

بازشناسی تومورها در تصاویر ماموگرام با استفاده از ویژگی‌های ماتریس هم‌رخداد و مبتنی بر ماشین بردار پشتیبان

سید مسعود هاشمی عمرآبادی^۱، محمدرضا احمدزاده^۲، علی حکمت‌نیا^۳
^۱- دانشجوی دکتری، دانشکده برق و کامپیوتر، دانشگاه صنعتی اصفهان، m_hashemi@ec.iut.ac.ir
^۲- استادیار، دانشکده برق و کامپیوتر، دانشگاه صنعتی اصفهان، ahmadzadeh@cc.iut.ac.ir
^۳- دانشیار، مرکز تحقیقات پردازش تصویر و سیگنال پزشکی، دانشگاه علوم پزشکی اصفهان، aekmatnia@med.mui.ac.ir

چکیده

در این مقاله الگوریتمی برای یافتن نواحی مشکوک و سپس بازشناسی تومورها در تصاویر ماموگرافی ارائه می‌شود. برای بازشناسی تومورها تعدادی ویژگی از تصاویر مربوط به نواحی مشکوک استخراج می‌شود. برای بدست آوردن ویژگی‌ها از ماتریس هم‌رخداد استفاده شده که کارآیی قابل قبولی دارد. برای کاهش تعداد ویژگی‌های مورد استفاده و افزایش بازدهی الگوریتم، یک روش برای انتخاب ویژگی‌های مؤثر با استفاده از الگوریتم ژنتیک پیشنهاد شده است. برای دسته‌بندی نهایی تومورها نیز از بین ابزارهای مختلف موجود، SVM پس از ارزیابی تعدادی از ابزارهای کلاس‌بندی به‌عنوان یکی از موفق‌ترین روش‌های کلاس‌بندی در این کاربرد انتخاب شده و مورد استفاده قرار گرفته است. درصد تشخیص مثبت صحیح بدست آمده توسط SVM برابر ۹۴/۵۹٪ و درصد تشخیص مثبت کاذب برابر ۲۲/۹۵٪ می‌باشد.

واژه‌های کلیدی

تصاویر ماموگرام، تومورهای سرطانی، ماتریس هم‌رخداد، ماشین بردار پشتیبان، الگوریتم ژنتیک، آنتروپی، ویژگی‌های هارالیک

۱- مقدمه

می‌گیرد الگوریتمی برای شناسایی نواحی مشکوک پیشنهاد شده است تا الگوریتم بازشناسایی تنها بر تصاویر خروجی از این الگوریتم اعمال شود. برای بازشناسی تومورها تعدادی ویژگی از تصاویر مربوط به نواحی مشکوک استخراج می‌شود. با استفاده از ویژگی‌های بدست آمده از ماتریس هم‌رخداد کارآیی قابل قبولی حاصل شد. برای کاهش تعداد ویژگی‌های مورد استفاده و افزایش بازدهی الگوریتم، یک روش برای انتخاب ویژگی‌های مؤثر با استفاده از الگوریتم ژنتیک پیشنهاد شده است. برای دسته‌بندی نهایی نیز از بین رده‌بندی‌کننده‌های مختلف موجود، SVM به‌عنوان یکی از مطمئن‌ترین روش‌های کلاس‌بندی در این کاربرد انتخاب شده و

سرطان سینه شایع‌ترین سرطان و اولین عامل مرگ و میر ناشی از سرطان در میان زنان جهان است. در حال حاضر ماموگرافی مناسب‌ترین روش تشخیص سرطان سینه محسوب می‌شود. اما به‌علت نوع بافت سینه و استفاده از تابش اشعه ایکس با تراکم پایین در تهیه ماموگرام‌ها، تصاویر از تباین پایینی برخوردار هستند. درضمن تومورها دارای اندازه‌ها و شکل‌های متفاوتی می‌باشند به همین دلیل تشخیص ضایعات به‌خصوص در مراحل اولیه تشکیل کار بسیار دشوار و خسته‌کننده‌ای است. آمار نشان می‌دهد تشخیص زودهنگام ضایعات در کاهش مرگ‌ومیر ناشی از سرطان نقش چشمگیری دارد. برای کاهش بیشتر نواحی که مورد پردازش قرار

۳- روش پیشنهادی برای یافتن نواحی مشکوک

روش پیشنهادی برای یافتن مناطق مشکوک و یا نواحی مورد توجه برای یافتن تومور شامل مراحل زیر است:

۱. یافتن مرز سینه و ماهیچه پکتورال و حذف تمام مصنوعات و پکتورال (این نواحی مشابه تومورها دارای روشنایی‌های زیادی هستند). برای یافتن مرز سینه از روش آستانه‌گذاری براساس آنتروپی و برای یافتن مرز ماهیچه پکتورال نیز از روش فیلترهای گابور استفاده شده است [۹، ۱۰].
۲. استفاده از عملگر Bottom Hat بر روی تصویر بدست‌آمده از مرحله اول به منظور حذف جزئیات روشن تصویر که دارای اندازه‌هایی کوچکتر از اندازه المان ساختاری هستند [۳].
۳. استفاده از عملگر Top Hat با المان ساختاری بزرگ بر روی تصویر بدست‌آمده از مرحله اول به منظور حذف جزئیات تاریک تصویر و حفظ مناطق روشن تصویر.
۴. تفاضل تصویر خروجی مرحله دوم از تصویر خروجی مرحله سوم؛ به این ترتیب تصویری که بدست می‌آید فقط شامل مناطقی خواهد بود که دارای روشنایی بیشتری نسبت به پیکسل‌های مجاور خود هستند.
۵. بافت سینه دارای رگ‌ها و مجاری شیر می‌باشد که دارای روشنایی‌های زیاد هستند و باید تا حد ممکن حذف شوند. برای حذف آنها از عملگر Opening استفاده می‌شود. اندازه در نظر گرفته شده برای المان ساختاری مربوطه باید دارای شعاع نسبتاً کمی باشد تا باعث حذف تومورهای با اندازه کوچک نشود [۳].
۶. پس از استفاده از عملگر Opening برای افزایش پیوستگی بین پیکسل‌ها، از عملگر Closing با المان ساختاری کمی کوچکتر از المان ساختاری مورد استفاده در مرحله ۵ استفاده می‌کنیم [۳].
۷. استفاده از تابع

$$I'(x, y) = \frac{I(x, y)}{\max(I(x, y)) + 1 - I(x, y)}$$

که روشنایی‌ها را به صورت غیر خطی تغییر می‌دهد و باعث واضح‌تر شدن مناطق روشن در تصویر می‌شود.

