

ارائه یک مدل هوشمند قطعه بندی جهت شناسایی همپوشانی سلولی در تصاویر دیجیتال مبتنی بر منطق فازی و تبدیل موجک گسسته جهت شناسایی سرطان معده

پانید تیموری^۱، مهدی مزینانی^{۲*} و راحیل حسینی^۳

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۳۹۸/۱۰/۲۵	سرطان معده در ایران اولین عامل و در سطح دنیا پنجمین عامل پیشتاز سرطان می باشد. اگر این بیماری در مراحل اولیه تشخیص داده شود، احتمال درمان بیشتر و هزینه های آن کاهش چشمگیری خواهد داشت. به دلیل پیچیدگی تصاویر پاتولوژیستی و چالش های اساسی موجود در این تصاویر نظیر کنتراست ضعیف بین سلول ها، هم پوشانی سلول ها و تناقض در رنگ آمیزی بافت، فرایند تشخیص این نوع بیماری با مشکل روبرو می شود؛ بنابراین در این پژوهش، مدل هوشمند جدیدی جهت حل این مشکلات ارائه گردیده است؛ به گونه ای که ابتدا از الگوریتمی مبتنی بر کلاسه بالا به پایین جهت بهبود کیفیت تصویر بهره گرفته شده و سپس با استفاده از روش های خوشه بندی فازی، تبدیل موجک گسسته، رشد منطقه و مکانیزم رای گیری جهت تشخیص سلول ها اقدام می شود. سپس با کمک یک روش مبتنی بر عملیات مورفولوژی پیشرفته، سلول های دارای هم پوشانی از یکدیگر جدا می شوند و در نهایت ویژگی های سلول، استخراج و به کمک الگوریتم ماشین بردار پشتیبان با کرنل (RBF)، طبقه بندی می گردند. الگوریتم ارائه شده بر روی ۹۶ تصویر دیجیتال میکروسکوپی بیماران بیمارستان بقیه الله اعمال گردیده و با روش تحلیل منحنی ROC ارزیابی شده است. نتایج به دست آمده توسط متخصص پاتولوژیست تأیید شده و دقت تشخیص سلول های سالم و سرطانی ۹۲،۱۲٪ و سلول های خوش خیم و بدخیم ۹۴،۱۴٪ می باشد که برای تشخیص زود هنگام این نوع سرطان امیدوارکننده است.
پذیرش مقاله: ۱۳۹۹/۰۸/۱۱	
واژگان کلیدی: منطق فازی، تبدیل موجک گسسته، مکانیزم رای گیری، سرطان معده، ماشین بردار پشتیبان.	

۱- مقدمه

معده در ایران اولین سرطان شایع بوده است و در جمهوری آذربایجان، شمال و شمال غربی ایران شیوع بالاتری نسبت به شرق و جنوب دارد [۴]. طبق پیش بینی های صورت گرفته در سال ۲۰۲۰ نرخ بروز سرطان معده در ایران می تواند به طور قابل ملاحظه ای افزایش یابد [۵]. سرطان معده انواع مختلفی دارد و علل ایجاد و رشد سریع این بیماری در مرحله اولیه، ناشی از زمینه ی ژنتیکی، رژیم غذایی، دخانیات، چاقی، سن، آلودگی به عفونت هلیکوباکتر

سرطان معده یکی از شایع ترین سرطان ها در سراسر جهان است [۱]. میزان بروز بیماری در مردان (۲۹/۱-۱۹/۲) در صد و در زنان (۱۵/۱-۹/۷) درصد است [۱]. سرطان معده شیوع بالایی در شرق آسیا [۲]، جنوب آمریکا، شیلی و پرو دارد [۳]. طبق اطلاعات ارائه شده متأسفانه کشور ایران یکی از کانون های شیوع این بیماری است، به گونه ای که بعد از سرطان سینه که در خانم ها رتبه ی اول را دارد، سرطان

* پست الکترونیک نویسنده مسئول: mahdi.mazini@qodsiau.ac.ir

۱. دانشجوی کارشناسی ارشد، گروه مهندسی برق، واحد شهرقدس، دانشگاه آزاد اسلامی، تهران، ایران

۲. استادیار، گروه مهندسی برق، واحد شهرقدس، دانشگاه آزاد اسلامی، تهران، ایران

۳. استادیار، گروه مهندسی کامپیوتر، واحد شهرقدس، دانشگاه آزاد اسلامی، تهران، ایران

است و تاکنون نرم‌افزاری مبتنی بر تصاویر پاتولوژیستی جهت تشخیص این نوع سرطان لحاظ نشده، استفاده از روش‌های هوشمند و پردازش تصویر جهت تشخیص زود هنگام و جلوگیری از مرگ بیمار ضروری به نظر می‌رسد. با توجه به پیچیدگی‌های تصاویر پاتولوژیستی و شرایط چالش برانگیز این تصاویر شامل کنتراست ضعیف بین سلول‌ها، هم‌پوشانی سلول‌ها، مرزهای سلول، تناقض در رنگ آمیزی بافت، تکنیک‌های پردازش تصویر را در فرایند تشخیص این نوع بیماری با چالش‌های زیادی روبرو می‌کند [۸]. هدف از این پژوهش ارائه‌ی یک روش هوشمند جهت تشخیص سرطان معده می‌باشد. با توجه به اینکه این بیماری از بیماری‌های همه‌گیر در کشورهای اروپا و آمریکا نبوده و تنها به صورت منطقه‌ای در نقاط خاص دیده می‌شود. لذا با توجه به آنکه در بررسی‌های صورت گرفته گزارشی از انجام پژوهش به روی تصاویر پاتولوژی معده مشاهده نشد؛ جهت تکمیل و مدیریت مسیر پژوهش، روش‌های مشابه موفق که بر روی تصاویر پاتولوژی سایر ارگان‌های بدن اعمال شده بودند، مورد بررسی قرار گرفته و بر این اساس، با بهبود روش‌های پیشین، روش نوینی جهت تشخیص بهتر این بیماری با استفاده از تصاویر پاتولوژی ارائه خواهد شد.

۲- مروری بر تحقیقات

شهریار و همکاران، روشی جهت تشخیص بافت تخمدان به صورت خودکار با استفاده از تصاویر رنگ‌آمیزی شده‌ی (P^{۶۳}) غیرشمارنده برای به حداقل رساندن مشاهدات متنوع متخصصان پاتولوژیست به ثبت رسانده‌اند. در این روش ابتدا تصاویر رنگ‌آمیزی شده‌ی (P^{۶۳}) غیرشمارنده با بزرگ‌نمایی ۴۰X، ۱۰۰X و ۲۰۰X برای شناسایی منطقه‌ی موردنظر در هسته‌های تخمدان بررسی شده و سپس مراحل پردازش، اعمال می‌شود که شامل: اصلاح تغییرات شدت روشنایی، عملکرد فیلتر، تقسیم‌بندی تصویر، عملکرد خوشه‌بندی برای تقسیم‌بندی تصاویر، شناسایی و در آخر طبقه‌بندی برای بهبود دقت می‌باشد و همچنین نتیجه‌ی دقت در شناسایی تصاویر (P^{۶۳}) غیرشمارنده با استفاده از روش طبقه‌بندی (SVM^{۱۱}) برای هر سه بزرگ‌نمایی ۴۰X، ۱۰۰X و ۲۰۰X آزمایش شده که

پیلوری، عوامل محیطی و... می‌باشد و اگر در مراحل اولیه مورد درمان مناسب قرار نگیرد به هایپرپلازی^۱ (پیش‌رویش تومور) منجر و اگر از رشد آن جلوگیری نشود به متاپلازی^۲ (تغییر ماهیت سلول در بافت) و سپس به دیسپلازی^۳ (رشد و گسترش سلول) تبدیل خواهد شد. در نتیجه با گذشت زمان وضعیت بیماری رشد می‌کند و به مرحله‌ی پیشرفته می‌رسد که ۹۵ درصد از تومورهای بدخیم اپیتلیال^۴ (مخاطی) شامل آدنوکارسینوم^۵ معده می‌باشد [۶]. به منظور تشخیص این بیماری پس از مطالعات کلینیکی و بررسی آزمایش خون و... چنانچه پزشک احتمال دهد که بیمار به سرطان معده مبتلا شده است، معاینه را انجام می‌دهد. در صورت پیشرفته نبودن سرطان روش‌های تشخیصی مانند: نمونه‌برداری از بافت و آندوسکوپی^۶ صورت می‌گیرد که متأسفانه تومورهای بدخیم در آندوسکوپی قابل تشخیص نیست. بنابراین از روش‌های غیرتهاجمی مثل رادیوگرافی، سی تی اسکن، ام آر آی و... تنها برای مشخص شدن میزان گسترش و رشد بیماری استفاده می‌شود. پرکاربردترین روش جهت تشخیص و تعیین سطح سرطان نمونه‌برداری از بافت معده می‌باشد که متخصص پاتولوژیست بافت نمونه برداری شده از معده بیمار را مورد بررسی قرار می‌دهد. به این منظور ابتدا نمونه‌ی بافت از بدن بیمار استخراج و پس از پرورش نمونه بافت مشکوک با رنگ‌آمیزی هماتوکسیلین و ائوزین^۷ (H&E) رنگ می‌شود [۷]. بعد از مراحل آماده سازی، متخصصان پاتولوژیست^۸ بیماری را با استفاده از میکروسکوپ‌های نوری و با بررسی نمونه‌های بافت، از نظر قواعد اشکال سلول، تغییر در شکل هسته^۹، اندازه‌ی هسته، وضع هسته و هسته‌ی میتوزی^{۱۰} تشخیص می‌دهند [۷]. با این حال به رغم دقت بالای این روش در تجزیه و تحلیل بافت، این فرایند کند بوده و می‌تواند نتایج آن به علت خطاهای انسانی و ابزاری نظیر خستگی و کسالت، تفاوت نظر چند متخصص و عوامل محیطی که مستعد ایجاد خطای بصری می‌شوند تحت تأثیر قرار گیرد؛ بنابراین با استفاده از پردازش تصویر می‌توان از مشکلات به وجود آمده جلوگیری کرد [۷]. از طرفی با توجه به اینکه سرطان معده یکی از شایع‌ترین بیماری‌ها در ایران

