

ارائه روش استخراج ویژگی شبه کوواریانسی مبتنی بر تبدیل موجک جهت کشف نشانگر حیاتی از الگوهای پروتئینی سرطان تخمدان

حسین منتظری کردی، محمدحسین میران بیگی و محمدحسن مرادی

رفتار ژنی درون یک سلول، بافت و یا ارگان حیاتی حاصل می‌شود، به کار برده شد [۲]. واژه پروتئین‌شناسی^۳ به علمی اطلاق می‌شود که امکان مقایسه کمی و کیفی و همچنین تمایز قابل‌شدن بین پروتئوم‌ها تحت شرایط مختلف (سالم یا نرمال بر حسب بیمار یا سرطانی) برای فهم فرآیندهای بیولوژیک (نظیر شناسایی عوامل منسوب به بیماری) را فراهم نماید [۳]. به همین خاطر حوزه فعالیت پروتئین‌شناسی می‌تواند به‌طور عام به هرگونه تکنولوژی و یا تکنیک پردازش اطلاعات که بر روی تحلیل الگوهای داده پروتئینی با ابعاد بالا تمرکز نماید، اطلاق گردد [۴].

طیف‌سنجی جرمی^۴ تکنیکی است که جهت تهیه اطلاعات درباره پروتئین‌ها و قسمت‌های تشکیل‌دهنده^۵ آن به کار می‌رود. طیف‌سنج جرم می‌تواند اطلاعات کمی و کیفی از یک عنصر، مولکول و یا ترکیب موجود زنده را ارائه دهد [۵]. لذا تحلیل طیف جرم می‌تواند یک روش سریع و ارزان در تشخیص بیماری تنها بر اساس یک نمونه از خون بیمار باشد و امکان بالقوه غربال‌گری سرطان را بدون هرگونه عوارض جانبی به هنگام تشخیص فراهم می‌سازد. در بین بیماری‌های مربوط به زنان، سرطان تخمدان یکی از عوامل اصلی مرگ و میر به شمار می‌رود. کاربرد طیف‌سنجی جرمی برای تشخیص سرطان تخمدان می‌تواند تأثیر به‌سزایی در سلامت و بهداشت عمومی جامعه زنان داشته باشد، اما جهت دستیابی به این هدف استفاده از نشانگرهای حیاتی^۶ جدید با قدرت تشخیص بالاتر ضروری است [۶].

از نقطه نظر مدل‌سازی، طیف جرم می‌تواند به‌عنوان داده عملکردی^۷ پیچیده در نظر گرفته شود که ویژگی‌های کلیدی مهم برای محققان، پیک‌های موجود در آن می‌باشد [۷]. پیک‌ها بیانگر حضور پروتئین‌ها و یا پپتیدها^۸ در الگوی پروتئینی حاصل از طیف‌سنجی جرمی می‌باشند. از سویی دیگر، داده طیف جرم سطری دارای ابعاد و همبستگی بالا بین نقاط طیف جرمی بوده و دارای مقدار قابل توجهی نویز می‌باشد که استفاده مستقیم از آن جهت استخراج ویژگی مناسب نمی‌باشد. لذا یک روش بهتر پیش‌پردازش که به‌همراه کاهش بعد داده و حذف نویز عمل انتخاب ویژگی را نیز به‌خوبی انجام دهد، نیاز خواهد بود.

توسعه روش‌های تشخیص سرطان مبتنی بر الگوی پروتئینی نمونه خون در گذشته بدون توجه به حذف اثر نویز و کاهش همبستگی بین

چکیده: تغییرات پاتولوژیک درون یک عضو حیاتی به‌صورت الگوهای پروتئینی در خون انعکاس می‌یابد. طیف‌سنجی جرمی به‌عنوان یک ابزار اندازه‌گیری قدرتمند در تولید الگوهای پروتئینی از خون شناخته شده است. داده‌های حاصل از این تکنیک به‌عنوان داده‌هایی با ابعاد و همبستگی بالا در نظر گرفته می‌شوند که ویژگی‌های کلیدی مهم برای محققان، پیک‌های موجود در طیف می‌باشند. با توجه به این خصوصیات رفتاری داده، یک روش تحلیل مناسب مثل تبدیل موجک نیاز می‌باشد. در این تحقیق، روش جدیدی جهت استخراج ویژگی از داده طیف جرمی با هدف کاهش بعد و همبستگی آن ارائه شده است که مبتنی بر تبدیل موجک گسسته و انتخاب ویژگی شبه‌کوواریانسی می‌باشد. با روش پیشنهادی، نشانگرهای حیاتی مناسب از مجموعه داده‌های مورد مطالعه برای سرطان تخمدان که از انستیتو ملی سرطان آمریکا اخذ شده است، از روی طیف جرمی بازسازی شده استخراج گردیده است که منجر به حصول نتایج تشخیصی بالا با استفاده از معیارهای ارزشیابی استاندارد شده است. با استفاده از روش‌های طبقه‌بندی مختلف، روش پیشنهادی جدید منجر به دقت تشخیص ۹۸٪ نرخ قطعیت ۹۷٪ و حساسیت ۹۸٪ درصد شده است.

کلید واژه: پروتئین‌شناسی، تشخیص الگو، تبدیل موجک گسسته، تابع وزن شبه‌کوواریانسی، نشانگر حیاتی.

۱- مقدمه

نیاز به توسعه ابزارهایی جهت تشخیص به موقع و زودرس بیماری سرطان یکی از مسایل اساسی تحقیقاتی روز دنیا می‌باشد. محققان پزشکی در این راستا تلاش‌هایی نموده‌اند که حاصل کوشش آنها منجر به ارائه برخی روش‌های تشخیصی شده است. از معایب عمده این روش‌ها می‌توان به عدم دقت کافی و همچنین هزینه بالای آنها اشاره نمود. در سال‌های اخیر این حقیقت آشکار شده است که تغییرات پاتولوژیک درون یک عضو از بدن به‌صورت الگوهای پروتئینی در خون انعکاس می‌یابد [۱]. از لحاظ واژه‌شناسی لغت پروتئوم^۱ برای اولین بار در سال ۱۹۹۴ جهت نام‌گذاری برای مجموعه کاملی از پروتئین‌ها که در نهایت از تغییر

این مقاله در تاریخ ۱۵ خرداد ماه ۱۳۸۶ دریافت و در تاریخ ۲۹ آذر ماه ۱۳۸۷ بازنگری شد.

حسین منتظری کردی، بخش مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، (email: hmontazery@modares.ac.ir).

محمدحسین میران بیگی، بخش مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، (email: Miranbhm@modares.ac.ir).

محمدحسن مرادی، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، (email: mhmoradi@aut.ac.ir).

1. Proteome

2. Transcription

3. Proteomics

4. Mass Spectrometry

5. Fragments

6. Biomarkers

7. Functional

8. Peptides

۲- مواد و پیش پردازش

در این مقاله از داده‌های پروفایل پروتئینی خونابه^۲ که توسط دستگاه طیف‌سنجی جرمی با تکنیک جذب-یونیزاسیون لیزری سطحی ارتقایافته زمان پروازی^۳ (SELDI-TOF) تهیه شده است، به‌عنوان داده ورودی برای الگوریتم‌های پردازشی استفاده شده است. سپس با استفاده از این داده‌ها غربالگری بیماران مبتلا به سرطان تخمدان و افراد تحت کنترل سالم به‌منظور دستیابی به نشانگرهای حیاتی متمایزکننده دو گروه انجام شده است.

