







دانشکده مهندسی برق و رباتیک

گروه الکترونیک

پایان نامه کارشناسی ارشد

استفاده از روش‌های بهینه‌سازی ازدحام ذرات برای انتخاب ویژگی‌های بهینه جهت

تشخیص میکروکلسیفیکیشن‌ها در تصاویر ماموگرافی

مهدی اکبری

استاد راهنما

دکتر سید علی سلیمانی

استاد مشاور

دکتر حسین خسروی

بهمن ماه ۱۳۹۳



تقدیم ہے

پدر و مادر مہربانم کہ زندگیم را دیدیون مهر و عطفوت آن نامی دانم

بمسرورم کہ نشان لطف الہی در زندگی من است

و خواہرانم کہ ہر اثنان ہمیشگی و پشتوانہ نامی زندگیم ہستند

## تقدیر و تشکر

پاس بی کران پروردگار یکتا را که بستی مان بخشید و به طریق علم و دانش را بنمونان شد و به بهنیشینی رحروان علم و دانش مستخرمان نمود و نموده چینی از علم و معرفت را روزی ان ساخت.

از استاد با کالات و شایسته جناب آقای دکتر علی سلیمانی ایوری که در کمال سعه صدر، با حسن خلق و فروتنی، از بیج گلی در این عرصه بر من دریغ نمودند و زحمت را بهمانی این رساله را بر عهده گرفتند؛ کمال تشکر و قدر دانی را دارم. همچنین از رهنمون های خردمندانه جناب آقای دکتر حسین خسروی نیز تشکر می کنم.

در انتها از اساتید محترم گروه الکترونیک دانشگاه صنعتی شاهرود، دوستان و هم کلاسی هایم و کسانی که در زندگی از سواد و خرد آن بهره گرفتند، تشکر می کنم.

## چکیده

سرطان سینه شایع‌ترین نوع سرطان و دومین عامل مرگ‌ومیر ناشی از سرطان در میان بانوان می‌باشد. در حال حاضر ماموگرافی یکی از بهترین روش‌های تشخیص زود هنگام سرطان سینه است. بیشترین ناهنجاری‌ها که ممکن است در سرطان سینه آشکار شوند، توده‌ها و میکروکلسیفیکیشن‌ها (MCها) می‌باشند. ناهنجاری‌های سینه با دامنه وسیعی از ویژگی‌ها معرفی می‌شود. مطالعات نشان می‌دهد ۱۵-۳۰ درصد سرطان‌های سینه، ممکن است به آسانی توسط رادیولوژیست‌هایی که تصاویر را می‌خوانند، نادیده گرفته شود. با پیشرفت‌هایی که در زمینه پردازش تصاویر دیجیتال به وقوع پیوسته است رادیولوژیست‌ها این فرصت را یافته‌اند که این درصد خطا را کاهش دهند. برای کمک به رادیولوژیست‌ها جهت تشخیص دقیق بیماری سامانه‌های شناسایی پزشکی (CAD) توسعه یافته‌اند.

یک سیستم تشخیص در حالت کلی شامل مراحل پیش‌پردازش، استخراج ویژگی و کلاسه‌بند است. در این پایان‌نامه دو روش جهت تشخیص MCها از تصاویر ماموگرافی پیشنهاد شده و کارایی این روش‌ها در مقایسه با روش‌های موجود ارزیابی شده است.

در هر دو روش پیشنهادی، پس از پیش‌پردازش تصاویر ماموگرافی، با استخراج ویژگی‌های بافتی حاصل از ماتریس هم‌رخداد (GLCM)، اقدام به دسته‌بندی افراد با طبقه‌بندهای ماشین بردار پشتیبان (SVM) و شبکه عصبی می‌نماییم. در نهایت در روش پیشنهادی دوم، با انتخاب ویژگی‌های بهینه با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات (PSO) و کلاسه‌بند ماشین بردار پشتیبان، بهترین نتیجه را در تشخیص MCها در تصاویر ماموگرافی بدست می‌آوریم.

**کلید واژه:** سرطان سینه، ماموگرافی، توده، میکروکلسیفیکیشن، ماتریس هم‌رخداد، کلاسه‌بند

ماشین بردار پشتیبان، کلاسه‌بند شبکه عصبی

## فهرست مطالب

۲	فصل اول
۲	۱-۱ سرطان سینه
۲	۲-۱ شیوع سرطان سینه
۴	۱-۳ علائم سرطان سینه
۵	۴-۱ علل سرطان سینه
۶	۵-۱ تعریف مساله و ضرورت تشخیص
۱۰	۶-۱ ساختار پایان نامه
۱۳	فصل دوم
۱۳	۱-۲ مقدمه
۱۴	۲-۲ ضایعه‌های توده‌ای
۱۶	۳-۲ کلسیفیکیشن‌ها
۱۸	۲-۴ ماموگرافی دیجیتال
۱۹	۱-۴-۲ موقعیت سینه در ماموگرافی دیجیتال
۲۰	۲-۴-۲ غربالگری سرطان سینه
۲۰	۵-۲ آلتراسونوگرافی (US)
۲۰	۶-۲ تصاویر با تشدید مغناطیسی (MRI)
۲۱	۷-۲ تحقیقات انجام شده



۳۲	..... فصل سوم
۳۲	..... ۱-۳ مقدمه
۳۳	..... ۲-۳ پیش پردازش
۳۴	..... ۱-۲-۳ نرمال سازی تصاویر.....
۳۴	..... ۲-۲-۳ عملیات ریخت شناسی
۳۵	..... ۱-۲-۲-۳ انبساط
۳۵	..... ۲-۲-۲-۳ ساییدگی
۳۴	..... ۳-۲-۲-۳ باز کردن و بستن
۳۵	..... ۳-۳ استخراج ویژگی
۳۵	..... ۱-۳-۳ تحلیل بافت.....
۳۹	..... ۲-۳-۳ قطعه بندی بافت.....
۳۸	..... ۳-۳-۳ تحلیل بافت تصاویر ماموگرافی
۴۱	..... ۱-۳-۳-۳ معرفی ماتریس های هم رخداد سطح خاکستری
۴۲	..... ۲-۳-۳-۳ الگوی دودوئی محلی
۴۶	..... ۴-۳ انتخاب ویژگی
۴۶	..... ۱-۴-۳ الگوریتم بهینه سازی ازدحام ذرات (PSO)
۵۰	..... ۱-۱-۴-۳ الگوریتم دودویی PSO
۴۹	..... ۲-۱-۴-۳ کاربردهای PSO

۴۹	.....	۵-۳ کلاسه‌بند
۵۰	.....	۱-۵-۳ ماشین بردار پشتیبان
۵۰	.....	۱-۱-۵-۳ ابر صفحه جداساز
۵۳	.....	۲-۱-۵-۳ ماشین بردار پشتیبان خطی
۵۵	.....	۲-۵-۳ شبکه عصبی مصنوعی
۵۷	.....	۱-۲-۵-۳ پرسپترون
۵۸	.....	۶-۳ معیارهای کارایی سیستم
۶۰	.....	۷-۳ جمع‌بندی
۶۰	.....	۴ فصل چهارم
۶۰	.....	۱-۴ مقدمه
۶۳	.....	۲-۴ پایگاه داده
۶۴	.....	۳-۴ پیش پردازش
۶۵	.....	۱-۳-۴ حذف نویز
۶۷	.....	۲-۳-۴ حذف اثر مصنوعی رادیوپک
۷۲	.....	۴-۴ قطعه‌بندی تصویر
۸۶	.....	۵-۴ استخراج ویژگی‌های بافت
۸۴	.....	۱-۵-۴ انتخاب ناحیه مورد توجه (ROI)
۸۹	.....	۱-۱-۵-۴ لزوم پردازش تصاویر ماموگرافی

- ۲-۵-۴ استخراج ویژگی‌های بافتی ..... ۸۹
- ۳-۵-۴ انتخاب ویژگی‌های بهینه با استفاده الگوریتم دودویی PSO ..... ۹۲
- ۱-۳-۵-۴ تابع برازندگی ..... ۹۴
- ۴-۵-۴ نرمال‌سازی ویژگی ..... ۹۶
- ۵-۵-۴ کلاسه‌بندی ..... ۹۷
- ۵ فصل پنجم ..... ۱۰۰
- ۱-۵ نتیجه‌گیری ..... ۹۸
- ۲-۵ پیشنهادها ..... ۹۹
- مراجع ..... ۱۰۳

## فهرست شکل‌ها

- شکل ۱-۱: آمار شیوع بیماری و مرگ و میر ناشی از رایج‌ترین سرطان‌ها در سال ۲۰۰۸ در میان زنان [۳]. ..... ۳
- شکل ۲-۱: آمار شیوع انواع سرطان‌ها رایج میان زنان در هر ۱۰۰ هزار نفر در تهران بین سال‌های (۱۳۷۷-۱۳۸۰) [۴]. ..... ۴
- شکل ۱-۲: تصویر یک توده سرطانی [۱۰]. ..... ۱۵
- شکل ۲-۲: شکل‌های مختلف یک توده سرطانی [۱۱]. ..... ۱۵
- شکل ۳-۲: لبه‌ها و مرزهای توده‌های سرطانی [۱۱]. ..... ۱۶
- شکل ۴-۲: تصویر یک خوشه MC [۱۰]. ..... ۱۸
- شکل ۵-۲: (a) تصویر ماموگرافی در زاویه CC، (b) تصویر ماموگرافی در زاویه MLO [۱۰]. ..... ۱۹
- شکل ۶-۲: بلوک دیاگرام روش پیشنهادی [۱۳]. ..... ۲۳
- شکل ۷-۲: نمای کلی از خوشه‌بندی MCها با استفاده از اطلاعات ریخت‌شناسی و الگوریتم PSO-LSSVM [۱۴]. ..... ۲۵
- شکل ۸-۲: نمایی کلی از سیستم CAD [۱۶]. ..... ۲۸
- شکل ۱-۳: مثالی از انبساط مجموعه A توسط مجموعه B [۱۸]. ..... ۳۵
- شکل ۲-۳: مثالی از ساییدگی مجموعه A توسط مجموعه B [۱۸]. ..... ۳۶
- شکل ۳-۳: نحوه قطعه‌بندی بافت در تصویر رنگی و خاکستری ..... ۴۰
- شکل ۴-۳: ارتباط فضایی بین پیکسل‌ها. d فاصله از پیکسل مورد نظر است. ..... ۴۲
- شکل ۵-۳: فرآیند ایجاد یک ماتریس GLCM. ماتریس سمت چپ ورودی و ماتریس سمت راست یک GLCM است. ..... ۴۲
- شکل ۶-۳: تغییرات ضریب اینرسی براساس تعداد تکرار ..... ۴۹

- شکل ۳-۷: به روز شدن سرعت و موقعیت ذره ..... ۴۹
- شکل ۳-۸: فلوجارت الگوریتم PSO ..... ۵۰
- شکل ۳-۹: ساختار کلی شبکه عصبی ..... ۵۶
- شکل ۳-۱۰: خروجی دسته‌بندی ..... ۵۹
- شکل ۴-۱: علامت گذاری MCها توسط رادیولوژیست ..... ۶۴
- شکل ۴-۲: (a, b, c, d) نویزهای دیجیتال در تصاویر ماموگرافی ..... ۶۶
- شکل ۴-۳: (a, b, c, d) حذف نویزهای دیجیتال در تصاویر ماموگرافی با فیلتر میانی دو بعدی .... ۶۷
- شکل ۴-۴: هیستوگرام بعد از اعمال استانه سراسری  $T=8$  ..... ۶۹
- شکل ۴-۵: STREL ..... ۷۰
- شکل ۴-۶: حذف اثرهای مصنوعی. (a) تصویر اصلی همراه با برجسب و علائم نشانه گذاری. (b) تصویر آستانه شده با مقدار  $T=18$ . (c) انتخاب اجزای بزرگ نسبت به کل فضا. (d) تصویر سیاه و سفید پس از حذف اثرهای مصنوع رادیوپیک. .... ۷۱
- شکل ۴-۷: برش حاشیه‌های شکل سینه. (a) تصویر باینری شکل سینه با جهت راست. (b) برش تصویر باینری شکل a از چپ و راست. (c) برش تصویر باینری شکل a از بالا و پایین. (d) تصویر باینری شکل سینه با جهت چپ. (e) برش تصویر باینری شکل d از چپ و راست. (f) برش تصویر باینری شکل d از بالا و پایین. .... ۷۳
- شکل ۴-۸: کد نوشته شده در متلب جهت تعیین جهت شکل سینه ..... ۷۵
- شکل ۴-۹: بهبود وضوح تصویر ماموگرام. (a) تصویر ماموگرام بعد از پیش پردازش. (b) هیستوگرام تصویر a. (c) وضوح بهبود یافته تصویر الف. (d) هیستوگرام تصویر c. .... ۷۶
- شکل ۴-۱۰: بخشبندی ماهیچه سینه با روش SRG. (a) تصویر بهبود یافته سینه با جهت راست. (b) تصویر باینری شکل a. (c) ماهیچه قطعه‌بندی شده شکل a با روش SRG. (d) تصویر بهبود یافته سینه با

- جهت چپ. (e) تصویر باینری شکل (f . d) ماهیچه قطعه‌بندی شده شکل d با روش SRG. ۷۸.....
- شکل ۱۱-۴: حذف ماهیچه سینه‌ای ناحیه شکل سینه. (a) تفریق تصویر دودویی شکل ۱۰-۴ c از
- شکل ۱۰-۴. b) (b) تفریق تصویر دودویی شکل ۱۰-۴ f از شکل ۱۰-۴. e) ۷۹.....
- شکل ۱۲-۴: تصویر دودویی قطعه‌بندی شده سینه. (a) شکل سینه با جهت راست بعد از عملیات ریخت شناسی. (b) شکل سینه با جهت چپ بعد از عملیات ریخت شناسی. ۸۰.....
- شکل ۱۳-۴: (a) ماهیچه سینه در شکل. که به شکل مثلث قائم الزاویه دیده شده برای سینه با جهت راست. (b) ماهیچه سینه در شکل. که به شکل مثلث قائم الزاویه دیده شده برای سینه با جهت چپ. ۸۱..
- شکل ۱۴-۴: (a) شکل سینه با جهت راست بعد از اعمال خط مستقیم ماهیچه سینه. (b) شکل سینه با جهت چپ بعد از اعمال خط مستقیم ماهیچه سینه. ۸۳.....
- شکل ۱۵-۴: (a) قطعه‌بندی ماهیچه سینه از تصاویر ماموگرام با جهت راست. (b) قطعه‌بندی ماهیچه سینه از تصاویر ماموگرام با جهت چپ. ۸۴.....
- شکل ۱۶-۴: (a) تصویر سیاه و سفید ماموگرام با جهت راست پس از حذف ماهیچه سینه. (b) هیستوگرام متناظر تصویر a. (c) تصویر سیاه و سفید ماموگرام با جهت چپ پس از حذف ماهیچه سینه. (d) هیستوگرام متناظر تصویر c. ۸۵.....
- شکل ۱۷-۴: استخراج مربع ROI از نمونه‌ها با اندازه‌های متفاوت. ۸۷.....
- شکل ۱۸-۴: نواحی ROI از ناهنجاری‌های خوش‌خیم تصاویر ماموگرام. (a) (کلسیفیکیشن) (b) توده مشخص (c) توده سوزنی شکل (d) توده III-defined (e) اعواج در معماری (f) توده نامتقارن. ۸۸.....
- شکل ۱۹-۴: نواحی ROI از ناهنجاری‌های بدخیم تصاویر ماموگرام. (a) کلسیفیکیشن (b) توده مشخص (c) توده سوزنی شکل (d) توده III-defined (e) اعواج در معماری (f) توده نامتقارن. ۸۹.....
- شکل ۲۰-۴: نواحی ROI بدون بخش‌بندی ناحیه زمینه. ۹۰.....
- شکل ۲۱-۴: نواحی ROI شکل ۲۰-۴ با بخش‌بندی ناحیه زمینه. ۹۱.....

شکل ۴-۲۲: ماتریسی تصادفی برای انتخاب ویژگی ..... ۹۳

شکل ۴-۲۳: نحوه انتخاب ویژگی بر اساس الگوی مورد نظر ..... ۹۴

شکل ۴-۲۴: محاسبه مقدار برازندگی بر اساس الگوها ..... ۹۵

## فهرست جداول

- جدول ۱-۲ : نتایج طبقه‌بند SVM [۱۳] ..... ۲۴
- جدول ۲-۲: نتایج کلاسه‌بندی خوشه‌های MC با الگوریتم PSO- LSSVM [۱۴] ..... ۲۶
- جدول ۱-۳: توصیفگرهای بافت برای ماتریس GLCM [۲۴] ..... ۴۳
- جدول ۱-۴: ویژگی‌های بافتی GLCM استخراج شده از نمونه ROI ..... ۹۲
- جدول ۲-۴: پارامترهای انتخاب شده برای PSO ..... ۹۶
- جدول ۳-۴: نتایج حاصل از کلاسه‌بندهای مختلف روی پایگاه داده DDSM ..... ۹۷
- جدول ۴-۴: نتایج حاصل از کلاسه‌بندهای مختلف روی پایگاه داده MIAS ..... ۹۸
- جدول ۵-۴: پارامترهای کارایی سیستم با دسته‌بندهای مختلف بدون انتخاب ویژگی‌های بهینه ... ۹۸
- جدول ۶-۴: پارامترهای کارایی سیستم با دسته‌بندهای مختلف با انتخاب ویژگی‌های بهینه با روش PSO ..... ۹۸



فصل اول:

مقدمه

# ۱ فصل اول

## ۱-۱ سرطان سینه

پستان عمدتاً از غدد تولید کننده شیر، کانال‌ها، چربی<sup>۱</sup> و بافت<sup>۲</sup> همبند اطراف مجراها، خون، عروق و غدد لنفاوی ساخته شده است. اکثر سرطان‌های سینه منشاء آنها در سلول‌های کانال‌ها می باشد. بدین معنی که در مرحله زودرس سرطان، سرطان هنوز محدود به کانال‌ها می باشد. هنگامی که هنوز به بافت چربی اطراف و احتمالاً به ارگان‌های دیگر پخش نشده است، سرطان سینه قابل درمان است. بنابراین، بسیاری از کشورها به جهت تشخیص سرطان در مراحل اولیه برنامه‌های غربالگری<sup>۳</sup> را شروع کرده‌اند [۱].

## ۲-۱ شیوع سرطان سینه

سرطان سینه<sup>۴</sup> یکی از شایع‌ترین انواع سرطان‌ها می باشد که ۳۲ درصد از سرطان‌های بانوان را تشکیل می دهد و در کل ۱۸ درصد مرگ ناشی از سرطان را شامل می شود. یکی از موثرترین راه‌های کاهش مرگ و میر ناشی از این بیماری، تشخیص آن در مراحل اولیه پیدایش است. تقریباً ۲۵ درصد از مرگ و میرهای ناشی از سرطان سینه بین سنین ۴۰ تا ۴۹ سال در زنان اتفاق می افتد. یک مطالعه توسط انجمن سرطان آمریکا (ACR)<sup>۵</sup> در سال ۲۰۰۳ نشان می دهد که در بین ۸ تا ۱۲ زن، یک نفر به سرطان سینه در طول زندگی مبتلا می شود [۱] و [۲]. مطابق با این آمار هر ۱۵ دقیقه یک زن به بیماری مبتلا به سرطان سینه تشخیص داده شده و یک نفر به علت این بیماری فوت می کند.

