

## An Experimental Investigation into the use of Ramanspectroscopy for the Diagnosis of Cancer

Z. Dehghani Bidgoli<sup>1</sup>, M. H. Miranbaygi<sup>2\*</sup>, R. Malekfar<sup>3</sup>, E. Kabir<sup>4</sup>, T. Khamechian<sup>5</sup>

<sup>1</sup>Ph.D Student, Biomedical Engineering Group, Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran, Dehghani\_zohreh@yahoo.com

<sup>2</sup>Associate Professor, Biomedical Engineering Group, Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran

<sup>3</sup> Associate Professor, Physics Group, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Iran, Malekfar@modares.ac.ir

<sup>4</sup> Professor, Electronics Group, Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran, Kabir@modares.ac.ir

<sup>5</sup> Associate Professor, Medical Pathology Group, Kashan University of Medical Sciences, Kashan, Iran, Khamechian-@yahoo.com

---

### Abstract

In this research, we investigated cancerous tissues from several organs of the human body using Raman spectroscopy. Different specimens with different pathologic labels (normal & cancerous) were borrowed from a pathology laboratory, and were investigated using two different Raman spectroscopy systems. Since one of the goals of this investigation was detection of cancer, independent of type of the system, we introduced some algorithms for removing systemic differences from the spectra. Then we removed noise and fluorescence signals using a new wavelet created with LWT.

The best classification result was 83% in differentiating between normal and cancerous specimens using the SVM classifier.

**Key words:** Raman Spectroscopy, cancer detection, classifying, feature extraction, LWT

---

\*Corresponding author

Address: Mohammad hoseinMiranbaygi, Biomedical Engineering Group, Electrical and Computer Engineering Department, Tarbiat Modares University, Tehran, Iran  
Tel: +982182884331  
Fax: +982182884325  
E-mail: Miranbmh@modares.ac.ir

## بررسی تجربی استفاده از طیف‌سنجی رامان در تشخیص سرطان

زهرة دهقانی بیدگلی<sup>۱</sup>، محمدحسین میران‌بیگی<sup>۲\*</sup>، رسول ملک‌فر<sup>۳</sup>، احسان اله کبیر<sup>۴</sup>، طاهره خامه‌چیان<sup>۵</sup>

<sup>۱</sup> دانشجوی دکترای مهندسی پزشکی، گروه مهندسی پزشکی، دانشکده برق و کامپیوتر، دانشگاه تربیت مدرس، تهران

Dehghani\_zohreh @ yahoo.com

<sup>۲</sup> دانشیار، گروه مهندسی پزشکی، دانشکده برق و کامپیوتر، دانشگاه تربیت مدرس، تهران

<sup>۳</sup> دانشیار، گروه فیزیک، دانشکده علوم پایه، دانشگاه تربیت مدرس، تهران Malekfar@modares.ac.ir

<sup>۴</sup> استاد، گروه الکترونیک، دانشکده برق و کامپیوتر، دانشگاه تربیت مدرس، دانشکده برق و کامپیوتر، Kabir@modares.ac.ir

<sup>۵</sup> دانشیار، گروه پاتولوژی پزشکی، دانشگاه علوم پزشکی کاشان، کاشان Khamechian\_t@yahoo.com

### چکیده

در این تحقیق به بررسی بافتهای سرطانی چندین اندام بدن با استفاده از طیف‌سنجی رامان لیزری پرداخته‌ایم. در این راستا نمونه‌های مختلف بافتی با تشخیص‌های پاتولوژیک متفاوت (سالم و سرطانی) از یک آزمایشگاه پاتولوژی به امانت گرفته و با دو سیستم طیف‌سنجی رامان بررسی شدند. از آنجایی که یکی از اهداف این مطالعه مستقل ساختن تشخیص از نوع سیستم طیف‌سنجی است؛ الگوریتم‌هایی برای حذف تفاوت‌های سیستمیک در طیف‌های بدست آمده از دو سیستم، ارائه شده است. پس از آن نوفه و فلورسانس موجود در طیف‌ها با موجک جدیدی که LWT آنرا ایجاد کرد، حذف و با چند طبقه‌بند مختلف، طبقه‌بندی شد. بهترین نتیجه حاکی از تمایز ۸۳٪ بین نمونه سالم و سرطانی با استفاده از طبقه‌بند SVM است.

کلیدواژگان: طیف‌سنجی رامان، تشخیص سرطان، طبقه‌بندی، استخراج ویژگی، LWT

\*عهده‌دار مکاتبات

نشانی: تهران، بزرگراه جلال آل احمد، دانشگاه تربیت مدرس، دانشکده برق و کامپیوتر، گروه مهندسی پزشکی، کدپستی: ۱۹۴-۱۴۱۱۵

تلفن: ۰۲۱۸۲۸۸۳۳۷۰، دورنگار: ۰۲۱۸۲۸۸۴۳۲۵، پیام‌نگار: miranbmh@modares.ac.ir

## ۱- مقدمه

امروزه یکی از بیماریهایی که جوامع بشری را با مشکلات فراوانی در بخش بهداشت و درمان، مددکاری و بهزیستی مواجه ساخته است، انواع مختلف سرطانهاست. این بیماری مسئول ۱۳٪ کل مرگ و میرها در جهان است. طبق گزارش انجمن سرطان آمریکا ۷/۶ میلیون نفر در سال ۲۰۰۷ بر اثر ابتلا به سرطان جان باختند. با وجود تفاوتی که بین انواع مختلف سرطانها وجود دارد، یک نکته بین همه مشترک است و آن این است که تشخیص زود هنگام این بیماری کمک شایانی به درمان بیمار خواهد کرد. با توجه به این امر، مطالعات بسیاری در جهت تشخیص و درمان هر چه سریعتر این بیماری انجام شده است. اما روشی که همواره به منظور تشخیص نهایی بکار می رود، آزمایش پاتولوژی است. این روش علیرغم توانایی هایی که در زمینه تشخیص انواع بیماریها دارد دارای مشکلاتی همچون، تهاجمی بودن (نیاز به نمونه برداری)، زمان پاسخ نسبتاً طولانی و نتایج وابسته به فرد- که به تجربه و تخصص پاتولوژیست بستگی دارد- است. لذا در سالهای گذشته، تکنیکهای مختلفی برای حل این مشکلات پیشنهاد شده است. در این میان، روشهای نوری از آنجایی که می توانند بصورت غیر تهاجمی و غیر مخرب بکار گرفته شوند، بسیار مورد توجه قرار گرفته اند. از جمله این روشها که در زمینه تشخیص تغییرات مولکولی و در نتیجه تشخیص سریع و صحیح ضایعات بافتی قابلیت بسیاری از خود نشان داده است، می توان به طیفسنجی رامان اشاره کرد.