۲-۱ داده‌های طیف‌سنجی جرمی

یک سیگنال طیف جرم، یک منحنی می‌باشد که محور افقی آن یا x نشان‌دهنده مقدار نسبت جرم به بار یک مولکول خاص است که با توجه به یونیزه‌شدن مولکول‌ها با یک بار الکتریکی، در واقع بیانگر جرم یک مولکول بر اساس واحد دالتون می‌باشد. محور عمودی طیف جرم یا y به تعداد مولکول‌های رسیده‌شده به آشکارساز با جرم یکسان اشاره دارد که در اینجا محور شدت نام‌گذاری شده است. محور شدت در اصل به میزان فراوانی یک مولکول با جرم معین در یک ترکیب اشاره دارد. بنابراین در منحنی طیف جرمی وجود هر مولکول یا پروتئین با جرم معین به‌صورت پیک‌هایی روی محور شدت نمایش داده می‌شود که شناسایی دقیق محل پیک روی محور نسبت جرم به بار در نهایت منجر به دستیابی نشانگر حیاتی می‌گردد.

در این تحقیق جهت تعیین الگوهای پروتئینی از خونابه که باعث تمایز و تشخیص نمونه‌های مبتلا به سرطان تخمدان از افراد تحت کنترل سالم شود، سه مجموعه داده طیف جرمی حاصل از SELDI-TOF مورد استفاده قرار گرفت. این مجموعه داده‌ها از انستیتو ملی سرطان آمریکا اخذ گردید که صحت این داده‌ها مورد تأیید محققان می‌باشد [۸] و [۱۴]. با توجه به توضیحات ارائه‌شده توسط انستیتو، مجموعه داده I با استفاده از تراشه پروتئینی H₄ تهیه شده است که مشتمل بر ۲۱۶ نمونه به تفکیک ۱۰۰ نمونه افراد سالم، ۱۰۰ نمونه افراد مبتلا به سرطان تخمدان و ۱۶ نمونه از مبتلایان با تومور خوش‌خیم می‌باشد. مجموعه داده II با استفاده از تراشه WCX₂ به‌دست آمده است که در این مجموعه افراد تحت مطالعه با نمونه‌های مجموعه I یکسان می‌باشند و در آماده‌سازی نمونه از تکنسین آزمایشگاه استفاده شده است. این مجموعه نیز مشتمل بر ۲۱۶ نمونه به تفکیک ۱۰۰ نمونه افراد سالم، ۱۰۰ نمونه بیماران مبتلا به سرطان تخمدان و ۱۶ نمونه از مبتلایان به بیماری با تومور خوش‌خیم می‌باشد. مجموعه III از روی نمونه خونابه افراد با جامعه آماری متفاوت از دو مجموعه I و II با استفاده از تراشه پروتئینی WCX₂ تهیه شده است با این تمایز که در مرحله آماده‌سازی نمونه از یک بازوی رباتیک به‌جای انسان کمک گرفته شده است. این مجموعه نیز مشتمل بر ۲۵۳ نمونه به تفکیک ۹۱ نمونه افراد سالم و ۱۶۲ نمونه بیماران مبتلا به سرطان تخمدان می‌باشد.

در این مجموعه داده‌ها، هر منحنی طیف جرم دارای ۱۵۱۵۴ نقطه روی محور نسبت جرم به بار (M/Z) می‌باشد که متناظر با آن یک نقطه روی محور شدت سیگنال وجود دارد. شکل ۱ سه نمونه از منحنی طیف جرمی را برای داده‌های تحت مطالعه نشان می‌دهد که به‌ترتیب مربوط به یک فرد سالم، دارای تومور خوش‌خیم و بیمار سرطان تخمدان

ویژگی‌های داده طیف جرمی انجام شده است [۸] و [۹]. در کارهای قبلی دستیابی به دقت‌های تشخیص بالا مد نظر قرار گرفت که نتایج به‌دست آمده به‌طور قابل ملاحظه‌ای به مجموعه داده‌های یادگیری و آموزش وابسته بوده است [۱۰] و [۱۱]. اکثر این روش‌ها به‌نوعی توسط مجموعه داده‌های یادگیری بایاس شده‌اند و نتایج حاصله به‌دنبال ارائه دقت‌های تشخیصی بالا بدون توجه به تکرارپذیری آن بوده است [۱۲]. در حال حاضر، کاهش ابعاد ویژگی داده با توجه به تعداد زیاد نقاط ویژگی و تعداد کم نمونه‌های داده از مشکلات اساسی این حوزه به‌شمار می‌آید [۱۳]. مشکل عمده دیگر حوزه پروتئین‌شناسی، عدم تکرارپذیری نتایج در یک مجموعه داده تحت مطالعه و تغییر زیاد در نتایج طبقه‌بندی با انتخاب تصادفی مجموعه‌های یادگیری و آموزش می‌باشد.

تبدیل موجک گسسته^۱ (DWT) یک ابزار مفید در کاهش بعد و حذف نویز از سیگنال به‌شمار می‌آید که می‌تواند در تحلیل داده پروتئینی مورد استفاده قرار گیرد. موجک‌ها به‌دلیل دارا بودن قابلیت تحلیل همزمان ویژگی‌های محلی و عام یک تابع، در پردازش سیگنال به‌طور گسترده‌ای مورد استفاده قرار گرفته‌اند. تبدیل موجک گسسته می‌تواند نگاهی به طیف جرمی بر روی یک فضای متعامد باشد که پایه موجک نامیده می‌شود. لذا تحلیل موجک می‌تواند نمایشی بدون نویز و کاهش بعد یافته از داده طیف جرمی ارائه دهد که فرآیند استخراج ویژگی را مؤثرتر و دقیق‌تر می‌سازد. این خاصیت می‌تواند ناشی از ویژگی‌های مناسبی از تبدیل موجک نظیر ساختار سلسله‌مراتبی و چندوضوحی، ضرایب غیر همبسته، توابع پایه متعامد و امکان انتخاب توابع پایه متعامد متنوع با خواص متفاوت باشد.

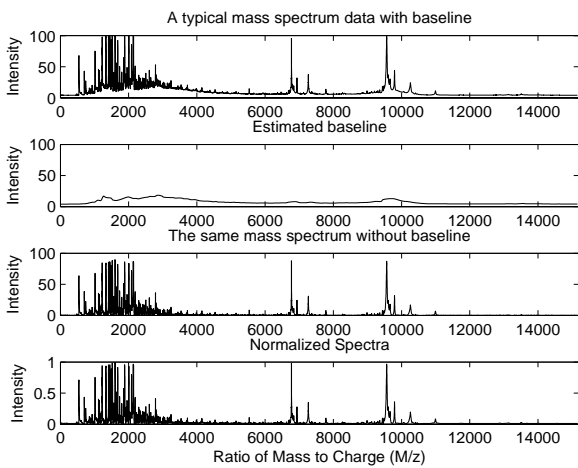
در این مقاله از یک روش پیش‌پردازش مبتنی بر تبدیل موجک جهت استخراج ویژگی از داده طیف جرمی بیماران سرطان تخمدان برای طبقه‌بندی گروه سالم از بیمار به‌منظور حصول به نشانگرهای حیاتی جدید استفاده شده است. یک روش جدید انتخاب ویژگی با هدف کاهش همبستگی بین نقاط داده طیف جرمی جهت گزینش زیرمجموعه‌ای از ویژگی‌ها به‌منظور کاهش بعد و دستیابی به نشانگر حیاتی با قابلیت تکرارپذیری ارائه شده است. پس از عمل کاهش بعد، از طیف بازسازی شده جهت دستیابی به نشانگرهای حیاتی و بررسی عملکرد طبقه‌بندی استفاده گردیده است. نتایج به‌دست آمده حاکی از این است که پروفایل‌های پروتئینی طیف‌سنجی جرمی توان تشخیص بیماران مبتلا به سرطان تخمدان از افراد سالم را دارا می‌باشند. لذا طیف جرم بازسازی شده مبتنی بر تبدیل موجک با روش ترکیبی استخراج و انتخاب ضرایب می‌تواند یک روش مفید در تشخیص سرطان تخمدان محسوب شود. با استفاده از تکنیک ارائه‌شده در این مقاله، برای مجموعه داده‌های تحت مطالعه دقت تشخیص برابر ۹۸، قطعیت تشخیص ۹۷ و حساسیت تشخیص ۹۸ درصد حاصل شده است.