سرطان سینه بیشترین عامل مرگ و میر زنان در اروپا است [۲]. در اروپا سرطان سینه ۱۹ درصد

---

<sup>1</sup> Fatty Tissue

<sup>2</sup> Tissue

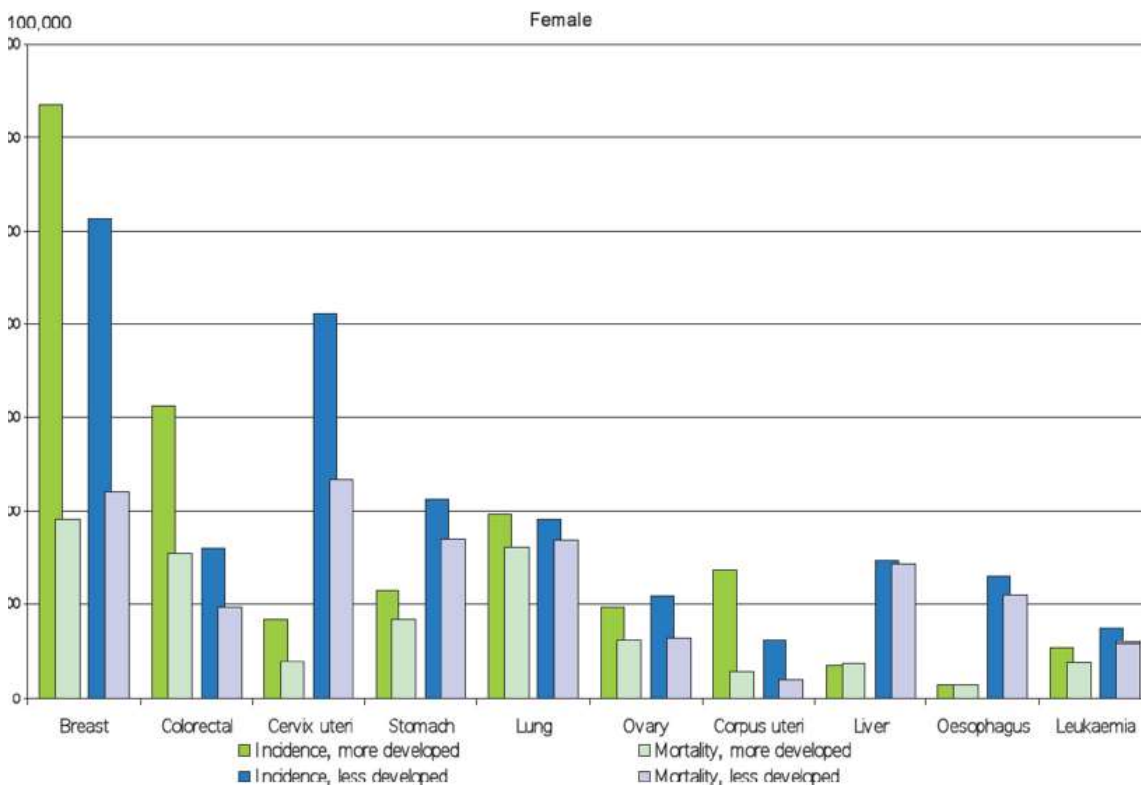
<sup>3</sup> Screening Program

<sup>4</sup> Breast Cancer

<sup>5</sup> American Cancer Society

مرگ‌های ناشی از سرطان را شامل می‌شود.

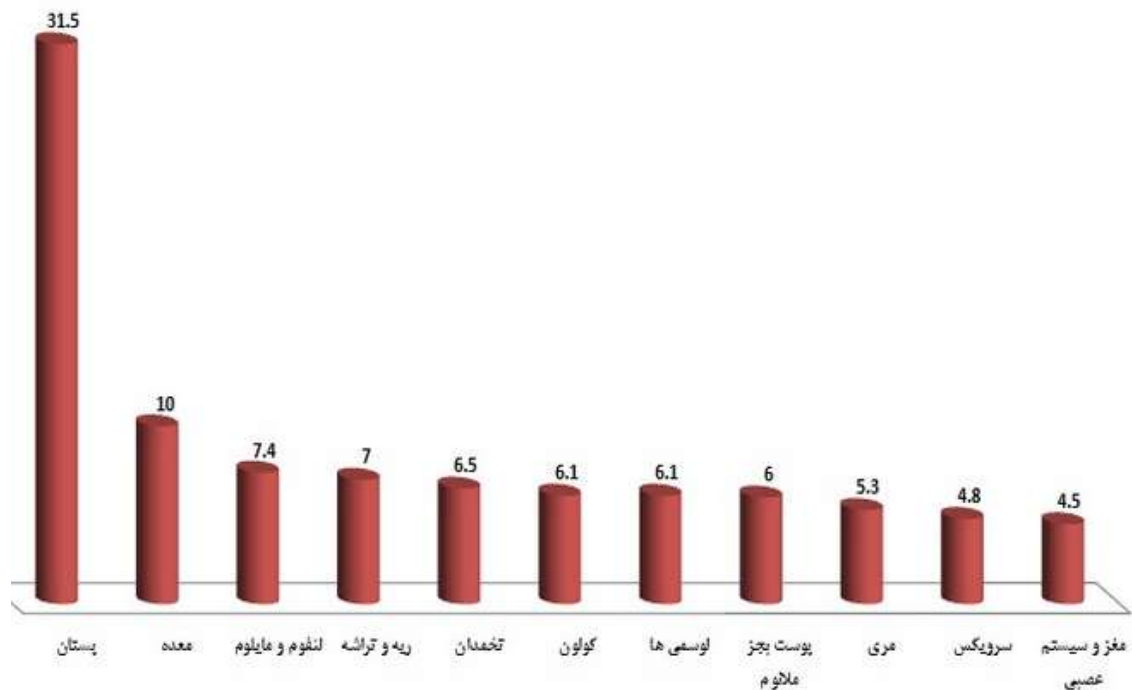
شکل ۱-۱ آمار شیوع بیماری و مرگ‌ومیر ناشی از رایج‌ترین سرطان‌ها در سال ۲۰۰۸ در میان زنان جهان را نشان می‌دهد، که سرطان سینه نسبت به سایر سرطان‌ها بیشترین مقدار را دارد [۳].



شکل ۱-۱: آمار شیوع بیماری و مرگ‌ومیر ناشی از رایج‌ترین سرطان‌ها در سال ۲۰۰۸ در میان زنان [۳].

در ایران گزارش مرکز تحقیقات سرطان سینه دانشگاه علوم پزشکی تهران (CRC)، میزان بروز انواع سرطان‌ها و مرگ‌ومیر آنها را ارایه نموده است. دوره اول برنامه ثبت سرطان شهر تهران جمعیتی در حدود ۷ میلیون نفر را در تهران بین سال‌های (۱۳۷۷-۱۳۸۰) پوشش داده است و در آن اطلاعات سرطان نزدیک به ۲۵۰ بیمارستان و آزمایشگاه تشخیصی پاتولوژی سراسر تهران به صورت فعال جمع آوری گردیده است. به طور خلاصه، طبق یافته‌های این پروژه، تعداد کل موارد سرطان در مردان و زنان به ترتیب ۱۸۵۲۸ و ۱۵۷۴۹ مورد بوده است. میزان خام بروز سرطان در زنان و مردان به ترتیب معادل ۱۰۹ و ۱۲۲ در زن و مرد در هر ۱۰۰ هزار نفر بوده است. همچنین، شایع‌ترین سرطان‌ها در

بین زنان، سرطان پستان، تخمدان، کولون، مغز و سیستم عصبی و شایع‌ترین سرطان‌ها در مردان، شامل سرطان معده، مثانه، پروستات، ریه، لنفوم، کولون و مری بوده است (شکل ۱-۲) [۴] و [۵].



شکل ۱-۲: آمار شیوع انواع سرطان‌ها رایج میان زنان در هر ۱۰۰ هزار نفر در تهران بین سال‌های (۱۳۷۷-۱۳۸۰) [۴]. هم‌چنان‌که رویداد سرطان سینه در یک دهه‌ی اخیر افزایش یافته است، میزان مرگ‌ومیر ناشی از آن در زنان برای همه‌ی رده‌های سنی کاهش یافته است. این کاهش به خاطر بکارگیری صفحات تصویر ماموگرافی<sup>۱</sup> است که امکان تشخیص و بهبودی سرطان سینه را در مراحل اولیه می‌دهد [۵].

### ۳-۱ علائم سرطان سینه

رایج‌ترین و مفیدترین روش برای تشخیص تومورهای سرطان در مراحل اولیه از طریق غربال‌گری ماموگرافی<sup>۲</sup> انجام می‌شود، که می‌تواند مشکلات و علائم بالقوه‌ای چون عدم تقارن بین سینه‌ها، اعوجاج در معماری<sup>۳</sup>، چگالی‌های متلاقی در ارتباط با بافت‌های خوش‌خیم، خوشه‌های

<sup>1</sup> Mammography Screening

<sup>2</sup> Mammogram Screening

<sup>3</sup> Architectural Distortion

میکروکلسیفیکیشن (MCC)<sup>۱</sup> و توده‌ها<sup>۲</sup> را تشخیص دهد. به مراتب رایج‌ترین ویژگی‌های عمومی در ارتباط با تومورهای سرطانی، MCCs و ضایعه‌های توده‌ای هستند، که در ادامه بیان خواهد شد.

## ۴-۱ علل سرطان سینه

اغلب سرطان‌های سینه علتی نامشخص دارند. شروع قاعدگی در سنین پایین‌تر، یائسگی در سنین بالاتر و تأخیر در تشکیل خانواده احتمال وقوع سرطان سینه را بالا می‌برند. استفاده از قرص‌های جلوگیری از بارداری و داروهای هورمونی، نوع تغذیه و شیوه زندگی ممکن است در این امر مؤثر باشند. اما تغییر شیوه زندگی احتمالاً اثر چندانی نخواهد داشت [۱].

بین ۵ تا ۱۰ درصد سرطان‌های سینه ارثی هستند. در این موارد ژن‌های مشخصی در سلول‌های ناسالم می‌باشند. شایع‌ترین ژن‌های معیوب در این زمینه دو ژن BRCA1 و BRCA2 هستند. اگر فرد یکی از ژن‌های ناسالم را به ارث برده باشد احتمال وقوع سرطان سینه و سرطان‌های دیگر بخصوص سرطان تخمدان در او بیشتر خواهد بود. تعداد زیادی از اعضای خانواده‌هایی که ژن‌های BRCA 1 و BRCA 2 را دارند به سرطان سینه در سنین جوانی دچار می‌شوند. در بسیاری از خانواده‌ها یک یا دو نفر دچار سرطان سینه می‌باشند. احتمال وقوع سرطان پستان در اعضای این خانواده‌ها نسبت به سایر افراد جامعه بیشتر است، اما به ندرت جهش‌های ژنی مرتبط به سرطان سینه در آنها مشاهده می‌شود [۱]. در بعضی موارد، اعضای این خانواده‌ها نیازمند غربالگری با ماموگرافی در سنین پایین‌تر هستند.

به طور خلاصه عوامل خطر زا در سرطان سینه را می‌توان در نمونه‌های زیر بیان نمود:

- سن بالا

---

<sup>1</sup> Micro Calcification Cluster

<sup>2</sup> Mass

- سطح بالای فاکتور رشد شبه انسولینی در خون زنان پیش از یائسگی<sup>۱</sup>
- سطح بالای استروژن در خون زنان قبل از یائسگی
- سابقه سرطان سینه در فامیل (تغییرات ارثی در ژن‌های BRCA1 و BRCA2 ( نام اختصاری برای سرطان سینه ۱ و سرطان سینه ۲) در بسیاری از مبتلایان به سرطان‌های سینه و تخمدان وجود داشته است)
- سابقه قبلی ابتلا به سرطان سینه و یا برخی موارد غیر طبیعی دیگر در بافت سینه
- افزایش مواجهه زنان با هورمون استروژن که در شرایط زیر رخ می‌دهد:
  - بروز اولین دوره قاعدگی قبل از ۱۳ سالگی
  - یائسگی بعد از ۵۱ سالگی
  - استفاده از درمان جایگزینی هورمون با استروژن
  - افرادی که در طول زندگی باردار نشده‌اند و یا اولین بارداری آنها بعد از ۳۰ سالگی بوده‌است.
- اضافه وزن و چاقی خصوصاً بعد از یائسگی
- مصرف الکل
- زندگی بی تحرک و فعالیت بدنی کم

## ۵-۱ تعریف مساله و ضرورت تشخیص

اگرچه امروزه، ویژگی‌های فیزیولوژیکی<sup>۲</sup> بیماری سرطان سینه به نسبت خوب تشخیص داده شده است ولی علل تغییرات ریخت شناسی<sup>۳</sup> سینه بر اثر این بیماری هنوز ناشناخته باقی مانده است. ماموگرافی دیجیتال یک روش به نسبت جدید برای تشخیص زودهنگام سرطان سینه است. که

<sup>۱</sup> Insulinlike Growth Factor

<sup>۱</sup> Physiological characteristics

<sup>۲</sup> Morphological

عملکرد آن بر مبنای چگالی انباشته شده بافت‌ها<sup>۱</sup> می‌باشد. این دلیلی بر موثر بودن این روش در تشخیص توده‌ها<sup>۲</sup> و میکروکلسیفیکیشن‌ها (MCs)<sup>۳</sup> می‌باشد [۶]. نتایج بالقوه ماموگرافی دیجیتال:

- قابلیت پردازش تصاویر

تصاویر پزشکی در فرمت دیجیتال به طور بهینه‌تر بیان می‌شوند. (بالا بردن وضوح، چگالی و .. با استفاده از توابع جداساز در نمایش یا ذخیره سازی تصویر)

- امکان استفاده از روش‌های تشخیص مبتنی بر رایانه

- ذخیره سازی تصویر و مدیریت آن

بزرگترین مشکل در ارتباط با برنامه‌های غربالگری ماموگرام<sup>۴</sup> درصد بالای خطای سرطان است. مطالعات نشان می‌دهد که در طی برنامه غربالگری، رادیولوژیست‌ها<sup>۵</sup> در تشخیص ۱۵ درصد سرطان سینه نسبت به شواهد قبلی ناموفق بوده‌اند [۷]. علاوه بر این، زمانی که کوچکترین نشانه‌ها در نظر گرفته می‌شود، تخمین نمونه‌های گم شده ناشی از خطای ادراکی<sup>۶</sup> به ۵۰ درصد افزایش می‌یابد مرجع. مطالعات ردیابی چشمی<sup>۷</sup>، این خطاهای غربالگری ماموگرام را به سه گروه دسته بندی می‌کند:

۱- خطاهای جست‌وجوگر: در این نمونه رادیولوژیست اختلال<sup>۸</sup> را نادیده می‌گیرد. آزمایش-های ردیابی چشمی نشان می‌دهد که با رویت چشمی نمی‌توان به ضایعه<sup>۹</sup> دست یافت.

۲- تشخیص خطا: ضایعه توسط رادیولوژیست‌ها دیده می‌شود اما زمان توقف کوتاه‌تر از یک

آستانه معین برای مثال یک ثانیه است.

---

<sup>1</sup> Tissue

<sup>2</sup> Mass

<sup>3</sup> Micro Classification

<sup>4</sup> Mammogram Screening Programmes

<sup>5</sup> Radiologist

<sup>6</sup> Perception

<sup>7</sup> Eye Tracker

<sup>8</sup> Abnormality

<sup>9</sup> Lesion

۳- خطا در تفسیر: این ضایعه‌ها آگاهانه توسط رادیولوژیست‌ها ارزیابی می‌گردد اما با یک تصمیم نامناسب عمل می‌شود.

به منظور شناخت اینکه آیا بافت سینه سرطانی است یا خیر، به طور معمول یک بافت برداری<sup>۱</sup> انجام می‌شود. طی فرآیند بافت برداری، بافت مشکوک سینه از بیمار جهت بررسی‌های تشخیصی برداشته می‌شود. در نهایت این بافت توسط یک پزشک آسیب شناس<sup>۲</sup> بررسی می‌شود [۵].

اگرچه ماموگرافی دیجیتال ثابت کرده است که ابزاری موثر در تشخیص سرطان سینه است، با این وجود تفسیر ماموگرام‌ها نیازمند تجربه و مهارت یک رادیولوژیست با تجربه است. مطالعات پزشکی نشان می‌دهد که مقدار پیشگویانه مثبت (PPV)<sup>۳</sup> (نسبت مجموع سرطان‌های سینه یافت شده به تعداد بافت برداری) تنها ۱۵-۳۰ درصد است. تصاویر ماموگرافی با کیفیت پایین، خستگی فیزیکی چشم‌ها، یافته‌های جدید رادیوگرافی و بسیاری از موارد دیگر سبب تشخیص نادرست یک ناهنجاری بدخیم<sup>۴</sup> به عنوان خوش‌خیم<sup>۵</sup> خواهد شد [۸].

ضایعه‌های توده‌ای و خوشه‌های میکروکلسیفیکیشن (MCها) دو نشانه مهم ماموگرافی مربوط به سرطان سینه است که در ۳۰-۵۰ درصد از طریق ماموگرافی یافت می‌شود. تشخیص MCها بسیار چالش‌انگیز است زیرا:

- MCها رسوبات کلسیم با ابعاد بسیار کوچک هستند که به صورت گروهی از دانه‌های روشن در ماموگرام ظاهر می‌شوند [۹].
- MCها اغلب در زمینه نامتناجس ظاهر می‌شوند که بیانگر بافت سینه هستند. بعضی از بخش‌های زمینه ویژگی‌هایی مانند آن را دارند که با بافت متراکم ممکن

---

<sup>1</sup> Biopsy

<sup>2</sup> Pathologist

<sup>3</sup> Positive Predictive Value

<sup>4</sup> Malignant

<sup>5</sup> Benign



است روشن تر از MCها در بافت چرب باشند [۹].

- بعضی از MCها وضوح کمتری نسبت به زمینه شان (مانند باف سینه، رگ های خونی، غدد شیری و چربی) دارند. به عبارت دیگر کثرت و اندازه MCها بسیار نزدیک به نويز و زمینه ناهمگن است [۹].

این پایان نامه نیز با تاکید بر تشخیص دقیق، زودهنگام و حساس در زمینه MCها، سعی در تفکیک اتوماتیک بیماران دارد. سیستم اتوماتیک تشخیص اختلال MCها مکانیزی ارائه می نماید که عملکرد سنتی تشخیص MCها که توسط معاینات پزشکی و تست های بالینی انجام می شود، را خودکار و سیستماتیک می نماید.

ضرورت انجام این پایان نامه و تحقیق این است که به منظور بهبود نرخ نتایج بافت برداری، روش های مبتنی بر رایانه برای تشخیص MCها و در نهایت دسته بندی ناهنجاری بدخیم و خوش خیم بکار گرفته شود. این روش ها سبب سهولت در انجام غربالگری اولیه و سپس خواندن ماموگرام و تحلیل آنها توسط رادیولوژیست شده است.

اگر سیستم تشخیص این بیماری کامپیوتری شود، می تواند به عنوان دستیار برای پزشکان عمل کند، به این صورت که در یک مورد خاص پزشک مشکوک به سایر بیماری های مربوط به سینه است در حالی که بیمار ممکن است مبتلا به سرطان بوده و این سیستم به او این هشدار را می دهد و پزشک باید در انجام معاینات خود این نتیجه گیری سیستم کامپیوتری را نیز مورد توجه قرار دهد و این امر تشخیص سریع تر این بیماری را سبب می شود. تشخیص سریع این بیماری باعث می شود بیمار هر چه زودتر تحت کنترل و مراقبت قرار گیرد و بدین ترتیب از عواقب و علائم منفی بیماری در آینده جلوگیری کند.

## ۱-۶ ساختار پایان نامه

این پایان نامه بر روی روش‌های تشخیص بافت‌های شامل MCها از بافت‌های افراد سالم متمرکز شده است. تحقیق حاصل در واقع ارزیابی قدرت تصاویر ماموگرام برای تفکیک افراد سالم از افراد دارای سرطان سینه می‌باشد. در این پایان نامه، ابتدا خلاصه‌ای از روش‌های تصویربرداری پزشکی و کاربرد هر کدام در تشخیص برخی از بیماری‌ها ارائه می‌گردد. به دلیل کارایی بهتر روش تصویربرداری ماموگرافی در تشخیص این سرطان، این روش تصویربرداری برای تحلیل‌های آتی انتخاب شده است.

ارزیابی اولیه نشان داد که به کمک برخی ویژگی‌های بافت تصاویر ماموگرام با درصد تشخیص ۷۰ درصد می‌توان افراد سالم را جدا نمود. در ادامه و با توجه به روش‌های ارائه شده در پیش‌پردازش و استخراج ویژگی این درصد تشخیص بالا رفت. در نهایت با استفاده از انتخاب و کاهش ویژگی‌ها از طریق الگوریتم بهینه‌سازی ازدحام ذرات و همینطور با استفاده از کلاس‌بندهای مناسب این درصد به بالای ۹۰ درصد رسید.

ساختار پایان نامه شامل ۵ فصل است که مختصری از فصل‌های آتی در ادامه آورده شده است:

فصل دوم - مروری بر فعالیت‌های انجام شده: این فصل به معرفی روش‌های مختلف تشخیص برخی بیماری‌ها و اختلالات مربوط به سرطان سینه، بیان مزایا و معایب هر کدام می‌پردازد. همچنین در این فصل، روش‌های تصویربرداری جدید از سینه و کاربرد آنها در تشخیص بیماری‌ها بررسی می‌گردد.

فصل سوم - مباحث تئوری: این فصل شامل مباحث و روش‌های استفاده شده جهت پیش‌پردازش، استخراج ویژگی و کاهش ویژگی‌های بدست آمده از تصاویر و در آخر کلاس‌بندها است.

فصل چهارم - پایگاه داده و روش پیشنهادی: در این فصل در ابتدا درباره پایگاه داده بحث شده و در نهایت درباره روش‌های پیشنهادی برای تشخیص میکروکلسیفیکیشن‌ها و درصدها و نتایج بدست

آمده از پایگاه داده‌های موجود در پایان‌نامه بررسی شده است.

فصل پنجم - در این فصل که آخرین فصل است نتیجه‌گیری نهایی و پیشنهاداتی برای کارهای

آینده و مقاطع بالاتر بیان شده است.