⁷ Hematoxylin and Eosin

⁸ Pathologist

⁹ Nuclei

¹⁰ Mitotic

¹¹ Support Vector Machine

¹ Hyperplasia

² Metaplasia

³ Dysplasia

⁴ Epithelial

⁵ Adenocarcinoma

⁶ Endoscopy

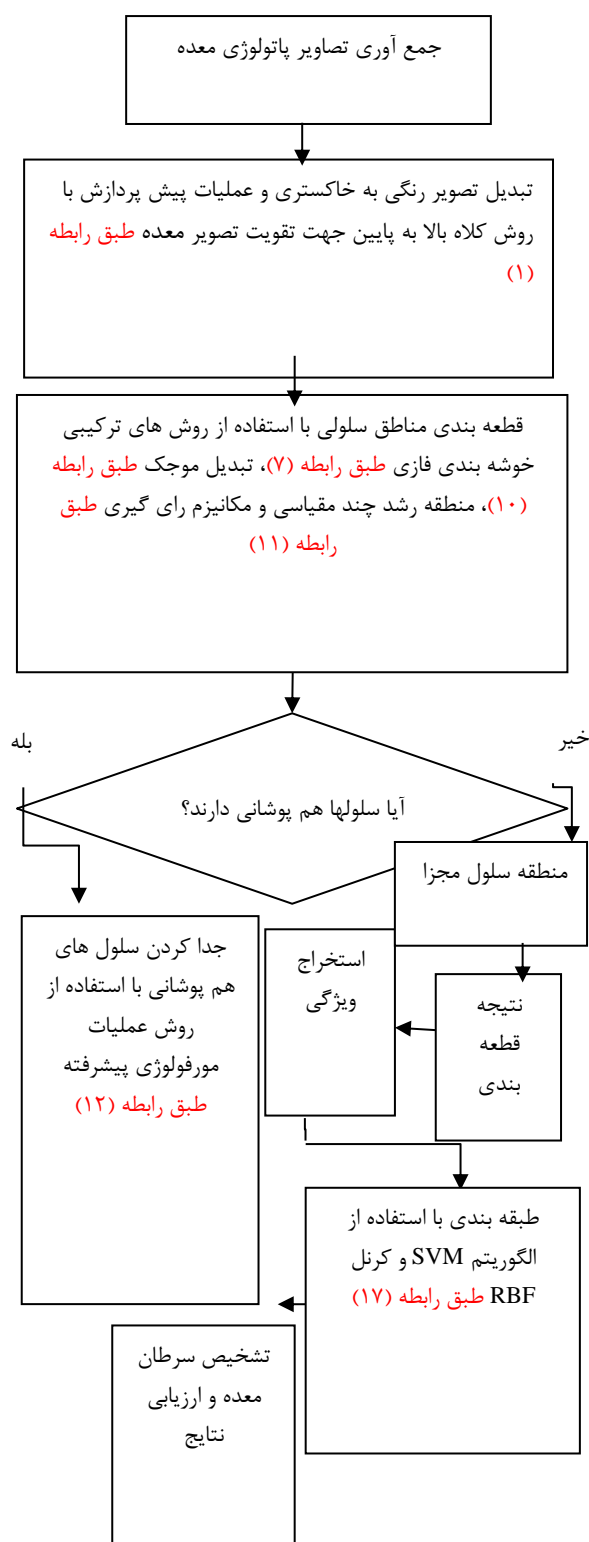
استفاده کرده‌اند؛ این مدل‌های هوشمند شامل: طبقه‌بندی ماشین‌بردار پشتیبان، تقسیم‌بندی آستانه‌ی تطبیقی در جهت پیدا کردن مقدار بهینه، فیلتر مورفولوژیکی جهت حذف اشکال در طبقه‌بندی به روش رأی اکثریت، فیلتر آب‌پخشان جهت دقت برش و ارائه‌ی یک تصویر روشن می‌باشند. فیلتر آب‌پخشان دقت مناسبی را در قطعه‌بندی نشان نمی‌دهد؛ همین‌طور ابزار پیشنهادی، دقت ۹۵٫۳ درصد را نشان می‌دهد [۱۱]. اسواتی و جاگانت، روش‌های رشد منطقه و الگوریتم فازی (FCM) را جهت قطعه‌بندی مناطق مورد نظر به دلیل راندمان بالای آنها در محاسبات ارائه نموده‌اند. یک تصویر، شامل اطلاعات برداری و جداسازی پیکسل به پیکسل است که می‌توان با استفاده از کمترین مانع، فاصله‌ی برداری را یافت؛ همچنین مقادیر شدت روشنایی پیکسل‌ها دارای ویژگی‌های متفاوتی می‌باشد که این روش نیازی به آموزش ندارد. حدود ۷۹۰۰ تصویر از ۸۲ بیمار جمع‌آوری شده است که دقت روش پیشنهادی، ۸۵ درصد را نشان می‌دهد. همچنین نتایج نشان می‌دهد که روش (FCM)، یک الگوریتم قدرتمند برای تشخیص سلول می‌باشد [۱۲]. اسدی‌امیری و حسن‌پور، یک روش پیش‌پردازش مناسب با استفاده از عملگر توان برای بهبود عملکرد JPEG 2000 ارائه داده‌اند؛ در این روش با افزایش وضوح تصویر در مرحله پیش‌پردازش، نواحی فرکانس بالای تصویر بیشتر رویت می‌شوند. در نتیجه ضرایب تبدیل موجک مربوط به این نواحی تصویر، مقادیر بزرگی می‌گیرند؛ حال اگر ضرایب موجک به صفحات بیتی تجزیه گردد، اطلاعات زیادی در صفحات بیتی پرارزش‌تر قرار گرفته است. از آنجایی که JPEG 2000 بسته به نرخ‌سازی مورد نظر، فقط چند صفحه بیتی پرارزش‌تر از ضرایب موجک کد می‌شوند، با این پیش‌پردازش اطلاعات فرکانس بالای تصویر نیز تا حد بیشتری حفظ می‌گردند؛ در نتیجه تصویر بازیابی شده کیفیت مطلوب‌تری خواهد داشت. در روش پیشنهادی به طور متوسط حدود ۳٫۵ درصد نرخ فشرده‌سازی را بهبود می‌دهد و در زمان کد‌گشایی، اتلاف زمانی وجود دارد که خود، مشکلاتی را به وجود می‌آورد [۱۳]. زارع و محمدزاده، دو الگوریتم آشکارساز، جهت شناسایی نویزهای ضربه‌ای در تصویر نویزدار معرفی کرده‌اند. در این تحقیق جهت بررسی

میزان دقت در این تصاویر بیش از ۹۰ درصد است و زمان محاسباتی، نسبتاً زمان خوبی است اما سرعت پایینی دارد [۸]. طبق پژوهش کائول و همکاران، از روش آستانه‌ی تطبیقی با استفاده از تصاویر میکروسکوپی سرطان پستان برای تقسیم‌بندی هسته و الگوریتم‌های خوشه بندی (FCM^۱, KM^۲) جهت شناسایی هسته در اشیای پیش‌زمینه و اعمال الگوریتم‌های طبقه‌بندی بیز ساده و درخت تصمیم‌گیری استفاده شده است. مشکل اصلی تقسیم‌بندی به روش آستانه، انتخاب درست آستانه با مقدار T است. بهترین روش طبقه‌بندی، الگوریتم نزدیک‌ترین همسایه (K-NN) و خوشه‌بندی فازی c میانگین (FCM) برای تصاویر منحصر به فرد می‌باشد و نتایج طبقه‌بندی، میزان ۹۳٫۱۱ درصد را نشان می‌دهد [۹]. وان و همکاران، یک روش ویژگی آماری مبتنی بر تبدیل موجک گسسته برای جدا کردن سلول‌های میتوزی از غیرمیتوزی در تصاویر هیستوپاتولوژی سرطان پستان ارائه نموده‌اند. هدف از این کار بهبود دقت طبقه‌بندی میتوز توسط مشخص کردن اشیای مورد نظر (سلول) در ارائه‌ی رزولوشن چندگانه‌ی مبتنی بر تبدیل موجک گسسته است که ویژگی‌های آماری را با داشتن تمایز میتوزی بهتر می‌کند. یک تبدیل موجک پیچیده‌ی دوگانه برای تجزیه کردن تیکه‌های تصویر به شکل چندمقیاسی انجام گرفته است. از آنجایی که هر دو سلول به عنوان اشیای کوچک با انواع مختلف شکل‌ها در تصاویر ظاهر می‌شوند، توصیف میتوز، یک مشکل چالش برانگیز است. این روش عملکرد خوبی را نسبت به روش‌های قطعه‌بندی نشان داده است. مجموعه داده، شامل ۵ اسلاید بیوپسی از بافت پستان با رنگ‌آمیزی (H&E) با بزرگنمایی ۴۰x می‌باشد. در هر اسلاید متخصص پاتولوژیست ۱۰ تصویر را مشخص و ویژگی‌های آماری مبتنی بر تبدیل موجک گسسته روی مجموعه داده شامل ۳۲۷ سلول میتوزی و ۴۰۶ سلول غیرمیتوزی از طریق الگوریتم (SVM) ارزیابی کرده است. نتایج نشان می‌دهد که تبدیل موجک گسسته روش پیشنهادی با دقت ۸۷٫۹۴ درصد و حساسیت ۸۶٫۸۰ درصد و ارزش سطح زیر منحنی ۰٫۹۴ درصد گزارش شده است [۱۰]. گوداس و همکاران، از مدل‌های هوشمند و روش‌های پردازش تصویر، جهت تشخیص سلول‌های سرطانی از تکنیک‌های تجزیه و تحلیل تصویر

³ K-Nearest Neighbor

¹ Fuzzy C-Mans

² K-Means



شکل ۱: بلوک دیاگرام روش پیشنهادی

سپس جهت استخراج مناطق مورد نظر (سلولها) از روش های مبتنی بر خوشه بندی فازی (FCM)، تبدیل موجک

روش پیشنهادی، نویزهای ضربه ای با سه نوع توزیع متفاوت مدل سازی به تصویر اصلی اضافه شده اند و تصویر بازسازی شده با تصویر اصلی مورد مقایسه قرار گرفته است؛ بنابراین برای ارزیابی نویزهای شناسایی شده به کمک الگوریتم های آشکار ساز پیشنهادی و کیفیت تصاویر بازسازی شده از شاخص صحت کلی و شاخص نسبت اوج سیگنال به نویز به ترتیب استفاده شده است. نتایج ارزیابی ها نشان می دهند که روش پیشنهادی بر مبنای دو الگوریتم آشکار ساز معرفی شده که به مراتب بهتر از روش فیلترینگ دو عبوری میانه می باشد و نسبت به یکی از روش های معرفی شده قبلی یعنی فیلترینگ دو عبوری میانه ی توافقی نتایج بهتری را نشان می دهد. کیفیت تصاویر بهینه سازی شده توسط روش پیشنهادی ایشان، بهتر از دو روش دیگر است؛ این افزایش کیفیت به طور میانگین و به ترتیب با افزایش ۲،۳۷ و ۴،۱۷ دسیبل در شاخص نسبت اوج سیگنال به نویز می باشد [۱۴]. در پژوهشی دیگر توسط رضایی پناه و همکاران، فرآیند تشخیص سرطان سینه با یک رویکرد دو مرحله ای ثبت شده است؛ در مرحله اول دو پارامتر ویژگی های موثر و تعداد نودهای لایه ی مخفی به منظور آموزش شبکه ی عصبی (MLP^۱) به صورت همزمان توسط یک الگوریتم ژنتیک، بهینه سازی می شوند. سپس با استفاده از ویژگی های انتخاب شده و تعداد نودهای لایه ی مخفی، یک مدل طبقه بندی بر مبنای شبکه ی عصبی (MLP) برای تشخیص سرطان سینه در مرحله دوم ایجاد می شود. در این مرحله از یک الگوریتم ژنتیک موازی (Fine Grain) بر مبنای پارامترهای بهینه سازی شده، برای تنظیم وزن های شبکه ی عصبی (MLP) استفاده می شود. ارزیابی آمایش ها نشان می دهد که روش پیشنهادی ایشان، در مقایسه با روش (GAANN^۲) روی مجموعه داده ی گلوبول سفید خون به نتایج بهتری رسیده است و دقت ۹۸،۷۲٪ را در حالت میانگین گزارش می کند.

۳- مراحل روش پیشنهادی

در این پژوهش ابتدا همانطور که در شکل (۱) نشان داده شده است پس از دریافت تصویر در روش پیشنهادی، از روش کلاه بالا به پایین برای بهبود کیفیت تصویر در قسمت پیش پردازش استفاده می شود.

² Genetic Algorithm Artificial Neural Networks

¹ Multi Layer Perceptron

که در آن نمادهای $f \circ b$ و $f \bullet b$ ، به ترتیب عملیات گشودن و بستن را بر تصاویر سیاه و سفید f با عنصر ساختار b را نشان می‌دهد.

تفریق تصویر اصلی از تصویر باز شده و بسته شده به تبدیل-کلاه بالا و کلاه پایین منجر می‌شود [۱۷]. تبدیل کلاه بالا (Top-Hat) معادل با کم کردن تصویر باز شده از تصویر اصلی مطابق با رابطه (۲) است:

$$T_{hat} = f - (f \circ b) \quad (2)$$

که در آن f نشان دهنده‌ی تصویر اصلی و b نشان دهنده‌ی عنصر ساختاری است و به طور مشابه تبدیل کلاه پایین (Bot-Hat) معادل با کم کردن تصویر بسته شده از تصویر اصلی f مطابق با رابطه‌ی (۳) است:^۲

$$B_{hat} = (f \bullet b) - f \quad (3)$$

تبدیل کلاه بالا برای ارتقای ناحیه‌های روشن و تبدیل کلاه پایین برای ارتقای ناحیه‌های تیره‌ی تصویر به کار می‌رود. با توجه به تصویر، از ترکیب‌های مختلف دو تبدیل کلاه بالا و پایین برای ارتقای جزئیات روشن یا تیره‌ی تصویر، استفاده می‌شود. روشی غیریکنواخت پس‌زمینه در قطعه‌بندی تصویر اثرات ناخوشایندی دارد که با به کارگیری این دو تبدیل می‌توان بر آن فائق آمد [۱۷].

۲-۳- قطعه‌بندی مناطق سلولی با استفاده از روش فازی، تبدیل موجک گسسته، رشد منطقه و مکانیزم رای‌گیری

پس از پیش‌پردازش، کنتراست بین سلول‌ها و پس‌زمینه تقویت می‌شود. اما سلول‌ها از پس‌زمینه جدا نمی‌شوند. با توجه به آن که کنتراست و اندازه‌ی سلول‌های مختلف، متفاوت هستند بنابراین برای استخراج مناطق مورد نظر، روش خوشه‌بندی فازی مورد استفاده قرار می‌گیرد. در این بخش ابتدا با استفاده از خوشه‌بندی فازی، قطعه‌بندی انجام و سپس برای افزایش دقت قطعه‌بندی از تبدیل موجک گسسته، منطقه رشد چندمقیاسی و مکانیزم رای‌گیری استفاده می‌شود.

در خوشه‌بندی فازی، یک تفکیک فازی صورت می‌گیرد؛ بدین معنا که هر داده با یک درجه‌ی عضویت به هر خوشه متعلق است. در الگوریتم خوشه‌بندی فازی (FCM)، ابتدا باید تعداد و مراکز خوشه‌ها توسط کاربر مشخص شوند.