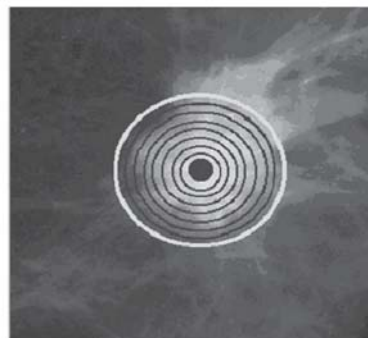
۸. پیدا کردن ماکزیمم‌های محلی در تصویر.
۹. در نظر گرفتن نقاط بدست‌آمده از مرحله ۸ به عنوان مرکز مناطق مشکوک و استفاده از روش رشد ناحیه^۲ برای یافتن مرز آن. هر نقطه ماکزیمم تا حداکثر شعاع ۲۰۰

مورد استفاده قرار گرفته است. در بخش ۲ روش‌های معمول جهت تشخیص نواحی مشکوک توضیح داده می‌شود. روش پیشنهادی برای یافتن این نواحی در بخش ۳ ارائه می‌شود. در بخش ۴ ویژگی‌های استخراج شده از ماتریس هم‌رخداد تشریح می‌شود. طبقه‌بندی بافت با استفاده از SVM در بخش ۵ مطرح و نتایج آن ارائه می‌گردد. در بخش ۶ خلاصه و نتیجه‌گیری از این پژوهش ارائه می‌شود.

۲- تشخیص نواحی مشکوک

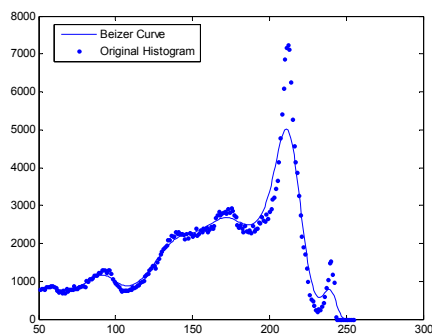
به دلیل اندازه بزرگ تصاویر ماموگرام و بالا بودن حجم محاسبات لازم برای یافتن مکان صدمات و بافت‌های بیمار، انجام عمل تشخیص برای تمام سطح ماموگرام کاری زمان‌بر است. بنابراین کاهش نواحی مورد بررسی اگر بدون از دست دادن اطلاعات با ارزش انجام گیرد به دقت، سرعت و حجم محاسبات کمک شایانی خواهد کرد. از سوی دیگر برخی ویژگی‌های تومورها بسیار شبیه نواحی بافت مثل پکتورال و یا برخی مصنوعات می‌باشد که لازم است برای کاهش درصد تشخیص مثبت نادرست، کاهش نواحی مورد بررسی به صورت پیش‌پردازش استفاده می‌شود. به یافتن نواحی مشکوک به سرطان، تعیین ناحیه مورد توج^۱ نیز گفته می‌شود.

یکی از روش‌هایی که در بسیاری از الگوریتم‌های تحقیقاتی پیشنهادی توسط گروه‌های مختلف از آن برای تشخیص نواحی مشکوک استفاده شده است به نام ROI Hunter [۱] می‌باشد. در این روش برای تعیین نواحی مورد توجه مراحل چهارگانه‌ای به صورت تکراری انجام می‌گیرد تا تمام سطح ماموگرام پیمایش شده و تمام مناطق مشکوک تعیین شوند. این مراحل عبارتند از: ۱- یافتن یک ماکزیمم محلی (مرکز یک ناحیه مشکوک). ۲- ترسیم دایره‌های هم مرکز حول نقطه ماکزیمم و افزایش شعاع دایره‌ها تا زمانی که نسبت متوسط روشنایی حلقه به ماکزیمم روشنایی از حد آستانه T_R بیشتر شود (شکل (۱)). ۳- حذف دایره نهایی از تصویر. ۴- تکرار الگوریتم تا یافتن تمامی ماکزیمم‌ها [۲].



شکل ۱- دسته دایره‌هایی با افزایش ۵ پیکسلی شعاع به دور یک ماکزیمم بدست آمده [۲]

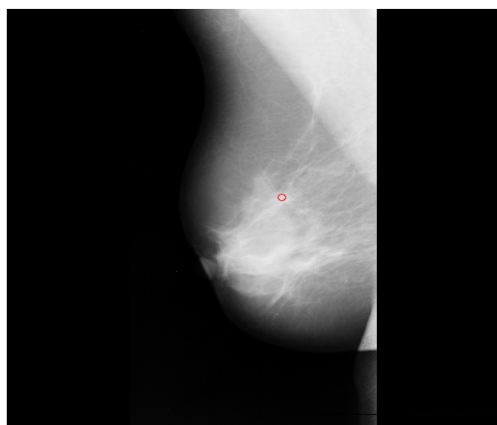
نمونه‌ای از یک هیستوگرام تقریب زده شده با این روش را در شکل (۲) می‌توان دید.



شکل ۲- یک هیستوگرام تصویر که با منحنی Bezier تقریب زده شده است.

بنابراین با اصلاح الگوریتم قبلی، پیش از مرحله ۹ یعنی الگوریتم رشد نواحی، هیستوگرام تصویر بدست آمده از مرحله ۷ را با استفاده از منحنی Bezier تقریب می‌زنیم. سپس نقطه مینیمی که بین اولین و دومین ماکزیمم در هیستوگرام تقریب زده شده وجود دارد را به‌عنوان حد آستانه در نظر می‌گیریم و ماکزیمم‌های محلی را که دارای مقادیر کمتر از این حد آستانه هستند را از تصویر حذف می‌نماییم.

با توجه به آمارهای داده شده الگوریتم ROI Hunter در هر تصویر به‌طور متوسط 9 ± 1 منطقه مشکوک تشخیص داده است [۲]. در الگوریتم پیشنهاد شده در این مقاله این تعداد به $4/5$ تا 5 منطقه در هر تصویر کاهش یافته است. در شکل (۳) نتیجه بدست آمده از اعمال این الگوریتم بر یک تصویر درحالی که تنها دارای سه عدد تشخیص مثبت اشتباه است نشان داده شده است.



(الف)

پیکسل و یا تا جایی که دارای روشنایی نصف مقدار ماکزیمم باشد رشد می‌یابد و مرز آن به‌عنوان مرز منطقه مورد توجه در نظر گرفته می‌شود. نواحی که از حد آستانه‌ای کوچک‌تر هستند نیز در این مرحله حذف می‌شوند.

لازم به‌ذکر است المان‌های ساختاری برای Top, Bottom Hat, Opening و Closing دیسک‌هایی دایره‌ای و با شعاع متناسب با رزولوشن تصویر انتخاب شده‌اند. برای پایگاه داده مورد استفاده این مقادیر به‌ترتیب عبارتند از: ۱۰۰، ۱۰۰، ۳۰ و ۲۰ پیکسل.

پس از اجرای الگوریتم پیشنهاد شده ملاحظه شد که تعداد نواحی بدست آمده زیاد می‌باشند. به‌منظور بهبود عملکرد الگوریتم از یک پیش‌پردازش مبتنی بر آستانه‌گذاری هیستوگرام تصویر استفاده گردید. این روش که مبتنی بر منحنی Bezier می‌باشد برای هموار نمودن هیستوگرام استفاده می‌شود.

