براساس استراتژی جستجو
۲۷	۴-۲-۲-۳ انواع روش‌های انتخاب ویژگی براساس هدف
۳۱	۵-۲-۲-۳ تکنیک‌های انتخاب ویژگی به کار رفته در این پایان نامه
۳۸	۳-۲-۳ مرحله دسته‌بندی داده‌های ریزآرایه
۳۸	۱-۳-۲-۳ دسته‌بندی کننده‌های منفرد
۳۹	۲-۳-۲-۳ دسته‌بندی کننده‌های ترکیبی (انسمبل‌ها)
۴۶	۳-۳-۲-۳ الگوریتم‌های دسته‌بندی کننده استفاده شده در این پژوهش
۵۷	۳-۳ جمع‌بندی

۵۸ فصل چهارم: روش‌های پیشنهادی و نتایج

۵۸	۱-۴ روش‌های پیشنهادی
۵۹	۱-۱-۴ روش پیشنهادی اول: انسمبلی از فیلترها و دسته‌بندکننده‌ها برای دسته‌بندی داده‌های ریزآرایه
۶۰	۱-۱-۴-۱ مرحله انتخاب ویژگی روش پیشنهادی اول
۶۴	۲-۱-۴-۱ مرحله دسته‌بندی روش پیشنهادی اول
۶۶	۲-۱-۴ روش پیشنهادی دوم: انسمبلی از فیلترها و راپرها برای دسته‌بندی داده‌های ریزآرایه
۶۸	۱-۲-۴-۱ مرحله انتخاب ویژگی روش پیشنهادی دوم
۶۹	۲-۲-۴-۱ مرحله دسته‌بندی روش پیشنهادی دوم
۷۳	۲-۴ بررسی و تحلیل نتایج روش‌های پیشنهادی
۷۴	۱-۲-۴ بسته‌های نرم‌افزاری مورد استفاده شده
۷۴	۲-۲-۴ چگونگی ارزیابی دسته‌بندکننده‌ها
۷۵	۳-۲-۴ تحلیل نتایج روش پیشنهادی اول

۷۵ ۱-۳-۲-۴ نتایج مرحله انتخاب ویژگی روش پیشنهادی اول
۸۲ ۲-۳-۲-۴ نتایج مرحله دسته‌بندی روش پیشنهادی اول
۸۶ ۴-۲-۴ تحلیل نتایج روش پیشنهادی دوم
۸۶ ۱-۴-۲-۴ نتایج مرحله انتخاب ویژگی روش پیشنهادی دوم
۹۱ ۲-۴-۲-۴ نتایج مرحله دسته‌بندی روش پیشنهادی دوم
۹۳ ۵-۲-۴ مقایسه نتایج دو روش پیشنهادی
۹۴ ۶-۲-۴ مقایسه کارایی روش‌های پیشنهادی با روش‌های پیشین
۹۶ ۳-۴ جمع‌بندی

