

روشی جدید برای ناحیه بندی تصاویر ماموگرافی دیجیتال و طبقه بندی تومورهای خوش خیم و بدخیم در سینه

حمزه حسن نیا^۱، مهدی چهل امیرانی^۲ و مرتضی ولی زاده^۳

چکیده

ماموگرافی رایج ترین و مؤثرترین روش غربالگری برای تشخیص سرطان پستان است. در این تحقیق، یک سیستم کمکی برای طبقه بندی تومورهای خوش خیم و بدخیم در تصاویر ماموگرافی دیجیتال ارائه شده است. در این روش ابتدا فیلتر میانه برای حذف نویز استفاده شده و سپس مصنوعات و ماهیچه‌ی پکتورال در صورت وجود حذف می‌شوند. برای ناحیه بندی ماموگرام و استخراج ناحیه‌های مورد نظر ابتدا یک الگوریتم جدید برای افزایش تابین نواحی مشکوک ارائه شده است که از تفاضل بهبود یافته تصویر اصلی و مکمل آن بهره می‌برد، سپس الگوریتم خوشه بندی C میانگین فازی بر مبنای هیستوگرام به تصویر اعمال شده و ناحیه‌های مورد نظر با دقتی مناسب استخراج می‌شوند. در مرحله‌ی بعد ویژگی‌های بافت و هندسی استخراج می‌شوند و در نهایت طبقه بندی‌های ماشین بردار پشتیبان خطی و درخت تصمیم برای دسته بندی ناحیه‌های مورد نظر به دو کلاس خوش خیم و بدخیم، استفاده می‌شوند. سیستم پیشنهادی بر روی تصاویر پایگاه‌های داده‌ی MIAS و DDSM آزمایش شده است. نتایج به دست آمده نشانگر این است که دقت سیستم پیشنهادی در مقایسه با تحقیقات پیشین امیدوار کننده است.

کلیدواژه‌ها

سرطان سینه، ماموگرافی، افزایش تابین، ناحیه بندی، استخراج ویژگی

۱ مقدمه

نوع بیماری، تشخیص دیر هنگام آن است. در ایران، این نوع سرطان شایع ترین سرطان در بین زنان بوده و سالانه حدود ۷۰۰۰ زن ایرانی به این سرطان مبتلا می‌شوند [۲-۳]. متأسفانه در کشور ما سن ابتلا به این سرطان ۵ تا ۱۰ سال پایین تر از میانگین جهان است [۴]. تشخیص زودهنگام سرطان سینه باعث تسهیل فرآیند درمان و افزایش احتمال درمان موفق می‌شود. به عبارت دیگر، هر چقدر این سرطان زودتر تشخیص داده شود، شانس زنده ماندن بیمار به صورت قابل توجهی افزایش می‌یابد. تشخیص به موقع (حداکثر ۵ سال پس از اولین تقسیم سلول سرطانی) شانس زنده ماندن بیمار را از ۵۶ درصد به ۸۶ درصد افزایش می‌دهد [۵-۶]. در حال حاضر ماموگرافی رایج ترین روش تصویربرداری از سینه است. این روش به تنهایی نرخ مرگومیر ناشی از سرطان را به اندازه‌ی ۲۵ تا ۳۰ درصد کاهش داده است [۷]. تفسیر تصاویر ماموگرافی به تجربه‌ی فردی وابسته است و

سرطان سینه یک بیماری رایج در کشورهای مختلف جهان محسوب می‌شود. طبق آمارهای منتشر شده، سرطان سینه بعد از سرطان ریه رایج ترین سرطان و پنجمین عامل مرگ‌های ناشی از سرطان محسوب می‌شود [۱]. بیشترین میزان مرگومیر در سنین ۴۹-۱۵ سال اتفاق می‌افتد. یکی از دلایل عدم درمان موفق این

این مقاله در اسفندماه ۱۳۹۷ دریافت، دومین بازنگری در اسفندماه ۱۳۹۸ و در فروردین ماه سال ۱۳۹۹ پذیرفته شد.

^۱ دانشجوی کارشناسی ارشد مهندسی برق و کامپیوتر، دانشگاه ارومیه

رایانامه: h.hassannia@urmia.ac.ir

^۲ دانشکده مهندسی برق و کامپیوتر، دانشگاه ارومیه

رایانامه: m.amirani@urmia.ac.ir و mo.valizadeh@urmia.ac.ir

دو جهت برای استخراج ویژگی از تصاویر ماموگرام استفاده کرده‌اند. نهایتاً برای دسته‌بندی بردارهای ویژگی استخراج شده، طبقه‌بندهای مختلفی از جمله ماشین بردار پشتیبان، شبکه‌های عصبی مصنوعی، K نزدیکترین همسایه و ... مورد استفاده قرار گرفته‌اند.

در مرجع [۲۰] پس از ناحیه‌بندی تصویر و یافتن نواحی مشکوک، از آن نواحی ویژگی‌های بافتی و هندسی استخراج می‌شود و از الگوریتم خوشه‌بندی FCM برای دسته‌بندی نواحی استفاده شده است. همچنین در تحقیقاتی FCM برای ناحیه‌بندی ماموگرام مورد استفاده قرار گرفته است. در [۲۱] از ترکیب عملگرهای مورفولوژی برای بهسازی تصویر استفاده شده است و از FCM برای ناحیه‌بندی تصویر استفاده شده است همچنین در [۲۲] از تبدیل ویولت گسسته برای بهسازی تصویر و از FCM برای ناحیه‌بندی تصویر استفاده شده است. مشکل اصلی استفاده از این الگوریتم خوشه‌یابی برای ناحیه‌بندی حجم محاسبات زیاد آن است.

در این مقاله، یک سیستم خودکار برای پیش‌بینی خوش‌خیم یا بدخیم بودن تومور سینه طراحی و ارزیابی می‌شود که از ترکیب ویژگی‌های بافت و هندسی بهره می‌برد. این سیستم با پیش‌بینی خودکار نوع تومور می‌تواند در انتخاب روش درمانی مناسب به پزشک کمک کند. با توجه به اینکه تصاویر ماموگرافی به صورت ذاتی نویزدار هستند، در روش پیشنهادی ابتدا نویز موجود در تصاویر ماموگرام کاهش داده می‌شود و سپس با استفاده از تکنیک‌های پردازش تصویر از جمله آستانه‌یابی و عملگرهای شکل‌شناسی مصنوعات^۸ و ماهیچه‌ی پکتورال^۹ در صورت وجود آن‌ها، حذف می‌شوند. برای ناحیه‌بندی ماموگرام و استخراج ناحیه‌های مورد نظر ابتدا از یک الگوریتم برای افزایش تباین نواحی مشکوک بهره گرفته و سپس از الگوریتم خوشه‌بندی C میانگین فازی بر مبنای هیستوگرام^{۱۰} استفاده می‌شود. پس از استخراج نواحی مورد نظر، تعدادی ویژگی بافتی و هندسی از این ناحیه‌ها استخراج شده و نهایتاً این نواحی با استفاده از طبقه‌بندهای مختلف، به دو کلاس خوش‌خیم و بدخیم دسته‌بندی می‌شوند و کارایی طبقه‌بندها و ویژگی‌های مورد استفاده ارزیابی می‌شود. در بخش دوم مقاله، سیستم پیشنهادی به طور کامل مورد بحث قرار می‌گیرد و در بخش سوم نتایج پیاده‌سازی و بحث‌های مربوط به آن آورده می‌شود. نهایتاً بخش چهارم به نتیجه‌گیری اختصاص داده شده است.

همچنین احتمال عدم تشخیص تومورهای سرطانی در اثر خستگی و خطای بصری رادیولوژیست وجود دارد. تفسیر تصاویر ماموگرافی کار آسانی نیست و بر اساس آمار مرکز ملی سرطان، ۱۰ تا ۳۰ درصد از تومورهای موجود در تصاویر ماموگرافی، توسط رادیولوژیست‌ها قابل تشخیص نیستند [۸-۹]. بنابراین ابزارهایی که در تشخیص، پیش‌بینی و طبقه‌بندی تومور سینه به رادیولوژیست کمک می‌کنند، مورد نیاز هستند. بدین منظور سیستم‌های تشخیص کمک کامپیوتری^۱ (CAD) ارائه می‌شوند که با استفاده از این سیستم‌ها می‌توان نظر ثانویه‌ای را برای کمک و پشتیبانی از تصمیمات، به پزشک و متخصصان ارائه داد [۱۰].