۳-۱- کاهش تعداد ماکزیمم‌های محلی با استفاده از

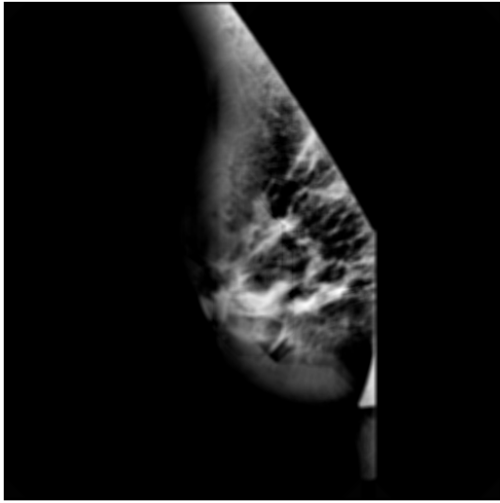
تقریب هیستوگرام به‌وسیله منحنی Bezier

هیستوگرام چگونگی توزیع روشنایی‌ها را در تصویر نشان می‌دهد. ولی گاهی اطلاعات لازم به‌دلیل وجود دندان‌های بسیار در هیستوگرام قابل حصول نیست. به‌همین منظور لازم است آن را هموار نمود. برای هموار کردن آن می‌توان از روش‌های نویز زدایی استفاده کرد. در این قسمت یک روش بسیار مناسب برای هموارسازی هیستوگرام معرفی می‌شود که مبتنی بر تقریب هیستوگرام توسط تابع Bezier می‌باشد [۲].

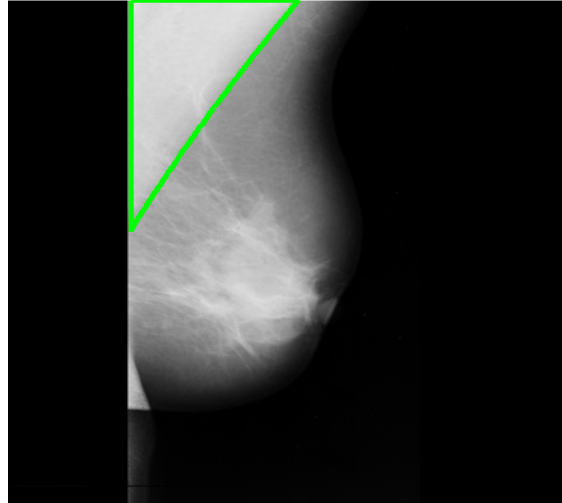
یکی از مهمترین خواص توابع Bezier این است که می‌توانند با هر تعداد نقطه کنترلی موجود تطبیق شده و همواره حالت تحدب و تقعر نقاط کنترلی خود را حفظ کنند. در ضمن تابع تقریب زده شده حتماً از نقاط اول و آخر کنترلی عبور می‌کند.

رابطه منحنی Bezier همیشه درجه‌اش یک درجه کمتر از تعداد نقاط کنترلی موجود است. برای تخمین هیستوگرام از آنجایی که حداکثر تعداد روشنایی‌ها برای تصاویر ۸ بیتی ۲۵۶ سطح است و این سطوح به‌عنوان نقاط کنترلی استفاده می‌شوند بنابراین ماکزیمم درجه مورد استفاده ۲۵۵ است. چنانچه نقاط کنترلی را با p_i نشان دهیم معادله منحنی Bezier از رابطه زیر بدست می‌آید. برای اطلاع از جزئیات بیشتر به مرجع [۸] مراجعه نمایید.

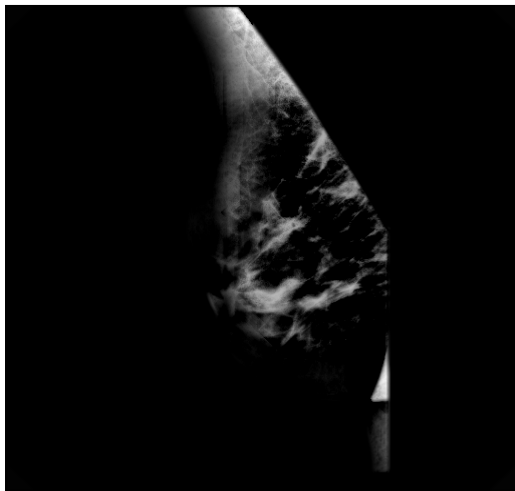
$$P(t) = \sum_{i=0}^{255} (1-t)^{255-i} t^i p_i, \quad 0 \leq t \leq 1 \quad (1)$$



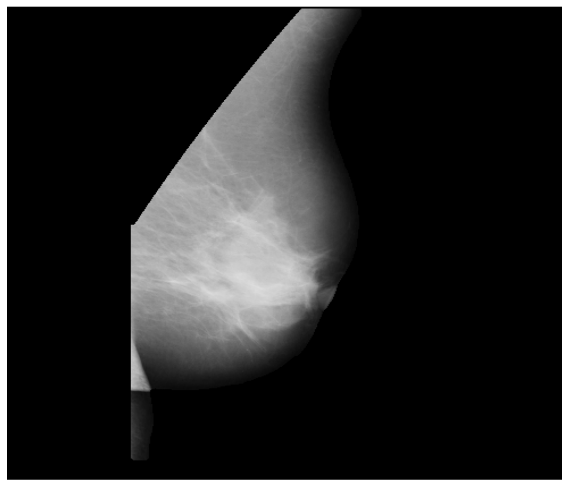
(ا)



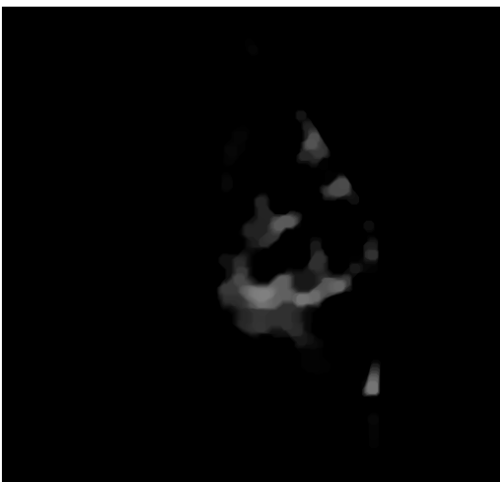
(ب)



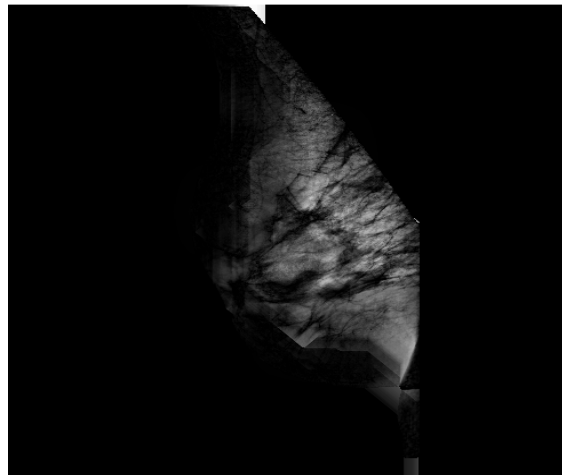
(ج)



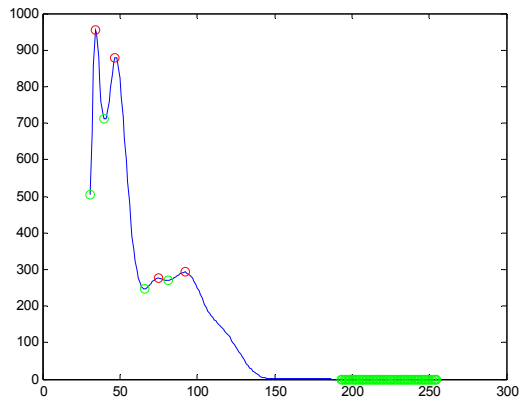
(د)