۹۷

فصل پنجم: نتیجه‌گیری

۹۷ ۱-۵ نتیجه‌گیری
۹۸ ۳-۵ کارهای آینده

۱۰۰ مراجع

فهرست شکل‌ها و نمودارها

- شکل ۱-۱: چگونگی سرطانی شدن سلول‌ها ۳
- شکل ۲-۱: یک نمونه میکروچیپ ۵
- شکل ۳-۱: مراحل تهیه مجموعه داده با کمک فناوری ریزآرایه ۷
- شکل ۱-۳: یک نمونه از ماتریس توصیف داده‌ها ۱۸
- شکل ۲-۳: مراحل تجزیه و تحلیل داده‌های ریزآرایه ۲۱
- شکل ۳-۳: انواع ویژگی‌ها ۲۳
- شکل ۴-۳: شیوه کارکرد راپر جهت ارزیابی مجموعه ویژگی ۳۷
- شکل ۵-۳: انتخاب دسته‌بند کننده ۴۱
- شکل ۶-۳: ترکیب نتایج ۴۱
- شکل ۷-۳: شیوه‌های مختلف ترکیب نتایج ۴۳
- شکل ۸-۳: شیوه‌های مختلف رأی‌گیری اکثریت ۴۴
- شکل ۹-۳: بهترین خط جدا کننده و بردارهای پشتیبان ۵۰
- شکل ۱۰-۳: بلوک دیاگرام الگوریتم ترکیبی بگینگ ۵۲
- شکل ۱۱-۳: بلوک دیاگرام الگوریتم ترکیبی دگینگ ۵۲
- شکل ۱۲-۳: بلوک دیاگرام الگوریتم ترکیبی آدابوست ۵۳
- شکل ۱۳-۳: بلوک دیاگرام الگوریتم ترکیبی جنگل تصادفی ۵۴
- شکل ۱۴-۳: بلوک دیاگرام الگوریتم ترکیبی استکینگ ۵۷
- شکل ۱-۴: بلوک دیاگرام روش پیشنهادی اول ۵۹
- شکل ۲-۴: حالات متفاوت انتخاب‌های ممکن در هنگام ترکیب نتایج دو فیلتر ۶۳
- شکل ۳-۴: بلوک دیاگرام روش پیشنهادی دوم ۶۷
- شکل ۴-۴: نمودار مقایسه حد آستانه پیشنهادی با حد آستانه یک‌دهم ویژگی‌ها برای فیلتر GSNR ۷۸
- شکل ۵-۴: نمودار مقایسه حد آستانه پیشنهادی با حد آستانه یک‌دهم ویژگی‌ها برای فیلتر MRMR ۷۸
- شکل ۶-۴: نمودار مقایسه حد آستانه پیشنهادی با حد آستانه یک‌دهم ویژگی‌ها برای فیلتر ReliefF ۷۹
- شکل ۷-۴: نمودار مقایسه دقت الگوریتم درخت تصمیم در هر یک از پنج حالت ممکن ترکیب فیلترها ۸۰
- شکل ۸-۴: نمودار مقایسه دقت الگوریتم نایویز در هر یک از پنج حالت ممکن ترکیب فیلترها ۸۰
- شکل ۹-۴: نمودار مقایسه دقت الگوریتم ماشین بردار پشتیبان در هر یک از پنج حالت ممکن ترکیب فیلترها ۸۱

- شکل ۴-۱۰: نمودار میزان صحت الگوریتم درخت تصمیم بر روی مجموعه حاصل شده از فیلترهای مختلف ۸۷
- شکل ۴-۱۱: نمودار میزان صحت الگوریتم نایوبیز بر روی مجموعه حاصل شده از فیلترهای مختلف ۸۸
- شکل ۴-۱۲: نمودار میزان صحت الگوریتم ماشین بردار پشتیبان بر روی مجموعه حاصل شده از فیلترهای مختلف ۸۸
- شکل ۴-۱۳: نمودار میزان دقت حاصل از طرح پیشنهادی دوم، قبل و بعد از اعمال راپرها ۹۲

فهرست جدول‌ها

- جدول ۱-۳: مشخصات مجموعه داده‌ها ۱۸
- جدول ۲-۳: مزایا و معایب انواع تکنیک‌های انتخاب ویژگی ۳۰
- جدول ۱-۴: ماتریس درهم‌ریختگی ۷۱
- جدول ۲-۴: ماتریس درهم‌ریختگی مرتبط با دسته‌بندکننده درخت تصمیم بر روی مجموعه داده MLL ۷۲
- جدول ۳-۴: تعداد ویژگی‌های منتخب پس از اعمال فیلتر عدم قطعیت متقارن ۷۶
- جدول ۴-۴: تعداد ویژگی‌های منتخب توسط هریک از فیلترها، پس از اعمال بر روی مجموعه ویژگی‌های تولید شده توسط فیلتر عدم قطعیت متقارن ۷۷
- جدول ۵-۴: تعداد ویژگی‌های منتخب و میانگین دقت حاصل شده (٪) توسط هریک از حالات ممکن ۸۲
- جدول ۶-۴: میزان دقت حاصل از اعمال هریک از الگوریتم‌های ترکیبی بر روی هریک از مجموعه داده‌ها (٪) ۸۴
- جدول ۷-۴: مقایسه میزان دقت روش پیشنهادی اول با الگوریتم‌های منفرد و ترکیبی (٪) ۸۵
- جدول ۸-۴: تعداد ویژگی‌های منتخب و میانگین دقت حاصل شده (٪) توسط فیلترهای متفاوت ۸۹
- جدول ۹-۴: تعداد ویژگی‌های منتخب توسط هر راپر ۹۰
- جدول ۱۰-۴: میزان دقت (٪) حاصل از طرح پیشنهادی دوم، قبل و بعد از اعمال راپرها ۹۲
- جدول ۱۱-۴: تعداد ویژگی‌ها و میزان دقت (٪) حاصل از دو روش پیشنهادی ۹۳
- جدول ۱۲-۴: مقایسه میزان دقت روش‌های پیشنهادی با پژوهش‌های پیشین (٪) ۹۵