تاکنون تحقیقات بسیاری برای طراحی یک سیستم CAD جهت تشخیص و طبقه‌بندی تومورهای سینه در تصاویر ماموگرافی انجام شده است. کشف توده یک مرحله‌ی مهم و مشکل چالش برانگیزی در سیستم‌های CAD مورد استفاده برای تشخیص سرطان پستان است که تا کنون نیز کاملاً حل نشده است [۱۱]. معمولاً در این سیستم‌ها، ابتدا ناحیه‌های مورد نظر^۲ (ROI) کشف می‌شوند سپس ناحیه و مرز توده با استفاده از روش‌های مختلفی مشخص شده و در نهایت طبقه‌بندی این نواحی صورت می‌گیرد. تا کنون از روش‌های مختلفی برای ناحیه‌بندی ماموگرام استفاده شده است. در مرجع [۱۲] از تلفیق تبدیل موجک و عملیات شکل‌شناسی^۳ برای تشخیص تومور بهره گرفته‌اند. همچنین مرجع [۱۳] پس از بهبود ماموگرام، از سه روش آستانه‌یابی اتسو^۴ [۱۴]، آستانه‌یابی محلی و خوشه‌بندی k-means برای ناحیه‌بندی ماموگرام و استخراج نواحی مشکوک استفاده کرده است. Sun و همکاران [۱۵] به منظور تشخیص توده روشی را پیشنهاد کردند که در آن از یک الگوریتم خوشه‌بندی C میانگین فازی^۵ تطبیقی استفاده می‌شود. در گام بعدی، برای دسته‌بندی تومورها لازم است ویژگی‌هایی از تومور استخراج شوند که بتوانند بین دو دسته‌ی مورد نظر تمایز ایجاد کنند. این ویژگی‌ها مربوط به روشنایی، بافت، مرز و شکل تومور هستند. Verma و همکارانش [۱۶] از ویژگی‌های مربوط به مرز و شکل تومور استفاده کرده‌اند. رحمانی و همکارانش [۱۷] نیز از دو گروه ویژگی‌های بافت‌شناسی و ریخت‌شناسی استفاده کرده‌اند، که برای استخراج ویژگی‌های ریخت‌شناسی از توصیف‌گرهای مرز تومور بهره گرفته‌اند.

در مرجع [۱۸] نیز از ویژگی‌های هندسی از جمله ویژگی‌های استخراجی از شکل تومور و ویژگی‌های بافتی به دست آمده از ماتریس‌های هم‌رخداد^۶ و طول تکرار^۷ سطوح خاکستری استفاده شده است. Midya و همکاران [۱۹] هم از ماتریس هم‌رخداد در

¹ Computer-aided diagnosis system

² Region of interest

³ Morphological operation

⁴ Otsu thresholding

⁵ Fuzzy c-means clustering

⁶ Grey-level co-occurrence matrix

⁷ Grey-level run length matrix

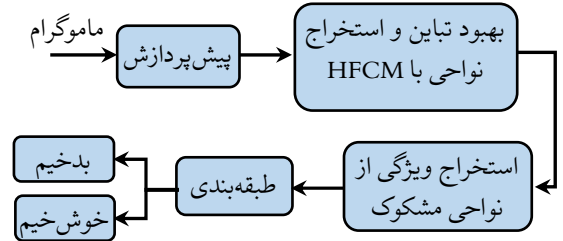
⁸ Artifacts

⁹ Pectoral muscle

¹⁰ Histogram based fuzzy c-means clustering

۲ روش پیشنهادی

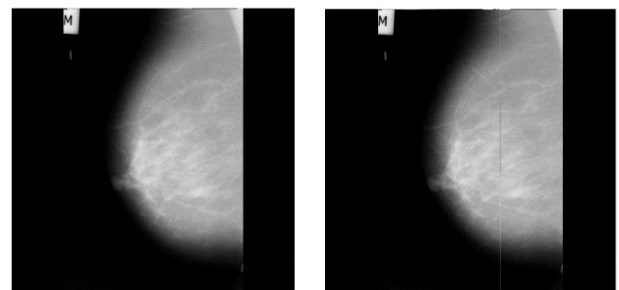
سیستم تشخیص کامپیوتری پیشنهادی از چهار قسمت اصلی تشکیل شده است. بلوک دیاگرام آن در شکل (۱) نشان داده شده است و تمام اجزای سیستم در این بخش با جزئیات کامل تشریح می شود.



شکل ۱: بلوک دیاگرام مراحل اصلی سیستم پیشنهادی

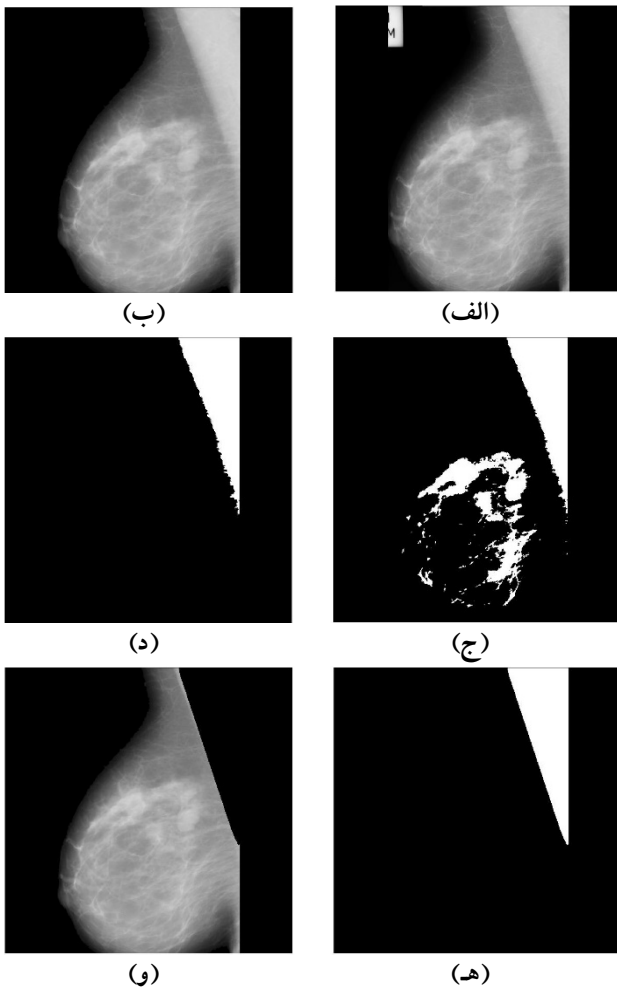
۲-۱ پیش پردازش

مرحله‌ی پیش پردازش شامل سه مرحله‌ی حذف نویز، حذف مصنوعات و حذف ماهیچه‌ی پکتورال در صورت وجود می باشد که در این بخش این مراحل با جزئیات نشان داده می شوند. دستگاه‌های تصویربرداری موجود و همچنین استفاده از اسکترهای معمولی که برای دیجیتال کردن تصاویر مورد استفاده قرار می گیرند، خطای گسسته سازی بزرگی ایجاد می کنند. بنابراین تصاویر ماموگرافی مورد استفاده شامل نویزهای زیادی از قبیل نویز حاصل از دیجیتالی کردن تصویر و نویز حاصل از فرآیند تصویربرداری می باشند که بر عملکرد طبقه بندی تأثیرگذار است. در واقع اگر نویز حذف نشود ویژگی‌ها تحت تأثیر نویز خواهند بود، لذا حذف نویز امری حیاتی است. در این تحقیق، نویز تصاویر با استفاده از فیلتر میانه با همسایگی 3×3 کاهش داده می شود. فیلتر میانه یک ابزار قدرتمند در حذف نویز سیگنال‌های دو بعدی است. لبه‌ها فاکتور بسیار مهمی در ناحیه بندی محسوب می شوند، به این علت که فیلتر میانه این قابلیت را دارد که بدون از بین بردن لبه‌ها نویز را حذف کند، نتایج نسبتاً بهتری را نسبت به سایر روش‌های حذف نویز ایجاد می کند [۲۳]. نتیجه‌ی اعمال فیلتر میانه به تصویر ماموگرافی در شکل (۲) نشان داده شده است.



