

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده : کامپیوتر و فناوری اطلاعات
گروه : هوش مصنوعی

تشخیص خودکار توده‌ها در تصاویر ماموگرافی

ریحانه امینی

استاد راهنما :

دکتر مرتضی زاهدی

استاد مشاور:

دکتر حمید حسن پور و دکتر علیرضا احمدی فرد

پایان نامه جهت اخذ درجه کارشناسی ارشد

شهریور ۹۲

دانشگاه صنعتی شاهرود

دانشکده : کامپیوتر و فناوری اطلاعات

گروه : هوش مصنوعی

پایان نامه کارشناسی ارشد خانم ریحانه امینی

تحت عنوان: تشخیص خودکار توده‌ها در تصاویر ماموگرافی

مورد

توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد

در تاریخ

مورد پذیرش قرار گرفت.

ارزیابی و با درجه

امضاء	اساتید مشاور	امضاء	اساتید راهنما

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی :		نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :

تقدیم بہ پدر بزرگوار و مادر مہربانم

و

ہمسر کرامیم کہ ہموارہ در طول تحصیل، تکیہ گاہ من در مواجہہ با مشکلات،

و وجودش مایہ دلگرمی من بودہ است

مشکر و قدردانی

سپاس خدایی را که نیکویی های آفرینش را برای ما برگزید و سپاس خدایی را که سیاهی ندانستن را از من زدود و هزار سپاس از برای او، به امید آنکه توفیق یابم جز خدمت به خلق او نکوشم.

ابتدا از زحمات و پشتیبانی بی دریغ و بی ثأبه استاد ارجمندم جناب آقای دکتر زاهدی که راهنمایی این تحقیق را بر عهده داشتند کمال مشکر را دارم.

و نیز سپاسگزارم از استادان گرامی جناب آقای دکتر علی اکبر پویان و جناب آقای دکتر حمید حسن پور که همواره در طول دوره کارشناسی و کارشناسی ارشد راهنمای من بوده اند.

تعهد نامه

اینجانب دانشجوی دوره کارشناسی ارشد رشته هوش مصنوعی دانشکده کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه تشخیص خودکار توده‌ها در تصاویر ماموگرافی تحت راهنمایی دکتر مرتضی زاهدی متعهد می‌شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « **Shahrood University of Technology** » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

سرطان سینه، شایع‌ترین نوع سرطان و دومین عامل مرگ‌ومیر ناشی از سرطان در میان زنان است. پیشگیری از این بیماری به دلیل ناشناخته بودن عوامل آن تقریباً غیرممکن به نظر می‌رسد. بنابراین شناسایی و تشخیص زودهنگام آن یکی از عوامل مهم و اساسی در درمان این بیماری است. تشخیص و درمان زودهنگام سرطان سینه باعث افزایش امید به زندگی شده و بیمار را برای مراجعه بعدی درمان آماده می‌کند. استفاده از روش ماموگرافی در حال حاضر رایج‌ترین راه تشخیص زودهنگام این بیماری است و درصد مرگ و میر را تا ۲۵ درصد کاهش داده است.

سامانه‌های تشخیص پزشکی (CAD)، به صورت گسترده برای شناسایی انواع ضایعات در تصاویر ماموگرافی به کار می‌روند. توده‌ها و MCها دو نوع مهم از این ضایعات هستند. در این تحقیق هدف طراحی، شبیه‌سازی، بررسی و مقایسه سیستم‌های CAD می‌باشد که قادر به شناسایی خودکار توده‌ها در تصاویر ماموگرافی با دقت بالایی باشند. علاوه بر توده کاربرد سامانه‌های تشخیص پزشکی در شناسایی MCها نیز به صورت کوتاه مورد بررسی قرار می‌گیرد.

روش پیشنهادی ما در این پایان نامه از ۴ مرحله اصلی تشکیل شده است. این ۴ مرحله، شامل پیش‌پردازش، استخراج نواحی مشکوک به توده، استخراج ویژگی از نواحی مشکوک و دسته‌بندی می‌باشد. همه مراحل در فصل‌های جداگانه به همراه جزئیات بحث می‌شود. در مرحله پیش‌پردازش راهکارهایی جهت بهبود کیفیت تصویر ماموگرافی و حذف مناطق اضافی همچون برجسب‌ها و ماهیچه پکتورال داده شده است. در محله دوم، ناحیه‌های مشکوک به توده به صورت خودکار و با استفاده از تحلیل هیستوگرام از تصویر ماموگرافی استخراج می‌شوند. در مرحله سوم انواع مختلف ویژگی‌های روشنایی، بافتی و شکلی معرفی می‌شود و نمونه‌هایی از آنها ارائه می‌شود. پس از آن ترکیب مناسبی از ویژگی‌های بافتی حاصل از ماتریس هم رخداد و ماتریس طول رخداد که منجر به دقت جداسازی بالا خواهد شد، پیشنهاد می‌شود. در مرحله آخر به منظور دسته‌بندی نواحی مشکوک، از سه طبقه‌بند پایه شبکه عصبی مصنوعی، K-NN و SVM استفاده شده است. با استفاده از تصاویر ماموگرافی از پایگاه داده DDSM نتایج را برای هر سه طبقه‌بند مورد بررسی قرار می‌دهیم.

در انتها به منظور افزایش دقت دسته‌بندی، از راهکار ترکیب اطلاعات در نتایج طبقه‌بندهای پایه استفاده می‌کنیم. برای این منظور روش OWA به کار گرفته شد. نتایج گزارش شده با استفاده از این روش حاکی از آن است که دقت دسته‌بندی نسبت به بهترین درصد در طبقه‌بندهای پایه بیش از ۴ درصد افزایش داشته است. همچنان استفاده همزمان از ویژگی‌های GLCM و GLRLM بیش از ۲٪ بهبود دقت، نسبت به شرایطی که از هر کدام از آنها به تنهایی استفاده شده است را به همراه دارد.

کلمات کلیدی:

سرطان سینه، ماموگرافی، سیستم CAD، توده، منطقه مورد توجه (ROI)، رشد ناحیه، ماتریس هم‌رخداد، ماتریس طول رخداد، ترکیب طبقه‌بندها، OWA

لیست مقالات مستخرج از پایان نامه

1. Reyhaneh Amini, Morteza Zahedi, Akbar Amirshakoori, **“Increasing accuracy in classification of suspicious areas in mammograms using classifiers combination”**, International Journal of Computer Applications (IJCA), 2013. (accept)

۲. ریحانه امینی، مرتضی زاهدی، **“دسته‌بندی الگوهای نرمال و سرطانی در تصاویر ماموگرافی به منظور تشخیص سرطان سینه”**، هشتمین کنفرانس ماشین بینایی و پردازش تصویر ایران (MVIP 2013). (accept)

فهرست مطالب

عنوان	صفحه
فصل اول : مقدمه و مفاهیم اولیه	۱
۱-۱ سرطان سینه	۲
۲-۱ سیستم‌های تصویربرداری برای معاینه سینه	۳
۱-۲-۱ تصویربرداری با اشعه ایکس (ماموگرافی)	۴
۲-۲-۱ تصویر برداری با تشدید مغناطیسی (MRI)	۸
۳-۲-۱ تصویربرداری با امواج فراصوتی (سونوگرافی)	۹
۳-۱ موارد استفاده ماموگرافی	۱۱
۴-۱ عوامل سرطانی قابل شناسایی در تصاویر ماموگرافی	۱۲
۱-۴-۱ توده‌ها (mass)	۱۲
۲-۴-۱ دسته‌های ریز کلسیمی (Micro Calcifications)	۱۶
۳-۴-۱ ناهنجاری ساختاری	۱۷
۴-۴-۱ عدم تقارن دو طرفه	۱۸
۵-۱ سامانه‌های تشخیص پزشکی (CAD)	۱۹
۱-۵-۱ مزایای استفاده از سیستم‌های CAD	۲۰
۲-۵-۱ نماهای چند گانه تکنیک‌های سیستم‌های CAD	۲۱
۳-۵-۱ شمای کلی سیستم‌های CAD	۲۲
فصل دوم : مروری بر کارهای گذشته در شناسایی ضایعه های MC و توده	۲۴
۱-۲ مقدمه	۲۵
۲-۲ روش‌های تشخیص MC ها	۲۵
۱-۲-۲ روش‌های مبتنی بر بهبود کیفیت تصویر	۲۶
۲-۲-۲ روش‌های مبتنی بر مدل‌های آماری	۲۶
۳-۲-۲ روش‌های تجزیه چند مقیاسی	۲۷
۴-۲-۲ روش‌های آموزش ماشین	۲۷
۳-۲ روش‌های تشخیص توده	۲۹

۳۰ Pixel-based روش‌های ۱-۳-۲
۳۱ Region-based روش‌های ۲-۳-۲
۳۲ انواع رویکردها در شناسایی توده‌ها.. ۴-۲
۳۶ ناحیه‌بندی تصاویر ماموگرافی ۵-۲
۳۶ global آستانه‌گذاری ۱-۵-۲
۳۷ local آستانه‌گذاری ۲-۵-۲
۳۷ رشد ناحیه ۳-۵-۲
۳۸ تطبیق الگو ۴-۵-۲
۳۸ متدهای فازی ۵-۵-۲
۳۹ تفاضل دو نمایه تصویر ۶-۵-۲
۴۰ فصل سوم : پیش‌پردازش بر روی تصاویر ماموگرافی و استخراج ROI
۴۱ ۱-۳ مقدمه
۴۱ ۲-۳ بهبود کیفیت تصویر
۴۲ global اصلاح هیستوگرام ۱-۲-۳
۴۴ local متدهای پردازشی ۲-۲-۳
۴۴ ۳-۲-۳ متدهای پردازش چند-مقیاسه
۴۵ ۳-۳ پیش‌پردازش استفاده شده بر روی تصاویر ماموگرافی
۴۶ ۱-۳-۳ بهبود تصاویر ماموگرافی
۴۷ ۲-۳-۳ حذف مصنوعات
۴۹ ۳-۳-۳ حذف ماهیچه پکتورال
۵۳ ۴-۳ استخراج نواحی مشکوک (ROIs)
۵۴ ۱-۴-۳ روش پیشنهادی برای استخراج خودکار ROIها
۵۶ فصل چهارم : استخراج ویژگی
۵۷ ۱-۴ مقدمه
۵۷ ۲-۴ انواع ویژگی
۵۸ ۳-۴ ویژگی‌های روشنایی
۵۹ ۴-۴ ویژگی‌های بافتی

۶۰	۱-۴-۴ ویژگی‌های حاصل از ماتریس هم‌رخداد
۶۴	۲-۴-۴ ویژگی‌های حاصل از ماتریس طول رخداد
۶۷	۵-۴ ویژگی‌های شکلی
۶۹	فصل پنجم: دسته‌بندی
۷۰	۱-۵ مقدمه
۷۱	۲-۵ طبقه‌بند SVM
۷۴	۳-۵ طبقه‌بند شبکه عصبی مصنوعی
۷۶	۴-۵ طبقه‌بند K-NN
۷۹	۵-۵ ترکیب طبقه‌بندها
۸۰	۱-۵-۵ روش OWA
۸۳	فصل ششم: نتایج شبیه‌سازی
۸۴	۱-۶ مقدمه
۸۵	۲-۶ پایگاه داده
۸۶	۳-۶ معیارهای ارزیابی
۸۹	۴-۶ داده‌های آموزشی و آزمایشی
۸۹	۵-۶ تاثیر ویژگی‌های مختلف در نتایج آزمایش
۹۰	۶-۵-۱ ویژگی‌های مورد استفاده
۹۱	۶-۶ تاثیر طبقه‌بندهای مختلف در نتایج آزمایش
۹۴	۷-۶ جمع بندی
۹۶	فصل هفتم: نتیجه‌گیری و ارائه پیشنهادات
۹۷	۱-۷ نتیجه‌گیری
۹۹	۲-۷ پیشنهاداتی برای ادامه تحقیق
۱۰۰	مراجع

فهرست شکل‌ها

صفحه	عنوان
۵.....	شکل (۱-۱): نمونه‌ای از یک دستگاه ماموگرافی
۶.....	شکل (۲-۱): یک تصویر ماموگرافی شامل دو نمای CC و MLO
۹.....	شکل (۳-۱): نمونه‌ای از یک تصویر MRI
۱۰.....	شکل (۴-۱): نمونه‌ای از یک تصویر اولتراسوند یا سونوگرافی سینه
۱۲.....	شکل (۵-۱): انواع شکل برای توده‌ها
۱۴.....	شکل (۶-۱): نمونه‌هایی از اشکال مختلف توده‌ها (الف) گرد (ب) تخم مرغی (ج) لاله‌ای (د) نامنظم
۱۵.....	شکل (۷-۱): انواع مرز توده‌ها
۱۵.....	شکل (۸-۱): نمونه‌ای از مرزهای مختلف توده‌ها (الف) خوش‌تعریف (ب) مبهم (ج) ریز لاله‌ای (د) بدتعریف (ه) سوزنی
۱۷.....	شکل (۹-۱): نمونه‌ای از تصویر ماموگرافی شامل MC
۱۸.....	شکل (۱۰-۱): تصویر ماموگرافی شامل اعوجاج ساختاری
۱۹.....	شکل (۱۱-۱): تصویر ماموگرافی شامل عدم تقارن
۲۲.....	شکل (۱۲-۱): تصاویر (a) و (b) نمونه‌ای از عدم تقارن را نشان می‌دهد و تصاویر (c) و (d) روشی برای یافتن این نامتقارن بودن است
۲۳.....	شکل (۱۳-۱): مراحل اصلی الگوریتم سامانه‌های تشخیص پزشکی CAD و سامانه‌های تشخیص عیب به کمک کامپیوتر CADx
۲۸.....	شکل (۱-۲): اعمال طبقه‌بند SVM برای تشخیص کلاستر MCها. تصویر سمت چپ نمونه‌ای از کلاستر MC در تصاویر ماموگرافی. تصویر سمت راست نتیجه اعمال طبقه‌بند SVM بر روی تصویر سمت چپ
۲۹.....	شکل (۲-۲): مراحل کلی الگوریتم‌های تشخیص توده
۳۳.....	شکل (۳-۲): جزییات ساختار روش‌های شناسایی توده

- شکل (۱-۳): نمونه‌ای از بهبود تصویر توسط متد تعدیل سازی هیستوگرام و ۴۳
هیستوگرام آن (ب) تصویر خروجی پس از تعدیل هیستوگرام
- شکل (۲-۳): بلوک دیاگرام پیش‌پردازش انجام شده ۴۵
- شکل (۳-۳): نمونه‌ای از تصاویر ماموگرافی شامل مصنوعات: شکل سمت راست) تصویر ۴۷
ماموگرافی دارای برجسب و شکل سمت چپ تصویر ماموگرافی دارای حاشیه
- شکل (۴-۳) حذف مصنوعات ناخواسته از تصاویر ماموگرافی. الف) تصویر اصلی ۵۹
ب) تصویر باینری با $T=18$ ج) حذف کلیه object ها به غیر از بزرگترین object در تصویر باینری
د) تصویر ماموگرافی که در آن برجسب موجود در تصویر الف حذف شده است.
- شکل (۵-۳) ماهیچه پکتورال در یک تصویر ماموگرافی از نمای MLO ۵۰
- شکل (۶-۳): حذف ماهیچه پکتورال. الف) تصویر ماموگرافی شامل ۵۲
ماهیچه پکتورال (ب) ماهیچه پکتورال ناحیه بندی شده توسط متد رشد ناحیه ج) تصویر ماموگرافی
با حذف ماهیچه پکتورال
- شکل (۷-۳) : دسته دایره هایی با افزایش ۵ پیکسلی شعاع به دور یک ماکزیمم بدست آمده ۵۴
- شکل (۸-۳): نمونه‌ای از استخراج ROI ها. الف) تصویر اصلی. ۵۵
ب و ج) ROI های استخراج شده از تصویر الف
- شکل (۱-۴): تصاویری با چندین نوع بافت مختلف ۶۰
- شکل (۲-۴): دو تصویر با بافت‌های مختلف و هیستوگرام‌های یکسان ۶۱
- شکل (۳-۴): چهار جهت مورد استفاده در ساخت ماتریس هم‌رخداد ۶۲
- شکل (۴-۴): چگونگی ایجاد ماتریس هم‌رخداد ۶۲
- شکل (۱-۵): اساس کار SVM ۷۱
- شکل (۲-۵): افزایش بعد برای رسیدن به نمایش داده‌هایی که به طور خطی تفکیک پذیر باشند ۷۳
- شکل (۳-۵): نمونه‌ای از ساختار یک شبکه عصبی مصنوعی MLP ۷۵
- شکل (۴-۵): اساس کار طبقه بند K-NN ۷۶
- شکل (۵-۵): یک نمای کلی از ترکیب طبقه‌بندها ۸۰
- شکل (۱-۶): نمودار کلی سیستم CAD پیشنهادی ۸۴

شکل (۶-۲): بررسی عملکرد روش OWA به ازاء مقادیر مختلف a ۹۴

فهرست جدول ها

عنوان	صفحه
جدول (۱-۱): احتمال خوش خیمی و بدخیمی برای اشکال مختلف توده.....	۱۳
جدول (۲-۱): احتمال خوش خیمی و بدخیمی برای مرزهای مختلف توده.....	۱۶
جدول (۱-۲): مقایسه دو رویکرد مبتنی بر ROI و مبتنی بر ناحیه بندی.....	۳۴
جدول (۱-۴): ویژگی های مبتنی بر روشنایی.....	۵۹
جدول (۲-۴): ویژگی های بافتی.....	۶۰
جدول (۳-۴): ویژگی های شکلی.....	۶۷
جدول (۱-۵): انواع کرنل های مشهور	۷۴
جدول (۱-۶): ساختار کانفیوژن ماتریس.....	۸۸
جدول (۲-۶): نتیجه CAD پیشنهادی با استفاده از ویژگی های GLCM	۹۰
جدول (۳-۶): نتیجه CAD پیشنهادی با استفاده از ویژگی های GLRLM	۹۰
جدول (۴-۶): نتیجه CAD پیشنهادی با استفاده از ترکیب ویژگی های GLRLM و GLCM	۹۰
جدول (۵-۶): نتایج آزمایش برای طبقه بند شبکه عصبی.....	۹۲
جدول (۶-۶): نتایج آزمایش برای طبقه بند 6-NN	۹۲
جدول (۷-۶): نتایج آزمایش برای طبقه بند SVM خطی.....	۹۳
جدول شماره (۸-۶): نتایج آزمایش برای طبقه بند SVM با هسته داخلی RBF.....	۹۳
جدول شماره (۹-۶) . نتایج آزمایش برای ترکیب طبقه بند های پایه	۹۴
جدول (۱۰-۶): نتایج معیارهای ارزیابی برای ۴ سیستم CAD مورد بررسی در این تحقیق	۹۵
جدول (۱۱-۶): مقایسه روش پیشنهادی با روش های دیگر	۹۵

فصل اول:

مقدمه و مفاهیم اولیه

۱-۱ سرطان سینه

سرطان یکی از بزرگترین تهدیدهای زندگی انسان است. و انتظار می‌رود که به یکی از عوامل اصلی مرگ و میر در چند دهه بعدی تبدیل شود. از هر ۴ مرگ در آمریکا یکی به علت سرطان بوده است. تخمین زده شده است ۵۷۷۱۹۰ نفر آمریکایی در سال ۲۰۱۲ بر اثر سرطان مرده‌اند، متناظر با بیش از ۱۵۰۰ نفر در روز [۱].

سرطان زمانی به وجود می‌آید که سلول‌های قسمتی از بدن بطور غیر قابل کنترل رشد کنند، به سرعت تقسیم شوند، به سمت بافت‌های مختلف بدن هجوم برند و در تمام بدن منتشر شوند. به مجموعه‌ای از این سلول‌های غیرقابل کنترل تومور گفته می‌شود. سلول‌ها، کوچکترین اجزای ساختمانی موجود زنده هستند که می‌توانند در بدن به طور مستقل فعالیت کنند. سلول‌های سالم در محدوده قوانین مشخصی رشد می‌کنند و رشد آنها نیز بعد از رسیدن به تکامل سلولی متوقف می‌شود ولی سلول‌های سرطانی خیلی سریعتر رشد می‌کنند و تا اندازه‌ای این رشد را ادامه می‌دهند که باعث تخریب سلول‌های سالم شوند. برخلاف سلول‌های سالم، سلول‌های سرطانی نمی‌توانند رشد خود را قطع کنند. این رشد و تکثیر غیرعادی نتیجه تغییرات یا جهش‌هایی در ماده ژنتیکی داخل آنها است. این وضعیت می‌تواند از والدین به ارث برسد یا اینکه در اثر مواجهه زیاد و طولانی مواد جهش‌زا ایجاد شود البته عوامل دیگری نیز در ایجاد این بی‌نظمی در رشد سلولی تأثیر دارند. سرطان به عنوان یک بیماری لاعلاج مدت‌هاست که بخش عظیمی از تحقیقات پزشکی و غیر پزشکی را به خود اختصاص داده است [۲].

یکی از رایج‌ترین سرطان‌ها به خصوص در بین زنان سرطان سینه است. سرطان سینه شایع‌ترین نوع سرطان و دومین عامل مرگ و میر ناشی از سرطان در میان زنان آمریکا است [۱]. در ایران نیز سرطان سومین عامل مرگ و میر است و در بین سرطان‌ها سرطان سینه شایع‌ترین سرطان در زنان

ایرانی است. لازم به ذکر است که سن ابتلا به سرطان سینه در ایران یک دهه پایین‌تر از کشورهای توسعه یافته است [۳].

مطالعات نشان می‌دهند که پیشگیری از سرطان سینه به دلیل ناشناخته بودن عوامل آن تقریباً غیرممکن به نظر می‌رسد. بنابراین شناسایی و تشخیص زودهنگام آن یکی از عوامل مهم و اساسی در درمان این بیماری است. تشخیص و درمان زودهنگام سرطان سینه باعث افزایش امید به زندگی شده و بیمار را برای مراجعه بعدی درمان آماده می‌کند.

برای اهداف درمانی ۵ مرحله برای این بیماری در نظر گرفته شده است. در مرحله صفر سلول‌های سرطانی پخش نشدنی هستند ولی این خطر وجود دارد که در زمان طولانی پخش شوند. در مرحله اول غده کمتر از ۲ سانتیمتر است و هنوز پخش نشده است. در مرحله دوم غده حدود ۲ سانتیمتر است ولی در حال پخش می‌باشد و یا اینکه حدود ۵ سانتیمتر است و هنوز پخش نشده است. در مرحله سوم غده سرطانی بیش از ۵ سانتیمتر است و به سمت غدد لنفاوی پخش شده است. در مرحله چهارم غده سرطانی در سرتاسر بدن پخش شده است.

۷۸ درصد موارد سرطان سینه در زنان بالای ۵۰ سال مشاهده شده است و ۶ درصد در زنان کمتر از ۴۰ سال گزارش شده است. در اکثر کشورهای پیشرفته مانند ایالات متحده، سرطان سینه مهم‌ترین عامل مرگ و میر در زنان بین ۲۰ تا ۵۹ سال می‌باشد.

۲-۱ سیستم‌های تصویربرداری برای معاینه سینه

برای معاینه سینه از تصویر برداری‌های مختلفی از جمله تصویر برداری با اشعه ایکس^۱، تشدید مغناطیسی^۲ و امواج فراصوت^۳ استفاده می‌شود [۴].

^۱ X-ray

^۲ Magnetic Resonance

^۳ Ultrasound

۱-۲-۱ تصویربرداری با اشعه ایکس (ماموگرافی)^۱

ماموگرافی یکی از روش‌های تصویربرداری با اشعه ایکس است که از پرتوهای اشعه ایکس دوز پایین^۲ با کنتراست^۳ بالا و رزولوشن بالا در تشخیص سرطان استفاده می‌کند. این متد تصویر برداری براساس تفاوت در جذب اشعه ایکس بین اجزای مختلف بافت سینه مانند بافت چربی، بافت تومور و غیره عمل می‌کند. ماموگرافی به عنوان یک روش مناسب برای تصویربرداری از ساختار درونی سینه، قابلیت آشکارسازی غدد داخلی سینه را قبل از ظهور فیزیکی بر عهده دارد [۵]. استفاده از روش ماموگرافی در حال حاضر رایج‌ترین راه تشخیص زودهنگام این بیماری است و درصد مرگ و میر را تا ۲۵ درصد کاهش داده است ولی با این حال تفسیر و تشریح تصاویر حاصل از ماموگرافی بسیار دشوار می‌باشد و براساس آمار رسمی مرکز ملی سرطان در آمریکا ۱۰ تا ۳۰ درصد غدد موجود در سینه بیمار در تصاویر ماموگرافی توسط رادیولوژیست قابل تشخیص نیستند [۶ و ۷].

این روش از حدود ۳۰ سال پیش به صورت امروزی ابداع شده و باعث سهولت در تشخیص و نیز بهبود درمان بیماران به علت تشخیص زودرس بیماری آنها شده است. آنچه باید به آن اشاره شود این است که خطرات ناشی از برخورد اشعه X با سینه در این روش آنقدر ناچیز است که می‌توان آن را نادیده گرفت. در حقیقت میزان سود ناشی از تشخیص زودرس توسط این روش بسیار بیشتر از خطرات احتمالی آن است. زیرا از یک طرف مقداری از اشعه X که به بافت‌ها نفوذ می‌کند بسیار کم است و از طرف دیگر، میزان اشعه لازم برای ایجاد سرطان بسیار بیشتر از اشعه‌ای است که توسط ماموگرافی به سینه تابانده می‌شود. حتی اگر یک خانم از ۴۰ سالگی تا ۹۰ سالگی هر سال ماموگرافی انجام دهد، میزان اشعه دریافتی برای ایجاد سرطان پستان بسیار ناچیز است.

¹ Mammography

² Low-dose

³ Contrast

در این روش یک تصویر سیاه و سفید از سینه تهیه می‌شود که لازم است توسط رادیولوژیست و پزشک جراح تفسیر شود. تفسیر تصویر ماموگرافی باید با توجه به شرایط معاینه و سایر آزمایشات بیمار انجام شود. برای انجام ماموگرافی بیمار در وضعیت ایستاده قرار می‌گیرد و در وضعیت افقی و عمودی، سینه فرد در بین صفحات مخصوص فشرده شده و تصویربرداری صورت می‌گیرد. این کار نه تنها باعث کاهش میزان اشعه دریافتی توسط بافت بدن می‌شود، یک تصویر واضح‌تر و قابل تفسیر را ایجاد می‌کند. البته این فشار فقط برای چند ثانیه وارد می‌شود.

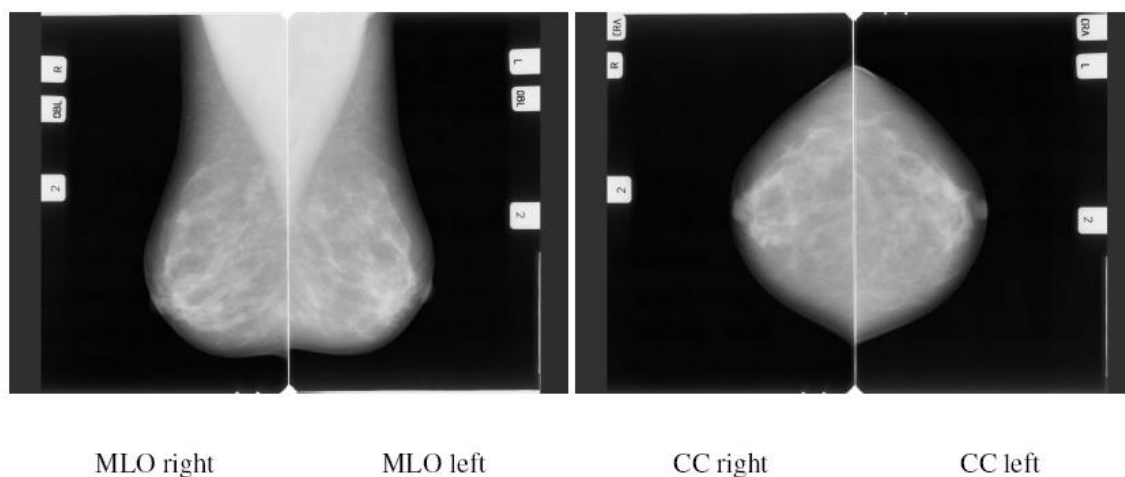


شکل (۱-۱): نمونه‌ای از یک دستگاه ماموگرافی

در ماموگرافی، چربی به رنگ خاکستری دیده می‌شود و بافتی از سینه که متراکم تر است، سفید به نظر می‌رسد. ذرات MC و توده‌های خوش‌خیم یا غیرسرطانی و نیز توده‌های بدخیم یا سرطانی نیز سفید دیده می‌شوند. ناهنجاری‌ها در ماموگرافی‌های زنان مسن‌تر و بعد از سن یائسگی را آسان‌تر

می‌توان تشخیص داد زیرا سینه‌ی آن‌ها عمدتاً از چربی تشکیل شده است، در نتیجه یک توده خوش‌خیم یا بدخیم که سفید رنگ است را در زمینه خاکستری مربوط به بافت چربی، به راحتی می‌توان شناسایی کرد.

برآمدگی^۱ سینه می‌تواند از زوایای مختلفی تصویربرداری شود. رایج‌ترین نماهای تصویر سینه در ماموگرام‌ها شامل MLO^۲ و CC^۳ هستند [۸]. نمای MLO از سمت اریب گرفته می‌شود و معمولاً با زاویه ۹۰ درجه است، اما نمای CC از بالا گرفته می‌شود. ماهیچه پکتورال^۴ در بالای نماهای MLO مشاهده می‌شود، در شکل (۲-۱) نمونه‌ای از این نماها آورده شده است.



شکل (۲-۱): یک تصویر ماموگرافی شامل دو نمای MLO و CC [۸]

ماموگرافی روی فیلم و ماموگرافی دیجیتال : ماموگرافی روی فیلم در واقع روش سنتی و معمول ماموگرافی است که در آن تصاویر اشعه ایکس به صورت مستقیم بر روی فیلم ذخیره می‌شود [۹]. ماموگرافی دیجیتال همزمان با پیشرفت‌هایی که در تکنولوژی کامپیوتر و صنایع الکترونیک نوری به وجود آمد، توسعه و رشد یافت. در ماموگرافی دیجیتال ابتدا تصویر اشعه ایکس از بافت سینه تولید

^۱ Projection
^۲ Medio-Lateral Oblique
^۳ Cranio-Caudal
^۴ Pectoral Muscle

می‌شود و سپس این تصویر را به صورت مستقیم و به شکل الکترونیکی در یک کامپیوتر ذخیره می‌کند [۹].