چکیده

نام خانوادگی: مروت	نام: محمد	شماره دانشجویی: ۹۱۱۴۲۰۴
عنوان پایان نامه: تشخیص سرطان از مجموعه داده‌های ریزآرایه با به‌کارگیری مدل‌های ترکیبی		
استاد راهنما: دکتر علی‌رضا عصاره		
استاد مشاور: دکتر بیتا شادگار		
درجه تحصیلی: کارشناسی ارشد	رشته: مهندسی کامپیوتر	گرایش: هوش مصنوعی
دانشگاه: شهید چمران اهواز	دانشکده: مهندسی	گروه: کامپیوتر
تاریخ فارغ التحصیلی: تیرماه ۱۳۹۳		تعداد صفحه: ۱۰۵
کلید واژه‌ها: فناوری ریزآرایه، دسته‌بندی داده‌های سرطان، تکنیک‌های ترکیبی، فیلترها، راپرها		
<p>چکیده:</p> <p>ارائه فناوری ریزآرایه در سال‌های اخیر، امکان مانیتورینگ بیان هزاران ژن را به‌طور همزمان فراهم نموده است. تحلیل‌هایی که در زمینه داده‌های ریزآرایه صورت گرفته است، بیانگر قدرت این فناوری در زمان تشخیص بسیاری از بیماری‌ها، از جمله سرطان است. چالشی که در این زمینه مطرح است، تعداد بالای ژن‌ها (ویژگی‌ها) و ازطرفی تعداد پایین نمونه‌ها است. در این پایان‌نامه، دو روش با دو رویکرد متفاوت به‌منظور ارائه یک سیستم کارای تشخیص سرطان ازطریق تحلیل داده‌های ریزآرایه ارائه شده است. در روش اول، چندین روش ترکیبی در راستای دستیابی به روشی مقاوم با یکدیگر ترکیب شده‌اند و در نهایت ترکیب نتایج توسط رأی‌گیری اکثریت صورت پذیرفته است. روش پیشنهادی دوم به‌منظور ارتقاء کارایی مرحله انتخاب ژن روش اول، از یک روش ترکیبی در مرحله انتخاب ژن سود جسته است که این روش ترکیبی، از ترکیب فیلترها و راپرها تشکیل شده است. درحقیقت به‌کارگیری فیلترها در گام ابتدایی این روش، زمینه را برای اعمال راپرها در گام بعدی در راستای بهبود دقت نتایج فراهم نموده است. نتایج حاصل از روش‌های پیشنهادی بر روی ۱۱ مجموعه داده ریزآرایه، بیانگر کارایی بسیار بالای روش‌های پیشنهادی در راستای تشخیص دقیق بیماری سرطان است.</p>		

مقدمه و طرح مسأله

۱

سالانه بیش از دوازده میلیون نفر از مردم سراسر دنیا به بیماری سرطان مبتلا شده و بیش از هفت میلیون نفر جان خود را از دست می‌دهند (روزانه حدود بیست هزار نفر از بیماری سرطان می‌میرند) [۱]. میلیون‌ها نفر از افراد خانواده بیماران هم بایستی از آنها مراقبت و پرستاری کنند، لذا خسارات روحی و مالی زیادی متحمل می‌شوند. سرطان هزاران سال پیش شناسایی شده است و بیش از پنجاه سال هست که دانشمندان به‌طور گسترده‌ای در این زمینه تحقیق می‌کنند، اما هنوز دارویی قطعی برای درمان این بیماری کشف نشده است. نکته‌ای که می‌تواند حیات بخش افراد مبتلا به این بیماری باشد، شناسایی سرطان در مراحل اولیه است؛ زیرا در صورتی که این بیماری کشنده، قبل از ورود به مراحل بحرانی شناسایی شود تا حدودی قابل درمان و کنترل است.

برای تعریف مسأله و توضیح اهمیت آن، نیاز به آشنایی با برخی مفاهیم اولیه وجود دارد که در ادامه به معرفی مختصری از آنها پرداخته می‌شود.