گسسته، منطقه رشد چندمقیاسی و مکانیزم رای‌گیری استفاده می‌شود و برای تفکیک سلول‌های دارای همپوشانی روشی مبتنی بر مورفولوژی ریاضی پیشرفته ارائه گردیده است. پس از تفکیک سلول‌ها، جهت طبقه‌بندی، ابتدا ویژگی‌های مربوط به بافت و سلول را به عنوان مجموعه ویژگی‌های اولیه استخراج و با استفاده از الگوریتم ژنتیک، مجموعه ویژگی بهینه انتخاب می‌شود و در نهایت این مجموعه ویژگی بهینه، توسط الگوریتم طبقه‌بندی کننده‌ی (SVM) با کرنل (RBF) برای دستیابی به دقت بالا، طبقه‌بندی می‌شود که در ادامه به توضیح روش پیشنهادی، پرداخته می‌شود.

۱-۳- پیش‌پردازش با استفاده از روش کلاه بالا به پایین^۱

در این مرحله به دلیل استفاده از مواد شیمیایی در طول آماده‌سازی اسلاید بیوپسی، شدت روشنایی و کیفیت تصاویر انواع بافت‌ها تغییر می‌کند، بنابراین روشی جهت افزایش کیفیت تصاویر میکروسکوپی ارائه می‌شود اما باید توجه داشت که یکی از اهداف اصلی این روش، حفظ کردن مرزهای تیز سلول در هنگام افزایش کیفیت تصاویر می‌باشد [۱۶]. برای شروع، ابتدا تصاویر رنگی به تصاویر خاکستری تبدیل می‌شوند سپس برای به دست آوردن اثر قطعه‌بندی بهتر، از روشی به نام تبدیل کلاه بالا به پایین (T_{Bhat}) جهت تقویت کنتراست بین سلول و پس‌زمینه استفاده می‌شود. مزیت اصلی این فیلتر نسبت به فیلترهای هموار ساز، حفظ لبه‌های سلول است چرا که استفاده از روش صاف کردن اثر نامطلوبی داشته و باعث از بین رفتن لبه‌ها می‌شود [۱۶].

دو مرحله‌ی مهم تبدیل کلاه بالا به پایین، کلاه بالا (T_{hat}) و کلاه پایین (B_{hat}) می‌باشد. کلاه بالا (T_{hat}) معمولاً در پس‌زمینه‌ی تاریک برای اشیای نسبتاً سبک و کلاه پایین (B_{hat}) در پس‌زمینه‌ی روشن برای اشیای نسبتاً بزرگ استفاده می‌گردد. فرض شود که f یک تصویر سیاه و سفید باشد، تقویت f به صورت رابطه‌ی (۱) تعریف می‌شود [۱۷]:

$$\begin{cases} f^{enh} = TB_{hat} = f + T_{hat} - B_{hat} \\ T_{hat} = f - (f \circ b) \\ B_{hat} = (f \bullet b) - f \end{cases} \quad (1)$$

^۱ Top-Bottom Hat

در برخورد با نقاط نویزی تصویر، فاصله‌ی بین نقاط نویزی و مقدار مرکز خوشه نسبتاً بزرگ است، که با تغییر قابل توجه‌ای بعد از تکرار، عضویت فازی نقطه‌ی نویزی تخصیص داده می‌شود و در نتیجه عملکرد قطعه‌بندی و کارایی آن بهبود می‌یابد.

تعیین تقریبی تعداد خوشه‌های طبقه‌ی c ، مقدار گیت نهایی e و حداکثر تعداد تکرار L_{max} ، نشان دهنده‌ی اطلاعات تصویر با ماتریس بردار دوبعدی $[F(m, n)]_{(m \times n) \times 2}$ است. تکرار تا زمانی ادامه می‌یابد که مقدار تغییر تابع هدف کمتر از آستانه ϵ شود، یا تعداد تکرار به ماکزیمم تعداد تکرارهای L_{max} قطعه‌بندی تصویر برسد.

در ادامه برای قطعه‌بندی بهتر، از تبدیل موجک گسسته استفاده می‌شود. تبدیل موجک، موجک‌های کوچکی است که موجک‌های فرکانس متغیر و مدت محدود، خوانده می‌شوند. تبدیل موجک گسسته^۱ (DWT) می‌تواند سیگنال یا تصویر را با بیان مجموعه‌ای از نسخه‌های تبدیل شده و دارای مقیاس، از یک موجک مادر^۲ مینا تجزیه کند [۱۸]. در ادامه به توضیحات آن پرداخته می‌شود.

تبدیل موجک گسسته، یک مجموعه توابعی مانند تابع موجک $\psi(x)$ و تابع مقیاس‌بندی $\varphi(x)$ است که تابعی از متغیرهای مجزای $1 - M; 2; \dots; 1$ را اعمال می‌کند. با توجه به اینکه تصویر دیجیتالی اغلب به صورت اطلاعات مجزای دوبعدی گرفته می‌شود، تبدیل موجک گسسته به تابع دوبعدی بسط داده می‌شود. در تابع دو-بعدی، تابع مقیاس‌بندی $\varphi(x; y)$ و سه موجک دوبعدی $(\psi^H(x; y); \psi^V(x; y); \psi^D(x; y))$ مورد نیاز هستند؛ این موجک‌ها، برای تغییرات عملکردی (تغییرات شدت یا سطح سیاه و سفید برای تصاویر) همراه با جهت‌های مختلف سنجیده شده‌اند که موجک‌های حساس به جهت نامیده می‌شوند (موجک ψ^H تغییرات در ستون‌ها، ψ^V تغییر در سطرها، و ψ^D تغییر در قطرها را اندازه‌گیری می‌کند). با توجه به آن که دابشیز^۳ دسته‌ای از موجک‌های واقعی فشرده را می‌سازد، راه حل واقعی برای معادله‌ی موجک ارائه داده است [۱۹].

تابع تصویر سیاه و سفید تقویت شده $f^{enh}(x; y)$ با اندازه‌ی $M \times N$ می‌تواند به دست آید. تبدیل موجک

کیفیت این الگوریتم به تعداد اولیه‌ی خوشه‌ها و مکان اولیه‌ی مراکز خوشه‌ها بستگی دارد و بسیار مهم است.

در الگوریتم خوشه‌بندی (FCM) فاصله‌ی بین پیکسل‌ها و مرکز کلاستر به صورت رابطه‌ی $\|x_j - v_i\|$ تعریف می‌شود، نتایج خوشه‌بندی بهینه را با تکرار پیوسته و بهینه‌سازی تابع هدف که برای FCM تعریف می‌شود طبق رابطه‌ی (۴) به دست می‌آید:

$$J_m(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m \|x_j - v_i\|^2 \quad (4)$$

که در آن $X = \{x_1, x_2, \dots, x_n\}$ مجموعه‌ی نمونه، n تعداد نمونه در فضای خوشه‌ها، رابطه‌ی $V = \{v_1, v_2, \dots, v_c\}$ مراکز خوشه در خوشه‌ی c_i ، پارامتر c_i تعداد خوشه و m میزان فازی بودن در (FCM) است و همچنین $\|x_j - v_i\|$ فاصله‌ی اقلیدسی را نشان می‌دهد که در خوشه‌بندی فازی $m > 1$ تعریف می‌شود. در روش خوشه-بندی فازی، طبق رابطه‌ی (۵)، نباید خوشه‌ها تهی باشند و طبق رابطه‌ی (۶)، مجموع درجه‌ی عضویت هر داده به همه‌ی خوشه‌ها باید برابر یک باشد:

$$0 < \sum_{j=1}^n u_{ij} < n \quad (5)$$

$$\sum_{i=1}^c u_{ij} = 1 \quad (6)$$

در این الگوریتم پس از انجام هر مرحله با توجه به تعیین مقادیر عضویت فازی مرکز خوشه‌ها تغییر می‌کنند و در مرحله بعد با تغییر مرکز خوشه‌ها، فاصله‌ی بین پیکسل و مرکز خوشه باید مجدداً تعیین شود. در نتیجه روابط باید به گونه‌ای باشد که سیستم همگرا شود. به این منظور، مجدداً فاصله‌ی بین پیکسل x_j و مرکز خوشه v_i به صورت $\|x_j^2 - v_i^2\|$ تعریف می‌شود که پس از تکرار اول، عضویت فازی پیکسل را به طور قابل توجهی تغییر می‌دهد و در نتیجه، همگرایی الگوریتم در چند تکرار به دست می‌آید و در نهایت قطعه‌بندی الگوریتم بهبود می‌یابد. تابع هدف (FCM) بهبود یافته به صورت رابطه‌ی (۷) تعریف می‌شود:

$$J_m(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m \|x_j^2 - v_i^2\|^2 \quad (7)$$

³ Daubechies

¹ Discrete Wavelet Transform

² Mother wavelet

شود، اتصال-۳۸ داشته باشد؛ همچنین ملزوم است هر دو معیار همزمان برآورده شوند.

تعیین R و T برای عملکرد منطقه‌ی رشد، حیاتی است. با توجه به مشخصات (اندازه‌ی مختلف شکل، لبه‌های ناپیوسته، سطوح خاکستری داخلی ناهموار، همپوشانی سنگین و...) تصاویر پاتولوژی معده، یک جفت R و T نمی‌توانند به خوبی، سلول‌های مختلف موجود در یک تصویر را نشان دهند که در این پژوهش سه مقیاس از R پذیرفته می‌شود.

برای حذف مناطق غیرسلولی از ۳ مقیاس استفاده می‌شود. مقیاس کوچک برای علامت‌گذاری تمام مناطق بدون همپوشانی به کار می‌رود. مقیاس بزرگ، برای تفکیک مناطق همپوشان و مقیاس متوسط، برای مکمل‌ها می‌باشد. آستانه توسط یک روش منطقه‌ی تکراری تعیین و با آستانه-ی کوچک T شروع می‌شود و هنگامی که تشابه اندازه و خاکستری، تیز شود متوقف می‌شود. مقیاس ۱ می‌تواند سطوح خاکستری داخلی ناهموار و لبه‌های ناپیوسته‌ی درون سلول را پردازش کند، اما محیط سلول‌های همپوشان از یکدیگر جدا نمی‌شوند. مقیاس ۳ می‌تواند محیط همپوشانی سلول را به درستی تفکیک کند؛ اما مشکلات سطوح خاکستری داخلی ناهموار و لبه‌های ناپیوسته شدیدتر هستند. مقیاس ۲ جبران ناکارآمدی مقیاس ۱ و ۳ است. این الگوریتم، سه مقیاس را می‌پذیرد و می‌تواند مسائل پیچیده و مختلف مربوط به تصاویر پاتولوژی را مدیریت کند. سپس هر یک از این روش‌ها با دقت قابل قبولی سلول‌ها را استخراج می‌کنند اما در عین حال چالش‌ها و محدودیت‌هایی را شامل می‌شود. از جمله این چالش‌ها، می‌توان به خطاهایی اشاره کرد که در پیکسل‌های کناری سلول‌ها مشاهده می‌شود؛ همچنین ممکن است، گیر افتادن روش خوشه‌بندی فازی در بهینه‌های محلی تصاویر، دقت مورد نظر را فراهم نکند. برای بهبود دقت روش شناسایی و افزایش قابلیت اطمینان، به جای استفاده از خروجی هر یک از این روش‌ها به تنهایی، از روش رای‌گیری^۴ برای مشخص نمودن برجسته‌ترین پیکسل‌های تصویر استفاده می‌شود. مناطق مورد نظر توسط یک مکانیزم رای‌گیری انتخاب می‌شوند که در این مقاله، در بین انواع روش‌های مختلف قطعه‌بندی، از الگوریتمی ابداعی به

گسسته متناظر نیز می‌تواند به صورت رابطه‌ی (۸ و ۹) تعریف شود [۲۰]:

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \quad (8)$$

$$\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f^{enh}(x, y) \varphi_{j_0, m, n}(x, y)$$

$$W_{\varphi}(j, m, n) = \frac{1}{\sqrt{MN}} \quad (9)$$

$$\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f^{enh}(x, y) \varphi_{j_0, m, n}(x, y) \psi_{j, m, n} =_{H.V.D}$$

که j_0 یک مقیاس شروع اختیاری است، و ضرایب $W_{\varphi}(j_0; m; n)$ تقریبی از $f^{enh}(x; y)$ در مقیاس j_0 می‌باشد و همچنین ضرایب $W_{\psi}^i(j; m; n)$ جزئیات افقی، عمودی و قطری را برای مقیاس‌های $j \geq j_0$ تعریف می‌کند. برای تشخیص لبه‌های سلول، برخلاف تبدیل موجک گسسته‌ی معکوس سنتی، جزئیات جهت، معکوس می‌شوند. نتیجه نهایی تبدیل موجک توسط رابطه‌ی (۱۰) به دست می‌آید [۲۰]:

$$f^W(x, y) \quad (10)$$

$$= \frac{1}{\sqrt{MN}} \sum_{i=H.V.D} \sum_{j=j_0}^{\infty} \sum_n \sum_m W_{\psi}^i(j, m, n) \psi_{j, m, n}^i(x, y) - \frac{1}{\sqrt{MN}} \sum_n \sum_m W_{\varphi}(j_0, m, n) \varphi_{j_0, m, n}(x, y)$$

مناطق مورد نظر می‌تواند به صورت نقاط پراکنده توسط $f^W(x; y)$ با چند مکان پس‌زمینه علامت‌گذاری شوند. این نقاط می‌توانند به صورت نقاط دانه‌ی^۱ الگوریتم منطقه‌ی رشد در نظر گرفته شوند. با توجه به آن که هنوز نقاط دانه‌ی اشتباه وجود دارد، برای دسترسی به نقاط علامت‌گذاری سلول واقعی، از فرسایش مورفولوژیکی با شعاع R عنصر سازنده استفاده می‌شود. مرحله‌ی بعدی، انتخاب معیار برای منطقه‌ی رشد است. در این پژوهش از دو معیار برای یک پیسکل جهت قرار گرفتن در منطقه‌ی رشد استفاده شده است. در اولین معیار باید تفاوت سطح سیاه و سفید مطلق بین پیسکل مجاور و دانه، کم‌تر از آستانه‌ی^۲ T باشد و در معیار دوم پیکسلی که باید در یکی از مناطق در نظر گرفته

³ 8-connected

⁴ Voting

¹ Seed

² Threshold

(f^{ove}) و بقیه به صورت تابع خروجی (f^{out}) ذخیره می‌شوند.