طیفسنجی رامان یک روش اسپکتروسکوپی بر اساس پراکندگی غیر الاستیک نور تک رنگی است که معمولاً از یک منبع نور لیزر تأمین می شود [۱، ۲]. هنگامی که نور تک رنگ لیزر، به یک نمونه برخورد می کند، بخشی از نور پراکنده می شود. در نور پراکنده شده علاوه بر فرکانس نور فرودی (پراکندگی رایلی) تابشهایی نیز با فرکانس متفاوت مشاهده می شود که پراکندگی رامان نام دارند. میزان اختلاف فرکانس نور تابشی و نور پراکنده شده، به فرکانس نوسان پیوندهای اتمی موجود در نمونه تحت تابش بستگی دارد. لذا می توان با ثبت میزان پراکندگی در هر فرکانس به ساختار مولکولی ماده تحت مطالعه پی برد؛ بدین ترتیب طیف رامان می تواند به عنوان اثر انگشت نمونه بکار

گرفته شود. از این روش می توان در شناسایی مواد مختلف از جمله نمونه های بیولوژیک استفاده کرد.

همانگونه که می دانیم، مجموعه داده های مرجع همچون قلب تشخیصهای پزشکی هستند. این طیف های مرجع در تشخیص سرطان در مدت زمان نسبتاً طولانی (ماهها یا سالها) بدست می آیند. بعلاوه کاربردهای پزشکی باید قابلیت انتقال از یک سیستم به سیستم دیگر را داشته باشند، بدین معنی که طیف های مرجع و طیف های بدست آمده برای یک کاربرد تشخیصی خاص، هر دو باید عاری از هرگونه مشخصات وابسته به سیستم باشند تا بتوان مجموعه داده های بدست آمده از یک سیستم را در جای دیگری نیز به کار برد [۳]. بدیهی است در غیر اینصورت نیاز به صرف مدت زمانی طولانی برای تهیه مجموعه داده های خاص آن سیستم است؛ که قطعاً این امر موجب کاهش کارایی این روش در تشخیصهای بالینی می شود. لذا هدف، توسعه الگوریتمی تعمیم پذیر است که صرف نظر از نوع سیستم رامان مورد استفاده قابل اعمال به طیف های بدست آمده؛ باشد. از این رو در این تحقیق به بررسی نمونه های سرطانی با استفاده از دو سیستم طیفسنجی کاملاً متفاوت پرداخته ایم و راهکارهایی نیز برای حذف مشخصات سیستمیک در آنها مطرح کرده ایم؛ که این مطالعه را از تحقیقات پیشین که محققان مختلف در این زمینه انجام داده اند، متمایز ساخته است. از سوی دیگر یکی از مشکلاتی که در بسیاری از موارد تشخیص سرطان وجود دارد، مشخص نبودن نوع بافتی است که ضایعه مورد نظر به آن تعلق دارد [۴-۶]. این امر ممکن است به دو دلیل روی دهد، اولاً حضور توده مشکوک در ناحیه ای بین اندامهای مختلف بگونه ای که تعیین بافت منشأ آن برای پزشک یا جراح امکان پذیر نباشد، مانند برخی از تومورهایی که در حفره شکمی ظاهر می شوند. ثانیاً در سرطانهای بسیار پیشرونده، برخی از سلولهای سرطانی از توده اولیه جدا و از طریق عروق خونی یا لنفاوی به دیگر نقاط بدن منتقل می شوند. سپس در مکان جدید لانه گزینی کرده و تومور سرطانی تشکیل می دهند (متاستاز). در نتیجه وقوع این پدیده، یک توده در بافتی مشاهده می شود که متعلق به آن بافت نیست و در واقع بافت سرچشمه آن ناشناخته است [۷، ۸]. لذا در چنین مواردی نمی توان از روشهای موجود تشخیص سرطان با طیفسنجی رامان- که برای

نمونه‌ها را پس از انتقال به آزمایشگاه رامان، از محلول فرمالین خارج کردیم و سپس با دو نمونه سیستم رامان موجود در آزمایشگاه (Almega و R3000) به اخذ طیف از نمونه‌ها پرداختیم. شایان ذکر است نمونه‌های سالم از بافت مجاور نواحی بدخیم بدست آمده است.

از ۱۲ نمونه سالم و سرطانی یاد شده، با استفاده از دو سیستم Almega و R3000 طیف‌سنجی انجام شد. از آنجایی که برخی نمونه‌ها (مانند پوست) دارای تفاوت‌های ظاهری در برخی قسمت‌ها هستند از بخش‌های مختلف آنها طیف‌سنجی انجام شد و در مجموع تعداد ۱۶ طیف از نمونه‌های ذکر شده با سیستم Almega اخذ شد. همچنین با سیستم R3000 نیز از هر نمونه دو طیف در دو توان ۱۲ و ۲۴ میلی وات بدست آمد. در این مورد نیز در برخی موارد از بخش‌های مختلف نمونه طیف‌سنجی انجام شد و در مجموع ۳۵ طیف با استفاده از این سیستم در دو توان مختلف برداشته شد. طیف‌های بدست آمده با Almega و R3000 بترتیب در شکل (۱) بخش‌های (الف) و (ب) نشان داده شده است.

لیزر سیستم Thermo Nicolet Almega - که ما در اینجا آنرا به اختصار Almega نامیده‌ایم - از نوع Nd:YLF بوده، که از هارمونیک دوم آن با طول موج ۵۳۲ نانومتر استفاده می‌شود. توان لیزر بکار رفته ۵۰ میلی وات و شعاع لیزر در فاصله کانونی  $100 \mu\text{m}$  است. طیف‌ها با ۳۲ بار اسکن در محدوده طیفی  $400 - 4000 \text{ cm}^{-1}$  با درجه تفکیک  $4 \text{ cm}^{-1}$  اخذ شده‌اند.

سیستم R3000 یک سیستم طیف‌سنجی رامان قابل حمل است. این سیستم از یک لیزر دیود با طول موج  $532 \text{ nm}$  استفاده می‌کند. توان لیزر به کار رفته بین ۱۲ تا ۲۴ میلی وات قابل تنظیم است. در این سیستم هر طیف پس از سه بار اندازه‌گیری در مدت زمان ۳۰ ثانیه در محدوده  $400 - 4300 \text{ cm}^{-1}$  با درجه تفکیک طیفی  $8 \text{ cm}^{-1}$  اخذ شده است.