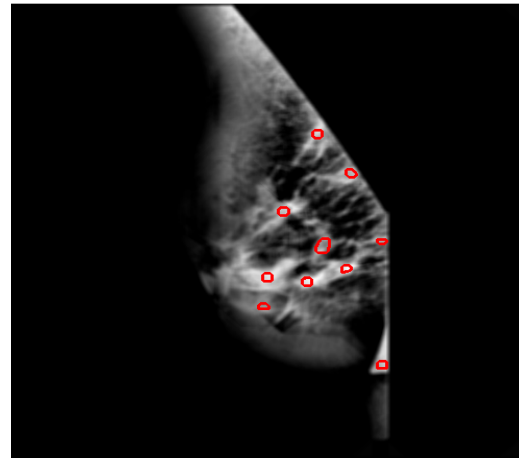
(ه)



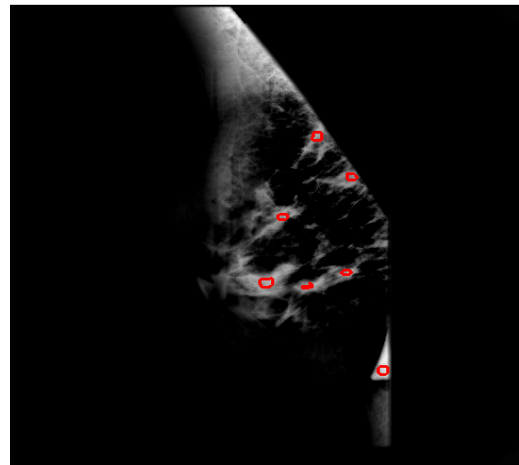
(و)



شکل ۳- الف- تصویر mdb207 از پایگاه داده mini-MIAS و محل تومور در آن، ب- تشخیص محل ماهیچه پکتورال، پ- تصویر پس از حذف پکتورال، ت- تصویر پس از اعمال عملگر Bottom Hat، ث- تصویر پس از اعمال عملگر Top Hat، ج- تفاضل دو تصویر قبل، چ- تصویر پس از اعمال Opening و Closing و تابع غیرخطی، ح- تخمین هیستوگرام، خ- ماکزیمم‌های محلی، د- کم کردن تعداد ماکزیمم‌ها، ذ- تصویر نهایی



(خ)



(د)

۴- ویژگی‌های استخراج شده از ماتریس هم‌رخداد

هارالیک و همکاران برای بررسی ساختار بافت‌های مختلف ویژگی‌هایی را براساس ماتریس همجواری پیشنهاد کردند که از موفق‌ترین روش‌های بررسی خواص بافت‌های مختلف است [۷ و ۴]. در این روش ابتدا ماتریس‌های هم‌رخداد برای فواصل و جهات مختلف محاسبه می‌شوند و سپس برای هر یک از ماتریس‌ها تعدادی ویژگی محاسبه می‌شود. ماتریس‌های هم‌رخداد برای زوایای ۰، ۴۵، ۹۰، ۱۳۵ و متوسط این چهار ماتریس و در فواصل ۱، ۵ و ۷ پیکسل (در مجموع ۱۵ ماتریس) ساخته می‌شوند. برای ماتریس $C_{m \times n}$ ویژگی‌های پیشنهادی هارالیک و همکاران زیر بدست آمده است [۵].

$$1. \text{ انرژی} = \sum_{i=1}^n \sum_{j=1}^m C_{ij}^2$$

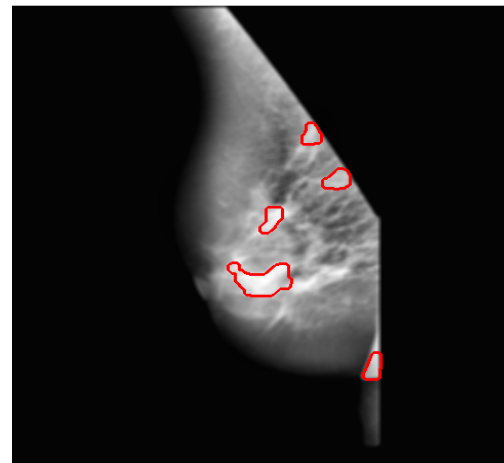
$$2. \text{ انتروپی} = -\sum_{i=1}^n \sum_{j=1}^m C_{ij} \log C_{ij}$$

$$3. \text{ کنتراست (تباین)} = \sum_{i=1}^n \sum_{j=1}^m |i-j| C_{ij}$$

$$4. \text{ مومنتوم اختلاف معکوس}^{\ddagger}$$

$$IDM = \sum_{i=1}^n \sum_{j=1}^m \frac{C_{ij}}{|i-j|^k}, i \neq j, k = 1, 2$$

$$5. \text{ یکنواختی}^{\circ}$$



(ذ)

تابعی است که مقدار مساحت زیر منحنی مشخصه عملکرد گیرنده^{۱۵} (ROC) را به عنوان خروجی می‌دهد. منحنی ROC حاصل ترسیم درصد تشخیص درست مثبت بر حسب درصد تشخیص اشتباه مثبت می‌باشد و هر چه مساحت زیر منحنی ROC به یک نزدیک‌تر باشد الگوریتم موفق‌تر است [۸]. نحوه کار به این صورت است که یک جمعیت اولیه با تعداد ۱۰۰ رشته که دارای بیت‌هایی با طول تعداد ویژگی‌ها هستند ساخته می‌شود. مثلاً در این تحقیق ۲۱۳ بیت برای هر رشته در نظر گرفته می‌شود. سپس این رشته‌ها که دارای کوروموزم‌هایی با مقادیر ۰ و ۱ هستند وارد الگوریتم می‌شوند. متناظر با هر یک از رشته‌ها اگر بیتی صفر باشد، ویژگی متناظر با آن بیت حذف می‌شود و ویژگی‌های متناظر با بیت‌های با مقدار یک برای تصمیم‌گیری استفاده می‌شوند و منحنی ROC با استفاده از آن ویژگی‌ها ترسیم می‌شود. بنابراین الگوریتم ژنتیک مورد استفاده ویژگی‌هایی را که باعث بدست آمدن بیشترین مقدار سطح زیر منحنی شوند را انتخاب خواهد کرد. در الگوریتم استفاده شده تعداد جمعیت اولیه ۱۰۰ رشته است و الگوریتم ۱۰۰ بار تکرار می‌شود. به منظور مقایسه، روش آنتروپی نیز جهت انتخاب ویژگی‌های مؤثر پیاده‌سازی گردید که نتایج آن در بخش بعدی بررسی می‌شود. برای این منظور از بهره اطلاعات^{۱۶} (IG) استفاده می‌نماییم. رابطه IG برای ویژگی A در مجموعه داده D به صورت زیر تعریف می‌شود.

$$Gain(D, A) =$$

$$Ent(D) - \sum_{v \in \text{values}(A)} \frac{|D_v|}{D} Ent(D_v) \quad (2)$$

که در این رابطه، آنتروپی به صورت زیر است.