در ماموگرافی دیجیتال، تصویر به صورت یک فایل دیجیتال کامپیوتری ذخیره می‌شود، نه روی فیلم رادیوگرافی. این تکنولوژی سودمندی‌های عملی مهمی نسبت به ماموگرافی معمولی دارد. از آن جایی که تصویر اشعه ایکس به صورت دیجیتال ذخیره می‌شود، متخصص رادیولوژی می‌تواند روی تصویر زوم کند تا آن را بزرگ کند یا می‌تواند سایه روشن تصویر را به نحوی تنظیم کند که ناهنجاری‌هایی را که به صورت معمول دیده نمی‌شدند، آشکار شوند. تصاویر ماموگرافی دیجیتال را می‌توان روی دیسک ثبت کرد. این امر ذخیره کردن آن‌ها را آسان و دسترسی بعدی به آن‌ها را تسهیل می‌کند. تصویر دیجیتال را می‌توان در هر پایانه کامپیوتری در هر کجای مرکز درمانی دید و برای مشاوره گرفتن به مراکز درمانی دیگر هم فرستاد. تصاویر دیجیتال پستان در کمتر از ۱۰ ثانیه آماده می‌شوند، در حالی که ماموگرافی معمول با استفاده از فیلم اشعه ایکس ۵ تا ۱۰ دقیقه زمان می‌برد.

هر کدام از روش‌های ماموگرافی دیجیتال و ماموگرافی فیلم مزایا و معایب مخصوص به خود را دارند. در مقایسه با ماموگرافی دیجیتال، ماموگرافی فیلم تعدادی محدودیت دارد [۱۰]، که این محدودیت‌ها شامل:

۱- طیف محدود پرتوهای اشعه ایکس

۲- پس از بدست آوردن تصویر از این طریق نمی‌توان کنتراست آن را تغییر داد

۳- پردازش فیلم آهسته است و مصنوعاتی^۱ را تولید می‌کند

همه این محدودیت‌ها محققان را ترغیب به توسعه تکنیک‌های پیشرفته برای ماموگرافی دیجیتال می‌کند. ماموگرافی دیجیتال بر این محدودیت‌ها غلبه می‌کند و مزایای بالقوه زیر را دارا می‌باشد [۱۰]:

¹ artifact

۱- بهبود کنتراست تصویر

۲- بهبود کیفیت تصویر

۳- دوز پایین تر اشعه ایکس

۴- نویز کمتر

همچنان یکی از مزایای عمده ماموگرافی دیجیتال این است که تصویر ماموگرافی حاصل از این روش را می توان توسط سیستم های تشخیص با کمک کامپیوتر یا همان سیستم های CAD تجزیه و تحلیل کرد.

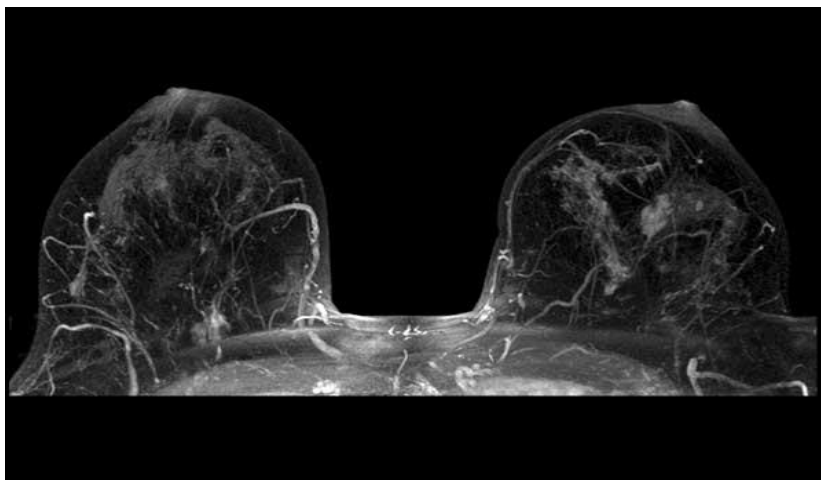
اگرچه ماموگرافی دیجیتال مزایای بالقوه بسیاری در مقابل روش سنتی ماموگرافی فیلم دارد، اما آزمایشات بالینی نشان می دهد که در مجموع سطح دقت تشخیص هر دو نوع ماموگرافی فیلم و ماموگرافی دیجیتال، هنگامیکه در تشخیص سرطان سینه به کار می روند شبیه به هم است [۱۱]. اما با این حال ماموگرافی دیجیتال ممکن است موثرتر باشد [۱۲] و [۱۳].

۱-۲-۲ تصویر برداری با تشدید مغناطیسی (MRI)

تصویربرداری با تشدید مغناطیسی (ام آر آی) با ایجاد یک حوزه قوی مغناطیسی تصاویری از بدن ایجاد می کند. ام آر آی بهترین جایگزین برای ماموگرافی است. ام آر آی برای تشخیص بعضی از سرطان ها که ممکن است ماموگرافی موفق به تشخیص آنها نشده باشد، حساس است. به علاوه ام آر آی توسط شناسایی مرحله بیماری می تواند به رادیولوژیست و دیگر متخصصان برای تعیین اینکه چگونه بیمار مبتلا به سرطان سینه را عمل کنند کمک برساند. و همچنان این متد برای تصویربرداری از سینه پس از جراحی یا اشعه درمانی بسیار مفید است. در هنگام انجام ام آر آی، بیمار باید به شکم روی میز اسکن بخوابد.

¹ Magnetic Resonance Imaging

برای بهبود کنتراست تصاویر سینه عمل تصویربرداری ام آر آی ممکن است به همراه تزریق به بدن بیمار انجام شود. این روش بر این فرضیه استوار است که پس از تزریق، اختلاف بین بافت‌های نرمال از غیرنرمال زیاد می‌شود. ام آر آی می‌تواند برای تمام زنانی که تصویر برداری ماموگرافی برای آنها مناسب نیست، از قبیل زنان جوان با سینه‌های متراکم و زنانی که ایمپلنت سیلیکونی کار گذاشته‌اند، انجام شود [۱۴].



شکل (۱-۳): نمونه‌ای از یک تصویر MRI

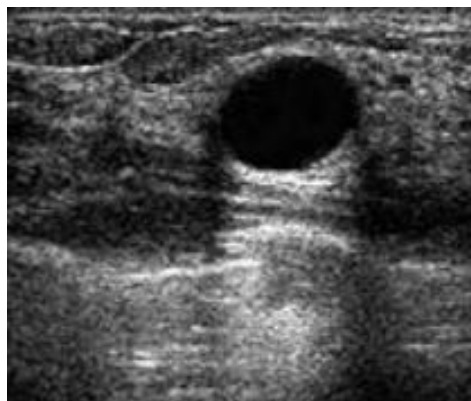
۳-۲-۱ تصویربرداری با امواج فراصوتی (سونوگرافی)

اولتراسوند یا سونوگرافی سینه فرآیندی است که از طریق امواج فوق العاده سریع صدا، تصاویری از بافت‌های داخلی سینه تولید کرده و در نهایت پس از شکل‌گیری به مانیتور ارسال می‌شود تا قابل مشاهده باشد. این روش بدون درد بوده و در آن اشعه به کار نمی‌رود. تشکیل تصاویر در این روش تصویربرداری نیاز به امواج با فرکانس معینی دارد اما عملکرد بدون اشعه ایکس آن باعث برتری اش بر ماموگرافی شده است. تکنیسین با استفاده از یک وسیله کوچک دستی به نام ترانسدایوسر، امواج را به طور مستقیم از طریق پوست به درون بدن می‌فرستد. برای کمک به انتقال امواج از یک ژل لیزکننده استفاده می‌شود.

این روش به ویژه در تعیین اینکه آیا یک منطقه متراکم در سینه یک توده جامد و احتمالاً سرطانی است و یا تنها یک کیست خوش خیم حاوی مایع، کاربرد دارد. امواج صوتی دارای فرکانس بالا از مایع عبور میکنند، بنابراین در سونوگرام یا همان تصویر ایجاد شده در سونوگرافی، کیست‌ها شبیه دوایر تیره به نظر می‌رسد.

حتی بدون ماموگرافی نیز پزشکان برای بررسی توده‌های سینه به ویژه در صورتی که زن باردار باشد، از سونوگرافی استفاده می‌کنند. اگرچه در ماموگرافی خطر ناچیزی وجود دارد اما نه پزشکان و نه بیماران آنها نمی‌خواهند که جنین در معرض اشعه قرار بگیرد، به همین دلیل روش سونوگرافی را انتخاب می‌کنند. تراکم سینه در نتیجه سونوگرافی تأثیری ندارد. بنابراین این روش در جستجوی ناهنجاری‌های سینه در زنان دارای سینه متراکم نیز مناسب است.

سونوگرافی را نمی‌توان به عنوان تست جایگزین ماموگرافی در غربالگری سرطان پستان به کار برد. این به آن دلیل است که اولتراسوند می‌تواند تنها بخش کوچکی از سینه را در زمان واحد نشان دهد. مهمتر اینکه اولتراسوند توانایی نشان دادن برخی MC ها را نیز ندارد.



شکل (۱-۴): نمونه‌ای از یک تصویر اولتراسوند یا سونوگرافی سینه

۳-۱ موارد استفاده ماموگرافی

ماموگرافی به عنوان یک ابزار آزمایشی برای تشخیص زودهنگام سرطان سینه در زمانی که بیمار هیچ نشانه‌ای ندارد استفاده می‌شود. همچنان ماموگرافی برای شناسایی و تشخیص بیماری‌های سینه در زنانی که نشانه‌هایی همچون لمس توده، درد و یا ترشحات در نوک سینه دارند نیز استفاده می‌شود.

ماموگرافی به دلیل اینکه توانایی تشخیص تغییر سینه ۲ سال قبل از ایجاد علائم را داراست نقش اساسی در تشخیص زود هنگام سرطان سینه بازی می‌کند. ماموگرافی غربالگری^۱ به هدف تشخیص علائم سرطان سینه در یک گروه فاقد نشانه‌های بیماری انجام می‌شود [۱۵].

دستورالعمل‌های کنونی از وزارت بهداشت و خدمات انسانی ایالات متحده^۲ و انجمن سرطان آمریکا^۳ غربالگری سالیانه زنان بالای ۴۰ سال را توصیه می‌کند. تحقیقات نشان داده است که ماموگرافی سالیانه منجر به تشخیص زودهنگام سرطان سینه می‌شود یعنی زمانی که آنها اکثراً قابل درمان می‌باشند و نیازی به برداشتن کامل سینه نمی‌باشد. انستیتوی ملی سرطان^۴، برای خانم‌هایی که سابقه سرطان سینه دارند و همچنین آنهایی که از لحاظ ژنتیکی ریسک بیشتری دارند آغاز غربالگری سرطان سینه را قبل از ۴۰ سالگی و همچنین با توالی بیشتر توصیه می‌کند [۱۵].

ماموگرافی تشخیصی^۵ جهت بررسی بیمارانی که یافته بالینی غیر طبیعی (همچون لمس توده) دارند استفاده می‌شود. ماموگرافی تشخیصی همچنین برای معاینه قسمت‌هایی که توسط ماموگرافی غربالگری غیرطبیعی و نگران‌کننده شناخته شده است، انجام می‌شود. همچنان بررسی می‌کند که آیا این مناطق مشکوک نیازمند تصویربرداری‌های اضافی از سینه برای تعیین اینکه آیا بیمار دچار سرطان سینه شده است، وجود دارد یا خیر [۱۴].

¹ Screening mammography

² Health and Human Services (HHS)

³ the American Cancer Society (ACS)

⁴ National Cancer Institute (NCI)

⁵ Diagnostic mammography

۴-۱ عوامل سرطانی قابل شناسایی در تصاویر ماموگرافی

۱-۴-۱ توده‌ها (mass)

توده‌ها معمولا با شکل و مرزشان دسته‌بندی می‌شوند. در این پایان نامه نیز تمرکز روی این علامت از سرطان سینه بر روی تصاویر ماموگرافی می‌باشد. از همین رو انواع مختلف آن را به تفصیل بیان می‌کنیم. از نظر شکل، توده‌ها به چهار دسته گرد^۱، بیضی^۲، لاله‌ای^۳ و نامنظم^۴ تقسیم‌بندی می‌شوند [۱۶]. شکل (۵-۱) طرحی از انواع این شکل‌ها را نشان می‌دهد. در شکل (۶-۱) نمونه‌های واقعی از هر کدام از آنها نشان داده شده است. وجود هر کدام از این اشکال در خوش‌خیمی و بدخیمی توده موثر است. جدول (۱-۱) احتمال خوش‌خیمی یا بدخیمی را برای هر کدام از این اشکال نشان می‌دهد.



نامنظم

لاله‌ای

تخم مرغی

گرد

شکل (۵-۱): انواع شکل برای توده‌ها [۱۶]

¹ Round

² Oval

³ Lobular

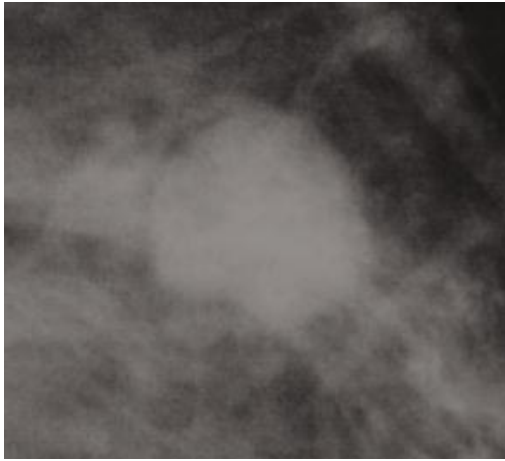
⁴ Irregular

جدول (۱-۱): احتمال خوش خیمی و بدخیمی برای اشکال مختلف توده [۱۶]

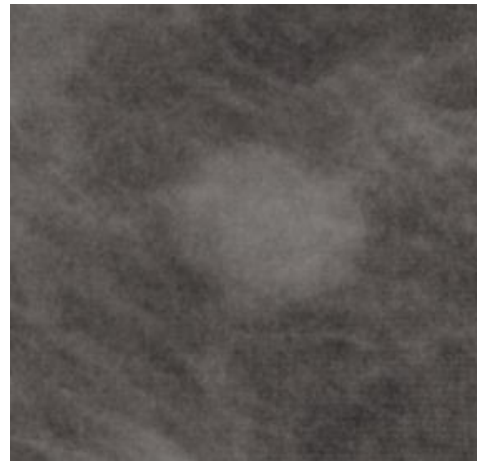
شکل	احتمال خوش خیم	مشکوک	احتمال بدخیم
گرد	×		
تخم مرغی	×		
لاله ای	×	×	
نامنظم			×

از نظر مرز، توده‌ها به پنج دسته خوش تعریف^۱، مبهم^۲، ریزلاله‌ای^۳، بد تعریف^۴، و سوزنی^۵ تقسیم می‌شوند که به ترتیبی که نام برده شدند، احتمال بدخیمی افزایش می‌یابد. در مرزهای خوش تعریف بین بافت مجاور و ضایعه انتقال روشنایی شدیدی وجود دارد. در مرزهای مبهم قسمتی از مرز، درون بافت مجاور قرار می‌گیرد و قابل تشخیص نیست. در مرزهای ریز لاله‌ای منحنی‌های کوچکی روی مرز مشاهده می‌شود. مرزهای بد تعریف به راحتی قابل تشخیص نیستند و در واقع پراکنده هستند.

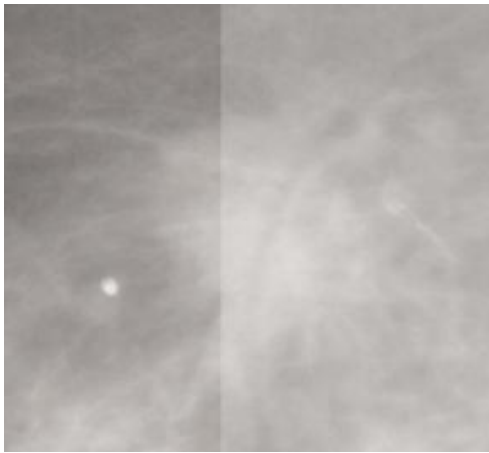
¹ Well-defined
² Obscured
³ Micro-lobulated
⁴ Ill-defined
⁵ Spiculated



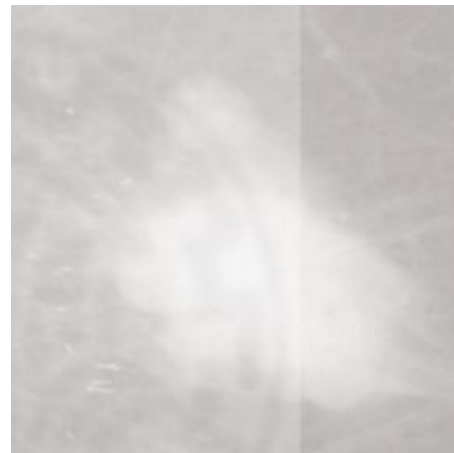
ب



الف



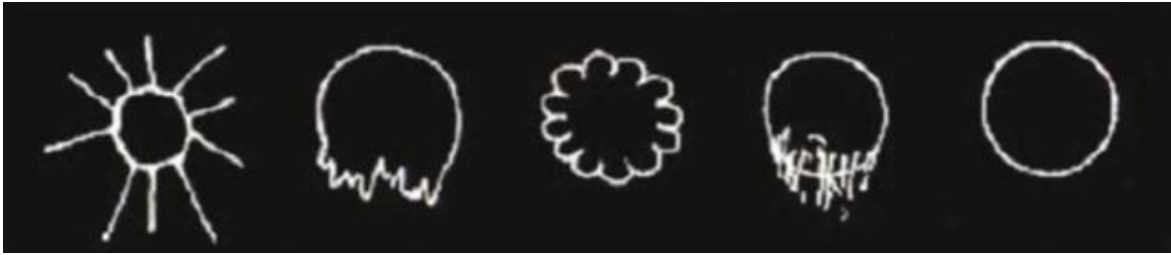
د



ج

شکل (۶-۱) : نمونه‌هایی از اشکال مختلف توده‌ها (الف) گرد ب) تخم مرغی ج) لاله‌ای د) نامنظم [۱۶]

مرزهای سوزنی با خطوط شعاعی که به اطراف می‌رود شناخته می‌شوند. می‌توان گفت که بیشتر توده‌های خوش‌خیم دارای مرزهای خوش تعریف هستند [۱۶]. جدول (۲-۱) احتمال خوش‌خیمی و بدخیمی را برای انواع مرز نشان می‌دهد. در شکل (۷-۱) طرح کلی انواع مرزها نمایش داده شده است. در شکل (۸-۱) مثالی از تصاویر ماموگرافی از هر نوع آورده شده است. در میان این مرزها نوع سوزنی به عنوان یکی از انواع بسیار خطرناک شناخته می‌شود و در مقالات مواردی را نیز به خود اختصاص داده است.



سوزنی

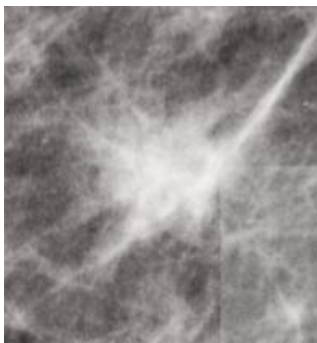
بد تعریف

ریز لاله‌ای

مبهم

خوش تعریف

شکل (۷-۱): انواع مرز توده‌ها [۱۶]



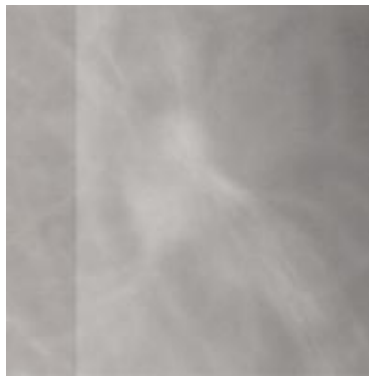
ج



ب



الف



ه



د

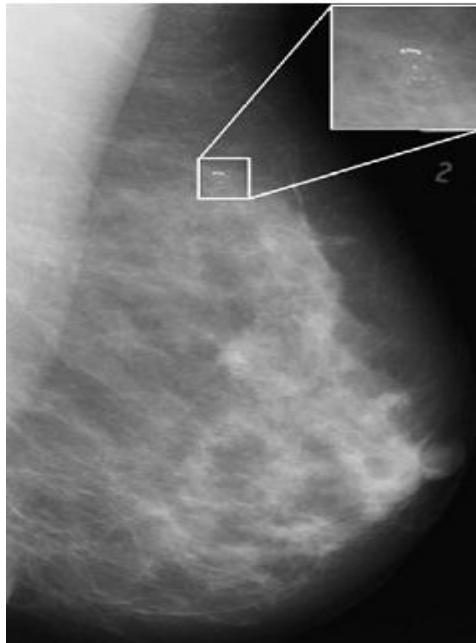
شکل (۸-۱): نمونه ای از مرزهای مختلف توده‌ها الف) خوش تعریف ب) مبهم ج) ریز لاله‌ای د) بد تعریف ه) سوزنی [۱۶]

جدول (۲-۱): احتمال خوش‌خیمی و بدخیمی برای مرزهای مختلف توده

مرز	احتمال خوش‌خیم	مشکوک	احتمال بدخیم
خوش‌تعریف	×		
مبهم		×	
ریزلاله‌ای		×	
بدتعریف		×	×
سوزنی			×

۲-۴-۱ دسته‌های ریز کلسیمی (Micro Calcifications)

یک دسته از نشانه‌هایی که می‌تواند در تصاویر ماموگرافی دلیلی بر بروز سرطان سینه باشد، ذرات ریز کلسیمی یا به اختصار MCها هستند. اینها ذرات بسیار کوچکی از جنس ترکیبات کلسیم می‌باشند که به دلایل متعددی از جمله سرطان سینه تولید می‌شوند. این ذرات در ماموگرام‌ها به صورت نقاط کوچک و روشنی دیده می‌شوند که اندازه‌ی آنها بین ۰.۱ تا ۱ میلی‌متر متغیر است. MCهایی که در بافت‌های بدخیم تولید می‌شوند، اغلب به صورت خوشه‌ای به چشم می‌خورند. رادیولوژیست‌ها خوشه‌های MC را با توجه به مشخصاتی مانند تعداد، شکل، اندازه و موقعیت آنها طبقه‌بندی می‌کنند، ولی به دلیل پایین بودن کنتراست این ذرات نسبت به زمینه‌ی اطراف آنها، تشخیص قطعی خوش‌خیم یا بدخیم بودن MCها همیشه امکان‌پذیر نخواهد بود و بنابراین پس از تصویربرداری، نمونه‌برداری از بافت‌های مشکوک تجویز می‌شود. توده‌های MC می‌توانند نماینده خوبی از سرطان سینه باشند. آنها در ۳۰٪ تا ۵۰٪ از موارد تشخیص داده شده با تصویربرداری ماموگرافی مشاهده می‌شوند [۱۷]. در شکل (۱-۹) نمونه‌ای از این ذرات نشان داده شده است. در دو دهه گذشته تحقیقات گسترده‌ای در روش‌های کامپیوتری تشخیص این ذرات در ماموگرام‌ها صورت گرفته است، مقاله [۱۸] این روش‌ها را بررسی کرده است.



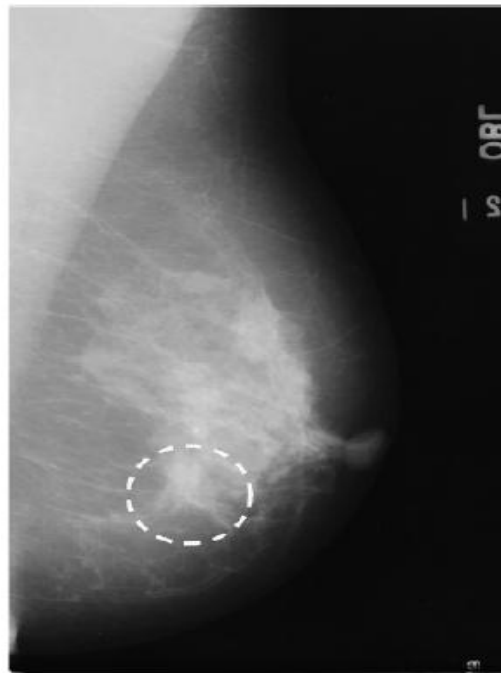
شکل (۱-۹): نمونه‌ای از تصویر ماموگرافی شامل MC

۳-۴-۱ ناهنجاری ساختاری

ناهنجاری ساختاری^۱ در [۱۹] به این صورت تعریف شده است: "بافت معمول از شکل خارج شده است در حالی که هیچ توده‌ای قابل مشاهده نیست، شعاع‌های سوزنی که از یک نقطه خارج می‌شوند از این نوع می‌باشند." یک تصویر ماموگرافی شامل ناهنجاری ساختاری در شکل (۱-۱۰) نشان داده شده است. اگرچه ناهنجاری‌های ساختاری کمتر از توده‌ها و MCها رایج هستند، اما سومین نشانه معمول سرطان سینه در ماموگرافی است. این نشانه در اغلب موارد به علت ظرافت و شکل پنهان آن، در طی معاینه غربالگری از بین می‌رود. وجود این نشانه در ۱۲٪ تا ۴۵٪ از سرطان‌های سینه چشم‌پوشی شده و یا به اشتباه تفسیر می‌شوند [۲۰ و ۲۱]. ناهنجاری ساختاری می‌تواند بیانگر تغییرات بدخیمی باشد به خصوص هنگامی که با ضایعات قابل مشاهده از جمله توده‌ها یا MCها همراه شود.

¹ Architectural distortion

متدهای تشخیص ناهنجاری ساختاری اغلب شامل الگوریتم‌های تشخیص توده می‌باشد. با این حال متدهایی که منحصرًا برای تشخیص ناهنجاری ساختاری طراحی شده‌اند عملکرد بهتری نسبت به متدهایی که برای تشخیص توده‌های سوزنی طراحی شده‌اند دارند [۲۲].



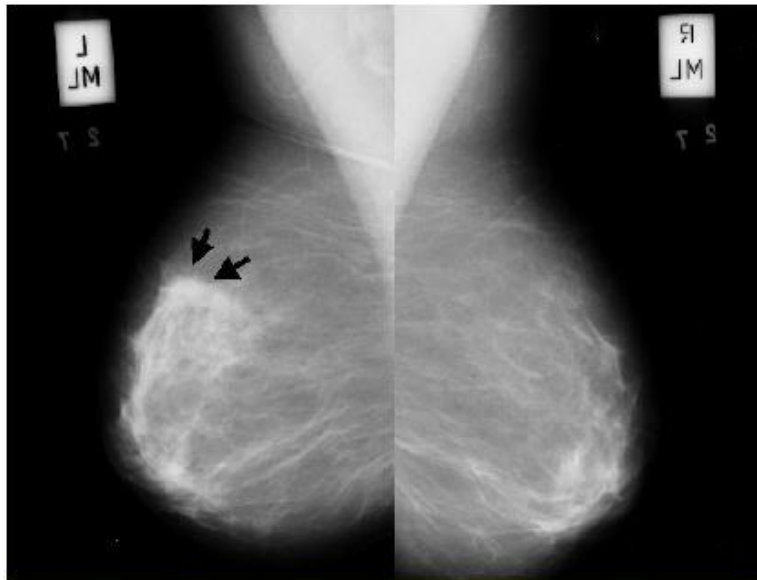
شکل (۱-۱۰): تصویر ماموگرافی شامل اعوجاج ساختاری [۲۲]

۴-۴-۱ عدم تقارن دو طرفه^۱

وجود عدم تقارن بین ماموگرام‌های سینه چپ و راست فرد مورد نظر، نشانه مهمی است که توسط رادیولوژیست برای تشخیص زود هنگام سرطان سینه مورد استفاده قرار می‌گیرد [۲۳ و ۲۴]. عدم تقارن اشاره به تغییر در بافت و چگالی یک قسمت، بدون حضور توده‌ای مشخص نسبت به تصویر دیگر دارد. تعداد مقالات چاپی در این زمینه نسبت به دیگر انواع ناهنجاری بسیار کمتر می‌باشد. روش مرسوم در بیشتر این کارها استفاده از هم‌ترازی دو تصویر است و دشواری کار نیز یافتن نقاط متناظر و همچنین

^۱ Bilateral asymmetry

عدم تقارن طبیعی در بعضی حالات می‌باشد. نمونه‌ای از عدم تقارن در تصویر ماموگرافی موجود در شکل (۱-۱۱) نشان داده است.



شکل (۱-۱۱): تصویر ماموگرافی شامل عدم تقارن [۲۴]

۵-۱ سامانه‌های تشخیص پزشکی (CAD)

یکی از مشکلات ماموگرام‌ها کنتراست نسبتاً پایین آنها می‌باشد. این موضوع باعث می‌شود که تفسیر تصاویر ماموگرافی برای رادیولوژیست مشکل شود. در واقع قضاوت آنها براساس آموزش، تجربه و معیارهای ذهنی می‌باشد. مطالعات نشان می‌دهد که تصاویر ماموگرافی دارای نرخ تشخیص مثبت کاذب^۱ (FP) و منفی کاذب^۲ (FN) بالایی هستند. این موضوع سبب می‌شود که تعدادی از زنانی که مبتلا به این بیماری نیستند تحت مراقبت‌های کلینیکی و نمونه‌برداری از سینه قرار بگیرند و یا تعدادی از زنان مبتلا به سرطان زمان مناسب برای درمان را از دست بدهند [۲۵-۲۸].