یک بافت خاص توسعه یافته است - استفاده کرد. بنابراین نیاز به توسعه روشی است که قادر به تشخیص نوع سرطان با منشأ نامعلوم باشد. با توجه به این امر در تحقیق حاضر نمونه‌های مورد مطالعه از بافتهای مختلف انتخاب شده است.

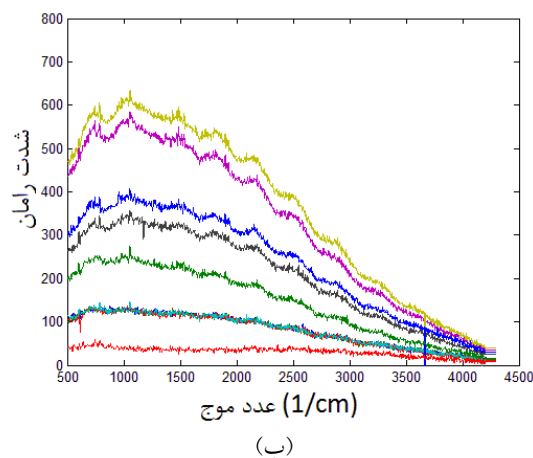
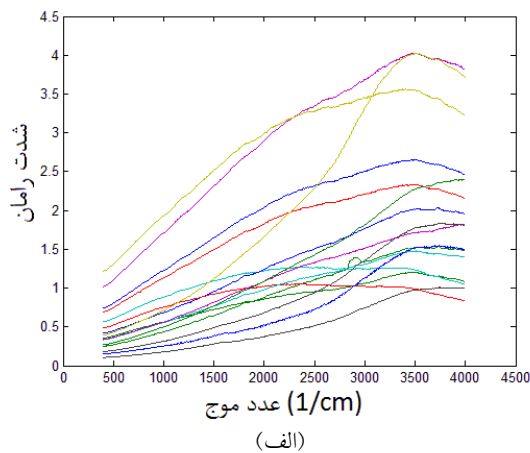
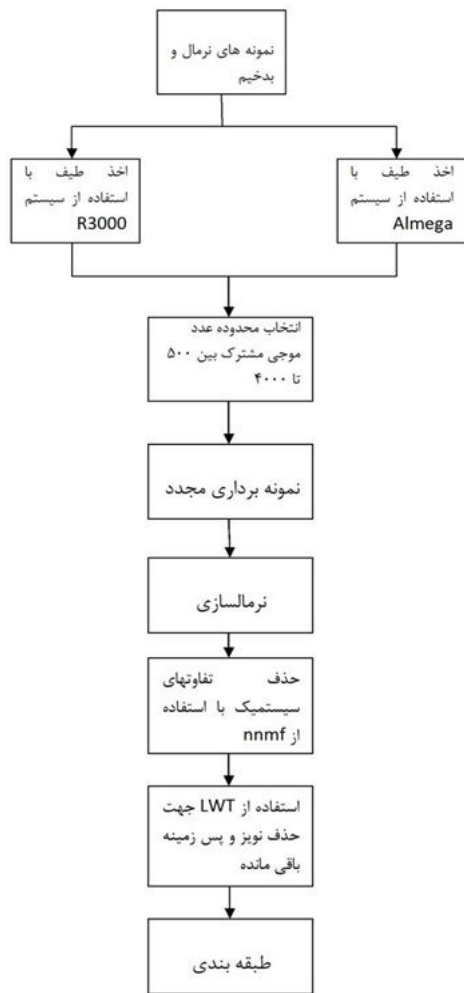
بافتهای بدخیم (سرطان) با افزایش مواد هسته‌ای، افزایش نسبت هسته به سیتوپلاسم، افزایش فعالیت تقسیمی، توزیع غیر طبیعی کروماتین و کاهش تمایز مشخص می‌شوند. افزایش تکثیر و کاهش بلوغ این سلولهای تمایز نیافته منجر به افزایش فعالیت متابولیکی می‌شود. افزایش فعالیت متابولیک نیز به رگ‌سازی سریع و تشکیل رگهای نشتی می‌انجامد.

از نظر بافت‌شناسی، نئوپلاسمها (تومورها) با تجمع و بی‌نظمی سلولی مشخص می‌شوند؛ اما این ویژگیهای عمومی سلولهای نئوپلاستیک موجب تغییرات خاص در مقادیر یا ترکیب اسید نوکلئیک، پروتئین، چربی و کربوهیدرات در بافت نیز می‌شود. برای مثال سلولهای نئوپلاستیک، لاکتات بیشتری در مقایسه با سلولهای سالم تولید می‌کنند. به طور خلاصه می‌توان گفت چندین نشانگر مولکولی در غشا، سیتوپلاسم، هسته و فضای بین سلولی وجود دارند که می‌توانند شاخص نئوپلازی باشند. مولکولهای نشانگر شامل پروتئینها، چربیها و اسیدهای نوکلئیک هستند و تغییر در خود نشانگر یا تعامل آن با مولکولهای دیگر مورد انتظار است. تغییرات سلولی و مولکولی - که در سرطان رخ می‌دهند - منجر به تفاوت در طیف‌های رامان بدست آمده از بافتهای سالم و سرطانی می‌شود.

پیش از این نیز محققان بسیاری به بررسی سرطان در نواحی مختلف بدن همچون پوست [۹-۱۱]، سین [۱۲، ۱۳]، روده بزرگ [۱۴]، کلیه [۱۵، ۱۶] و دهانه رحم [۱۷]، با استفاده از طیف‌سنجی رامان پرداخته‌اند. از بررسی این مطالعات می‌توان در یافت که وجوه اشتراکی در تشخیص سرطان اندامهای مختلف بدن وجود دارد؛ اما مطالعه مستقلی تا کنون در این زمینه انجام نشده است.

## ۲- توصیف داده

در این تحقیق ۱۲ نمونه بافتی سالم و سرطانی (بدخیم) تثبیت شده در فرمالین از بافتهای پستان، پوست، طحال، روده بزرگ و تیروئید، بدست آمده از افراد مختلف همراه نتایج تشخیصهای پاتولوژیک آنها، از آزمایشگاه پاتولوژی به امانت گرفته شد. این



شکل (۱)- طیف های بدست آمده از بافتهای مختلف سالم و سرطانی با استفاده از سیستم (الف) Almega و (ب) R3000 (چند نمونه)

شکل (۲)- روندنمای مراحل پردازش طیفها

علاوه بر تفاوت های فوق، برخی از مشکلات همچون نوفه<sup>۱</sup>، پس زمینه قوی و همچنین تعداد نقاط داده فراوان بین آنها مشترک است. لذا برای غلبه بر این مشکلات مطابق روندنمای شکل (۲) مراحل مختلف پردازش طیفها انجام می شود که در ادامه به توضیح هر یک خواهیم پرداخت.