در ادامه مقاله، در بخش دوم مواد و مجموعه داده مورد استفاده در تحقیق توصیف شده و همچنین مراحل پیش‌پردازش اعمال شده به داده طیف جرمی، تشریح شده است. در بخش سوم با ارائه مقدمه‌ای راجع به استخراج ویژگی با تبدیل موجک گسسته، روش جدید انتخاب زیرمجموعه ویژگی معرفی می‌شود. بخش چهارم نتایج حاصل از تحلیل داده طیف جرمی سرطان تخمدان با روش ارائه‌شده را در بردارد و سرانجام در بخش پنجم بحث و نتیجه‌گیری از تحلیل داده طیف جرمی با رهیافت جدید ارائه می‌شود.

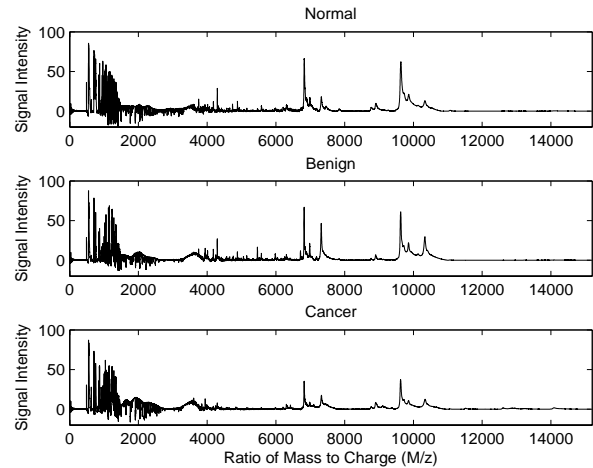
2. Serum

3. Surface Enhanced Laser Desorption/Ionization Time-of-Flight

1. Discrete Wavelet Transform



شکل ۲: یک نمونه از سیگنال طیف جرمی قبل و بعد از حذف نویز خط زمینه به همراه منحنی نرمالیزه شده از همان طیف جرم.



شکل ۱: نمونه‌ای از منحنی طیف جرمی برای افراد سالم، تومور خوش خیم و سرطانی.

جدول ۱: نحوه توزیع نمونه‌ها در سه مجموعه داده تحت مطالعه.

| مجموعه داده | تعداد نمونه‌های سالم | تعداد نمونه‌های سرطانی | تعداد نمونه‌های با تومور خوش خیم | تعداد ویژگی‌ها در هر طیف جرم |
|-------------|----------------------|------------------------|----------------------------------|------------------------------|
| مجموعه I | ۱۰۰ | ۱۰۰ | ۱۶ | ۱۵۱۵۴ |
| مجموعه II | ۱۰۰ | ۰ | ۱۶ | ۱۵۱۵۴ |
| مجموعه III | ۹۱ | ۱۶۲ | ۰ | ۱۵۱۵۴ |

به ترتیب مقادیر حداقل و حداکثر شدت در هر سیگنال را نشان می‌دهند. شکل ۲ نمونه‌ای از منحنی طیف جرم را قبل از حذف خط زمینه و بعد از حذف آن نشان می‌دهد که در این شکل خط زمینه تخمین زده شده به همراه طیف نرمالیزه شده نیز مشاهده می‌شود.

۳- روش استخراج و انتخاب ویژگی

۳-۱ استخراج ویژگی

همان طوری که در بخش‌های قبلی ذکر شد، طیف جرم داده‌ای، عملکردی با رفتار تیز است که ویژگی‌های کلیدی آن پیک‌ها می‌باشند. از این رو، روش تحلیلی می‌بایست توانایی در نظر گرفتن رفتارهای محلی و کلی منحنی طیف را دارا بوده و همچنین قابلیت استخراج ویژگی‌های وابسته به شکل سیگنال را داشته باشد. این واقعیت ما را بر می‌انگیزاند تا از روشی مانند موجک جهت استخراج ویژگی از داده استفاده نماییم تا توانایی استخراج این اطلاعات بالقوه از طیف جرم حاصل شود. موجک‌ها خانواده‌ای از توابع پایه متعامد یک‌ه می‌باشند که می‌توانند جهت بسط توابع دیگر مورد استفاده قرار گیرند. در تحلیل موجک، یک سیگنال یا بردار (نظیر منحنی طیف جرم) به مجموعه‌ای از تقریبات^۱ و جزئیات^۲ که شکل ساده‌تری از تابع می‌باشند، با مقیاس‌بندی و شیفت‌دادن تجزیه می‌شوند. در تبدیل موجک چندوضوحی^۳ فضای نگاشت مربوط به تقریب سطح اول، به دو زیرفضای تقریب و جزئیات دیگر تبدیل می‌شود [۱۷]. با این روش، تبدیل جهت حذف اثر نویز کارایی بهتری داشته و هر چقدر که در بین سطوح به سمت پایین می‌رویم، ضرایب ویژگی‌های تیز بزرگ‌تر مانده و ضرایب کوچک‌تر به ویژگی‌های هموارتر مربوط می‌گردد. در فضای توابع با انرژی محدود، $L^2(\mathcal{R})$ یک موجک پایه متعامد با

با اثبات بیماری به روش بیوپسی می‌باشد. جدول ۱ نیز نحوه توزیع نمونه‌ها را در این سه مجموعه از داده نشان می‌دهد.

۲-۲ مرحله پیش پردازش

مدل مفهومی زیر را می‌توان برای داده طیف‌سنجی جرمی در نظر گرفت. فرض می‌شود که n طیف مشاهده شده است، که هر یک به‌طور یکنواخت در بازه زمانی به طول T از زمان پرواز در زمان‌های $t_j, j=1, K, T$ نمونه‌برداری شده‌اند. یک مدل برای سیگنال طیف جرمی $y_i(t_j)$ می‌تواند به صورت زیر ارائه شود [۷] و [۱۵]

$$y_i(t_j) = B_i(t_j) + N_i S_i(t_j) + \varepsilon_{ij} \quad (1)$$

که $i=1, 2, K, n$ و $t=1, 2, K, m$ مربوط به پالس ساعت زمان پرواز آشکارساز جهت محاسبه نسبت جرم به بار می‌باشد. $y_i(t_j)$ شدت سیگنال مشاهده شده برای طیف در زمان t_j است. $B_i(t_j)$ به یک خطای سیستم ناشی از جذب و بازپس‌دادن انرژی توسط ماده جاذب به نمونه بیولوژیک مربوط می‌شود که به عنوان خطای زمینه شناخته می‌شود. $S_i(t_j)$ پیک‌های واقعی سیگنال را نشان می‌دهد که توسط ضریب N_i مقیاس‌بندی شده است. ε_{ij} معرف نویزهای دیگر سیستم از جمله اغتشاش منابع الکتریکی و منبع لیزر می‌باشد که در این مدل گوسی با میانگین صفر و واریانس σ_y وابسته به زمان نمونه‌برداری فرض می‌شود.