^۱ False Positive

^۲ False Negative

دوباره خوانی^۱ تصاویر روشی است که سهم نمونه‌های سرطانی تشخیص داده نشده را کاهش می‌دهد [۲۷ و ۲۸]. ایده اصلی این است که از دو رادیولوژیست در خواندن استفاده شود. در مرجع [۲۸] نشان داده شده است که دوباره خوانی می‌تواند حساسیت و تاثیر تصویربرداری را در تشخیص افزایش دهد. به هر حال حجم کاری و هزینه مربوط به دوباره‌خوانی زیاد است. به جای آن، سامانه‌های تشخیص پزشکی (CAD) به عنوان "چشم دوم رادیولوژیست" به رادیولوژیست‌ها در تفسیر تصاویر ماموگرافی کمک می‌کند [۲۹]. به علت محدودیت‌های دید بشر، کامپیوترها نقش مهمی را در تشخیص علائم اولیه سرطان بازی می‌کنند. ترکیبی از طرح CAD و دانش متخصصان تا حد زیادی دقت تشخیص را بهبود می‌بخشد. حساسیت تشخیص بدون CAD در حدود ۸۰٪ است و با CAD تا ۹۰٪ افزایش می‌یابد [۳۰].

استفاده از روش‌ها و تکنیک‌های علوم کامپیوتر^۲، تشخیص الگو^۴، پردازش تصویر^۵ و هوش مصنوعی^۶ در تشخیص و تعیین خودکار سرطان سینه از روی تصاویر ماموگرافی در سیستم‌های CAD، باعث کم شدن خطاهای انسانی و افزایش سرعت تشخیص می‌شود. سیستم‌های CAD به رادیولوژیست‌ها در تعیین محل بیماری‌های بدخیم در تصاویر ماموگرافی دیجیتال کمک می‌کنند. این سیستم‌ها اولین بار در دهه ۱۹۹۰ استفاده شدند و سپس به مرور زمان با افزایش مزایا و دقت این سیستم‌ها استفاده از آنها افزایش پیدا کرد [۳۱].

۱-۵-۱ مزایای استفاده از سیستم‌های CAD

با استفاده از یک سیستم CAD به جای دو رادیولوژیست تنها یک رادیولوژیست برای خواندن تصاویر مورد نیاز است. استفاده از CADها حجم کاری کارشناسان را کاهش می‌دهد. ثابت شده است که

¹ Double reading

² Computer Aide Detection

³ Computer Science

⁴ Pattern Recognition

⁵ Image processing

⁶ Artificial intelligence

سیستم‌های CAD، احتمال تشخیص سرطان سینه را در مراحل اولیه آن افزایش می‌دهد. برای مثال Morton et al. در مرجع [۳۲] نشان داده‌اند که استفاده از سیستم‌های CAD باعث می‌شود که میزان تشخیص سرطان سینه در تصاویر ماموگرافی ۷.۶۲٪ افزایش یابد.

تشخیص نهایی تمامی انواع بیماری‌های سینه وابسته به نمونه‌برداری^۱ است. در بیشتر موارد تصمیم برای نمونه‌برداری بر مبنای یافته‌های حاصل از ماموگرافی است. نتایج نمونه‌برداری نشان می‌دهد که ۶۵-۹۰٪ از سرطان‌های مشکوک شناسایی شده توسط ماموگرافی، مربوط به تومورهای خوش خیم خیم است. در نتیجه یکی از مزایای دیگر استفاده از سیستم‌های CAD، کمترین افزایش در میزان نمونه‌برداری در موارد خوش خیم است.

همچنین Bern et al. در مرجع [۳۳] نشان داده‌اند که CADها تشخیص درست سرطان سینه را با افزایش میزان حساسیت رادیولوژیست‌ها به مقدار ۲۱.۲٪ بهبود می‌بخشند.

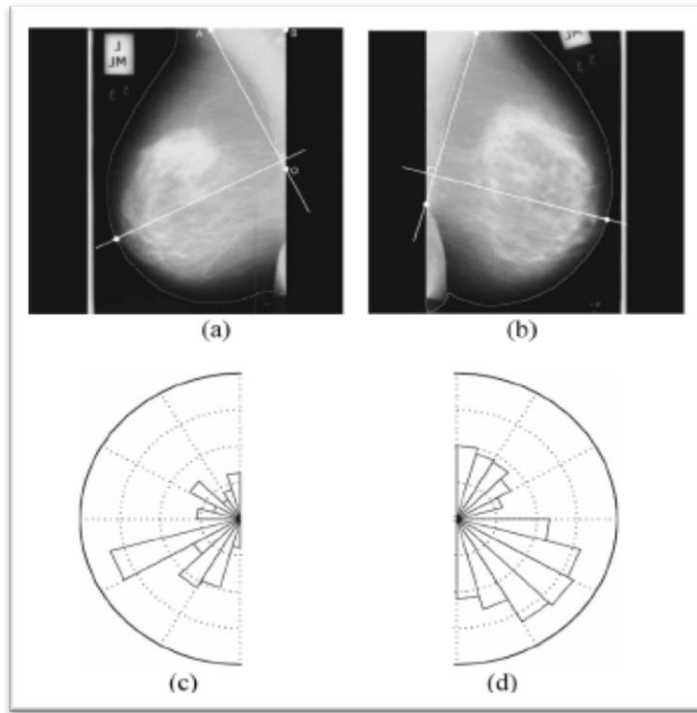
۲-۵-۱ نماهای چند گانه تکنیک‌های سیستم‌های CAD

به طور کلی حساسیت سیستم‌های CAD در تشخیص توده‌های سرطانی نسبتاً بالا است. برای کاهش نرخ مثبت تشخیص‌های غلط تلاش‌های زیادی برای گسترش تکنیک‌های CAD ای که از اطلاعات زمینه و تقارن و نماهای چند گانه استفاده می‌کنند وجود دارد [۴].

اکثر سیستم‌های CAD تنها از اطلاعات یک نمای تصویر در زمان استفاده می‌کنند. در مواردی که بهبود کارایی CADها مورد نظر است نتایج همه نماها در پایان به طور مستقل ترکیب می‌شوند. به عبارت دیگر رادیولوژیست از اطلاعات تمام نماها در تفسیر خود استفاده می‌کند. آنها اینگونه آموزش داده شده‌اند که از مقایسه تقارن میان تصویر سینه چپ و راست نواحی مشکوک را تشخیص دهند (شکل ۱-۱۲).

^۱ Biopsy

مقایسه زودگذر هر دو جفت ماموگرام سبب می‌شود که نواحی جدید و یا در حال رشد سرطان در ماموگرام تشخیص داده شود. این حقیقت که استفاده از اطلاعات تصاویر MLO و CC به طور همزمان دقت تشخیص سرطان را بهبود می‌بخشد، امروزه برای همه شناخته شده است.



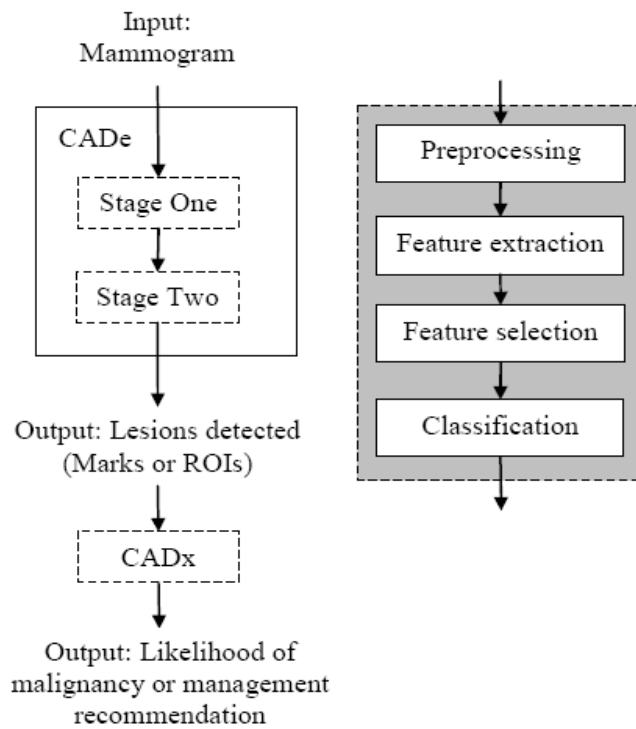
شکل (۱-۱۲): تصاویر (a) و (b) نمونه ای از عدم تقارن را نشان می‌دهد و تصاویر (c) و (d) روشی برای یافتن این نامتقارن بودن است [۴].

۳-۵-۱ شمای کلی سیستم‌های CAD

در بیشتر سیستم‌های CAD پیشرفته، چند قدم اساسی و متداول وجود دارد که باید به منظور یافتن نواحی سرطانی انجام شود. مراحل فوق در شکل (۱-۱۳) آورده شده است [۳۴].

بیشتر الگوریتم‌های سیستم‌های تشخیص سرطان CAD، دو مرحله اساسی دارند. در مرحله اول، هدف یافتن ضایعه‌های مشکوک در یک حساسیت بالا است. در مرحله بعدی هدف کاهش نرخ مثبت کاذب است بدون اینکه حساسیت تغییر چشمگیری داشته باشد. نرخ مثبت کاذب درصد نمونه‌های

نرمال و فاقد بیماری بوده که توسط سیستم CAD به عنوان بیماری تشخیص داده شده است. در بعضی از دیدگاه‌ها، ممکن است بعضی از مراحل شامل روش‌های خیلی ساده‌ای شوند و یا به کلی حذف گردند.



شکل (۱-۱۳): مراحل اصلی الگوریتم سامانه‌های تشخیص پزشکی CAD و سامانه‌های تشخیص عیب به کمک کامپیوتر CADx [۳۴].

بیشتر الگوریتم‌های سیستم‌های تشخیص عیب^۱ CADx با تشخیص نواحی مشکوک (ROI)^۲ شروع می‌شوند. مراحل طراحی چنین سیستمی نیز در شکل (۱-۱۳) آورده شده است [۳۵]. هر یک از گروه‌های پژوهشی ممکن است بر روی اجزای مختلفی از مسئله کار کنند، به عنوان مثال خیلی از الگوریتم‌های CAD موجود کار خود را با ناحیه‌بندی دستی ROI ها شروع می‌کنند.

^۱ Diagnosis
^۲ Region Of Interest

فصل دوم:

مروری بر کارهای گذشته در شناسایی ضایعه های MC و توده

۱-۲ مقدمه

در حال حاضر ماموگرافی موثرترین روش برای تشخیص زودهنگام و غربالگری سرطان سینه است. MCها و توده‌ها دو نشانه اولیه مهم از این بیماری در تصاویر ماموگرافی هستند [۳۶]. در این فصل مرور مختصری بر الگوریتم‌های معرفی شده در مقالات متعدد برای تشخیص این دو نشانه بیان می‌شود.

تشخیص توده‌ها به دلیل ویژگی‌هایشان که می‌تواند مبهم و یا شبیه به بافت نرمال سینه باشد بسیار دشوارتر از MCها است. توده‌ها کاملاً نامحسوس هستند و اغلب در مناطق متراکم از بافت سینه رخ می‌دهند، لذا تشخیص و دسته‌بندی خودکار توده‌ها چالش برانگیزتر است، به همین دلیل است که در این پایان‌نامه به طور خاص تمرکز روی شناسایی این نشانه از سرطان سینه بر روی تصاویر ماموگرافی است.

۲-۲ روش‌های تشخیص MCها

همانطور که در مرجع [۴] اشاره شده است تشخیص MCها را می‌توان در ۴ گروه کلی دسته‌بندی کرد:

- (۱) روش‌های مبتنی بر بهبود کیفیت تصویر
- (۲) روش‌های مبتنی بر مدل‌های آماری
- (۳) روش‌های تجزیه چند مقیاسی
- (۴) روش‌های آموزش ماشین

۲-۲-۱ روش‌های مبتنی بر بهبود کیفیت تصویر

ایده اصلی این روش‌ها با این انگیزه به وجود آمده است که MC ها نقاطی روشن‌تر از محیط اطرافشان هستند. بنابراین این روش‌ها بر روی بهبود کنتراست تصویر کار می‌کنند و سپس با استفاده از یک آستانه، MCها را از اطراف آن جداسازی می‌کنند. برای مثال در مرجع [۳۷] بهبود کیفیت با استفاده از تکنیک تصویر اختلافی^۱ به همراه اپراتورهای مورفولژیکی جهت کاهش نرخ False Positive انجام می‌گیرد. تصویر اختلافی با استفاده از دو فیلتر به دست می‌آید. فیلتر اول برای بهبود کنتراست MCها و فیلتر دوم برای فشرده‌سازی آن‌ها به کار می‌روند. یک ایده متداول در روش‌ها، اعمال روش‌های بهبود کیفیت تصویر به عملیات پردازش تصویر در مرحله‌ی preprocessing یا post-processing است. مزیت این کار راحتی انجام، و پیاده‌سازی و تاثیر آن در عملیات و کاربردهای کلینیکی می‌باشد.

۲-۲-۲ روش‌های مبتنی بر مدل‌های آماری

ایده اصلی این روش‌ها براساس تفاوت‌های آماری MCها با محیط اطرافشان است. برای مثال در مرجع [۳۸] تفاوت‌ها در مشخصات آماری مرتبه‌های بالا برای مثال skewness (moment مرتبه سوم) و kurtosis (moment مرتبه چهارم) استفاده شده است. با این فرض که نواحی شامل MC دارای توزیع مشابه گوسی و نواحی شامل MC دارای توزیع مشابه غیر گوسی است، می‌توان MCها را از محیط اطرافشان جداسازی کرد. چرا که توزیع گوسی دارای skewness و kurtosis برابر با صفر می‌باشد و توزیع غیرگوسی دارای skewness و kurtosis غیر صفر است.

در بعضی از تحقیقات اخیر از مدل پنهان مارکوف (MRF) در تشخیص MCها استفاده شده است. استفاده از مدل MRF برای جداسازی تصویر نسبت به دیگر مدل‌های آماری بسیار ارجحیت دارد، زیرا

¹ Difference Image

توانایی این مدل در کاراکتربندی توزیع چگانی حوزه فضایی^۱ تصویر بالا است. به هر حال تخمین یک توزیع اولیه مناسب در این مورد یک موضوع مخاطره‌آمیز محسوب می‌شود.

۳-۲-۲ روش‌های تجزیه چند مقیاسی

در این دسته از روش‌ها تلاش بر این است که تفاوت‌های فرکانسی میان لکه روشن MC و اطرافش را به عنوان ویژگی متمایز کننده، به کار بگیرند. به طور معمول تبدیل موجک به طور بسیار گسترده برای تشخیص MCها مورد استفاده قرار گرفته است. در مرجع [۳۹] از تبدیل un-decimated bi-orthogonal wavelet استفاده شده است که در آن MCها به صورت اشکال گوسی دوار با عرض‌های متغیر در مقیاس‌های متفاوت نمایش داده می‌شوند. سپس با استفاده از اعمال یک تابع جداساز fisher ضرایب وزندهی می‌شود و از زیرباندهای^۲ دوم و سوم برای بهبود تشخیص و جداسازی MCها استفاده می‌شود.

در تحقیقات اخیر در مرجع [۴۰] نشان داده شده است که استفاده از تبدیل موجک Matzinger به جای تبدیل‌های موجک Daubechies برای تشخیص MCها نتایج بهتری بدست می‌دهد.

در اینجا به دلیل اهمیت و تاثیر تبدیل موجک و تجزیه چندمقیاسی یک دسته جداگانه از روش‌های تشخیص به آن‌ها اختصاص داده شد. این روش‌ها به طور متداول در مرحله استخراج ویژگی وبه عنوان ورودی برای طبقه‌بندها استفاده می‌شوند.

۴-۲-۲ روش‌های آموزش ماشین

هدف روش‌های آموزش ماشین یافتن وابستگی‌های متقابل داده‌ها است. در زمینه تشخیص MC، مساله اساسی عمل کردن به صورت یک طبقه‌بند باینری است که هدف آن این است که مشخص کند

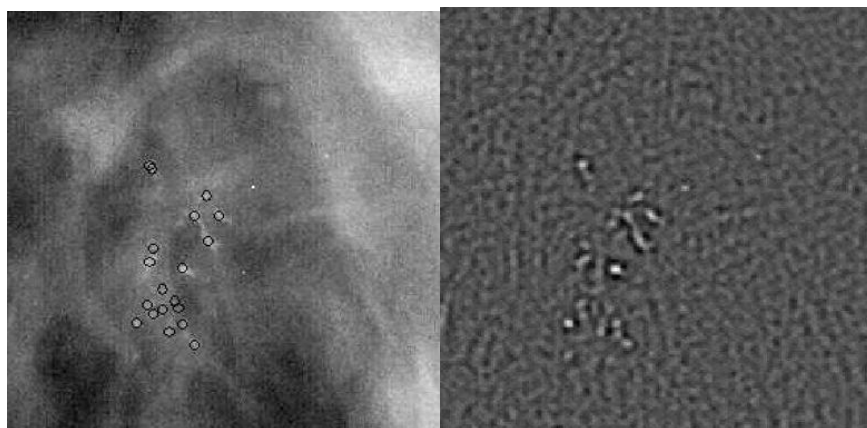
¹ Spatial

² Sub band

آیا MC در یک پیکسل موجود است یا خیر. برای مثال، در مرجع [۴۱] یک شبکه عصبی دولایه استفاده شده است. ویژگی‌های این طبقه‌بند اجزای فرکانسی موجک، مشخصات آماری (gray-scale) و ویژگی‌های شکل هستند. لایه اول در شبکه عصبی پیکسل‌هایی را که پتانسیل MC بودن را دارند شناسایی می‌کند و لایه دوم MC را به عنوان یک شی تشخیص می‌دهد.

روش‌های آموزش ماشین سهم زیادی از تحقیقات اخیر را شامل می‌شود. ثابت شده است که طبقه‌بندهایی که براساس روش‌های آموزش ماشین طراحی شده‌اند بسیار قدرتمند هستند. پیشرفت‌های اخیر در آموزش ماشین در زمینه الگوریتم‌های شناخته شده SVM^۱ هستند.

در شکل (۱-۲) یک مثال از اعمال طبقه‌بند SVM بر روی تصویر ماموگرافی دیده می‌شود که در آن MCها به صورت مکان‌های روشن در خروجی طبقه‌بند مشخص شده است. گزارش شده است که SVMها بالاترین دقت را در تشخیص MCها دارند.



شکل (۱-۲): اعمال طبقه‌بند SVM برای تشخیص کلاستر MCها. تصویر سمت چپ نمونه‌ای از کلاستر MC در

تصاویر ماموگرافی. تصویر سمت راست نتیجه اعمال طبقه‌بند SVM بر روی تصویر سمت چپ [۴]

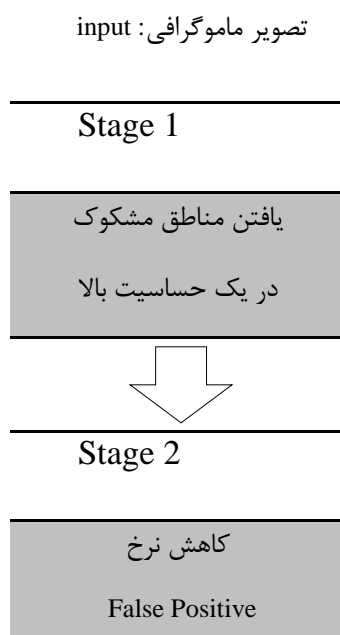
^۱ Support Vector Machine

۳-۲ روش‌های تشخیص توده

توده‌های با حاشیه‌های سوزنی احتمال بسیار بالایی برای بدخیمی سرطان می‌باشند. به همین دلیل بسیاری از متدها به طور خاص بر روی تشخیص توده‌های سوزنی متمرکز شده‌اند. توده‌های سوزنی توسط خطوط منشعب شده از حاشیه توده‌ها شناسایی می‌شوند. اما با این وجود به این دلیل که همه توده‌های بدخیم سوزنی نیستند در نتیجه تشخیص توده‌های غیر سوزنی نیز با اهمیت است. بیشتر الگوریتم‌های تشخیص توده از دو مرحله اساسی تشکیل می‌شوند شکل (۲-۲):

(۱) تشخیص نواحی مشکوک در تصاویر ماوگرافی

(۲) طبقه‌بندی نواحی مشکوک به عنوان ناحیه نرمال یا توده سرطانی



شکل (۲-۲): مراحل کلی الگوریتم‌های تشخیص توده

مرحله اول با حساسیت بسیار بالایی طراحی شده است و دارای نرخ مثبت کاذب بالایی است، چراکه انتظار می‌رود در مرحله بعدی تعداد بسیار زیادی از آنها حذف شوند. الگوریتم‌های مرحله اول تشخیص توده‌ها معمولاً Pixel-based یا Region-based هستند [۴۲ و ۴۳]. هدف مرحله دوم کاهش تعداد نرخ مثبت کاذبی است که در پایان مرحله قبل تولید شده است.

۱-۳-۲ روش‌های Pixel-based

در دیدگاه‌های pixel-based، ابتدا ویژگی‌ها برای هر پیکسل استخراج می‌شود و سپس پیکسل‌ها به دو دسته مشکوک و یا نرمال دسته‌بندی می‌شوند [۴۲]. این کار ممکن است به سادگی و با استفاده از اعمال یک مقدار آستانه برای feature ها، و یا با استفاده از تکنیک‌های دسته‌بندی پیچیده انجام شود. واژه pixel-based گمراه‌کننده است به این دلیل که ویژگی‌های هر پیکسل از همسایگی محلی آن پیکسل استخراج می‌شود. در نهایت پیکسل‌های مشکوک توسط پیکسل‌های متصل به هم در یک ناحیه گروه‌بندی می‌شوند. این نکته بسیار مهم است که پیکسل‌هایی که توسط این الگوریتم‌ها به عنوان دسته مشکوک گروه‌بندی می‌شوند لزوماً نشان‌دهنده بدخیمی نیستند. دسته‌بندی مناطق مشکوک به توده‌های خوش‌خیم و بدخیم، مسئله متفاوتی است.

تحقیقات زیادی در زمینه روش‌های pixel-based تاکنون صورت گرفته است. برای مثال در مرجع [۴۴]، ویژگی‌ها^۱ و مشخصات لبه‌های نزدیک به هم محلی از نواحی مورد علاقه (ROI) استخراج می‌شوند و یک طبقه‌بند درختی با تصمیم‌گیری باینری برای دسته‌بندی مناطق مشکوک از نواحی نرمال به کار می‌رود. نویسندگان این مقاله ۱۰۰٪ حساسیت را برای این روش گزارش داده است.

در مرجع [۴۵]، یک شماتیک چند-رزولوشن برای تشخیص توده‌ها ارائه شده است. در این روش ابتدا تصویر به نمایش چند رزولوشن تجزیه می‌شود و چهار ویژگی برای هر پیکسل در هر رزولوشن

^۱ feature

استخراج می‌شود. نویسنده در آزمایشات خود از ۱۹ عدد تصویر ماموگرافی از بانک تصاویر MIAS استفاده کرده است.

۲-۳-۲ روش‌های Region-based

در روش‌های فوق ابتدا ناحیه‌های مشکوک از تصویر اولیه جدا می‌شوند و سپس ویژگی‌ها از هر ناحیه مشکوک استخراج می‌شود. در نهایت این نواحی هستند که در دو گروه سرطانی و یا نرمال دسته‌بندی می‌شوند.

تاکنون مقالات زیادی در این زمینه ارائه شده است. برای مثال در سال ۲۰۱۱ در مرجع [۴۶] یک الگوریتم Region-based برای شناسایی توده معرفی شده است. در این روش ویژگی‌ها بافتی از ماتریس‌های هم-رخداد^۱ مربوط به نواحی مشکوک استخراج می‌شود. در نهایت این ویژگی‌ها به عنوان ورودی دسته‌بندی کننده شبکه عصبی پرسپترون چند لایه برای جداسازی نواحی مربوط به توده از نواحی سالم استفاده می‌شود. طبق گزارش، عملکرد این روش بر روی تصاویر پایگاه داده DDSM ارزیابی شده است. و نرخ حساسیت ۹۱.۶۷ درصد و اختصاصیت ۸۴.۱۷ گزارش شده است. علی‌رغم نرخ حساسیت بالای این سیستم، یکی از مشکل‌های آن، دستی جدا شدن ناحیه‌های مشکوک در مرحله نخست است. خودکار نبودن آن و در نتیجه ورود خطاهای ناشی از دخالت انسان از جمله ضعف‌های این سامانه است.

همچنان در مرجع [۴۷] یک روش دیگر برای شناسایی توده‌ها معرفی شد است. در این روش برای یافتن ناحیه‌های مشکوک از متدهای ناحیه‌بندی مختلفی از قبیل Otsu استفاده شده است. با استفاده از ویژگی‌های استخراج شده از ماتریس هم-رخداد مربوط به ناحیه‌های مشکوک و طراحی یک درخت تصمیم، ناحیه‌های مربوط به توده شناسایی شده است. مزیت این سامانه، خودکار، ساده و

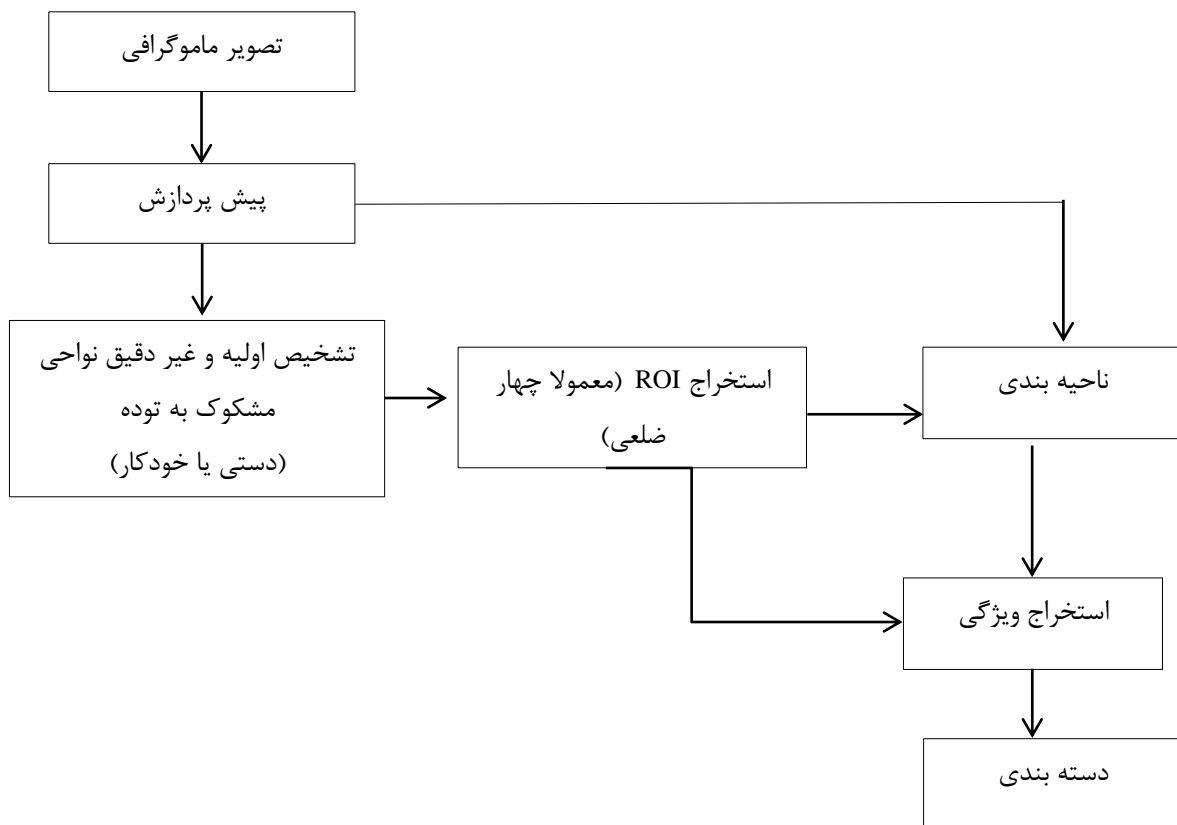
^۱ Co- Occurrence martrix

سریع بودن آن و ضعف آن انتشار خطای ناحیه بندی در شناسایی توده‌ها و عدم تطابق شرط‌های درخت تصمیم برای هر نوع تصویر ماموگرافی است.

نیما صهبا و همکارانش در مرجع [۴۸] یک متد ترکیبی از روش‌های تشخیص توده مبتنی بر پیکسل و مبتنی بر ناحیه ارائه داده‌اند. در این روش پس از پیش‌پردازش تصاویر ماموگرافی، ناحیه‌های مشکوک به توده توسط اپراتورهای مورفولوژیکی و متدهای آستاته تطبیق پذیر استخراج می‌شوند. در آخر برای کاهش نرخ مثبت کاذب به وجود آمده یک بردار ویژگی مفید بر مبنای تبدیل ranklet به دست آمده و به طبقه‌بند SVM داده می‌شود. عملکرد این الگوریتم بر روی پایگاه داده MIAS و پایگاه داده بیمارستان امام خمینی تهران ارزیابی شده است. حساسیت و اختصاصیت متد پیشنهادی به ترتیب ۷۴٪ و ۹۱٪ گزارش شده است. متد پیشنهادی در تشخیص توده با شکل‌های مختلف، دارای درجه بالایی از robustness است.

۴-۲ انواع رویکردها در شناسایی توده‌ها

شناسایی توده‌ها مانند بسیاری از روش‌های یادگیری آماری براساس استخراج ویژگی می‌باشد. بنابراین می‌توان تمام روش‌های موجود را در این زمینه مشترک دانست. ساختار بیشتر روش‌های شناسایی را می‌توان مانند شکل (۲-۳) دانست. البته ممکن است در بین روش‌های مختلف تفاوت‌های جزئی در این ساختار مشاهده شود. برای مثال پیش پردازش می‌تواند روی تصویر اصلی انجام شود و یا اینکه تنها روی ناحیه مورد نظر (ROI) انجام گیرد.



شکل (۲-۳): جزئیات ساختار روش‌های شناسایی توده.

یکی دیگر از مواردی که می‌تواند باعث تفاوت روش‌های شناسایی شود، شناسایی نواحی مشکوک (ROI) می‌باشد. در مقالات علمی از هر دو نوع شناسایی خودکار و دستی نواحی مشکوک می‌توان مواردی را یافت. در [۴۹] شناسایی ROIها به صورت خودکار انجام می‌گیرد و در [۴۶] ROIها به صورت دستی جدا می‌شوند. در حال حاضر شناسایی نواحی مشکوک همچنان از بحث‌های باز سامانه های تشخیص پزشکی می‌باشد و کارهای زیادی روی آن انجام می‌شود مانند [۵۰].