$$Ent(D) = \sum_{i=1}^c -p_i \log(p_i) \quad (3)$$

۴-۲- نتایج بدست آمده از استخراج ویژگی‌ها و انتخاب

ویژگی‌های مؤثر

پس از معرفی روابط و تکنیک‌هایی که برای استخراج ویژگی‌های مختلف از آنها استفاده شده، به بررسی نتایج حاصل از دسته‌بندی انجام شده بر روی تصاویر مربوط به مناطق مشکوک با استفاده از هر دسته از این ویژگی‌ها می‌پردازیم. برای انجام دسته‌بندی و تصمیم‌گیری نهایی از شبکه‌های عصبی استفاده شده است. شبکه عصبی مورد استفاده از نوع پرسپترون چندلایه است که با استفاده از روش پس انتشار خطا آموزش داده می‌شود [۶]. تعداد لایه‌های استفاده شده سه لایه است. تعداد نرون‌های لایه اول به تعداد ویژگی‌های منتخب، یک نرون در لایه خروجی و تعداد نرون‌های لایه پنهان با انجام آزمایش‌های متعدد انتخاب شده است. شبکه

$$Homogeneity = \sum_{i=1}^n \sum_{j=1}^m \frac{C_{ij}}{1+|i-j|} \quad 6. \text{ مجموع متوسط}^6$$

$$SumMean = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m (iC_{ij} + jC_{ij}) \quad 7. \text{ واریانس}$$

$$Var = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m ((i-\mu)^2 C_{ij} + (j-\mu)^2 C_{ij}) \quad 8. \text{ همبستگی}^7$$

$$Correlation = \frac{\sum_{ij} (i-\mu)(j-\mu)C_{ij}}{\sqrt{\text{var}(i)\text{var}(j)}} \quad 9. \text{ حداکثر احتمال}^8$$

$$MP = \max_{i,j} C_{ij} \quad 10. \text{ گرایش دسته}^9$$

$$CT = \sum_{i=1}^n \sum_{j=1}^m (i+j-2\mu)^k C_{ij}, k=2,3$$

$$RL2 = \sum_{i=1}^n \frac{(C_i - C_{ij})^2 C_{ij}}{C_i^2} \quad 11. \text{ احتمال تکرار با طول دو}^{10}$$

$$Inertia = \sum_{i=1}^n \sum_{j=1}^m (i-j)^2 C_{ij} \quad 12. \text{ اینرسی}$$

$$Intensity = \sum_{i=1}^n \sum_{j=1}^m ij C_{ij} \quad 13. \text{ شدت روشنایی}^{11}$$

۱۳ ویژگی فوق برای ۱۵ ماتریس هم‌رخداد (جمعاً ۱۹۵ ویژگی) محاسبه می‌شود. علاوه بر این ویژگی‌های مربوط به هیستوگرام تصویر و ویژگی‌هایی که مستقیماً از تصویر گرفته شده‌اند نیز اضافه شد. ویژگی‌های متوسط سیگنال، انحراف معیار، کشیدگی^{۱۲}، عدم تقارن^{۱۳} (کجی)، عامل Clearance، عامل ضربه^{۱۴}، عامل Crest و عامل k از هیستوگرام تصویر و ویژگی‌های انرژی، آنتروپی، بعد فرکتالی تصویر، عرض از مبدأ بدست آمده از نمودار مربوط به بعد فرکتالی، عامل Crest، عامل Clearance، عامل ایمپالس، عامل K، متوسط روشنایی‌های تصویر و همبستگی از خود تصویر استخراج شدند. بنابراین در مجموع ۲۱۳ ویژگی استخراج شده تشکیل بردار ویژگی را می‌دهند.

۴-۱- روش پیشنهاد شده برای انتخاب ویژگی‌ها

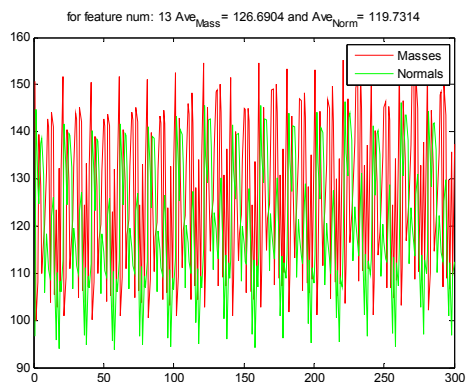
از آنجا که تعداد ویژگی‌ها زیاد است لازم است برای کارایی بالاتر طبقه‌بندی‌کننده ویژگی‌های مؤثر انتخاب و بقیه حذف شوند. در این تحقیق از الگوریتم ژنتیک به این منظور استفاده شده است. از نکات مهم، انتخاب تابع هدف، ماکزیمم نمودن تابع هدف و نحوه کدگذاری داده‌ها است. تابعی که برای این منظور استفاده شده است

۱۱۵-۱۰-۱ بهترین نتیجه را بر روی داده‌های آموزش داشته است. برای آموزش این شبکه از ۵۰ تصویر مربوط به نواحی مشکوک بدست آمده از الگوریتم پیشنهادی در فصل سوم استفاده شده است که دارای ۲۰ تصویر دارای تومور و ۳۰ تصویر بدون تومور می‌باشد. آموزش انجام شده به نحوی است که اگر تصویر دارای تومور باشد خروجی یک و در غیر این صورت خروجی صفر داشته باشد. بنابراین اعدادی که در خروجی این شبکه خواهیم داشت اعدادی به صورت $0 \leq p \leq 1$ هستند. برای تست کارایی نیز از ۶۰ تصویر شامل ۱۸ تومور و ۴۲ بافت سالم استفاده شده است. تمام تصاویر این تحقیق از پایگاه داده MIAS استفاده شده است.

برای بررسی میزان کارایی حاصل از هر تعداد از ویژگی‌ها از سطح زیر منحنی ROC استفاده شده است. برای بدست آوردن منحنی ROC اعداد خروجی بدست آمده از شبکه عصبی را با اعدادی بین ۰ و ۱ آستانه‌گذاری می‌نماییم و برای هر یک از این سطوح درصد تشخیص درست مثبت و درصد تشخیص اشتباه مثبت را بدست می‌آوریم.

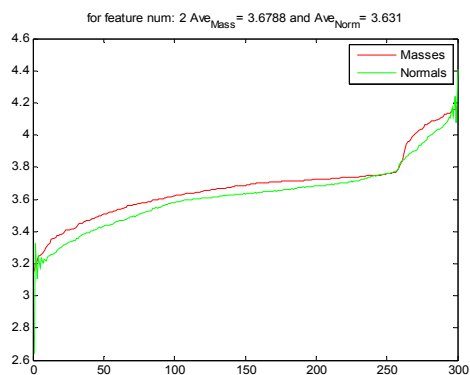
۴-۳- برآورد کارایی ویژگی‌های حاصل از ماتریس‌های هم‌خداد

برای آنکه بتوان مقادیر بدست آمده هر ویژگی را برای بافت‌های سرطانی و سالم با یکدیگر مقایسه کرد در شکل (۴) نمودار مقادیر ویژگی‌ها نشان داده شده است. برای آنکه میزان تفاوت آشکارتر باشد مقادیر از کوچک به بزرگ مرتب شده‌اند. بدون انجام مرتب‌سازی همان‌طور که در شکل (۴) دیده می‌شود مقایسه مقادیر تقریباً غیرممکن است. مقدار متوسط هر یک از ویژگی‌ها نیز در نمودارهای شکل (۵) نمایش داده شده است.