با توجه به این مدل، چند مرحله پیش‌پردازش به منحنی طیف جرم اعمال شد. در ابتدا هر سیگنال طیف به‌طور جداگانه جهت حذف نویز خط زمینه مورد پردازش قرار گرفت که روش استفاده شده در [۱۶] توصیف شده است. سپس طیف‌های بدون خط زمینه جهت حذف اثر مقیاس با استفاده از رابطه زیر نرمالیزه شدند

$$NS = \frac{S - \min(S)}{\max(S) - \min(S)} \quad (2)$$

که NS به مقدار شدت طیف نرمالیزه شده و S به مقدار آن قبل از نرمالیزاسیون اشاره دارد. در این رابطه عبارات $\max(S)$ و $\min(S)$

1. Spiky
2. Approximations
3. Details
4. Multiresolution

طیف جرم می‌باشد، یک تابع وزن شبه کوواریانسی به نقاط ویژگی اعمال می‌شود. اثر این تابع به گونه‌ای است که در صورت همبسته بودن نقاط با اولین نقطه ویژگی منتخب به آن یک وزن پایین اختصاص می‌دهد. فرض می‌شود که دو کلاس الگوی ω_1 و ω_2 به ترتیب با تعداد عضو N_1 و N_2 موجود باشد. هر عضو از این کلاس‌ها به صورت $X_j = (x_{j1}, x_{j2}, \dots, x_{jK}, x_{jp})$, $j = 1, 2$ ویژگی هر بردار از یک کلاس الگو برابر p در نظر گرفته شده است. مقادیر میانگین و واریانس برای هر کلاس با استفاده از روابط زیر محاسبه می‌شود

$$\mu_{jk} = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ik}, \quad k = 1, 2, K, p \quad (8)$$

$$\sigma_{jk} = \left[\frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ik} - \mu_{jk})^2 \right]^{\frac{1}{2}}, \quad k = 1, 2, K, p \quad (9)$$

با محاسبه مقادیر میانگین و واریانس، معیار فیشرفر جهت رتبه‌بندی اندیس‌های ویژگی، k از رابطه زیر به دست می‌آید

$$F_k = \frac{(\mu_{1k} - \mu_{2k})^2}{\sigma_{1k}^2 + \sigma_{2k}^2}, \quad k = 1, 2, K, p \quad (10)$$

با استفاده از معیار فیشرفر جهت رتبه‌بندی، اندیس‌های ویژگی با حداکثر مقدار تابع F انتخاب می‌شوند. از آنجایی که در طیف‌سنجی جرمی هر مولکول خاص علاوه بر پیک‌زدن در مکان واقعی نسبت جرم به بار خود در مکان‌های دیگری نیز دارای پیک‌های خفیف‌تر خواهد بود [۲۲]، لذا بین نقاط ویژگی در بردار الگو همبستگی وجود دارد. در روش پیشنهادی جهت کاهش همبستگی از تابع وزن شبه کوواریانسی برای محاسبه تابع فیشرفر به صورت زیر استفاده شده است

$$F_k^0 = \frac{(\mu_{1k} - \mu_{2k})^2}{\sigma_{1k}^2 + \sigma_{2k}^2} w_k, \quad k = 1, 2, K, p \quad (11)$$

در رابطه فوق w_k یک ضریب وزنی می‌باشد که به طور جداگانه پس از هر بار انتخاب یک اندیس ویژگی محاسبه می‌شود. در این روش وزن‌دهی [۲۳] تا [۲۵]، تمام داده‌های موجود در کلاس‌های الگو را در یک ماتریس داده $\mathbf{D}_{N \times P}$ قرار می‌دهیم. این ماتریس دارای P ستون و $N = N_1 + N_2$ سطر خواهد بود. در ابتدا یک پارامتر طول برای هر متغیر تعریف می‌شود

$$\lambda_j = \left[\frac{1}{N} \sum_{i=1}^N (d_{ij})^2 \right]^{\frac{1}{2}}, \quad j = 1, 2, K, P \quad (12)$$

در این رابطه d_{ij} به هر درایه از ماتریس داده اشاره دارد. سپس با استفاده از پارامتر طول، ماتریس داده مطابق رابطه زیر نرمالیزه خواهد شد

$$d(\lambda)_{ij} = \frac{d_{ij}}{\lambda_j} \quad (13)$$

در رابطه بالا جهت جلوگیری از تأثیر نقاط با شدت خیلی کم در ماتریس داده می‌توان یک پارامتر تصحیح، α ، را به تابع طول اضافه کرد. با فرض توزیع یکنواخت برای تمام ویژگی‌های ماتریس داده، تابع وزن شبه کوواریانسی از رابطه زیر محاسبه می‌شود

$$\mathbf{C} = \frac{1}{N} \mathbf{D}(\lambda)^T \mathbf{D}(\lambda) \quad (14)$$

در رابطه فوق، عملگر $(\bullet)^T$ به ترانهاده ماتریس اشاره دارد. در روش پیشنهادی، ابتدا اولین اندیس ویژگی از روی (۱۰) محاسبه می‌شود که برابر با اندیس حداکثر مقدار این تابع می‌باشد. فرض می‌شود که اولین

شبه‌تبدیل دادن و تغییر مقیاس تابع موجک مادر Ψ به صورت $\Psi_{j,k}(x) = 2^{-j/2} \Psi(2^{-j}x - k)$ به دست می‌آید که در این رابطه j و k مقادیر صحیح می‌باشند. تابع $f(x)$ به طول N را می‌توان توسط تابع مقیاس $\Phi(x)$ (که تابع موجک $\Psi(x)$ از روی آن ساخته می‌شود) و همچنین تابع موجک مادر به صورت زیر بسط داد

$$f(x) = \langle f, \Phi \rangle \Phi(x) + \sum_{j=0}^J \sum_{k=-\infty}^{\infty} \langle f, \Psi_{j,k} \rangle \Psi_{j,k}(x) \quad (3)$$

مجموعه $\{\langle f, \Phi \rangle, \langle f, \Psi_{j,k} \rangle\}$ به ازای $j = 0, 1, K, 2^J, k = 0, 1, K, 2^j - 1$ مجموعه ضرایب موجک نامیده می‌شود. جهت حذف اثر نویز با استفاده از تبدیل موجک شیوه آستانه‌دهی ضرایب استفاده می‌شود. دو روش آستانه‌دهی وجود دارد که به نام‌های آستانه‌دهی سخت و نرم شناخته می‌شوند [۱۸]. روابط زیر تعاریف ریاضی این دو روش را ارائه می‌دهند

$$\eta_H(w, t) = \begin{cases} w & |w| \geq t \\ 0 & |w| < t \end{cases} \quad (4)$$

$$\eta_S(w, t) = \begin{cases} w-t & w \geq t \\ 0 & |w| < t \\ w+t & w \leq -t \end{cases} \quad (5)$$

در روابط فوق w و t به ترتیب به مقادیر ضرایب موجک و آستانه تعیین شده جهت تعدیل ضرایب اشاره دارند. در این مقاله از روش آستانه‌دهی سخت با نگاه داشتن ضرایب تقریب در آخرین سطح تحلیل استفاده شده است که مقدار آستانه مناسب با روش داناهو-جانستون^۱ از رابطه زیر محاسبه می‌شود [۱۹] و [۲۰]

$$t = \sigma^2 \sqrt{2 \ln N} \quad (6)$$

$$\sigma^2 = \frac{\text{median}(\bar{w})}{0.6745} \quad (7)$$

در روابط فوق N طول سیگنال و $\bar{w} = \{|w_1|, |w_2|, \dots, |w_n|\}$ می‌باشد. از آنجایی که هدف استخراج ضرایب مرتبط با ویژگی‌های مهم در داده می‌باشد، لذا از روش آستانه‌دهی سخت استفاده شده است.