شکل ۲: اعمال فیلتر میانه به تصویر ماموگرافی. الف) تصویر اصلی، ب) تصویر بعد از اعمال فیلتر میانه

اغلب تصاویر ماموگرافی دارای مصنوعات از قبیل برجسب‌ها و حاشیه‌هایی در بالای تصویر هستند. به منظور حذف این مصنوعات و استخراج ناحیه‌ی سینه، ابتدا تصویر را با استفاده از یک آستانه‌ی تجربی ($T=18$)، باینری کرده و تمام اشیا موجود در تصویر باینری برجسب گذاری می شوند، سپس مساحت همه‌ی آن‌ها محاسبه و بزرگترین شی که مربوط به ناحیه‌ی سینه است، انتخاب می شود. سپس از عملگرهای شکل شناسی به منظور پر کردن حفره‌های موجود در ناحیه سینه استفاده می شود و بدین ترتیب ماسک سینه بدست می آید. نهایتاً با اعمال ماسک حاصل و حذف اطلاعات خارج از ناحیه سینه در تصویر اصلی، مصنوعات از ماموگرام حذف می شوند.



شکل ۳: حذف ماهیچه‌ی پکتورال از تصویر ماموگرافی. الف) تصویر اصلی، ب) تصویر بعد از اعمال فیلتر میانه و حذف برجسب، ج) تصویر باینری با آستانه‌ی اتسو، د) اشیا (Object) انتخاب شده در تصویر باینری با توجه به مکان ماهیچه‌ی پکتورال، ه) بدست آوردن پوش محدب شی، ه) تصویر ماموگرافی که در آن ماهیچه‌ی پکتورال حذف شده است.

همچنین در برخی از تصاویر ماموگرافی یک منطقه‌ی روشن به شکل مثلث و متراکم در گوشه‌ی سمت بالای تصویر مشاهده می شود که به سمت قفسه سینه و تا پایین منطقه‌ی اصلی سینه کشیده می شود و به آن ماهیچه‌ی پکتورال می گویند. این ماهیچه از

استخراج می‌شوند. نتایج ناحیه‌بندی بر روی دو تصویر از پایگاه‌های داده مورد استفاده در شکل‌های (۵) و (۶) نشان داده شده است.

الگوریتم خوشه‌بندی C میانگین فازی، یکی از روش‌های پرکاربرد در ناحیه‌بندی تصاویر ماموگرافی است. یکی از معایب این روش، زیاد بودن محاسبات مورد نیاز برای اجرای آن است. زیرا این الگوریتم برای ناحیه‌بندی تصویر، مرکز و تابع درجه‌ی عضویت را پیکسل به پیکسل محاسبه می‌کند که کار زمان‌بری است و ناحیه‌بندی تصویر را غیرعملی می‌کند. برای حل این مشکل می‌توان از خوشه‌بندی C میانگین فازی بر مبنای هیستوگرام استفاده کرد و به جای خوشه‌بندی پیکسل‌ها، سطوح روشنایی که تعداد آن‌ها خیلی کمتر از تعداد پیکسل‌ها است، خوشه‌بندی می‌شوند. با این کار تعداد محاسبات مورد نیاز برای ناحیه‌بندی تصویر به طور چشم‌گیری کاهش می‌یابد. بنابراین در این مقاله از روش خوشه‌بندی بر مبنای هیستوگرام برای استخراج نواحی مشکوک به تومور استفاده شده است. تابع هزینه الگوریتم خوشه‌بندی C میانگین فازی به صورت زیر تعریف می‌شود [۲۵]:

(۱)

که در آن H هیستوگرام تصویری با L سطح خاکستری است. همچنین در رابطه (۱)، I بیانگر سطوح روشنایی، v_k نماینده یا مرکز خوشه‌ی kام، L تعداد سطوح خاکستری و u_{kl} میزان تعلق سطح روشنایی lام در خوشه‌ی kام را نشان می‌دهد. علامت $\| \cdot \|$ میزان تشابه (فاصله) نمونه با (از) مرکز خوشه می‌باشد که می‌توان از هر تابعی که بیانگر تشابه نمونه و مرکز خوشه باشد، استفاده کرد. همچنین m یک عدد حقیقی بزرگ‌تر از یک است که در اکثر موارد برای m عدد ۲ انتخاب می‌شود. در واقع در این روش محاسبات درجه‌ی عضویت H(l) پیکسل، تنها به محاسبه‌ی درجه عضویت سطح خاکستری lام کاهش می‌یابد. برای کمینه کردن تابع هزینه در خوشه‌بندی C میانگین فازی بر مبنای هیستوگرام، تابع درجه عضویت و مرکز خوشه‌ها مطابق روابط (۲) و (۳) به‌طور تکراری محاسبه می‌شوند و در هر مرحله، تابع هزینه توصیف شده با رابطه (۱) نسبت به مرحله قبل کاهش می‌یابد. این فرآیند تا همگرایی الگوریتم خوشه‌بندی ادامه پیدا می‌کند.

$$u_{kl} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{kl}}{d_{jl}} \right)^2} \quad (2)$$

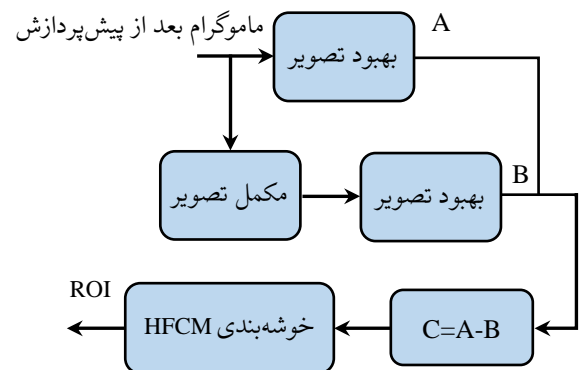
$$v_k = \frac{\sum_{l=1}^L u_{kl}^2 H(l) l}{\sum_{l=1}^L u_{kl}^2} \quad (3)$$

که در رابطه‌ی (۲)، d_{kl} بیانگر فاصله‌ی بین مرکز خوشه‌ی kام و سطح خاکستری lام است.

نظر روشنایی و خصوصیات بافتی به نواحی آسیب دیده و سرطانی بسیار شبیه است. همچنین تصاویر ماموگرام دارای ابعاد بزرگی هستند که پردازش کل تصویر را برای یافتن تومور زمان‌بر می‌کند، لذا به منظور کاهش نواحی مورد پردازش، حذف ماهیچه‌ی پکتورال از بافت سینه ضروری به نظر می‌رسد. بدین منظور ماموگرام را با استفاده از روش آستانه‌یابی اتسو باینری کرده و تمام اشیا موجود در تصویر باینری را برچسب‌گذاری کرده و مرکز همه‌ی آن‌ها تعیین می‌شود. با توجه به اینکه این ماهیچه همیشه در بالای تصویر و در سمت راست یا چپ قرار دارد، آن شی که در داخل این ناحیه قرار دارد را انتخاب کرده و سپس پوش محدب آن محاسبه می‌شود و نهایتاً اطلاعات موجود در پوش محدب از ماموگرام حذف می‌شوند. این مراحل در شکل (۳) نشان داده شده است.

۲-۲ ناحیه‌بندی

پس از مرحله‌ی پیش‌پردازش، تصویر آماده‌ی ناحیه‌بندی می‌شود. هدف ناحیه‌بندی استخراج نواحی مورد نظر است، طوری که شامل تمام کاندیدهای مشکوک توده در تصویر ماموگرام باشد. بلوک دیاگرام الگوریتم پیشنهادی برای ناحیه‌بندی ماموگرام و استخراج نواحی مشکوک در شکل (۴) نشان داده شده است.