مسئله دیگری که باعث تفاوت بین روش‌های مختلف می‌گردد، در وجود یا عدم وجود ناحیه‌بندی می‌باشد. باید توجه داشت که این ناحیه‌بندی با استخراج ROI متفاوت است، از استخراج ROI معمولا یک ناحیه چهار ضلعی که توده در آن قرار دارد بدست می‌آید در حالی که ناحیه بندی مرز دقیق توده را شناسایی می‌کند. روش‌هایی مانند [۴۶] این ناحیه‌بندی را انجام نمی‌دهند و ویژگی‌ها را به طور مستقیم از ROI بدست می‌آورند. در حالی که در بعضی روش‌های دیگر مانند [۴۷] ناحیه‌بندی پس از

پیش پردازش و قبل از مرحله استخراج ویژگی صورت می‌گیرد. در جدول (۱-۲) مقایسه‌ای بین این دو روش صورت گرفته است.

آنچه مسلم است، این است که انجام ناحیه‌بندی، استخراج ویژگی‌هایی - مربوط به شکل و مرز- را میسر می‌سازد که در صورت عدم انجام آن قادر به استخراج آنها نخواهیم بود. می‌توان گفت که ایده اصلی روش‌هایی که این مرحله را انجام نمی‌دهند، جلوگیری از انتشار خطای ناحیه‌بندی در شناسایی توده‌ها می‌باشد. زیرا به همان ترتیب که ناحیه‌بندی خوب استخراج ویژگی‌های مناسب را ممکن می‌سازد، عدم دقت در ناحیه‌بندی نیز باعث منحرف شدن شناسایی می‌گردد. در بعضی از روش‌ها برای جلوگیری از خطای ناحیه‌بندی، این امر به صورت دستی توسط رادیولوژیست‌های خبره صورت می‌گیرد که البته این امکان همواره میسر نیست.

جدول (۱-۲): مقایسه دو رویکرد مبتنی بر ROI و مبتنی بر ناحیه‌بندی

مزایا	معایب
<ul style="list-style-type: none"> • ساده • عدم وجود خطای ناحیه بندی 	<ul style="list-style-type: none"> • دقت کمتر
<ul style="list-style-type: none"> • استخراج ویژگی های مربوط به شکل 	<ul style="list-style-type: none"> • انتشار خطای ناحیه بندی • زمان بیشتر

رویکرد این پایان‌نامه مبتنی بر استخراج ROI است و ROIها نیز به صورت خودکار استخراج می‌شوند. در فصل بعدی روش پیشنهادی برای استخراج خودکار ROIها با نرخ بالایی از حساسیت ارائه خواهد شد. اما با این حال روش‌های مبتنی بر ناحیه‌بندی را نیز نمی‌توان نادید گرفت. در بخش بعدی مروری بر روش‌های ناحیه‌بندی موجود در مقالات متعدد بیان می‌شود.

مساله دیگری که باعث تفاوت بین روش‌های مختلف شناسایی توده می‌گردد، در استخراج و انتخاب ویژگی‌های نواحی مشکوک است. بر طبق آنکه در مقالات متعدد چه ویژگی‌هایی انتخاب شده است، فضای ویژگی به سه زیر فضا تقسیم می‌شود: (۱) ویژگی‌های مبتنی بر شدت روشنایی (۲) ویژگی‌های هندسی (۳) ویژگی‌های بافتی^۱. لازم به ذکر است که تنها روش‌هایی قادر به استخراج ویژگی‌های هندسی هستند که مبتنی بر ناحیه‌بندی باشند. در غیر این صورت چون مرز و شکل نواحی مشکوک به توده مشخص نیست، تنها از دو دسته دیگر از ویژگی‌ها می‌توان استفاده کرد.

به عنوان مثال در مرجع [۵۱] برای تشخیص سریع توده‌ها از ویژگی‌های مربوط به سطوح روشنایی استفاده است. و همچنان علی باقری خلیج در مرجع [۵۲] یک مجموعه ویژگی‌های مرزی^۲ برای دسته‌بندی توده‌ها معرفی کرده است که کمترین وابستگی را به متد ناحیه‌بندی دارد. این نویسنده به منظور دسته‌بندی توده‌ها از ترکیبی از ویژگی‌های مرزی و بافتی استفاده کرده است.

یک دسته از ویژگی‌های مهم بافتی استفاده شده در مقالات متعدد به منظور شناسایی توده در تصاویر ماموگرافی، ویژگی‌های حاصل از ماتریس‌های هم-رخداد است. به عنوان مثال در مراجع [۴۶، ۴۷، ۵۳] از این ویژگی‌ها استفاده شده است.

برای شناسایی ROIها در این پایان‌نامه از ترکیب ویژگی‌های روشنایی و ویژگی‌های بافتی حاصل از ماتریس‌های هم-رخداد و ماتریس‌های طول-رخداد^۴ استفاده شده است.

مساله مهم دیگری که باعث تفاوت بین روش‌های مختلف شناسایی توده می‌گردد، نوع طبقه‌بندی می‌باشد. روش‌های دسته‌بندی که تا کنون به صورت عمده توسط محققان در زمینه شناسایی توده‌ها مورد استفاده قرار گرفته است شامل شبکه‌های عصبی مصنوعی [۴۶، ۵۳، ۵۴]، روش‌های ماشین‌های بردار پشتیبان [۴۸، ۵۲]، بیز [۵۵، ۵۶] و درخت تصمیم دودویی [۴۴، ۴۷] می‌باشد.

¹ Geometric

² Texture

³ margin

⁴ Run-length matrix

در این تحقیق به منظور دسته‌بندی ROIها، الگوریتمی کارآمد ارائه شده است که از ترکیب نتایج طبقه‌بندی‌های مختلف استفاده می‌کند. شرح این الگوریتم به تفصیل در فصل پنجم بیان خواهد شد. استفاده از این الگوریتم در این نوشتار به منظور دسته‌بندی ROIها منجر به افزایش قابل قبولی از دقت خواهد شد.

موارد دیگری را علاوه بر مواردی که تا کنون ذکر شده است، می‌توان برای متفاوت بودن رویکردهای شناسایی توده در نظر گرفت. اما ما به ذکر مهمترین آنها بسنده کرده‌ایم.

۵-۲ ناحیه‌بندی تصاویر ماموگرافی

با وجود اینکه رویکرد ما در شناسایی توده استخراج ROI است اما در اینجا به منظور آشنایی خواننده خلاصه‌ای از روش‌های ناحیه‌بندی را نیز بیان می‌کنیم. تا به اکنون متدهای زیادی برای ناحیه‌بندی تصاویر ماموگرافی مورد استفاده قرار گرفته است در این نوشتار ۶ زیرگروه مهم و پرکاربرد از آنها را معرفی می‌کنیم.

۱-۵-۲ آستانه‌گذاری global

متدهای آستانه‌گذاری global به صورت گسترده در ناحیه‌بندی ماموگرام‌ها مورد استفاده قرار می‌گیرند [۵۷، ۵۸]. متد آستانه‌گذاری global بر مبنای اطلاعات global از قبیل هیستوگرام است. از آنجایی که توده‌ها روشن‌تر از عضلات اطراف خود هستند، آستانه‌گذاری یک متد مفید برای ناحیه‌بندی تصاویر ماموگرافی محسوب می‌شود. هیستوگرام یک منطقه همراه با اختلال، دارای اوج^۱‌های اضافی است، در حالیکه یک منطقه سالم تنها دارای یک اوج است. پس از اینکه مقدار آستانه global به دست آمد، اشیا مورد نظر از زمینه جدا خواهند شد.

^۱ Peak

در این گونه متدها نرخ مثبت کاذب (FP) و منفی کاذب (FN) بسیار بالا است. متدهایی که تنها وابسته به آستانه‌گذاری global هستند، برای تعیین ROI ها مناسب نیستند. به این دلیل که تصاویر ماموگرافی نماهای ۲ بعدی از سینه‌های ۳ بعدی هستند، مناطق بافت‌های هم‌پوشان^۱ شامل ۳ نوع بافت هستند: منطقه چربی^۲، منطقه چربی و غده ای^۳ و منطقه متراکم که ممکن است روشن‌تر از توده‌ها باشد. اغلب در بیشتر متدها، خروجی آستانه‌گذاری global به عنوان ورودی مرحله بعد مورد استفاده قرار می‌گیرد.

۲-۵-۲ آستانه‌گذاری local

آستانه‌گذاری local می‌تواند نتایج حاصل از آستانه‌گذاری global را بهتر کند و یا خود به تنهایی برای تعیین مناطق مشکوک استفاده شود. به این دلیل که مقدار آستانه در این گونه متدها به صورت محلی برای هر پیکسل بر مبنای مقادیر intensity پیکسل‌های اطراف آن تعیین می‌شود، برای تشخیص توده بهتر از آستانه‌گذاری global عمل می‌کنند. در متدهای آستانه‌گذاری local دو متغیر باید در نظر گرفته شود: سایز پنجره و مقدار آستانه. آستانه‌گذاری local ممکن است به عنوان پیش‌پردازش برای الگوریتم‌های دیگر، مانند میدان تصادفی مارکوف^۴ استفاده شود [۵۹].

۳-۵-۲ رشد ناحیه

یکی از متدهای مشهور برای ناحیه‌بندی توده‌ها در تصاویر ماموگرافی رشد ناحیه است [۶۰ و ۶۱]. ایده اصلی این الگوریتم در ابتدا یافتن یک مجموعه پیکسل‌های دانه^۵ و سپس رشد دادن و به هم پیوستن

¹ Overlapping

² Fatty

³ Glandular

⁴ Markov Random Field

⁵ Seed

تکراری پیکسل‌هایی است که دارای خصوصیات مشابه هستند. این روش گرچه ساده به نظر می‌رسد دارای سه مشکل عمده است:

۱. انتخاب دانه‌های اولیه که نمایندگان خوبی برای نواحی مورد علاقه باشد
۲. انتخاب معیار مناسب برای الحاق نقاط به نواحی مختلف در حین فرایند رشد
۳. تعیین قاعده توقف

وقتی هیچ پیکسل دیگری معیارهای پیوستن به یک ناحیه را برآورده نکند باید رشد آن ناحیه متوقف شود. رشد ناحیه ممکن است به صورت global و یا local اعمال شود.

۴-۵-۲ تطبیق الگو

تطبیق الگو^۱ یکی از شایع‌ترین متدها برای ناحیه‌بندی تصاویر پزشکی و از جمله تصاویر ماموگرافی است. این متد از دانش قبلی تصاویر ماموگرافی استفاده می‌کند و توسط الگوهای اولیه^۲، توده‌های احتمالی را از میان پس‌زمینه جدا می‌کند. الگوهای اولیه از توده‌های احتمالی، بر مبنای خصوصیات یا ویژگی‌های فیزیکی توده‌های هدف^۳ ایجاد می‌شود [۶۳ و ۶۱]. زمانیکه دانش قبلی در مورد سائز توده‌ها در دسترس نیست، دامنه‌ای از سائز برای الگوها در نظر گرفته می‌شود [۶۴]. عملکرد این متد وابسته به تعیین نمونه‌های اولیه است. اگر نمونه‌های اولیه به درستی انتخاب شوند، نتایج خوبی از این متد حاصل می‌شود.

۵-۵-۲ متدهای فازی

به دلیل اینکه کنتراست تصاویر ماموگرافی بسیار پایین است و مرزهای بین بافت نرمال و توده غیرشفاف است، ممکن است متدهای ناحیه‌بندی سنتی و قدیمی به خوبی کار نکنند. اعمال متدهای

¹ Template Matching

² Prototypes

³ Target

فازی در ناحیه بندی تصاویر ماموگرافی در مراجع [۶۵-۶۸] مطالعه شده است. منطق فازی در مرجع [۶۵] به منظور جداسازی توده‌های مشکوک معرفی شده است. این الگوریتم در ابتدا یک مقدار عضویت فازی برای هر پیکسل اختصاص می‌دهد و سپس در هر گام مقدار خطا محاسبه می‌شود و مقادیر عضویت فازی به‌روز می‌شود [۶۶]. همچنان تاثیر پیکسل‌های همسایه در قانون به‌روزرسانی در نظر گرفته می‌شود. این الگوریتم هنگامی متوقف می‌شود که خطای صفر بدست آید. پس از آن هر پیکسل به یکی از ناحیه‌های تیره (زمینه) یا روشن (توده) نسبت داده می‌شود. ثابت شده است که تئوری مجموعه فازی به همراه الگوریتم‌های مبتنی بر بافت برای دسته‌بندی توده‌ها بسیار مفید هستند [۶۵].

۶-۵-۲ تفاضل دو نمایه تصویر

تفاضل دو نمایه تصویر^۱ برای تعیین مناطق مشکوک استفاده می‌شود [۶۷، ۷۰]. این متد همچنان متد عدم تقارن^۲ نیز نامیده می‌شود، بر مبنای تقارن بین سینه‌های چپ و راست عمل می‌کند. پیاده‌سازی این متد آسان است و اختلاف بین تصاویر ماموگرافی سینه چپ و راست به عنوان مناطق مشکوک در نظر گرفته می‌شود. اما نکته‌ای که حائز اهمیت است این است که، منطبق کردن دقیق سینه‌های راست و چپ بسیار مشکل است.

¹ Bilateral

² Asymmetry

فصل سوم:

پیش پردازش بروی تصاویر ماکروگرافی و استخراج ROI

۱-۳ مقدمه

پیش از اقدام به پردازش تصاویر ماموگرافی به هدف تشخیص توده، لازم است یک سری پیش-پردازش‌هایی بر روی تصاویر مورد استفاده صورت گیرد. این عملیات در مقالات متعدد متفاوت است. اغلب عملیات انجام شده شامل حذف مناطق ناخواسته از تصویر، بهبود کیفیت تصویر و حذف نویز می‌باشد. به طور کلی جزییات این عملیات بسته به تصویر مورد استفاده و مشخصات آن می‌تواند متفاوت باشد.

مناطق ناخواسته شامل مناطقی می‌شوند که جز بافت اصلی سینه محسوب نمی‌شوند و حذف آنها تاثیری بر نتیجه عملکرد سیستم CAD ندارند. به این ترتیب تنها بافت اصلی سینه از هر تصویر جداسازی می‌شود. حذف این مناطق موجب کاهش نرخ مثبت کاذب در سیستم‌های CAD خواهد شد. برخی از مقالات در همان ابتدا تصاویر را به صورت دستی برش می‌دهند تا کلیه مناطق ناخواسته را از تصویر حذف کنند [۵۳]. اما مقالاتی هم هستند که توسط روش‌هایی به صورت خودکار این مناطق را از تصویر حذف می‌کنند [۷۱].

در این فصل پس از معرفی روش‌های بهبود تصاویر ماموگرافی، پیش‌پردازش‌های صورت گرفته در این نوشتار و متد پیشنهادی مربوط به استخراج اتوماتیک نواحی مشکوک به توده بیان می‌شود.

۲-۳ بهبود کیفیت تصویر

تصاویر ماموگرافی معمولاً دارای وضوح و کنتراست^۱ ضعیفی هستند. به همین دلیل معمولاً افزایش و بهبود کیفیت تصویر در اولین قدم از پیاده‌سازی سیستم CAD ضروری به نظر می‌رسد. تکنیک‌های بهبود تصویر به هدف تقویت کیفیت و خوانایی تصاویر ماموگرافی و یا تشخیص مناطق غیر عادی به

^۱ Contrast

کار برده می‌شوند. تصاویر بهبود یافته برای انجام عملیات خاص و همچنین اعمال توابع هدف بهتر از تصاویر اولیه است. روش‌های بهبود تصاویر ماموگرافی به گروه‌های زیر دسته‌بندی می‌شود:

(۱) متدهای اصلاح هیستوگرام global

(۲) متدهای پردازشی local

(۳) متدهای پردازش چند-مقیاسه^۱

۱-۲-۳ متدهای اصلاح هیستوگرام global

ایده اصلی متدهای اصلاح هیستوگرام global، اختصاص دادن مقدار جدید روشنایی^۲ برای هر پیکسل به منظور ایجاد یک توزیع یکنواخت جدید از روشنایی است. شیوه‌ای از اصلاح هیستوگرام global که به طور معمول مورد استفاده قرار می‌گیرد، تعدیل هیستوگرام^۳ (HE) است [۷۲].

فرض کنید $H(i)$ هیستوگرام تصویری با سایز $M \times N$ است، و $[G_{min}, G_{max}]$ دامنه شدت روشنایی تصویر است. متد تعدیل‌سازی هیستوگرام، شدت روشنایی تصویر اصلی (I_{org}) را طبق رابطه (۱-۳)، به شدت روشنایی تصویر جدید (I_{new}) تبدیل می‌کند.

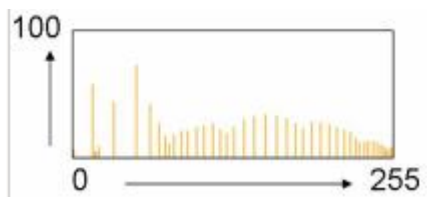
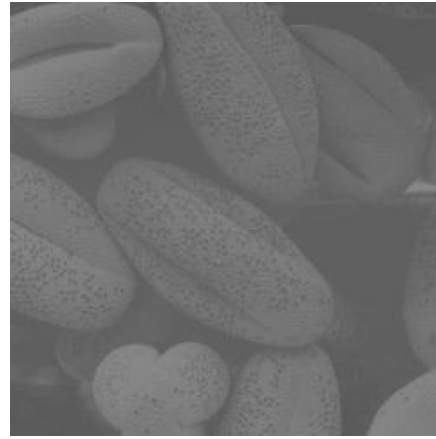
$$I_{new} = G_{min} + (G_{max} - G_{min}) \times \frac{\sum_{i=0}^{I_{org}} H(i)}{(M \times N)} \quad (1-3)$$

عملکرد متد تعدیل‌سازی هیستوگرام برای بهبود تصویری با کنتراست پایین بسیار ساده و موثر است. زمانی که می‌گوییم کنتراست تصویری پایین است این بدان معنا است که اختلاف بین کمترین و بیشترین شدت روشنایی تصویر کم است. تعدیل‌سازی هیستوگرام موجب می‌شود کنتراست تصویر ورودی تا حد ممکن افزایش یابد. به عنوان مثال شکل (۱-۳) تصویری را قبل و بعد از اعمال تعدیل‌سازی هیستوگرام نشان می‌دهد.

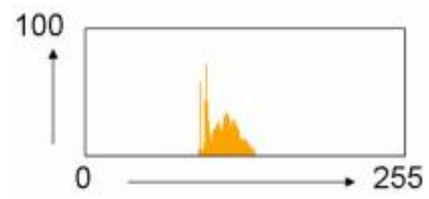
¹ Multi-scale

² Intensity

³ Histogram Equalization (HE)



ب



الف

شکل (۱-۳): نمونه ای از بهبود تصویر توسط متد تعدیل سازی هیستوگرام. الف) تصویر ورودی و هیستوگرام آن ب)

تصویر خروجی پس از تعدیل هیستوگرام

یکی دیگر از متدهای اصلاح هیستوگرام global، هیستوگرام انبساطی^۱ است [۷۳]. و از تابع انتقال

خطی رابطه (۲-۳) استفاده می کند :

$$I_{new} = G_{min} + (I_{org} - I_{min}) \times (G_{max} - G_{min}) / (I_{max} - I_{min}) \quad (2-3)$$

که در آن $[I_{min}, I_{max}]$ دامنه شدت روشنایی تصویر اصلی و $[G_{min}, G_{max}]$ دامنه شدت روشنایی

تصویر جدید است.

¹ Stretching

۲-۲-۳ متدهای پردازشی local

متدهای پردازش local نیز برای افزایش کنتراست تصویر، مورد مطالعه قرار می‌گیرند. روش‌های بسیاری برای افزایش کنتراست به وسیله تغییر شدت روشنایی پیکسل وجود دارد. یک راه براساس نگاشت غیرخطی می‌باشد (روش‌های Non-Sigmoid, Bi-linear, local histogram technique, continuous و غیره) [۷۲، ۷۳-۷۵]. پیاده‌سازی می‌تواند مبتنی بر ویژگی باشد، ویژگی‌های محلی ممکن است با تشخیص لبه و یا با استفاده از اطلاعات آماری محلی همچون میانگین محلی، انحراف استاندارد و غیره بدست آید. نگاشت‌های غیرخطی ممکن است: (۱) بر مبنای گرادیان و توابع غیر خطی [۷۲-۷۴، ۷۶]. (۲) بر مبنای تعدیل‌سازی هیستوگرام تطبیق‌پذیر^۱ باشد [۷۷، ۷۸].

۳-۲-۳ متدهای پردازش چند-مقیاسه

برخی متدهای موجود برای بهبود تصاویر، مبتنی بر تبدیل ویولت^۲ است [۷۹، ۸۰]. مراحل کلی این متدها به این شرح است: (۱) تصویر ماموگرافی با استفاده از ویولت‌ها تبدیل می‌شوند (۲) ضرایب برای بهبود ویژگی‌های توده تغییر داده می‌شوند. (۳) تصویر بهبود یافته با استفاده از تبدیل معکوس ویولت بدست می‌آید. این متدها از اطلاعات جهت‌دار در هر scale استفاده می‌کنند. تفاوت پیاده‌سازی این نوع متدها در توابع پایه^۳ (ویولت‌های مادر) و روش تعیین ضرایب^۴ است. این متدها برای انتخاب ویژگی‌های محلی به منظور بهبود کیفیت منعطف هستند و همچنان قادر به حذف نویز نیز می‌باشند.

^۱ Adaptive Histogram Equalization

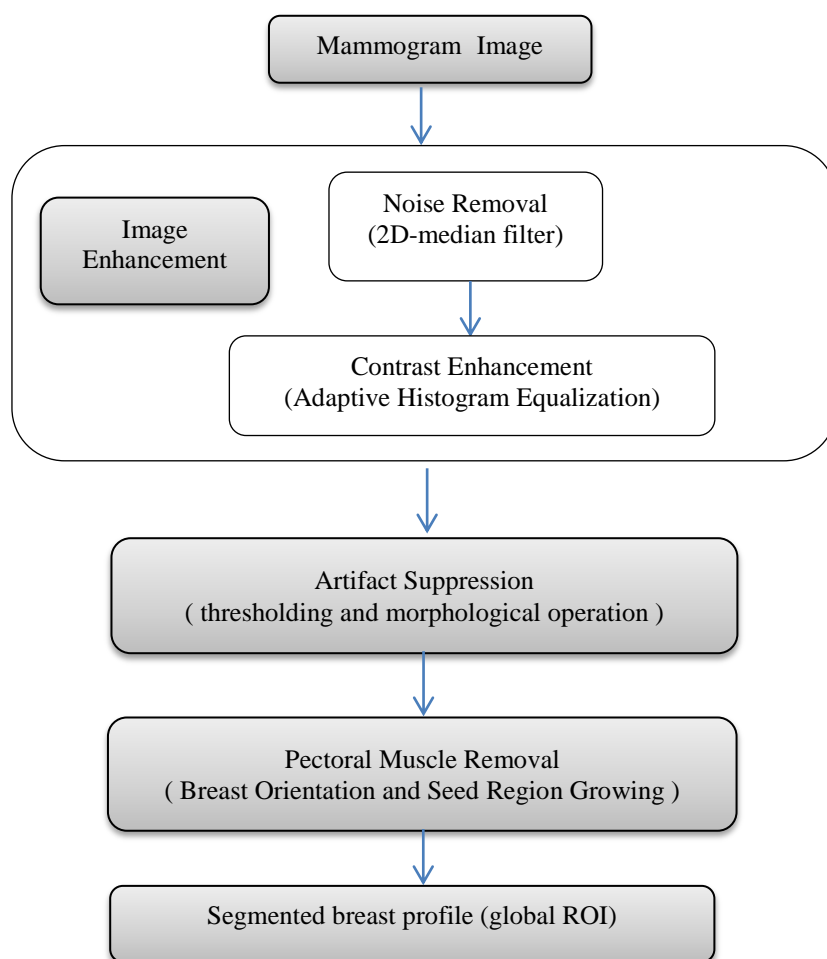
^۲ Wavelet

^۳ Basis Function

^۴ Coefficient

۳-۳ پیش پردازش استفاده شده بر روی تصاویر ماموگرافی

در این بخش قصد داریم عملیات پیش پردازش انجام شده بر روی تصاویر استفاده شده در این پایان-نامه را مورد بررسی قرار دهیم. این عملیات به ترتیب شامل بهبود تصویر، حذف مصنوعات^۱ و حذف ماهیچه پکتورال^۲ است. شکل (۲-۳) بلوک دیاگرام این عملیات را نشان می دهد.



شکل (۲-۳): بلوک دیاگرام پیش پردازش های انجام شده

¹ Artifact

² Pectoral muscle

۳-۳-۱ بهبود تصاویر ماموگرافی

همان طور که در اوایل فصل نیز اشاره شد، تصاویر ماموگرافی معمولاً دارای وضوح و کنتراست^۱ پایینی هستند و در نتیجه تفسیر آنها بسیار مشکل است. بنابراین معمولاً افزایش و بهبود کیفیت تصاویر ماموگرافی جهت تشخیص توده‌های سرطانی به صورت چشمی و یا استخراج بهتر نواحی مشکوک ضروری به نظر می‌رسد. عملیات بهبود کیفیت تصاویر در این نوشتار شامل دو مرحله حذف نویز و افزایش کنتراست است.

الف) حذف نویز

تصاویر ماموگرافی مورد استفاده شامل نویزهای زیادی از قبیل نویزهای حاصل از دیجیتالی کردن تصویر و نویزهای حاصل از فرایند تصویربرداری می‌باشد. در این پایان‌نامه برای حذف نویز از فیلتر میانه^۲ استفاده شده است. در فیلتر میانه هر پیکسل تصویر خروجی با مقدار میانه در همسایگی 3×3 تصویر ورودی جایگزین می‌شود.

فیلتر میانه یک ابزار قدرتمند در حذف نویز سیگنال‌های دوبعدی است. لبه‌ها فاکتور بسیار مهمی در ناحیه‌بندی محسوب می‌شود. به این علت که فیلتر میانه این قابلیت را دارد که بدون از بین بردن لبه‌ها نویز را حذف کند، نتایج نسبتاً بهتری از دیگر فرایندهای حذف نویز ایجاد می‌کند.

ب) افزایش کنتراست

برای بهبود کنتراست تصویر در این پایان‌نامه از متد تعدیل سازی هیستوگرام تطبیقی^۳ استفاده شده است. تفاوت آن با تعدیل سازی هیستوگرام معمولی در محاسبه تطبیقی هیستوگرام‌های متفاوت است که هر کدام متناظر با بخش مجزایی از تصویر است، که از آنها برای توزیع مقادیر جدید روشنایی تصویر استفاده می‌کند. در طول این فرایند کنتراست هر پیکسل وابسته به همسایگی محلی آن پیکسل به صورت تطابق یافته افزایش می‌یابد. این روش برای افزایش کنتراست محلی تصویر و

¹ Contrast

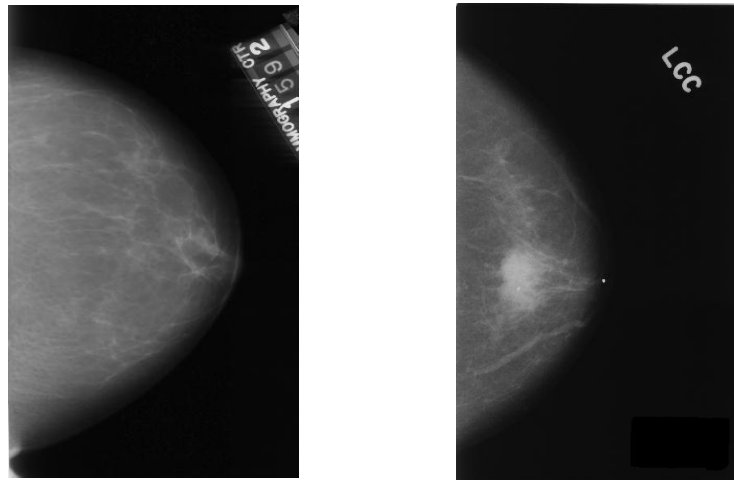
² Median Filter

³ Adaptive Histogram Equalization (AHE)

برجسته کردن جزئیات موثر است و همچنان به کاهش نویزهای تولید شده در مناطق همگن نیز کمک می‌کند.

۲-۳-۳ حذف مصنوعات

اغلب تصاویر ماموگرافی مورد استفاده دارای مصنوعاتی^۱ از قبیل برجسب‌ها^۲ و حاشیه‌هایی^۳ در بالای تصویر می‌باشند (شکل ۳-۳). به دلیل اینکه این مصنوعات در طول فرایند تشخیص توده اختلال ایجاد خواهند کرد، لازم است در مرحله پیش‌پردازش حذف شوند.



شکل (۳-۳) : نمونه ای از تصاویر ماموگرافی شامل مصنوعات: شکل سمت راست) تصویر ماموگرافی دارای برجسب و شکل سمت چپ تصویر ماموگرافی دارای حاشیه

¹ Artifact

² Lable

³ Edge

به منظور حذف مصنوعات از تصاویر مورد استفاده، از متد آستانه‌گذاری^۱ به همراه اپراتورهای مورفولوژی^۲ استفاده می‌کنیم. مراحل اجرا به همراه توضیح هر مرحله در ادامه آورده شده است:

(۱) تبدیل تصویر سطح خاکستری به تصویر باینری. با توجه به اینکه اپراتورهای مورفولوژی تنها بر روی تصاویر باینری اعمال می‌شوند، باید در مرحله اول تصاویر را توسط یک مقدار آستانه global باینری کنیم. مقدار آستانه تجربی مناسب با توجه به مقدارهای متفاوت که بر روی همه تصاویر امتحان شده است، ($T=18$) یا مقدار نرمال شده آن ($T=0.0706$) می‌باشد.

(۲) برچسب گذاری تمامی object های موجود در تصویر باینری. این object ها شامل مصنوعات موجود در تصویر و ناحیه اصلی سینه می‌باشد.

(۳) محاسبه مساحت تمامی object ها (ناحیه‌ها) در تصویر باینری. مساحت هر object متناظر است با تعداد واقعی همه پیکسل‌های موجود در آن ناحیه.