برای جدا کردن سلول‌های همپوشانی، از عملیات مورفولوژی پیشرفته استفاده شده است. دو نوع از اپراتورهای اصلی مورفولوژی، فرسایش^۱ و افزایش^۲ می‌باشند؛ منظور از عملیات افزایش عملیاتی است که باعث افزایش ابعاد اجزای داخل تصویر به اندازه‌ی عنصر افزایش (یک یا چند پیکسل) می‌باشد و عملیات فرسایش دقیقاً برعکس عملیات افزایش است که در این عملیات معمولاً نقاط ناخواسته‌ی تصویر باینری حذف می‌شوند و سایر اجزای تصویر نیز به اندازه‌ی یک یا چند پیکسل نازک‌تر خواهند شد. این عملیات، تابع $\{0; 1\} \rightarrow I = u$ را به تصویر باینری و تابع $\{1\} \rightarrow I^{-1} = u$ را به پیکسل‌های پس زمینه تبدیل می‌کند. فرسایش و افزایش تصویر I باینری توسط عنصر ساختار $S \in Z \times Z$ به عنوان رابطه‌ی (۱۲) تعریف شده‌اند:

فرسایش: (۱۲)

$$Uf - S = \{(X|\forall_s) \in S. X + s \in Uf\}$$

افزایش

$$Uf + S = \{X + (s|x) \in I \wedge s \in S\}$$

با توجه به آن که سلول‌ها، محدب و مدور هستند، در طول مرزهای همپوشان سلول‌ها گوشه‌های مقعر وجود دارد. عملیات فرسایش مورفولوژی می‌تواند سلول‌های همپوشان را توسط گوشه‌های مقعر جدا و عملیات افزایش مورفولوژی می‌تواند سلول‌های فرسوده را بازسازی کند. عملیات مورفولوژی سنتی، از اندازه‌ی عنصر ساختار یکسان بر روی تمام سلول‌ها استفاده می‌کند که منجر به نتایج نامطلوب برای سلول‌هایی می‌شود که درجه‌ی همپوشانی متفاوتی دارند. یک عملیات مورفولوژی بهبود یافته برای سلول‌های همپوشان پیشنهاد می‌شود؛ روش پیشنهادی طبق مراحل زیر، اندازه‌ی مناسب عنصر ساختار را برای تفکیک سلول‌های همپوشان دارای اعوجاج کم‌تر انتخاب می‌کند.

مرحله‌ی ۱- شعاع ماکزیمم و شعاع فرسایش اولیه‌ی عنصر ساختار برحسب شعاع بزرگ‌ترین منطقه در تابع خروجی f^{out} تعیین می‌شود.

مرحله‌ی ۲- عملیات فرسایش با عنصر ساختار و شعاع فرسایش ($Kero^3$) بر روی منطقه‌ی تابع همپوشان اعمال

نام رای‌گیری استفاده شده است. این الگوریتم از تلفیق الگوریتم‌های موجود و رای گرفتن از تمامی آنها اقدام به جمع‌آوری رای‌ها کرده و رای خود را مطابق با رای اکثریت تعیین می‌کند [۲۱]؛ که مراحل این الگوریتم در زیر توضیح داده شده است.

مرحله‌ی ۱- دایره بودن و محیط منطقه در هر مقیاس طبق رابطه‌ی (۱۱) محاسبه می‌شود:

$$e = \frac{4\pi \times S_i}{C_j \times C_i} \quad (11)$$

مرحله‌ی ۲- مناطقی که محیط، بزرگ‌تر از 100 پیکسل برای هر مقیاس باشد به ترتیب به صورت ماتریس تابع مقیاس ۱، تابع مقیاس ۲ و تابع مقیاس ۳ ذخیره می‌شود. مرحله‌ی ۳- یکی از مناطق در تابع مقیاس ۱ انتخاب و با رشد منطقه علامت‌گذاری می‌شود. همان مکان در سه مقیاس توسط رشد مناطق ۲ و ۳ علامت‌گذاری می‌شود.

بزرگ‌ترین دایره سه منطقه را به عنوان بهترین منطقه انتخاب کرده و آن به صورت تابع مناطق مورد نظر f^{ROIS} ذخیره می‌شود. سه منطقه در ماتریس متناظر حذف خواهد شد.

مرحله‌ی ۴- تکرار مراحل ۲ و ۳ تا زمانی که توابع سه مقیاس خالی شوند ادامه می‌یابد [۲۱].

۳-۳- جداکردن سلول‌های همپوشانی با استفاده از عملیات مورفولوژی پیشرفته

بیشتر سلول‌های جدا شده، در بخش قبلی استخراج شده‌اند و مناطق باقی مانده، از مناطق همپوشانی و غیرسلولی تشکیل شده است. سلول‌های جدا شده و مناطق همپوشانی، طبق مراحل زیر می‌توانند از یکدیگر جدا شوند: مرحله‌ی ۱- دایره‌ای بودن هر منطقه در تابع مناطق مورد نظر (f^{ROIS}) محاسبه می‌شود.

مرحله‌ی ۲- مناطقی که دایره‌ی آن‌ها بزرگ‌تر از 0.7 است به صورت ماتریس تابع خروجی (f^{out}) و سایر مناطق به صورت مناطق قطعه‌بندی نشده ذخیره می‌شوند.

مرحله‌ی ۳- متوسط محیط مناطق در تابع خروجی (f^{out}) به صورت محیط اولیه‌ی آستانه‌ی (AG) محاسبه می‌شود.

مرحله‌ی ۴- محیط هر منطقه در تابع قطعه‌بندی نشده (f^{usg}) با AG مقایسه می‌شود و مناطقی که محیط آن‌ها بزرگ‌تر از $AG*1.5$ است به صورت ماتریس همپوشان

³ K-Erosion

¹ Erosion

² Dilation

شده است و ماتریس فرسایش یافته، بدست می‌آید. مرحله ۳- اگر تعداد مناطق متصل افزایش یابد، باید به مرحله ۴ رجوع کرد؛ در غیر این صورت $k_{ero} = k_{ero} + 2$ قرار داده می‌شود و به مرحله ۲ باز می‌گردد. مرحله ۴- مناطق تابع فرسایش با استفاده از عملیات فرسایش مورفولوژی و عنصر ساختار و شعاع فرسایش بازسازی می‌شود و مناطق بازسازی شده به صورت تابع خروجی f^{out} ذخیره خواهد شد.

۳-۴- استخراج ویژگی‌های سلول

تفاوت‌های خاصی در مشخصات ظاهری بین سلول‌های نرمال و سرطانی وجود دارد؛ چهار ویژگی شکلی از جمله مساحت^۱، محیط^۲، گریز از مرکز و دایره‌ای بودن آن‌ها محاسبه می‌شود. اطلاعات کامل برای این ویژگی‌ها به شرح زیر است:

مساحت: این ویژگی با محاسبه‌ی تعداد پیکسل‌های داخل منطقه‌ی سلول سنجیده می‌شود.

محیط: تعداد کل پیکسل‌های لبه‌ی سلول را محیط تشکیل می‌دهد.

گریز از مرکز: گریز از مرکز، بیضی بودن سلول و گشتاور^۳ دوم، منطقه‌ی سلول را مشخص می‌کند. مقادیر این ویژگی، بین ۰ و ۱ متغیر و شکلی که گریز از مرکز آن ۰ باشد، دایره و اگر ۱ باشد، خط است.

دایره‌ای بودن: این ویژگی، با X مساحت 4π / محیط به توان ۲ تعریف می‌شود که تشابه بین منطقه‌ی سلول و یک دایره را نشان می‌دهد؛ مقادیر بین ۰ و ۱ متغیر هستند و دایره‌ای بودن یک دایره برابر با ۱ است.

۳-۴-۱- استخراج ویژگی‌های بافت

در استخراج ویژگی‌های بافت، روش‌های آماری می‌توانند برای تحلیل توزیع فضایی مقادیر خاکستری با محاسبه‌ی ویژگی‌های محلی در هر نقطه از تصویر استفاده شوند و مجموعه‌ای از ویژگی‌های آماری را از توزیع ویژگی‌های محلی استخراج کنند. می‌توان چیدمان فضایی سطوح خاکستری پیکسل را در یک تصویر میکروسکوپی دیجیتال شرح داد. اولین ویژگی‌های آماری، میانگین و انحراف معیار هیستوگرام سطح سیاه و سفید هستند. فرض شود که Z یک متغیر تصادفی برای سطوح سیاه و سفید و $i =$

۳-۵- طبقه‌بندی با استفاده از الگوریتم SVM با کرنل RBF

ماشین‌بردار پشتیبان (SVM) یک الگوریتم طبقه‌بندی می‌باشد و به عنوان یکی از بهترین تکنیک‌های دسته‌بندی، شناخته و برخلاف الگوریتم‌های خوشه‌بندی در دسته‌ی یادگیری با نظارت قرار می‌گیرد؛ که شامل دو فاز آموزش و تست و دو حالت خطی و غیرخطی می‌باشد. داده‌ها در الگوریتم ماشین‌بردار پشتیبان به دو کلاس با استفاده از ابرصفحه تقسیم می‌شوند؛ به این داده‌ها برچسب مثبت یک و منفی یک زده می‌شود که آن‌ها را داده‌های آموزشی می‌نامند. سپس داده‌های تست در دو کلاس ایجاد می‌شوند تا نتیجه‌ی خوبی به وجود آید. هدف (SVM) پیدا کردن ابرصفحه‌ی منحصر به فردی است که برای آن حاشیه‌ی جداسازی، ماکزیمم باشد. می‌توان داده‌ها را با استفاده از نگاشت داده به یک فضای ویژگی انتقال داد و آن‌ها را بصورت خطی از یکدیگر جدا نمود؛ این فضای جدید باید دو شرط زیر را قبول کند: (۱) تبدیل، غیرخطی باشد (۲) ابعاد فضای ویژگی به اندازه‌ی کافی، بزرگ باشد. انجام محاسبات در فضای ویژگی می‌تواند، به دلیل ابعاد زیاد پرهزینه باشد اما در حالت کلی ابعاد این فضا بی‌نهایت است و برای غلبه بر این مشکل از کرنل استفاده می‌شود. کرنل-های بسیاری در الگوریتم (SVM) وجود دارد که سریع-ترین آموزش، مربوط به کرنل خطی و همچنین معروف‌ترین و پرکاربردترین کرنل، کرنل (RBF^4) می‌باشد. کرنل (RBF) منحنی‌های نرمال در اطراف نقاط داده را با هم جمع می‌کند تا بتواند یک مرز تصمیم‌گیری خوبی را بیان کند؛ با استفاده از کرنل (RBF)، فضای ویژگی از طریق

³ Moment

⁴ Radial Base Function

¹ Area

² Perimeter

بر اساس این مشاهده و مفهوم، دو ویژگی، مطلوباند و به صورت زیر توصیف می‌شوند: (۱) نمونه‌ها در یک کلاس باید در منطقه‌ی مشابهی در فضای ویژگی نگاشت شوند (۲) نمونه‌ها در کلاس‌های مختلف باید در منطقه‌های مختلفی نگاشت شوند.