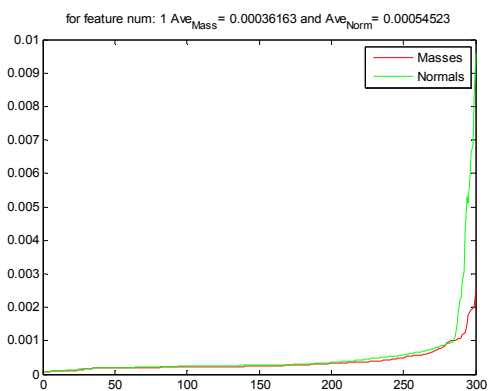


(ب)

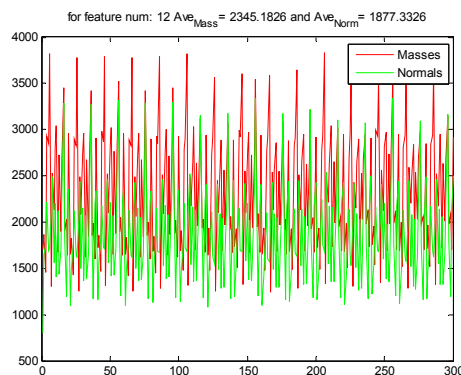
شکل ۴-الف- مقادیر مجموع میانگین‌ها برای تصاویر تومور و سالم، ب- مقادیر واریانس تصاویر برای بافت تومور و سالم



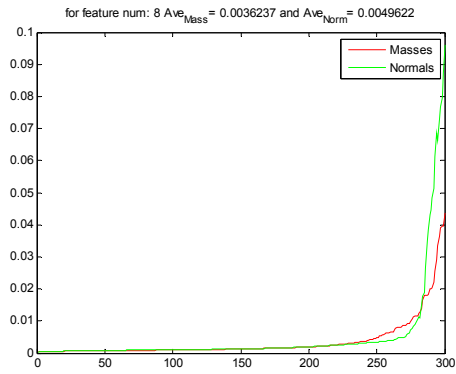
(الف)



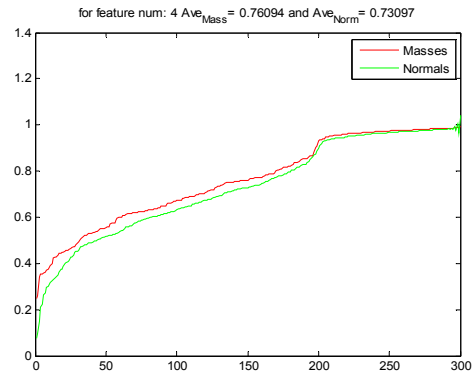
(ب)



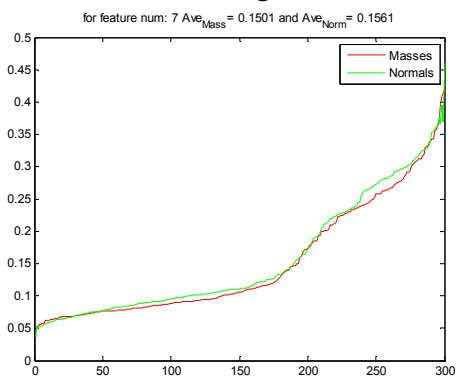
(الف)



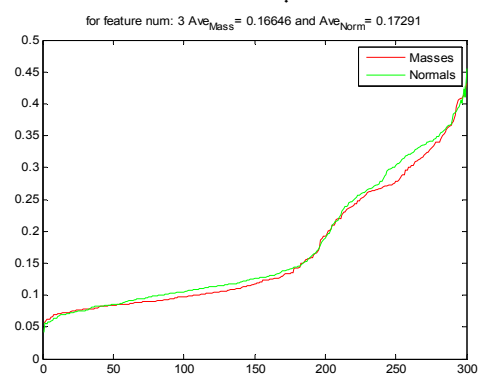
(چ)



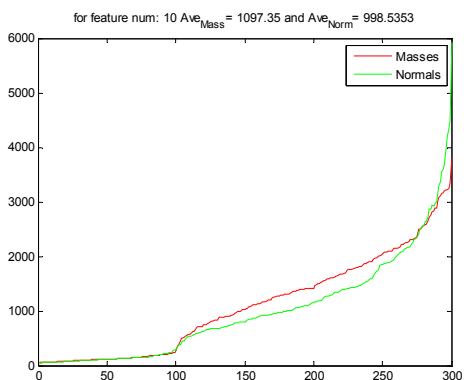
(پ)



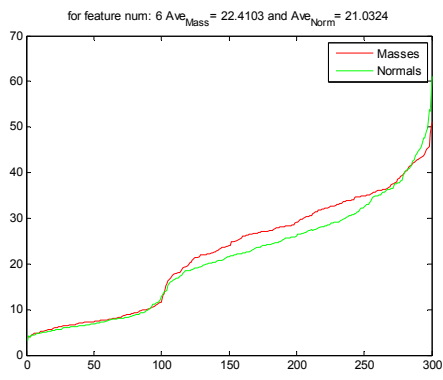
(ج)



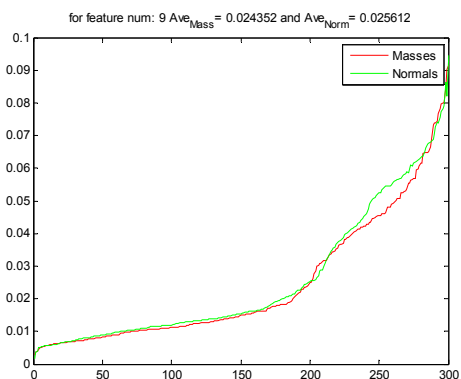
(ب)



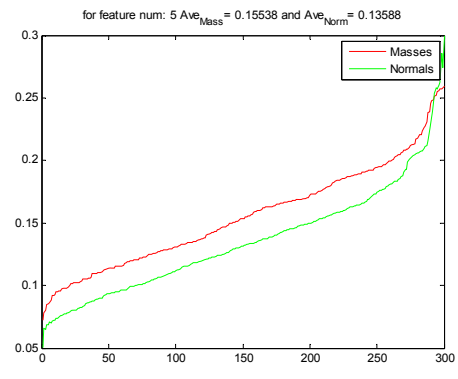
(خ)



(ت)

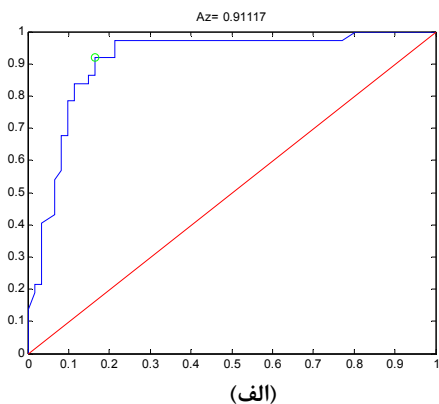


(د)