شکل ۴: بلوک دیاگرام الگوریتم پیشنهادی برای ناحیه‌بندی تصاویر ماموگرافی

هدف از مرحله‌ی اول الگوریتم پیشنهادی برای ناحیه‌بندی، افزایش تباین^۲ نواحی مشکوک نسبت به سایر پیکسل‌ها و تسهیل ناحیه‌بندی است. در این مرحله ابتدا تباین تصویر حاصل از مرحله‌ی پیش‌پردازش با استفاده از تعدیل تطبیقی هیستوگرام^۳ [۲۴]، بهبود داده می‌شود. همچنین مکمل تصویر مذکور را تولید کرده و تباین آن نیز با استفاده از تعدیل تطبیقی هیستوگرام، بهبود داده می‌شود و سپس تفاضل دو تصویر حاصل از بهبود تباین، محاسبه می‌شود. تفاضل دو تصویر نواحی تومور را برجسته‌تر می‌سازد. نهایتاً با اجرای الگوریتم خوشه‌بندی C میانگین فازی بر مبنای هیستوگرام بر روی تصویر حاصل از تفاضل، نواحی مشکوک

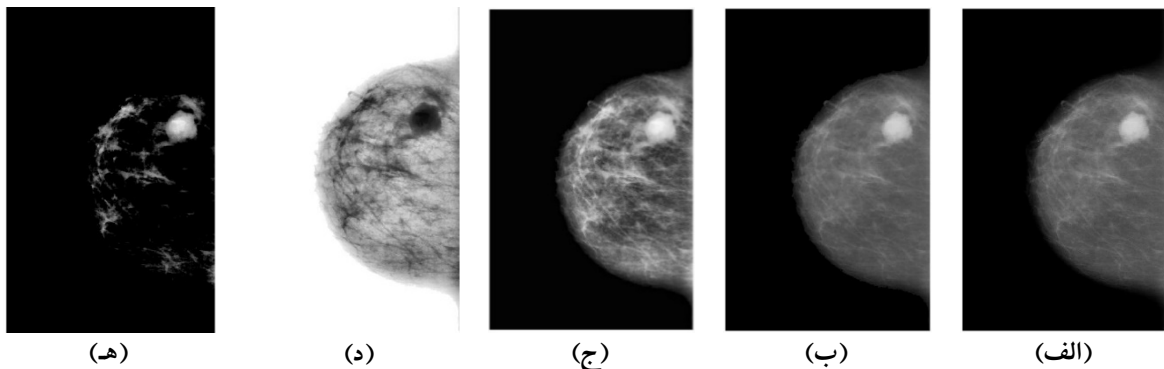
¹ Convex hull

² Contrast

³ Adaptive histogram equalization



شکل ۵: اجرای مرحله اول الگوریتم ناحیه‌بندی نشان داده شده در شکل ۴ و آماده کردن تصویر جهت اعمال خوشه‌بندی. (الف) تصویر اصلی از پایگاه داده‌ی MIAS، (ب) تصویر اصلی بعد از عملیات پیش‌پردازش، (ج) نتیجه‌ی بهبود تباین تصویر (ب)، (د) مکمل تصویر (ب) بعد از بهبود تباین، (ه) نتیجه‌ی تفریق دو تصویر (ج) و (د).



شکل ۶: اجرای مرحله اول الگوریتم ناحیه‌بندی نشان داده شده در شکل ۴ و آماده کردن تصویر جهت اعمال خوشه‌بندی. (الف) یک نمونه تصویر از پایگاه داده‌ی DDSM، (ب) تصویر اصلی بعد از عملیات پیش‌پردازش، (ج) نتیجه‌ی بهبود تباین تصویر (ب)، (د) مکمل تصویر (ب) بعد از بهبود تباین، (ه) نتیجه‌ی تفریق دو تصویر (ج) و (د).

در نهایت با اعمال عملگر شکل‌شناسی بازکردن^۳ و بستن^۴ به تصویر باینری، به ترتیب مؤلفه‌های مزاحم حذف شده و حفره‌های موجود در ناحیه مشکوک پر می‌شوند و یک تصویر باینری تولید می‌شود که در آن نواحی مشکوک با دقت خوبی استخراج شده است. بعد از استخراج تومور، ناحیه‌ی مورد نظر که شامل تومور و ناحیه‌ی اطراف آن است، استخراج می‌شود. از ناحیه مورد نظر سه دسته ویژگی بر اساس شکل، مرز و بافت تومور استخراج می‌شوند. در این تحقیق، برای استخراج مرز تومور از عملگر لبه‌یاب کنی استفاده شده است. ناحیه‌های مورد نظر استخراج شده از تصاویری که الگوریتم ناحیه‌بندی بر روی آن‌ها اجرا شده است، در شکل (۸) نمایش داده شده است.

۲-۳ استخراج ویژگی

یکی از مهم‌ترین مراحل در سیستم‌های تشخیص کامپیوتری جهت کلاس‌بندی تومورها، مرحله‌ی استخراج ویژگی است چرا که مجموعه‌های ویژگی مختلف بر عملکرد طبقه‌بندی تأثیر می‌گذارد

در این تحقیق تعداد خوشه‌ها به صورت تجربی برای تصاویر پایگاه‌های داده MIAS^۱ [۲۶] و DDSM^۲ [۲۷]، برابر با ۴ تنظیم شده است. بنابراین پس از انجام خوشه‌بندی، پیکسل‌های تصاویر پایگاه‌های داده فوق ۴ خوشه تقسیم می‌شوند. شکل (۷) نتیجه اعمال خوشه‌بندی روی یک تصویر نمونه از پایگاه داده DDSM پس از بهسازی آن را نشان می‌دهد که در هر یک از تصاویر، پیکسل‌های متعلق به یک خوشه نشان داده شده است. روشن‌ترین خوشه بیانگر نواحی مشکوک است و سایر خوشه‌ها، پیکسل‌های با سطوح روشنایی متوسط در تصویر پردازش شده را نشان می‌دهند.

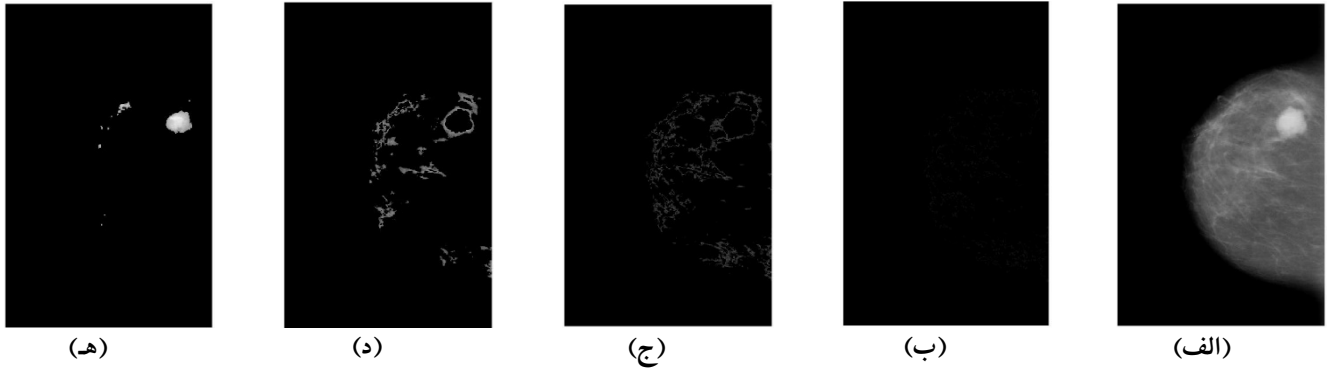
با توجه به اینکه نواحی سرطانی در تصویر ماموگرافی از سایر نواحی روشن‌تر است، خوشه‌ای که دارای روشن‌ترین مرکز (سطح شدت بیشتر) است، به عنوان خوشه متناظر با ناحیه مشکوک انتخاب می‌گردد. سپس با تخصیص سطح روشنایی ۱ به پیکسل‌های قرار گرفته در روشن‌ترین خوشه و تخصیص سطح روشنایی ۰ به سایر پیکسل‌ها، یک تصویر باینری تولید می‌شود.