یک پارامتر کامل σ باید طبق رابطه‌ی (۱۴ و ۱۵) یافت شود:

$$k(x_l^{(i)}; x_k^{(i)}; \sigma) \approx 1 \quad (14)$$

$$k(x_l^{(i)}; x_k^{(i)}; \sigma) \approx i \neq j. \quad (15)$$

در واقع، σ برای تعیین مقدار کرنل (RBF) است. اگر پارامتر σ نزدیک به ۰ باشد، مقادیر عملکرد مربوط به هسته تقریباً به ۰ (حد پایین عملکرد) هسته (RBF) نزدیک است؛ این بدان معنی است که تمام نمونه‌های موجود در یک فضای ویژگی تقریباً عمود بر هم هستند. با افزایش مقادیر، هسته (RBF) با توجه به نمونه‌هایی که با استفاده از فاصله اقلیدسی در فضای اصلی، نزدیکتر هستند، سریع عمل می‌کند. از آنجا که σ نزدیک به بی‌نهایت است، مقادیر عملکرد هسته مربوطه تقریباً نزدیک به ۱ و نمونه‌های موجود در فضای ویژگی نزدیک به یک نقطه ثابت هستند. دو معیار برای اندازه‌گیری قابلیت تفکیک‌پذیری در فضای ویژگی پیشنهاد شده است. (۱) میانگین مقادیر محاسبه شده توسط تابع هسته‌ای در نمونه‌هایی که کلاس مشابه دارند به صورت رابطه‌ی (۱۶) می‌باشد.

$$\omega(\sigma) = \frac{1}{\sum_{i=1}^L N_i^2} \sum_{i=1}^L \sum_{l=1}^L \quad (16)$$

$$\sum_{k=1}^L k(x_l^{(i)}; x_k^{(i)}; \sigma)$$

N_i تعداد نمونه‌های آموزشی در کلاس i است و پارامتر σ باید به گونه‌ای تعیین شود که $\omega(\sigma)$ نزدیک به ۱ و کمتر از ۱ باشد (۲) میانگین مقادیر محاسبه شده توسط تابع هسته‌ای (RBF) در نمونه‌هایی که کلاس مختلف دارند به صورت رابطه‌ی (۱۷) می‌باشد:

یک تبدیل غیرخطی به دست می‌آید.

در این پژوهش از یک الگوریتم طبقه‌بندی جدید، مبتنی بر ماشین‌بردار پشتیبان با کرنل (RBF) استفاده شده که در این بخش، این الگوریتم شرح داده شده است.

یک روش اتوماتیک برای انتخاب پارامتر بخش مرکزی کرنل (RBF) (که با APR^۲ اشاره شده است) برای تسریع زمان آموزش (SVM) حاشیه‌ی نرم^۳ استفاده شده است [۲۲]. این روش با موفقیت، پارامترهای هسته‌ی اصلی مناسبی را برای (SVM) انتخاب و نسبت به روش اعتبار سنجی متقابل (ضربدری) (k-fold) زمان کم‌تری را صرف می‌کند. در ادامه مفهوم این مدل ترسیم می‌گردد:

فرض شود که $\{x_j^{(i)}\}_{j=1; \dots; N_i} \subset R^d$ مجموعه نمونه‌های آموزشی در کلاس i هستند که N_i تعداد نمونه‌های آموزشی در کلاس i و i تابع $i = 1; 2; \dots; L$ و L تعداد کلاس‌ها می‌باشد. هسته (RBF)، یک تابع هسته‌ی محبوب است که در الگوریتم‌های یادگیری مختلف هسته‌ای استفاده و به‌طور خاص، در دسته‌بندی ماشین‌بردار پشتیبانی مورد استفاده قرار می‌گیرد. هسته (RBF) در دو نمونه x و x' ، به عنوان بردارهای ویژگی در برخی از فضای ورودی، به صورت رابطه‌ی (۱۳) تعریف می‌شود:

$$k(x; x'; \sigma) = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (13)$$

$\|x - x'\|^2$ به عنوان فاصله‌ی اقلیدسی مربع بین دو بردار ویژگی، تعریف شده است زیرا مقدار هسته‌ی (RBF) دارای فاصله می‌باشد که مقدار آن، بین صفر (در حد) و یک (هنگامی که $x = x'$) کاهش می‌یابد. فضای ویژگی هسته تعداد بی شماری ابعاد دارد و $x; x' \in R^d$ و $\sigma \in R -$ و $\{0\}$ پارامترهای متناظر می‌باشند. تابع هسته‌ای^۵ (RBF) $k(x_\rho^{(i)}; x_\rho^{(j)}; \sigma) = 1$ (۱) دو ویژگی دارد: (۱) یعنی هنجار^۶ هر نمونه در فضای ویژگی^۷ ۱ است (۲) $0 < x_\rho^{(i)}; x_\rho^{(i)}; \sigma \leq 1$ (یعنی مقدار کسینوس دو نمونه‌ی آموزشی $x_\rho^{(j)}$ و $x_\rho^{(i)}$ در فضای ویژگی، از صفر بیشتر و کم‌تر مساوی ۱ است که شباهت بین این دو نمونه را تعیین می‌کند).

⁵ Kernel Function

⁶ Norm

⁷ Feature Space

¹ Kernel

² Annual Percentage Rate

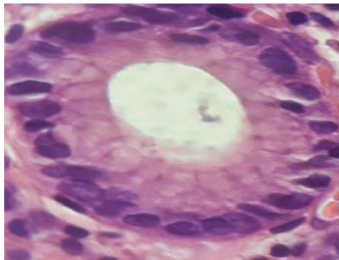
³ Soft Margin SVM (Support Vector Machine)

⁴ K-fold Cross Validation

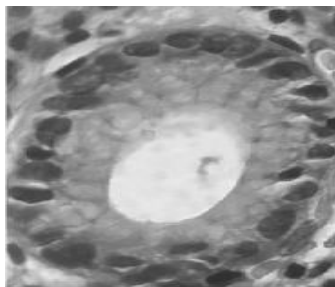
۲-۴- قطعه‌بندی مناطق سلولی

در این مرحله برای جداسازی سلول‌ها از پس‌زمینه از روش خوشه‌بندی فازی (FCM) استفاده شده است.

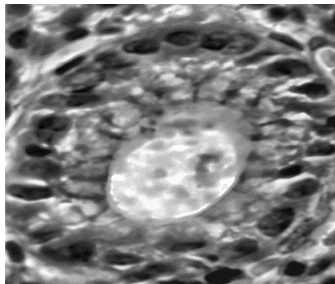
در خوشه‌بندی فازی دو نوع پارامتر ورودی و خروجی مورد بررسی قرار می‌گیرد؛ پارامترهای ورودی شامل تعداد خوشه‌ها، میزان خطا و تعداد تکرارها می‌باشد؛ که در شبیه‌سازی، تعداد خوشه‌ها اهمیت بسیاری دارد بنابراین در این پژوهش، مطابق شکل (۳) تعداد ۲ خوشه به عنوان پس‌زمینه و سلول تعیین شده که علت آن حاصل شدن نتایج بهتر می‌باشد. پارامترهای خطا و تکرار هم نقش به‌سزایی در این پژوهش دارند که نشان می‌دهند هر چقدر تعداد تکرارها بالا رود زمان برنامه افزایش و خطاها کاهش می‌یابد.



(الف)



(ب)



(ج)

شکل ۲: مراحل تصویر بهبود یافته با استفاده از الگوریتم کلاه بالا به پایین

الف) تصویر رنگی بافت معده، ب) تبدیل تصویر رنگی به تصویر خاکستری، ج) تصویر بهبود یافته با استفاده از الگوریتم کلاه بالا به پایین

$$b(\sigma) = \frac{1}{\sum_{i=1}^L \sum_{j=1, j \neq i}^L N_i N_j} \sum_{i=1}^L \sum_{j=1, j \neq i}^L \sum_{l=1}^L \sum_{k=1}^L k \left(x_l^{(i)} \cdot x_k^{(j)} \cdot \sigma \right) \quad (17)$$

بنابراین σ باید به گونه‌ای تعیین شود که $b(\sigma)$ به ۱ نزدیک شود. در نهایت برای روش طبقه‌بندی (SVM) از کرنل (RBF) استفاده و در بخش بعدی نتایج آن با کرنل خطی و بدون کرنل مقایسه می‌شود و با استفاده از منحنی (ROC^۱) دقت نهایی مورد ارزیابی قرار می‌گیرد.

۴- بررسی نتایج

در این بخش، مطابق روش پیشنهادی بخش قبلی، به شرح نتایج شبیه‌سازی پرداخته خواهد شد. در ابتدا از روش کلاه بالا به پایین جهت بالا بردن کیفیت تصویر، استفاده و سپس موضوع استخراج سلول (قطعه‌بندی مناطق) با روش‌های خوشه‌بندی فازی، تبدیل موجک گسسته، رشد منطقه و مکانیزم رای‌گیری به صورت جداگانه مورد بررسی قرار می‌گیرد. سپس سلول‌های هم‌پوشانی با روش مورفولوژی پیشرفته، جداسازی و در مرحله‌ی بعدی ویژگی سلول‌ها استخراج می‌شود و به الگوریتم ماشین‌بردار پشتیبان با کرنل (RBF) برای طبقه‌بندی سلول‌های سالم و سرطانی، آموزش داده خواهد شد.

۱-۴- پیش‌پردازش تصویر

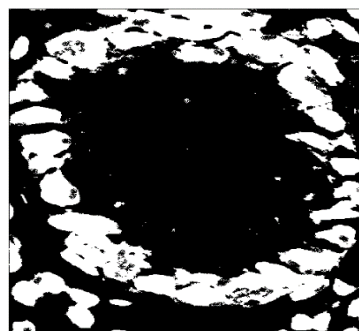
مطابق شکل (الف-۲) تصویر رنگی، سه‌بعدی و ابعاد آن نسبتاً بزرگ می‌باشد به همین جهت مدت زمان محاسبات به طول می‌انجامد؛ بنابراین مطابق شکل (ب-۲) تصویر رنگی را خاکستری کرده، تا حجم محاسباتی کاهش و سرعت آن بالا رود. برای یکنواختی و افزایش کنتراست بین سلول (مناطق تیره) و پس‌زمینه (مناطق روشن)، از یکی از عملیات‌های مورفولوژی به نام کلاه بالا به پایین استفاده می‌شود [۲۲]. با توجه به اینکه در تصویر سلولی، اشیای کوچک با پس‌زمینه‌ی تاریک وجود دارد، تبدیل کلاه بالا (T_{hat}) می‌تواند مناطق روشن تصویر را با عملیات گشودن استخراج کند؛ حال آن که تبدیل کلاه پایین (B_{hat}) برعکس آن عمل می‌کند [۲۳]. در نتیجه با توجه به شکل (ج-۲) مشاهده می‌شود که تصویر، درصد وضوح بالاتری را نسبت به قبل نشان می‌دهد.

^۱ Receiver Operating Characteristic

مد نظر نیستند؛ بنابراین برای تقویت مناطق مورد نظر سلولی توسط ترکیب روش ذکر شده، چند مکان پس‌زمینه علامت‌گذاری و این مناطق با روش الگوریتم منطقه رشد، گسترش داده می‌شوند. هدف از انجام این روش، گسترش نقاط پیشنهاد شده می‌باشد تا سطح هر سلول را در هر مقیاس تعیین کند. برای شروع، تابع اولیه با دو ورودی و یک خروجی تعریف می‌شود. ورودی اول، تصویر بدست آمده از روش تبدیل موجک گسسته و خوشه‌بندی فازی می‌باشد که نقاطی از آن به صورت اولیه علامت‌گذاری و نقاط پیش‌فرض، باعث رشد منطقه و در مقیاس‌های سه‌گانه تعریف شده است. ورودی دوم، بررسی می‌کند که آیا برنامه، در هر مرحله، نتایج را نشان می‌دهد یا خیر. برای رشد نواحی، ابتدا مناطق مورد نظر تعیین و تعداد نواحی برابر با ۲۰۰ و بیشترین آستانه ۱۰ در نظر گرفته و با استفاده از روش منطقه‌ی تکراری، آستانه‌ی مورد نظر محاسبه می‌شود که با کم‌ترین آستانه شروع و تا زمانی که تشابه اندازه و مناطق خاکستری تیز شوند، پایان می‌یابد. پس از این مراحل، منطقه‌ی اولیه انتخاب و پیکسل‌های موجود در منطقه بررسی می‌شوند و تا زمانی که مناطق سلولی گسترش یابند، این مرحله ادامه می‌یابد. هر پیکسل موجود در این منطقه بررسی می‌شود؛ که آیا پیکسل‌های اطراف قابلیت وصل شدن به این دسته را دارند یا خیر که در صورت وصل شدن به این پیکسل‌ها، این منطقه گسترش می‌یابد.

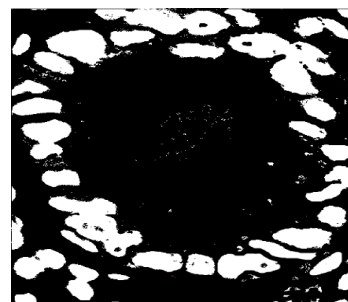
اگر تعداد پیکسل‌ها در این مرحله افزایش نداشت، مقدار فاصله‌ی بین آن‌ها تا یک واحد افزایش و دوباره برنامه اجرا می‌شود؛ این روال تا زمانی که تعداد پیکسل‌ها به حد نصاب و مقدار آستانه به حداکثر خود برسد ادامه می‌یابد. روش رشد منطقه در سه مقیاس بررسی می‌شود؛ که در آن، محیط و ویژگی‌های هر منطقه‌ی سلولی محاسبه و سپس مناطقی که کمتر از ۳۰۰ پیکسل باشد حذف می‌گردد. مقیاس ۱، طبق شکل (۵-الف) می‌تواند سطوح خاکستری داخلی ناهموار و لبه‌های ناپیوسته‌ی درون هسته‌ی سلول را پردازش کند، اما محیط هسته‌های همپوشان با یکدیگر جمع می‌شود. مقیاس ۳، طبق شکل (۵-ج) می‌تواند محیط چسبندگی را به درستی تفکیک کند، اما مشکلات سطوح خاکستری داخلی ناهموار و لبه‌های ناپیوسته شدیدتر هستند. مقیاس ۲، طبق شکل (۵-ب) جبران ناکارآمدی مقیاس ۱ و ۳ است.