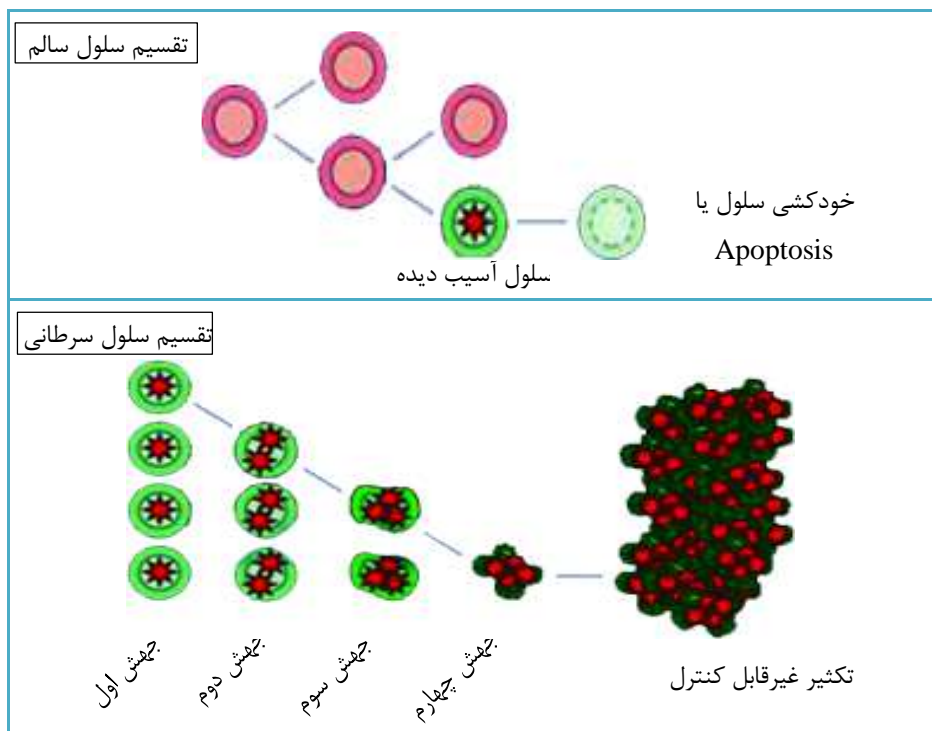
۱-۱ بیولوژیکی مولکولی

اساس علم بیولوژی، درک بیولوژی مولکولی سلول است که امروزه یک حیطه پررونق تحقیقات هست [۲]. کاربردهای پزشکی بیولوژی مولکولی سلول، امروزه شهره عام و خاص است. این علم با این هدف که با جمع‌آوری داده‌ها و پیدا کردن الگوهای مابین آنها به کشف و درمان بیماری‌ها پردازد، به سرعت به سوی نظام‌های مبتنی بر داده در حرکت است. باتوجه به تلاش پژوهشگران زمینه بیولوژیکی مولکولی، در سال‌های اخیر با سیل عظیمی از داده‌ها در این زمینه مواجه هستیم. لذا ابزارها و الگوریتم‌های محاسباتی، نقش بسیار مهمی در شناسایی الگوها بازی می‌کنند. از این‌رو، محققین در زمینه هوش مصنوعی با به‌کارگیری الگوریتم‌های داده‌کاوی موجود در یادگیری ماشین و تشخیص الگو می‌توانند کمک شایانی را به دانشمندان حوزه بیولوژیکی مولکولی کنند [۵].

۲-۱ بیولوژی سرطان

سرطان یک بیماری مدرن نیست؛ در گذشته هم وجود داشته است؛ اما امروزه به دلیل افزایش آلودگی‌های محیط زیست، افراد زیادی به کام مرگ کشیده می‌شوند و بر همین اساس تعداد بیش‌تری در مقایسه با گذشته به پزشک مراجعه می‌کنند [۱]. سرطان، یعنی رشد کنترل ناشدنی سلول‌ها در ناحیه‌ای از بدن؛ علی‌رغم این‌که چه عاملی باعث سرطان شده و چه آسیبی به سلول‌ها رسیده باشد، این وجه مشترک تمامی سرطان‌ها است. در دوران جنینی، سلول‌های سالم به سرعت رشد می‌کنند. اما این اتفاق فقط تا زمانی رخ می‌دهد که فرد به بلوغ کامل می‌رسد. در دوان بلوغ نیز در بیش‌تر نقاط بدن، سلول‌ها زمانی رشد می‌کنند و تقسیم می‌شوند که لازم است سلول‌های تازه را جایگزین سلول‌های فرسوده، مرده یا در حال مرگ کنند. در مواردی هم که بافت آسیب می‌بیند (زخم، شکستگی و ...)، سلول‌های سالم به سرعت رشد کرده و تقسیم می‌شوند تا بافت آسیب دیده را ترمیم کنند.