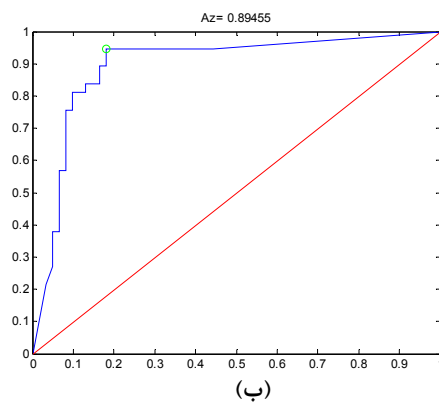


(ز)

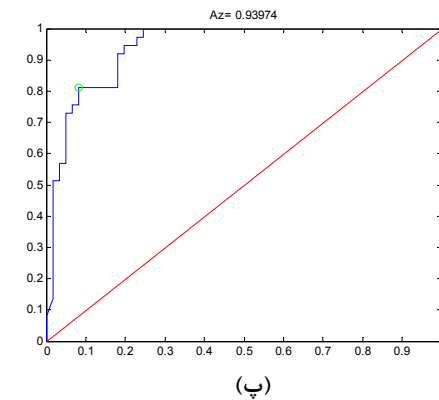
ویژگی‌ها است. همان‌طور که از منحنی‌های بدست آمده مشهود است انتخاب ویژگی‌ها با استفاده از الگوریتم ژنتیک بازدهی را تا حد خوبی افزایش داده است. ویژگی‌هایی که توسط الگوریتم ژنتیک به‌عنوان ویژگی‌های مناسب انتخاب شده در جدول (۱) آمده است. ردیف آخر تعداد ماتریس‌هایی که ویژگی در آنها استفاده شده را نشان می‌دهد.



(الف)

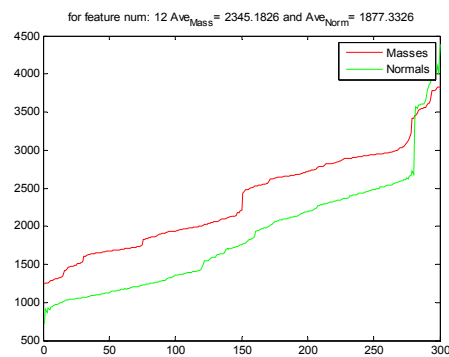


(ب)

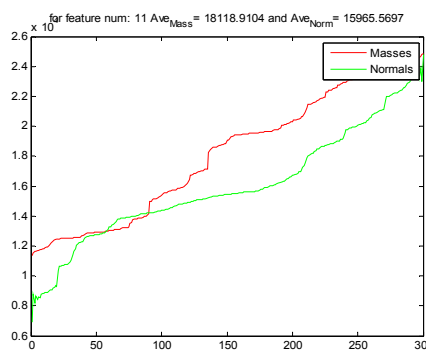


(پ)

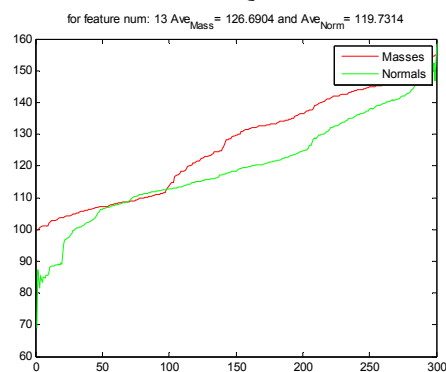
شکل ۶- منحنی ROC برای ویژگی‌های ماتریس هم‌رخداد الف- با وجود تمام ویژگی‌ها، ب- با استفاده از ویژگی‌های انتخاب شده با استفاده از انتروپی، پ- با استفاده از ویژگی‌های انتخابی توسط GA



(د)



(ر)



(ز)

شکل ۵- مقادیر مربوط به الف- انرژی، ب- انتروپی، پ- IDM، ت- همبستگی، ث- Cluster Tendency، ج- تباین، چ- یکنواختی، ح- حداکثر احتمال، خ- RL2، د- اینرسی، ذ- روشنایی، ر- واریانس، ز- مجموع متوسط

در شکل (۶) منحنی‌های ROC برای این ویژگی‌ها ارائه می‌شود. منحنی شکل (۶- الف) مربوط به حالتی است که از تمام ویژگی‌ها استفاده شود، منحنی شکل (۶- ب) نتیجه بکارگیری الگوریتم انتخاب ویژگی‌ها با استفاده از روش مبتنی بر انتروپی و منحنی شکل (۶- پ) مربوط به استفاده از الگوریتم ژنتیک برای انتخاب

جدول ۱- ویژگی‌های انتخاب شده توسط GA برای ویژگی‌های ماتریس هم‌رخداد

نام ویژگی	ویژگی‌های مربوط به ماتریس‌های هم‌رخداد													ویژگی‌های تصویر و هیستوگرام					
	انرژی	انترویی	IDM	همبستگی	Cluster	تباين	یکپارچگی	حد اکثر احتمال	RL2	ایترسی	روشنایی	وارانس	مجموع متوسط	انرژی تصویر	فاکتور K	متوسط تصویر	تصویر	هیستوگرام	Clarence
تعداد ماتریس	۹	۸	۷	۱۰	۸	۸	۴	۹	۷	۹	۱۲	۷	۱۰	۱	۱	۱	۱	۱	۱

جدول ۲- نتایج حاصل از دسته‌بندی با هسته‌های متفاوت

نوع خطا نوع هسته	بدون انتخاب		انتخاب GA	
	TPF	FPF	TPF	FPF
خطی	٪۹۴/۵۹	٪۳۱/۱۵	٪۹۷/۳۰	٪۲۶/۲۳
چند جمله‌ای درجه ۲	٪۹۴/۵۹	٪۲۲/۹۵	٪۹۴/۵۹	٪۲۲/۹۵
چند جمله‌ای درجه ۳	٪۹۴/۵۹	٪۲۷/۸۷	٪۸۹/۱۹	٪۲۱/۳۱
MLP	-	-	٪۴۳/۲۳	٪۷۷
RBF	٪۱۳/۵۷	٪۸/۲	٪۷۶/۲۷	٪۱۶/۳۹

رسیدیم. لازم بذکر است گرچه در بهترین نتیجه کاهش ویژگی‌ها تأثیر نداشته است ولی به دلیل تعداد کمتر ویژگی‌ها زمان طبقه‌بندی کاهش می‌یابد. درصد خطاهای حاصل از دسته‌بندی با استفاده از SVM و هسته‌های مختلف در جدول (۲) آورده شده است.