بنابراین به روش سعی و خطا میزان خطاها را ۰,۰۰۱ و تعداد تکرارها ۶۶ در نظر گرفته شده است. پارامترهای خروجی شامل مرکز خوشه‌ها و تصویر خوشه‌بندی می‌باشد. در این مرحله، برای آماده‌سازی بردار داده‌ها، تصویر، با یک پنجره‌ی ۳*۳ فیلتر و به صورت بردار، نمایش داده و سپس مقدار اولیه برای مراکز خوشه‌ها و میزان هزینه‌ی کل محاسبه و یک حلقه‌ی تکرار به تعداد تکرارهای مجاز اضافه و تا زمانی که میزان خطا تا حد تعیین شده پایین نیاید این حلقه ادامه می‌یابد و هنگامی که به خطای مورد نظر برسد، الگوریتم خاتمه پیدا می‌کند. پس از این مراحل مقدار جدیدی برای مراکز خوشه تولید و متناظر با آن، میزان هزینه‌ی جدید محاسبه می‌شود که متعاقباً می‌توان میزان خطا را به دست آورد. در هر مرحله، برنامه، یک خروجی تولید می‌کند که نشان دهنده‌ی میزان تغییرات در مقدار هزینه است.



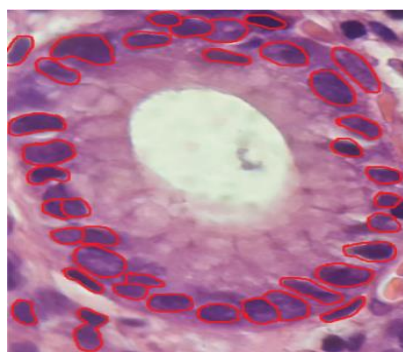
شکل ۳: تصویر خروجی حاصل از قطعه‌بندی فازی (FCM) که شامل ۲ خوشه است؛ رنگ سفید مربوط به سلول‌ها و رنگ سیاه مربوط به پس‌زمینه می‌باشد.

برای بالا بردن دقت قطعه‌بندی مناطق سلولی، دو روش خوشه‌بندی فازی و تبدیل موجک گسسته، طبق روش پیشنهادی با هم ترکیب شده‌اند و نتایج ترکیب این دو روش در شکل (۴) نشان داده شده است.



شکل ۴: خروجی تصویر حاصل از قطعه‌بندی تبدیل موجک گسسته و خوشه‌بندی فازی (FCM)

بعد از ترکیب دو روش فوق هنوز نقاط سلولی، مشخص و



(الف)



(ب)

شکل ۶: خروجی مناطق مورد نظر با روش مکانیزم رای گیری (الف) تصویر اصلی (ب) تصویر خروجی مناطق سلولی

۳-۴- جدا کردن سلول‌های هم‌پوشانی

مناطق مورد نظر با استفاده از روش‌های بیان شده بدست آمده و باید مناطقی که در آن سلول‌های تکی و سلول‌های هم‌پوشانی وجود دارند را از هم تفکیک و به دو تصویر سلول‌های تکی مطابق شکل (۷-الف) و سلول‌های هم‌پوشانی مطابق شکل (۷-ب) تبدیل کرد. تابع با یک ورودی و یک خروجی تعریف می‌شود. ورودی این تابع، تصویر نهایی مرحله قبل و خروجی آن تصویر بهبود یافته‌ی مناطق سلولی می‌باشد.

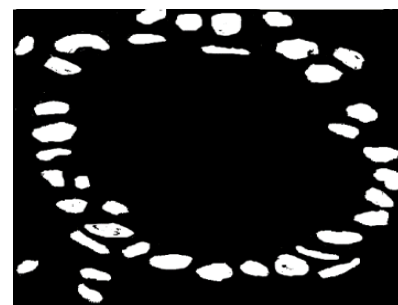
در این روش، ابتدا میزان دایره‌ای بودن هر منطقه سنجیده و مناطقی که دایره‌ی آن‌ها بزرگ‌تر از ۰,۷ باشد در خروجی قرار می‌گیرد و سایر مناطق به عنوان مناطق هم‌پوشانی ذخیره می‌شوند. مناطقی که دارای مساحت بیش از ۱,۵ باشد، به صورت سلول‌های هم‌پوشانی؛ و بقیه به صورت تابع خروجی محاسبه می‌شود که در شکل (۷) نشان داده شده است. حال باید مناطقی که سلول‌های هم‌پوشانی دارند را با استفاده از عملیات مورفولوژی افزایش و فرسایش از هم جدا کرد. در روند فوق مقدار فرسایش تا حد امکان برای مناطق سلولی انجام می‌شود؛ و اگر این فرسایش باعث چند تکه شدن ناحیه شود، برنامه خاتمه یافته و سراغ مناطق



(الف)



(ب)



(ج)

شکل ۵: بررسی رشد منطقه‌ی سلولی در سه مقیاس: (الف)

مقیاس ۱ (ب) مقیاس ۲ (ج) مقیاس ۳

در ادامه، از ترکیب ۳ مقیاس گفته شده، مناطق سلولی مورد نظر، به دست آمده اما برای استخراج مناطق سلولی دقیق‌تر و حذف مناطق غیرسلولی از روش مکانیزم رای-گیری استفاده می‌شود. در این روش دایره‌ای بودن و محیط سلول، در هر مقیاس محاسبه و مناطقی که بزرگ‌تر از ۱۰۰ پیکسل برای هر مقیاس باشد به ترتیب در ماتریس مقیاس ۱ و ۲ و ۳ ذخیره و یکی از مناطق در مقیاس ۱ انتخاب و همچنین در مقیاس ۲ و ۳ علامت‌گذاری می‌شود. سپس بزرگ‌ترین دایره را به عنوان بهترین منطقه انتخاب و به صورت تابع بهبود یافته از خروجی مناطق سلولی ذخیره و سه منطقه‌ی دیگر را حذف می‌کند. در انتها طبق شکل (۶-الف)، تصویر نهایی توسط خطوط قرمز، نواحی سلول‌ها را مشخص می‌کند. با توجه به شکل (۶-ب) نتایج حاصله، نشان دهنده‌ی این واقعیت است که این مرحله باعث حذف نویز شده و این عمل به روند طبقه‌بندی تصویر کمک شایانی کرده است.

۵- روش ارزیابی و منحنی ROC

ماتریس در هم ریختگی^۱ حاوی اطلاعاتی می باشد که از دو طبقه بندی واقعی و پیش بینی شده، توسط سیستم طبقه بندی انجام گرفته است. عملکرد این سیستم ها معمولاً با استفاده از داده های ورودی در ماتریس، ارزیابی می شوند. جدول ۱ ماتریس در هم ریختگی را برای طبقه بندی دو کلاس نشان می دهد [۲۴،۲۵].

جدول ۱: ماتریس در هم ریختگی [۲۵].

		Predicted	
		Positive	Negative
Actual	Positive	True Positives (TP)	False Negatives (FN)
	Negative	False Positives (FP)	True Negatives (TN)

عناصر ماتریس در هم ریختگی به شرح زیر است:
 TN^2 : بیانگر تعداد نمونه هایی است که دسته ی واقعی آن ها نمونه های سالم (منفی) بوده و الگوریتم طبقه بندی نیز دسته ی آن ها را به درستی، سالم (منفی) تشخیص داده است.

TP^3 : بیانگر تعداد نمونه هایی است که دسته ی واقعی آن ها سرطانی (مثبت) بوده و الگوریتم طبقه بندی نیز دسته ی آن ها را به درستی، سرطانی (مثبت) تشخیص داده است.

FP^4 : بیانگر تعداد نمونه هایی است که دسته ی واقعی آن ها سالم (منفی) بوده و الگوریتم طبقه بندی، دسته ی آن ها را به اشتباه، سرطانی (مثبت) تشخیص داده است.

FN^5 : بیانگر تعداد نمونه هایی است که دسته ی واقعی آن ها سرطانی (مثبت) بوده و الگوریتم طبقه بندی، دسته ی آن ها را به اشتباه، سالم (منفی) تشخیص داده است.

معیار مهم دیگری که برای تعیین میزان کارایی یک طبقه بندی استفاده می شود، معیار (AUC) است.

معیار (AUC) نشان دهنده ی ارزیابی سطح زیر منحنی (ROC)^۶ می باشد که هر چه مقدار این معیار مربوط به یک طبقه بندی بزرگ تر باشد کارایی نهایی، مطلوب تر ارزیابی می شود. منحنی ارزیابی (ROC) نمودارهای دوبعدی هستند که در آن ها (TPR)^۸ نرخ تشخیص صحیح دسته ی

بعدی می رود. شکل (۸) نتیجه ی جدا شدن سلول ها را نشان می دهد. در هر تکرار تعداد نواحی منتصب به سلول ها به روزرسانی شده و این روند برای تمامی مناطق دارای هم پوشانی انجام می شود. در انتها، برای کل مناطق، این دو بخش دوباره نمایش داده می شود تا بخش هایی که از هم پوشانی در آمده است دیده شود



(الف)



(ب)

شکل ۷: (الف) مناطق سلول های تکی. (ب) مناطق سلول های هم پوشانی.



شکل ۸: جدا کردن سلول های هم پوشانی با استفاده از روش عملیات مورفولوژی پیشرفته

⁵ False Negative

⁶ Area Under Curve

⁷ Receiver Operating Characteristic

⁸ True Positive Rate

¹ Confusion Matrix

² True Negatives

³ True Positive

⁴ False Positive

کلاس‌بندی، داده‌های بدون برچسب به الگوریتم داده می‌شود و آن به صورت اتوماتیک کلاس مربوطه را مشخص می‌کند. به دلیل حجم زیاد سلول‌های استخراج شده و به صورت تصادفی، با بررسی آزمون و خطا، میزان داده‌های آموزشی ۷۰ درصد و داده‌های آزمایشی ۳۰ درصد گرفته شده است تا بتوان درصد دقت نهایی را بالا برد و همینطور زمان ارزیابی را کاهش داد. معیارهایی مانند صحت، ویژگی، حساسیت و میانگین خطای ایجاد شده برای اعتبار سنجی عملکرد، مورد استفاده قرار می‌گیرد. نتایج میانگین مربع خطا برای آموزش ماشین بردار پشتیبان با کرنل (RBF) در ۱۰۰ دور تکرار، مورد بررسی قرار گرفته است. خطای آموزش بین ۰،۰۱۵ و ۰،۰۴۱ می‌باشد که نشان دهنده‌ی آموزش مناسب است. نتایج میانگین مربع خطا برای تست ماشین بردار پشتیبان با کرنل (RBF) در ۱۰۰ دور تکرار، مورد بررسی قرار گرفته است. میانگین مربع خطا در تکرار ۱۵ام دارای کمترین مقدار و برابر ۰ است و در تکرار ۴۵ام میانگین مربع خطا دارای بیشترین مقدار و برابر ۰،۰۵۷۰ می‌باشد که نشان دهنده‌ی عملکرد مناسب این الگوریتم می‌باشد. برای یافتن مناسب‌ترین کرنل برای الگوریتم ماشین بردار پشتیبان، آزمون‌های مختلف با تعداد تکرار ۳۰ صورت گرفته است. با توجه به نتایج، بهترین کرنل از بین کرنل‌های مختلف، کرنل (RBF) می‌باشد و مقدار بهترین پارامتر σ در بین این $\{2^{16} \dots 2^7\}$ مجموعه داده انتخاب می‌شود. الگوریتم ماشین بردار پشتیبان با کرنل (RBF) با $\sigma=15$ در نتایج میانگین مربع خطای داده‌های تست، کم‌ترین مقدار ۰،۱۸-۰،۲۲ را دارد. اما در سیگماهایی با مقادیر ۵، ۱۰ و ۱۶ خطا بیشتر بوده است؛ پس به عنوان مناسب‌ترین کرنل با سیگمای ۱۵ به کار می‌رود.