پس به طور کلی در حالت نرمال، سلول‌های سالم بدن طبق برنامه‌ی منظمی رشد می‌کنند، تقسیم می‌شوند و می‌میرند؛ اما سلول‌های سرطانی، برخلاف سلول‌های عادی نمی‌میرند. بلکه بی‌وقفه رشد می‌کنند، تقسیم می‌شوند و به‌طور مداوم سلول غیرعادی تولید می‌کنند. شکل ۱-۴ مراحل چگونگی سرطانی شدن سلول را نمایش می‌دهد. در ابتدا عواملی مانند ویروس، مشکلات محیطی و ... باعث جهش دی-ان-ای می‌شوند، سپس سلول‌های جهش یافته رشد نموده و منجر به تومورهای کوچک می‌شوند؛ در مراحل بعدی این تومورهای کوچک رشد بیش‌تری یافته و منجر به آسیب‌رسانی به بافت اصلی می‌شوند [۵].



شکل ۱-۱: چگونگی سرطانی شدن سلول‌ها

در چند سال اخیر با پیشرفت سریع تکنولوژی، بسیاری از عوامل مولکولی این بیماری شناخته شده است و امید برای یافتن داروهای مؤثر افزایش یافته است؛ اما عواملی که در پیدایش سرطان دخالت مستقیم و یا غیرمستقیم دارند بسیار زیاد و نامحدود هستند. به‌همین دلیل، درمان کامل این بیماری دشوار و در بسیاری موارد غیرممکن است. نکته بسیار مهمی که در این میان

حائز اهمیت است، تشخیص دقیق، دسته‌بندی و پیش‌بینی به‌موقع است. لذا به‌کارگیری تکنولوژی‌های جدید و کارا به‌جهت پیش‌بینی دقیق، بسیار مورد اهمیت است.

۳-۱ فناوری ریزآرایه

فناوری ریزآرایه^۱ در سال ۱۹۹۶ توسط شاننا^۲ با هدف مطالعه همزمان ژن‌های موجود در دو بافت سالم و سرطانی معرفی گردید [۶]. این فناوری، امکان بررسی همزمان بسیاری از فعل و انفعالات زیستی را فراهم می‌کند و در دو زمینه ژنومیکس (مطالعه مجموعه ژن‌های موجود زنده) و پروتئومیکس (مطالعه مجموعه پروتئین‌های موجود زنده) کاربردهای وسیعی دارد. بازده این روش بسیار بالا بوده و در مدت زمان کوتاهی قادر به تحلیل میزان قابل توجهی از اطلاعات است. به‌منظور پردازش اطلاعات حاصل از ریزآرایه، بیوانفورماتیک می‌تواند حامی بسیار خوبی برای این تکنولوژی باشد و نظر به پیشرفت‌های شگرف علم بیوانفورماتیک در دهه‌های اخیر، می‌توان چشم‌انداز خوبی را برای این فناوری انتظار داشت.

فناوری ریزآرایه این فرصت را در اختیار دانشمندان قرار می‌دهد تا با اندازه‌گیری میزان ام-آر-ان-ای^۳ تولیدی در هر سلول، میزان اختصاصی بیانی ژن‌ها را سنجیده و در صورت لزوم با یکدیگر مقایسه نمایند. بر این اساس بسیاری از افراد، این فناوری را انقلاب و تحولی بزرگ در تشخیص بسیاری از بیماری‌ها از جمله سرطان می‌دانند [۷].

ریزآرایه در حقیقت تکه‌ای همانند لام میکروسکوپ با ابعادی در حدود ۴×۲ سانتی‌متر است. بر روی هر ریزآرایه بعضاً تا حدود ۸۰۰۰ نوع رشته منفرد با قطر ۷۵ تا ۱۰۰ میکرومتر موجود است که در فاصله‌ای حدود ۱۵۰ میکرومتر به‌صورت ماتریسی قرار گرفته است [۳].