## فصل دوم:

مروری بر فعالیت‌های انجام شده

## ۲ فصل دوم

### ۱-۲ مقدمه

روش‌های تصویر برداری که برای داده‌های بافت سینه انجام شده است روش‌های مختلفی هستند که در این فصل از پایان‌نامه به آنها اشاره شده است. به دلایلی که در ادامه بیان می‌گردد به دنبال استخراج ویژگی<sup>۱</sup> از تصاویر ماموگرافی هستیم. بعد از استخراج ویژگی و انتخاب ویژگی<sup>۲</sup> با روش‌هایی که گفته شده است، با استفاده از یک طبقه‌بند مناسب اقدام به طبقه‌بندی<sup>۳</sup> و متمایز ساختن افراد سالم از افراد بیمار کرده است. در پایان نتایج و درصدهای بدست آمده ذکر شده است. همچنین در انتهای این فصل مروری بر برخی از آخرین رویکردها و تحقیقات انجام شده در زمینه تشخیص MCها خواهد شد.

در حالت کلی یک سیستم تشخیص بیماری با استفاده از تصاویر پزشکی، شامل مراحل پیش پردازش، استخراج ویژگی و طبقه‌بندی است. هدف پیش پردازش حذف تمام عناصری است که برای فرایند تشخیص مفید نیست. معمولاً پیش پردازش، شامل عملیاتی برای بهبود<sup>۴</sup> تصاویر، قبل از انجام استخراج ویژگی از تصاویر می‌باشد.

### ۲-۲ ضایعه‌های توده‌ای

توده از یک گروه از سلول‌های خوشه‌بندی<sup>۵</sup> تشکیل شده است که اطراف بافت‌ها متراکم هستند. توده را می‌توان نشانه‌ای از وضعیت پستان و یا سرطان دانست. اندازه، شکل و لبه‌های توده به رادیولوژیست در ارزیابی احتمال وجود سرطان کمک می‌کند (شکل ۲-۱ و شکل ۲-۲) [۱۰].

---

<sup>1</sup> Feature extraction

<sup>2</sup> Feature selection

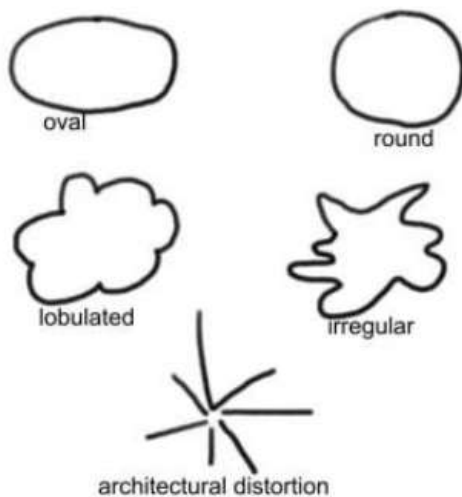
<sup>3</sup> Classification

<sup>4</sup> Enhance

<sup>5</sup> Cluster



شکل ۱-۲: تصویر یک توده سرطانی [۱۰]



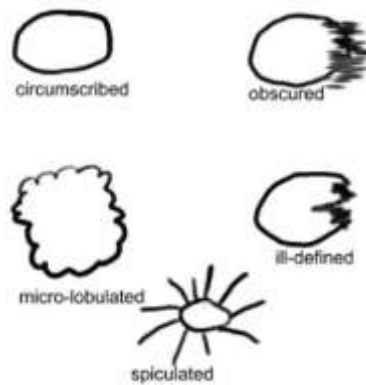
شکل ۲-۲: شکل‌های مختلف یک توده سرطانی [۱۱]

رادیلوژیست‌ها لبه‌ها و مرزهای توده‌ها را در ۵ گروه دسته‌بندی کرده‌اند [۱۱]: (شکل ۲-۳)

- لبه‌های مشخص<sup>۱</sup>: از روی شکل حدود آن کاملاً مشخص می‌شود که بین ناحیه و بافت صدمه دیده قرار می‌گیرد.

<sup>۱</sup> Circumscribe

- لبه‌های مبهم و تیره<sup>۱</sup>: این لبه‌ها دیده نمی‌شوند و یا در مجاورت یکدیگر روی بافت قرار گرفته‌اند.
  - لبه‌های دارای زائده‌های ریز<sup>۲</sup>: دارای محیط دایره‌ای موجدار هستند که در امتداد لبه‌های توده قرار گرفته است.
  - لبه‌های III-defined: این لبه‌ها بطور ناچیزی مشخص و پراکنده هستند.
  - لبه‌های سوزنی شکل<sup>۳</sup>: بوسیله شعاع‌های خطوط کم چربی علامت گذاری می‌شود.
- طبق این دسته‌بندی، شکل لبه‌های سوزنی شکل و III-defined احتمال بیشتری برای وجود توده‌های بدخیم دارد.



شکل ۲-۳: لبه‌ها و مرزهای توده‌های سرطانی [۱۱]

## ۲-۳ کلسیفیکیشن‌ها

کلسیفیکیشن‌ها<sup>۴</sup> رسوبات کلسیم<sup>۵</sup> هستند که در تصاویر ماموگرافی به صورت لکه‌های سفید دیده می‌شود. آنها می‌توانند به عنوان ذخایر نسبتاً بزرگ کلسیم باشند. کلسیفیکیشن‌ها به دو صورت میکرو

---

<sup>1</sup> Obscure  
<sup>2</sup> Micro- Lobulate  
<sup>3</sup> Spiculate  
<sup>4</sup> Calcifications  
<sup>5</sup> Calcium



و ماکرو هستند که معمولاً ماکروکلسیفیکیشن<sup>۱</sup> ناشی از مواردی مانند پیر شدن عروق پستان، آسیب های قدیمی و یا التهاب هستند. وجود خوشه های میکروکلسیفیکیشن (MC)<sup>۲</sup> در تصاویر ماموگرافی می تواند بعنوان یکی از علائم تشخیص سرطان سینه در بانوان باشد. MCها بافت های استخوانی شده ای هستند که در تصاویر ماموگرافی بصورت دانه های درخشان شامل چند پیکسل<sup>۳</sup> روشن تر نسبت به پیکسل های مجاور ظاهر می شوند. البته به علت بافت نسبتاً نامتجانس زمینه ماموگرام ها شناسایی این دانه ها براحتی صورت نمی گیرد به همین دلیل استفاده از سیستم های تشخیص مبتنی بر رایانه می تواند در شناسایی و تشخیص MCها کمک قابل توجهی را به رادیولوژیست ها بنماید (شکل ۲-۴) [۱۱].

ضخامت MCها از ۵۰ تا چند صد میکرون است که معمولاً بصورت یک خوشه بنظر می آیند. در این موارد آنها بر اساس سایز، شکل، تعداد و اعوجاج تحلیل می شوند. براساس قانون عام MCها با اندازه غیر نرمال، شکل های بزرگتر، گرد و بیضی احتمال بیشتری برای فرآیندهای خوش خیم و همچنین MCهای کوچکتر، غیر معمول، چند ریختی و چند شاخه ای اغلب حاکی از پیشرفت بدخیمی است [۱۱].

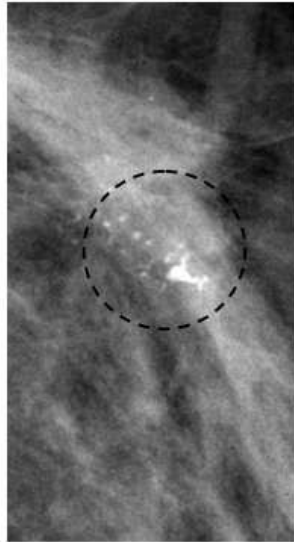
این پایان نامه بر روی روش های تشخیص MCها متمرکز شده است.

---

<sup>۱</sup> Macrocalcification

<sup>۲</sup> Microcalcification

<sup>۳</sup> Pixel



شکل ۲-۴: تصویر یک خوشه MC [۱۰]

## ۴-۲ ماموگرافی دیجیتال

شناسایی و تشخیص سرطان سینه در مراحل اولیه آن احتمال موفقیت در درمان و بهبود کامل بیمار را افزایش می‌دهد. ماموگرافی آزمون تشخیص این بیماری در این مرحله می‌باشد. از لحاظ آماری این شیوه تصویر برداری احتمال مرگ ناشی از سرطان سینه را ۲۰-۳۵ درصد کاهش می‌دهد. در ماموگرافی دیجیتال<sup>۱</sup> شبیه ماموگرافی معمولی از اشعه ایکس<sup>۲</sup> استفاده می‌شود. اما در اینجا به جای اینکه تصویر روی فیلم ثبت شود، به صورت بایت‌های کامپیوتری کدبندی می‌شود. در واقع علاوه بر استفاده از ابزارهای بیشتر و مانیتورهای کامپیوتری، بیمار تفاوت دیگری احساس نمی‌کند. ماموگرافی‌های دیجیتالی نسبت به ماموگرافی‌های قدیمی، از اشعه کمتری استفاده می‌کنند [۱۲].

رادیولوژیست‌ها به تدریج از نرم افزاری استفاده می‌کنند که به آنها در تفسیر تصاویر دیجیتالی کمک می‌کند. توانایی افزایش کنتراست، بزرگنمایی و دستکاری تصویر به رادیولوژیست‌ها این توانایی را می‌بخشد که تفاوت‌های پنهان را که در فیلم رادیولوژی قابل مشاهده نیستند، مشاهده کنند.

---

<sup>۱</sup> Digital Mammography

<sup>۲</sup> X-ray

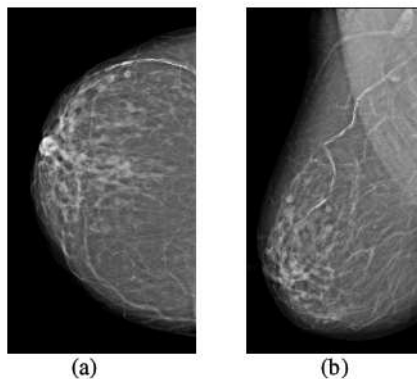
انواع زیادی از نشانه های غیر نرمال در پستان وجود دارد که به نوعی به وجود سرطان سینه اشاره می کند از جمله:

- عدم تقارن<sup>۱</sup> بین تصاویر چپ و راست سینه
- اشکال در شکل نرمال سینه
- وجود MCها در سینه
- وجود توده در سینه

عموما پذیرفته می شود که از یک ماموگرافی با کیفیت بالا، استخراج این چهار ویژگی را قادر سازد [۱۲].

## ۱-۴-۲ موقعیت سینه در ماموگرافی دیجیتال

تصاویر اشعه ایکس از سینه می تواند از زوایای مختلف گرفته شود. در عملیات تشخیصی بیشتر از زوایای CC<sup>۲</sup> و MLO<sup>۳</sup> استفاده می شود که بترتیب در شکل ۲-۵ (a) و شکل ۲-۵ (b) نمایش داده شده است [۱۰].



شکل ۲-۵: (a) تصویر ماموگرافی در زاویه CC، (b) تصویر ماموگرافی در زاویه MLO [۱۰]

<sup>1</sup> Asymmetric

<sup>2</sup> Craniocaudal

<sup>3</sup> Mediolateral-oblique

## ۲-۴-۲ غربالگری سرطان سینه

دانشگاه رادیولوژی امریکا (ACR)<sup>۱</sup> تصاویر پزشکی را یکی از پیشرفت‌های قابل توجه در قرن اخیر اعلام نموده است. در یک مطالعه اخیر نشان داده شده است که تشخیص تومورهای<sup>۲</sup> سرطانی در آمریکا به وسیله ماموگرافی در سال ۱۹۸۴-۱۹۷۴ در حدود ۲ درصد و در سال‌های ۱۹۹۹-۱۹۹۵ حدود ۳۶ درصد پیشرفت داشته است [۱۲].

برنامه غربالگری ماموگرافی در حال حاضر بهترین روش رادیولوژیک برای تشخیص زود هنگام سرطان سینه است. غربالگری از زنان بدون علامت جهت گرفتن عکس با اشعه ایکس (ماموگرافی) در تاریخ‌های مشخص انجام می‌شود. محدوده سن زنان در برنامه‌های مختلف متفاوت است. پس از یائسگی، بافت پستان به تدریج متراکم‌تر می‌شود، بنابراین تشخیص سرطان آسان‌تر است. اجرای برنامه‌های غربالگری سرطان سینه نرخ مرگ‌ومیر را در هلند بطور قابل توجهی کاهش داد. در سال ۱۹۸۶-۱۹۸۸، قبل از غربالگری، میزان مرگ و میر سرطان سینه در زنان هلندی ۷۴-۵۵ ساله ۱۰۵,۲ در هر ۱۰۰۰۰۰ بود در سال ۲۰۰۱ پس از غربالگری، این نرخ به ۸۵,۳ در هر ۱۰۰۰۰۰ رسید [۱۲].

زمانی که ماموگرافی دیجیتال سودمند واقع نشود و یا جهت بررسی‌های بیشتر از روش‌های تصویر برداری دیگری چون آلتراسونوگرافی (US)<sup>۳</sup> و تصاویر با تشدید مغناطیسی (MRI)<sup>۴</sup> استفاده می‌شود.

## ۲-۵ آلتراسونوگرافی (US)

در این روش از امواج صوتی برای بررسی توده‌های سینه استفاده می‌شود. سونوگرافی معمولاً در

<sup>1</sup> American College of Radiology

<sup>2</sup> Tumor

<sup>3</sup> Ultrasonography

<sup>4</sup> Magnetic Resonance Imaging

مواردی انجام می‌شود که پزشک به هر دلیل بررسی‌های بیشتری را صلاح بداند، مثلاً در مواقعی که بافت سینه در خانمی متراکم و سفت باشد و یا سن بیمار کمتر از ۳۵ سال باشد. اما بیشترین کاربرد سونوگرافی برای افتراق توده‌های توپر از توده‌های حاوی مایع و کیست‌های سینه است. سونوگرافی در تشخیص شبه توده‌ها (توده‌های طبیعی که ناشی از به هم چسبیدگی سلول‌های طبیعی سینه هستند)، از توده‌های واقعی که غیرطبیعی هستند نیز کمک کننده است [۱].

## ۲-۶ تصاویر با تشدید مغناطیسی (MRI)

یک روش تصویربرداری بدون درد دیگر MRI است که برای تشخیص ماهیت توده‌های پستانی به کار می‌رود. در این روش از یک میدان مغناطیسی استفاده می‌شود. اگرچه پیش از این کاربردهای MRI بسیار محدود بود ولی امروزه موارد کاربرد آن بیشتر شده و حداقل شامل موارد زیر می‌شود:

- بررسی سینه در افرادی که دارای پروتز سینه هستند و معاینه سینه یا انجام ماموگرافی یا سونوگرافی در آن‌ها با مشکلاتی روبروست.
- بررسی سینه در افرادی که قبلاً تحت عمل جراحی سینه قرار گرفته‌اند. در این موارد انجام معاینات یا ماموگرافی دقت کمتری دارد.
- هرگونه شک تشخیصی که علی‌رغم انجام معاینه، ماموگرافی و سونوگرافی برطرف نشده باشد.
- در افرادی که بعد از درمان سرطان سینه، شک به عود وجود دارد ولی سایر روش‌های تشخیصی کاربرد زیادی ندارد.

در این پایان‌نامه از تصاویر ماموگرافی دیجیتال، برای تشخیص MCها استفاده شده است. زیرا این تصاویر مانند تصاویر MRI و US، شامل مراحل پیچیده پردازش داده‌ها نبوده و در عین حال تصاویری

با کیفیت بالا و با ثبات و وضوح بهتری در مقایسه با داده‌های MRI و US می‌باشند [۱۲].

## ۷-۲ تحقیقات انجام شده

گستره روش‌های استفاده شده در زمینه تشخیص MCها بسیار وسیع است. تشخیص MCها را می‌توان در چهار گروه تقسیم کرد [۹]:

(۱) روش‌های مبتنی بر بهبود کیفیت تصویر

(۲) روش‌های مبتنی بر مدل‌های آماری

(۳) روش‌های تجزیه چند مقیاسی

(۴) روش‌های آموزش ماشین

در این بخش تنها مروری بر برخی از رویکردها و تحقیقات انجام شده خواهیم داشت.

در [۱۳] به ارائه روشی جهت تشخیص خودکار خوشه‌های میکروکلسیفیکیشن‌ها (MC)<sup>۱</sup> در

نواحی مورد توجه (ROI)<sup>۲</sup> استخراج شده از تصاویر ماموگرافی اشعه‌ی X می‌پردازد.

بلوک دیاگرام روش پیشنهادی [۱۳] در شکل ۲-۶ نشان داده شده است. در اولین گام وضوح

تصویر ورودی از طریق پوشش غیر نوک تیز<sup>۳</sup> بهبود می‌یابد. در گام دوم ناحیه‌ی ROI به دو سطح

تجزیه می‌شود و تصویر بازسازی شده از حذف زیر باند فرکانس پایین بدست می‌آید. در گام سوم

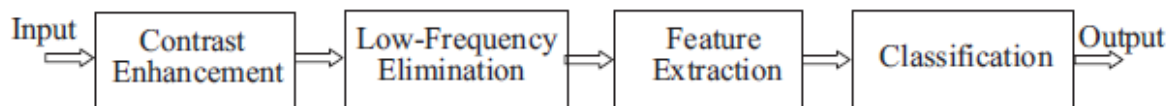
ویژگی‌های بافتی آماری از نواحی ROI استخراج شده و جهت طبقه‌بندی به SVM داده می‌شود.

---

<sup>1</sup> Microcalcification

<sup>2</sup> Regions Of Interest

<sup>3</sup> Unsharp Masking



شکل ۲-۶: بلوک دیاگرام روش پیشنهادی [۱۳]

ROI در موقعیت‌های مختلف در پیکسل‌های  $50 \times 50$  انتخاب می‌شود. در کل ۵۷ ناحیه استفاده می‌شود که ۳۳ عدد از آنها حاوی MC و مابقی بدون MC هستند.

MCها در نواحی ROI در مقایسه با زمینه دارای فرکانس بالاتری هستند. بنابراین حذف اجزای فرکانس پایین در تبدیل کنترلت<sup>۱</sup>، این اطمینان را می‌دهد که وضوح MCها بهبود یافته است. در این گام تصویر ماموگرافی از طریق تحلیل‌های تفکیک پذیری به اجزای فرکانسی‌اش تجزیه می‌شود. استخراج ویژگی‌های آماری<sup>۲</sup> و بافتی<sup>۳</sup> از نواحی ROI ساختار MCها را آشکار خواهد نمود. در روش پیشنهادی [۱۳]، ROI از یک فیلتر بانک<sup>۴</sup> عبور داده می‌شود و به زیر باندهای سطح دو تجزیه می‌شود.

ویژگی‌های کشیدگی<sup>۵</sup>، عدم تقارن<sup>۶</sup>، بی‌نظمی<sup>۷</sup> و معیارهای انرژی<sup>۸</sup> از توزیع MCها محاسبه شده و جهت طبقه‌بندی به SVM داده می‌شود.

برای تشخیص اینکه آیا نواحی ROI حاوی MC هستند یا خیر و طبقه‌بندی صحیح نواحی از بردار ماشین پشتیبان استفاده می‌گردد.

از ۵۷ تصویر ROI، ۱ تا ۲۱ از نوع خوش‌خیم، ۲۲ تا ۳۳ بدخیم و مابقی نرمال هستند. نتایج در جدول ۱-۲ نشان داده شده است.

<sup>1</sup> Contourlet Transform  
<sup>2</sup> Statistical  
<sup>3</sup> Tissue Characteristics  
<sup>4</sup> Bank Filter  
<sup>5</sup> Kurtosis  
<sup>6</sup> Skewness  
<sup>7</sup> Entropy  
<sup>8</sup> Energy Criteria

جدول ۱-۲ : نتایج طبقه‌بند SVM [۱۳]

		Ground truth	
		With MC	Without MC
Classification result	With MC	11.22 (TP)	0.97 (FP)
	Without MC	0.78 (FN)	15.03 (TN)

در [۱۴] به طبقه‌بندی خوشه‌های خوش‌خیم و بدخیم MC در تصاویر ماموگرافی با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات (PSO) و حداقل مربعات بردار ماشین پشتیبان (LSSVM)<sup>۱</sup> می‌پردازد.