۳-۲ روش انتخاب ویژگی شبه کوواریانسی

روش‌های انتخاب ویژگی به طور کلی به دو دسته فیلتری و مبتنی بر یادگیری^۲ تقسیم می‌شوند. در روش‌های فیلتری، انتخاب‌کننده ویژگی مستقل از الگوریتم یادگیری ویژه مورد استفاده در طبقه‌بندی می‌باشد و از آن به عنوان یک فیلتر جهت رد یا قبول نقاط ویژگی نامناسب بهره گرفته شده است. به عبارت دیگر در روش‌های مبتنی بر یادگیری، انتخاب‌کننده ویژگی به عنوان یک معیار وابسته به الگوریتم یادگیری معین در انتخاب ویژگی‌های مناسب شرکت می‌کند [۲۱]. در این مقاله، یک روش فیلتری انتخاب ویژگی شبه کوواریانسی جهت گزینش پیک‌های مناسب در فضای نگاشت تبدیل موجک از داده طیف جرمی با هدف کاهش همبستگی بین نقاط ویژگی ارائه شده است.

در روش پیشنهادی از یک تست آماری به عنوان تابع اندازه‌گیر فاصله به نام معیار فیشرفر استفاده شده است. با استفاده از این تابع فاصله، ابتدا هر ویژگی در فضای نگاشت رتبه‌بندی می‌شود و اولین اندیس ویژگی با حداکثر مقدار رتبه انتخاب می‌گردد. در مرحله بعدی، جهت کاهش میزان همبستگی بین نقاط ویژگی که ناشی از طبیعت ذاتی منحنی

1. Donoho - Johnstone
2. Wrapper

تشخیص بیماری استفاده شد [۲۹]. در مجموعه داده تحت مطالعه، T_P تعداد تشخیص مثبت بیماری برای افراد بیمار، F_N تعداد تشخیص منفی بیماری برای افراد بیمار، F_P تعداد تشخیص مثبت بیماری برای افراد سالم و T_N تعداد تشخیص منفی بیماری برای افراد سالم تعریف شده است. با استفاده از معرف‌های بالا، معیارهای تشخیص به صورت دقت A_c ، قطعیت S_p و حساسیت S_n با روابط زیر تعریف می‌شود

$$A_c = \frac{T_P + T_N}{T_P + F_N + F_P + T_N} \quad (17)$$

$$S_p = \frac{T_N}{T_N + F_P} \quad (18)$$

$$S_n = \frac{T_P}{T_P + F_N} \quad (19)$$

داده‌های مجموعه یادگیری پس از بازسازی و انتخاب ویژگی با روش پیشنهادی، جهت طبقه‌بندی و نشان‌دادن عدم وابستگی نشانگرها به نوع طبقه‌بند همراه مجموعه آزمایش توسط طبقه‌بندهای متداولی مورد بررسی قرار گرفت. در این بررسی، انتخاب نشانگر حیاتی در فضای داده‌های بازسازی شده صورت گرفته است و عمل آموزش و آزمون توسط مجموعه داده‌های یادگیری و آزمایش در فضای اصلی انجام شده است. طبقه‌بندهای مورد استفاده عبارتند از: تحلیل ممیز خطی (LDA) ، K همسایه نزدیک تر $(K-NN)$ ، و ماشین بردار پشتیبان (SVM) [۳۰].

با استفاده از روش پیشنهادی، از هر مجموعه تقریب و جزئیات در فضای ویولت ۳ ضریب انتخاب شد و سپس طیف جرم از روی آن بازسازی گردید. سپس، تعداد ۱۰ اندیس جرم به عنوان نشانگرهای حیاتی از طیف جرم بازسازی شده برای ارزیابی عملکرد تشخیص انتخاب شد.

شکل ۳ توزیع اندیس‌های ضرایب منتخب ویولت با سطح تحلیل ۱۰ و ویولت مادر داییشیز مرتبه ۴ را نشان می‌دهد که همچنین دو نمونه طیف جرمی بازسازی شده از روی ضرایب زنده پس از عمل انتخاب در این شکل مشاهده می‌شود. این دو طیف از مجموعه داده III می‌باشد که پس از عمل کاهش بعد و حذف نویز با روش پیشنهادی در متن مقاله بازسازی شده‌اند. در مسایل بازشناسی الگو و طبقه‌بندی، استفاده از تمامی نقاط بردار سیگنال ورودی غیر عملی می‌باشد لذا به کارگیری یک روش مؤثر کاهش بعد و انتخاب ویژگی مفید خواهد بود.

جدول ۲ نتایج حاصل از طبقه‌بندی داده‌های طیف جرمی لیست شده در جدول ۱ را نشان می‌دهد. از آنجایی که در مرحله اول هدف یافتن نشانگرهای حیاتی با بیشترین قدرت تشخیص در بین مجموعه‌های تحت مطالعه بود، لذا از طبقه‌بند پایه تحلیل ممیز خطی (LDA) استفاده شد. همان طوری که در جدول ۲ مشاهده می‌شود، نتایج حاصله با روش پیشنهادی دارای عملکرد خوبی با توجه به معیارهای تعریف شده در (۱۷) تا (۱۹) می‌باشد.

در مرحله دوم جهت نشان‌دادن پایداری نشانگرهای منتخب از نظر عدم وابستگی به نوع طبقه‌بند، از چند طبقه‌بند متداول استفاده شد. جدول ۳ نتایج حاصل از طبقه‌بندی داده‌های طیف جرمی را با استفاده از

3. Accuracy
4. Specificity
5. Sensitivity
6. Linear Discriminate Analysis
7. K - Nearest Neighbors
8. Support Vector Machines

اندیس منتخب را با پارامتر p_1 نشان دهیم. جهت محاسبه اندیس وزن برای (۱۱) به منظور به دست آوردن اندیس دوم ویژگی، از رابطه زیر استفاده می‌شود

$$w_{i,2} = \begin{bmatrix} c_{ii} & c_{ip_1} \\ c_{p_1i} & c_{p_1p_1} \end{bmatrix}, \quad C = [c_{ij}]_{n \times n} \quad (15)$$

در این رابطه زیرنویس ۲ به انتخاب اندیس ویژگی دوم از بردارهای ویژگی اشاره دارد و $w_{i,2}$ اندازه‌ای جهت استقلال متغیرهای i و p_1 از یکدیگر می‌باشد که در اینجا وزن نام‌گذاری شده است. مقدار وزن برای ویژگی‌های همبسته مقدار کم و غیر همبسته مقدار زیاد خواهد داشت. عملگر $|\bullet|$ به مقدار دترمینان درایه‌های منتخب از ماتریس همبستگی اشاره دارد. در حالت کلی برای $z \geq 2$ ، تابع دترمینان به صورت زیر تعریف می‌شود

$$w_{i,j} = \begin{bmatrix} c_{ii} & c_{ip_1} & K & c_{i,p_{j-1}} \\ c_{p_1i} & c_{p_1p_1} & K & c_{p_1,p_{j-1}} \\ M & M & M & M \\ c_{p_{j-1}i} & K & K & c_{p_{j-1},p_{j-1}} \end{bmatrix}, \quad j = 2, 3, K, v \quad (16)$$

با استفاده از این روش، ابتدا اندیس اول ویژگی با (۱۰) محاسبه می‌شود و سپس در هر بار انتخاب ویژگی می‌بایست مقدار وزن از (۱۶) محاسبه شده و در (۱۱) قرار داده شود. با محاسبه وزن جهت یافتن اندیس‌های بعدی فقط به یک تابع حداکثر نیاز خواهد بود که اندیس حداکثر مقدار در (۱۱) نشان‌دهنده ویژگی دوم منتخب می‌باشد. این عمل تا محاسبه اندیس‌های ویژگی مورد انتظار ادامه می‌یابد که تعداد این اندیس‌ها با تعریف یک مقدار خطای طبقه‌بندی می‌تواند توسط کاربر تعیین شود. علت نام‌گذاری شبه‌کوواریانس بدین خاطر است که (۱۴) شباهت زیادی به ماتریس کوواریانس دارد اما به طور دقیق با آن برابر نمی‌باشد.