(۴) حذف تمامی object ها به جز بزرگترین object. بزرگترین object در تصویر باینری مربوط به ناحیه اصلی سینه است.

(۵) اعمال اپراتورهای مورفولوژیکی به تصویر باینری حاصل از مرحله قبل. هدف این اپراتورها کاهش اعوجاج^۳ و حذف پیکسل‌های تنها^۴ (۱های تنها که به وسیله صفر محاط شده‌اند) در تصویر است.

(۶) ضرب تصویر باینری حاصل از مرحله قبل در تصویر سطح خاکستری اصلی.

¹ Thresholding

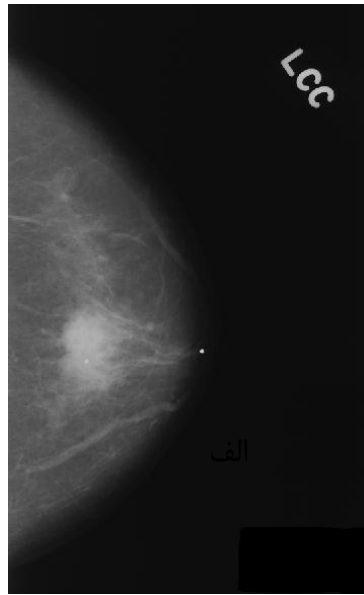
² Morphological

³ Distortion

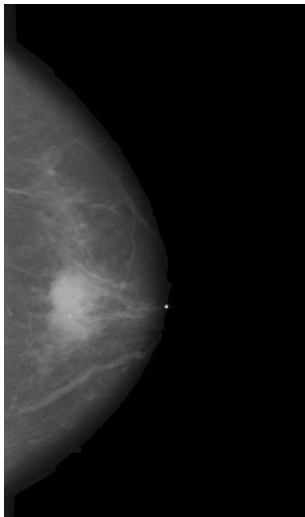
⁴ Isolated



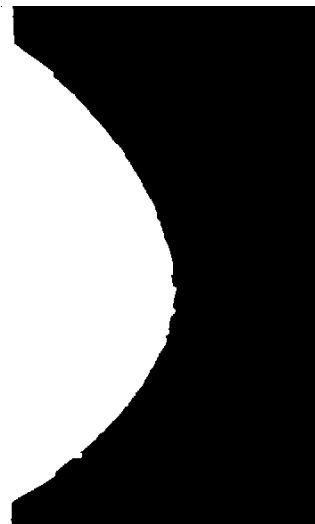
ب



الف



د



ج

شکل (۳-۴): حذف مصنوعات ناخواسته از تصاویر ماموگرافی. الف) تصویر اصلی ب) تصویر باینری با $T=18$ ج) حذف کلیه object ها به غیر از بزرگترین object در تصویر باینری د) تصویر ماموگرافی که در آن برچسب موجود در تصویر الف حذف شده است.

۳-۳-۳ حذف ماهیچه پکتورال

ماهیچه پکتورال در اغلب نماهای MLO از ماموگرافی به صورت یک منطقه روشن مثلثی متراکم در گوشه تصویر مشاهده می‌شود، که به سمت قفسه سینه و تا پایین منطقه اصلی سینه کشیده می‌شود.

شکل (۳-۵). روشنایی و خصوصیات بافتی ناحیه مربوط به ماهیچه پکتورال بسیار نزدیک به روشنایی و خصوصیات بافتی نواحی مشکوک در ماموگرام است که تومورها و عوارض سرطانی اغلب در این نواحی رخ می‌دهند. بنابراین اگر پردازش با وجود ماهیچه پکتورال انجام گیرد باعث افزایش درصد تشخیص مثبت کاذب می‌شود.

از سوی دیگر، تصاویر ماموگرام دارای ابعاد بزرگی هستند که پردازش کل تصویر را برای یافتن تومور بسیار زمان بر می‌کند لذا به منظور کاهش نواحی مورد پردازش تصویر، جداسازی ماهیچه پکتورال از بافت سینه ضروری به نظر می‌رسد.



شکل (۳-۵): ماهیچه پکتورال در یک تصویر ماموگرافی از نمای MLO

ناحیه بندی اتوماتیک ماهیچه پکتورال مبتنی بر رشد ناحیه

به منظور یافتن ناحیه مربوط به ماهیچه پکتورال از تصاویر شامل نماهای MLO، از متد ناحیه بندی اتوماتیک مبتنی بر رشد ناحیه استفاده می‌کنیم.

ایده اصلی متد رشد ناحیه براساس روشنایی این است که در یک ناحیه، روشنایی‌ها مقادیر نزدیک به هم دارند. این روش از یک پیکسل دانه رشد خود را شروع می‌کند. سه مرحله اصلی این الگوریتم به صورت زیر است:

۱. انتخاب دانه

۲. بررسی پیکسل‌های همسایه و اضافه کردن آنها در صورت شباهت

۳. برای هر پیکسل جدید که اضافه می‌شود، مرحله ۲ را تکرار کن، تا زمانی که پیکسلی باقی نمانده باشد.

انتخاب دانه به صورت خودکار انجام می‌شود، فرض بر این است که در یک تصویر فقط یک ماهیچه پکتورال وجود دارد، بنابراین تنها یک دانه لازم است. این دانه باید به عنوان نقطه‌ای در ماهیچه پکتورال در نظر گرفته شود.

برای انتخاب دانه باید موقعیت^۱ سینه (چپ یا راست) در هر تصویر ماموگرافی تعیین شود. به عبارت دیگر باید مشخص شود در هر تصویر سینه در سمت راست و یا چپ تصویر وجود دارد. بنابراین مجموع ۱۰ ستون اول و ۱۰ ستون آخر از تصویر باینری سینه را محاسبه کنیم و با ایجاد یک شرط ساده موقعیت سینه در هر تصویر را تعیین کنیم. اگر جمع ۱۰ ستون اول بزرگتر شد ($sum_{first} > sum_{last}$)، سینه در سمت چپ تصویر قرار دارد و بالعکس اگر جمع ۱۰ ستون آخر بزرگتر شد ($sum_{first} < sum_{last}$)، سینه در سمت راست تصویر قرار دارد.

حال پس از تعیین موقعیت سینه در تصویر به راحتی می‌توان دانه را در مکان مناسب خود یعنی درون ماهیچه پکتورال در تصویر سطح خاکستری جای داد. فرض کنید (m, n) بیانگر اندازه تصویر باشد. حال اگر سینه در موقعیت چپ تصویر قرار دارد، دانه در سطر و ستون ۵ از تصویر و همچنان اگر دانه در موقعیت راست تصویر قرار دارد، دانه در سطر ۵ و ستون $(n-5)$ قرار می‌گیرد. بدین صورت موقعیت سینه و انتخاب دانه به صورت کاملاً اتوماتیک و خودکار تعیین می‌شود.

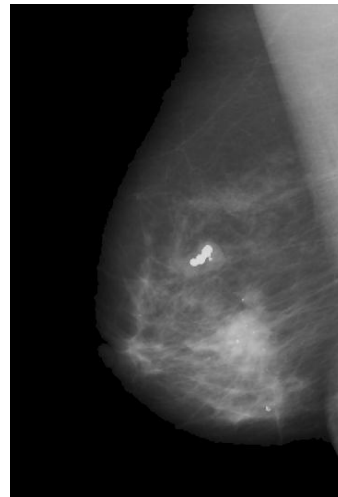
مسئله مهم دیگر معیار شباهت است که مشخص می‌کند آیا پیکسل جدید به ناحیه اضافه شود یا خیر. این معیار به صورت تجربی بر روی ۲۰ نمونه تصویر بدست می‌آید. دانه به تمام پیکسل‌های

¹ Orientation

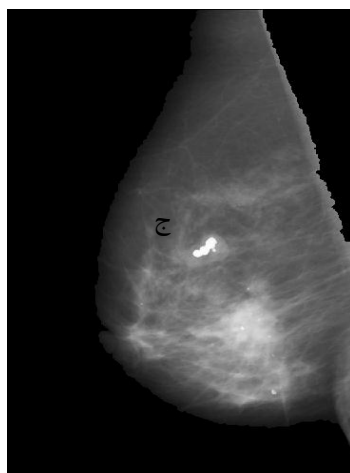
همسایه باقی مانده حرکت داده می‌شود. تفاوت بین مقدار روشنایی آن پیکسل و میانگین پیکسل‌های دانه محاسبه می‌شود. اگر این مقدار کوچکتر از معیار شباهت مورد نظر بود، آن پیکسل به ناحیه پکتورال اختصاص داده می‌شود. هنگامیکه این مقدار بزرگتر از معیار بود، فرایند رشد متوقف می‌شود. حال برای بدست آوردن تصویر نهایی که در آن ماهیچه پکتورال حذف شده است کافیست ابتدا نتیجه متد اتوماتیک رشد ناحیه از تصویر باینری تفریق شود و نتیجه آن در تصویر خاکستری اصلی ضرب شود شکل (۳-۶).



ب



الف



ج

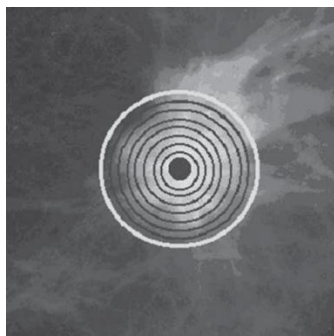
شکل (۳-۶): حذف ماهیچه پکتورال. الف) تصویر ماموگرافی شامل ماهیچه پکتورال ب) ماهیچه پکتورال ناحیه بندی شده توسط متد رشد ناحیه ج) تصویر ماموگرافی با حذف ماهیچه پکتورال

۳-۴ استخراج نواحی مشکوک (ROIs)

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 زمان بر 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. به یافتن نواحی مشکوک به سرطان، تعیین ناحیه مورد توجه (ROI) نیز گفته می‌شود.

مرحله دوم از طرح‌های CAD، جداکردن ناحیه‌های مشکوک از زمینه است، که ممکن است شامل توده باشد. ناحیه مشکوک، منطقه‌ای روشن‌تر از اطرافش، دارای تراکم‌های یکنواخت با سایز متغیر و مرزهای فازی است [۵۰]. این مرحله یک مرحله بسیار ضروری و مهم است که تعیین کننده میزان حساسیت کل سیستم است. بنابراین کاهش نواحی مورد نظر اگر بدون از دست دادن اطلاعات با ارزش انجام گیرد به دقت، سرعت و حجم محاسبات کمک شایانی خواهد کرد. روش‌های استخراج ROI نیازی به مشقت در یافتن مکان توده‌ها ندارد بلکه هدف یافتن تمام مناطق مشکوک به توده حتی با تعدادی نرخ مثبت کاذب (FP) است. FP ها در مراحل بعد از بین خواهند رفت.

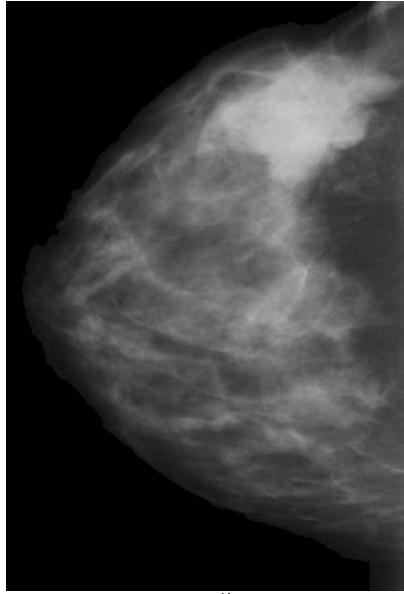
یکی از روش‌هایی که در بسیاری از الگوریتم‌های تحقیقاتی پیشنهادی توسط گروه‌های مختلف از آن برای تشخیص نواحی مشکوک استفاده شده است، ROI Hunter می‌باشد [۸۱]. در این روش برای تعیین نواحی مورد توجه مراحل چهار گانه‌ای به صورت تکراری انجام می‌گیرد تا تمام سطح ماموگرام پیمایش شده و تمام مناطق مشکوک تعیین شوند. 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 ۳-۷). ۳- حذف دایره‌های نهایی از تصویر. ۴- تکرار الگوریتم تا یافتن تمامی ماکزیمم‌ها.



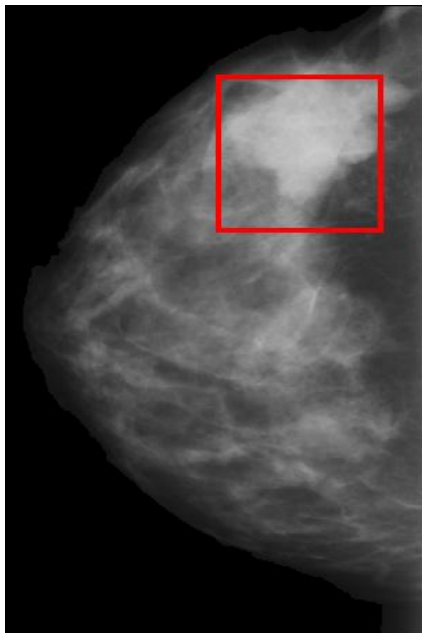
شکل (۷-۳) : دسته دایره هایی با افزایش ۵ پیکسلی شعاع به دور یک ماکزیمم بدست آمده [۸۱]

۱-۴-۳ روش پیشنهادی برای استخراج خودکار ROIها

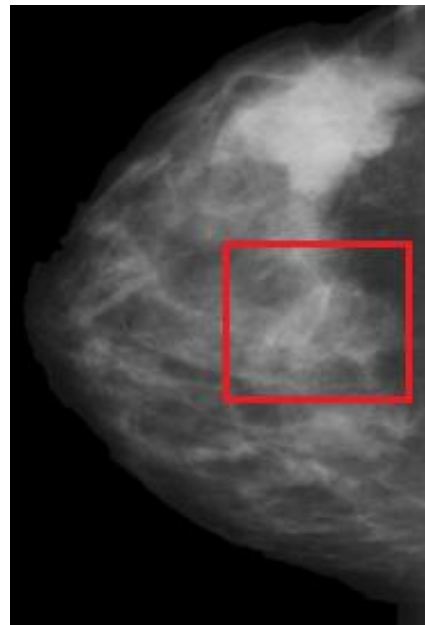
در این تحقیق استخراج ROI مبتنی بر تجزیه و تحلیل هیستوگرام صورت می‌پذیرد. با توجه به اینکه توده‌ها روشن‌تر از محیط اطراف خود هستند، نقطه ماکزیمم محلی تیز غیر معمول در هیستوگرام ایجاد می‌کنند. تجزیه و تحلیل نقاط ماکزیمم در هیستوگرام و استخراج مناسب نواحی مربوط به نقاط ماکزیمم منجر به استخراج ROIها در تصاویر ماموگرافی خواهد شد. لازم به ذکر است که این روش دارای حساسیت بالایی است و در نتیجه نرخ مثبت کاذب (FP) بالایی دارد. شکل (۸-۳).



الف



ج



ب

شکل (۸-۳): نمونه‌ای از استخراج ROI ها. الف) تصویر اصلی. ب و ج) ROI های استخراج شده از تصویر الف

فصل چهارم :

استخراج ویرگی

۱-۴ مقدمه

در فصل قبل پیش‌پردازش و استخراج ROI در تصاویر ماموگرافی مورد بررسی قرار گرفت. اساس کار دسته‌بندی با روش‌های آماری استخراج ویژگی از داده‌ها می‌باشد. در این فصل انواع ویژگی‌هایی که در تصاویر ماموگرافی بیشتر مورد استفاده قرار می‌گیرند، معرفی می‌شود. جزئیات ساختار روش‌های شناسایی توده در شکل (۲-۳) نشان داده شده است و در آن جایگاه استخراج ویژگی بعد از استخراج ROI مشخص است. استخراج و انتخاب ویژگی یک مرحله کلیدی در تشخیص توده است.

۲-۴ انواع ویژگی

ویژگی‌ها از خصوصیات (ROI) مثل اندازه، شکل، تراکم، روشنایی، بافت و غیره استخراج می‌شود. فضای ویژگی به علت تفاوت گسترده بافت‌های نرمال و گوناگونی ناهنجاری‌ها، بسیار بزرگ و پیچیده است. اما تنها برخی از آنها مناسب و قابل توجه هستند. استفاده از ویژگی‌های فراوان ممکن است عملکرد الگوریتم را پایین آورد و همچنین پیچیدگی طبقه‌بندها را افزایش دهد. برخی از ویژگی‌های زائد برای بهبود عملکرد طبقه‌بند باید حذف شوند.

مرجع [۸۲] تحقیقی از تکنیک‌های تجزیه و تحلیل ویژگی برای استخراج ویژگی از تصاویر ماموگرافی ارائه کرده است.

تا به اکنون ویژگی‌های مفید بسیاری برای تصاویر ماموگرافی پیشنهاد شده است بر طبق آنکه چه نوع ویژگی‌هایی انتخاب شده است، فضای ویژگی به سه زیر فضا تقسیم می‌شود:

- (۱) ویژگی‌های روشنایی^۱
- (۲) ویژگی‌های شکلی^۲
- (۳) ویژگی‌های بافتی^۱

^۱ Intensity
^۲ Shape

از نظر رادیولوژیست‌ها هر کدام از این ویژگی‌ها معنی خاصی دارند و در تصمیم‌گیری برای شناسایی توده موثر هستند. ویژگی‌های استخراج شده از خصوصیات سطح خاکستری، شکل، بافت و عضلات اطراف آن می‌تواند به عنوان یک توصیف ریاضی بیان شود و برای دسته‌بندی مناطق مشکوک مفید واقع شود. اما پیش‌بینی اینکه چه ویژگی و یا چه ترکیبی از ویژگی‌ها منجر به نتایج بهتری در دسته‌بندی خواهد شد بسیار مشکل است. معمولا ترکیبات متفاوتی از ویژگی‌ها در عملکردهای متفاوت نتیجه می‌شود. بنابراین اغلب با انتخاب زیرمجموعه بهینه از ویژگی‌ها از میان تعداد زیادی از ویژگی‌های ممکن روبه‌رو هستیم. راهنمای کلی برای انتخاب ویژگی‌های قابل توجه، شامل ۴ نکته اساسی می‌شود [۸۳]: افتراق^۲، اطمینان^۳، استقلال^۴ و بهینگی^۵.

فرایند استخراج و انتخاب ویژگی برای تشخیص توده می‌تواند بر مبنای تجزیه و تحلیل مولفه اصلی^۶ (PCA) [۸۴]، تجزیه و تحلیل جداسازی خطی^۷ (LDA) [۸۶ و ۸۵] و الگوریتم ژنتیک^۸ (GA) [۷۸] باشد.

۳-۴ ویژگی‌های روشنایی

این دسته از ویژگی‌ها در میان سه گروه ویژگی دیگر استخراج شده از تصاویر ماموگرافی ساده‌تر هستند. لیستی از پرکاربردترین ویژگی‌های روشنایی که در مقالات متعدد شناسایی توده از آنها استفاده شده است، در جدول (۴-۱) مشاهده می‌شود.

¹ Textural

² Discrimination

³ Reliability

⁴ Independence

⁵ Optimality

⁶ Principle Component Analysis (PCA)

⁷ Linear Discriminate Analysis (LDA)

⁸ Genetic Algorithm (GA)

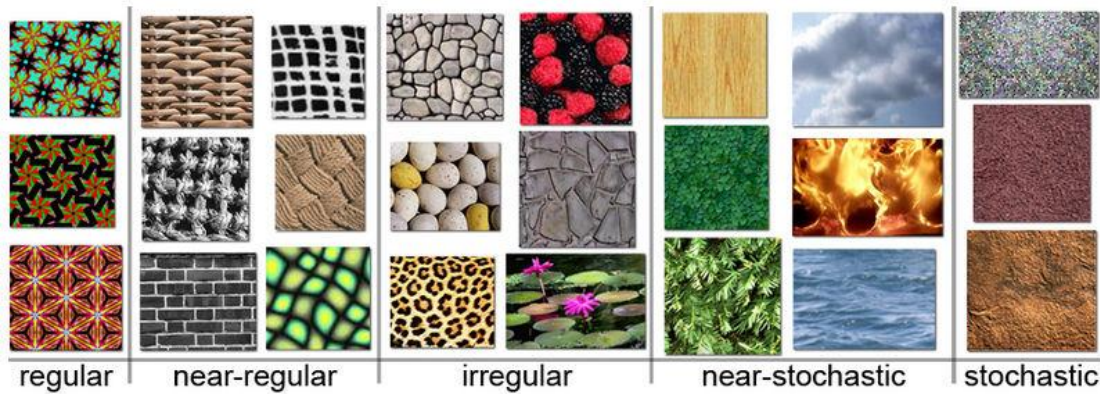
جدول (۴-۱): ویژگی‌های مبتنی بر روشنایی [۸۸]

Sub-spaces	Feature descriptions
Intensity Features	<p>F1: Contrast measure of ROI</p> <p>F2: Average grey level of ROIs (Mean)</p> <p>F3: Standard derivation of ROIs (variance)</p> <p>F4: Skewness of ROIs</p> <p>F5: Kurtosis of ROIs</p>

ویژگی F1 بیانگر کنتراست ROI ، F2 مربوط به میانگین روشنایی ROI ، F3 مربوط به انحراف معیار روشنایی ROI ، F4 مربوط به کجی هیستوگرام مربوط به ROI و F5 مربوط به کشیدگی هیستوگرام مربوط به ROI است.

۴-۴ ویژگی‌های بافتی

بررسی بافت یکی از مباحث حائز اهمیت در حوزه تحلیل تصاویر پزشکی است. بافت یک تصویر، شیوه توزیع سطوح روشنایی را در تصویر مشخص می‌کند. از آنجا که تعریف دقیق و کاملی برای بافت وجود ندارد، معمولاً بافت یک تصویر را با ویژگی‌هایی مثل نرم یا زمخت، منظم یا تصادفی و مانند آن مشخص می‌کنیم. چنین مشخصاتی را می‌توان با کمک برخی ویژگی‌ها بیان کرد. در شکل (۴-۱) چند نوع بافت مختلف مشاهده می‌شود. برای استخراج ویژگی‌های بافتی انتخاب‌های بسیار زیادی در متون مختلف پیشنهاد شده است [۸۹]. تعدادی از آنها که در این نوشتار مورد استفاده قرار گرفته است، در جدول (۴-۲) لیست شده است.



شکل (۴-۱): تصاویری با چندین نوع بافت مختلف

جدول (۴-۲): ویژگی‌های بافتی

Sub-spaces	Feature descriptions
Texture Features	<u>GLCM</u>
	F6: correlation
	F7: contrast
	F8: energy
	F9: homogeneity
	<u>GLRLM</u>
	F10: Short-run emphasis
	F11: Long-run emphasis
	F12: Gray-level non-uniformity
	F13: Run length non-uniformity
	F14: Run percentage

۱-۴-۴ ویژگی‌های حاصل از ماتریس هم‌رخداد

اطلاعات حاصل از ویژگی‌های آماری مرتبه اول (هیستوگرام ROI) مربوط به توزیع سطوح روشنایی در یک تصویر هستند و هیچ اطلاعاتی درباره موقعیت نسبی سطوح روشنایی و یا به عبارتی نحوه چینش آنها نسبت به یکدیگر بدست نمی‌دهد. مثلاً دو تصویر موجود در شکل (۴-۲) با وجود داشتن بافت‌های متفاوت دارای هیستوگرام‌های یکسان هستند.

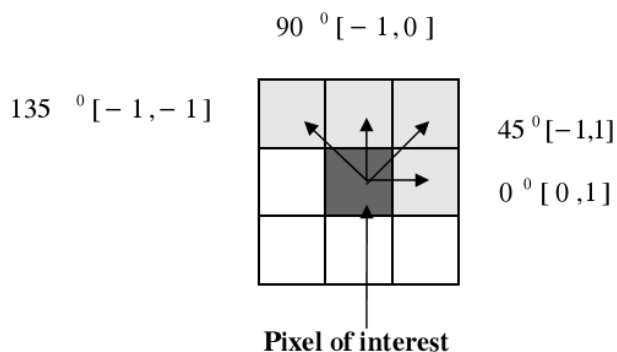


شکل (۴-۲): دو تصویر با بافت‌های مختلف و هیستوگرام‌های یکسان

اطلاعات مربوط به مکان را می‌توان با استفاده از هیستوگرام‌های مرتبه دوم بدست آورد. در محاسبه این هیستوگرام‌ها به‌جای اندازه‌گیری مقدار روشنایی هر پیکسل به تنهایی، پیکسل‌ها را به صورت جفت در نظر می‌گیریم. این هیستوگرام‌ها را که فاصله دو پیکسل و جهت قرارگیری آن دو نسبت به یکدیگر در آنها نقش اساسی بازی می‌کند، با استفاده از ماتریس‌هایی به نام ماتریس هم‌رخداد محاسبه می‌کنند.

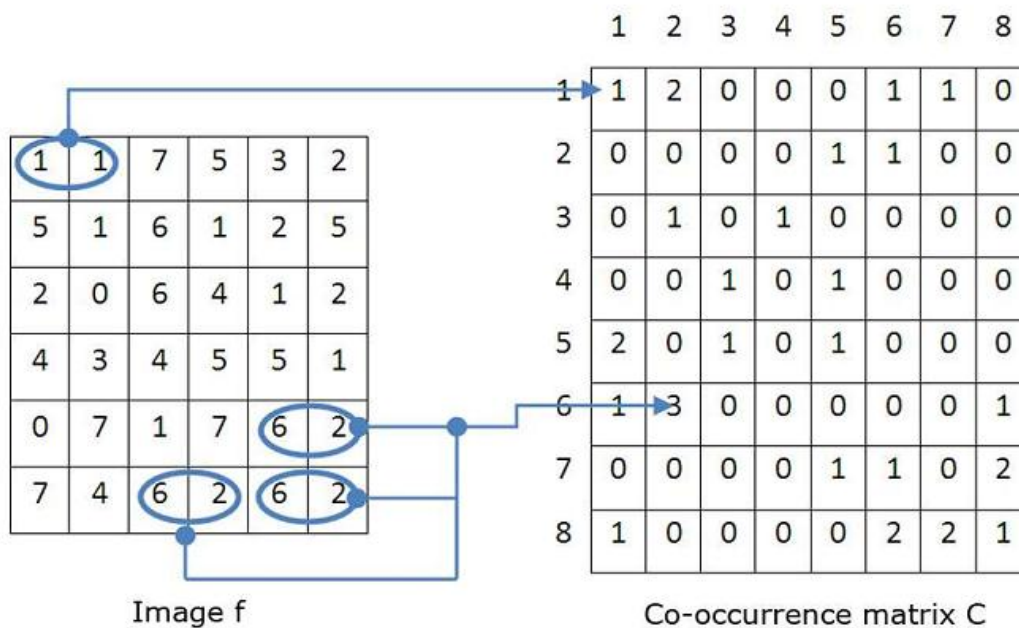
ماتریس هم‌رخداد، نرخ رخداد مقادیر خاکستری دو پیکسل وابسته از تصویر را که در فاصله و جهت بخصوصی از یکدیگر قرار گرفته‌اند، بیان می‌کند. ماتریس هم‌رخداد $P_{i,j}$ ، ماتریس مربعی است که عناصر آن نشان‌دهنده فرکانس نسبی دو پیکسل، یکی با سطح خاکستری i و دیگری با سطح خاکستری j در فاصله خاص و در یک جهت تعیین شده در تصویر می‌باشند. زاویه‌های ممکن بین دو پیکسل ۰، ۴۵، ۹۰ و ۱۳۵ درجه می‌باشند، شکل (۴-۳). تعداد سطوح خاکستری ممکن در تصویر، اندازه ماتریس P را تعیین می‌کند.

شکل (۴-۴) چگونگی ایجاد ماتریس هم‌رخداد را نشان می‌دهد [۸۹]. در این شکل در ماتریس خروجی هم‌رخداد زاویه صفر و فاصله یک، عضو (۱و۱) شامل مقدار ۱ است چون فقط یک نمونه در تصویر ورودی وجود دارد که دو پیکسل مجاور مقادیر ۱ و ۱ دارند.



شکل (۳-۴): چهار جهت مورد استفاده در ساخت ماتریس هم‌رخداد

عضو (۶ و ۲) شامل مقدار ۲ است چون دو نمونه وجود دارد که پیکسل‌های مجاور افقی مقادیر ۶ و ۲ دارند و عضو (۳ و ۱) در ماتریس هم‌رخداد مقدار صفر دارد چون هیچ نمونه‌ای در تصویر وجود ندارد که پیکسل‌های مجاور افقی دارای مقادیر ۳ و ۱ باشد.



شکل (۴-۴): چگونگی ایجاد ماتریس هم‌رخداد

با توجه به تعریف ماتریس هم‌رخداد روشن است که این ماتریس یک ماتریس متقارن است که می‌توان از این واقعیت برای کاهش بار محاسباتی و فضای مورد نیاز استفاده کرد.

• استخراج ویژگی از ماتریس هم‌رخداد:

بعد از تشکیل ماتریس هم‌رخداد بایستی به دنبال استخراج ویژگی‌های مناسبی بود که متمایز کننده یک کلاس از دیگری باشند. چنانچه دیده می‌شود بنا به سطوح خاکستری مختلف نحوه قرارگیری درایه‌های ماتریس فرق می‌کند. ویژگی‌هایی که بتواند چینش‌های گوناگون را متمایز کند باید مد نظر قرار گیرند.

اولین بار هارلیک از ماتریس هم‌رخداد برای استخراج ویژگی بافتی از تصویر به منظور عیب‌یابی میوه گریپ فرت استفاده کرد. هر چقدر مقادیر پیکسل‌های تصویر به همدیگر نزدیک‌تر باشند، به همان میزان تجمع زیادتری روی قطر اصلی ماتریس به وجود خواهد آمد. مزیت استفاده از این ماتریس بر هیستوگرام ساده تصویر این است که در مقایسه با هیستوگرام ساده، در آن اطلاعات مکانی پیکسل‌ها نیز در نظر گرفته می‌شود. به طوریکه هر چه توزیع مقادیر خاکستری گسترده‌تر باشد، واریانس زیادتری در ماتریس دیده خواهد شد.