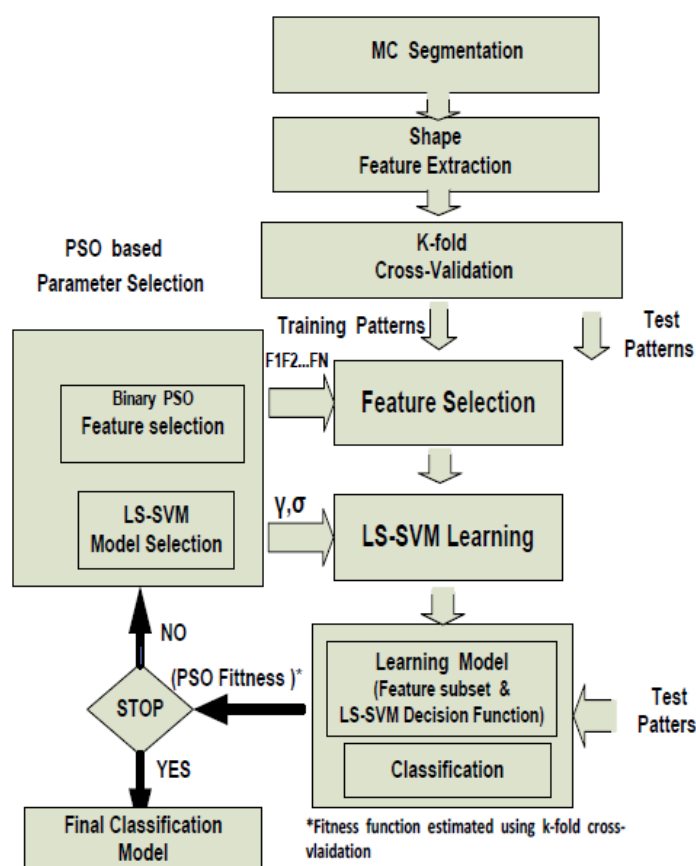
انتخاب ویژگی و بهینه‌سازی پارامترهای طبقه‌بند یکی از گام‌های مهم سامانه‌های تشخیص پزشکی برای ماموگرافی می‌باشد. انتخاب بهینه برای قالب ویژگی‌ها، پارامترهای کرنل<sup>۲</sup> و تنظیم ثابت طبقه‌بند جهت دستیابی به عملکرد مناسب از LSSVM بسیار تعیین کننده خواهد بود. [۱۴] یک سیستم CADx پیشنهاد می‌دهد که براساس آن بدون کاهش ابعاد ویژگی‌ها، با انتخاب ویژگی‌های بهینه در فضای ویژگی‌ها با استفاده از الگوریتم PSO پارامترهای طبقه‌بند را بهبود می‌دهد. در انتها روش PSO و LSSVM با رقبای خود یعنی GA و SVM مقایسه شده‌اند.

در شکل ۷-۲ نمای کلی از مراحل خوشه‌بندی MCها با استفاده از اطلاعات ریخت‌شناسی و الگوریتم PSO- LSSVM نشان داده شده است.

<sup>1</sup>Least-Square Support Vector Machine

<sup>2</sup> Kernel





شکل ۲-۷: نمای کلی از خوشه‌بندی MC با استفاده از اطلاعات ریخت‌شناسی و الگوریتم PSO-LSSVM [۱۴] پایگاه داده این مقاله انجمن تحلیل تصاویر ماموگرافی (MIAS)<sup>۱</sup> معرفی شده است. مجموعه اول ۲۵ تصویر شامل ۱۲ تصویر بیمار از نوع خوش‌خیم سرطان و ۱۳ تصویر بیمار از نوع بدخیم سرطان پستان، مجموعه دوم ۳۵ تصویر شامل ۱۷ تصویر بیمار از نوع خوش‌خیم سرطان و ۱۸ تصویر بیمار از نوع بدخیم سرطان سینه می‌باشد.

نتایج الگوریتم PSO-LSSVM با تابع کرنل RBF برای تشخیص خوشه‌های خوش‌خیم و بدخیم سرطان با استفاده از اطلاعات ریخت‌شناسی<sup>۲</sup> در جدول ۲-۲ بیان شده است. این نتایج برای ۱۳ قالب ویژگی حاصل از ماتریس‌های هم‌رخداد بدست آمده که بهترین دقت برای یک سری از ویژگی‌ها پررنگ‌تر نشان داده شده است.

<sup>1</sup> Mammography Image Analysis Society

<sup>2</sup> Morphology

جدول ۲-۲: نتایج کلاسه‌بندی خوشه‌های MC با الگوریتم PSO-LSSVM [۱۴]

$\gamma$	$\sigma$	N	Shape Features	TPF	TNF	Accuracy	Fitness
604.63	5.55	6	F9,F18,F23,F27,F31,F33	1.0	0.87	0.93	0.2481
481.13	5.67	7	F9,F18,F23,F24,F27,F31,F33	1.0	0.90	0.95	0.2429
928.68	4.83	9	F4,F7,F8,F15,F16,F23,F26,F31,F32	0.93	1.0	0.97	0.2642
<b>3634116</b>	<b>13.28</b>	<b>13</b>	<b>F3,F6,F7,F8,F13,F17,F19,F23,F26,F29,F30,F31,F32</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>0.2974</b>
229.84	3.19	4	F9,F18,F23,F27	0.90	0.90	0.90	0.2586

فرشته تاج نیا در پایان نامه ارشد خود در [۱۵] از ویژگی‌های بافتی حاصل از ماتریس‌های هم‌رخداد، ویژگی‌های فرکانسی و روشنایی تصویر استفاده نموده است. در نهایت این ویژگی‌ها به عنوان ورودی دسته‌بند شبکه عصبی پرسپترون چند لایه و ماشین بردار پشتیبان استفاده شده است. عملکرد این الگوریتم بر روی پایگاه داده DDSM، نرخ دقت ۸۵-۹۴ درصد را برای این سیستم‌ها گزارش می‌دهد.

در [۱۶] به طبقه‌بندی<sup>۱</sup> میکروکلسیفیکیشن‌ها (MC)<sup>۲</sup> با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات (PSO)<sup>۳</sup> پرداخته است.

در [۱۶] طبقه‌بندی بر پایه ۴ گام صورت گرفته است:

۱- با استفاده از فیلترهای فرکانسی متوسط تصاویر صاف‌تر و یکنواخت‌تر شده و بافت ماهیچه‌ها از تصاویر حذف می‌گردد. سپس با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات (PSO) با روش میدان تصادفی مارکو (MRF)<sup>۴</sup> به بخش‌بندی MCها در تصاویر اصلاح شده ماموگرافی می‌پردازد.

۲- ماتریس هم‌رخداد<sup>۵</sup> برای استخراج ویژگی‌های بافتی<sup>۶</sup> از تصاویر بخش‌بندی<sup>۱</sup> شده ساخته می‌-

<sup>1</sup>Classification

<sup>2</sup>Microcalcification

<sup>3</sup>Particle Swarm Optimization

<sup>4</sup>Markov Random Field

<sup>5</sup>Co- occurrence

<sup>6</sup>Texture Feature

شود. ویژگی‌های Haralick استخراج شده و ۱۴ ویژگی برای هر تصویر خواهیم داشت. این

۱۴ ویژگی به ۴ گروه طبق مشخصاتشان تقسیم‌بندی می‌شوند.

۳- الگوریتم ژنتیک (GA) و الگوریتم بهینه‌سازی ازدحام ذرات (PSO) برای انتخاب ویژگی‌های

بهینه<sup>۲</sup> از هر ۴ گروه ویژگی پیشنهاد شده است.

۴- ویژگی‌های انتخاب شده جهت تغذیه شبکه انتشار رو به عقب (BPN) استفاده می‌گردد. وزن

نرون<sup>۳</sup> شبکه BPN با استفاده از PSO استخراج و بروز می‌شوند.

در شکل ۲-۸ نمایی کلی از سیستم<sup>۴</sup> CAD معرفی شده در [۱۶] آورده شده است.

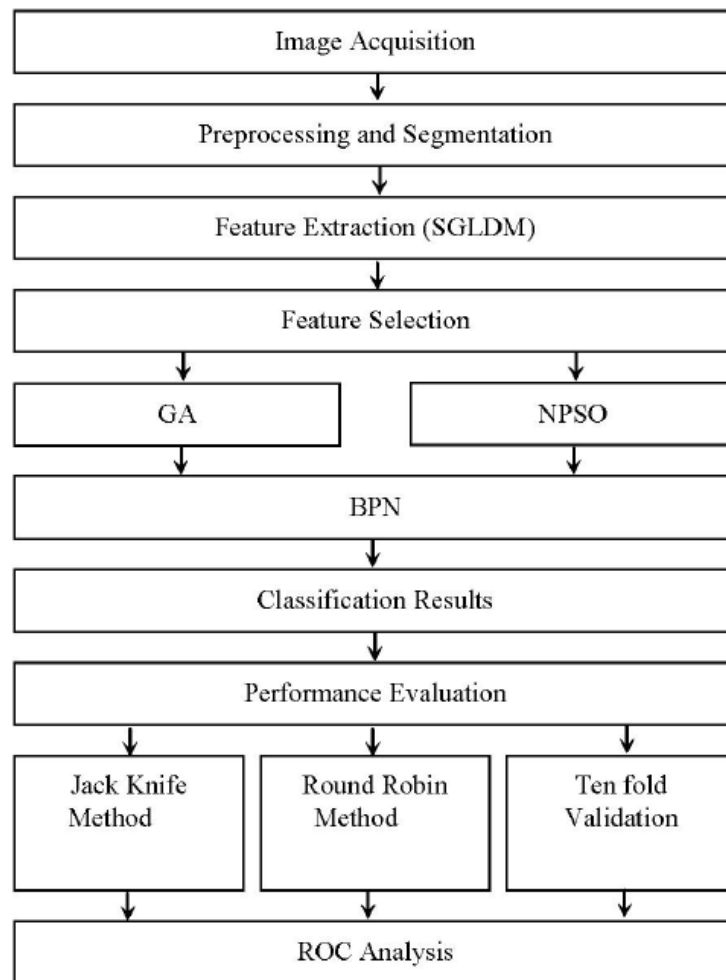
---

<sup>1</sup> Segmentation

<sup>2</sup> Optimal Feature

<sup>3</sup> Neuron

<sup>4</sup> Computer- Aided Diagnosis



شکل ۲-۸: نمایی کلی از سیستم CAD [۱۶]

ویژگی‌های بافتی جهت طبقه‌بندی MCها استخراج می‌شود، برای این منظور از اطلاعات آماری<sup>۱</sup> تصاویر با استفاده از ماتریس‌های هم‌رخداد<sup>۲</sup> انجام می‌شود. ویژگی‌های محاسبه شده از ماتریس‌های هم‌رخداد شامل میانگین شدت روشنایی نقاط مربوط به هر MCها، میانگین مقادیر چگالی نوری نقاط داخل هر MCها، ممان مرتبه چهارم چگالی نوری، واریانس شعاع هر MCها، آنترופی<sup>۳</sup> هیستوگرام مرتبه دوم و ... می‌باشد [۱۶].

در [۱۶]، در مجموع ۱۴ ویژگی بدست آمده که آن‌ها را در ۴ گروه طبقه‌بندی کرده است: از نظر

<sup>۱</sup> Statistical

<sup>۲</sup> Co-occurrence Matrices

<sup>۳</sup> Entropy

هم‌جنسی و همگنی<sup>۱</sup>، درشتی و زبری<sup>۲</sup>، دوره‌ای و تناوبی<sup>۳</sup> بودن و سایر.

با استفاده از الگوریتم ژنتیک (GA) و الگوریتم بهینه‌سازی ازدحام ذرات (PSO) ویژگی‌های بهینه انتخاب می‌گردند و به عنوان ورودی به شبکه BPN داده می‌شود. تعیین وزن اولیه شبکه و بروز رسانی وزن‌ها توسط الگوریتم PSO صورت می‌گیرد.

پایگاه داده این مقاله انجمن تحلیل تصاویر ماموگرافی (MIAS)<sup>۴</sup> معرفی شده است که ۳۲۰ تصویر شامل ۲۰۶ تصویر سالم، ۶۳ تصویر بیمار از نوع خوش‌خیم<sup>۵</sup> سرطان و ۵۱ تصویر بیمار از نوع بدخیم<sup>۶</sup> سرطان سینه<sup>۷</sup> می‌باشد.

نتایج حاصل از طبقه‌بند شبکه BPN با استفاده از روش Jack- Knife ، روش Round- Robin و ۱۰ روش اعتبارسنجی دیگر مورد آزمایش قرار گرفته است، همچنین نتایج با منحنی ROC<sup>۸</sup> مقایسه شده‌اند. نتایج نشان می‌دهد که الگوریتم PSO در انتخاب ویژگی‌های بهینه نسبت به الگوریتم GA بهتر عمل کرده است [۱۶].

---

<sup>1</sup> Homogeneity

<sup>2</sup> Coarseness

<sup>3</sup> Periodicity

<sup>4</sup> Mammography Image Analysis Society

<sup>5</sup> Benign

<sup>6</sup> Malign

<sup>7</sup> Breast Cancer

<sup>8</sup> Receiver Operating Characteristic



فصل سوم:

مباحث توری

## ۳ فصل سوم

در این فصل مباحث تئوری مورد نیاز در فصل بعد ارائه می‌شود. مراحل پیش‌پردازش<sup>۱</sup>، استخراج ویژگی<sup>۲</sup>، انتخاب ویژگی<sup>۳</sup> و کلاسه‌بندی در این فصل بررسی می‌شوند.

### ۳-۱ مقدمه

با بررسی وضعیت تصاویر ماموگرافی و به کمک روش‌های شناسایی آماری الگو می‌توان بیماری سرطان سینه را به صورت اتوماتیک و با هدف دستیاری و کمک به پزشکان تشخیص داد. هدف مسائل شناسایی آماری الگو، دسته‌بندی اتوماتیک داده‌ها در دسته‌ها و کلاس‌های متمایز براساس ویژگی‌های منحصر به فرد داده‌های مربوط به اشخاص می‌باشد. الگوریتم‌های یادگیری ماشین، روش‌های کاربردی و آماری را برای پیدا کردن و استخراج ویژگی‌های مناسب از نمونه‌ها فراهم می‌نمایند. برای کمک به رادیولوژیست‌ها جهت تشخیص دقیق بیماری سامانه‌های شناسایی پزشکی (CADE)<sup>۴</sup> و سامانه‌های تشخیص پزشکی (CADx)<sup>۵</sup> توسعه یافته‌اند.

به کمک انواع ویژگی‌های استخراجی از تصاویر ماموگرافی، می‌توان سیستم‌های منحصر به فردی جهت تشخیص MCها طراحی کرد. برای تشخیص MCها از سیستم‌هایی شامل یک مرحله پیش‌پردازش، استخراج ویژگی و در نهایت آموزش جهت ساختن مدل سیستم و سپس یک مرحله تست برای دسته‌بندی یک نمونه تست جدید می‌باشد، استفاده خواهد شد. بنابراین، ابتدا در فاز آموزش، یک مدل مناسب براساس نمونه‌ها (افراد) موجود در دسته آموزش ساخته می‌شود. سپس به کمک مدل ساخته شده، دسته صحیح یک نمونه تست جدید تشخیص داده می‌شود.

---

<sup>1</sup> Preprocessing

<sup>2</sup> Feature extraction

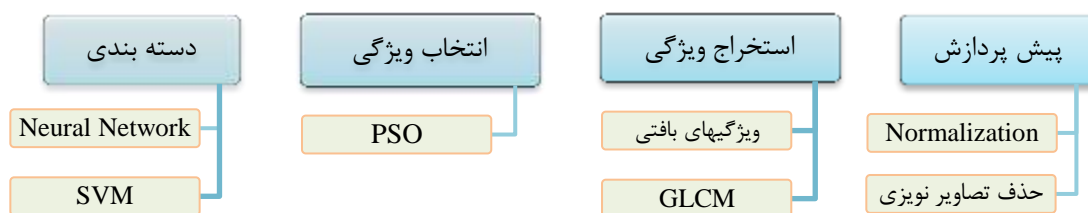
<sup>3</sup> Feature selection

<sup>4</sup> Computer-aided detection

<sup>5</sup> Computer-aided diagnosis



دیاگرام روش پیشنهادی دسته‌بندی تصاویر ماموگرافی جهت تشخیص MCها در شکل ۳-۱ نشان داده شده‌است. در روش پیشنهادی، ابتدا تصاویر ماموگرافی از پایگاه داده استخراج شده و سپس پیش‌پردازش می‌شوند. پس از پیش‌پردازش و آماده‌سازی تصاویر، ابتدا اقدام به استخراج ویژگی با استفاده از برخی روش‌های آنالیز بافت تصویر می‌کنیم. پس از استخراج ویژگی و کاهش ویژگی، با استفاده از کلاس‌بندهای شبکه عصبی و SVM اقدام به دسته‌بندی افراد مبتلا و افراد سالم می‌کنیم. جزئیات مربوط به هر مرحله در بخش‌های بعد تشریح خواهد شد.



شکل ۳-۱: دیاگرام روش پیشنهادی دسته‌بندی تصاویر ماموگرافی جهت تشخیص MCها

## ۳-۲ پیش پردازش

پیش‌پردازش یکی از قسمت‌های مهم و اساسی در هر سیستم پردازش الگو است، مهم از این منظر که عدم اعمال روش‌های مناسب در این مرحله باعث انتشار خطا در کل سیستم شده و کارایی سیستم را در نهایت تحت تاثیر قرار می‌دهد.

پیش پردازش تصاویر ماموگرافی دیجیتال معمولاً شامل حذف نویز و اثرهای تصنعی رادیوپک<sup>۱</sup> است. روش‌هایی چون ایجاد پنجره خط مستقیم<sup>۲</sup> پنجره‌های شش گوشه‌ای<sup>۳</sup> مرجع به عنوان تکنیک‌های فیلتر غیر خطی معرفی شده‌اند [۱۱]. تکنیک‌های فیلتر غیر خطی در حذف نویز موفق‌آمیزتر از

<sup>۱</sup> Radiopaque

<sup>۲</sup> Straight line windowing

<sup>۳</sup> Hexagonal Window

رویکردهای خطی نشان داده شده‌اند. تکنیک‌های دیگری چون فیلترهای میانی<sup>۱</sup> و حفظ کننده لبه‌ها<sup>۲</sup> نیز استفاده شده‌اند [۱۰] و [۱۷].

در این پایان‌نامه جهت حذف اثرهای تصنعی رادیوپک‌ها از زمینه ماموگرام از اعمال فیلتر میانی دو بعدی<sup>۳</sup> استفاده می‌شود.

### ۱-۲-۳ نرمال‌سازی تصاویر

هدف اصلی در پیش پردازش تصاویر پزشکی ارتقا کیفیت تصویر است و روش‌های مختلفی برای این کار وجود دارد که متناسب با کاربرد استفاده می‌شود. در اینجا چون هدف ما استخراج ویژگی از بافت تصویر است این روش بسیار موثر می‌باشد. نرمال‌سازی<sup>۴</sup> مکانی در واقع شامل انتقال داده‌های همه افراد به یک فضای استریوتاکتیک است [۱۲].

در روش نرمال‌سازی داده‌ها به وسیله کمینه کردن مجموع مربعات تفاضلات بین تصویری که باید نرمال شود و تصویری که به عنوان الگو قرار داده شده است، صورت می‌پذیرد. در واقع نرمال‌سازی یا یکسان‌سازی اندازه تصاویر باعث می‌شود، که ابعاد و اندازه تمام تصاویر هم اندازه شود [۱۲].

### ۲-۲-۳ عملیات ریخت‌شناسی

عملیات ریخت‌شناسی<sup>۵</sup> در پردازش تصاویر دیجیتال راهی برای استخراج اجزای تصویر با هدف استخراج ویژگی الگوهای تصویر است. عملیات اساسی ریخت‌شناسی عبارت‌اند از انبساط<sup>۶</sup>، ساییدگی<sup>۷</sup>، باز<sup>۸</sup> و بستن<sup>۱</sup>. این عملیات اساسی هر کدام دارای کاربردهای متفاوتی هستند [۱۸].

<sup>1</sup> Median filtering

<sup>2</sup> Edge- preserving

<sup>3</sup> Two Dimensional Median Filtering

<sup>4</sup> Normalization

<sup>5</sup> Morphology

<sup>6</sup> Dilation

<sup>7</sup> Erosion

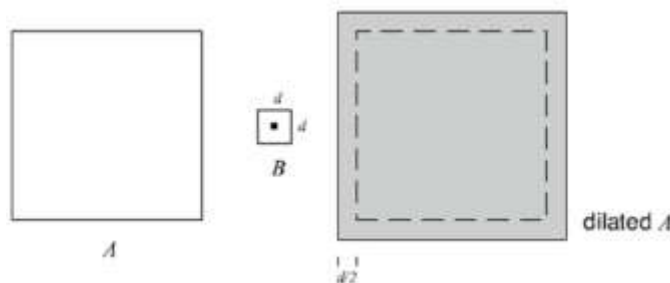
<sup>8</sup> Opening

### ۱-۲-۲-۳ انبساط

انبساط عملیاتی است که شی‌های موجود در تصویر را ضخیم‌تر می‌کند. ضخیم شدن توسط شکلی به نام عنصر سازنده کنترل می‌شود. عنصر سازنده را می‌توان با ماتریسی از صفرها و یک‌ها یا به صورت مجموعه‌ای از پیکسل‌های با مقدار یک نشان داد. فرمول ریاضی انبساط تصویر دودویی  $A^z$  با عنصر سازنده  $B$  در رابطه ۱-۳ نشان داده شده است [۱۸].