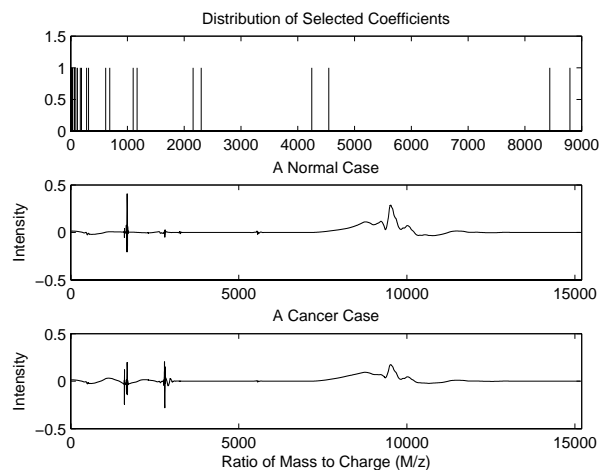
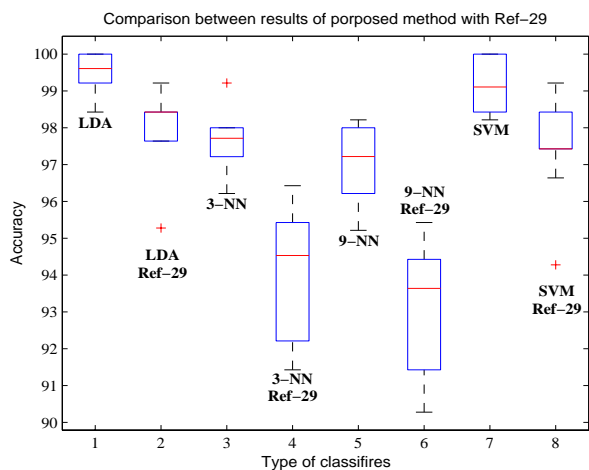
۴- نتایج و بحث

جهت نشان‌دادن کارایی روش پیشنهادی، داده‌های طیف‌سنجی جرمی بیماری سرطان تخمدان که در جدول ۱ توصیف شده‌اند، مورد تحلیل قرار گرفت. تابع موجک مادر انتخاب شده در این بررسی، داییشیز^۱ مرتبه ۴ بوده است که مطابق با کارهای گزارش شده قبلی استفاده شده است [۲۶]. با توجه به کار انجام شده در [۲۷] و [۲۸]، تحلیل تبدیل موجک با روش چندوضوحی و تا سطح تجزیه ۱۰ انجام شد.

در ابتدا داده‌های طیف‌سنجی جرمی، به دو گروه کنترل و سرطانی تقسیم شدند که گروه کنترل شامل نمونه‌های سالم و خوش‌خیم می‌باشد. جهت بررسی عملکرد طبقه‌بندها و ارزیابی دقت تشخیص نشانگرهای حیاتی منتخب، در تقسیم داده‌های هر گروه به زیرمجموعه‌های یادگیری و آزمایش از روش ارزیابی متقابل K نگهدار^۲ استفاده شد. با انتخاب $K=10$ داده‌های هر مجموعه به طور تصادفی با تکرار ۱۰ به دو گروه یادگیری و آزمایش با نسبت ۵۰ درصد تقسیم شدند. دلیل دیگر انتخاب تصادفی این مجموعه‌ها، کالیبره کردن طبقه‌بندها نسبت به تغییر مکان ویژگی‌ها در داده طیف جرمی ناشی از خطای جابه‌جایی طیف‌سنج جرمی می‌باشد.

در این تحقیق از معیارهای استاندارد جهت بررسی مؤثر بودن تکنیک

1. Daubichies Wavelet
2. K - Hold Cross Validation



شکل ۴: مقایسه نتایج روش پیشنهادی با روش [۲۹] با نمودار جعبه‌ای واریانس تغییرات دقت تشخیص در مجموعه داده III.

شکل ۳: توزیع اندیس‌های ضرایب منتخب و بولت و طیف‌های جرمی بازسازی شده از روی ضرایب زنده مربوط به یک نمونه از فرد سالم و بیمار.

جدول ۲: نتایج حاصل از طبقه‌بندی با استفاده از نشانگرهای منتخب برای مجموعه داده‌های طیف جرمی با استفاده از طبقه‌بند پایه LDA.

| مجموعه داده | T_P | F_P | F_N | T_N | S_N (%) | S_P (%) | A_C (%) |
|-------------|-------|-------|-------|-------|-----------|-----------|-----------|
| مجموعه I | ۴۸ | ۳ | ۲ | ۵۵ | ۹۶ | ۹۴٫۸ | ۹۵٫۴ |
| مجموعه II | ۵۰ | ۰ | ۰ | ۵۸ | ۱۰۰ | ۱۰۰ | ۱۰۰ |
| مجموعه III | ۸۱ | ۰ | ۰ | ۴۶ | ۱۰۰ | ۱۰۰ | ۱۰۰ |

جدول ۳: نتایج حاصل از طبقه‌بندی با استفاده از نشانگرهای منتخب برای مجموعه داده‌های طیف جرمی با استفاده از طبقه‌بندهای گوناگون.

| مجموعه داده | نوع طبقه‌بند | T_P | F_P | F_N | T_N | S_N (%) | S_P (%) | A_C (%) |
|-------------|--------------|-------|-------|-------|-------|-----------|-----------|-----------|
| مجموعه I | LDA | ۴۸ | ۳ | ۲ | ۵۵ | ۹۶ | ۹۴٫۸ | ۹۵٫۴ |
| | ۳-NN | ۴۷ | ۴ | ۳ | ۵۴ | ۹۴ | ۹۳٫۱ | ۹۳٫۵ |
| | ۹-NN | ۴۷ | ۶ | ۳ | ۵۲ | ۹۴ | ۸۹٫۷ | ۹۱٫۷ |
| | SVM | ۴۹ | ۲ | ۱ | ۵۶ | ۹۸ | ۹۶٫۵ | ۹۷٫۲ |
| مجموعه II | LDA | ۵۰ | ۰ | ۰ | ۵۸ | ۱۰۰ | ۱۰۰ | ۱۰۰ |
| | ۳-NN | ۵۰ | ۱ | ۰ | ۵۷ | ۱۰۰ | ۹۸٫۳ | ۹۹٫۳ |
| | ۹-NN | ۵۰ | ۱ | ۰ | ۵۷ | ۱۰۰ | ۹۸٫۳ | ۹۹٫۳ |
| | SVM | ۵۰ | ۰ | ۰ | ۵۸ | ۱۰۰ | ۱۰۰ | ۱۰۰ |
| مجموعه III | LDA | ۸۱ | ۰ | ۰ | ۴۶ | ۱۰۰ | ۱۰۰ | ۱۰۰ |
| | ۳-NN | ۸۰ | ۲ | ۱ | ۴۴ | ۹۸٫۸ | ۹۵٫۷ | ۹۷٫۴ |
| | ۹-NN | ۸۰ | ۳ | ۱ | ۴۳ | ۹۸٫۸ | ۹۳٫۵ | ۹۷ |
| | SVM | ۸۰ | ۱ | ۱ | ۴۵ | ۹۸٫۸ | ۹۷٫۸ | ۹۹ |

علاوه بر کاهش همبستگی، خطر حذف نقاط ویژگی مهم غیر همبسته را نیز کاهش می‌دهد که در روش [۲۹] مد نظر قرار نگرفته است. با استفاده از روش پیشنهادی، زیرمجموعه‌ای از ویژگی‌ها انتخاب شد که با کاهش بعد فضای اولیه بردارهای الگوی ورودی، میزان همبستگی ذاتی بین نقاط مختلف این بردارها را کاهش داد. نشانگرهای منتخب نشان دادند که علاوه بر داشتن قدرت تشخیص بالا در بین مجموعه داده‌های مختلف، دارای عدم وابستگی به نوع طبقه‌بند می‌باشند. در تمامی مراحل انتخاب ضرایب، بازسازی و انتخاب نشانگر از روش ارزیابی متقابل K نگهدار برای محاسبه خطای طبقه‌بندی استفاده شد. جهت محاسبه پارامترهای تشخیص کلی، از میانگین نتایج طبقه‌بندهای مختلف در بین سه مجموعه داده استفاده شد. با انتخاب ۱۰ نشانگر حیاتی برای هر مجموعه داده، عملکرد روش پیشنهادی به‌طور متوسط در بین طبقه‌بندهای استفاده‌شده جهت ارزیابی منجر به حصول دقت تشخیص ۹۸، قطعیت تشخیص ۹۷ و حساسیت تشخیص ۹۸ درصد گردید.