^۱ Microarray Technology

^۲ Schena

^۳ Messenger Ribonucleic Acid (MRNA)

توالی این رشته‌ها که شناساگر^۱ نام دارد از قبل مشخص بوده است و بسته به نوع آزمایش انتخاب می‌شود. هر شناساگر دارای میلیون رشته منفرد یکسان بوده است که وقتی بر روی اسلاید قرار می‌گیرد، نقطه^۲ نامیده می‌شود و به این اسلایدهای آماده، میکروچیپ گفته می‌شود. یک نمونه میکروچیپ در شکل ۱-۲ آورده شده است.



شکل ۱-۲: یک نمونه میکروچیپ

۱-۳-۱ بیان ژنی و فناوری ریزآرایه

فناوری ریزآرایه ابزاری بسیار قدرتمند است که امکان بررسی بیان هزاران ژن به صورت همزمان و شناسایی فعل و انفعالات پروتئینی را فراهم می‌کند. این فناوری شامل دو زیرمجموعه عمده ریزآرایه دی-ان-ای^۳ و پروتئین هست [۴].

با استفاده از ریزآرایه دی-ان-ای، امکان بررسی بیان هزاران ژن به صورت همزمان فراهم می‌شود. اهداف این گونه از تحلیل‌های ژنی عبارت است از:

- چگونگی تاثیر بیان هر ژن منفرد بر بیان ژن‌های دیگر
- چگونگی بیان ژن^۴ در سلول‌های سالم و بیمار

به‌عنوان مثال، انواع سرطان نشانه‌های مورفولوژیکی مشابهی دارند، بنابراین با استفاده از

داده‌های بیان ژن می‌توان شیوه‌های تشخیصی مستقیمی را فراهم کرد.

¹ Prob

² Spot

³ Deoxyribonucleic Acid (DNA)

⁴ Gene Expression

ازسوی دیگر، آرایه پروتئینی نوعی روش اندازه‌گیری هست که می‌تواند متخصصان پزشکی را در اندازه‌گیری وجود پروتئین‌ها در نمونه‌های زیستی، مانند خون یاری دهد. این روش باتوجه به پتانسیل بالای خود، می‌تواند یکی از ارزشمندترین روش‌های پروتئومیکس^۱ محسوب شود. امروزه به خوبی مشخص شده است که پیچیدگی پروتئوم انسان بیش‌تر از ژنوم او است. از این‌رو سرمایه‌گذاری تجاری در این زمینه از بیوتکنولوژی بسیار مهم هست.

روال تهیه مجموعه داده با کمک ریزآرایه دی-ان-ای همان‌طور که در شکل ۱-۳ نیز آورده شده است، بدین‌نحو است که پس از نمونه‌گیری، در ابتدا نمونه‌های ام-آر-ان-ای به سی-دی-ان-ای^۲ تبدیل می‌شوند و سپس با فلئورسانت^۳ نشان‌دار می‌گردند. پس از آن، نمونه‌های نشان‌دار شده با هم مخلوط شده و بعد از آن اجازه داده می‌شود که با مولکول‌های دی-ان-ای سطح اسلاید هیبرید شوند. در ادامه شستشو انجام داده می‌شود و فقط رشته‌های منفرد که با شناساگرهای مکملشان دورگه شده‌اند، باقی می‌مانند. درنهایت با تاباندن اشعه لیزر با طول موج‌های مشخص به سطح ریزآرایه توسط دستگاه‌های خاص مانند CCD^۴ عکس‌برداری می‌شود و تصاویر به کامپیوتر داده می‌شود تا تجزیه و تحلیل برطبق لکه‌های به‌جا مانده انجام گرفته شود که در این مرحله از تکنیک‌های پردازش تصویر کمک گرفته می‌شود.