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad 1-3$$

در این رابطه  $\emptyset$  مجموعه تهی و  $B$  عنصر سازنده می‌باشند. انبساط  $A$  و  $B$  در واقع مجموعه‌ای شامل تمام مکان‌های مبدأ عنصر سازنده است که  $B$  ابتدا انعکاس یافته و انتقال یافته، حداقل در یک عنصر از  $A$  هم‌پوشانی می‌کنند (شکل ۱-۳).



شکل ۱-۳: مثالی از انبساط مجموعه  $A$  توسط مجموعه  $B$  [۱۸]

### ۲-۲-۲-۳ ساییدگی

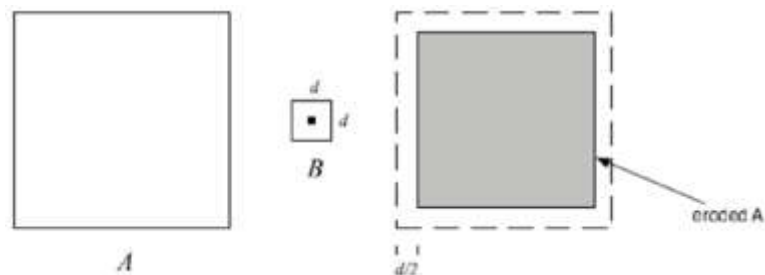
ساییدگی شی‌های موجود در تصویر را کوتاه یا نازک می‌کند. مانند روش انبساط نازک سازی با استفاده از یک عنصر سازنده کنترل می‌شود. فرمول ریاضی ساییدگی در رابطه ۲-۳ نشان داده شده است.

<sup>1</sup> Closing

<sup>2</sup> Binary

$$A \ominus B = \{z | (B)_z \cap A^c\} \quad ۲-۳$$

با توجه به این رابطه، ساییدگی A توسط B، مجموعه‌ای از تمام مکان‌های مبدأ عنصر سازنده است که هیچ بخشی از B با پس‌زمینه A هم‌پوشانی ندارد (شکل ۲-۳) [۱۸].



شکل ۲-۳: مثالی از ساییدگی مجموعه A توسط مجموعه B [۱۸]

### ۳-۲-۲-۳ باز کردن و بستن

ترکیب عملیات انبساط و ساییدگی با هم‌دیگر عملیات باز و بسته کردن ریخت‌شناسی را به وجود می‌آورد. فرمول عملیات باز کردن در رابطه ۳-۳ نشان داده شده است [۱۸].

$$A \circ B = (A \ominus B) \oplus B \quad ۳-۳$$

در این رابطه در ابتدا تصویر A توسط عنصر سازنده B ساییده شده و سپس انبساط توسط همان عنصر سازنده انجام می‌گیرد. در باز کردن ناحیه‌هایی از شی که نمی‌توانند شامل عنصر سازنده باشند را به طور کامل حذف می‌کند، منحنی‌های موجود در شی را هموار و اتصال‌های نازک را می‌شکند و برجستگی‌های نازک را حذف می‌سازد.

فرمول ریاضی عملیات بستن در رابطه ۴-۳ نشان داده شده است.

$$A \blacksquare B = (A \oplus B) \ominus B \quad ۴-۳$$

عملیات بستن با توجه به رابطه (۴-۱۹)، ابتدا عمل انبساط و سپس عمل ساییدگی را انجام می‌دهد. بستن همانند باز کردن منحنی‌های موجود در شی را هموار می‌سازد. برخلاف باز کردن، عمل

بستن شکستگی‌های باریک را به هم متصل می‌کند و شکاف‌های باریک و سوراخ‌های کوچک‌تر از عنصر سازنده را پر می‌کند.

### ۳-۳ استخراج ویژگی

یکی از مهم‌ترین قسمت‌های انجام شده در روش‌های کلاسه‌بندی، کاهش ویژگی‌های بدست آمده بعد از استخراج ویژگی می‌باشد. در بسیاری از موارد ویژگی‌های بدست آمده از بافت تصاویر بسیار زیاد و غیر مفید می‌باشد و همین امر باعث ایجاد اشتباه و پایین آمدن دقت کلاسه‌بندی می‌شود [۱۹]. معمولاً در تصاویر پزشکی بردار ویژگی دارای اندازه بزرگی است و اگر بتوان ویژگی‌های مورد نظر را انتخاب کرد و یا اندازه بردار ویژگی‌ها را کاهش داد می‌توان دقت تشخیص در کلاسه‌بندیها را بالا برد.

برای درک روش‌های مطرح شده در این پایان‌نامه، لازم است دانشی از مفاهیم تحلیل و استخراج بافت تصاویر جهت استخراج ویژگی‌ها داشته باشیم که در ادامه بیان خواهد شد.

### ۱-۳-۳ تحلیل بافت

بافت یک تصویر، شیوه توزیع سطوح روشنایی را در تصویر مشخص می‌کند. از آنجایی که تعریف دقیق و کاملی برای بافت وجود ندارد، معمولاً بافت یک تصویر را با ویژگی‌هایی مثل نرم یا زمخت<sup>۱</sup>، منظم یا تصادفی و مانند آن مشخص می‌کنیم. چنین مشخصاتی را می‌توان با کمک برخی ویژگی‌ها بیان کرد. در شکل ۱-۲ می‌توانید تصاویر با بافت‌های مختلف را مشاهده نمایید [۲۰].

---

<sup>۱</sup> Rough



شکل ۱-۲: تصاویری با چندین نوع بافت [۲۰]

در تعداد زیادی از الگوریتم های پردازش تصویر<sup>۱</sup> و بینایی ماشین<sup>۲</sup> یک فرض ساده و از پیش تعیین شده وجود دارد به این صورت که روشنایی در تمامی ناحیه های تصویر یکنواخت<sup>۳</sup> است، این در حالی است که در اکثر تصاویر گرفته شده از طبیعت این فرض صادق نیست، برای مثال تصویر گرفته شده از سطح یک چوپ دارای توزیع یکنواخت نور در ناحیه های تصویر نیست اما یک الگوی خاص در این تصویر تکرار می شود که به آن بافت بصری<sup>۴</sup> گفته می شود. این الگوها می تواند حاصل ویژگی های فیزیکی سطح نظیر خشونت (زمختی) و یا بازتاب مختلف و متنوع رنگ ها از سطح باشد.

- ما بافت را با دیدن می توانیم تشخیص دهیم اما تعریف بافت سخت است، پیچیدگی کار به دلیل وجود تعداد زیاد و متنوعی از تعریف های مختلف برای بافت است که توسط محققین بینایی ماشین گزارش شده است.
- تعریف بافت ظاهرا دارای یک تضاد<sup>۵</sup> است، از یک طرف از ویژگی بافت تصویر به طور رایج برای پردازش های اولیه در اطلاعات بصری به ویژه با هدف طبقه بندی استفاده می شود، مجموعه ای از تعریف های مختلف برای بافت توسط محققین مختلف فرموله و ارائه شده است که هر کدام برای یک کاربرد خاص مورد استفاده قرار می گیرد و کاربرد

<sup>1</sup> Image processing

<sup>2</sup> Machine vision

<sup>3</sup> Uniform distribution

<sup>4</sup> Visual texture

<sup>5</sup> Conflict

عمومی برای هر یک از فرمول‌های مختلف بافت وجود ندارد. به طور کاملاً ساده می‌توان بافت تصویر را به عنوان یک تابع از تنوع مکانی در شدت پیکسل‌ها در نظر گرفت که می‌تواند در کاربردهای متنوع و مختلف مفید واقع شود اما از طرف دیگر تاکنون هیچ موفقیتی در تعریف دقیقی از بافت وجود ندارد [۲۱].

بافت مهمترین نشانه بصری در تشخیص ناحیه‌های مشابه و یکسان در تصاویر است که طبقه بندی بافت<sup>۱</sup> نیز نامیده می‌شود، هدف نهایی طبقه بندی بافت در نهایت تولید یک نقشه کلاس بندی شده از تصویر ورودی است. ما همچنین می‌توانیم مرزهای بافت را تشخیص دهیم. حتی اگر سطح تصویر را کلاس بندی نکنیم، این مفهوم از کاربرد بافت دومین مسئله است که تعداد زیادی از محققین بینایی ماشین به آن می‌پردازند که تحت عنوان قطعه بندی تصویر<sup>۲</sup> از آن یاد می‌شود. هدف نهایی قطعه بندی تصویر ترسیم مرزهای بافت است. سنتز بافت<sup>۳</sup> اغلب برای کاربردهای متناسب با فشرده سازی تصویر، علاوه بر این از سنتز بافت برای کاربردهای متناسب با گرافیک کامپیوتری استفاده می‌شود جایی که هدف ارائه یک سطح با بالاترین شباهت به همان سطح در دنیای واقعی است [۲۱].

### ۲-۳-۳ قطعه بندی بافت

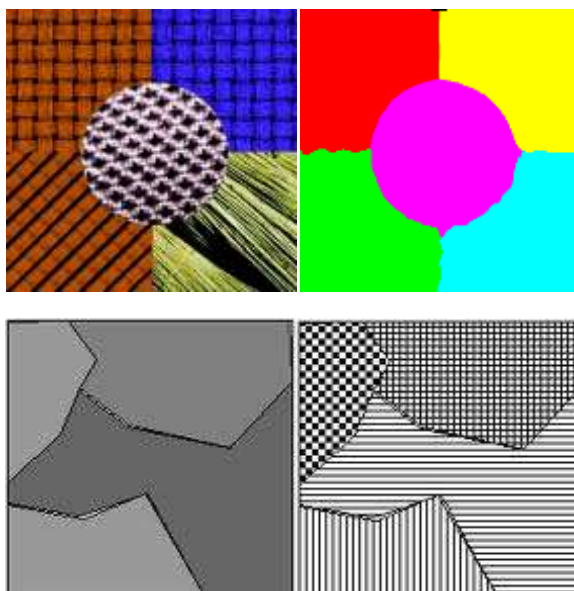
در قطعه بندی بافت تصویر بر مبنای بافت آن به بخش‌های مختلفی تقسیم می‌گردد. شکل ۳-۳ یک نمونه مشخص از چنین ناحیه بندی را برای تصاویر رنگی و همچنین برای تصاویر خاکستری نمایش می‌دهد. بعد از مشخص کردن بافت تصویر باید معیاری را تعریف کنیم که قطعات تصویر بر مبنای آن شکل گیرند [۲۲].

---

<sup>1</sup> Texture classification

<sup>2</sup> Texture segmentation

<sup>3</sup> Texture synthesis



شکل ۳-۳: نحوه قطعه‌بندی بافت در تصویر رنگی و خاکستری

این مسئله یکی از مسائل چالش برانگیز در دنیای بینایی ماشین است و مدت بسیار زیادی است که در این زمینه‌ی مهم کار می‌شود، انسان به راحتی و با کمترین مشکل می‌تواند یک تصویر را ناحیه بندی کند این در حالی است که به دلیل عدم وجود یک تعریف جامع ریاضی برای ناحیه بندی بافت هنوز راه حل جامعی برای این مسئله وجود ندارد، اما می‌توان ناحیه بندی را به صورت کلی تر برای تصویر هم انجام داد به طور کلی هدف از ناحیه بندی تصویر یافتن اشیا و به عبارت بهتر اطلاعات مرتبط در تصاویر است [۲۲].

### ۳-۳-۳ تحلیل بافت تصاویر ماموگرافی

تحلیل بافت تصاویر ماموگرافی دیجیتال از آنجا حائز اهمیت است که این بافت‌ها حاوی الگوهای ویژه-ای از ناهنجاری‌های سرطان سینه هستند. زیرا ماموگرافی دیجیتال، تصاویر سینه با تفکیک پذیری بالا را تولید می‌کند. جایی که بافت‌ها نقش مهمی را ایفا می‌کنند، توصیف‌گر بافت نیز برای انتخاب مجموعه‌ای از ویژگی‌های بافتی مجزا و شایسته نیاز است. ویژگی‌های بافتی شامل داده‌هایی در مورد



توزیع فضایی<sup>۱</sup> چگالی پیکسل‌ها در نواحی تعریف شده تصاویر خاکستری هستند [۲۳].

چندین روش تحلیل بافت در ارتباط با تشخیص سرطان سینه مبتنی بر رایانه از روی تصاویر ماموگرافی معرفی شده است، که در این پایان‌نامه تنها آخرین و موفق‌ترین آنها بیان می‌گردد. دو رویکرد اخیر مبتنی بر بافت به شرح زیر است:

- ماتریس هم رخداد سطح خاکستری (GLCM)<sup>۲</sup>
- الگوی دودویی محلی<sup>۳</sup>

### ۱-۳-۳-۳ معرفی ماتریس‌های هم رخداد سطح خاکستری

ماتریس‌های هم رخدادی (GLCM) یکی از روش‌های محبوب جهت آنالیز بافت و تخمین ویژگی‌های مرتبط با ممان‌های دوم آماری است که بر اساس پیکسل‌های همسایه تصویر تشکیل می‌شوند که در سال ۱۹۷۳ توسط هارالیک و همکاران معرفی شد [۲۴] و [۲۵]. ماتریس‌های هم رخدادی برای زوایای ۰، ۴۵، ۹۰ و ۱۳۵ درجه و فواصل ۱، ۳، ۵، ۷، ۹، ۱۱ و ۱۳ پیکسل می‌توان ساخته شوند.

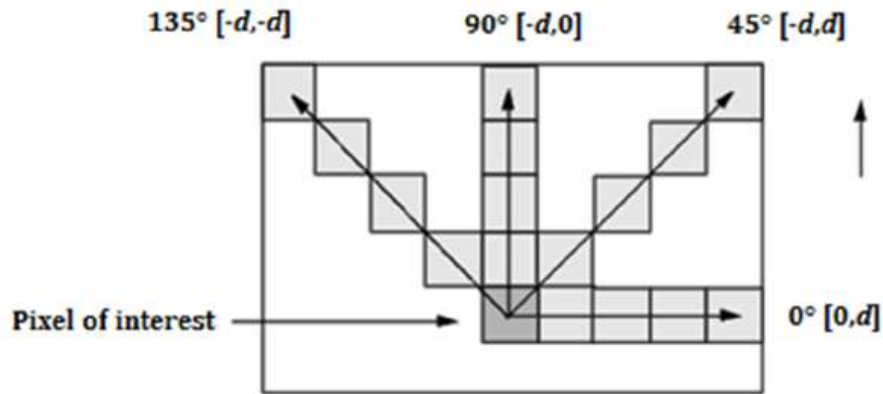
شکل ۳-۴ نشان می‌دهد که ارتباط فضایی بین پیکسل‌ها با دو پارامتر تعریف می‌شود یکی فاصله بین پیکسل‌ها (d) و دیگری زاویه (θ) [۲۶].

---

<sup>1</sup> Spatial Distribution

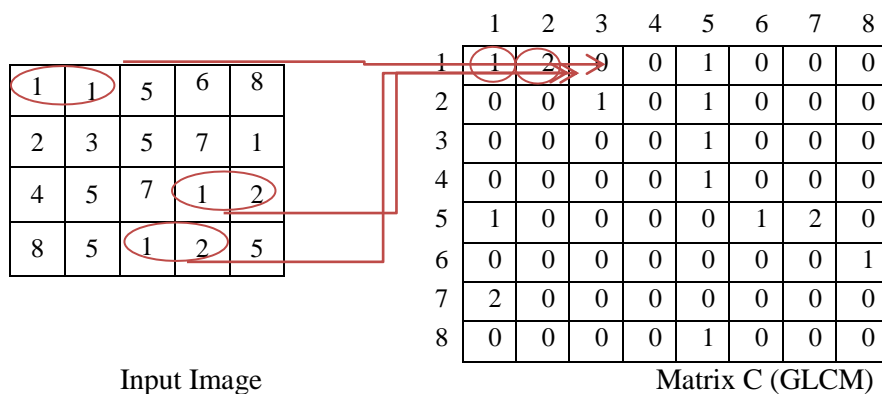
<sup>2</sup> Grey-level Co-occurrence Matrices

<sup>3</sup> Local Binary Patterns



شکل ۳-۴: ارتباط فضایی بین پیکسل‌ها.  $d$  فاصله از پیکسل مورد نظر است.

شکل ۳-۵ فرآیند ایجاد یک ماتریس GLCM را نشان می‌دهد. همسایگی در فاصله یک پیکسل و زاویه صفر درجه برای پیکسل‌های تصویر تعیین شده است. در ماتریس خروجی درایه‌ی (۱ و ۱)، یک شده زیرا تنها یک نمونه با مقادیر ۱ و ۱ در دو پیکسل مجاور افقی تصویر ورودی وجود دارد. به طور مشابه در ماتریس خروجی درایه‌ی (۱ و ۲)، دو شده زیرا تنها دو نمونه با مقادیر ۱ و ۲ در دو پیکسل مجاور افقی تصویر ورودی وجود دارد.



شکل ۳-۵: فرآیند ایجاد یک ماتریس GLCM. ماتریس سمت چپ ورودی و ماتریس سمت راست یک GLCM است. برای ماتریس GLCM با مرتبه‌ی  $m \times n$ ، ۱۴ توصیف‌گر ویژگی توسط هارالیک و همکاران ارائه شده که در جدول ۳-۱ آورده شده است [۲۴] و [۲۵].  $P(i, j)$  بیانگر  $i, j$  امین درایه ماتریس  $C$  شکل ۳-۵ می‌باشد.

جدول ۳-۱: توصیف‌گرهای بافت برای ماتریس GLCM [۲۴]

1	Energy	$\sum_i \sum_j \{p(i,j)\}^2$
2	Contrast	$\sum_{n=0}^{Ng-1} n^2 \{ \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j),  i-j =n \}$
3	Correlation	$\sum_i \sum_j \frac{(ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ <p>Where, <math>\mu_x, \mu_y, \sigma_x, \sigma_y</math> are the means and standard deviations of <math>p_x, p_y</math>, the partial probability density functions.</p>
4	Sum of Square: Variance	$\sum_i \sum_j (i - \mu)^2 p(i,j)$
5	Inverse Difference Moment	$\sum_i \sum_j \frac{p(i,j)}{1 + (i-j)^2}$
6	Sum Average	$\sum_{i=2}^{2Ng} i p_{x+y}(i)$ <p>Where x and y are the coordinates ( row and column ) of any entry in the co-occurrence matrix , and <math>p_{x+y}(i)</math> is the probability of co-occurrence matrix coordinates summing to x+y.</p>
7	Sum Variance	$\sum_{i=2}^{2Ng} (i - s_{ent})^2 P_{x+y}(i)$
8	Sum Entropy	$- \sum_{i=2}^{2Ng} P_{x+y}(i) \log(P_{x+y}(i)) = S_{ent}$
9	Entropy	$- \sum_i \sum_j p(i,j) \log(p(i,j))$
10	Difference Variance	$\sum_{i=0}^{Ng-1} i^2 p_{x-y}(i)$
11	Difference Entropy	$- \sum_{i=2}^{Ng-1} P_{x-y}(i) \log(P_{x-y}(i))$
12	Information Measure of Correlation 1	$\frac{H_{XY} - H_X - H_Y - 1}{\max\{H_X, H_Y\}}$ <p>Where , <math>H_X</math> and <math>H_Y</math> are the entropies of <math>P_X</math> and <math>P_Y</math> such that :</p> $H_{XY} = - \sum_i \sum_j p(i,j) \log(p(i,j))$ $H_{XY1} = - \sum_i \sum_j p(i,j) \log\{p_x(i)p_y(j)\}$ $H_{XY2} = - \sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\}$
13	Information Measure of Correlation 2	$\exp[-2(H_{XY2} - H_{XY})]^{1/2}$

14	Maximum correlation coefficient	$\sqrt{\text{second largest eigenvalue of } Q}$ <p style="text-align: center;">Where,</p> $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$
----	---------------------------------	---

### ۳-۳-۲ الگوی دودوئی محلی

روش الگوی دودوئی محلی (LBP)<sup>۱</sup> یکی از روش‌های مناسب برای توصیف بافت تصویر است، که از آن برای استخراج ویژگی در جهت طبقه‌بندی بر اساس بافت تصویر استفاده می‌شود. این روش در سال‌های اخیر به طور گسترده‌ای در کاربردهای مختلف استفاده شده است. جداسازی مناسب، تغییر ناپذیری در تغییرات سطح خاکستری، مقاوم بودن در برابر تغییرات روشنایی، کارایی و سادگی محاسبات، از ویژگی‌های مهم این روش تحلیل بافت تصویر است [۲۷].

این روش در ابتدا برای استخراج ویژگی از بافت تصاویر مختلف مانند الیاف آلی، چوب و پارچه طراحی شده است. چند دهه بعد از توسعه، مشخص شد برای استخراج ویژگی از انواع دیگر بافت تصاویر، مانند شناسایی چهره مفید است. همچنین در تقسیم‌بندی تصویر و دیگر برنامه‌های کاربردی مفید است [۲۸].