۶- نتیجه‌گیری

در این مقاله در ابتدا روشی ارائه گردید که بتواند بدون از دست دادن تومورها تا حد ممکن دیگر نواحی را از حوزه جستجو برای تومورها حذف نماید. از آنجا که تصاویر ماموگرام بزرگ می‌باشند تست همه‌ی نواحی برای وجود تومور بسیار وقت می‌باشد. از سوی دیگر برخی از بافت‌های سالم و نیز مصنوعات دارای ویژگی‌هایی می‌باشند که می‌توانند با تومورها اشتباه شوند. بنابراین بدست آوردن نواحی مشکوک ضمن کاهش زمان اجرای الگوریتم بازشناسی تومور، باعث کاهش درصد تشخیص مثبت نادرست نیز می‌شود. این الگوریتم شامل یافتن مرز سینه و مرز پکتورال به‌منظور محدود نمودن فضای اولیه به خود سینه و سپس بدست آوردن نواحی مشکوک در سینه می‌باشد. اجرای مراحل ۹ گانه پیشنهادی باعث

۵- طبقه‌بندی بافت با استفاده از SVM

یکی از انواع ابزارهایی که برای کلاس‌بندی داده‌ها به دو دسته مختلف در سال‌های اخیر بسیار استفاده می‌شود ماشین بردار پشتیبان (SVM¹⁷) می‌باشد. البته به دلیل آن‌که خروجی این شبکه‌ها صفر یا یک می‌باشد نمی‌توان برای خروجی آنها منحنی ROC ترسیم کرد. بنابراین فقط درصد تشخیص مثبت صحیح و درصد تشخیص مثبت کاذب را ذکر می‌نماییم. این روش با استفاده از یک هسته، داده‌ها را به فضایی می‌برد که در آن فضا می‌توان داده‌ها را توسط یک فوق صفحه به دو دسته جداگانه تقسیم کرد. سپس در آن فضای داده فوق صفحه‌ای را که بهترین تقسیم‌بندی را انجام می‌دهد و حتی‌الامکان با فاصله یکسان از دو دسته قرار می‌گیرد را می‌یابد. هسته‌ای که در این تحقیق پس از بررسی چندین هسته انتخاب شد، هسته چندجمله‌ای از درجه دو می‌باشد. با استفاده از این هسته و با استفاده از تمام ویژگی‌ها به درصد تشخیص مثبت صحیح ۰/۹۴۶ و درصد تشخیص مثبت کاذب ۰/۲۳ رسیدیم. نکته جالب در مورد این شبکه این است که با استفاده از ویژگی‌هایی که توسط الگوریتم ژنتیک و یا براساس انترویی انتخاب شده بودند نیز به همان درصدهای مربوط به استفاده از کل ویژگی‌ها

- [5] Kurani A.S, Xu D.H, Frust J; **“Co-Occurrence Matrices for Volumetric Data”**, Computer Graphics and Imaging CGIM 2004, Vol. 12, No. 3, pp. 426-430, 2004.
- [6] Martin, Hagan, Howard, Demuth, Mark H. Beale. **Neural Network Design**. Oklahoma State University, 2002.
- [7] Haralick R.M, Shanmugam K, and Dinstein I; **“Textural Features for Image Classification”**, IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-3, No. 6, pp. 610-621, 1973.

[۸] سید مسعود هاشمی عمروآبادی، طراحی سیستم بازشناسی اتوماتیک تومورهای سرطانی از روی تصاویر ماموگرام، پایان نامه کارشناسی ارشد، دانشکده برق و کامپیوتر، دانشگاه صنعتی اصفهان، ۱۳۸۷.

[۹] سید مسعود هاشمی عمروآبادی، محمدرضا احمدزاده، علی حکمت نیا **“آشکار سازی مرز سینه در تصاویر ماموگرام با استفاده از روش‌های مبتنی بر بخش‌بندی بافت و آستانه گذاری”**، پنجمین کنفرانس ماشین بینایی و پردازش تصویر ایران، دانشگاه تبریز، ص ۳۲۹-۳۳۵، ۱۴-۱۶ آبان ۱۳۸۷.

[۱۰] سید مسعود هاشمی عمروآبادی، محمدرضا احمدزاده، علی حکمت نیا، **“تشخیص خودکار مرز ماهیچه پکتورال در تصاویر ماموگرام با استفاده از مدل ترکیب گوسی و مدل مارکوف”**، پنجمین کنفرانس ماشین بینایی و پردازش تصویر ایران، دانشگاه تبریز، ص ۳۳۶-۳۴۲، ۱۴-۱۶ آبان ۱۳۸۷.

۹- پی‌نوشت‌ها

- 1- Region of Interest (ROI)
- 2- Region Growing
- 3- Contrast
- 4- Inverse Difference Moment (IDM)
- 5- Homogeneity
- 6- SumMean
- 7- Correlation
- 8- Maximum Probability (MP)
- 9- Cluster Tendency
- 10- Texture Probability of Run Length of 2
- 11- Intensity
- 12- Kurtosis
- 13- Skewness
- 14- Impulse Factor
- 15- Receiver Operating Characteristic (ROC) curve
- 16- Information Gain (IG)
- 17- Support Vector Machine

گردید درصد نواحی مشکوک اشتباه نسبت به روشی که در اکثر مقالات از آن استفاده نموده‌اند بطور متوسط از ۹ به ۴ در هر تصویر کاهش یابد.

بکارگیری آمارگان مرتبه اول و دوم ماتریس هم‌رخداد به‌عنوان ویژگی و نیز برخی ویژگی‌های خود تصویر و هیستوگرام آن توانست درصد تشخیص مثبت صحیح ۰/۹۴۶ و درصد تشخیص مثبت کاذب ۰/۲۳ ارائه نماید که نسبت به کارهای مشابه نتیجه بسیار خوب و قابل قبولی می‌باشد. علاوه بر ویژگی‌های مورد استفاده طبقه‌بندی‌کننده ماشین بردار پشتیبان در بدست آوردن این نتیجه بسیار مؤثر می‌باشد به‌نحوی که دیگر طبقه‌بندی‌کننده‌ها نتیجه پائین‌تری ارائه می‌دادند.

تحقیقات به‌منظور افزایش درصد تشخیص مثبت صحیح و کاهش درصد تشخیص مثبت کاذب ادامه دارد. این تحقیقات در زمینه استخراج دیگر انواع ویژگی‌ها و ترکیب آنها و نیز ترکیب طبقه‌بندی‌کننده‌ها می‌باشد.

۷- تقدیر و تشکر

این تحقیق با پشتیبانی مالی مرکز تحقیقات پردازش تصویر و سیگنال پزشکی دانشگاه علوم پزشکی اصفهان و امکانات تحقیقاتی دانشکده برق و کامپیوتر دانشگاه صنعتی اصفهان انجام شده است.

۸- مراجع

- [1] F. Fauchi, G. Raso, Rl. Magro; **“A Massive Lesion Detection Algorithm in Mammo-Graphy”**, Physica Medica, Vol. 21, No. 1, pp. 21-28, 2005.
- [2] H. Qi, W. E Snyder; **“Lesion Detection and Characterization in Digital Mammo-Graphy by Bezier Histograms”**, The 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, No. 29, pp. 1021-1024, 1998.
- [3] H. Farzin, H. Abrishami Moghaddam, Moin, MohammadShahram. **“New Method for Blood Vessels Segmentation in Retinal Images”** Majlesi Journal of Electrical Engineering [Online], Vol. 1 No. 1, 8 June 2009.
- [4] Magoulas G.D, Karkanis S.A, Karras D.A, and Vrahatis M.N; **“Comparison Study of Descriptors for Training Neural Network classifier”**, The 19th Annual International Conference of the IEEE Engineering in Medicine and Biology society, Vol. 3, No. 30, pp. 1359, 1362, 1997.