در نمونه برداری مجدد تعداد نقاط داده برابر ۱۰۰۰ انتخاب شده است و بهنجارش برای هر طیف بین بیشینه و کمینه آن مطابق با رابطه (۱) انجام شده است:

$$SN = \frac{S - \min(S)}{\max(S) - \min(S)} \quad (1)$$

در رابطه فوق SN مقدار شدت طیف بهنجار، S مقدار طیف قبل از بهنجارش و  $\min(S)$  و  $\max(S)$  بترتیب حداقل و حداکثر شدت سیگنال هستند.

### ۳- همانندسازی طیفها

همانگونه که در شکل (۱) مشاهده می شود، طیف های بدست آمده از دو سیستم دارای تفاوت های اساسی هستند که ناشی از تفاوت دو سیستم است. تفاوت های این دو سیستم را که در طیف های بدست آمده از آنها منعکس شده است، می توان به صورت زیر برشمرد:

۱. محدوده عدد موجی متفاوت
۲. درجه تفکیک طیفی متفاوت
۳. شدت های کلی متفاوت
۴. قالب پس زمینه متفاوت

<sup>۱</sup>Noise

داده شده در طیف‌ها و  $W$  ماتریس نشان‌دهنده میزان مشارکت هر منبع در هر طیف است. در مطالعه حاضر ۵۱ طیف دارای ۱۰۰۰ نقطه، طیفی ماتریس داده ورودی را تشکیل می‌دهند و تعداد منابع برابر ۵ در نظر گرفته شده است. راه حل رایج برای یافتن  $W$  و  $H$  کمینه کردن اختلاف بین  $V$  و  $WH$  طبق رابطه (۲) است:

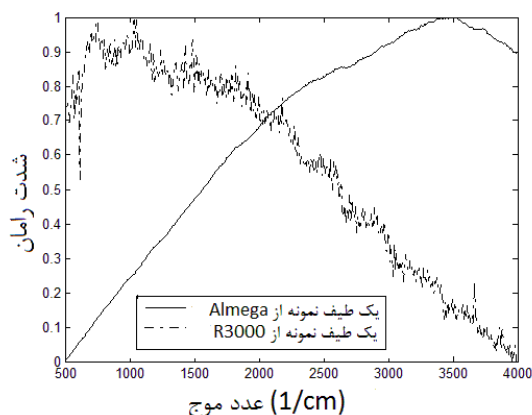
$$\min_{W,H} f(W,H) \equiv \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m (V_{ij} - (WH)_{ij})^2 \quad (2)$$

$$W_{ia} \geq 0, H_{bj} \geq 0 \quad \forall i, a, b, j$$

در مطالعه حاضر از الگوریتم ALS<sup>۲</sup> برای یافتن کمینه تابع  $f$  استفاده شده است. این الگوریتم بطور متناوب یکی از ماتریسهای  $W$  یا  $H$  را ثابت فرض کرده و دیگری را بهبود می‌بخشد. در واقع  $W^{k+1}$  را به گونه‌ای می‌یابد که  $f(W^{k+1}, H^k) \leq f(W^k, H^k)$  و سپس  $H^{k+1}$  را طوری انتخاب می‌کند که  $f(W^{k+1}, H^{k+1}) \leq f(W^{k+1}, H^k)$  باشد [۱۸].

منابع برآورد شده پس از ۱۰۰ تکرار در شکل (۴) نشان داده شده است. سپس میزان مشارکت هر منبع در هر مجموعه طیف بررسی شد. شکل (۵) میزان حضور این ۵ منبع را در ۵۱ طیف مورد مطالعه، نشان می‌دهد. در مجموعه طیف‌های مورد مطالعه - همانطور که در شکل نشان داده شده است - ۱۶ طیف نخست مربوط به سیستم Almaga و مابقی مربوط به سیستم R3000 است. از شکل پیداست که میزان مشارکت منبع دوم در طیف‌های Almaga و میزان منبع سوم در طیف‌های R3000 تقریباً صفر هستند، و از آنجایی که نمونه‌ها یکسان هستند و فقط نوع سیستم متفاوت است، به نظرمی رسد این منابع وابسته به نوع سیستم باشند و نه نمونه؛ لذا این دو منبع را از مجموعه منابع حذف و بار دیگر طیف‌ها را با سه منبع باقیمانده بازسازی کردیم. نتایج پس از باز سازی در شکل (۶) نشان داده شده است. همانگونه که در این شکل مشاهده می‌شود تفاوتها تا حدود زیادی برطرف شده است.

طیف‌های بدست آمده پس از اعمال پردازشهای مراحل ۱ تا ۳ در شکل (۳) نشان داده شده‌اند. همانگونه که در این شکل مشاهده می‌شود همچنان طیف‌های بدست آمده از دو سیستم دارای تفاوتی قابل ملاحظه‌ای هستند، که ناشی از تفاوت پس‌زمینه و نوفه در آنها است.



شکل (۳) - دو نمونه از طیف‌های بدست آمده پس از اعمال نمونه‌برداری مجدد و بهنجارش

از آنجایی که این طیف‌ها از دو سیستم متفاوت اخذ شده‌اند و هر سیستم با توجه به ادوات نوری خاص خود همچون لنزها، شبکه پراش، آشکارساز و... یک تابع تبدیل روی طیف واقعی نمونه اعمال می‌کند که خاص خود آن سیستم است؛ لذا - همانگونه که در شکل (۳) نشان داده شده است - طیف‌های بدست آمده از دو سیستم کاملاً با یکدیگر متفاوت است. با فرض خطی بودن اثر عوامل فوق در شکل‌گیری پس‌زمینه می‌توان با استفاده از الگوریتمهای ICA، منبع مربوط به پس‌زمینه را در این دو دسته طیف تشخیص داد و با حذف آن، مشکل متفاوت بودن پس‌زمینه را در این دو دسته طیف حل کرد. بدلیل مثبت بودن ماهیت طیف‌های رامان و فلورسانس، در این مرحله از الگوریتم NNMF<sup>۳</sup> که به تخمین منابع نامنفی می‌پردازد، استفاده کردیم.