تجزیه و تحلیل بافت<sup>۲</sup>، براساس الگوی دودوئی محلی LBP، به تازگی بیان شده است که برای بسیاری از برنامه‌های حوزه بینایی ماشین<sup>۳</sup> دارای قدرت عالی تمییز است [۲۸].

در این روش شدت روشنایی ۸ همسایگی روی عملگر را با شدت روشنایی پیکسل مرکزی مقایسه می‌شود. اگر مقدار شدت روشنایی هر یک از این ۸ پیکسل همسایه مقدارش از مقدار شدت روشنایی پیکسل مرکزی بزرگتر یا مساوی باشد با یک جایگزین می‌شود در غیر این صورت مقدار آن صفر است. در پایان مقدار شدت روشنایی پیکسل مرکزی با جمع کردن مقدار وزن دار باینری پیکسل-

<sup>۱</sup> Local Binary Pattern(LBP)

<sup>۱</sup>Texture analyze

<sup>۲</sup>Machin vision

های همسایه بدست می‌آید و این پنجره به پیکسل‌های مجاور منتقل می‌شود. با گرفتن هیستوگرام از کل تصویر بدست آمده توصیف کننده‌ای برای بافت تصویر بدست می‌آید. در زیر رابطه‌هایی آمده که به وسیله آنها به راحتی می‌توان این روش را تشریح نمود.

$$\text{Sing}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad ۵-۳$$

$$LBP_{N,R} = \sum_{N=0}^{N-1} \text{sign}(g_n - g_c) 2^N \quad ۶-۳$$

در رابطه ۵-۳، تابع  $\text{sign}(x)$  تابع علامت می‌باشد که در آن برای  $x$ ‌های بزرگتر یا مساوی صفر مقدارش یک می‌شود و برای  $x$ ‌های کوچکتر از صفر مقدارش صفر می‌شود.

در رابطه ۶-۳،  $N$  تعداد همسایه‌ها و  $R$  شعاع همسایگی،  $g_n$  مقدار شدت روشنایی پیکسل همسایه و  $g_c$  مقدار شدت روشنایی پیکسل مرکزی می‌باشد و در نهایت همانطور که در رابطه ۶-۳ آمده است، تابع علامت با ضرب شدن در توان‌هایی از ۲ به صورت باینری در می‌آید.

در حوزه بینایی کامپیوتر، LBP یک روش موثر قوی برای استخراج اطلاعات از ریخت شناسی است. حتی در باز شناسی رگ انگشتان<sup>۱</sup> نیز کاربرد دارد [۲۹].

بر اساس آنچه در مورد روش‌های تحلیل بافت تصاویر ماموگرافی مرور شد و مطالعات پیشین، ویژگی‌های بافتی<sup>۲</sup> و آماری<sup>۳</sup> مهمترین کشف کننده‌ها برای نیل به هدف شناسایی الگوی تصاویر دیجیتال ماموگرام است. بنابراین این پایان‌نامه، ماتریس‌های GLCM را برای استخراج بافت از تصاویر ماموگرام با استفاده از واصف‌گرهای بافت مرجع [۲۴] اعمال می‌نماید.

---

<sup>۱</sup> Finger vein

<sup>۲</sup> Texture Feature

<sup>۳</sup> Statistical Feature

## ۳-۴ انتخاب ویژگی

در بسیاری از موارد ویژگی‌های بدست آمده از بافت تصاویر بسیار زیاد و غیر مفید می‌باشد و همین امر باعث ایجاد اشتباه و پایین آمدن دقت کلاسه‌بند می‌شود. معمولاً در تصاویر پزشکی بردار ویژگی دارای اندازه بزرگی است و اگر بتوان ویژگی‌های مورد نظر را انتخاب کرد و یا اندازه بردار ویژگی‌ها را کاهش داد می‌توان دقت تشخیص در کلاسه‌بندها را بالا برد.

طی دهه گذشته شمار زیادی روش‌های انتخاب ویژگی بهینه ماموگرام آشکار شده و به تصاویر ماموگرام دیجیتال جهت تشخیص توده‌ها و میکروکلسیفیکیشن‌ها اعمال شده است. این روش‌ها اغلب الگوریتم ژنتیک (GA)<sup>۱</sup>، آنالیز تمایزات خطی (LDA)<sup>۲</sup> و الگوریتم بهینه‌سازی ازدحام ذرات (PSO)<sup>۳</sup> است [۱۰].

رویکرد این پایان‌نامه بکارگیری روش‌های بهینه‌سازی ازدحام ذرات (PSO) برای انتخاب ویژگی-های بهینه از تصاویر ماموگرام جهت تشخیص MCها می‌باشد که در ادامه بیان خواهد شد.

### ۳-۴-۱ الگوریتم بهینه‌سازی ازدحام ذرات (PSO)

الگوریتم بهینه‌سازی ازدحام ذرات (PSO) اولین بار در سال ۱۹۹۵ توسط James Kennedy و Russell Eberhart ارائه شد [۳۰]. در تدوین این روش از رفتار اجتماعی دسته‌های پرواز پرندگان و شنای گروهی ماهی‌ها الهام گرفته شده که با استفاده از یک سری روابط ساده فرمول بندی شده است [۳۱].

مانند هر الگوریتم تکاملی، الگوریتم PSO با ایجاد یک جمعیت تصادفی از ذرات شروع می‌شود. مشخصات هر ذره در گروه بر اساس مجموعه‌ای از پارامترها تعیین می‌شود که باید مقادیر بهینه آنها

<sup>1</sup> Genetic Algorithm

<sup>2</sup> Linear Discriminant Analysis

<sup>3</sup> Particle Swarm Optimization

تعیین شود. در این روش هر ذره یک نقطه از فضای جواب مساله را نشان می‌دهد. هر کدام از ذرات دارای حافظه‌اند یعنی بهترین موقعیتی که در فضای جست‌وجو به آن می‌رسند را به خاطر می‌سپارند. حرکت هر ذره در دو جهت انجام می‌شود: ۱- به سوی بهترین موقعیتی که تا کنون اختیار کرده ۲- به سوی بهترین موقعیتی که همه ذرات تا کنون اختیار کرده‌اند. در این روش تغییر موقعیت هر ذره در فضای جست‌وجو تحت تاثیر تجربه و دانش خود و همسایگانش است [۳۱].

فرض کنید در یک مسئله خاص،  $D$  بعدی داریم و  $i$  امین ذره از گروه می‌تواند با یک بردار سرعت و یک بردار موقعیت نمایش داده شود. تغییر موقعیت هر ذره با تغییر در ساختار موقعیت و سرعت قبلی امکان پذیر است. هر ذره اطلاعاتی شامل بهترین مقداری که تاکنون به آن رسیده (بهینه شخصی<sup>۱</sup>) و موقعیت  $X_t$  را داراست. این اطلاعات حاصل مقایسه تلاش‌هایی است که هر ذره برای یافتن بهترین جواب انجام می‌دهد. همچنین هر ذره بهترین جوابی را که تاکنون در کل گروه بدست آمده است، از مقایسه مقادیر بهینه ذرات مختلف می‌شناسد (بهینه فراگیر<sup>۲</sup>). هر ذره برای رسیدن به بهترین جواب سعی می‌کند موقعیت خود را با استفاده از اطلاعات زیر تغییر دهد [۳۲]:

- موقعیت کنونی ( $X^t$ )
- سرعت کنونی ( $V^t$ )
- فاصله بین موقعیت کنونی و بهینه شخصی
- فاصله بین موقعیت کنونی و بهینه فراگیر

بدین ترتیب سرعت هر ذره و به تبع آن موقعیت جدید آن به صورت رابطه‌های ۷-۳ و ۸-۳ تغییر می‌کند.

$$V_i^{t+1} = wV_i^t + c_1 \text{rand}(0,1)(pbest_i - X_i^t) + c_2 \text{rand}(0,1)(gbest_t - X_i^t) \quad 7-3$$

---

<sup>1</sup> Personal Best  
<sup>2</sup> Global best

$$X_i^{t+1} = X_i^t + V_i^{t+1}$$

$V_i^{t+1}$  سرعت ذره  $i$  در تکرار جدید،  $V_i^t$  سرعت ذره  $i$  در تکرار فعلی،  $X_i^{t+1}$  موقعیت ذره  $i$  در تکرار جدید،  $X_i^t$  موقعیت کنونی ذره  $i$ ،  $pbest_i$  بهترین موقعیتی که ذره  $i$  تاکنون اختیار کرده است و  $gbest_t$  بهترین موقعیت بهترین ذره (بهترین موقعیتی که تمام ذرات تا کنون اختیار کرده‌اند) است.  $rand(0,1)$  یک عدد تصادفی بین صفر و یک است که برای حفظ تنوع و گوناگونی گروه به کار می‌رود.  $C_1$  و  $C_2$  به ترتیب پارامترهای شناختی و اجتماعی هستند. انتخاب مقدار مناسب برای این پارامترها منجر به تسریع همگرایی الگوریتم و جلوگیری از همگرایی زودرس در بهینه‌های محلی می‌شود. که انتخاب مقدار بزرگتری برای پارامتر شناختی  $C_1$  نسبت به پارامتر اجتماعی  $C_2$  مناسب‌تر است، اما بایستی همیشه شرط  $C_1 + C_2 \leq 4$  رعایت شود. پارامتر  $w$  اینرسی وزنی نام دارد که برای تضمین همگرایی در دسته ذرات به کار می‌رود اینرسی وزنی جهت کنترل تاثیر سوابق سرعت‌های پیشین بر سرعت‌های جاری مورد استفاده قرار می‌گیرد. بر اساس تحقیقات انجام شده، مقداری بین  $0.4$  و  $0.7$  برای  $w$  مناسب است [۳۲]. در این پایان‌نامه این مقدار را بر اساس تعداد تکرار<sup>۱</sup> با رابطه  $3-9$  تعیین می‌کنیم که مقدار  $w_{initial}$  برابر  $0.9$  می‌باشد.

$$W(t) = W_{initial} - \frac{0.5 * iteration}{max\ iteration}$$

$$W(t) = (W_{initial} - W_{final}) * (max\ iteration - iteration) + 0.2$$

۹-۳

$$W_{initial} = 0.9 \quad , \quad W_{final} = 0.2$$

$$W(t+1) = W(t) + dw \quad \text{Where} \quad dw =$$

$$\frac{W_{min} - W_{max}}{max\ iteration}$$

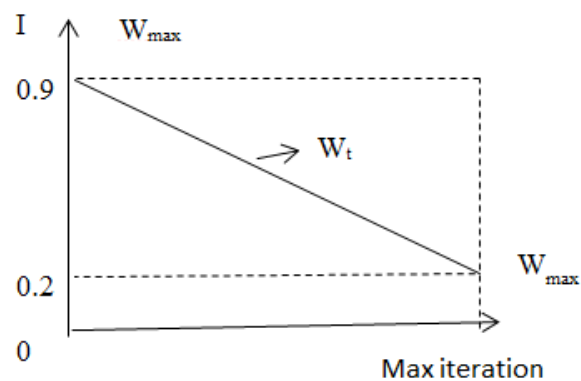
مقدار  $W$  همچنین در کاهش زمان نیز موثر بوده و طبق تعریف بین  $W_{max}$  و  $W_{min}$  قرار می‌گیرد

که هر چقدر مقدار  $W$  زیاد باشد یعنی به  $W_{max}$  نزدیک، مقدار تکرار کاهش یافته و بالعکس. طرحی که

برای این منظور توسط ابره‌ارت و همکارش معرفی شده در شکل ۳-۶ نشان داده شده است.

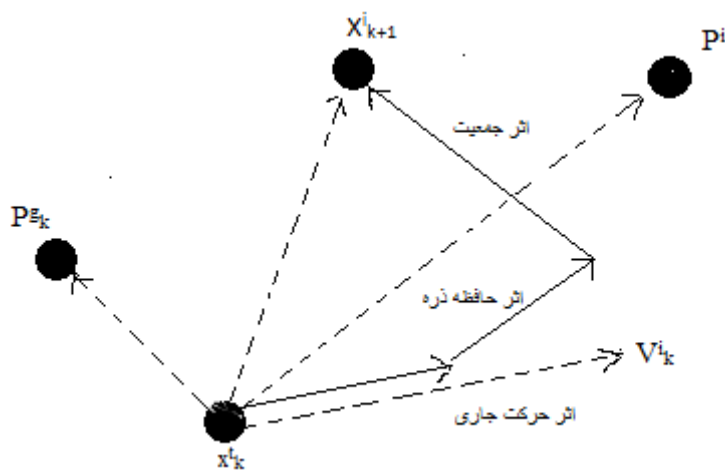
<sup>1</sup> Iteration





شکل ۳-۶: تغییرات ضریب اینرسی براساس تعداد تکرار

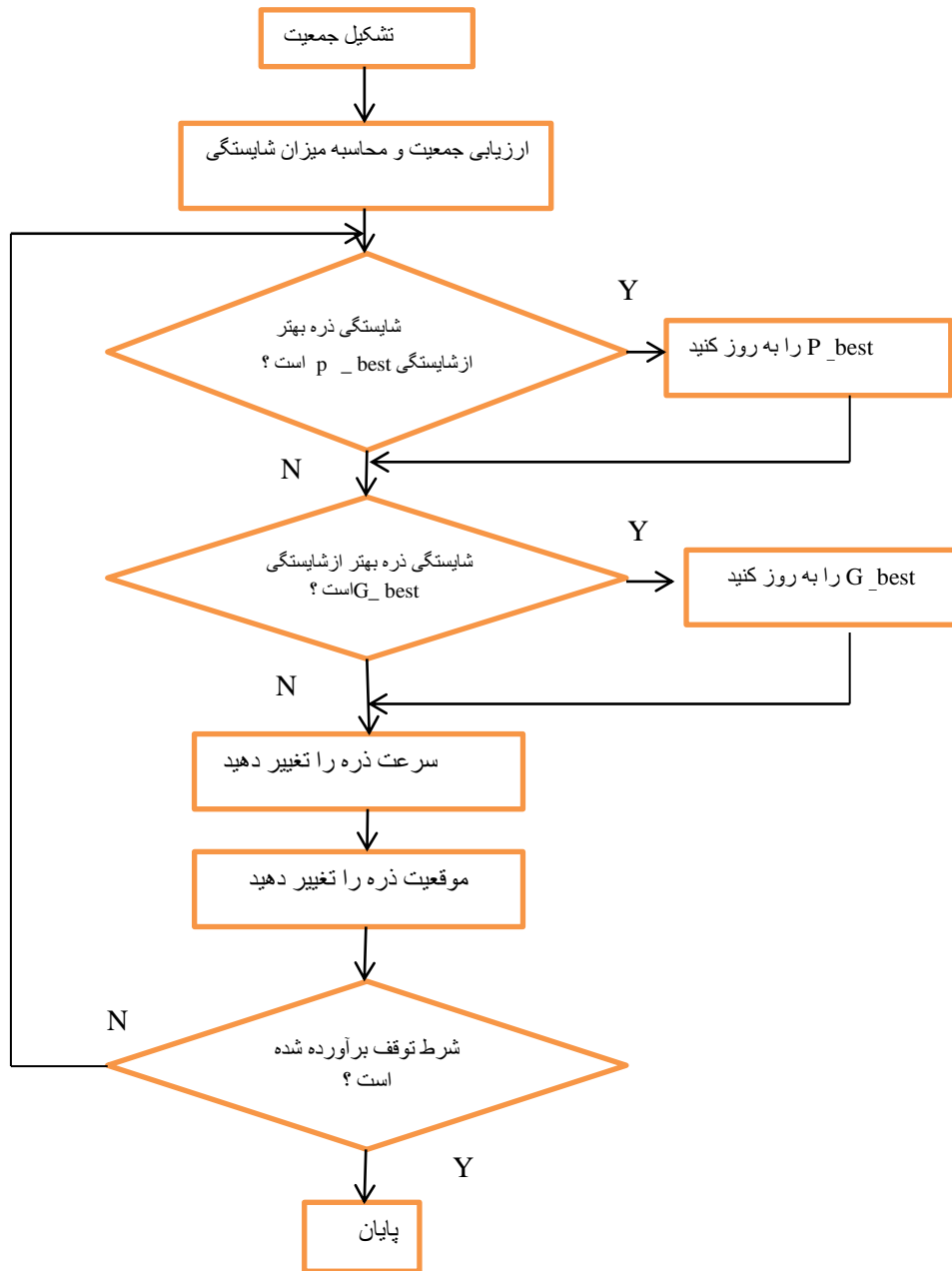
در شکل ۳-۷ نحوه تأثیر روابط به روزرسانی سرعت و موقعیت بر موقعیت هر ذره و همچنین فلوچارت الگوریتم PSO در شکل ۳-۸ به تصویر کشیده شده است.



شکل ۳-۷: به روز شدن سرعت و موقعیت ذره

شرط توقف را میتوان به چند صورت تعریف نمود، مثلاً بر روی حداکثر تعداد تکرار، یا اینکه ماکزیمم تغییر در بهترین شایستگی طبق رابطه ۳-۱۰ برای تعداد معینی از حرکت جمعیت (s)، کمتر از تفرانس تعریف شده باشد [۳۳].

$$|f(p_k^s) - f(p_{k-q}^s)| \leq \varepsilon \quad \text{where} \quad \varepsilon = 10^{-5} \quad \& \quad q = 1, 2, \dots, s \quad ۳-۱۰$$



شکل ۳-۸: فلوجارت الگوریتم PSO

### ۳-۴-۱-۱ الگوریتم دودویی PSO

برای حل مسائل گسسته روش دودویی این الگوریتم توسط آقای کندی در سال 1997 ارائه شد و به PSO شهرت یافت [۳۴]. استفاده از الگوریتم دودویی PSO همچنین زمان محاسبات را کاهش می‌دهد. تفاوت عمده این روش با الگوریتم اصلی در روابط بروزرسانی موقعیت و سرعت می‌باشد که ابتدا باید مقدار بردار سرعت توسط تابع سیگموئید به بازه [۰,۱] منتقل شود سپس موقعیت جدید ذره

محاسبه شود.

موقعیت هر ذره در رشته دودویی با دو مقدار صفر و یک در هر بعد مشخص می‌شود و وضعیت انتخاب ویژگی را نمایش می‌دهد. هر ذره بر اساس رابطه‌های ۱۱-۳ و ۱۲-۳ به روزرسانی می‌شود که سرعت طبق همان رابطه ۷-۳ بدست می‌آید.

$$S(V_i^{t+1}) = 1 / (1 + e^{(-V_i^{t+1})}) \quad 11-3$$

$$X_i^{t+1} = \begin{cases} 1 & \text{if } Rand < S(V_i^{t+1}) \\ 0 & \text{Else} \end{cases} \quad 12-3$$

### ۲-۱-۴-۳ کاربردهای PSO

الگوریتم PSO در حوزه وسیعی از حل مسائل از جمله در بهینه سازی توابع مشکل و چند متغیره، با سرعت و کارایی بالا قابل استفاده است. در ادامه جهت آشنایی فهرست وار به برخی از کاربردهای PSO اشاره می‌شود:

- انتخاب ویژگی‌های بهینه
- شبکه عصبی مصنوعی
- بازشناسی الگو
- کنترل سیستم‌های فازی و ..

### ۵-۳ کلاسه‌بند

آخرین مرحله در تشخیص MCها کلاسه‌بندی است. روش‌های مختلفی برای کلاسه‌بندی وجود دارد که همه آنها مبتنی بر کاربرد، ماهیت داده‌ها و همچنین تعداد داده‌ها می‌باشد. برای مثال زمانی که تعداد داده‌های ورودی مناسب باشد می‌توان از شبکه عصبی برای کلاسه‌بندی استفاده کرد در حالی که در مواردی که تعداد داده‌های ورودی زیاد نباشد ماشین بردار پشتیبان پیشنهاد می‌شود [۳۵].

در این پایان‌نامه تمرکز اصلی روی درصد تشخیص MCها است. با نگاهی به ویژگی‌های استخراج شده و نتایج حاصل از تصاویر می‌توان استنباط کرد که کلاسه‌بند ماشین بردار پشتیبان سازگاری مناسبی با این داده‌ها دارد لذا این کلاسه‌بند پیشنهاد می‌شود.

### ۳-۵-۱ ماشین بردار پشتیبان

یک دسته‌بند، تابعی است که دسته‌ای از اشیا را در کلاس‌های مختلف قرار می‌دهد. در اینجا تصاویر ماموگرام افراد به عنوان شی و کلاس‌ها به صورت بیمار و فرد سالم می‌باشد. بنابراین هدف دسته‌بند پیدا کردن یک قاعده برای تعیین کلاس مناسب مشاهده جدید  $x$  براساس داده‌های آموزش می‌باشد.