هارلیک [۹۰]، ۱۴ ویژگی بر روی این ماتریس تعریف نمود. بسیاری از این ویژگی‌ها مشابه می‌باشند. چهار ویژگی متداول که از ماتریس هم‌رخداد P استخراج می‌شوند به صورت روابط زیر بیان می‌شوند:

F6: همبستگی^۱

همبستگی، معیاری از چگونگی همبستگی پیکسل به همسایه‌اش روی کل تصویر است. بازه‌ی مقادیر ۱ تا -۱ است، که متناظر با همبستگی‌های مثبت و منفی کامل است. اگر انحراف استاندارد صفر باشد، این معیار تعریف نمی‌شود. m_r و m_c و σ_r و σ_c به ترتیب میانگین و انحراف استاندارد است که در امتداد سطرها و ستون‌های ماتریس هم‌رخداد P محاسبه شدند.

$$COR = - \sum_i^m \sum_j^n \frac{(i - m_r)(j - m_c)P_{ij}}{\sigma_r \sigma_c} \quad (1-4)$$

^۱ correlation

F7: کنتراست

معیاری از کنتراست بین یک پیکسل و همسایه آن روی کل تصویر است. بازه‌ی مقادیر از صفر تا $(k-1)^2$ است، k بعد ماتریس هم‌رخداد می‌باشد.

$$CON = \sum_i^m \sum_j^n (i-j)^2 P(i,j) \quad (2-4)$$

F8: انرژی

معیاری از یکنواختی است که در بازه‌ی $[0,1]$ است. انرژی برای تصویر ثابت، ۱ است. گ

$$ENR = \sum_i^m \sum_j^n P(i,j)^2 \quad (3-4)$$

F9: همگنی

نزدیکی مکانی توزیع عناصر در G را به قطر، اندازه‌گیری می‌کند. بازه‌ی مقادیر $[0,1]$ است، و ماکزیمم وقتی به دست می‌آید که G ماتریس قطری باشد.

$$HOM = \sum_i^m \sum_j^n \frac{P(i,j)}{1+|i-j|} \quad (4-4)$$

۲-۴-۴ ویژگی‌های حاصل از ماتریس طول-رخداد

این کمیت مجموعه‌ای از پیکسل‌های پشت سرهم است که دارای روشنایی یکسان باشند. طول هر رخداد را تعداد پیکسل‌های موجود در آن رخداد نشان می‌دهند. در واقع با استفاده از این مفهوم تعداد پیکسل‌هایی را که در هر سطح روشنایی مانند l به صورت پشت‌سرهم در یک تصویر در گروه‌های تکی، دوتایی، سه‌تایی و غیره آمده‌اند مشخص می‌کنیم [۹۱]. به عنوان مثال تصویر I را در نظر بگیرید:

$$I = \begin{bmatrix} 0 & 0 & 2 & 2 \\ 1 & 1 & 0 & 0 \\ 3 & 2 & 3 & 3 \\ 3 & 2 & 2 & 2 \end{bmatrix}$$

اگر فرض کنیم که سطوح روشنایی $\{0, 1, 2, 3\}$ باشند، برای هر یک از جهات $0^\circ, 45^\circ, 90^\circ$ و 135° درجه می‌توانیم ۴ ماتریس طول‌رخداد مانند Q_{RL} تعریف کنیم. هر عنصر این ماتریس یعنی $Q_{RL}(i,j)$ ، تعداد دفعاتی را که سطح روشنایی i در جهت تعیین شده در تصویر با طول رخدادی به طول z ظاهر شده است را مشخص می‌کند. $i = 1, 2, 3, \dots, L$ و $j = 1, 2, 3, \dots, R$ است. که R حداکثر مقدار ممکن برای طول رخداد و L تعداد کل سطوح روشنایی در تصویر مورد نظر است. به این ترتیب هر ماتریس طول‌رخداد یک ماتریس $L \times R$ خواهد بود. به عنوان مثال ماتریس طول‌رخداد تصویر I با زاویه صفر و 45° درجه عبارت است از:

$$Q_{RL}(0^\circ) = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix}$$

$$Q_{RL}(45^\circ) = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix}$$

• استخراج ویژگی از ماتریس طول‌رخداد:

حال براساس ماتریس طول‌رخداد می‌توانیم ویژگی‌های زیر را برای معرفی ساختار بافت یک تصویر مشخص کنیم:

F10 : Short-Run Emphasis

$$SRE = \frac{\sum_{i=1}^L \sum_{j=1}^N (Q_{RL}(i,j)/j^2)}{\sum_{i=1}^L \sum_{j=1}^N Q_{RL}(i,j)} \quad (5-4)$$

F11: Long-Run Emphasis

$$LRE = \frac{\sum_{i=1}^L \sum_{j=1}^N (Q_{RL}(i, j)j^2)}{\sum_{i=1}^L \sum_{j=1}^N Q_{RL}(i, j)} \quad (۶-۴)$$

F12: Gray-Level Nonuniformity

$$GLNU = \frac{\sum_{i=1}^L [\sum_{j=1}^N Q_{RL}(i, j)]^2}{\sum_{i=1}^L \sum_{j=1}^N Q_{RL}(i, j)} \quad (۷-۴)$$

F13: Run Length Nonuniformity

$$RLN = \frac{\sum_{i=1}^N [\sum_{j=1}^L Q_{RL}(i, j)]^2}{\sum_{i=1}^L \sum_{j=1}^N Q_{RL}(i, j)} \quad (۸-۴)$$

F14: Run Percentage

$$RP = \frac{\sum_{i=1}^L \sum_{j=1}^N Q_{RL}(i, j)}{RL} \quad (۹-۴)$$

در SRE، مخرج تعداد کل طول رخدادها در ماتریس را نشان می‌دهد که برای مثال فوق به ترتیب ۹ و ۱۶ است. ویژگی SRE از آنجا که بر j^2 تقسیم می‌شود، تاکید بر طول‌رخدادهای کوچک دارد. در حالی که تاکید LRE بر طول‌رخدادهای طولانی‌تر است و در واقع هرچه بافت تصویر نرم‌تر باشد مقدار آن بزرگتر است. در مورد GLNU، جمله موجود در براکت تعداد کل طول‌رخدادها برای هر سطح روشنایی است. وقتی طول‌رخدادها به صورت یکنواخت در سطوح روشنایی توزیع شده باشند، آنگاه مقدار GLNU کوچک خواهد بود. همینطور RLN نیز معیاری برای اندازه‌گیری یکنواختی طول‌رخداد است. در رابطه محاسبه RP، پارامتر L تعداد کل طول‌رخدادهای ممکن در تصویر است. اگر همه run‌ها دارای مقادیر یکسان باشند، L با تعداد کل پیکسل‌ها برابر خواهد بود.

۵-۴ ویژگی‌های شکلی

ویژگی‌های شکلی به نام‌های ویژگی‌های مورفولوژیکی یا ویژگی‌های هندسی نیز بیان می‌شوند. این گروه از ویژگی‌ها بر مبنای شکل ROI محاسبه می‌شوند. برای معرفی یک شکل، ویژگی‌های بسیار زیادی از خواص ساده هندسی تا پیچیده وجود دارد. ویژگی‌های ساده هندسی با آن که ضروری می‌باشند، تنها برای تمایز بین شکل‌های بسیار متفاوت به کار می‌آیند. لازم به ذکر است به علت این که متد پیشنهاد شده در این نوشتار مبتنی بر استخراج ROI است، امکان استخراج ویژگی‌هایی مربوط به شکل توده وجود ندارد. اما در این قسمت برای آشنایی بیشتر خواننده، تعدادی از ویژگی‌های مهم و پر کاربرد از ویژگی‌های شکلی معرفی می‌شود. این ویژگی‌ها در جدول (۳-۴) لیست شده است.

جدول (۳-۴): ویژگی‌های شکلی

Sub-spaces	Feature descriptions
shape Features	F15: Area measure F16: Circularity measure F17: convexity measure F18: solidity measure

F15: اندازه ناحیه

که با A نمایش داده می‌شود و مساحت ناحیه را نشان می‌دهد.

F16: میزان دایرگی^۱

این معیار میزان شباهت یک شکل به دایره را نشان می‌دهد و از رابطه زیر محاسبه می‌شود:

$$Circularity = \frac{A}{A_c} \quad A_c = \frac{P^2}{4\pi} \quad (۱۰-۴)$$

^۱ Circularity measure

که A مساحت شکل است و A_c مساحت دایره‌ای با محیط شکل اصلی می‌باشد. مقدار A_c برای دایره ۱ است و هر چه شکل، شباهت کمتری به دایره داشته باشد، این مقدار کمتر می‌شود.

F17: میزان محدبی^۱

این ویژگی طبق رابطه (۴-۱۲) محاسبه می‌شود. که $P_{convex-Hull}$ ، محیط کوچکترین شکل محدبی می‌باشد که شکل اصلی را در بر می‌گیرد و P محیط شکل اصلی می‌باشد.

$$convexity = P_{convex-Hull} / P \quad (۴-۱۱)$$

F18: میزان سفتی^۲

این معیار نیز مربوط به محدبی می‌باشد، با این تفاوت که مساحت را بررسی می‌کند و از رابطه زیر بدست می‌آید:

$$solidity = A / A_{Convex-Hull} \quad (۴-۱۲)$$

که A مساحت شکل است و $A_{Convex-Hull}$ مساحت Convex – Hull است.

¹ Convexity

² Solidity

فصل پنجم:

دسته بندی

۵-۱ مقدمه

در این فصل وارد مسئله دسته‌بندی می‌شویم. دسته‌بندی یعنی نسبت دادن یک کلاس به هر یک از بردارهای ویژگی که توسط اندازه‌گیری از هر یک از سمپل‌های یک مسئله خاص بدست آمده‌اند. این نسبت دادن کلاس می‌تواند با یک روش یادگیری بانظارت یا بدون نظارت انجام شود. در یادگیری بدون نظارت، هیچ اطلاعاتی در مورد برچسب‌های کلاس‌های اندازه‌گیری شده قابل دسترس نیست. اما در یادگیری با نظارت، یک مجموعه داده وجود دارد که در آن اندازه‌گیری‌ها دارای برچسب هستند. در CADهای شناسائی توده، بیشتر یادگیری با نظارت ترجیح داده می‌شود. در این مطالعه، به منظور نشان دادن کارائی ویژگی‌های پیشنهاد شده در کاربردهای CAD، ۳ نوع روش طبقه‌بند با نظارت، معرفی شده و قدرت آنها در دسته‌بندی با یکدیگر مقایسه خواهد شد.

به منظور دسته‌بندی ROIها براساس ویژگی‌های پیشنهاد شده می‌توان از متدهای گوناگونی استفاده نمود. همانگونه که در فصل ۲ نیز بیان شد تا به اکنون مقالات متعدد به این منظور از متدهای طبقه‌بندی زیادی همچون انواع شبکه‌های عصبی مصنوعی، SVM، درخت‌های تصمیم دودویی، بیزین، LDA و غیره استفاده کرده‌اند.

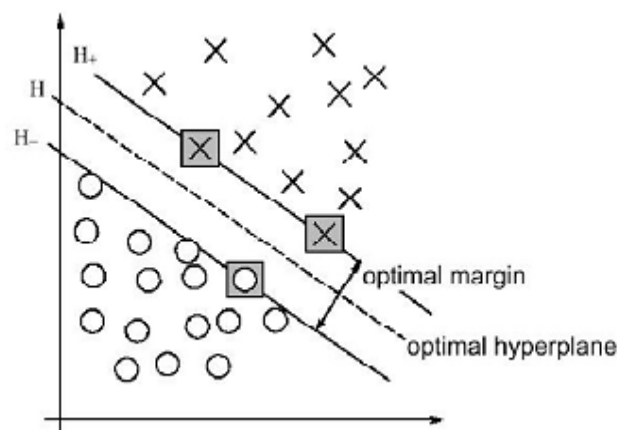
در این پایان‌نامه در ابتدا به منظور مقایسه عملکرد طبقه‌بندهای مختلف بر روی ویژگی‌های پیشنهادی، از سه روش طبقه‌بند پایه SVM، KNN، و شبکه عصبی مصنوعی MLP استفاده شده است. سپس به منظور افزایش دقت طبقه‌بندی، از راهکار ترکیب اطلاعات حاصل از طبقه‌بندهای پایه استفاده شده است.

این فصل شامل ۴ قسمت کلی می‌شود. در ۳ قسمت اول، اصول اولیه طراحی، مشخصات و پیاده‌سازی سه طبقه‌بند پایه SVM، KNN و MLP مورد بررسی قرار گرفته است. و قسمت آخر به معرفی راهکار ترکیب اطلاعات حاصل از طبقه‌بندهای پایه پرداخته می‌شود.

۲-۵ طبقه‌بند SVM^۱

SVM یک روش بسیار قدرتمند در دسته‌بندی است. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نشان داده است. مبنای کاری طبقه‌بندی کننده SVM، دسته‌بندی خطی داده‌ها است. اولین مسئله‌ای که SVM باید به آن پردازد این است که کدام خط مناسب‌ترین گزینه برای تقسیم‌بندی خطی داده‌ها است. SVM در تقسیم خطی داده‌ها به دنبال خطی است که تا آنجا که ممکن است از داده‌های هر دو کلاس بیشترین فاصله را داشته باشد.

SVM در سال ۱۹۹۲ توسط وینک معرفی شد [۹۳]. اخیراً این مدل برای حل مسائل زیادی به کار رفته است. در این پایان نامه از این سیستم یادگیری برای تشخیص نواحی شامل توده استفاده خواهد شد. ایده اصلی یک طبقه‌بند SVM در شکل (۵-۱) نمایش داده شده است:



شکل (۵-۱): اساس کار SVM.

در این قسمت به طور خلاصه روش کار SVM شرح داده می‌شود. برای مطالعه جزئیات بیشتر می‌توان به مراجع [۹۴ و ۹۵] مراجعه کرد.

الف. SVM خطی و تفکیک‌پذیر

^۱ Support -Vector Machine (SVM)

شکل (۱-۵) ساده‌ترین حالت را نشان می‌دهد که در آن بردارهای ویژگی (x) با یک ابرصفحه مجزاساز خطی از هم جدا می‌شوند. نمونه‌های آموزش $\{(x_i, d_i)\}_{i=1}^N$ را در نظر می‌گیریم، که x_i الگوی ورودی نمونه‌ی i ام و $d_i \in \{-1, 1\}$ خروجی مطلوب است.

فرض می‌شود الگوهای $+1$ و الگوهای -1 به صورت خطی تفکیک‌پذیر می‌باشند. معادله‌ی سطح تصمیم‌گیری (ابر صفحه جدا کننده) به شکل زیر است:

$$w^T + b = 0 \quad (۱ - ۵)$$

که در آن x بردار ورودی، w بردار وزن و b بایاس آن است. برای بردار وزن w و بایاس b ، جداسازی توسط صفحه‌ی رابطه (۱-۵) انجام می‌شود. با انتخاب w و b های مختلف، صفحه‌های مختلفی می‌تواند به عنوان صفحه مجزاساز انتخاب شود. فاصله‌ی نزدیک‌ترین نقاط داده به صفحه جدا کننده، حاشیه‌ی جداسازی نامیده می‌شود (M). هدف SVM آن است که ابر صفحه‌ای را به عنوان سطح تصمیم‌گیری پیدا کند که این حاشیه‌ی جداسازی را بیشینه کند. اگر w_0 و b_0 مقادیر بهینه‌ی بردار وزن و بایاس را نشان دهند، ابر صفحه‌ی بهینه‌ی چند بعدی به شکل زیر نشان داده می‌شود:

$$w_0^T x_i + b_0 = 0 \quad (۲ - ۵)$$

جفت پارامترهای w_0 و b_0 برای این هدف باید شرایط زیر را برآورده کند.

$$w_0^T x_i + b_0 \geq +1 \quad \text{for } d_i = +1 \quad (۳ - ۵)$$

$$w_0^T x_i + b_0 \leq -1 \quad \text{for } d_i = -1$$

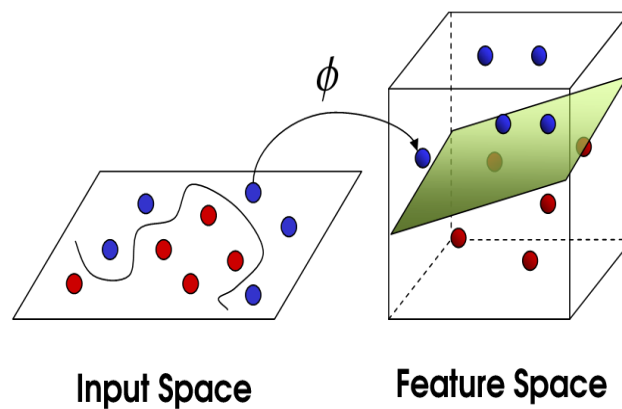
بنابراین صفحه مثبت را به شکل $w_0^T x_i + b_0 = +1$ و صفحه منفی را به شکل $w_0^T x_i + b_0 = -1$ نمایش می‌دهیم.

مقدار حاشیه‌ی جداسازی^۱ (M) را از رابطه (۴-۵) محاسبه می‌کنیم. این نکته واضح است که بیشینه‌سازی حاشیه‌ی جداسازی بین دو کلاس معادل است با کمینه‌سازی نرم اقلیدسی بردار وزن w .

$$M = \frac{2}{\|w\|} \quad (۴ - ۵)$$

ب. SVM غیرخطی

SVM خطی برای دسته‌بندی الگوهای یک مسئله دو کلاسه، از مرزهای جدا کننده خطی و از یک ابرصفحه استفاده می‌کنند. با این وجود، می‌توان مسائلی که به صورت خطی جداپذیر نیستند را با نگاشت نقاط نمونه‌ها به فضایی با بعد بیشتر با استفاده از یک نگاشت غیر خطی حل نمود. یعنی یک نگاشت $\phi: R^d \rightarrow H$ انتخاب شود به گونه‌ای که بعد H از d بیشتر باشد. سپس به دنبال ابرصفحه متمایز کننده در این فضا می‌گردیم که معادل سطح غیرخطی متمایز کننده در فضای R^d است.



شکل (۲-۵): افزایش بعد برای رسیدن به نمایش داده‌هایی که به طور خطی تفکیک‌پذیر باشند

به طور کلی SVM بردارهای ورودی را با استفاده از یک کرنل به یک فضای بزرگتری می‌برد که تفکیک‌پذیری خطی در آن بهتر و قاطعانه‌تر انجام می‌شود. با استفاده از کرنل‌های مختلف، می‌توان روش‌های فراگیری مختلفی با انواع سطوح تصمیم‌گیری دلخواه ساخت. به این ترتیب است که SVM

^۱ Margin

به عنوان یک کلاس جدید از طبقه‌بندی کننده‌ها و ماشین‌های تشخیص الگو در می‌آید. کرنل‌های مختلفی وجود دارند که برخی از نمونه‌های مشهور آن در جدول (۵-۱) آمده است.

جدول (۵-۱): انواع کرنل‌های مشهور

Kernel	Formula
Sigmoid	$K(u,v)=\tanh(au.v+b)$ (۵-۵)
Polynomial	$K(u,v)= (1 + u.v)^2$ (۶-۵)
RBF	$K(u,v)=\exp(-\alpha\ u - v\ ^2)$ (۷-۵)

۳-۵ طبقه‌بند شبکه عصبی مصنوعی

شبکه‌های عصبی به صورت یک شبکه توزیع شده موازی بر مبنای فرآیند یادگیری زیست شناختی مغز انسان است که کاربردهای بسیاری در زمینه طبقه‌بندی دارد. شبکه عصبی مصنوعی به طور کلی از تعداد زیادی نرون مصنوعی تشکیل شده و امکان تصمیم‌گیری برای تعیین کلاس بردارهای ورودی را فراهم می‌نماید. بسته به نوع ارتباط مابین نرون‌ها ساختارهای گوناگونی برای شبکه عصبی معرفی شده است. از میان آنها ساختار MLP^۱ بیش از سایرین در CAD های تشخیص توده مورد استفاده قرار می‌گیرد [۴۶ و ۵۴]. برای دسته بندی ROIها در این پایان نامه، یک شبکه عصبی چندلایه روبه جلو (MLP) با یک لایه مخفی پیاده سازی شده است.

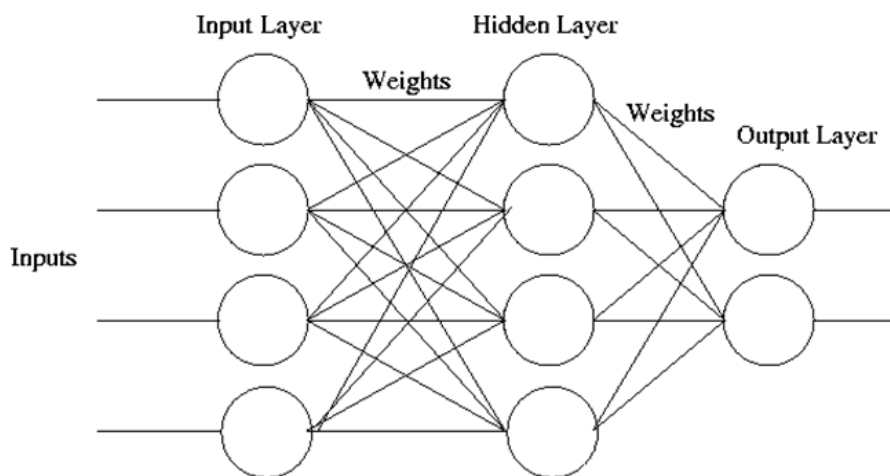
یک پرسپترون یا یک شبکه تک لایه قابلیت عملیاتی و دسته‌بندی محدودی دارد یعنی تنها در مواردی کاربرد دارد که داده‌های تحت دسته‌بندی به سادگی و توسط رسم یک خط از هم‌دیگر جداپذیر باشند. در حالیکه در دنیای واقعی اکثر مسائل ما به این صورت نبوده و مرز داده‌ها درون

¹ Multi Layer Perceptron

یکدیگر هم پوشانی دارند لذا به توابع غیرخطی و دسته‌بندهای قدرتمندی که بتواند داده‌های غیر خطی را از هم جدا سازد نیاز است. شبکه‌ی عصبی MLP دارای یک ساختار چند لایه می‌باشد که علاوه بر لایه ورودی و خروجی مشتمل بر چند لایه مخفی یا میانی می‌باشد. می‌توان این گونه گفت که یک شبکه عصبی بنا بر تعداد لایه‌هایی که در ساختار خود استفاده می‌کند می‌تواند یک پرسپترون و یا یک شبکه MLP باشد.

توانایی جداسازی داده‌های غیرخطی توسط اضافه کردن لایه‌های مخفی به شبکه داده می‌شود. لایه‌های مخفی مستقیماً نه به داده‌های ورودی متصلند و نه به خروجی‌ها. وظیفه آنها انجام یک سری عملیات و محاسبات برای دسته‌بندی است. یعنی چنانچه محاسبات ما پیچیده بوده و در یک لایه قابل انجام نباشد محاسبات در لایه‌های مخفی انجام می‌گیرد. در شکل (۳-۵) مثالی از یک ساختار شبکه عصبی MLP را با یک لایه مخفی مشاهده می‌کنید.

شبکه‌های عصبی MLP دو وجه عملیات دارند. اولین آن وجه جلورونده است، که متشکل از بیان یک الگو توسط نودهای ورودی و عبور سیگنال‌ها از طریق شبکه به منظور گرفتن خروجی است. وجه دوم عملیات یادگیری است که متشکل است از بیان الگوهای ورودی و یافتن پارامترهای شبکه (وزن‌ها)، به نحوی که فاصله بین خروجی محاسبه شده و خروجی مطلوب مینیمم شود.



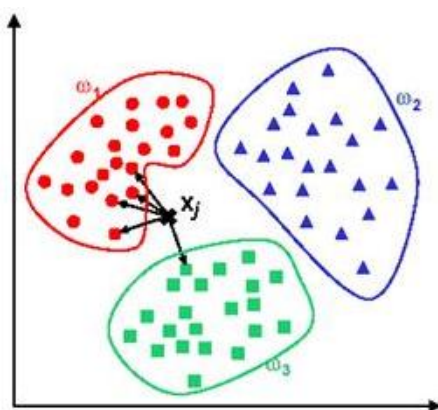
شکل (۳-۵): نمونه‌ای از ساختار یک شبکه عصبی مصنوعی MLP

۴-۵ طبقه‌بند K-NN

یکی از بهترین مدل‌های ارائه شده در روش‌های آماری، روش K-NN^۱ می‌باشد که بر پایه شمارش و مفهوم فاصله استوار است.

الگوریتم K-NN یکی از روش‌های طبقه‌بندی عضو گروهی، موسوم به یادگیرنده‌های مبتنی بر نمونه است. این روش طبقه‌بندی یکی از پرکاربردترین روش‌های طبقه‌بندی در مسائل یادگیری ماشین است که نمونه‌ها را بر اساس میزان شباهت آنها با نمونه‌های مشخص طبقه‌بندی می‌کند. ایده اصلی این تکنیک این است که به یک داده تست مشاهده‌نشده^۲، کلاس مربوط به نزدیک‌ترین همسایه موجود در مجموعه آموزشی را نسبت دهیم شکل (۴-۵). میزان شباهت هر نمونه نامشخص در واقع میزان نزدیک بودن آن نمونه نامشخص به نمونه‌های مشخص می‌باشد که به آن نمونه‌های مشخص، مجموعه آموزشی می‌گویند.

به طور کلی برای طبقه‌بندی داده‌هایی که تغییرپذیری‌های محلی دارند، لازم است دیدی از همسایگی هر نمونه برای طبقه‌بندی‌اش داشته باشیم که این دید با استفاده از مفهوم معیار شباهت محقق می‌شود.



شکل (۴-۵): اساس کار طبقه‌بند K-NN

^۱ K-Nearest Neighbors (K-NN)

^۲ unseen

اندازه گیری فاصله (معیار شباهت):

کارائی الگوریتم K-NN وابسته به انتخاب یک تابع مناسب برای اندازه گیری فاصله (معیار شباهت) است. در بسیاری از موارد، انتخاب این معیار بستگی به داده‌ها یا موضوع مسئله مورد بررسی ما دارد، بطوریکه هیچ اندازه گیری جهانی که بهترین راه حل برای تمام مسائل دسته بندی را ارائه بدهند وجود ندارد. در ادامه ۴ معیار شباهت مهم و پر کاربرد معرفی شده است [۹۶].

الف) معیار شباهت کسینوسی

معیار شباهت کسینوسی یک تابع فاصله برداری است که فاصله بین دو بردار (k, j) را براساس زاویه بین آنها محاسبه می کند. این معیار شباهت که برای اولین بار توسط هیرست به کار برده شد، هنگامی به کار می رود که داده تنک، پراکنده و نامتقارن باشد. محاسبه کسینوس زاویه بین دو بردار مشخص می کند که آیا این دو بردار تقریباً در جهت یکسانی نقطه گذاری شده اند یا نه. نتایج معیار شباهت کسینوسی هنگامی برابر با ۱ است که زاویه اندازه گیری شده ۰ باشد، که در این صورت این دو بردار به جز در مقدار بزرگی شان، با یکدیگر یکسان هستند. هنگامی این فاصله کمتر از ۱ است که این زاویه هر مقدار دیگری داشته باشد. مقدار این معیار شباهت وقتی ۰ است که دو بردار با هم زاویه ۹۰ درجه تشکیل دهند و این یعنی اینکه آنها هیچ جمله ای را به اشتراک نمی گذارند. این فاصله با به کارگیری یک ضرب نقطه ای و مقدار بزرگی به صورت زیر بیان می شود:

$$\text{Cosin_Metric}(k, j) = \cos \theta = \frac{p \cdot q}{|p| |q|} = \frac{\sum_{i=1}^n p_i * q_i}{\sqrt{\sum_{i=1}^n (p_i)^2} * \sqrt{\sum_{i=1}^n (q_i)^2}} \quad (۸-۵)$$

ب) معیار شباهت چبیشف^۱

در ریاضیات، فاصله چبیشف یا متریک ماکزیمم، متریکی است بر روی فضای برداری، که در آن فاصله بین دو بردار از اختلافشان در اروشاد هر یک از ابعاد دستگاه مختصات بزرگتر است [۵۷]. این فاصله از رابطه زیر بدست می‌آید:

$$D_{\text{Chebyshev}}(k, j) = \max(|p_i - q_j|) \quad (۹-۵)$$

ج) فاصله اقلیدسی^۲

پرکاربردترین معیار شباهت برای اندازه‌گیری فاصله، معیار شباهت فاصله اقلیدسی است. هنگامی که داده متراکم یا پیوسته است، این فاصله بهترین معیار جهت اندازه‌گیری فاصله است. در واقع با داشتن دو بردار p و q در صفحه xy ، اگر این دو برار در راستای محورها باشند، فاصله اقلیدسی بین آنها برابر با طول وتر مثلث قائم‌الزاویه خواهد بود. این فاصله از رابطه زیر بدست می‌آید:

$$\text{Euclidean Distance}(k, j) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (۱۰-۵)$$

د) معیار فاصله مربعی^۳ (فاصله ماهالانوبیس)

فاصله مربعی به عنوان فاصله در محور افقی بعلاوه فاصله در محور عمودی محاسبه می‌شود. فاصله مربعی همیشه بزرگتر و یا مساوی با صفر است. اندازه این فاصله برای نقاط همانند صفر و برای نقاطی که شباهت کمی را نشان می‌دهند، مقدار آن بالا خواهد بود. در بسیاری از موارد، این اندازه‌گیری فاصله موجب نتایجی مشابه با فاصله اقلیدسی می‌شود. این فاصله از رابطه زیر محاسبه می‌شود:

¹ Chebyshev

² Euclidean Distance

³ City block

$$D_{city\ block}(k, j) = \sum_{i=1}^n |p_i - q_i| \quad (11-5)$$

۵-۵ ترکیب طبقه‌بندها

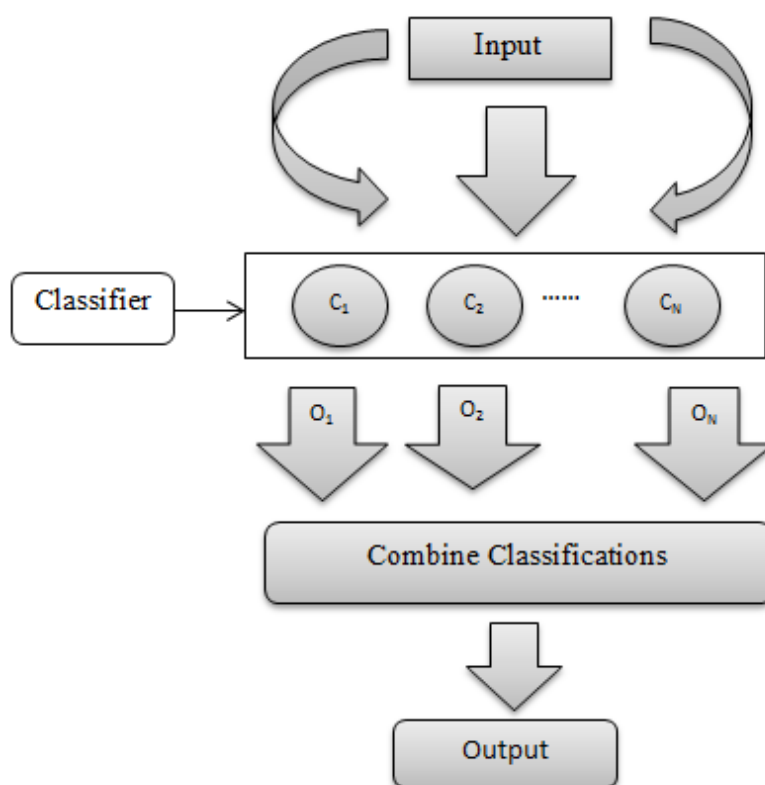
در سطوح مختلف سیستم های CAD مانند برداشت داده، استخراج ویژگی و نتایج دسته‌بندی، می‌توان فرآیند ترکیب اطلاعات را به کارگرفت. در این پژوهش بر ترکیب اطلاعات در سطح طبقه‌بندها تمرکز شده است شکل (۵-۵). برای این منظور اعلام نظر نهایی در مورد کلاس هر ورودی با تصمیم‌گیری براساس نتایج چند طبقه‌بند مختلف انجام می‌شود.