نتایج روش ارائه شده جهت دسته‌بندی سلول‌های سرطانی به دو دسته‌ی خوش‌خیم و بدخیم ارزیابی شده است. کل سلول‌های استخراج شده از داده‌ی آزمایشی، ۵۳۱ سلول می‌باشد که ۲۴۴ سلول سرطانی در دسته‌ی خوش‌خیم (منفی) و ۲۸۷ سلول سرطانی در دسته‌ی بدخیم (مثبت) دسته‌بندی شده‌اند. مقدار دقت کلی در روش پیشنهادی برابر با ۹۴،۳۵ است. نتیجه‌ی ماتریس در هم ریختگی در اعمال الگوریتم (SVM) با کرنل (RBF) توسط روش‌های

مثبت روی محور Y و (FPR)^۱ نرخ تشخیص غلط دسته‌ی مثبت روی محور X را نشان می‌دهد. نقاط متعددی در فضای (ROC) قابل توجه است که بطور کلی نقطه‌ای که در شمال غربی‌تر این فضا قرار گرفته باشد، فضای (ROC) بهترین دقت را دارد؛ همچنین منحنی‌های ارزیابی، رفتار یک طبقه‌بندی کننده را بدون توجه به توزیع دسته‌ها یا هزینه‌ی خطا نشان می‌دهند، بنابراین کارایی طبقه‌بندی را از این عوامل جدا می‌کنند.

۱-۵- معرفی پایگاه داده پژوهش

دیتای مورد استفاده در جهت ارزیابی مدل ارائه شده شامل ۹۶ عدد تصاویر دیجیتالی پاتولوژی بافت معده‌ی مربوط به افراد بیمار و سالم می‌باشد که از بیمارستان بقیه‌الله به کمک یکی از متخصصین پاتولوژیست با استفاده از دوربین دیجیتالی با مدل سامسونگ S5 تهیه گردیده است. تصاویر گرفته شده با استفاده از میکروسکوپ نوری مدل (optika b-500) با بزرگنمایی‌های ۲۰x، ۴۰x و ۱۰۰x بدست آمده است. نمونه‌های بیوپسی بافت معده از بیماران گرفته شده؛ شامل ۳۲ عدد تصویر سالم دیجیتالی، ۳۱ عدد تصویر نیمه-سرطانی (آدنوم) دیجیتالی و ۳۳ عدد تصویر سرطانی (آدنوکارسینوم) دیجیتالی از بافت معده می‌باشد.

۲-۵- تحلیل و بررسی نتایج مدل هوشمند

پیشنهادی

در ابتدا نمونه‌های بافت معده که شامل ۳۲ نمونه سالم، ۳۱ نمونه سرطانی خوش‌خیم و ۳۳ نمونه سرطانی بدخیم از بافت معده می‌باشد به صورت جداگانه به روش پیشنهادی که در بخش‌های قبلی شرح داده شده است، اعمال و سپس ۱۰ ویژگی بافتی و ویژگی‌های سلولی از کل مجموعه داده استخراج و در نهایت تمام اطلاعات استخراج شده از سلول‌ها با هم جمع می‌شوند. تعداد کل سلول‌های استخراج شده از روش قطعه‌بندی از ۹۶ نمونه‌ی بافت معده شامل ۲۷۴۹ سلول که ۹۸۱ سلول سالم، ۸۲۷ سلول سرطانی خوش‌خیم و ۹۴۱ سلول سرطانی بدخیم است. آموزش در روش (SVM) به این معنی است که تعداد ویژگی‌های مربوط به سلول‌های سالم، خوش‌خیم و بدخیم که از روش قطعه‌بندی بدست آمده است به تابع داده می‌شود و الگوریتم بر اساس برچسب‌گذاری‌های صورت گرفته پارامترهای خود را به روز رسانی می‌کند؛ سپس برای

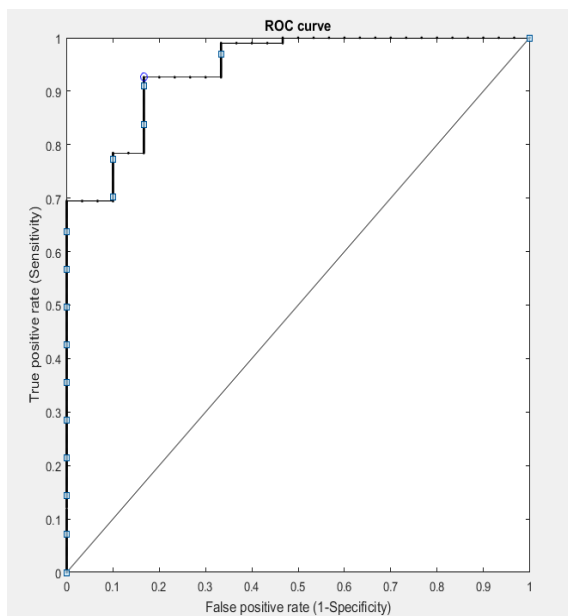
¹ False Positive Rate

دسته‌ی سالم (منفی) و دسته‌ی سرطانی (مثبت) در جدول ۳ نشان داده شده است.

جدول ۳: نتایج ماتریس در هم ریختگی با استفاده از روش پیشنهادی برای دسته‌ی سالم (منفی) و دسته‌ی سرطانی

		Predicted	
		Positive	Negative
Actual	Positive	True Positives (TP)=271	False Negatives (FN)=38
	Negative	False Positives (FP)=27	True Negatives (TN)=489

نتایج حاصل از ارزیابی روش ارائه شده جهت طبقه‌بندی سلول‌های سرطانی به دو دسته‌ی سالم و سرطانی با کمک نمودار (ROC) در شکل (۱۰) نشان داده شده است و در ارزیابی صورت گرفته، سطح زیر نمودار بیانگر دقت ۹۲,۴۸ می‌باشد.



شکل ۱۰: نمودار (ROC) با استفاده از روش پیشنهادی در دسته‌ی سرطانی (مثبت) و دسته‌ی سالم (منفی)

۶-مقایسه مدل‌های هوشمند پیشنهادی

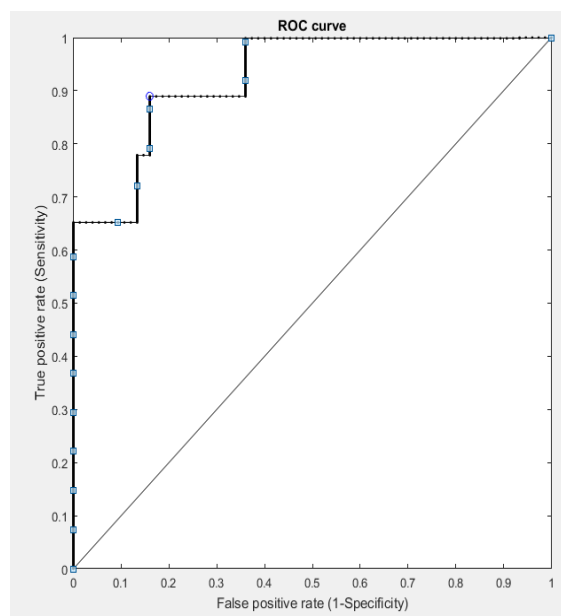
با توجه به جدول ۴ در تصاویر حاصل از اعمال الگوریتم (SVM) با کرنل خطی توسط روش‌های قطعه‌بندی در حالت‌های سالم و سرطانی، میزان کارایی سیستم، مقدار نسبتاً مناسبی را نشان داده است که زمان اجرای این روش ۳۲ ثانیه می‌باشد (زمان مناسبی می‌باشد). اما مقدار زمان اجرای تصاویر حاصل از اعمال الگوریتم (SVM) توسط

قطعه‌بندی برای دسته‌ی خوش خیم (منفی) و دسته‌ی بدخیم (مثبت) در جدول ۲ نشان داده شده است.

جدول ۲: نتایج ماتریس در هم ریختگی با استفاده از روش پیشنهادی برای دسته‌ی خوش خیم (منفی) و دسته‌ی بدخیم

		Predicted	
		Positive	Negative
Actual	Positive	True Positives (TP)=225	False Negatives (FN)=19
	Negative	False Positives (FP)=11	True Negatives (TN)=276

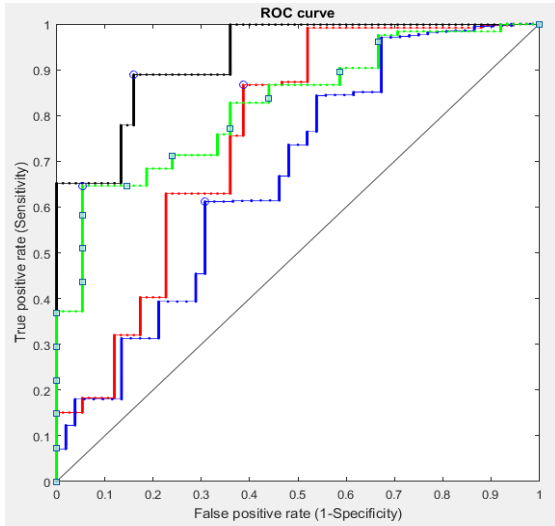
نتایج حاصل از ارزیابی روش ارائه شده جهت دسته‌بندی سلول‌های سرطانی به دو دسته‌ی بدخیم و خوش خیم با کمک نمودار (ROC) در شکل (۹) نشان داده شده است و در ارزیابی صورت گرفته، سطح زیر نمودار بیانگر دقت ۹۴,۱۴ می‌باشد.



شکل ۹: نمودار (ROC) با استفاده از روش پیشنهادی در دسته‌ی بدخیم (مثبت) و دسته‌ی خوش خیم (منفی)

نتایج روش ارائه شده جهت طبقه‌بندی سلول‌های سرطانی به دو دسته‌ی سالم و سرطانی ارزیابی شده است. کل سلول‌های استخراج شده از داده‌ی آزمایشی، ۸۲۵ سلول می‌باشد که ۳۰۹ سلول سالم (منفی) و ۵۱۶ سلول سرطانی (مثبت) طبقه‌بندی شده‌اند. مقدار دقت کلی برابر با ۹۲,۱۲ است. نتیجه‌ی ماتریس در هم ریختگی از اعمال الگوریتم (SVM) با کرنل (RBF) توسط روش‌های قطعه‌بندی برای

هر ۴ حالت (سالم، سرطانی، خوش خیم و بدخیم) ۶۱ ثانیه می‌باشد که میزان کارایی سیستم، مقدار و زمان مناسبی را نشان نمی‌دهد. بررسی‌ها نشان می‌دهد که روش پیشنهادی (الگوریتم (SVM) با کرنل (RBF) توسط روش‌های قطعه-بندی) نسبت به سایر روش‌ها عملکرد بسیار مناسبی، داشته و همچنین مدت زمان شبیه‌سازی ۲۰ ثانیه بوده است.



- طبقه‌بندی SVM با کرنل RBF توسط قطعه‌بندی FCM
- طبقه‌بندی SVM توسط روش‌های قطعه‌بندی
- طبقه‌بندی SVM با کرنل خطی توسط روش‌های قطعه‌بندی
- روش پیشنهادی

شکل ۱۱: نتایج مقایسه‌ی منحنی (ROC) سه روش با روش پیشنهادی در نمونه‌های سالم و سرطانی

جدول ۵: نتایج مقایسه‌ی سه روش با روش پیشنهادی در

نمونه‌های خوش خیم و بدخیم

نمونه‌های خوش خیم و بد خیم				
AUC %	specificity %	sensitivity %	Accuracy %	روش‌ها
۶۹,۱۶	۷۴,۱۳	۶۵,۵۵	۶۹,۳۰	طبقه‌بندی SVM توسط روش‌های قطعه-بندی
۸۸,۰۶	۸۲,۰۶۴	۹۰,۷۰	۸۷	طبقه‌بندی SVM با کرنل خطی توسط روش‌های قطعه‌بندی
۵۸,۱۹	۶۴,۲۳	۴۹,۷۸	۵۸	طبقه‌بندی SVM با کرنل RBF توسط قطعه‌بندی FCM
۹۴,۱۴	۹۶,۱۶	۹۲,۲۱	۹۴,۳۵	روش پیشنهادی

روش‌های قطعه‌بندی، ۴۶ ثانیه و زمان اجرای تصاویر حاصل از الگوریتم (SVM) توسط الگوریتم فازی (FCM) با کرنل (RBF) ۵۳ ثانیه می‌باشد که میزان کارایی سیستم، مقدار و زمان مناسبی را نشان نمی‌دهد. بررسی‌ها نشان می‌دهد که روش پیشنهادی (الگوریتم (SVM) با کرنل (RBF) توسط روش‌های قطعه‌بندی) نسبت به سایر روش‌ها عملکرد بسیار مناسبی، داشته و همچنین مدت زمان شبیه‌سازی ۲۵ ثانیه بوده که این امر موجب بهبود عملکرد سیستم در زمان اعمال تصاویر با تعداد زیاد می‌شود.