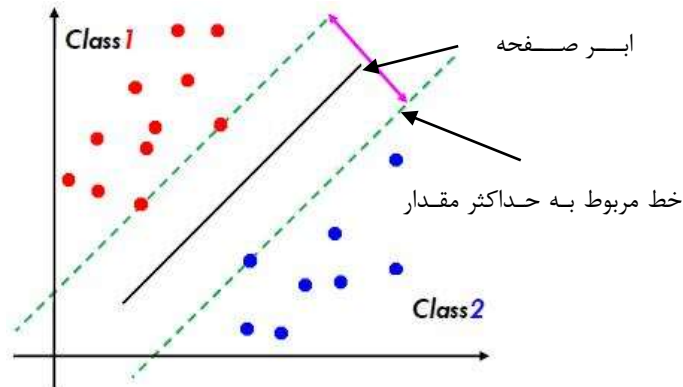
ماشین برداری پشتیبان می‌تواند به خوبی در مسائل طبقه‌بندی دو کلاسه تا چند کلاسه استفاده گردد. برای سادگی درک قضیه می‌توان از مسائل طبقه‌بندی دو کلاسه استفاده نمود. هدف از این گونه مسائل ایجاد یک معیار طبقه‌بندی است که برای نمونه‌های نادیده به خوبی قابل استفاده باشد و در عین حال از قابلیت تعمیم خوبی برخوردار باشد [۳۶].

### ۳-۵-۱-۱ ابر صفحه جداساز

در این گونه مسائل هدف ایجاد یک معیار طبقه‌بندی برای نمونه‌های دو کلاسه است به نحوی که به خوبی قابل استفاده باشد و در عین حال از قابلیت تعمیم و تفکیک خوبی برخوردار باشد، هدف اصلی است.

به عنوان مثال، شکل ۳-۵ در نظر گرفته می‌شود. همان‌گونه که مشاهده می‌شود برای داده‌های شکل ۳-۵ صفحات جداساز خطی وجود دارد که قادر به جداسازی داده‌ها می‌باشند اما تنها

یکی از آن‌ها دارای حاشیه<sup>۱</sup> (فاصله بین صفحه جداساز و نزدیکترین نقاط) بیشینه خواهد بود. این طبقه‌بندی کننده خطی، ابر صفحه جداساز<sup>۲</sup> بهینه نامیده می‌شود که انتظار می‌رود بتواند مرز به دست آمده را به تمام محدوده‌های ممکن تعمیم دهد [۳۷]. ابر صفحه جداساز بهینه در شکل ۵-۳ مشاهده می‌شود.



شکل ۵-۳: صفحه جداساز بهینه با حداکثر مقدار حاشیه

### ۲-۱-۵-۳ ماشین بردار پشتیبان خطی

فرض می‌شود مسئله‌ای برای جداسازی مجموعه نمونه‌های آموزشی که متعلق به دو کلاس

$$w^T x_i + b \begin{cases} \geq 1 \\ \leq -1 \end{cases} \text{ که جداگانه هستند، وجود دارد به گونه‌ای که}$$

$$D = \{(x', y'), \dots, (x', y')\}, x \in R^n, y \in \{-1, 1\}$$

عموما در مسائل جداسازی خطی برای برداری مانند  $x$  وزنی مانند  $w$  در نظر گرفته شود،

به گونه‌ای که این وزن بتواند بردارها را به خوبی در کلاس مربوط به خود طبقه‌بندی نماید. در این

حالت برای صفحه جداساز بهینه رابطه ۳-۱۳ در نظر گرفته می‌شود.

<sup>1</sup> Margin

<sup>2</sup> Hyperplane

$$w^T x + b = 0$$

رابطه بین بردار  $x$  و وزنی مانند  $w$  در این حالت به صورت ضرب داخلی بیان می‌گردد. ضمناً علامت  $T$  ترانهاده ماتریس وزن را نشان می‌دهد.

ابر صفحه‌ای که در این گونه از مسائل برای طبقه‌بندی داده‌ها استفاده می‌شود باید دارای دو ویژگی خاص باشد: اول اینکه دارای کمترین میزان خطای ممکن باشد و از سویی دیگر، از داده‌های هر کلاس بیشترین فاصله ممکن را داشته باشد. در این حالت اگر رابطه (۳-۳) برای صفحه جداساز در نظر گرفته شود، داده‌های آموزشی در بالا و پایین این صفحه قرار خواهند گرفت که به ترتیب برای  $y=1$ ،  $w^T x + b > 0$  و برای  $y=-1$ ،  $w^T x + b < 0$  خواهد بود [۳۸]. بر اساس شرایط بیان شده، زمانی مجموعه‌ای از نقاط به صورت بهینه با یک صفحه جداسازی می‌شوند که:

۱- بدون اشتباه در کلاس مربوط به خود قرار گرفته باشند.

۲- فاصله بین نزدیکترین نقاط هر کلاس داده تا صفحه جدا کننده بیشینه باشد.

دسته‌بند SVM از مجموعه الگوریتم‌هایی است که در شاخه روش‌های بر پایه هسته<sup>۱</sup> از یادگیری ماشین قرار می‌گیرد و این نکته علاوه بر ایجاد کارایی، نقطه ضعفی برای آن به حساب می‌آید. از مشکلات SVM می‌توان به وابستگی این دسته‌بند به هسته مناسب به منظور دسته‌بندی اشاره نمود. وظیفه اصلی هسته انتقال مناسب داده‌ها از فضای ورودی به فضای ویژگی است. در این صورت داده‌هایی که در فضای ورودی به سختی دسته‌بندی می‌شدند، در فضای بردار ویژگی تعریف شده توسط SVM و تابع انتقال هسته، قابل جداسازی می‌باشند [۳۹].

با در نظر گرفتن اهمیت تابع هسته، مدل‌های ریاضی گوناگونی برای آن ارائه شده‌است که در جدول ۳-۱ تعدادی از آنها را ملاحظه می‌نمایید. این مدل‌ها، پیچیدگی‌های متفاوتی از نظر زمان اجرا

<sup>۱</sup> Kernel-Base Methods

و پیاده‌سازی دارند و با توجه به نوع داده‌های ورودی، کارایی‌های متفاوتی را از خود به نمایش می‌گذارند.

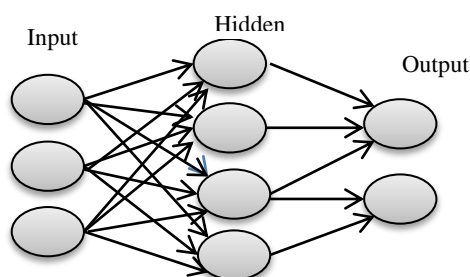
جدول ۳-۱: توابع کرنل در فضای ویژگی

تابع کرنل	نوع طبقه بندی
$K(x_i, x_j) = (x_i^T x_j)^\rho$	خطی از درجه $\rho$
$K(x_i, x_j) = (x_i^T x_j + 1)^\rho$	چند جمله ای از درجه $\rho$
$K(x_i, x_j) = e^{-\frac{\ x_i - x_j\ ^2}{2\sigma^2}}$	گوسین یا نمایی
$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + \mu)$	پرسپترون چند لایه
$K(x_i, x_j) = \frac{\sin((n+1/2)(x_i - x_j))}{2 \sin((x_i - x_j) / 2)}$	دریکله برای مسائل شرایط مرزی

### ۲-۵-۳ شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی<sup>۱</sup> سیستم‌ها و روش‌های محاسباتی نوینی هستند که جهت یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده به کار می‌روند. ایده اصلی این‌گونه شبکه‌ها الهام‌گرفته از شیوه کارکرد سیستم عصبی زیستی، برای پردازش داده‌ها و اطلاعات به منظور یادگیری و ایجاد دانش است.

<sup>1</sup> Artificial Neural Network



شکل ۳-۹: ساختار کلی شبکه عصبی

این سیستم از شمار زیادی عناصر پردازشی فوق العاده به هم پیوسته، با نام نرون<sup>۱</sup> تشکیل شده است. این نرون‌ها برای حل یک مسأله با هم هماهنگ عمل می‌کنند و توسط سیناپس‌ها<sup>۲</sup> اطلاعات را منتقل می‌کنند. هر شبکه دارای لایه‌های ورودی، خروجی و لایه میانی یا مخفی است و این لایه‌ها از تعدادی نرون تشکیل شده‌اند. بین هر لایه و لایه دیگر، نرون‌ها به یکدیگر متصل هستند که این اتصال‌ها دارای وزنی هستند. نرون‌ها نیز خودشان هم دارای وزنی هستند که به این وزن، بایاس<sup>۳</sup> می‌گویند [۴۰].

مقادیر این وزن‌ها بین  $[-1, 1]$  و یا  $[0, 1]$  است و یادگیری شبکه بر اساس به روز رسانی این وزن‌ها با الگوریتم‌های یادگیری می‌باشد. اطلاعات یک شبکه عصبی در وزن‌های آن ذخیره می‌شود. داده‌های ورودی در لایه ورودی قرار می‌گیرند و سپس در وزن‌های مربوط به خود ضرب شده و مجموع حاصل ضربشان به یک تابع غیر خطی به نام تابع فعال سازی می‌رود و خروجی تابع فعال سازی به عنوان ورودی به لایه دیگر می‌رود و فرآیند ذکر شده انجام می‌شود. نرون‌های لایه میانی در اکثر اوقات غیر خطی و نرون‌های لایه خروجی بسته به نوع شبکه می‌توانند خطی یا غیر خطی باشند.

---

<sup>1</sup> Neuron  
<sup>2</sup> Synapse  
<sup>3</sup> Bias



## ۳-۵-۲-۱ پرسپترون

نوعی از شبکه عصبی بر مبنای یک واحد محاسباتی به نام پرسپترون<sup>۱</sup> ساخته می‌شود. پرسپترون برداری از ورودی‌هایی با مقادیر حقیقی را گرفته و ترکیب خطی از این ورودی‌ها را محاسبه می‌کند. اگر حاصل از یک مقدار آستانه بیشتر باشد خروجی پرسپترون برابر با ۱ و در غیر این صورت معادل ۰ خواهد بود.

این نوع شبکه‌ها به دو نوع کلی پرسپترون تک لایه (تنها یک لایه مخفی) و پرسپترون چند لایه (مخفی) تقسیم می‌شوند. پرسپترون‌ها به وسیله اتصالات بین لایه‌ای تنظیم شده‌اند و نرون‌ها در لایه  $i$  به نرون‌های لایه  $i+1$  متصل هستند. این نوع شبکه دارای یک یا چند لایه مخفی بین نرون‌های ورودی و خروجی می‌باشد [۴۱].

شبکه عصبی پرسپترون چند لایه<sup>۲</sup> از جمله شبکه‌های رو به جلو<sup>۳</sup> است که نگاشتی از مجموعه داده ورودی به مجموعه داده خروجی می‌باشد. شبکه عصبی چند لایه، دارای چند لایه است که این لایه‌ها به یکدیگر وصل شده‌اند و این لایه‌ها شامل لایه ورودی، لایه خروجی و یک یا چند لایه میانی هستند. این شبکه، اصلاح شده شبکه پرسپترون خطی است و قادر است داده‌های جدا ناپذیر خطی را تفکیک کند. در این نوع از شبکه‌ها یادگیری پروسه‌ای است که طی آن شبکه عصبی پارامترهای خود را به گونه‌ای تغییر می‌دهد که بهترین شبیه‌سازی محیط را انجام دهد. در شبکه‌های عصبی با پروسه تعریف شده تغییر اوزان در زمان باعث یادگیری می‌شود.

هدف استفاده شبکه پرسپترون چند لایه این است که شبکه‌ای طرح شود که ابتدا با داده‌های آموزشی موجود شبکه آموزش ببیند. سپس با اعمال ورودی‌های تست، کلاس این ورودی را تعیین کند این کار تشخیص الگو می‌باشد [۴۱].

<sup>1</sup> Perceptron

<sup>2</sup> Multilayer Perceptron

<sup>3</sup> Feed forward

### ۳-۶ معیارهای کارایی سیستم

در روش ارائه شده برای تشخیص سرطان، ابتدا ویژگی‌ها از تصاویر مجموعه آموزش و تست استخراج می‌شوند. سپس سیستم با استفاده از ویژگی‌های تصاویر آموزش، آموزش داده می‌شود. به کمک طبقه‌بندها، کلاس یک فرد تست جدید که در گروه آموزش نبوده مشخص می‌گردد. در یک مسئله دسته‌بندی با دو کلاس مثبت و منفی، پس از طبقه‌بندی نمونه‌ها چهار حالت ممکن است ایجاد شود. این چهار خروجی که در شکل ۳-۱۰ نشان داده شده است، به صورت زیر تعریف می‌شود: [۴۲][۴۳]:

۱.  $TN$ : تعداد افراد سالم که به درستی دسته‌بندی شده‌اند<sup>۱</sup>.

۲.  $FP$ : تعداد افراد سالم که به عنوان افراد دارای اختلال تشخیص داده شده‌اند<sup>۲</sup>.

۳.  $TP$ : تعداد افراد بیمار که به درستی دسته‌بندی شده‌اند<sup>۳</sup>.

۴.  $FN$ : تعداد افراد بیمار که به عنوان فرد سالم تشخیص داده شده‌اند<sup>۴</sup>.

معیارهای ارزیابی دیگر که برای ارزیابی کارایی روش‌های ارائه‌شده استفاده می‌شوند عبارتند از: اختصاصیت<sup>۵</sup>، حساسیت<sup>۶</sup>، نرخ تعادل<sup>۷</sup>، صحت<sup>۸</sup>، دقت<sup>۹</sup> و خطا<sup>۱۰</sup> که در ادامه نحوه محاسبات این معیارها بررسی می‌شود.

---

<sup>1</sup> True negative  
<sup>2</sup> False positive  
<sup>3</sup> True Positive  
<sup>4</sup> False Negative  
<sup>5</sup> Specificity  
<sup>6</sup> Sensitivity  
<sup>7</sup> Balanced Rate (BR)  
<sup>8</sup> Precision  
<sup>9</sup> Accuracy  
<sup>10</sup> Error

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

شکل ۳-۱۰: خروجی دسته‌بندی

۱. **اختصاصیت:** این معیار، درصد افراد سالم که به درستی سالم تشخیص داده شده‌اند را مشخص می‌کند و به صورت زیر محاسبه می‌شود.

$$Specificity = \frac{TN}{TN + FP} \quad ۱۴-۳$$

۲. **حساسیت:** این معیار درصد افراد بیماری که به درستی بیمار تشخیص داده شده‌اند را مشخص می‌کند و به صورت زیر محاسبه می‌شود.

$$Sensitivity = \frac{TP}{TP + FN} \quad ۱۵-۳$$

۳. **نرخ تعادل:** در دسته‌بندی، فرض کنید که تعداد افراد سالم ۹۰ و تعداد بیماران ۱۰ باشد. اگر پس از تشخیص تمام افراد به عنوان فرد سالم تشخیص داده شوند، درصد تشخیص روش مربوطه ۹۰٪، درصد تشخیص افراد سالم ۱۰۰٪ و درصد تشخیص افراد بیمار ۰٪ خواهد بود. گرچه سیستم درصد تشخیص بالایی دارد ولی نتوانسته بیماران را به درستی تشخیص دهد. در مقابل این، میانگین درصد تشخیص افراد سالم و بیمار (معیار نرخ تعادل) یک معیار مناسب برای ارزیابی سیستم ارائه می‌نماید که به صورت رابطه زیر محاسبه می‌شود.

$$BR = \frac{Sensitivity + Specificity}{2} \quad ۱۶-۳$$

۴. **صحت:** این معیار درصد افراد بیماری که به درستی بیمار تشخیص داده شده‌اند را به کل افراد بیمار مشخص می‌کند.

$$Precision = \frac{TP}{TP + FP} \quad ۱۷-۳$$

۵. **دقت:** دقت یا نرخ دسته‌بندی یکی از معیارهای ارزیابی دسته‌بندی‌کننده‌ها است. این معیار درصد افراد بیمار و سالمی که به درستی تشخیص داده شده‌اند را به کل افراد موجود در پایگاه داده، مشخص می‌کند و هر چه این معیار بزرگ‌تر باشد، دقت و عملکرد سیستم بالاتر است.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad ۱۸-۳$$

۶. **خطا:** این معیار درصد افراد بیمار و سالمی که به درستی تشخیص داده نشده‌اند را به کل افراد موجود در پایگاه داده، مشخص می‌کند و هر چه این معیار کوچک‌تر باشد، عملکرد سیستم بهتر است.

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \quad ۱۹-۳$$

### ۷-۳ جمع‌بندی

در این فصل مباحث تئوری که برای سیستم تشخیص MCها مورد نیاز می‌باشد را به صورت اجمالی مورد بحث قرار دادیم. در ادامه، در فصل چهارم نتایج آزمایشات و روش‌های گفته شده را با درصد تشخیص افراد سالم از افراد دارای بیماری سرطان سینه با ارائه پیشنهاداتی بر روی دو پایگاه داده [۴۴] و [۴۵] مورد بحث و بررسی قرار می‌دهیم.

فصل چهارم:

روش پیشنهادی و نتایج

## ۴ فصل چهارم

### ۱-۴ مقدمه

در فصل‌های دوم و سوم روش اتوماتیک برای تشخیص بافت‌های شامل MCها از بافت‌های افراد سالم مرور شد.

در هر سیستم بازشناسی یا کلاسه‌بندی، مانند تشخیص چهره، رگ انگشت، اثر انگشت یا بیماری و یا غیره، یکی از مهم‌ترین موضوعات، فراهم کردن پایگاه داده مناسب برای آموزش و ارزیابی سیستم است.

این موضوع در تصاویر پزشکی از اهمیت بسیار بالایی برخوردار است چون به غیر از ارزیابی و بررسی این تصاویر به دست مهندسين و محققين، در سراسر دنيا بايد توسط پزشكان مورد تاييد قرار گيرد و سپس اجازه استفاده به محققين داده شود. در سال‌های اخیر کارهایی در زمینه ایجاد پایگاه داده برای تشخیص سرطان سینه انجام شده است.

بعد از معرفی پایگاه داده، در این پایان نامه روشی برای تشخیص MCها با استفاده از روش‌های ریخت‌شناسی بافت تصویر ارائه شده است. برای ارزیابی روش پیشنهادی از دو پایگاه داده انجمن تحلیل تصاویر ماموگرافی DDSM<sup>1</sup> و انجمن تحلیل تصاویر ماموگرافی MIAS<sup>2</sup> استفاده شده است. برای مقایسه با سایر روش‌های موجود در این زمینه، نتایج ارائه شده است. در هر پایگاه داده، ۷۰٪ تصاویر به طور تصادفی برای آموزش<sup>3</sup> سیستم و ۳۰٪ مابقی برای آزمایش<sup>4</sup> انتخاب شده‌اند.

---

1 Digital database for screening mammography

2 Mammography Image Analysis Society

3 Train

4 Test

## ۲-۴ پایگاه داده

با توجه به اهمیت این بیماری در سطح جهانی، چندین پایگاه داده برای تحقیقات علمی جمع آوری شده است. این کار یک فرصت مهم برای تمامی محققان در سراسر جهان به وجود می‌آورد تا به مطالعه و تحقیق درباره تغییرات ریخت‌شناسی ماموگرام‌های متعدد بپردازند.

این پایگاه‌های داده شامل موارد زیر است:

۱. انجمن تحلیل تصاویر ماموگرافی (MIAS)<sup>۱</sup>
۲. DDSM<sup>۲</sup>
۳. بخش رادیولوژی بیمارستان برونسون (BMH)<sup>۳</sup>
۴. UMMC<sup>۴</sup>
۵. LLNL/UCSF<sup>۵</sup>
۶. Nijmegen<sup>۶</sup>

در این پایان‌نامه از ۱۲۰ تصویر ماموگرام‌های دیجیتال DDSM و MIAS به عنوان ورودی‌های استاندارد در روش پیشنهادی استفاده شده است.

پایگاه داده DDSM با همکاری بیمارستان عمومی ماساچوست<sup>۷</sup> گرفته شده است، که شامل بیش از ۳۰۰۰ تصویر با نمای MLO<sup>۸</sup> و CC<sup>۹</sup> است.

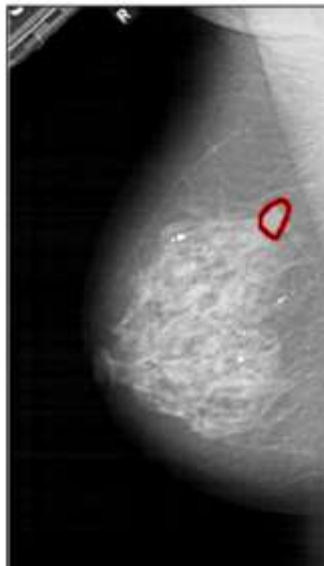
پایگاه داده MIAS شامل بیش از ۳۲۲ تصویر با نمای MLO است.