³ Opening

⁴ Closing

¹ Mammographic Image Analysis Society

² Digital Database for Screening Mammography



شکل ۷: (الف) یک نمونه تصویر از مجموعه داده DDSM (ب-ه) نتیجه اعمال خوشه‌بندی HFCM روی تصویر (الف)

$$d(n) = \frac{\sqrt{(x(n)-x_0)^2 + (y(n)-y_0)^2}}{\max(d(n))}, n=1,2,\dots,N \quad (4)$$

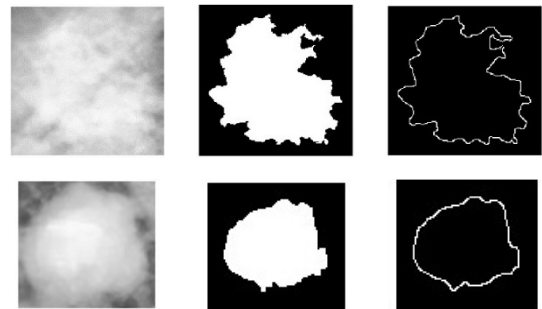
در رابطه‌ی بالا x_0 و y_0 مختصات مرکز تومور، $(x(n), y(n))$ مختصات پیکسل مرزی n ام، N تعداد کل پیکسل‌های مرزی و $\max(d(n))$ بزرگترین طول شعاعی استخراج شده است. ویژگی‌های تعریف شده بر اساس فاصله‌ی شعاعی نرمالیزه شده در جدول (۱) آورده شده است [۳۰]. همچنین به منظور استخراج ویژگی‌های مرتبط با شکل تومور، نیاز به محاسبه‌ی مساحت و محیط خود تومور و پوش محدب تومور است. پوش محدب تومور عبارت است از کوچکترین چند ضلعی محدبی که تومور را در بر می‌گیرد. مساحت تومور و پوش محدب تومور که به ترتیب با A و A_c در فرمول‌های زیر نشان داده شده‌اند برابر با مجموع تعداد پیکسل‌های تومور و پوش محدب آن در تصویر باینری هستند. همچنین محیط تومور و پوش محدب آن که به ترتیب با P و P_c نشان داده می‌شوند برابر با مجموع پیکسل‌های مرزی تومور و مرز پوش محدب آن می‌باشند. ویژگی‌های استخراج شده از شکل تومور نیز در جدول (۲) آورده شده است [۳۱-۳۳].

جدول (۱): ویژگی‌های مرزی استفاده شده در این مقاله

میانگین طول شعاعی نرمالیزه شده	$d_{avg} = \frac{1}{N} \sum_{n=1}^N d(n)$
انحراف استاندارد طول شعاعی نرمالیزه شده	$\frac{1}{N-1} \sqrt{\sum_{n=1}^N (d(n)-d_{avg})^2}$
شاخص ناهمواری	$Roughness = \frac{1}{N} \sum_{n=1}^N d(n)-d(n+1) $
نسبت مساحت طول شعاعی	$Ratio\ of\ area = \frac{1}{d_{avg} \cdot N} \sum_{n=1}^N (d(n)-d_{avg})$

و عملکرد این سیستم‌ها شدیداً به نحوه‌ی استخراج ویژگی وابسته است. استخراج ویژگی‌های مفیدی که قادر به تفکیک طبقات خوش‌خیم و بدخیم هستند بسیار مهم است [۲۸]. ویژگی‌های استخراج شده باید حاوی اطلاعات مفیدی از تومورها بوده و توصیف‌کننده‌های خوبی از اطلاعات موجود در شکل، مرز و بافت تومورها باشند.

تومورهای با شکل دایروی و بیضوی با مرز صاف، محدود و مشخص غالباً خوش‌خیم هستند درحالی‌که تومورهای بدخیم معمولاً مرز سوزنی شکل، ناصاف و کدر دارند [۲۹]. در این تحقیق از دو گروه ویژگی‌های هندسی و بافت استفاده شده است. ویژگی‌های هندسی که شامل ویژگی‌های شکل و ویژگی‌های مرز هستند، مشخصات هندسی و مرزی تومور را توصیف می‌کنند. ویژگی‌های بافت مورد استفاده نیز شامل دو دسته ویژگی مستخرج از ماتریس هم‌رخداد و ماتریس طول تکرار سطوح خاکستری است.



شکل ۸: ناحیه‌های مورد نظر استخراجی شامل بافت، شکل و مرز تومور

۱-۳-۲ ویژگی‌های هندسی

برای استخراج ویژگی‌های مرتبط با مرز تومور، مختصات پیکسل‌های مرزی تعیین شده و یک معیار فاصله‌ی شعاعی با استفاده از آنها تعریف می‌شود. برای محاسبه‌ی فاصله‌ی شعاعی ابتدا مختصات مرکز تومور تعیین می‌شود، سپس فاصله‌ی اقلیدسی نقاط مرزی از مرکز تومور اندازه‌گیری شده و در نهایت نرمالیزه می‌گردد.

درایه‌ی $q(i,r)$ ، فراوانی رخدادهایی را نشان می‌دهد که در آنها r پیکسل با شدت روشنایی i و در جهت مزبور در کنار هم قرار گرفته‌اند. متغیر i مربوط به شدت روشنایی و محدوده آن متناسب با ماکزیمم شدت روشنایی است و متغیر r مربوط به طول تکرار است که محدوده آن با حداکثر مقدار ممکن برای طول تکرار تعیین می‌شود. در این تحقیق برای هر ناحیه‌ی موردنظر ماتریس طول تکرار در جهت ۰ درجه به دست آمده و سپس ۱۱ ویژگی مطابق تحقیق [۳۵] استخراج شده است.

جدول (۳): ویژگی‌های بافت استخراج شده از ماتریس هم‌رخداد سطح خاکستری (GLCM)

کنتراست (Contrast)	$CON = \sum_{i=1}^n \sum_{j=1}^n (i-j)^2 p(i,j)$
انرژی (Energy)	$ENR = \sum_{i=1}^n \sum_{j=1}^n p(i,j)^2$
همگنی (Homogeneity)	$HOM = \sum_{i=1}^n \sum_{j=1}^n \frac{p(i,j)}{1+ i-j }$
همبستگی (Correlation)	$COR = \frac{\sum_{i=1}^n \sum_{j=1}^n (i-j)(i+j)p(i,j)}{n_1 n_2}$
آنترپی (Entropy)	$ENT = -\sum_{i=1}^n \sum_{j=1}^n p(i,j) \log_2 p(i,j)$

جدول (۴): ویژگی‌های بافت استخراج شده از ماتریس طول تکرار سطح خاکستری (GLRLM)

قوه تکرار کوتاه (Short Run Emphasis)	قوه تکرار سطوح خاکستری پایین (Low Grey-Level Run Emphasis)
قوه تکرار بلند (Long Run Emphasis)	قوه تکرار سطوح خاکستری بالا (High Grey-Level Run Emphasis)
عدم یکنواختی سطوح خاکستری (Grey-Level Nonuniformity)	قوه تکرار کوتاه سطوح خاکستری پایین (Short Run Low Grey-Level Emphasis)
عدم یکنواختی طول تکرار (Run Length Nonuniformity)	قوه تکرار کوتاه سطوح خاکستری بالا (Short Run High Grey-Level Emphasis)
درصد تکرار (Run Percentage)	قوه تکرار بلند سطوح خاکستری پایین (Long Run Low Grey-Level Emphasis)
قوه تکرار بلند سطوح خاکستری بالا (Long Run High Grey-Level Emphasis)	

جدول (۲): ویژگی‌های شکل استفاده شده در این مقاله

میزان دایرگی (Circularity)	$Circularity = \frac{4\pi A}{P^2}$
میزان محدب (Convexity)	$Convexity = \frac{P_c}{P}$
میزان سختی (Solidity)	$Solidity = \frac{A}{A_c}$
میزان تراکم (Compactness)	$Compactness = \frac{2\sqrt{zA}}{P}$
گریز از مرکز (Eccentricity)	$E = \sqrt{1 - \left(\frac{\text{Minor axis length}^2}{\text{Major axis length}^2}\right)}$
شاخص دایروی (Roundness)	$R = \frac{4A}{\pi(\text{Major axis length}^2)}$
میزان پیچیدگی (Complexity)	$Complexity = \frac{P}{\sqrt{A}}$
طول محور اصلی (Major axis length)	
طول محور فرعی (Minor axis length)	

۲-۳-۲- ویژگی‌های مستخرج از ماتریس هم‌رخداد

ماتریس هم‌رخداد، نرخ رخداد مقادیر خاکستری دو پیکسل مرتبط از تصویر که در فاصله و جهت بخصوصی از یکدیگر قرار گرفته‌اند، را بیان می‌کند. ماتریس هم‌رخداد P ماتریس مربعی است که درایه‌ی $P(i,j)$ آن نشان‌دهنده‌ی فراوانی وقوع زوج پیکسل‌هایی است که یکی از آنها با سطح خاکستری i و دیگری با سطح خاکستری j در فاصله‌ی خاص و در جهت تعیین شده از یکدیگر قرار گرفته‌اند. زاویه‌های ممکن بین دو پیکسل ۰، ۴۵، ۹۰ و ۱۳۵ درجه هستند. در این تحقیق برای هر ناحیه‌ی مورد نظر، ماتریس هم‌رخداد در جهت صفر درجه به دست آمده و سپس ویژگی‌های مورد نظر از این ماتریس استخراج شده است. برای محاسبه‌ی ویژگی‌ها ابتدا باید ماتریس هم‌رخداد نرمالیزه شده و سپس معادلات مربوط محاسبه شوند. برای نرمال سازی ماتریس هم‌رخداد، مقدار هر درایه به مجموع مقادیر همه درایه‌های آن ماتریس تقسیم می‌گردد. M بعد ماتریس هم‌رخداد است که برابر با بیشینه مقدار سطح خاکستری موجود در تصویر است. همچنین i و j به ترتیب سطر و ستون خانه‌ی مورد نظر، p ماتریس هم‌رخداد نرمال شده را نشان می‌دهد. ویژگی‌های استخراجی از ماتریس هم‌رخداد [۳۴] پس از نرمال سازی آن در جدول (۳) نشان داده شده است.