نکته قابل توجه در استفاده از این راهکارها، انتخاب مناسب طبقه‌بندهای پایه است. افزایش تعداد طبقه‌بندهای پایه، لزوماً دقت دسته‌بندی را بهتر نخواهد کرد. نکته مهم در انتخاب طبقه‌بندهای پایه، ساختار آموزش متفاوت آنها و در نتیجه اختلاف نظر^۱ آنها در نتایج دسته‌بندی است. به عنوان مثال استفاده از دو طبقه‌بند آماری K-NN و بیزین با هم مناسب نخواهد بود. زیرا دارای ساختارهای آموزش یکسان هستند. اما استفاده از SVM در کنار یک طبقه‌بند آماری همچون K-NN مناسب خواهد بود، زیرا دارای روندهای آموزشی کاملاً متفاوت هستند.

در این پایان‌نامه از سه طبقه‌بند پایه SVM، K-NN و شبکه عصبی استفاده شده است. مهمترین دلیل برای انتخاب طبقه‌بند شبکه عصبی، خواص غیر پارامتری آن است. بر خلاف روش‌های طبقه‌بندی آماری مثل K-NN، در شبکه عصبی به هیچگونه اطلاعاتی درباره توزیع احتمالاتی اصلی نیازی نیست. از آنجا که شبکه عصبی می‌تواند پارامترهای خود را در مرحله آموزش با استفاده از مثال‌هایی آموزش دهد، می‌توان در موارد عملی و کاربردی مثل تشخیص توده‌ها که داده‌ها نایب هستند از آن استفاده کرد.

¹ Diversity

تاکنون راهکارهای مختلفی برای اجماع نظر بین طبقه‌بندها ارائه شده است [۹۷]. در این میان روش‌های مبتنی بر تصمیم‌گیری فازی دارای عملکرد مناسب به همراه حجم محاسباتی پایین هستند که فاکتورهای مثبت در زمینه کاربردهای تشخیص سرطان محسوب می‌شوند. OWA از جمله روش‌های تصمیم‌گیری فازی مورد استفاده در ترکیب طبقه‌بندها می‌باشد، که در ادامه در مورد آن توضیحاتی بیان شده است.



شکل (۵-۵): یک نمای کلی از ترکیب طبقه‌بندها

۱-۵-۵ روش OWA

روش OWA^۱ از جمله روش‌های تصمیم‌گیری فازی مورد استفاده در ترکیب طبقه‌بندها می‌باشد. این روش اولین بار توسط Yager معرفی شد تا به کمک آن بتوان به یک اجماع بین امتیازهای متفاوت

^۱ Ordered Weighted Averaging (OWA)

داده شده براساس معیارهای مختلف دست یافت [۹۸]. عملگر OWA در فضای n بعدی یک نگاشت $R^n \rightarrow R$ انجام می‌دهد که به صورت رابطه (۱۲-۵) تعریف می‌شود.

$$OWA(x_1, x_2, \dots, x_n) = \sum_{j=1}^n w_j x_{\sigma(j)} \quad (12-5)$$

که در آن x_1, x_2, \dots, x_n نتایج هریک از طبقه‌بندها می‌باشد که ورودی فرآیند OWA می‌باشد و سیگما یک جایگشت از ورودی‌هاست که آنها را به صورت زیر مرتب می‌کند:

$$x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$$

w_j ضرایب اثر هر ورودی است که همگی غیرمنفی بوده و شرط زیر را ارضاع می‌نمایند.

$$\sum_{i=1}^n w_i = 1 \quad (13-5)$$

نتایج تحقیقات نشان می‌دهد که این روش به علت قدرت تطابق پذیری بالا بسیار کارآمد می‌باشد [۹۸]. با استفاده از روش OWA می‌توان به انواع راه‌های اجماع بر ورودی‌ها مانند حداکثر، حداقل و یا میانگین دست یافت. تمامی این شرایط با تنظیم مقدار ضرایب w میسر خواهد شد. به منظور تعیین ضرایب w براساس توابع نمایی دو راهکار خوشبینانه و بدبینانه معرفی شده است [۹۹].
خوشبینانه:

$$\begin{aligned} w_1 &= a; \\ w_2 &= a(1-a); \\ w_3 &= a(1-a)^2; \\ &\dots \\ w_{n-1} &= a(1-a)^{n-2}; \\ w_n &= (1-a)^{n-1} \end{aligned} \quad (14-5)$$

بدینانه:

$$\begin{aligned}w_1 &= a^{n-1}; \\w_2 &= (1-a)a^{n-2}; \\w_3 &= (1-a)a^{n-3}; \\&\dots \\w_{n-1} &= (1-a)a; \\w_n &= (1-a)\end{aligned}\tag{۱۵-۵}$$

در روابط (۱۴-۵) و (۱۵-۵) پارامتر a عددی متعلق به بازه $[0, 1]$ می‌باشد. در صورتی که پاسخ صحیح دسته‌بندی برای هر ورودی را X_d بنامیم خطای تصمیم‌گیری طبقه‌بندی‌های مورد استفاده برابر است با:

$$e = \frac{1}{2} \left(\sum_{j=1}^n w_j x_{\sigma(j)} - X_d \right)^2\tag{۱۶-۵}$$

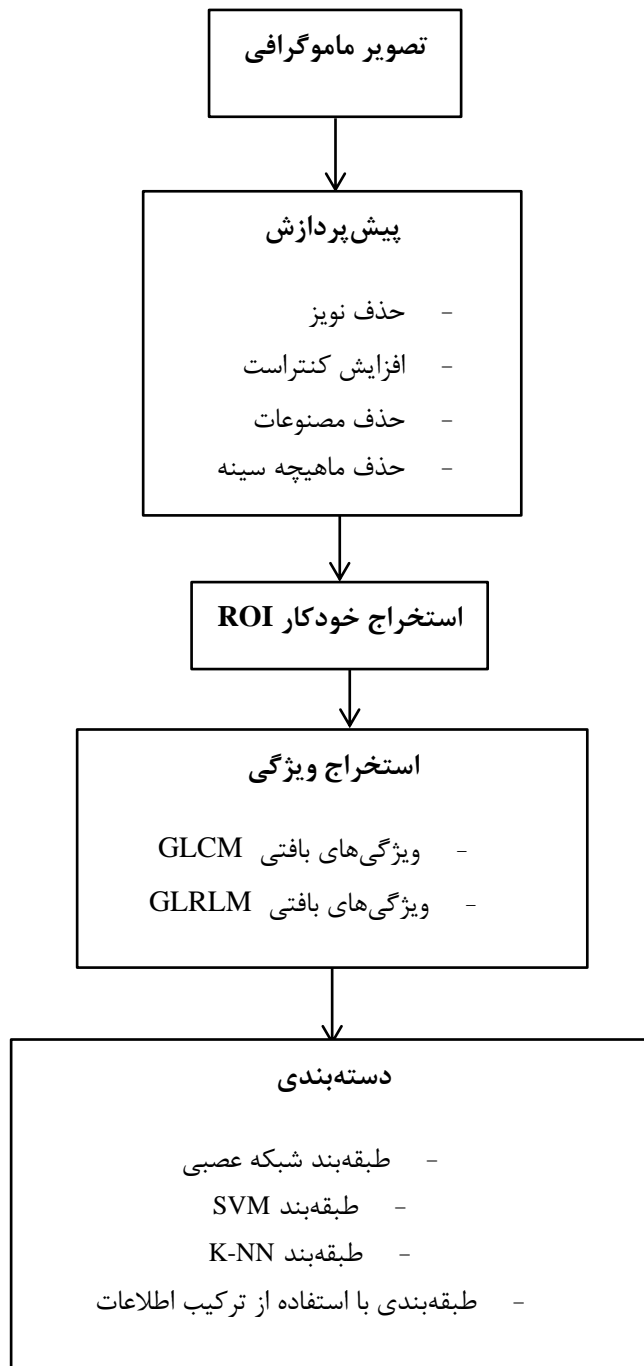
بنابراین ضرایب عملگر OWA باید طوری تعیین گردد که مقدار خطا کمینه شود. طبق روابط (۱۴-۵) و (۱۵-۵) به منظور بهینه‌سازی ضرایب، مقدار پارامتر a را از صفر تا یک تغییر می‌دهیم تا ضرایبی که به ازاء آن کمینه خطا ایجاد شود بدست آید.

فصل هشتم:

نتایج شیشه سازی

۱-۶ مقدمه

در این قسمت، روش استفاده شده در سیستم CAD پیشنهادی به طور کامل شرح داده می‌شود. نمودار کلی این روش در شکل (۱-۶) رسم شده است.



شکل (۱-۶): نمودار کلی سیستم CAD پیشنهادی

در فصل‌های قبل، تمام قسمت‌های این نمودار به طور کامل توضیح داده شده است. در این فصل به ارائه نتایج حاصل از شبیه‌سازی سیستم‌های CAD تشخیص توده پرداخته می‌شود. در ابتدا پایگاه داده استفاده شده و معیارهای ارزیابی مورد استفاده برای سنجش سیستم‌های CAD معرفی می‌شود. در ادامه تاثیر ویژگی‌های مختلف در نتایج آزمایش را بیان می‌کنیم. پس از آن نتایج شبیه‌سازی سیستم‌های CAD تشخیص توده بر روی تصاویر پایگاه داده مرتبط با هر یک از طبقه‌بندی‌های پایه و جزئیات مربوط به اعمال هر روش مورد بررسی قرار گرفته است.

در نهایت نیز از روش رویکرد ترکیب اطلاعات حاصل از هر یک از طبقه‌بندی‌ها استفاده کرده و به بالاترین سطح دقت در دسته‌بندی خواهیم رسید. برای این منظور از روش OWA استفاده کرده‌ایم.

۲-۶ پایگاه داده

در این قسمت به معرفی پایگاه داده DDSM^۱ که در این نوشتار مورد استفاده قرار گرفته است، خواهیم پرداخت. پایگاه داده‌ای DDSM با همکاری بیمارستان عمومی ماساچوست^۲ (D.Kopans,R.Moore)، دانشگاه فلوریدای جنوبی^۳ (K.Bowyer) و آزمایشگاه ملی سندیا^۴ (P.Kegelmeyer) تهیه شده است. هدف‌های اصلی این پروژه یکی تسهیل انجام تحقیقاتی برای توسعه الگوریتم‌های کامپیوتری که به تصویربرداری کمک می‌کنند و دیگری توسعه الگوریتم‌های تشخیص بیماری و همچنین کمک‌های آموزشی و تحصیلی می‌باشند.

این پایگاه داده شامل بیش از ۳۰۰۰ تصویر ماموگرافی است که با اسکنرهای LUMISYS و HOWTEK به صورت دیجیتال درآورده شدند. عمق تصاویر ۱۲ یا ۱۶ بیت می‌باشد. تصاویر به صورت Lossless JPEG کد شده اند و رمز گشایی آنها در محیط LINUX امکان پذیر است.

^۱ Digital Database Screening Mammography

^۲ Massachusetts General Hospital

^۳ University of South Florida

^۴ Sandia National Laboratories

این پروژه شامل ۲۵۰۰ نمونه مطالعه شده است. هر نمونه شامل دو تصویر از هر سینه، به همراه اطلاعاتی در مورد بیمار (شامل سن بیمار در زمان مطالعه، شدت بیماری و نوع بیماری و ...) و اطلاعاتی راجع به تصویر (نوع اسکنر، رزولوشن تصویر و ...) می‌باشد. در هر نمونه در این دادگان، ۴ تصویر چپ و راست از نماهای MLO و CC وجود دارد. موارد غیرعادی یک فایل Overlay دارند که مشخصات ضایعه را در خود دارد. در این فایل طرح کلی (غیردقیق) از مرز توده به صورت یک کد زنجیری نیز موجود است که توسط رادیولوژیست‌ها کشیده شده است. از دیگر مواردی که در این فایل موجود است می‌توان به نوع توده و نوع مرز و درجه پیشرفت اشاره کرد (در مورد سرطان) و البته خوش خیمی و بدخیمی نیز در آن مشخص شده است.

۳-۶ معیارهای ارزیابی

در این تحقیق به منظور بررسی نتایج حاصل از سیستم‌های CAD شبیه‌سازی شده، چند معیار مهم مورد ارزیابی قرار گرفته است. این معیارها که در کاربردهای مشابه نیز مورد استناد قرار گرفته شده است [۵۴]، عبارتند از دقت (Accuracy)، حساسیت (Sensitivity) و اختصاصیت (Specificity). هر چه مقادیر این سه معیار به یک نزدیک‌تر باشند، عملکرد سیستم بهتر خواهد بود.

برای محاسبه معیارهای ذکر شده، به دو دسته‌بندی نیاز است. یکی دسته‌بندی واقعی و دیگری دسته‌بندی پیش‌بینی شده براساس مدل. مثلاً در مورد تشخیص سرطان سینه، دسته‌بندی‌ها عبارتند از یکی نظر متخصصان و رادیولوژیست‌ها در جداسازی نواحی سرطانی و نواحی نرمال و دوم دسته‌بندی نواحی براساس خروجی سیستم CAD. براساس اطلاعات بدست آمده از مدل مربوطه، پارامترهایی تعریف می‌شود. این پارامترها عبارتند از :

- TP^1 : تعداد نواحی که از نظر متخصصان و رادیولوژیست ها شامل توده هستند و توسط سیستم CAD نیز توده شناسایی شده‌اند.

- TN^2 : تعداد نواحی که از نظر متخصصان و رادیولوژیست‌ها نرمال و فاقد بیماری هستند و توسط سیستم CAD نرمال شناسایی شده‌اند.

- FP^3 : تعداد نواحی که از نظر متخصصان و رادیولوژیست ها نرمال و فاقد بیماری هستند اما توسط سیستم CAD به عنوان توده شناسایی شده‌اند.

- FN^4 : تعداد نواحی که از نظر متخصصان و رادیولوژیست ها شامل توده هستند اما توسط سیستم CAD نرمال شناسایی شده‌اند.

حال با توجه به پارامترهای تعریف شده در بالا، معیارهای دقت، حساسیت و اختصاصیت به صورت رابطه‌های (۱-۶) تا (۳-۶) تعریف می‌شوند [۱۴].

$$Accuracy = (ACC) = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (1-6)$$

$$Sensitivity = (SN) = \frac{TP}{TP+FN} \quad (2-6)$$

$$Specificity = (SP) = \frac{TN}{(FP+TN)} \quad (3-6)$$

حساسیت بدین معنی است که چقدر سیستم CAD موردنظر ما توانسته است سرطان را تشخیص دهد. این همان معنی TP/P می‌باشد (P برابر است با تعداد کل موارد سرطانی با استفاده از یک روش دقیق و به کمک متخصصان و رادیولوژیست‌ها). به عبارت دیگر حساسیت، تعداد سرطانی‌های درست انتخاب

¹ True Positive

² True Negative

³ False Positive

⁴ False Negative

شده توسط سیستم CAD، به نسبت کل مواردی که واقعا سرطانی هستند را نشان می‌دهد. بنابراین این پارامتر می‌تواند معیاری برای اندازه‌گیری نرخ موفقیت مدل در تشخیص بیماری باشد [۴۷].

اختصاصیت به طور دقیق نشان‌دهنده تعداد موارد غیرسرطانی است که توسط سیستم CAD، منفی در نظر گرفته شده است. به عبارت دیگر اختصاصیت معیاری از توانایی سیستم CAD برای انتخاب درست مواردی که فاقد سرطان است، می‌باشد.

و در آخر دقت، بیانگر کیفیت دسته‌بندی باینری است. در حالیکه حساسیت معیاری تنها مرتبط با موارد سرطانی و اختصاصیت معیاری تنها مرتبط با موارد غیر سرطانی است، این معیار تعادلی بین دو معیار حساسیت و اختصاصیت ایجاد می‌کند.

خروجی هر یک از سیستم‌های CAD شبیه‌سازی شده در این تحقیق، ماتریسی به نام کانفیوژن ماتریس است. کانفیوژن ماتریس یک ماتریس ۲×۲ است که ستون‌های آن بیانگر نتایج پیش‌بینی شده بر اساس سیستم CAD و سطرهای آن بیانگر نتایج واقعی است. و در نتیجه هر کدام از آرایه‌های آن متناظر با پارامترهای TP، TN، FP و FN است. جدول (۶-۱) ساختار و عناصر یک کانفیوژن ماتریس را بیان کرده است.

جدول (۶-۱): ساختار کانفیوژن ماتریس

Actual	Predicted	
	Cancer (Positive)	Normal (Negative)
Cancer (Positive)	True Positive (TP)	False Negative (FN)
Normal (Negative)	False Positive (FP)	True Negative (TN)

۴-۶ داده‌های آموزشی و آزمایشی

مجموعه داده‌های آموزشی و آزمایشی که در طبقه‌بندها مورد استفاده قرار گرفته است از تعداد ۲۵۰ عدد ROI (نمونه) تشکیل شده است. به ازای هر ROI، یک بردار ویژگی استخراج می‌شود. ROI ها از تصاویر ماموگرافی پایگاه داده DDSM استخراج شده‌اند. تصاویر مورد نظر به صورت رندوم و از هر دو نمای CC و MLO برگزیده شده‌اند.

برای دسته‌بندی و آموزش، لازم است که نمونه‌ها را به دو قسمت کاملاً جداگانه آموزشی و آزمایشی، تقسیم‌بندی کنیم، تا علاوه بر انجام آموزش، بتوانیم قابلیت تعمیم طبقه‌بند مورد نظر را برای فضای مشاهده نشده بسنجیم. تعداد نمونه‌های تولید شده از تصاویر پایگاه داده DDSM، ۲۵۰ نمونه می‌باشد که از این تعداد، ۸۰ نمونه به صورت رندوم به عنوان نمونه‌های تست در نظر گرفته شده است. نسبت داده‌های آموزشی به داده‌های آزمایشی حدود ۷۰ به ۳۰ است.

۵-۶ تاثیر ویژگی‌های مختلف در نتایج آزمایش

در مقاله‌های متعدد به منظور شناسایی توده از ویژگی‌های بافتی GLCM استفاده شده است [۴۶ و ۴۷ و ۵۴]. در این پایان‌نامه به منظور معرفی بهتر بافت توده، علاوه بر این ویژگی‌ها از ویژگی‌های بافتی GLRLM نیز استفاده شده است. این دو گروه ویژگی بافتی در فصل ۴ معرفی و شیوه محاسبه آنها بیان شده است.

در این قسمت قصد داریم تاثیر انتخاب هر کدام از این گروه ویژگی‌ها را در شناسایی توده مورد بررسی قرار دهیم. به این منظور از طبقه‌بند SVM با هسته داخلی RBF استفاده کرده‌ایم. در جدول-های (۲-۶)، (۳-۶) و (۴-۶) تاثیر هر کدام از این دو گروه ویژگی و ترکیب آنها در سیستم CAD پیشنهادی مورد بررسی قرار گرفته است.

جدول (۶-۲): نتیجه CAD پیشنهادی با استفاده از ویژگی‌های GLCM

کانفیوژن ماتریس		دقت	ویژگی
۹	۳	٪۹۰	ویژگی‌های GLCM
۵	۶۳		

جدول (۶-۳): نتیجه CAD پیشنهادی با استفاده از ویژگی‌های GLRLM

کانفیوژن ماتریس		دقت	ویژگی
۹	۳	٪۸۸.۷۵	ویژگی‌های GLRLM
۶	۶۲		

جدول (۶-۴): نتیجه CAD پیشنهادی با استفاده از ترکیب ویژگی‌های GLCM و GLRLM

کانفیوژن ماتریس		دقت	ویژگی
۱۰	۲	٪۹۲.۵	ترکیب ویژگی‌های GLCM و GLRLM
۴	۶۴		

همانگونه که در ۳ جدول بالا مشاهده می‌شود، سطح دقت ویژگی‌های GLCM در جداسازی نواحی سرطانی و نرمال از ویژگی‌های بافتی GLRLM بیشتر است. و همچنان سطح دقت ترکیب این ویژگی‌ها با هم از دقت هر کدام از دو گروه ویژگی بیشتر است. در واقع توانستیم با استفاده از ترکیب این دو گروه ویژگی با هم، ٪۲.۵ دقت تشخیص را در مقایسه با شرایطی که از ویژگی‌های GLCM به تنهایی استفاده شده است، بالا ببریم.

۱-۵-۶ ویژگی‌های مورد استفاده

در این قسمت ویژگی‌های مورد استفاده در این تحقیق را ذکر می‌کنیم.

- ویژگی‌های کنتراست، انرژی، همگنی و همبستگی که از ماتریس هم‌رخداد مربوط به ROIها با زاویه ۰، ۴۵، ۹۰ و ۱۳۵ درجه و فاصله یک استخراج شده‌اند.

- ویژگی‌های SRE، LRU، GLNU، RLN و RP که از ماتریس طول‌رخداد مربوط به ROIها با زاویه ۰، ۴۵، ۹۰ و ۱۳۵ درجه و فاصله یک استخراج شده‌اند.

در نتیجه جمعاً ویژگی‌های مورد استفاده ما در این تحقیق ۳۶ ویژگی بافتی شامل ۱۶ ویژگی استخراج شده از ماتریس هم‌رخداد و ۲۰ ویژگی استخراج شده از ماتریس طول‌رخداد است.

۶-۶ تاثیر طبقه‌بندهای مختلف در نتایج آزمایش

همانطور که در فصل ۵ نیز ذکر شد، عملیات دسته‌بندی در این تحقیق در ابتدا از طریق سه طبقه‌بند پایه SVM، KNN و MLP انجام می‌شود. پس از استخراج ویژگی، طبقه‌بندهای مختلف توسط مجموعه داده‌های آموزشی، آموزش داده شده و از مجموعه‌های آزمایشی به دست آمده برای تست آنها استفاده شده است. لازم به ذکر است که در آزمایش‌های صورت گرفته نمونه‌های آموزشی به صورت رندوم انتخاب شده‌اند و همچنین نمونه‌های آموزشی و آزمایشی کاملاً از هم متمایز هستند. در این قسمت پس از بیان جزئیات مربوط به اعمال هر روش، نتایج عملکرد حاصل از سیستم CAD شبیه‌سازی شده از طریق هر یک از این سه طبقه‌بند ارائه و در آخر با هم مقایسه شده است. لازم به ذکر است که بردار ویژگی استفاده شده برای هر سه طبقه‌بند یکسان است.

الف) نتایج آزمایش بر روی سیستم CAD مربوط به طبقه‌بند شبکه عصبی

مشخصات شبکه عصبی پیاده‌سازی شده در این تحقیق به شرح زیر است:

- نوع شبکه عصبی: شبکه عصبی سه لایه (MLP) پیش‌شو
- تعداد ورودی‌ها: ۳۶ برابر با طول بردار ویژگی
- تعداد نرون خروجی: ۱ (سرطانی=۱ و نرمال=۰)
- تعداد نرون‌های مخفی: ۲۴ (دو سوم مجموع نرون‌های ورودی و خروجی)
- تعداد نمونه‌های آموزشی: ۱۷۰ (تقریباً ۷۰ درصد)

- تعداد نمونه‌های آزمایشی: ۸۰ (تقریباً ۳۰ درصد)

- روش آموزش شبکه: روش پس انتشار خطا

نتایج عملکرد سیستم CAD پیشنهادی با متد طبقه‌بند MLP با ترکیب انتخاب شده برای آموزش و تست به صورت جدول (۵-۶) به دست آمده است.

جدول (۵-۶): نتایج آزمایش برای طبقه‌بند شبکه عصبی

کانفیوژن ماتریس		دقت	حساسیت	اختصاصیت
۹	۳	%۸۷.۵	%۷۵	%۸۹.۷
۷	۶۱			

ب) نتایج آزمایش بر روی سیستم CAD مربوط به طبقه‌بند K-NN

نتایج عملکرد سیستم CAD پیشنهادی با متد طبقه‌بند K-NN با مقدار بهینه شده تعداد همسایگی برابر ۶ و پیاده‌سازی بر روی داده‌های آزمایشی، به صورت جدول (۶-۶) به دست آمده است. در این مطالعه برای اندازه‌گیری فاصله داده آزمایشی از داده‌های آموزشی، از معیار شباهت فاصله اقلیدسی استفاده شده است که در بخش (۳-۵) معرفی شده است.

جدول (۶-۶): نتایج آزمایش برای طبقه‌بند 6-NN

کانفیوژن ماتریس		دقت	حساسیت	اختصاصیت
۸	۴	%۹۱.۲۵	%۶۶.۶۶	%۹۵.۵۸
۳	۶۵			

ج) نتایج آزمایش بر روی سیستم CAD مربوط به طبقه‌بند SVM

SVM را یک بار به صورت خطی و بار دیگر با هسته داخلی RBF (رابطه (۷-۵)) پیاده‌سازی کرده- ایم. نتایج عملکرد سیستم CAD پیشنهادی با متد طبقه‌بند SVM خطی و SVM با هسته داخلی RBF به ترتیب در جدول (۷-۶) و (۸-۶) آورده شده است. همانطور که مشخص است، طبقه‌بند

SVM با هسته داخلی RBF در جداسازی نواحی نرمال از سرطانی در مقایسه با SVM خطی عملکرد بهتری داشته است.

جدول (۶-۷): نتایج آزمایش برای طبقه‌بند SVM خطی

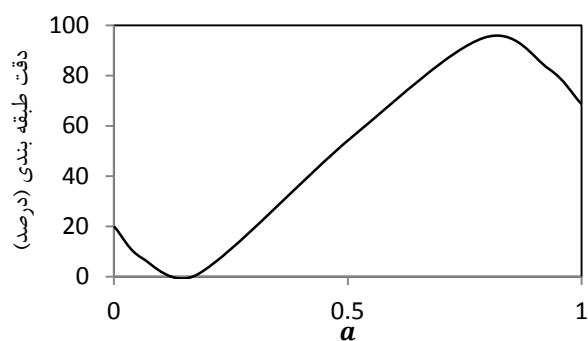
کانفیوژن ماتریس		دقت	حساسیت	اختصاصیت
۷	۵	%۸۷.۵	%۵۸.۳۳	%۹۲.۶۴
۵	۶۳			

جدول (۶-۸): نتایج آزمایش برای طبقه‌بند SVM با هسته داخلی RBF

کانفیوژن ماتریس		دقت	حساسیت	اختصاصیت
۱۰	۲	%۹۲.۵	%۸۳.۳۳	%۹۴.۱۱
۴	۶۴			

د) نتایج ترکیب طبقه‌بندها با الگوریتم OWA

با بردارهای ویژگی یکسان هر یک از سه طبقه‌بند پایه به طور جداگانه آموزش دیدند. حال پس از آموزش، در ازای یک داده آزمایشی یکسان هر کدام از آنها پاسخ خود را مربوط به کلاس آن داده تست مورد نظر دارند. در بعضی موارد پاسخ هر سه طبقه‌بند با هم یکسان و در بعضی موارد دیگر با هم متفاوت است. حال ما در این قسمت توسط متد OWA که در فصل ۵ به تفصیل توضیح داده شده است، اطلاعات حاصل از هر یک از سه طبقه‌بند را برای یک ورودی یکسان با هم ترکیب می‌کنیم. ضرایب w در الگوریتم OWA بر اساس راهکار خوشبینانه تعیین شده است، رابطه (۵-۱۴)، و برای دست یافتن به مقادیر بهینه این ضرایب، پارامتر a را از صفر تا ۱ با گام ۰.۰۱ تغییر داده‌ایم. طبق شکل (۶-۱۲) بیشینه دقت طبقه‌بندی به ازاء $a = 0.79$ حاصل شده است. بنابراین ضرایب عملگر OWA بر اساس این مقدار α محاسبه می‌گردد.



شکل (۶-۲): بررسی عملکرد روش OWA به ازاء مقادیر مختلف a

نتایج عملکرد سیستم CAD پیشنهادی با استفاده از الگوریتم OWA با رویکرد خوشبینانه بر روی خروجی حاصل از طبقه‌بندی‌های پایه، به صورت جدول (۶-۹) به دست آمده است.

جدول (۶-۹): نتایج آزمایش برای ترکیب طبقه‌بندی‌های پایه

کانفیوژن ماتریس		دقت	حساسیت	اختصاصیت
۱۱	۱	%۹۶.۲	%۹۱.۶۶	%۹۷.۰۵
۲	۶۶			

۶-۷ جمع بندی

به منظور مقایسه عملکرد طبقه‌بندی‌های مختلف، همه نتایج بدست آمده را در جدول (۶-۱۰) جمع کرده‌ایم. همانگونه که در این جدول مشاهده می‌شود، نتایج عملکرد حاصل از الگوریتم OWA از ۳ طبقه‌بندی دیگر بهتر است. و هر ۳ معیار ارزیابی مورد بررسی، برای این الگوریتم دارای بالاترین مقدار است. مقایسه‌ای از عملکرد سیستم CAD پیشنهادی، در مقایسه با روش‌های دیگر شناسایی توده نیز در جدول (۶-۱۱) آمده است.