با فرض اینکه ماتریس غیر منفی داده ورودی،  $V \in \mathbb{R}^{n \times m}$ ، و یک عدد صحیح مثبت از پیش تعریف شده  $r$  موجود باشد، هدف از nmmf پیدا کردن دو ماتریس غیر منفی  $W \in \mathbb{R}^{n \times r}$

$$V \approx WH \quad \text{به گونه‌ای است که: } H \in \mathbb{R}^{r \times m}$$

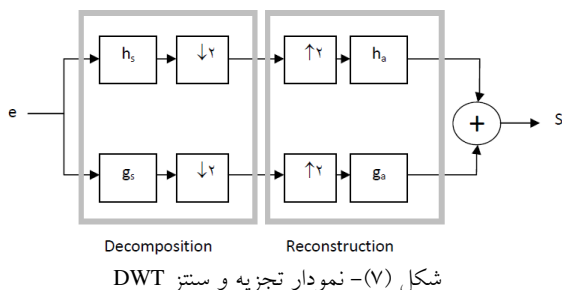
اگر هر سطر از  $V$  نشان‌دهنده یک طیف دارای  $m$  نقطه طیفی باشد، ماتریس  $H$  نشان‌دهنده  $r$  منبع غیر منفی تشخیص

<sup>2</sup> Non Negative Matrix Factorization

<sup>3</sup> Alternating non-negative Least Squares

جداسازی دو طبقه می‌شدند؛ لذا تصمیم گرفتیم با استفاده از LWT<sup>۴</sup> موجکی ایجاد کنیم که بتواند علاوه بر حذف پس‌زمینه، کارایی طبقه‌بندی را نیز افزایش دهد.

در این راستا لازم است ابتدا ساختار یک موجک گسسته، DWT<sup>۵</sup>، را معرفی کرده، سپس به نحوه ایجاد یک موجک جدید دارای این ساختار با استفاده از لیفتینگ<sup>۶</sup>، (LWT)، بپردازیم. هر DWT با چهار صافی  $h_a$  و  $g_a$  برای آنالیز و  $h_s$  و  $g_s$  برای سنتز تعریف می‌شود. از جمله شروطی که این صافی‌ها باید داشته باشند، شرط بازسازی کامل است. این شرط را می‌توان بصورت تساوی  $s$  و  $e$  در نمودار شکل (۷) توصیف کرد [۱۹-۲۱].



این امر به دو شرط ۳ و ۴ منجر می‌شود که شرایط بازسازی کامل شناخته می‌شوند.

$$H_s(z)H_a(z^{-1}) + G_s(z)G_a(z^{-1}) = 2z^{-d} \quad (۳)$$

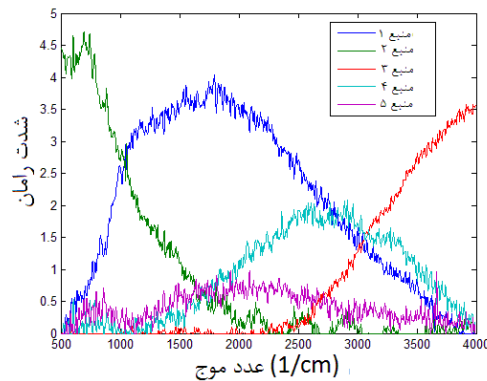
$$H_s(z)H_a(-z^{-1}) + G_s(z)G_a(-z^{-1}) = 0 \quad (۴)$$

در رابطه فوق  $H_a, H_s, G_a, G_s$  نشان‌دهنده تبدیل  $Z$  صافی‌های  $h_a, h_s, g_a, g_s$  است.

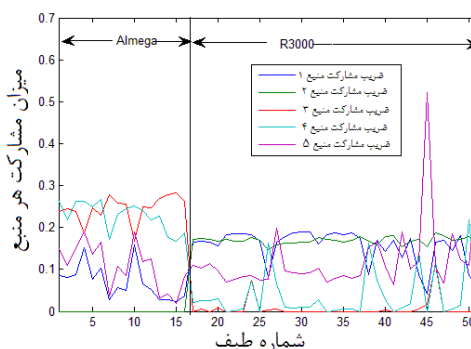
صافی‌هایی که شرایط فوق را داشته باشند با نام چهارتایی‌های دو تعامدی<sup>۷</sup> شناخته می‌شوند. اساس لیفتینگ نیز ایجاد یک چهارتایی دو تعامدی از نمونه موجود با اعمال زنجیره متناهی از گامهای لیفتینگ (ELS) اولیه<sup>۸</sup> یا ثانویه<sup>۹</sup> است. ELS اولیه، با استفاده از روابط ۵ و ۶ از مجموعه  $(H_a, G_a, H_s, G_s)$  یک مدل جدید  $(H_a^N, G_a, H_s, G_s^N)$  می‌سازد [۲۱].

$$H_a^N = H_a(z) - G_a(z)S(z^{-2}) \quad (۵)$$

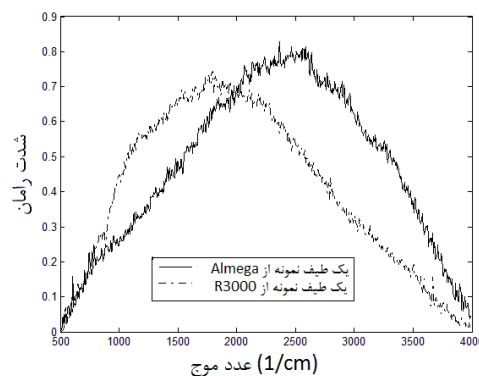
$$G_s^N = G_s(z) + H_s(z)S(z^2) \quad (۶)$$



شکل (۴) - منابع برآورد شده پس از اعمال NNMF



شکل (۵) - میزان مشارکت منابع تخمین زده شده در هر طیف



شکل (۶) - دو نمونه از طیف‌های بازسازی شده پس از حذف منابع دوم و سوم

#### ۴- حذف فلورسانس پس‌زمینه

در ادامه به منظور حذف پس‌زمینه باقی‌مانده از طیف‌های بدست آمده، از تبدیل موجک استفاده کردیم. اما از آنجایی که هیچیک از موجک‌های موجود قادر به حذف پس‌زمینه و نوفه به صورت مناسبی نبودند و عملاً موجب کاهش کارایی طبقه‌بندی در