در روابط همبستگی، μ_1 ، μ_2 ، σ_1 و σ_2 به ترتیب میانگین و انحراف استاندارد است که در امتداد سطرها و ستون‌های ماتریس هم‌رخداد محاسبه شدند.

۲-۳-۳- ویژگی‌های مستخرج از ماتریس طول تکرار

ماتریس طول تکرار فراوانی تعداد پیکسل‌هایی که دارای روشنایی یکسان هستند و پشت سرهم قرار گرفته‌اند را نشان می‌دهد. اگر ماتریس طول تکرار در جهت مشخص با q نمایش داده شود،

۴-۲ طبقه‌بندی

فلوچارت، ساختاری درخت مانند را جهت اخذ تصمیم و تعیین کلاس و طبقه‌ی یک داده‌ی خاص به ما ارائه می‌کند. همان‌طور که از نام آن مشخص است، این درخت از تعدادی گره و شاخه تشکیل شده است به گونه‌ای که برگ‌ها، کلاس‌ها را نشان می‌دهند و گره‌های میانی هم بر اساس یک یا چند صفت مشخص برای تصمیم‌گیری به کار می‌روند. درخت تصمیم یک مدل خود توصیف است، بدین معنی که به‌تنهایی و بدون حضور یک فرد متخصص در آن حوزه، نحوه‌ی دسته‌بندی را به‌صورت گرافیکی نشان می‌دهد و به دلیل همین سادگی و قابل فهم بودن، روش محبوب در داده‌کاوی محسوب می‌شود. البته در مواردی که تعداد گره‌های درخت زیاد باشد، نمایش گرافیکی و تفسیر آن می‌تواند کمی پیچیده باشد. در این تحقیق از درخت تصمیم با حداکثر ۴ اشعاب استفاده شده است.

در این تحقیق، برای طبقه‌بندی بردارهای ویژگی به دو کلاس خوش‌خیم و بدخیم از طبقه‌بندهای ماشین بردار پشتیبان خطی^۱ و درخت تصمیم^۲ استفاده شده و قدرت آن‌ها در طبقه‌بندی با هم مقایسه شده است. در سیستم پیشنهادی به صورت تصادفی ۷۰ درصد از داده‌های مربوط به نوع خوش‌خیم و داده‌های نوع بدخیم به‌عنوان داده‌های آموزشی جهت تعلیم به طبقه‌بندها اعمال می‌شود. در مرحله‌ی آخر ۳۰ درصد باقیمانده از هر دو کلاس به عنوان داده‌های آزمایش مورد استفاده قرار می‌گیرد و طبقه‌بند باید به طور خودکار خوش‌خیمی و بدخیمی داده‌های آزمایش را پیش‌بینی کند و دقت طبقه‌بندی را محاسبه می‌کنیم. در ادامه طبقه‌بندهای مذکور را به اختصار معرفی و کارایی آن‌ها را بررسی می‌کنیم.

۱-۴-۲ ماشین بردار پشتیبان خطی

ماشین بردار پشتیبان یک روش بسیار قدرتمند در طبقه‌بندی است. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نشان داده است. مبنای کار طبقه‌بند SVM، طبقه‌بندی خطی داده‌ها است. اولین مسئله‌ای که SVM باید به آن بپردازد این است که کدام خط مناسب‌ترین گزینه برای تقسیم‌بندی خطی داده‌ها است. در واقع به دنبال خطی است که تا حد امکان از داده‌های هر دو کلاس بیشترین فاصله را داشته باشد. SVM در سال ۱۹۶۰ توسط وپنک معرفی شد [۳۶]. در این روش با استفاده از یک الگوریتم بهینه‌سازی، نمونه‌های تشکیل دهنده مرزهای کلاس‌ها بدست می‌آیند و با استفاده از آن‌ها یک مرز تصمیم‌گیری خطی بهینه برای جدا سازی کلاس‌ها محاسبه می‌شود. به نمونه‌های مرزی، بردارهای پشتیبان می‌گویند و برای محاسبه‌ی مرز تصمیم دو کلاس کاملاً جدا از هم، از روش حاشیه-ی بهینه استفاده می‌شود. این حاشیه به گونه‌ای محاسبه می‌شود که تمام نمونه‌های کلاس +۱ در یک طرف مرز و تمام نمونه‌های کلاس -۱ در طرف دیگر مرز واقع شوند و مرز تصمیم به گونه‌ای است که فاصله‌ی نزدیکترین نمونه‌های آموزشی هر دو کلاس از هم در راستای عمود بر مرز تصمیم تا حد امکان حداکثر شود. اگر بخواهیم مرز تصمیم را به طور موازی توسط دو خط از دو طرف گسترش دهیم تا از بین نزدیکترین نمونه‌های دو کلاس بگذرد مرز تصمیم‌گیری بهینه به وجود می‌آید.

۲-۴-۲ درخت تصمیم

درخت تصمیم یکی از الگوریتم‌ها و روش‌های محبوب در حوزه‌ی طبقه‌بندی یا دسته‌بندی داده‌ها است. درخت تصمیم که هدف اصلی آن، طبقه‌بندی داده‌هاست، مدلی در داده‌کاوی است که مشابه

۵-۲ پایگاه داده

در این تحقیق از تصاویر دو پایگاه داده‌ی DDSM و MIAS استفاده شده است. پایگاه MIAS حاوی ۳۲۲ تصویر دیجیتالی مربوط به پستان‌های راست و چپ ۱۶۱ زن مختلف می‌باشد. از این ۳۲۲ پستان، ۱۱۳ پستان سرطانی هستند، که از این تعداد ۲۳ پستان حاوی خوشه‌های کلسیفیکاسیون^۳، ۵۶ پستان حاوی توده^۴، ۱۹ پستان حاوی درهم‌ریختگی ساختاری^۵ و ۱۵ پستان نامتقارن است. ممکن است برخی از ماموگرام‌ها حاوی بیش از یک آسیب باشند. این ۱۱۳ پستان سرطانی شامل ۱۲۱ آسیب هستند که از این تعداد ۶۷ آسیب خوش‌خیم و ۵۴ آسیب بدخیم هستند. در این تحقیق از تصاویر شامل آسیب‌های توده، درهم‌ریختگی ساختاری و عدم تقارن^۵ که در این پایگاه داده وجود دارند، استفاده شده است.

پایگاه داده‌ی DDSM یک پایگاه عمومی با تعداد تصاویر زیاد است که معمولاً محققان یک زیرمجموعه از تصاویر آن را انتخاب می‌کنند و آزمایش‌های خود را بر روی آن انجام می‌دهند. در این پایگاه برای هر بیمار، چهار تصویر چپ و راست از نماهای MLO و CC وجود دارد. در نمای CC پستان از بالا تصویربرداری می‌شود. این نما هم در ماموگرافی معمولی و هم در ماموگرافی تشخیصی به عمل می‌آید. در نمای CC باید بافت غده‌های پستان کاملاً قابل رؤیت شوند. بافت چربی نزدیک عضلات پستان باید به‌صورت نوارهای تیره مشخص باشد و در پشت آنان باید عضلات قفسه سینه دیده شود. MLO هم نمایی تقریباً با زاویه‌ی ۴۵ درجه است. در این نما عضلات ریه باید از بالا تا ۱/۴ فوقانی خارجی پستان و زیربغل تا زیر نوک پستان یا حتی پایین‌تر قابل رؤیت باشد.