جدول (۶-۱۰): نتایج معیارهای ارزیابی برای ۴ سیستم CAD مورد بررسی در این تحقیق

classifier	دقت	حساسیت	اختصاصیت
Neural Network	%۸۷.۵	%۷۵	%۸۹.۷
K-NN	%۹۱.۲۵	%۶۶.۶۶	%۹۵.۵۸
SVM	%۹۲.۵	%۸۳.۳۳	%۹۴.۱۱
OWA	%۹۶.۲	%۹۱.۶۶	%۹۷.۰۵

جدول (۶-۱۱): مقایسه روش پیشنهادی با روش‌های دیگر

	پایگاه داده	ویژگی‌ها	طبقه‌بند	دقت	حساسیت	اختصاصیت
متد پیشنهادی	DDSM	GLCM & GLRLM	OWA	%۹۶.۲	%۹۱.۶۶	%۹۷.۰۵
مرجع [۴۷]	MIAS	GLCM	Decision Tree	%۸۵	%۷۰	%۱۰۰
مرجع [۵۳]	MIAS	GLCM	RBF NN	%۸۷	%۹۴	%۹۰
مرجع [۴۶]	DDSM	GLCM	MLP NN	%۸۷.۹۱	%۹۱.۶۷	%۸۴.۱۷

فصل ہفتم :

نتیجہ گیری و رائے پیشہادات

۱-۷ نتیجه‌گیری

هدف ما در این پایان نامه مطالعه سیستم‌های CAD جهت تشخیص توده‌ها در بافت‌های سینه بود. پس از بررسی روش‌های متداول در طراحی سیستم‌های CAD و روش‌های موجود در تشخیص توده‌ها در تصاویر ماموگرافی، به طراحی و شبیه‌سازی یک سیستم CAD پرداختیم. سیستم CAD پیشنهادی شامل ۴ مرحله اصلی بود.

در روش پیشنهادی ما ابتدا در دو مرحله پیش‌پردازش و استخراج ناحیه‌های مشکوک، ناحیه‌های اضافه بدون از دست رفتن اطلاعات مفید، از حوزه جستجو برای توده‌ها حذف شدند. به دست آوردن ناحیه‌های مشکوک علی‌رغم کاهش زمان اجرا، باعث کاهش نرخ حساسیت سیستم نیز شد. لازم به ذکر است که کلیه اقدامات انجام شده در این دو مرحله به صورت خودکار صورت گرفت. خودکار بودن کلیه مراحل شناسایی و در نتیجه عدم دخالت انسان در شناسایی بیماری، سبب کاربردی شدن روش ما در سیستم‌های تجاری شد.

استفاده همزمان از دو ماتریس GLCM و GLRLM در مرحله استخراج ویژگی توانست ترکیب مناسب و موثری از ویژگی‌ها را به وجود آورد. ماتریس GLCM به تنهایی در بسیاری از مقالات مرتبط مورد استفاده قرار گرفته بود [۵۴،۴۷،۴۶]. پس از استفاده همزمان از دو ماتریس GLCM و GLRLM در روش پیشنهادی خود توانستیم دقت را از ۹۰ به ۹۲.۵ درصد ارتقا دهیم. دلیل این امر می‌تواند اطلاعات بیشتری باشد که داده‌های بدست‌آمده از این روش برای طبقه‌بندها در مرحله آموزش فراهم می‌کند. استفاده همزمان از این ویژگی‌ها، در حالی دقت را افزایش می‌دهد که سرعت دسته‌بندی کاهش می‌یابد.

در آخرین مرحله از روش پیشنهادی خود به منظور طبقه‌بندی ROIها، از سه طبقه‌بند ساده شبکه عصبی مصنوعی MLP، K-NN و SVM استفاده شد. نتایج تست این ۳ طبقه‌بند در تشخیص توده‌ها با اعمال بر روی تصاویر ماموگرافی مورد بررسی قرار گرفت. لازم به ذکر است که هر ۳ طبقه‌بند با

مجموعه‌های یکسانی از داده‌های آموزش و تست و همچنان شرایط برابر از نظر نوع بردارهای ویژگی ورودی مورد بررسی قرار گرفتند. دقت حاصل از آزمایش این سه طبقه‌بند، برحسب درصد، به ترتیب برابر با ۸۷.۵، ۹۱.۲ و ۹۲.۵ است. از میان این سه طبقه‌بند، طبقه‌بند SVM نتایج دقیق‌تری را نسبت به دو طبقه‌بند دیگر ارائه داد.

در انتها به منظور افزایش دقت دسته‌بندی، از راهکار ترکیب اطلاعات در نتایج طبقه‌بندهای پایه استفاده کردیم. برای این منظور روش OWA به کار گرفته شد. به کمک این روش دقت نتایج دسته‌بندی تا ۹۶.۲ درصد افزایش یافت. دقت نتایج حاصل از این روش نسبت به بهترین درصد در طبقه‌بندهای پایه حدود ۴ درصد افزایش داشت.

با استفاده از OWA علاوه بر دقت، دو معیار حساسیت و اختصاصیت که دو معیار بسیار مهم در سیستم‌های پزشکی محسوب می‌شود نیز به مقدار قابل قبولی افزایش یافت. (حساسیت ۹۱.۶۶٪ و اختصاصیت ۹۷.۰۵ درصد). در راهکارهای ترکیب اطلاعات، اعلام نظر نهایی در مورد کلاس هر ورودی با تصمیم‌گیری براساس نتایج چند طبقه‌بند مختلف انجام می‌شود. در نتیجه طبقه‌بندها می‌توانند نقاط ضعف همدیگر را پوشش دهند. این امر را می‌توان دلیل برتری این روش دانست.

همچنین لازم به ذکر است که عملکرد روش پیشنهادی در مقایسه با روش‌های دیگر شناسایی توده [۴۷ و ۴۸ و ۵۳] بسیار قابل قبول بوده است.

۲-۷ پیشنهاداتی برای ادامه تحقیق

- ✓ به منظور ترکیب نتایج حاصل از طبقه‌بندهای پایه می‌توان از روش‌های دیگری به جز OWA به عنوان مثال روش انتگرال‌های فازی (Sugeno and Choquest fuzzy integral) استفاده کرد.
- ✓ پس از مرحله آخر روش پیشنهادی، زمانیکه یک ROI به عنوان توده شناسایی شد، می‌توان از ویژگی‌های شکلی و مرزی به منظور تعیین خوش‌خیمی و بدخیمی توده استفاده کرد.
- ✓ در این پایان‌نامه برای هر ROI چهار ماتریس هم‌رخداد در ۴ جهت (۰ و ۴۵ و ۹۰ و ۱۳۵ درجه) ساخته شد. به نظر می‌رسد ماتریس پنجم به صورت میانگین ۴ ماتریس دیگر به تنهایی و یا در کنار چهار ماتریس دیگر می‌تواند نرخ شناسایی توده‌ها را افزایش دهد که نیاز به بررسی و مطالعه دارد.

مراجع

[¹] R. Siegel, D. Naishadham, and A. Jemal, (2012), “cancer statistics, 2012,” CA: A Cancer Journal for Clinicians, vol. 62, no. 1, pp. 10-29.

[²] www. Wikipedia.com

[3] Harirchi I, Karbakhsh M, Kashefi A, Momtahan AJ, (2004), ” Breast Cancer in Iran: result of a multicenter study”, Asian Pac j Cancer Prev; 5(1), pp. 24-70.

[4] Jinshan Tang, IEEE, Rangara M, Jun Xu, Issam EI Naqa, Yongyi Yong, (2009), “ Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances” , IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 13, NO. 2.

[5] K. H. Ng and M. Muttarak, (2003) “Advances in mammography have improved early detection of breast cancer ” J. Hong Kong College Radiol, vol. 6, no. 3, pp. 126–131.

[6] Baines CJ, McFarlane DV, Miller AB, (1990) “ The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the national breast screening study “, Invest Radiol, 25(9), pp. 971-976.

[7] Wallis MG, Walsh MT, Lee JR, (1991) “A review of false negative mammography in a symptomatic population”, Clin Radiol, 44(1), pp. 5-13.

[8] Saskia van England, (2006), PHD. Thesis, “Detection Of mass lesions in mammograms by using multiple view”.

[9] (2005). [Online].Available: <http://www.cancer.gov/cancertopics/factsheet/DMISTQ> and A

[10] W. Yang, (2006), “Digital mammography update” Biomed. Imag. Intervention J., vol. 2, no. 4, pp. 45–12.

[11] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. Baum, S. Acharyya, E. Conant, L. Fajardo, L. Bassett, C. D’Orsi, R. Jong, and M. Rebner, (2005), “Diagnostic performance of digital versus film mammography for breast cancer screening,” New England J. Med., vol. 353, no. 17, pp. 1773– 1783.

- [12] D. Spurgeon, (2005), "Digital mammography is more accurate only for certain groups of women," Br. Med. J., vol. 331, no. 7518, pp. 653–653.
- [13] M. Del, P. Mantellini, S. Ciatto, R. Bonardi, F. Martinelli, B. Lazzari, and N. Houssami, (2007), "Full-field digital versus screen-film mammography: Comparative accuracy in concurrent screening cohorts", Amer. J. Roentgenol., vol. 189, no. 4, pp. 860–866.
- [14] Yao Yao , (2004), " segmentation of breast cancer mass in mammograms and detection using magnetic resonance imaging", IEEE image processing Soc. J, pp: 561-567.
- [15] <http://www.radiologyinfo.org/en/info.cfm?pg=mammo>
- [۱۶] باقری خلیج، علی، ۱۳۹۱، پایان نامه ارشد، "دسته‌بندی ضایعات در تصاویر ماموگرافی"، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف.
- [17] D. B. Kopans, (2007), "Breast Imaging", 3rd ed. Baltimore, MD: Williams & Wilkins.
- [18] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, (2003), "computer-aided detection and classification of microcalcifications in mammograms: a survey", Pattern Recognition, vol. 36, no.12, pp. 2967-2991.
- [19] American College of Radiology (ACR), Illustrated Breast Imaging Reporting and Data System (BI-RADS), 3rd ed. Reston, VA: Amer. Coll. Radiol , 1998.
- [20] B. C. Yankaskas, M. J. Schell, R. E. Bird, and D. A. Desrochers, (2001), "Reassessment of breast cancers missed during routine screening mammography: A community based study," Amer. J. Roentgenol., vol. 177, no. 3, pp. 535–541.
- [21] H. Burrell, A. Evans, A. Wilson, and S. Pinder, (2001), "False-negative breast screening assessment: What lessons we can learn?," Clin. Radiol., vol. 56, no. 5, pp. 385–388.
- [22] R.M. Rangayyan, F.J. Ayres, J.E.L. Desautels, (2007), "A Review of Computer-Aided Diagnosis of Breast Cancer: Toward the Detection of Subtle Signs", Journal of the Franklin Institute, Vol. 344, Issues 3-4, pp. 312-348.

- [23] Muthukarthigadevi. R, Anand. S, (2013), "Detection of architectural distortion in mammogram image using wavelet transform ", International Conference on Information Communication and Embedded Systems (ICICES), pp. 638-643.
- [24] Daniel Rodrigues Ericeira, Aristófanés Corrêa Silva, (2013), "Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions", VOL. 43, pp. 987-999.
- [25] R.Bird, T.Wallace, and B.Yankaskas, (1992), " Analysis of cancer missed at screening mammography" Radiology, vol.184, no.3 , pp.613-617.
- [26] K. Kerlikowske, P. Carney, B. Geller, M. Mandelson, S. Taplin, K. Malvin, V. Ernster, N. Urban, G. Cutter, R. Rosenberg, and R. Ballard-Barbash, (2000), " performance of screening mammography among woman with and without a first-degree relative with breast cancer", Ann.Internal Med., vol. 133 , no. 11, pp. 855-863.
- [27] J. Brown, S. Bryan , R. Warren, (1996), " Mammography screening : an incremental cost effectiveness analysis of double versus single reading of mammograms", VOL. 312, no. 7034, pp. 809-812.
- [28] R. Warren and S.Duffy , (1995) " compression of single and double reading of mammograms, and change in effectiveness with experience" Br.J.Radiol., vol. 68, no. 813, pp. 958-692.
- [29] M. L. Giger, (2000), "computer-aided diagnosis of breast lesion in medical image ", Comput. Sci.Eng., vol. 2, no. 5, pp. 39-45
- [30] K. Doi,(1996) "Computer-aided diagnosis : potential usefulness in diagnostic radiology and telemedicine", Proceedings of National Forum, pp.9–13 .
- [31] N. Karssemeijer and J. H. Hendriks, (1997), " computer-assisted reading of mammograms" Eur. Radiol., vol. 7, no. 5, pp. 743-748.
- [32] M.Morton , D.Whaley , K.Brandt and K.Amrami , (2006) "Screening mammograms: interpretation with computer-aided detection--prospective evaluation" Radiology, vol. 239, no. 2, pp. 375-383.

- [33] R. Brem, J. Baum, M. Lechner, S. Kaplan, S. Souders, L. Naul, and J. Hoffmeister, (2003), “ improvement in sensitivity of screening mammography with computer-aided detection : a multi institutional trial” *Amer. J. Roentgenol.*, vol. 181, no 3, pp. 687-693.
- [34] Jelen Bozek, Kresimir Delac, Mislav Grgic, (2008), “ Computer –Aided Detection and Diagnosis of Breast Abnormalities in Digital Mammography” ,50th International Symposium ELMAR, 10-12, Zadar, Croatia, pp. 2-3.
- [35] L.M. Bruce, R.R. Adhmai, (1999), “Classifying Mammographic Mass Shapes Using the Wavelet Transform Modulus-Maxima Method” , *IEEE Transactions on Medical Imaging*, vol. 18, Issue 12, pp. 1170-1177.
- [36] H.D. Cheng, X.P. Cai, X.W. Chen, L.M. Hu, X.L. Lou, (2003), “Computer-aided detection and classification of microcalcifications in mammograms: a survey”, *Pattern Recognition* , pp. 2967–2991.
- [37] R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, (1995) “computer-aided detection of clustered microcalcifications on digital mammograms” *Med. Biol. Eng. Comput.*, vol. 33, no. 2, pp. 174-178.
- [38] M. N. Gurcan, Y. Yardimci, A. E. Cetin, and R. Ansari, (1997), “ detection of microcalcifications in mammograms using higher order statistics”, *IEEE Signal Process. Lett.*, vol. 4, no. 8, pp. 213-216.
- [39] R. N. Strickland and H. Hahn,(1996), “ Wavelet transform for detection microcalcifications in mammograms ”, *IEEE Trans. Med. Image.*, vol. 15, no.2, pp. 218-229.
- [40] G. Lemaur, K. Drouiche, and J. DeConinck, (2003), “ Highly regular wavelets for the detection of clustered microcalcifications in mammograms”, *IEEE Trans. Med. Image.*, vol. 22, no. 3, pp. 393-401.
- [41] S. Yu and L. Guan, (2000), “ A CAD system for the outomatic detection of clustered microcalcifications in digitized mammogram films: *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 115-126.

- [42] M. P. Sampat, M. K. Markey, and A. C. Bovik, (2005), “ Computer-aided detection and diagnosis in mammography” in Handbook of Image and Video Processing, A.C. Bovik, Ed., 2nd ed. New York: Academic, pp. 1195-1217.
- [43] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark,(1995), “ Markov random field for tumor detection in digital mammography” IEEE Trans. Med. Image., vol. 14, no. 3, pp.565-576.
- [44] W. P. Kegelmeyer, Jr., J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper,(1994) “Computer-aided mammographic screening for speculate lesion ” Radiology, vol. 191, no. 2, pp. 331-337.
- [45] S. Liu, C. F. Babba, and E. J. Delp, (2001), “ Multiresolution detection of speculated lesions in digital mammograms” IEEE Trans.Image Process., vol. 10, no. 6, pp. 874-884.
- [46] Al Mutaz M. Abdolla, Safaai Dress, Nazar Zaki, (2011), “Detection of Masses in Digital Mammogram Using Second Order Statistics and Artificial Neural Network”, International Journal of Computer Science & Information Technology (IJCSIT), Vol.3 No.3
- [47] A Mohd. Khuzi, R Besar, NN Ahmad, WMD Wan Zaki, (2009), " Identification of masses in digital mammogram using gray level co-occurrence matrices", Biomedical Imaging and Intervention Journal.
- [48] Nima Sahba, Vahid Tavakoli, Alireza Ahmadian, Masoumeh Giti, (2009), “mammography mass detection: a multi-stage hybrid approach”.
- [49] F. Fauchi, G. Raso, Rl. Magro, (2005), “A Massive Lesion Detection Algorithm in Mammography”, Physica Medica, Vol. 21, No. 1, pp. 21-28.
- [50] Nevin H.Eltonsy, Georgia D. Tourassi, Adel S.Elmaghraby, (2007), “A Concentric Morphology Model for the Detection of Masses in Mammography”, IEEE TRANSACTION ON MEDICAL IMAGING, VOL. 26, NO. 06, P. 880-889.
- [51] I. Christoyianni, E. Dermatas, G. Kokkinakis, (2000), “Fast detection of Masses in computer- aided mammography”, IEEE Signal Process.Mag.17(1), pp. 54–64.
- [52] Ali Bagheri-Khaligh, Ali Zarghami, and M. T. Manzuri-Shalmani, (2012), “Novel Margin Features for Mammographic Mass Classification”.

- [53] M. Jirari, (2005), “ A computer aided detection system for digital mammograms based On radial basis functions and feature extraction techniques”, 27th International Conference of the IEEE Engineeringin Medicine and BiOlogy Society EMBS, 6 pages, Shanghai, China.
- [54] R. Nithya, B. Santh, (2011), "Classification of Normal and Abnormal Patterns in Digital Mammograms for Diagnosis of Breast Cancer", International Journal of Computer Applications, VOL 28, No.6, PP. 21-25.
- [55] B. Zheng, Y.H. Chang, X.H. Wang, W.F. Good, D. Gur, (1999), “Application Of a Bayesian belief network in a computer-assisted diagnosis Scheme for mass detection, SPIE Conference on Image Processing, 3661(2), 1553–1561.
- [56] B. Zheng, Y. H. Chang, X. H. Wang, W. F. Good, D. Gur, (1999), “Application Of a Bayesian belief network in a computer-assisted diagnosis Scheme for mass detection, SPIE Conference on Image Processing, 3661(2), 1553–1561.
- [57] V. Gimenez, D. Manrique, J. Rios, A. Vilarrasa, (1999), “Iterative method for Automatic detection of masses in digital mammograms for computer Aided diagnosis”, Proceedings of SPIE—The International Society For Optical Engineering 3661, vol.II , pp.1086–1093.
- [58] T. Matsubara, H. Fujita, S. Kasai, M. Goto, Y. Tani, T. Hara, T. Endo, (1997), “Development of new schemes for detection and analysis of Mammographic masses”, Intel. Inform. Systems, 63–66
- [59] H.D. Li, M. Kallergi, L.P. Clarke, V.K. Jain, R.A. Clark, (1995), “Markov Random field for tumor detection in digital mammography”, IEEE Trans. Med. Imaging14(3), 565–576.
- [60] M. Kallergi, K. Woods, L.P. Clarke, W. Qian, R.A. Clark, (1992) , “Image segmentation in digital mammography: comparison of local thresholding and region growing algorithm" , Comput. Med. Imaging Graph. 16 (5), 231–323
- [61] Y.J. Lee, J.M. Park, H.W. Park, (2000), “Mammographic mass detection by adaptive thresholding and region growing”, Int. J. Imaging Systems Technol. 11 (5), 340–346.

- [62] S.M. Lai, X. Li, W.F. Biscof, (1989), "On techniques for detecting Circumscribed masses in mammograms", IEEE Trans. Med. Imaging 18(4), 377–386.
- [63] G.D. Tourassi, Rene Vargas-Voracek , (2003), " Computer-assisted detection Of mammographic masses: a template matching scheme based on Mutual information", Med. Phys. 30, 2123–2130
- [64] A.S. Constantinidis, M.C. Fairhurst, F. Deravi, M. Hanson, C.P. Wells, C. Chapman-Jones, (1999), "Evaluating classification strategies for Detection of circumscribed masses in digital mammograms", Proceedings of 7th International Conference on Image Processing And its Applications, pp.435–439.
- [65] M. Sameti, R.K. Ward, (1996) "A fuzzy segmentation algorithm for Mammogram partitioning", Digital Mammography, Elsevier, Amsterdam , pp.471–474.
- [66] M. Sameti, R.K. Ward, B. Palcic, J. Morgan-Parkes, (1997), " Texture feature Extraction for tumor detection in mammographic images", IEEE PacificRim Conference on Comm.,Computers and Signal Processing, 1997, pp.831–834.
- [67] M. Sameti, R.K. Ward, J.Morgan-Parkes, B. Palcic, (2000), " A method For detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm", Proceedings of the 19th International Conference IEEE E/MBS , pp.513–516.
- [68] D. Guliato, R.M. Ranayyan, W.A. Carnielli, J.A. Zuffo, J.E.L. Desautels, (1998), "Detection of breast tumors in mammograms by fuzzy region growing", The IEEE Proceedings of the Annual International Conference on Engineering in Medicine and Biology Society, vol. 2, pp. 1002–1005.
- [69] A.J. Mendez, P.G. Tahoces, M.J. Lado, M. Souto, J.J. Vidal, (1998), "Computer-aided diagnosis : automatic detection of malignant masses In digitized mammograms", Med. Phys. 25(6), 957–964.
- [70] A. Hadjarian, J. Bala, S.A. Gutta, S. Trachiotis, P. Pachowicz, (1998), "The Fusion of supervised and unsupervised techniques for segmentation Of abnormal regions ", N. Karssemeijer, M. Thijssen, J. Hendriks, L. Erning(Eds.), Digital Mammography, Kluwer, Amsterdam, pp.299–302.

- [71] S. Meenalosini, J. Janet, E. Kannan, (2012), "A Novel Approach in Malignancy Detection of Computer Aided Diagnosis", American Journal of Applied Sciences, Vol. 9, pp. 1020-1029.
- [72] H. D. Cheng, X. J. Shi, (2004), "A simple and effective histogram equalization Approach to image enhancement", Digital Signal Process. 14(2) 158–170.
- [73] H. Li, K. Liu, Y. Wang, S.C. Lo, (1997), "Nonlinear filtering enhancement And histogram modeling segmentation of masses for digital Mammograms", The IEEE 18th Annual International Conference on Bridging Disciplines for BioMedicine, vol.3, pp.1045–1046.
- [74] J.K Kim, J.M. Park, K.S. Song, H.W. Park, (1997), "Adaptive Mammographic image enhancement using first derivative and local statistics ", IEEE Trans. Med. Imaging 16(5), pp. 495–502.
- [75] K. Woods, L.P. Clarke, R. Welthuizen, (1991), "Enhancement of digitized Mammograms using a local thresholding technique", Annual International Conference of the IEEE Engineering in Medicine and Biology Society13, vol.1, pp.114–115.
- [76] S. Singh, R. Al-Mansoori, (2000), "Identification of regions of interest in Digital mammograms", J. Intell. Systems10 (2), 183–210.
- [77] E. D. Pisano, S. Zong, B. M.Hemminger, M. DeLuca, R.E. Johnston, K. Muller, M.P. Braeuning, S.M. Pizer, (1998), "Contrast limited adaptive Histogram equalization image processing to improve the detection Of simulated spiculations in dense mammograms" , J. Digital Imaging 11(4), 193–200.
- [78] D. Braccialarghw, G.H. Kaufmann, (1996) "Contrast enhancement of Mammographic features: a comparison of four methods ",Opt. Eng. 35(1)76–80.
- [79] J. Lu, D.M. Healy, J.B. Weaver, (1994), "Contrast enhancement of medical Images using multi scale edge representation", Opt. Eng. 33(7), 2151–2161.
- [80] A. Laine, J. Fan, W. Yang, (1995), "Wavelets for contrast enhancement Of digital mammography" ,IEEE Eng. Med. Biol. 14(5) , 536–550.

- [81] F. Fauchi, G. Raso, Rl. Magro; (2005), “A Massive Lesion Detection Algorithm in Mammography”, *Physica Medica*, Vol. 21, No. 1, pp. 21-28.
- [82] F.F. Yin, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, (1994), “ Computerized detection of masses in digital mammograms”, *Investigation of feature-analysis techniques* , *J. Digital Imaging* , 18–26.
- [83] H. Li, Y. Wang, K.J.R. Liu, S.B. Lo, M.T. Freedman, (2001), “ Computerized Radiographic mass detection- part II: lesion site selection by Morphological enhancement and contextual segmentation”, *IEEE Trans. Med. Imaging*20(4), 302–313
- [84] K. Bovis, S. Singh, J. Fieldsend, C. Pinder, (2000), “Identification of masses In digital mammograms with MLP and RBF nets”, *Proceedings Of the IEEE-INNS-ENNS International Joint Conference on Neural Networks Com*, pp.342–347.
- [85] Z. Huo, M.L. Giger, C.J. Vyborny, F.I. Olopade, D.E. Wolverton, (1998), "Computer-aided diagnosis: analysis of mammographic parenchyma Patterns and classification of masses on digitized mammograms", *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, No. 2, pp.1017–1020.
- [86] B. Sahiner, N. Petrick, H.P. Chan, (2001), “Computer-aided characterization Of mammographic masses: accuracy of mass segmentation and its Effects on characterization“, *IEEE Trans. Med. Imaging*20 (12), 1275–1284.
- [87] B. Zheng, Y.H. Chang, X.H. Wang, W.F. Good, (1999), “Comparison of Artificial neural network and Bayesian belief network in a computer- Assisted diagnosis scheme for mammography”, *IEEE International Conference on Neural Networks*, pp.4181–4185.
- [88] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du, (2005) “Approaches for automated detection and classification of masses in mammograms”, *Pattern Recognition*, Elsevier, Vol 39, pp. 646–668.
- [89] Anupa Maria Sabu, D.Narain Ponraj, Poongodi, (2012), “Textural Features Based Breast Cancer Detection: A Survey”, *Journal of trends in computing and information science*, VOL. 3, NO. 9.

[90] R.M. Haralick, K. Shanmugan, I. Dinstein, (1973), "Textural features for image classification", IEEE Trans. SMC 3 ,pp. 610–621.

[91] S. Theodoridis, K. Koutroumbas, (2003), "Pattern Recognition", Academic Press; 4 edition.

[92] S. Lindsay, (2002), "A tutorial on Principal Components Analysis".

[93] Boser, B. E., Guyon, I. M., Vapnik, V., (1992), "A Training Algorithm for Optimal Margin Classifiers", Fifth Annual Workshop on Computational Learning Theory, ACM, (1992).

[94] C. J. Burges, (1998), "A tutorial on support vector machines for pattern recognition ", Knowledge Discovery and Data Mining, vol. 2, pp. 121–167 .

[95] Hwanjo Yu , Sungchul Kim, (2009), "SVM Tutorial: Classification, Regression, and Ranking", *Handbook of Natural Computing* .

[۹۶] اسمعیلی، معصومه، ۱۳۹۱، پایان نامه ارشد، "شناسایی حرکات ارادی انسان با استفاده از سیگنال‌های EEG"، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود .

[97] Tulyakov S., Jaeger S., Govindaraju V., Doermann D., (2008), "Review of Classifier Combination Methods", Machine Learning in Document Analysis and Recognition Studies in Computational Intelligence Vol. 90, pp. 361-386.

[98] Yager R. R., (1988), " On ordered weighted averaging aggregation operators in multi criteria decision making", Man and Cybernetics, Vol. 18, pp. 183-190.

[99] Yager, R. R., & Filev, D. P. (1999), "Induced ordered weighted averaging operators". Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 29(2), 141-150.

Abstract

Breast cancer is the most common cancer and the second leading cause of cancer deaths among women. Studies show that it is not nearly possible to prevent of breast cancer since its causes are not known. So, the early detection of the stage of cancer plays the key role in treatment of this cancer. The early detection and treatment of breast cancer increase the life expectancy and stimulate the patient for next visit. Now, mammography is the most common way of early detection of this cancer and the death rate has been reduced by 25%.

Computer-Aided Diagnosis (CAD) systems are widely used for detection of various kinds of abnormalities in mammography images. Masses and Microcalcifications are two type important of these abnormalities in mammography images. This study aim to plan, simulate, investigate and compare CAD systems ableing automatically determine the masses im mammography images with high accuracy. In addition to the mass, the application of CAD in MCs detection are considered in short term.

Our proposed method in this thesis consists of four main steps. These steps include preprocessing, extraction of the suspected areas of mass, feature extraction and classification. All steps will be discussed in detail in different chapters. In preprocessing step, some solutions for image enhancement of mammography and deletion of extra parts such as labels and pectoral muscle are present. In second step, suspicious areas are extracted from mammography image by histogram analysis. Then, Different types of intensity features, texture features and shape features are introduced. And samples of them are expressed. effective combination of texture features resulting from GLCM and GLRLM matrix will be suggested and it will result in high accuracy segregation . in the last stage to classification of suspicious areas, three basic classifier MLP, K-NN and SVM are used. We evaluate the results of these classifiers using DDSM database mammography image. Finally In order to increase the accuracy of classification, we used combination of basic classifiers output. OWA method is used for this purpose.

The reported results of this method show that the accuracy of classification is increased by 4% compared with the best result in basic classifiers. Also, simultaneous use of GLCM and GLRLM features shows 2% accuracy enhancement compare with the time when each of them is used alone.

Key-Words: breast cancer, mammography, Computer-Aide Detection (CAD), Region Of Interested (ROI), Region Growing, Gray Level Co-occurrence Matrix(GLCM), Gray Level Run-Length Matrix (GLRLM), combination of classifiers, OWA



Shahrood University of Technology

Faculty of Computer and IT Engineering

M. Sc, Thesis

Auotomatic Mass Detection in Mammography images

Reyhaneh Amini

Supervisor:

Dr. Morteza Zahedi

Sep 2013