طبقه‌بندهای گوناگون نشان می‌دهد. در این مرحله، نشانگرهای منتخب در مرحله قبل به هر یک از طبقه‌بندها داده شد و نتایج با استفاده از معیارهای ارزیابی مورد بررسی قرار گرفت. نتایج به‌دست آمده حاکی از آن است که نشانگرهای حیاتی منتخب از روی طیف جرم بازسازی شده از قدرت تفکیک بالایی در بین مجموعه داده‌های جدول ۱ برخوردار بوده و وابستگی به نوع طبقه‌بند مشاهده نمی‌شود.

به‌منظور مقایسه عملکرد روش شبه کوواریانس با روش‌های مشابه جهت استخراج نشانگر حیاتی از داده طیف جرمی، از [۲۹] استفاده شد. شکل ۴ یک نمودار جعبه‌ای^۱ از میزان واریانس تغییرات در دقت تشخیص به‌دست آمده از طبقه‌بندهای گوناگون با استفاده از داده‌های مجموعه III را نشان می‌دهد. با توجه به این شکل، نتایج حاصله از روش پیشنهادی دارای تغییرات کمتر نسبت به روش [۲۹] می‌باشد که جهت کاهش همبستگی از یک پنجره فاصله استفاده نموده‌اند. روش شبه کوواریانس

proteomic tests for ovarian cancer," *J. of National Cancer Institute*, vol. 97, no. 4, pp. 307-309, Feb. 2005.

- [13] M. Hilario and A. Kalousis, "Approaches to dimensionality reduction in proteomic biomarker studies," *Briefings in Bioinformatics*, vol. 9, no. 2, pp. 102-118, Feb. 2008.
- [14] K. A. Baggerly, J. S. Morris, and K. R. Coombes, "Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments," *Bioinformatics*, vol. 20, no. 5, pp. 777-785, 22 Mar. 2004.
- [15] G. M. Boratyn, M. L. Merchant, and J. B. Klein, "Utilization of human expert techniques for detection of low-abundant peaks in high-resolution mass spectra," *28th IEEE EMBS Annual Int. Conf.*, pp. 5798-5801, New York City, US, 30 Aug-3 Sep. 2006.
- [16] A. G. Hanbury and J. Serra, "Morphological operators on the unit circle," *IEEE Trans. Image Processing*, vol. 10, no. 12, pp. 1842-1850, Dec. 2001.
- [17] S. Mallat, "A wavelet tour of signal processing," Academic Press, 1998.
- [18] D. L. Donoho and I. M. Johnstone, "Threshold selection for wavelet shrinkage of noisy data," in *Proc. 16th Annual Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 24a-25a, Nov. 1994.
- [19] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613-627, May 1995.
- [20] J. Ojanen, T. Miettinen, J. Heikkonen, and J. Rissanen, "Robust denoising of electrophoresis and mass spectrometry signals with minimum description length principle," *Federation of European Biochemical Societies Lett.*, vol. 570, no. 1-3, pp. 107-113, 2004.
- [21] G. Frosini, B. Lazzarini, and F. Marcelloni, "A modified fuzzy C-means algorithm for feature selection," in *Proc. of 19th Int. Conf. of the North American Fuzzy Information Processing Society, NAFIPS'2000.*, Atlanta, US, pp. 148-152, Jul. 2000.
- [22] E. D. Hoffman and V. Stroobant, *Mass Spectrometry: Principles and Applications*, John Wiley and Sons Ltd., 2001.
- [23] W. Windig and J. Guilment, "Interactive self-modeling mixture analysis," *Analytical Chemistry*, vol. 63, no. 14, pp. 1425-1432, 15 Jul. 1991.
- [24] L. Cao, P. B. Harrington, and J. Liu, "SIMPLISMA and ALS applied to tow-way nonlinear wavelet compressed ion mobility spectra of chemical warfare agent simulates," *Analytical Chemistry*, vol. 77, no. 8, pp. 2575-2586, Apr. 2005.
- [25] S. A. Astakhov, H. Stogbauer, A. Kraskov, and P. Grassberger, "Monte carlo algorithm for least dependent non-negative mixture decomposition," *Analytical Chemistry*, vol. 78, no. 5, pp. 1620-1627, 2006.
- [26] M. Vannucci, N. Sha, and P. J. Brown, "NIR and mass spectra classification: bayesian methods for wavelet-based feature selection," *Chemometrics and Intelligent Laboratory Systems*, vol. 77, no. 1-2, pp. 139-148, May 2005.
- [27] H. Montazery Kordy, M. H. Miranbaygi, and M. H. Moradi, "Ovarian cancer diagnosis using discrete wavelet transform based feature extraction from serum proteomic patterns," in *Cairo Int. Biomedical Engineering Conf.*, vol. 1, pp. 139-142, Cairo, Egypt, Dec. 2006.
- [28] H. Montazery Kordy, M. H. Miranbaygi, and M. H. Moradi, "Diagnosis of prostate cancer by wavelet based feature extraction method using blood proteomic patterns," in *Proc. 13th Iranian Conf. in Biomedical Engineering*, Tehran, Iran, Jan. 2007.
- [29] L. Li, H. Tang, Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R. A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," *Artificial Intelligence in Medicine*, vol. 32, no. 2, pp. 71-83, Mar. 2004.
- [30] B. Wu, T. Abbott, D. Fishman, W. McMurray, G. Mor, K. Stone, D. Ward, K. Williams, and H. Zhao, "Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data," *Bioinformatics*, vol. 9, no. 13, pp. 1636-1643, Jul. 2003.

حسین منتظری کردی تحصیلات خود را در مقاطع کارشناسی مهندسی برق- الکترونیک در سال ۱۳۷۸ از دانشکده فنی دانشگاه مازندران، کارشناسی ارشد مهندسی برق- مهندسی پزشکی در سال ۱۳۸۲ از دانشگاه صنعتی شریف، و دکترای مهندسی برق- مهندسی پزشکی در سال ۱۳۸۸ از دانشگاه تربیت مدرس به پایان رسانده است. نامبرده از سال ۱۳۸۷ تا کنون با گروه مهندسی برق و کامپیوتر دانشگاه آزاد اسلامی واحد آیت الله املی همکاری داشته است. زمینه های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنالهای گسسته، فیلترهای وقتی، تحلیل زمان- فرکانس، بازشناسی الگو، پردازش تصویر، و داده شناسی حیاتی.

۵- نتیجه گیری

در این تحقیق امکان تشخیص سرطان تخمدان با تحلیل داده طیف جرمی خونابه مورد بررسی قرار گرفت. تبدیل موجک گسسته به عنوان یک ابزار پیش پردازش جهت کاهش بعد و حذف نویز از منحنی طیف جرم به کار گرفته شد. همچنین، یک روش شبه کوواریانس انتخاب زیرمجموعه ویژگی برای تعیین ضرایب مهم در فضای تبدیل موجک جهت بازسازی سیگنال به منظور کشف نشانگر حیاتی ارائه گردید.