^۲ Calcification

^۳ Mass

^۴ Architectural Distortion

^۵ Bilateral Asymmetry

^۱ Linear support vector machine

^۲ Decision tree

همچنین نرخ خطای سیستم به ازای این طبقه‌بند قابل قبول است و سیستم عملکرد خوبی در تشخیص موارد خوش‌خیم و بدخیم دارد. همچنین سیستم پیشنهادی به ازای طبقه‌بند SVM خطی علیرغم اینکه دارای نرخ خطای بالاتری نسبت به سیستم با طبقه‌بند درخت تصمیم است ولی در کل دارای نرخ دقت قابل قبولی است. لازم به ذکر است که این نتایج بعد از چندین بار اجرا و میانگین گیری به دست آمده‌اند و هر دو طبقه‌بند با مجموعه‌های یکسانی از داده‌های آموزش و تست و همچنین شرایط برابر از نظر نوع بردارهای ویژگی ورودی مورد بررسی قرار گرفتند.

جدول (۶): عملکرد سیستم پیشنهادی به ازای طبقه‌بندهای مختلف

پایگاه داده	طبقه‌بند	دقت %	انحراف معیار دقت	FNR	FPR
MIAS	SVM خطی	۹۰/۸۱	۱/۲۹	۱۰	۸/۴۸
	درخت تصمیم	۹۱/۷۳	۱/۴۲	۸/۵۷	۷/۶۹
DDSM	SVM خطی	۹۰/۶۱	۱/۵۲	۸/۱۸	۱۰/۶
	درخت تصمیم	۹۱/۲۳	۱/۱۸	۸/۵۷	۹

در جدول (۷) عملکرد چندین تحقیق گذشته آورده شده است و مشاهده می‌شود که عملکرد سیستم پیشنهادی نسبت به سیستم‌های پیشنهادی در این تحقیق‌ها دارای نرخ تشخیص بالاتری است.

جدول (۷): مقایسه‌ی عملکرد سیستم پیشنهادی با تحقیقات پیشین

تحقیق	پایگاه داده	طبقه‌بند	دقت %
روش پیشنهادی	MIAS	درخت تصمیم	۹۱/۷۳
	DDSM	درخت تصمیم	۹۱/۲۳
Saki et al (2012)	MIAS	OWBPE	۸۹/۲۸
Verma et al (2012)	DDSM	MLP NN	۸۸/۷۵
Buciu et al (2011)	MIAS	PSVM	۸۲/۳
Rojas and Nandi (2009) [39]	DDSM	Bayesian, FLD, SVM	۸۱
Midya and Chakraborty (2015)	DDSM	ANN	۸۱/۲۳
Al-Masni et al (2017)	DDSM	Fully Convolutional NN	۸۵/۵۲
Hamidinekoo et al (2017) [40]	DDSM	Convolutional NN	۸۷
Rampun et al (2018) [41]	DDSM	Ensemble Convolutional NN	۸۰/۴

۴ جمع‌بندی و نتیجه‌گیری

تحقیقات برای استفاده از سیستم‌های تشخیص به کمک رایانه در پزشکی، در سال‌های اخیر پیشرفت چشم‌گیری داشته است. تا کنون سیستم‌های مختلفی به کمک روش‌های هوش مصنوعی برای دسته‌بندی تومورهای پستان به دو دسته‌ی خوش‌خیم و بدخیم طراحی شده‌اند. هدف از طراحی و توسعه‌ی چنین سیستم‌هایی،

در این تحقیق، ۱۱۵ تصویر حاوی توده‌ی خوش‌خیم و ۱۱۵ تصویر حاوی توده‌ی بدخیم از این پایگاه داده انتخاب شده‌اند.

۳ یافته‌ها

به‌منظور بررسی نتایج حاصل از سیستم‌های تشخیص کامپیوتری چند معیار مهم مورد ارزیابی قرار می‌گیرند که در ادامه تعاریف مهم و معیارهای ارزیابی عملکرد سیستم‌ها توضیح داده می‌شود. پاسخ مثبت صحیح (TP): عبارت است از تعداد تومورهایی که توسط سیستم، بدخیم تشخیص داده شده و نتایج پاتولوژی نیز بدخیمی آن‌ها را تأیید می‌کند.

پاسخ مثبت کاذب (FP): عبارت است از تعداد تومورهایی که توسط سیستم بدخیم تشخیص داده شده ولی نتایج پاتولوژی آن‌ها خوش‌خیم است.

پاسخ منفی صحیح (TN): عبارت است از تعداد تومورهایی که توسط سیستم، خوش‌خیم تشخیص داده شده و نتایج پاتولوژی نیز خوش‌خیمی آن‌ها را تأیید می‌کند.

پاسخ منفی کاذب (FN): عبارت است از تعداد تومورهایی که توسط سیستم خوش‌خیم تشخیص داده شده ولی نتایج پاتولوژی آن‌ها بدخیم است.

به منظور ارزیابی عملکرد سیستم‌های تشخیص کامپیوتری، پارامترهایی در نظر گرفته می‌شود که با مشاهده‌ی مقادیر آن‌ها می‌توان میزان موفقیت سیستم پیشنهادی را در پیش‌بینی صحیح مورد ارزیابی قرار داد. در این تحقیق از پارامترهای دقت^۱، FNR و FPR استفاده شده است که در جدول (۵) روابط مربوطه آورده شده است.

جدول (۵): معیارهای ارزیابی عملکرد سیستم پیشنهادی

معیار	دقت	FNR	FPR
رابطه	$\frac{TP+TN}{TP+FP+TN+FN}$	$\frac{FN}{FN+TP}$	$\frac{FP}{FP+TN}$

دو مقدار TN و TP مهم‌ترین مقادیری هستند که باید بیشینه شوند تا کارایی طبقه‌بندی به حداکثر برسد و مهم‌ترین عامل برای تعیین کارایی یک الگوریتم طبقه‌بندی، معیار دقت است. همچنین FNR و FPR معیارهایی برای توصیف خطای سیستم هستند. FPR نشان‌دهنده‌ی این است که چند درصد از تومورهای خوش-خیم به اشتباه به‌عنوان بدخیم دسته‌بندی شده‌اند. FNR هم بیانگر این است که چند درصد از تومورهای بدخیم توسط سیستم به اشتباه خوش‌خیم طبقه‌بندی شده‌اند.

عملکرد سیستم پیشنهادی بر روی دو پایگاه داده‌ی MIAS و DDSM در جدول (۶) نشان داده شده است. مشاهده می‌شود که طبقه‌بند درخت تصمیم دارای نرخ دقت طبقه‌بندی بیشتری است.

¹ Accuracy

- the detection of subtle signs." *Journal of the Franklin Institute* 344.3-4 (2007): 312-348.
- [9] Wallis, M. G., M. T. Walsh, and J. R. Lee. "A review of false negative mammography in a symptomatic population." *Clinical radiology* 44.1 (1991): 13-15.
- [10] Al-masni, Mohammed A., et al. "Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network." *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017.
- [11] Akselrod-Ballin, Ayelet, et al. "A region based convolutional network for tumor detection and classification in breast mammography." *Deep learning and data labeling for medical applications*. Springer, Cham, 2016. 197-205.
- [12] Wang, Zhiqiong, et al. "Breast tumor detection in digital mammography based on extreme learning machine." *Neurocomputing* 128 (2014): 175-184.
- [13] Khuza, A. Mohd, et al. "Identification of masses in digital mammogram using gray level co-occurrence matrices." *Biomedical imaging and intervention journal* 5.3 (2009).
- [14] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." *IEEE transactions on systems, man, and cybernetics* 9.1 (1979): 62-66.
- [15] X. Sun, W. Qian, and D. Song, "Ipsilateral-mammogram computer-aided detection of breast cancer," *Computerized Medical Imaging and Graphics the Official Journal of the Computerized Medical Imaging Society*, vol. 28, no. 3, pp. 151-158, 2004.
- [16] Verma, Brijesh, Peter McLeod, and Alan Klevansky. "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer." *Expert systems with applications* 37.4 (2010): 3344-3351.
- [17] RAHMANI, SERYASAT OMID, Javad Haddadnia, and ZADEH HOSSEIN GHAYOUMI. "Assessment of a novel computer aided mass diagnosis system in mammograms." (2016): 31-41.
- [18] Chaieb, Ramzi, et al. "Image features extraction for masses classification in mammograms." *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*. IEEE, 2014.
- [19] Midya, Abhishek, and Jayasree Chakraborty. "Classification of benign and malignant masses in mammograms using multi-resolution analysis of oriented patterns." *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015.
- [20] Lashkari, Amir Ehsan, and Mohammad Firouzmand. "Early breast cancer detection in thermogram images using AdaBoost classifier and fuzzy C-Means clustering algorithm." *Middle East Journal of Cancer*. (2016): 113-124
- [21] Basha, S. Saheb, and K. Satya Prasad. "AUTOMATIC DETECTION OF BREAST CANCER MASS IN MAMMOGRAMS USING MORPHOLOGICAL OPERATORS AND FUZZY C--MEANS

جایگزین کردن آنها به جای متخصصان نیست بلکه این سیستم‌های دسته‌بندی در واقع به عنوان یک ابزار کمکی در اختیار رادیولوژیست‌ها قرار می‌گیرند و هدف این است که به رادیولوژیست در شناسایی خوش‌خیمی و بدخیمی تومور کمک کنند.