جدول ۴: نتایج مقایسه‌ی سه روش با روش پیشنهادی در

نمونه‌های سالم و سرطانی

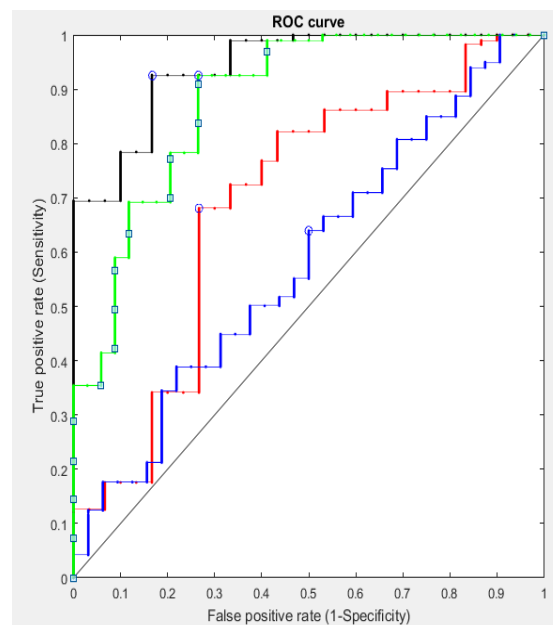
نمونه‌های سالم و سرطانی				
AUC %	specificity %	sensitivity %	Accuracy %	روش‌ها
۷۵,۵۶	۷۰,۴۲	۸۰,۰۴	۷۵,۴۰	طبقه‌بندی SVM توسط روش‌های قطعه‌بندی
۸۲,۳۹	۸۵,۳۱	۸۱,۱۵	۸۲,۴۲	طبقه‌بندی SVM با کرنل خطی توسط روش‌های قطعه-بندی
۶۷,۸	۶۶,۸۲	۶۷,۲۵	۶۷,۰۳	طبقه‌بندی SVM با کرنل RBF توسط قطعه‌بندی FCM
۹۲,۴۸	۹۴,۷۴	۸۷,۷۰	۹۲,۱۲	روش پیشنهادی

برای نشان دادن دقت و کارایی مدل پیشنهادی در مقایسه با مدل‌های دیگر، نمودار (ROC) آنها در شکل (۱۱) نشان داده شده و همانطور که مشاهده می‌شود، روش پیشنهادی نسبت به سایر روش‌های دیگر در هر ۴ حالت (سالم، سرطانی، خوش خیم و بدخیم) از مساحت زیر نمودار و دقت و زمان بالاتری برخوردار است. با توجه به جدول ۵ در تصاویر حاصل از اعمال الگوریتم (SVM) با کرنل خطی توسط روش‌های قطعه‌بندی در حالت‌های خوش خیم و بدخیم، میزان کارایی سیستم، مقدار نسبتاً مناسبی را نشان داده است که زمان اجرای این روش ۳۰ ثانیه می‌باشد (زمان مناسبی می‌باشد). اما مقدار زمان اجرای تصاویر حاصل از اعمال الگوریتم (SVM) توسط روش‌های قطعه‌بندی، ۴۲ ثانیه و زمان اجرای تصاویر حاصل از الگوریتم (SVM) توسط الگوریتم فازی (FCM) با کرنل (RBF) در

بهبودسازی شده است تا در کشورهای در حال توسعه که در آن‌ها سرطان معده شایع است موثر باشد. این پژوهش به مشکلاتی از جمله: خطاهای متخصصان پاتولوژیست و کاهش قابل توجهی در هزینه‌های ارائه‌ی خدمات پزشکی که به نوبه‌ی خود از اهمیت بالایی برخوردار است، می‌پردازد. نتایج حاصل از شبیه‌سازی در ۴ روش بر روی تصاویر پاتولوژی، جهت تشخیص صحیح نمونه‌های سالم از نمونه‌های سرطانی بافت معده مورد ارزیابی قرار گرفته است. میزان کارایی سیستم در تصاویر حاصل از الگوریتم (SVM) با کرنل خطی توسط روش‌های قطعه‌بندی در حالت‌های سالم، سرطانی، خوش‌خیم و بدخیم مقدار نسبتاً مناسبی را نشان داده است؛ اما در تصاویر حاصل از الگوریتم (SVM) توسط روش‌های قطعه‌بندی و در تصاویر حاصل از الگوریتم (SVM) با کرنل (RBF) توسط روش (FCM) در هر ۴ حالت میزان کارایی سیستم، مقدار مناسبی را نشان نمی‌دهد. بررسی‌ها نشان می‌دهد که روش پیشنهادی در حالت سالم و سرطانی با دقت ۹۲،۱۲ و در حالت خوش‌خیم و بدخیم با دقت ۹۴،۳۵ نسبت به سایر روش‌ها عملکرد بسیار مناسبی را نشان داده است. همچنین مدت زمان شبیه‌سازی کاهش یافته، که این امر موجب بهبود عملکرد سیستم در زمان اعمال تصاویر با تعداد زیاد می‌شود.

با توجه به اینکه مطالعات کمی در خصوص تشخیص سرطان معده با روش‌های هوشمند پردازش تصویر صورت گرفته است و هدف این پژوهش کمک به کاهش خطاهای بصری متخصصان پاتولوژیست در تشخیص سلول‌های سالم از سلول‌های سرطانی معده می‌باشد بنابراین با استفاده از یک روش مناسب به طبقه‌بندی درست سلول‌های سرطانی از سالم پرداخته شده اما تا کنون روشی جهت درجه‌بندی سلول‌های سرطانی در تصاویر پاتولوژی ارائه نشده است؛ می‌توان امیدوار بود که در آینده یک روش ترکیبی با دقت بالا و کاهش زمان محاسباتی، جهت درجه‌بندی سرطانی معده ارائه کرد که بتواند در درجه‌بندی سلول‌های سرطانی بدخیم در جهت رفع ضعف‌های پاتولوژیست‌ها، کمک به سزایی کند.

همانطور که در شکل (۱۲) مشاهده می‌شود منحنی (ROC) نشان می‌دهد که روش پیشنهادی نسبت به سایر روش‌های دیگر در هر ۴ حالت (سالم، سرطانی، خوش‌خیم و بدخیم) از مساحت زیر نمودار و دقت و زمان بالاتری برخوردار است.



● طبقه‌بندی SVM با کرنل RBF توسط قطعه‌بندی FCM
 ● طبقه‌بندی SVM توسط روش‌های قطعه‌بندی
 ● طبقه‌بندی SVM با کرنل خطی توسط روش‌های قطعه‌بندی
 ● روش پیشنهادی
 شکل ۱۲: نتایج مقایسه‌ی منحنی (ROC) سه روش با روش پیشنهادی در نمونه‌های خوش‌خیم و بدخیم

۷- نتیجه‌گیری

نیاز به تشخیص دقیق و به موقع در درمان بیماری‌ها به خصوص در سرطان معده موجب به کارگیری تحقیقات گسترده و تخصصی توسط محققین در زمینه‌ی تشخیص پزشکی شده است. زیرا عدم تشخیص دقیق و به موقع این بیماری باعث می‌شود، سالانه جان میلیون‌ها انسان گرفته شود. این پژوهش یک سیستم هوشمند را پیشنهاد می‌کند که با استفاده از الگوریتم طبقه‌بندی (SVM) با کرنل (RBF) توسط روش‌های قطعه‌بندی جهت صحت و دقت بیش‌تر و همچنین نزدیکی بیش‌تر به تشخیص متخصصین،

مراجع

- [1] M.S. Sierra, P. Cueva, L.E. Bravo, and D. Forman, "Stomach cancer burden in Central and South America", the International Journal of Cancer Epidemiology, Detection and Prevention, Vol.4, No.1, pp.62-73, 2016.
- [2] X. Yang, L. Liu, Fa. Fang, D. Redati, and H. Wang, "Correlation between cellular immune function and prognosis of gastric cancer", Department of Gastrointestinal Surgery, Cancer Hospital, Affiliated to Xinjiang Medical University, china, pp.1275-1282, 2017.
- [3] E. Norero, E.A. Vega, C. Diaz, G. Cavada, M. Ceroni, C. Martinez, E. Briceno, F. Araos, P. Gonzalez, S. Baez, E. Vinuela, M. Caracci, and A. Diaz, "Improvement in postoperative mortality in elective gastrectomy for gastric cancer: Analysis of predictive factors in 1066 patients from a single centre", European Journal of Surgical Oncology, Vol.43, No.7, pp.1330-1336, 2017.
- [4] N. Amoori, S. Mahdavi, and M. Enayatrads, "Epidemiology and trend of stomach cancer mortality in Iran", International Journal of Epidemiologic Research, Vol.3, No.3, pp. 268-275, 2016.
- [5] Y. Veisani, and A. Delpisheh, "Survival rate of gastric cancer in Iran; a systematic review and meta-analysis", Gastroenterology and Hepatology from Bed to Bench, Ilam, Iran, Vol.9, pp.78-86, 2015.
- [6] W.L. Lamps, A.M. Bellizzi, W.L. Frankel, S.R. Owens, and R.K. Yantiss, "Neoplastic Gastrointestinal Pathology", springer, New York, July 2015.
- [7] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential", IEEE reviews in biomedical engineering, IEEE, vol. 7, pp.97-114, 2013.
- [8] TM. Shahriar Sazzad, L.J. Armstrong, and A.K. Tripathy, "An Automated Ovarian Tissue Detection Approach Using Type P63 Non-Counter Stained Images to Minimize Pathology Experts Observation Variability", IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, Malaysia, pp.1-5, 2017.
- [9] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images". Computers in Biology and Medicine, ScienceDirect, Vol.43, No.10, pp.1563-1572, 2013.
- [10] T. Wan, X. Liu, J. Chen, and Z. Qin, "Wavelet-based statistical features for distinguishing mitotic and non-mitotic cells in breast cancer histopathology", IEEE International Conference on Image Processing (ICIP), Paris, France, pp.1-5, 2015.
- [11] T. Goudas, and I. Maglogiannis, "Cancer Cells Detection and Pathology Quantification Utilizing Image Analysis Techniques", Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, vol.34, pp.4418-4421, 2012.
- [12] M.A. Aswathy, and M. Jagannath, "Detection of breast cancer on digital histopathology images: Present status and future possibilities". Informatics in Medicine Unlocked at Science Direct, India, Vol.8, pp.74-79, 2016.
- [۱۳] سکینه اسدی امیری و حمید حسن پور، "ارائه روشی برای پیش پردازش تصویر جهت بهبود عملکرد JPEG 2000 در فشرده سازی تصویر"، نشریه مدل سازی در مهندسی، دوره ۱۵، شماره ۴۸، بهار ۱۳۹۶، صفحه ۱-۱۲.
- [۱۴] اصغر زارع و علی محمدزاده، "حذف نویز ضربه‌ای از تصاویر دیجیتالی مبتنی بر تخمین توزیع مکانی نویزها"، نشریه مدل سازی در مهندسی، دوره ۱۲، شماره ۳۹، زمستان ۱۳۹۳، صفحه ۱-۱۷.
- [۱۵] امین رضایی پناه و علی مبارکی، سعید بحرانی خادمی، "بهینه سازی شبکه عصبی MLP با استفاده از الگوریتم ژنتیک موازی FinGrain برای تشخیص سرطان سینه"، نشریه مدل سازی در مهندسی، دوره ۱۷، شماره ۵۷، تابستان ۱۳۹۸، صفحه ۱۲-۱۲.
- [16] X. Baia, X. Chen, F. Zhou, Zh. Liu, and B. Xue, "Multiscale top-hat selection transform based infrared and visual image fusion with emphasis on extracting regions of interest", Infrared Physics and Technology, Vol. 60, pp. 81-93, 2013.
- [17] H. Hassanpour, N. Samadiani, and M. Salehi, "Using morphological transforms to enhance the contrast of medical images", The Egyptian Journal of Radiology and Nuclear Medicine, Vol.46, Issue 2, pp.481-489, 2015.
- [18] C. Goswami Jaideva, and K. Chan Andrew, "Fundamentals of wavelets: theory, algorithms, and applications", 2th ed, Wiley, USA, 2011.

- [19] A. Khare, and U.S. Tiwary, "Daubechies complex wavelet transform based technique for denoising of medical images". International Journal of Image and Graphics, Vol.7, No.40, pp. 663–687, 2007.
- [20] S.I. Niwas, P. Palanisamy, K. Sujathan, and E. Bengtsson, "Analysis of nuclei textures of fine needle aspirated cytology images for breast cancer diagnosis using Complex Daubechies wavelets", Signal Processing, Elsevier, Vol.93, Issue.10, pp.2828–2837, 2013.
- [21] L. I. Kuncheva, "Combining pattern classifiers: methods and algorithms", John Wiley & Sons, 2004.
- [22] L. Cheng-Hsuan, L. Chin-Teng, K. Bor-Chen, and H. Hsin-Hua, "An automatic method for selecting the parameter of the normalized kernel function to support vector machines", International Conference on Technologies and Applications of Artificial Intelligence, Hsinchu, Taiwan, pp.226-232, 2011.
- [23] W. Wang, H. Song, and Q. Zhao, "A modified Watersheds Image Segmentation Algorithm for Blood Cell", International Conference on Communications, Circuits and Systems, Guilin, China, pp.451-452, 2007.
- [24] K. Rodenacker, and E. Bengtsson, "A feature set for cytometry on digitized microscopic images", the Journal of the European Society for Analytical Cellular Pathology, Vol.25, No.1, pp.1-36, 2003.
- [25] J. Davis, and M. Goadrich, "The Relationship between Precision-Recall and ROC Curves", international conference on Machine learning, New York, NY, USA, pp.233-240, 2006.