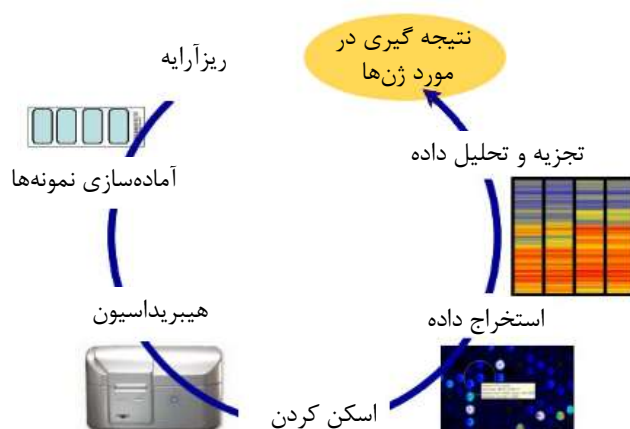
در مرحله نتیجه‌گیری، سطح روشنایی نقاط حائز اهمیت است و نشان‌دهنده میزان فعالیت و سطح بیان ژن‌های مورد نظر است. به‌عنوان نمونه، لکه‌های سبز نشان‌دهنده بیان بیش‌تر ژن نمونه دوم نسبت به نمونه اول است

¹ Proteomics

² Complementary DNA

³ Fluorescent

⁴ Charged Coupled Device



شکل ۱-۳: مراحل تهیه مجموعه داده با کمک فناوری ریزآرایه

۱-۳-۲ کاربردهای ریزآرایه دی-ان-ای

امروزه در علم پزشکی، ریزآرایه دی-ان-ای به عنوان ابزاری شناخته شده است که با کمک آن می‌توان اطلاعاتی مفید را در مورد ژن‌ها یافت و در بسیاری از کاربردها از آن کمک گرفت. برخی از مهم‌ترین کاربردهای ریزآرایه دی-ان-ای عبارتند از:

- بررسی بیان ژن و تغییرات آن در اثر عواملی مانند درمان، عوامل بیماری‌زا، آسیب سلول [۸].
- هیبرید سازی مقایسه‌ای ژنوم: تعیین محتوای ژنوم موجودات زنده و مقایسه آنها با یکدیگر [۹].
- شناسایی چندشکلی تک نوکلئوتیدی: با استفاده از این تکنولوژی می‌توان چندشکلی‌های تک نوکلئوتیدی را در میان جمعیت‌های مختلف مطالعه کرد [۱۰].

۴-۱ اهداف تحقیق

- بر اساس آنچه بیان شد، می‌توان اهداف این تحقیق را به صورت زیر بیان نمود:
- بهبود قابلیت طبقه‌بندی تشخیص سرطان با استفاده از تکنیک‌های رایج در هوش مصنوعی
 - ارائه روشی کارآمد و اتوماتیک به جهت پیش‌بینی احتمال سرطان در داده‌های ریزآرایه
 - به‌کارگیری خلاقیت در الگوریتم‌های موجود به منظور افزایش کارایی
- در این راستا این پایان‌نامه، دو سیستم خودکار به منظور پیش‌بینی احتمال سرطان پیشنهاد و ارائه می‌کند. اصول، مفاهیم و ویژگی‌های اصلی این چارچوب به طور نظری تشریح می‌شوند و سپس جهت ارزیابی عملی سیستم، مدل پیشنهادی با دیگر روش‌های موجود مقایسه می‌شود.

۵-۱ ساختار پایان‌نامه

در این پایان‌نامه، سیستم پیشنهادی به جهت پیش‌بینی احتمال سرطان با کمک تکنیک‌های یادگیری ماشین مورد بررسی قرار گرفته است که در زیر شرح مختصر مباحث انجام گرفته در هر فصل آورده شده است:

فصل ۲: در این فصل مروری بر تحقیقات پیشین و چالش‌های موجود در این زمینه مورد بررسی قرار می‌گیرد.

فصل ۳: این فصل دربرگیرنده مبانی تحقیق، تشریح مجموعه داده‌های مورد استفاده و شرح تکنیک‌ها و الگوریتم‌های به‌کارگرفته شده در روش پیشنهادی به منظور تحلیل داده‌های ریزآرایه‌ای سرطان است.

فصل ۴: این فصل نکات اصلی و کلی در مورد چگونگی طراحی و پیاده‌سازی سیستم‌های پیشنهادی و نتایج حاصله از روش‌های پیشنهادی را شرح می‌دهد.

فصل ۵: در این فصل نتیجه‌گیری کلی از پایان‌نامه انجام می‌شود و همچنین پیشنهادهایی برای ادامه‌ی کار در آینده ارائه می‌گردد.