در این مقاله با بهسازی تصویر ماموگرام و استفاده از HFCM ناحیه مشکوک به تومور با دقت خوبی از سایر بافتها جدا می‌شوند و سپس ترکیبی از ویژگی‌های بافت و هندسی از ناحیه مشکوک استخراج می‌شود که با طبقه بندی این ویژگی‌ها می‌توان تومورهای خوش‌خیم را از بدخیم متمایز کرد. سیستم پیشنهادی بر روی مجموعه داده‌ی MIAS و DDSM آزمایش شده است. به منظور ارزیابی سیستم پیشنهادی و مقایسه‌ی آن با سیستم‌های پیشنهاد شده در تحقیقات پیشین، شاخص‌های دقت، FNR و FPR به‌عنوان معیار ارزیابی در نظر گرفته شد. نرخ تشخیص حاصل قابل قبول و بالاتر از سایر الگوریتم‌های رایج است. نتایج نشان دهنده‌ی این است که سیستم پیشنهادی در طبقه‌بندی نوع تومور موفق عمل کرده است. یکی از محدودیت‌های روش پیشنهادی، نیاز به ناحیه‌بندی دقیق تومور است و اگر به هر دلیلی مرز تومور درست استخراج نشود منجر به افزایش خطای طبقه بندی خواهد شد.

مراجع

- [1] de Oliveira, Júlia Episichina Engrácia, Arnaldo de Albuquerque Araújo, and Thomas M. Deserno. "Content-based image retrieval applied to BI-RADS tissue classification in screening mammography." *World journal of radiology* 3.1 (2011): 24.
- [2] Asadi, Sajjad, et al. "A Study about Global Burden of Disease: Calculation and Analysis of Years of Life Lost (YLL) In Iran." *Adv Biol Res* 9 (2015): 156-65.
- [3] Taghavi, Afsoon, et al. "Increased trend of breast cancer mortality in Iran." *Asian Pacific Journal of Cancer Prevention* 13.1 (2012): 367-370.
- [4] Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin* 2016; 66(1):7- 30.
- [5] Lee, Carol H., et al. "Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer." *Journal of the American college of radiology* 7.1 (2010): 18-27.
- [6] Alipoor, M., and J. Haddadnia. "An accurate intelligent breast cancer diagnosis system." (2009): 33-40.
- [7] Oliver, A. *Automatic mass segmentation in mammographic images*. Diss. PhD thesis, Universitat de Girona (España) 2008, 2007.
- [8] Rangayyan, Rangaraj M., Fabio J. Ayres, and JE Leo Desautels. "A review of computer-aided diagnosis of breast cancer: Toward

- images." *Biomedical Signal Processing and Control* 6.4 (2011): 370–378.
- [39] Rojas-Domínguez, Alfonso, and Asoke K. Nandi. "Development of tolerant features for characterization of masses in mammograms." *Computers in Biology and Medicine* 39.8 (2009): 678–688.
- [40] Hamidinekoo, Azam, et al. "Investigating the effect of various augmentations on the input data fed to a convolutional neural network for the task of mammographic mass classification." *Annual Conference on Medical Image Understanding and Analysis*. Springer, Cham, 2017.
- [41] Rampun, Andrik, et al. "Breast Mass Classification in Mammograms using Ensemble Convolutional Neural Networks." *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2018.
- CLUSTERING." *Journal of Theoretical & Applied Information Technology* 5. 6 (2009).
- [22] Sinare, Swati S., and Pratap S. Vikhe. "Fuzzy C-means Clustering Algorithm for Breast Cancer Detection. *International Journal of Advanced and Innovative Research*" (2012).
- [23] Charate, A. P., and S. B. Jamge. "The preprocessing methods of mammogram images for breast cancer detection." *Int. J. Recent Innov. Trends Comput. Commun. (IJRITCC)* 5.1 (2017): 261–264.
- [24] Pizer, Stephen M., et al. "Adaptive histogram equalization and its variations." *Computer vision, graphics, and image processing* 39.3 (1987): 355–368.
- [25] Yang, Yong, and Shuying Huang. "Image segmentation by fuzzy c-means clustering algorithm with a novel penalty term." *Computing and Informatics* 26.1 (2012): 17–31.
- [26] Suckling, John, et al. "The mammographic image analysis society digital mammogram database." *Excerpta Medica. International Congress Series*. Vol. 1069. 1994.
- [27] Heath, Michael, et al. "The digital database for screening mammography." *Proceedings of the 5th international workshop on digital mammography*. Medical Physics Publishing, 2000.
- [28] Azli, Effa Adrina, et al. "Mass classification in mammograms using neural network." *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*. IEEE, 2017.
- [29] Cheng, Heng-Da, et al. "Approaches for automated detection and classification of masses in mammograms." *Pattern recognition* 39.4 (2006): 646–668.
- [30] Kilday, Judy, Francesco Palmieri, and Martin D. Fox. "Classifying mammographic lesions using computerized image analysis." *IEEE transactions on medical imaging* 12.4 (1993): 664–669.
- [31] Yang, Mingqiang, Kidiyo Kpalma, and Joseph Ronsin. "A survey of shape feature extraction techniques." (2008): 43–90.
- [32] Khurana, Garima, Garima Joshi, and Jatinderpal Kaur. "Static hand gestures recognition system using shape based features." *Engineering and Computational Sciences (RAECS), 2014 Recent Advances in*. IEEE, 2014.
- [33] Xu, Juan, et al. "Predicting human gaze beyond pixels." *Journal of vision* 14.1 (2014): 28–28.
- [34] Haralick, Robert M., and Karthikeyan Shanmugam. "Textural features for image classification." *IEEE Transactions on systems, man, and cybernetics* 6 (1973): 610–621.
- [35] Galloway, Mary M. "Texture analysis using grey level run lengths." *NASA STI/Recon Technical Report N 75* (1974).
- [36] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992.
- [37] Saki, Fatemeh, et al. "Fast opposite weight learning rules with application in breast cancer diagnosis." *Computers in biology and medicine* 43.1 (2013): 32–41.
- [38] Buciu, Ioan, and Alexandru Gacsadi. "Directional features for automatic tumor classification of mammogram



حمزه حسن‌نیا تحصیلات مقطع کارشناسی خود را در سال ۱۳۹۴ در رشته‌ی مهندسی برق مخابرات دانشگاه بناب به پایان رساند و مدرک کارشناسی ارشد خود را در رشته‌ی مهندسی برق مخابرات سیستم در سال ۱۳۹۷ از دانشگاه ارومیه دریافت نمود. زمینه‌های تحقیقاتی مورد علاقه ایشان پردازش تصویر و کاربردهای آن است.



مهدی چهل امیرانی دوره کارشناسی مهندسی برق- الکترونیک را در دانشگاه ارومیه در سال ۱۳۷۲ و کارشناسی ارشد و دکتری خود را در رشته مهندسی برق- مخابرات در دانشگاه علم و صنعت ایران به ترتیب در سال‌های ۱۳۷۶ و ۱۳۸۸ به پایان رساند. ایشان در حال حاضر دانشیار گروه مخابرات دانشگاه ارومیه می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان بازشناخت الگو، پردازش تصویر و سیگنال و نیز مخابرات امن است.



مرتضی ولی زاده دوره کارشناسی مهندسی برق-الکترونیک را در دانشگاه صنعتی شاهرود و کارشناسی ارشد و دکتری خود را در رشته مهندسی برق-الکترونیک در دانشگاه تربیت مدرس به پایان رسانده است. ایشان در حال حاضر استادیار گروه مخابرات دانشگاه ارومیه می‌باشد. زمینه های تحقیقاتی مورد علاقه ایشان پردازش تصویر و سیگنال، بینایی ماشین و یادگیری عمیق است.