<sup>۴</sup> Lifting Wavelet Transform  
<sup>۸</sup> primal <sup>۹</sup> dual

<sup>۵</sup> Discrete Wavelet Transform

<sup>۶</sup> Lifting

<sup>۷</sup> biorthogonal quadruplets

$$LS = \begin{matrix} 'd' & [ -2.4255] & [ 0] \\ 'd' & [ 2.895, -0.561] & [-1] \\ 'p' & [-0.976, -0.016] & [ 0] \\ 'd' & [ 5.557e-05, 5.557e-05] & [ 0] \\ 'p' & [ 0.990, -0.002] & [ 0] \\ 'd' & [ 0.00044, 0.00044] & [ 5] \\ 'p' & [ -0.0430] & [ 2] \\ & [ 2.3155] & [ 0.4319] & [ ] \end{matrix}$$

در طرح لیفتینگ فوق هر سطر نشان دهنده یک گام لیفتینگ است. حرف اول نوع گام، p برای اولیه و d برای ثانویه، اعداد داخل کروشه ضرایب چندجمله‌ای لرننت و عدد سوم مرتبه چندجمله‌ای است.

صافی‌های بهینه بدست آمده، در شکل (۸) نشان داده شده‌اند. پس از آن با استفاده از صافی‌های بدست آمده به پردازش طیف‌ها پرداختیم. بخش‌های (الف) و (ب) شکل (۹) مراحل تجزیه بترتیب برای یک نمونه از طیف‌های سیستم Almega و R3000 نشان می‌دهند.

طیف‌های بدست آمده پس از اعمال پردازش LWT با کمک موجک جدید و تجزیه به ۵ سطح و بازسازی با استفاده از سطوح چهارم و پنجم جزئیات، در شکل (۱۰) نشان داده شده است.

در عبارات فوق S یک چند جمله‌ای لرننت<sup>۱۰</sup> است.

ELS ثانویه نیز به شکل مشابهی، با استفاده از روابط ۷ و ۸ از مجموعه  $(H_a, G_a, H_s, G_s)$  یک مدل جدید  $(H_a, G_a, H_s, G_s)$  می‌سازد [۲۱].

$$H_s^N = H_s(z) + G_s(z)T(z^2) \quad (7)$$

$$G_a^N = G_a(z) - H_a(z)T(z^{-2}) \quad (8)$$

در عبارات فوق نیز T یک چند جمله‌ای لرننت است. یک چندجمله‌ای لرننت با رابطه (۹) تعریف می‌شود:

$$C(Z) = c_1 Z^{pmax} + c_2 Z^{pmax-1} + \dots + c_{end} Z^{pmin} \quad (9)$$

آنچه در ایجاد صافی‌های مناسب اهمیت دارد، انتخاب صحیح درجه و ضرایب چند جمله‌ای‌های لرننت و در ELS‌های اولیه و ثانویه و تعداد و ترتیب این گامها است.

برای یافتن صافی‌هایی که به بهترین شکل بتوانند نوفه و خط زمینه را تخمین بزنند و حذف کنند، از الگوریتم بهینه‌سازی جستجوی الگو<sup>۱۱</sup> به منظور تعیین درجه و ضرایب چند جمله‌ای‌های لرننت استفاده کرده‌ایم. در الگوریتم جستجوی الگو که برای یافتن ضرایب LS<sup>۱۲</sup> استفاده شد، تابع هدف بصورت نشان داده شده در رابطه (۱۰)، اندازه بردار اختلاف سیگنال بدست آمده،  $x_{rec}$  و سیگنال مطلوب،  $x_d$  معرفی شده است:

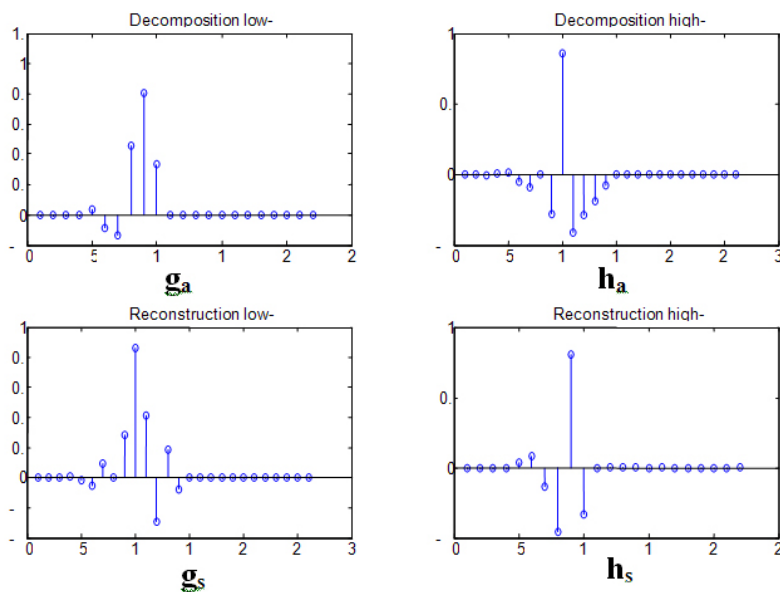
$$f = \|x_{rec} - x_d\| \quad (10)$$

در تابع هدف فوق سیگنال مطلوب،  $(x_d)$ ، سیگنال یک نمونه استاندارد با طیف از پیش شناخته شده (استواستات) در نظر گرفته شده؛ سپس طیف همان نمونه با استفاده از سیستم رامان اخذ شده و به جای ورودی برای بهینه‌سازی بکار رفته است. در هر مرحله طیف ورودی با اعمال یک زنجیره LS به ۵ سطح تجزیه شده است و با استفاده از سطوح چهارم و پنجم جزئیات بازسازی شده  $(x_{rec})$  و با سیگنال استاندارد مقایسه می‌شود. سرانجام این زنجیره به گونه‌ای انتخاب می‌شود که سیگنال بازسازی شده و سیگنال مطلوب کمترین اختلاف را طبق رابطه (۹) داشته باشند.