با استفاده از سیگنال طیف جرمی بازسازی شده، کارایی روش پیشنهادی بر روی مجموعه داده های تحت مطالعه مورد بررسی قرار گرفت. جهت بررسی عملکرد نشانگرهای حیاتی منتخب در طبقه بندی افراد سالم از بیمار و عدم وابستگی این ویژگی ها به نوع طبقه بندی، از روش های طبقه بندی متداولی بهره گرفته شد. برای مجموعه داده های در دسترس، عملکرد روش پیشنهادی منجر به حصول دقت تشخیص ۹۸٪ قطعیت تشخیص ۹۷٪ و حساسیت تشخیص ۹۸٪ درصد گردید.

بررسی عملکرد روش پیشنهادی انتخاب زیرمجموعه ویژگی بر اساس پایه های دیگر تبدیل موجک گسسته و امکان ترکیب ضرایب در فضای تبدیل موجک در آینده مد نظر می باشد. همچنین ارائه روش های دیگر انتخاب زیرمجموعه ویژگی نیز از موضوعات مورد علاقه در فعالیت های آتی می باشد.

مراجع

- [1] E. F. Petricoin and L. A. Liotta, "SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer," *Analytical Biotechnology, Science Direct*, vol. 15, no. 1, pp. 24-30, Feb. 2004.
- [2] H. Kuruma, S. Egawa, and *et al.*, "Proteome analysis of prostate cancer," *Prostate Cancer and Prostatic Disease*, vol. 8, no. 1, pp. 14-21, 2005.
- [3] T. P. Conrads, M. Zhou, E. F. Petricoin III, L. Liotta, and T. D. Veenstra, "Cancer diagnosis using proteomics patterns," *Expert Rev. Mol. Diagn.*, vol. 3, no. 4, pp. 411-420, 2003.
- [4] E. F. Petricoin III, D. K. Ornstein, C. P. Paweletz, A. M. Ardekani, P. S. Hackett, B. A. Hitt, A. Velasco, C. Trucco, L. Wiegand, K. Wood, C. B. Simone, P. J. Levine, W. M. Linehan, M. R. Emmert-Buck, S. M. Steinberg, E. C. Kohn, and L. A. Liotta, "Serum proteomic patterns for detection of prostate cancer," *J. of National Cancer Institute*, vol. 94, no. 20, pp. 1576-1578, Oct. 2002.
- [5] E. J. Finehout and K. H. Lee, "An introduction to mass spectrometry applications in biological research," *Biochemistry and Molecular Biology Education*, vol. 32, no. 2, pp. 93-100, 2004.
- [6] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. J. Thun, "Cancer statistics, 2008" *CA Cancer J. Clin.*, vol. 58, pp. 71-96, 2008.
- [7] J. S. Morris, K. R. Coombes, J. Koomen, K. A. Baggerly, and R. Kobayashi, "Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum," *Bioinformatics*, vol. 21, no. 9, pp. 1764-1775, 2005.
- [8] E. F. Petricoin III, A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta, "Use of proteomic patterns in serum to identify ovarian cancer," *The Lancet*, vol. 359, pp. 572-577, Feb. 2002.
- [9] B. L. Adam, A. Vlahou, O. J. Semmes, and G. L. Wright, "Proteomic approaches to biomarker discovery in prostate and bladder cancers," *Proteomics*, vol. 1, no. 10, pp. 1264-1270, Oct. 2001.
- [10] Y. Qu, B. L. Adam, Y. Yasui, M. D. Ward, S. Nasim, P. F. Schellhammer, Z. Feng, O. J. Semmes, and G. L. Wright, "Boosted decision tree analysis of SELDI mass spectral serum profiles discriminates prostate cancer from noncancer patients," *Clinical Chemistry*, vol. 48, no. 10, pp. 1835-1843, Oct. 2002.
- [11] J. M. Sorace and M. Zhan, "A data review and re-assessment of ovarian cancer serum proteomic profiling," *BMC Bioinformatics*, vol. 4, no. 24, pp. 1-13, Jun. 2003.
- [12] K. A. Baggerly, J. S. Morris, S. R. Edmonson, and K. R. Coombes, "Signal in noise: evaluating reported reproducibility of serum

محمد حسن مرادی کارشناسی خود را در مهندسی برق-الکترونیک در سال ۱۳۶۷ از دانشکده فنی دانشگاه تهران و کارشناسی ارشد مهندسی برق-الکترونیک در سال ۱۳۶۹ از همان دانشگاه اخذ نمود. وی در سال ۱۳۷۴ دکترای خود را در رشته مهندسی برق-مهندسی پزشکی از دانشگاه تربیت مدرس دریافت نمود و از آن زمان تا کنون عضو هیات علمی دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر می‌باشد. تا کنون پنجاه مقاله در مجلات علمی پژوهشی داخل و خارج از کشور و بیش از صد و پنجاه مقاله کنفرانس‌های معتبر و بین‌المللی داخل و خارج از کشور منتشر نموده است. کتاب "آنالیز سیگنال مشتمل بر فیلتر بانک‌ها، تبدیل ویولت و بازنمایی زمان-فرکانس" توسط وی در سال ۱۳۸۰ ترجمه شده‌است. نامبرده ده‌ها پروژه کارشناسی ارشد و دکتری را سرپرستی نموده و در چند طرح تحقیقاتی مشارکت داشته‌است. پروژه کارشناسی ارشد "طراحی و ساخت دستگاه انتقال تلفنی سیگنال‌های قلبی" به سرپرستی ایشان برنده جایزه سوم جشنواره جوان خوارزمی در سال ۱۳۸۰ و ثبت شده به عنوان یک اختراع در داخل کشور می‌باشد. زمینه‌های تحقیقاتی مورد علاقه وی اندازه‌گیری، ثبت، پردازش و تحلیل سیگنال‌های حیاتی و تصاویر پزشکی، کاربرد سیستم‌های فازی و فازی عصبی در پردازش و بکارگیری تبدیل ویولت، مالتی ویولت و تبدیلات زمان-فرکانس، طیف‌های مرتبه بالا، شبکه‌های ویولت و ویونت است. وی هم اکنون دانشیار مهندسی پزشکی دانشگاه صنعتی امیرکبیر می‌باشد.

محمدحسین میران‌بیگی کارشناسی خود را در مهندسی برق-کنترل در سال ۱۹۸۹ میلادی از دانشگاه بیرمنگام در انگلستان دریافت نمود. وی بترتیب در سالهای ۱۹۹۱ و ۱۹۹۴ میلادی کارشناسی ارشد مهندسی برق-طراحی سیستم‌های دیجیتال و دکترای مهندسی برق-ابزار دقیق مهندسی پزشکی خود را از دانشگاه UMIST انگلستان اخذ نمودند. دکتر میران‌بیگی از سال ۱۳۷۴ در دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس مشغول به فعالیت گردید و هم‌اکنون نیز دانشیار گروه مهندسی پزشکی این دانشکده می‌باشد. تاکنون از ایشان مقالات متعددی در مجلات علمی و پژوهشی داخل و خارج از کشور و همچنین کنفرانس‌های معتبر بین‌المللی داخلی و خارجی به چاپ رسیده‌است. زمینه‌های علمی مورد علاقه نامبرده متنوع بوده و شامل موضوعاتی عبارتند از: مدل‌سازی سیستم‌های بیولوژیکی، مطالعه برهم کنش لیزر با بافت‌های بیولوژیکی، و ابزار دقیق مهندسی پزشکی می‌باشد.

Archive of SID