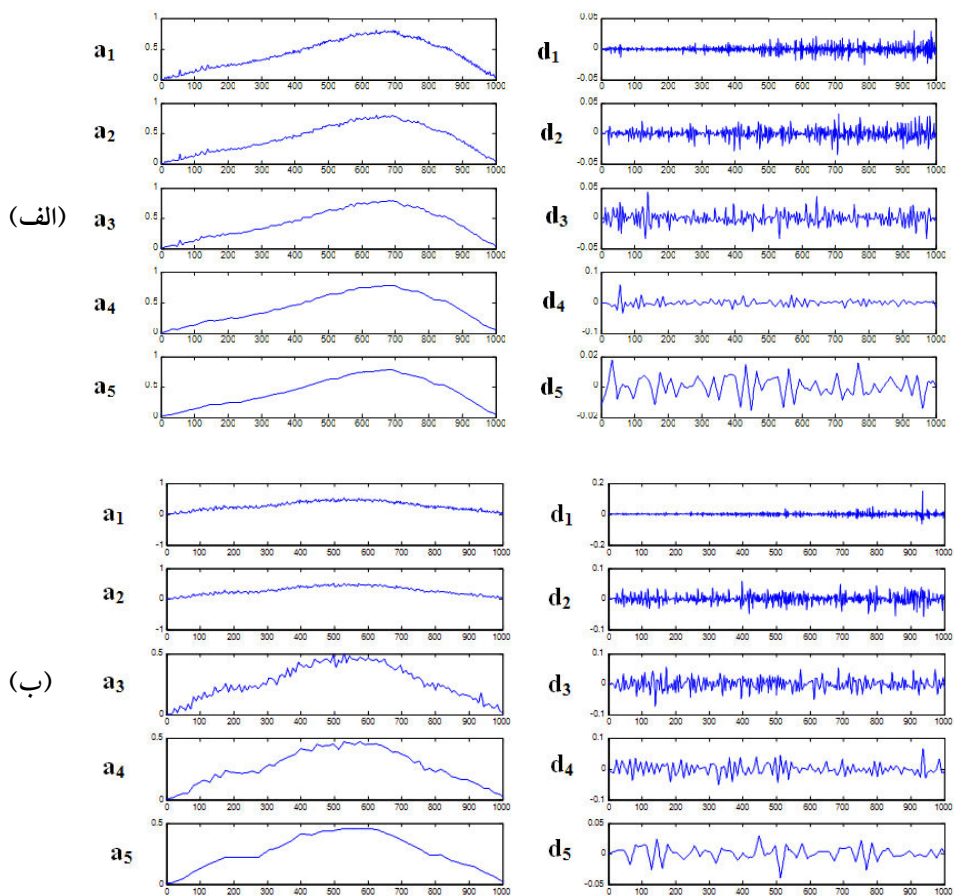
پس از انجام بهینه‌سازی زنجیره LS بدست آمده بصورت زیر است:

<sup>10</sup> Laurent<sup>11</sup> pattern search<sup>12</sup> Lifting scheme





شکل (۸) - صافی‌های بهینه بدست آمده



شکل (۹) - مراحل تجزیه برای یک نمونه طیف سیستم (الف) Almega (ب) R3000

طیف‌ها بین دو طبقه سالم و سرطانی اقدام کردیم. در طبقه‌بندی به منظور آموزش و آزمایش طبقه‌بند از روش «یکی را کنار بگذار»<sup>۱۳</sup> استفاده کردیم. این روش در مواردی که تعداد داده‌ها کم است؛ روش مناسبی محسوب می‌شود. در مقایسه SVM طبقه‌بندهای مختلف بکار رفته، (جدول ۱)، طبقه‌بند کارایی بهتری از خود نشان داده است. همانگونه که در جدول توانستیم به نرخ SVM مشاهده می‌شود، با استفاده از طبقه‌بند تشخیص صحیح<sup>۱۴</sup> ۸۳٪ با حساسیت<sup>۱۵</sup> ۸۹٪ و ویژگی<sup>۱۶</sup> ۷۳٪ در تمایز نمونه‌های سالم و سرطانی بافتهای مختلف و با طیف‌های بدست آمده از دو سیستم متفاوت، دست یابیم.

جدول (۱) - نتایج بدست آمده از طبقه‌بندهای مختلف در تمایز

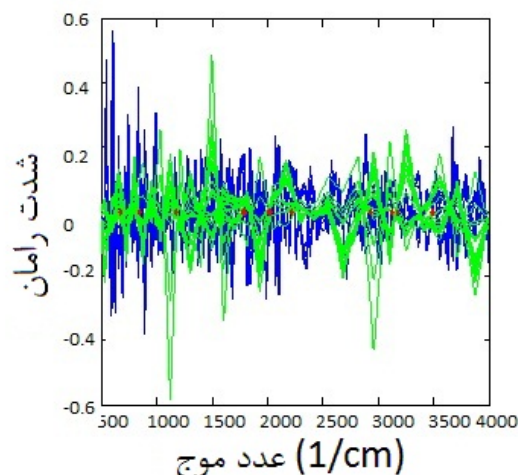
نمونه‌های سالم و سرطانی بدخیم

NN	KNN	SVM	LDA	
۰/۷۵۰۰	۰/۷۷۰۸	۰/۸۳۳۳	۰/۷۵۰۰	صحت
۰/۸۲۷۶	۰/۸۲۷۶	۰/۸۹۶۶	۰/۷۹۳۱	حساسیت
۰/۶۳۱۶	۰/۶۸۴۲	۰/۷۳۶۸	۰/۶۸۴۲	ویژگی

## ۶- نتیجه‌گیری

نتایج بدست آمده از این تحقیق، حاکی از قابلیت‌های روش طیف‌سنجی رامان در تشخیص سرطان است. بعلاوه استفاده از دو سیستم طیف‌سنجی کاملاً متفاوت در این تحقیق و رفع تفاوت‌های سیستمیک در طیف‌ها، راه حلی را برای رفع مشکل تعمیم‌پذیری روشهای تشخیصی در سیستمهای مختلف ارائه می‌کند. همچنین با توجه به بررسی بافتهای مختلف بصورت همزمان در فرایند تشخیص سرطان، توانسته‌ایم عوامل تفاوت‌های بافتی را از فرایند تشخیص حذف کنیم و به ویژگی‌هایی دست یابیم که صرفاً ناشی از سرطانی شدن بافت است.

بدیهی است برای دستیابی کامل به هدف تشخیص سرطان مستقل از بافت و نوع سیستم طیف‌نگاری، باید مطالعات گسترده‌تری از نظر تنوع بافتی و سیستمی انجام شود، که در برنامه پژوهشی آینده این گروه تحقیقاتی قرار گرفته است.



شکل (۱۰) - طیف‌های بدست آمده پس از اعمال LWT (طیف‌های آبی طبقه سرطانی و طیف‌های سبز طبقه سالم و \* ویژگی‌های بدست آمده از آمارگان T)

## ۵- طبقه‌بندی

در مواردی که تعداد نقاط ویژگی از تعداد نمونه‌ها خیلی بیشتر باشد، استفاده از روشهای داده کاوی به منظور استخراج و انتخاب ویژگی حائز اهمیت خواهد بود؛ زیرا استفاده از تمام ویژگی‌ها عملی نیست و موجب کاهش کارایی مدل می‌شود.

با فرض اینکه ماتریس داده ورودی،  $D_{N \times M}$  با  $N$  نمونه و  $M$  ویژگی وجود دارد، هدف استخراج ویژگی پیدا کردن یک زیر فضا با  $m$  ویژگی؛  $(\mathbb{R}^m)$ ، از فضای مشاهدات دارای  $M$  ویژگی،  $(\mathbb{R}^M)$ ، خواهد بود؛ بگونه‌ای که نقاط این زیر فضا از نظر آماری دارای اختلاف معنی‌دار باشند.

برای تعیین میزان تفکیک‌پذیری دو طبقه، به بررسی فاصله دو طبقه با استفاده از آمارگان  $T$ ، پرداختیم. در آزمون آمارگان  $T$ ، با فرض توزیع طبیعی نمونه‌ها در هر طبقه، از رابطه (۱۱) برای محاسبه آمارگان  $T$  استفاده کردیم [۲۲].

$$T = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (11)$$

در رابطه فوق،  $\mu$ ،  $S$  و  $n$  به ترتیب میانگین، واریانس و تعداد نمونه‌های جوامع آماری ۱ و ۲ است. سپس با استفاده از آمارگان  $T$ ، تعداد ۱۵ ویژگی از کل نقاط داده را انتخاب کردیم، ویژگی‌های بدست آمده با استفاده از این آزمون در شکل (۱۰) با \* نشان داده شده است.

سپس با استفاده از ویژگیهای بدست آمده به طبقه‌بندی

<sup>13</sup> Leave one out

<sup>14</sup> Correct rate

<sup>15</sup> Sensitivity

<sup>16</sup> Specificity

## مراجع

- Spectroscopy; *J Molecular Liquids*, 2008; 141: 145-148.
- [13] Abramczyk H., Surmacki J., Brozek-Pluska B., Morawiec Z., Tazbir M., The Hallmarks Of Breast Cancer By Raman Spectroscopy; *J Molecular Structure*, 2009: 175-182.
- [14] Beljebbar A., Bouche O., Diebold M. D., Guillou P. J., Palot J. P., Eudes D., Manfait M., Identification Of Raman Spectroscopic Markers For The Characterization Of Normal And Adenocarcinomatous Colonic Tissues; *J Critical Reviews In Oncology/Hematology* , 2009; 72: 255-264.
- [15] Wills H., Kast R., Stewart C., Rabah R., Pandya A., Poulik J., Auner G., Klein M. D., Raman Spectroscopy Detects And Distinguishes Neuroblastoma And Related Tissues In Fresh And (Banked) Frozen Specimens; *J Pediatric Surgery*, 2009; 44: 386-391.
- [16] Wills H., Kast R., Stewart C., Sullivan B., Rabah R., Poulik J., Pandya A., Auner G., Klein M. D., Diagnosis Of Wilm's Tumor Using Near-Infrared Raman Spectroscopy; *J Pediatric Surgery*, 2009; 44: 1152-1158.
- [17] Lyng F. M., Faolain E. O., Meade A. D., vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool; *Experimental and molecular pathology*, 2007; 82: 121-129.
- [18] Lin C. J., Projected Gradient Methods for Non-negative Matrix Factorization; *Neural Computation*, October 2007, vol. 19.
- [19] Mallat S., *A wavelet tour of signal processing*: Academic Press, 1998.
- [20] Strang G., Nguyen T., *Wavelets and filter banks*: Wellesley-Cambridge Press, 1996.
- [21] Sweldens W., The Lifting Scheme: a Construction of Second Generation of Wavelets; *SIAM J. Math. Anal.*, 1998; 29: 511-546.
- [22] Hollander M., Wolfe D., *Nonparametric Statistical Methods*: Wiley, 1999.
- [1] Lewis I. R., *Handbook of Raman spectroscopy*: Marcel Dekker, 2001.
- [2] Twardowski J., Anzenbacher P., *Raman and IR Spectroscopy in Biology and Biochemistry*: Ellis Horwood, 1994.
- [3] Wolthuis R., Schut T. C. B., Caspers P. J., Buschman H. P. J., Romer T. J., Bruining H. A., Puppels G. J., *Raman spectroscopic methods for in vitro and in vivo tissue characterization; in fluorescent and luminescent probes for biological activity*, 2nd ed: Academic Press, 1999.
- [4] Itani Y., Itoh K., Adachi S., Koyama M., Kimura T., Chin R., Takemura T., Malignant Epithelial Tumor of Unknown Origin of the Broad Ligament; *Arch Gynecol Obstet*, 2002; 267: 113-116.
- [5] VandeWouw A. J., Jansen R. L. H., Speel E. J. M., Hillen H. F. P., *The Unknown Biology of the Unknown Primary Tumour: A Literature Review*; *Annals of Oncology*, 2003; 14: 191-196.
- [6] Ismael G., Azambuja E. D., Awada A., *Molecular Profiling of a Tumor of Unknown Origin*; *The New England Journal Of Medicine*, 2006; 1071-1072.
- [7] Cotran R. S., Kumar V., Robbins S. L., *Pathologic Basis of Diseases*, 8th ed.: Philadelphia, Pa : Saunders Elsevier, 2010.
- [8] Tannock I. F., Hill R. P., Bristow R. G., Harrington L., *The Basic Science Of Oncology*, 4th ed.: McGraw-Hill 2004.
- [9] Gniadecka M., Wulf H. C., Mortensen N. N., Nielsen O. F., Christensen D. H., *Diagnosis Of Basal Cell Carcinoma By Raman Spectroscopy*; *J Raman spectroscopy*, 1997; 28: 125-129.
- [10] Sigurdsson S., Philipsen P. A., Hansen L. K., Larsen J., Gniadecka M., *Detection Of Skin Cancer By Classification Of Raman Spectra*; *IEEE transactions on biomedical engineering*, 2004; 51: 1784-1793.
- [11] Park A., Jung I. W., Baek S. J., Kim J. Y., Shin D., *An enhancement on the detection of skin cancer based on band approach of confocal raman spectra; in International symposium on information technology conference*, 2007: 273-277.
- [12] Brozek-Pluska B., Placek I., Kurczewski K., Morawiec Z., Tazbir M., Abramczyk H., *Breast Cancer Diagnostics By Raman*