---

<sup>1</sup> Mammography Image Analysis Society  
<sup>2</sup> Digital database for screening mammography  
<sup>3</sup> The radiology department at Bronson Methodist Hospital  
<sup>4</sup> University Malaya Medical Center  
<sup>5</sup> Lawrence Livermore National Laboratories  
<sup>6</sup> National Expert and Training Centre  
<sup>7</sup> Massachusetts General Hospital  
<sup>8</sup> Mediolateral-oblique  
<sup>9</sup> Craniocaudal

دیگر ویژگی‌های این پایگاه داده:

- تصاویر به سه گروه خوش‌خیم، بدخیم و سالم دسته‌بندی شده‌اند.
- اندازه تصاویر ۱۰۲۴\*۱۰۲۴ پیکسل است.
- محل ناهنجاری‌های خوش‌خیم و بدخیم توسط رادیولوژیست‌های خبره نشانه گذاری شده‌اند (شکل ۱-۴).



شکل ۱-۴: علامت گذاری MCها توسط رادیولوژیست

همچنین در [۴۴] و [۴۵] جزئیاتی از پایگاه داده‌ای که فراهم شده وجود دارد، که اطلاعاتی از پایگاه داده را در اختیار محققین قرار می‌دهد. این اطلاعات شامل تهیه کنندگان پایگاه داده، انتشارات، بودجه و از همه مهم‌تر حجم نمونه‌ها است، که تعداد افراد سالم و افراد مبتلا نشان می‌دهد.

### ۳-۴ پیش پردازش

تصاویر ماموگرافی دیجیتال اغلب حاوی اثرهای مصنوعی<sup>۱</sup> به شکل برجسب<sup>۲</sup>، علایم نشانه گذاری<sup>۳</sup> و

<sup>1</sup> Artifact

<sup>2</sup> Label

<sup>3</sup> Marker



گوشه<sup>۱</sup> در زمینه خود می‌باشند. این اثرهای مصنوعی معمولاً رادیوپک<sup>۲</sup> می‌باشند که برای تابش<sup>۳</sup> شفاف<sup>۴</sup> نیستند. بیشترین مشکل در بخش بندی دقیق تصاویر سینه ناشی از وجود این اثرهاست که ممکن است نتیجه الگوریتم را تغییر دهد. پیش پردازش تصاویر ماموگرافی شامل حذف نویز و اثرهای مصنوعی رادیوپک به منظور خنثی نمودن پیکسل‌های سیاه انجام می‌گیرد.

### ۱-۳-۴ حذف نویز

تصاویر سیاه و سفید ماموگرافی با استفاده از جعبه ابزار پردازش تصاویر متلب قابل نمایش است. این تصاویر از نظر ریاضی با یک ماتریس<sup>۵</sup> مرتبه  $1024 * 1024$  که مقادیری در محدوده (۰,۲۵۵) دارند بیان می‌گردند.

همان‌گونه که در شکل ۲-۴ نشان داده شده است نویزهای دیجیتالی مانند خط‌های افقی و عمودی در بیشتر تصاویر ماموگرام ظاهر می‌شوند. این نویزها با اعمال فیلتر میان‌دو بعدی<sup>۶</sup> در یک همسایگی سه به سه حذف می‌شوند.

---

<sup>1</sup> Wedge

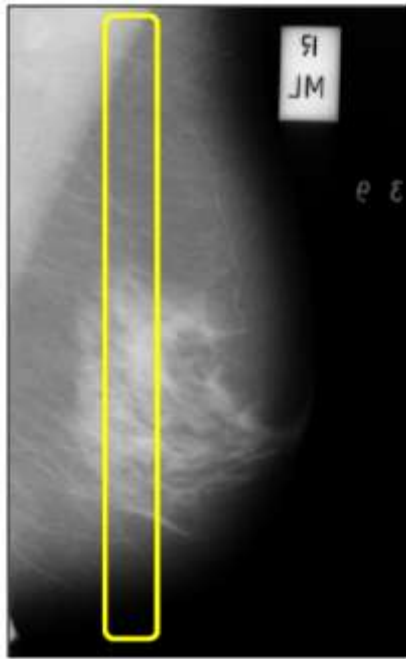
<sup>2</sup> Radiopaque

<sup>3</sup> Radiation

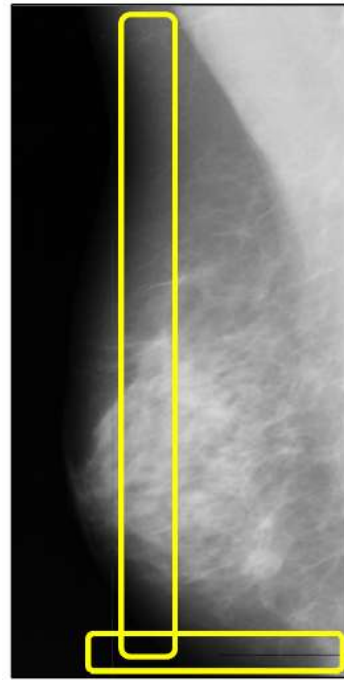
<sup>4</sup> Transparent

<sup>5</sup> Matrix

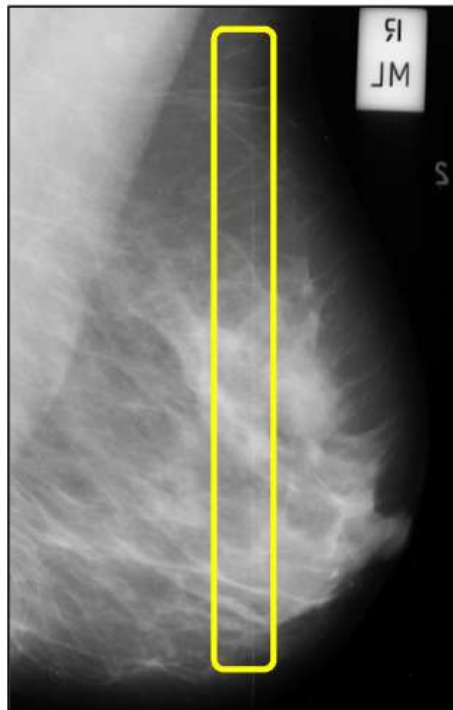
<sup>6</sup> Two Dimensional Median Filtering



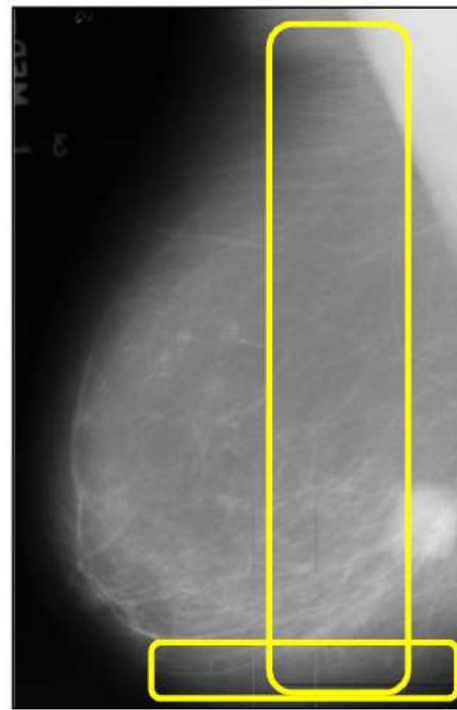
a



b



c



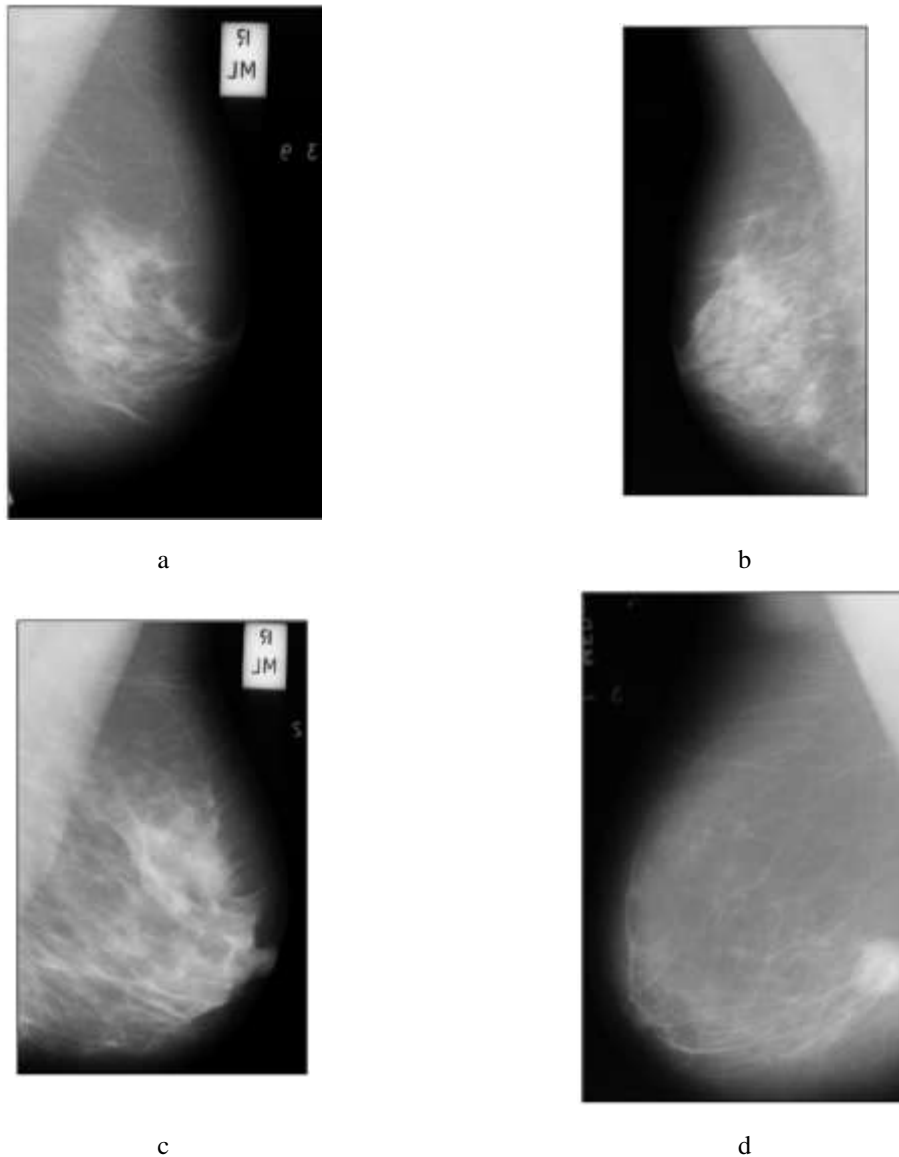
d

شکل ۴-۲: (a, b, c, d) نویزهای دیجیتال در تصاویر ماموگرافی

در رویکرد فیلتر میانی دو بعدی، هر پیکسل خروجی محتوای میانگین مقادیر همسایگی سه در

سه پیکسل مورد نظر در تصویر است. لبه‌های تصویر با صفر جایگزین می‌شوند (رنگ سیاه). این کار بر

روی تصویر اصلی اثر نخواهد داشت زیرا ناحیه ROI از تصویر شامل لبه‌ها و کرانه‌های آن نمی‌شود.  
 شکل ۳-۴ تصاویر ماموگرافی شکل ۲-۴ را بعد از اعمال فیلتر مذکور نشان می‌دهد.



شکل ۳-۴: a, b, c, d حذف نویزهای دیجیتالی در تصاویر ماموگرافی با فیلتر میانی دو بعدی

### ۲-۳-۴ حذف اثر مصنوعی رادیوپیک

این اثرهای مصنوعی ذرات ریزی هستند که در تصاویر ماموگرافی به اشتباه میکروکلسیفیکیشن به نظر می‌رسند. حذف اثر مصنوعی رادیوپیک و ناحیه‌ی زمینه در تصویر ماموگرام، همگنی ناحیه را

افزایش داده و همچنین قابلیت اطمینان<sup>۱</sup> و نیرومندی<sup>۲</sup> جدا نمودن شکل سینه از زمینه را بهبود می-دهد. الگوریتم حذف اثر مصنوعی رادیوپیک بر پایه‌ی اطلاعات ریخت شناسی<sup>۳</sup> ناحیه، از [۴۶]، [۴۷] و [۴۸] گرفته شده است.

به منظور استفاده از رویکرد ریخت شناسی ناحیه معرفی شده در [۴۶]، لازم است تصاویر ماموگرام سیاه و سفید<sup>۴</sup> (۰ و ۲۵۵) به فرمت دودویی<sup>۵</sup> (۰، ۱) تغییر یابد. جهت این کار یک آستانه<sup>۶</sup> سیاه و سفید برای تصاویر تعیین می‌گردد تا بدون اینکه اطلاعات تصویر سینه از بین برود اثرهای مصنوعی را از زمینه جدا نماییم.

معمولاً از روش سعی و خطا برای تعیین آستانه سراسری<sup>۷</sup> (T) مجموعه تصاویر ماموگرام استفاده می-شود، بطوریکه عملکرد بخش‌بندی مذکور با آستانه‌های ممکن برای همه تصاویر ارزیابی گردد. بررسی چشمی تصاویر ماموگرافی نشان می‌دهد آستانه سراسری (T) باید ۱۸ باشد.

شکل ۴-۴ به ترتیب هیستوگرام تصاویر ماموگرام شکل ۳-۴ را نشان می‌دهد.

---

<sup>1</sup> Reliability

<sup>2</sup> Robustness

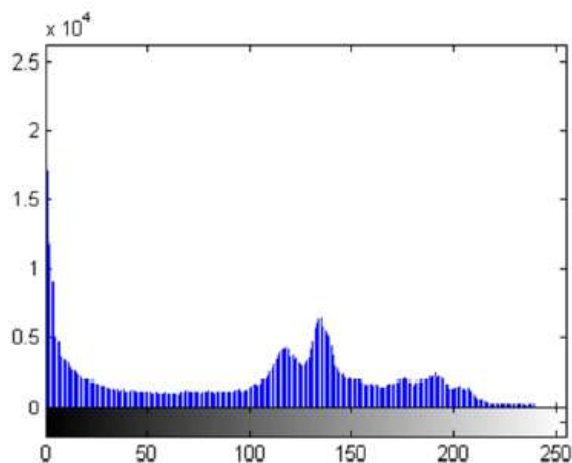
<sup>3</sup> Morphology

<sup>4</sup> Grayscale

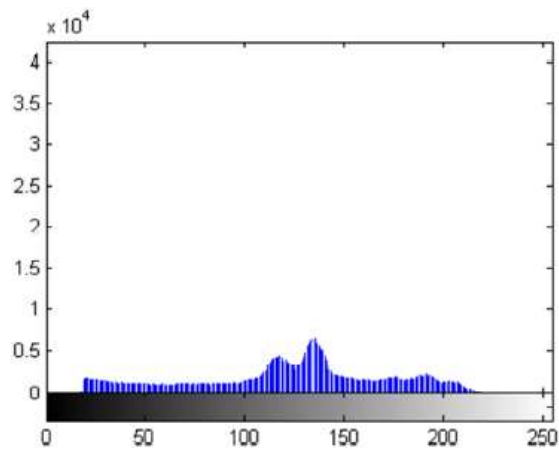
<sup>5</sup> Binary

<sup>6</sup> Thresholding

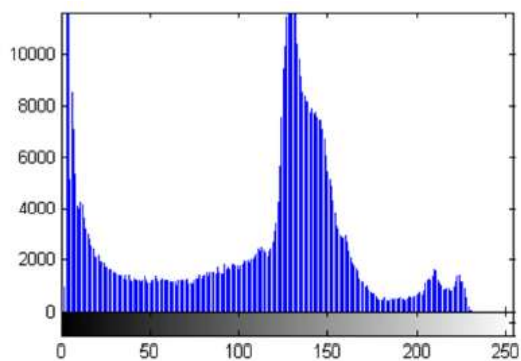
<sup>7</sup> Global



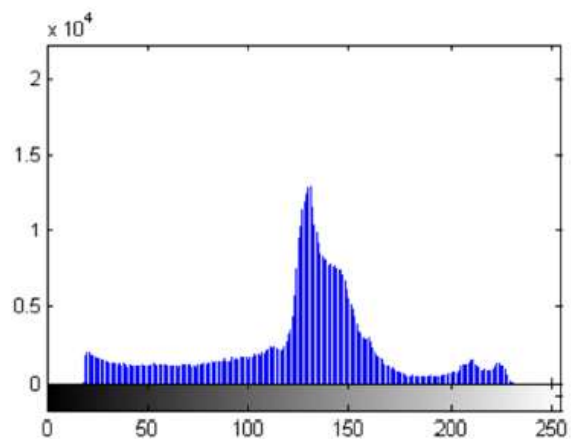
a



b



c



d

شکل ۴-۴: به ترتیب هیستوگرام تصاویر ماموگرام شکل ۴-۳ بعد از اعمال آستانه سراسری  $T=8$ .

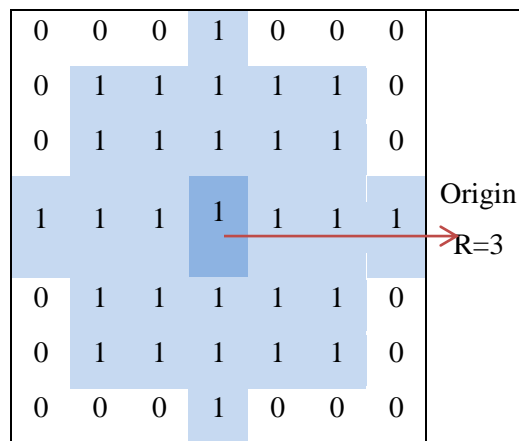
الگوریتم حذف اثر مصنوعی رادیوپیک:

۱. تبدیل تصویر سیاه و سفید به معادل تصویر باینری آن (شکل ۴-۶ b).
۲. شناسایی اجزا بزرگ تصویر (یک‌هایی که با صفر احاطه شده‌اند) و حذف آنها با استفاده از هشت همسایگی برای ماتریس دو بعدی (شکل ۴-۶ c).
- در این همسایگی ۸ در ۸، آرایه‌ی یک که با صفر احاطه شده‌اند به صفر تغییر می‌یابند. یک مثال از پیکسل ایزوله شده در ماتریس  $X$  در رابطه ۴-۱ آورده شده است:

$$X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

۳. الگوریتم دیگری که میتوان جهت نرم تر کردن نویز به تصویر باینری اعمال نمود اینست که همه درایه‌های "۱" تصویر باینری را بررسی کرده و اگر در همسایگی سه در سه آن پنج یا تعداد بیشتری "۱" بود بدون تغییر می‌ماند در غیر این صورت به "۰" تغییر یابد (شکل ۵-۴).

۴. همانگونه که در شکل ۵-۴ نشان داده شده، تصویر دودویی با المان STREL<sup>۱</sup> فرسایش<sup>۲</sup> یافته است (بخش ۳-۲-۲-۲ ببینید). از روش سعی و خطا مقدار بهینه شعاع R برای مجموعه تصاویر ماموگرام R=3 یافت شده است.



شکل ۵-۴: STREL

۵. با المانی مشابه STREL در گام ۴، تصویر دودویی گشوده<sup>۳</sup> شده است (بخش ۳-۲-۲-۱ ببینید).

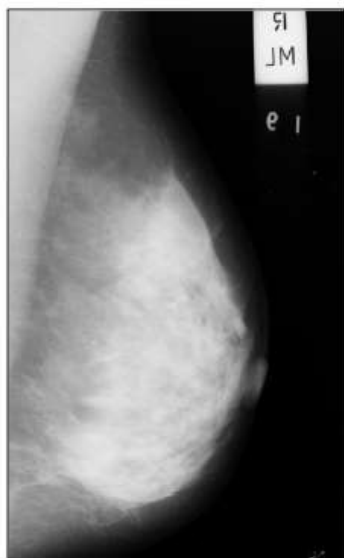
۶. حفره‌های تصویر دودویی با الگوریتم مشابه پر می‌شوند.

۷. در انتها تصویر دودویی بدست آمده را در تصویر سیاه و سفید اصلی ضرب کرده، نتیجه تصویر ماموگرام سیاه و سفید بدون اثرهای مصنوعی خواهد بود (شکل ۶-۴ d).

<sup>1</sup> Flat disk-shaped morphological structuring element

<sup>2</sup> Erode

<sup>3</sup> Dilation



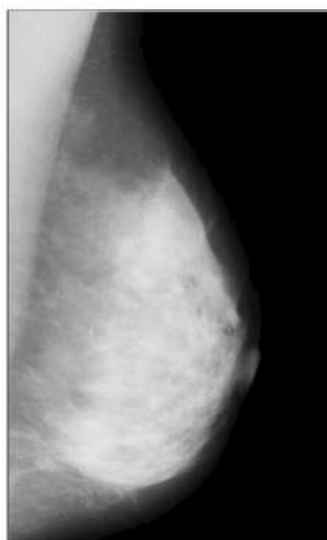
a



b



c



d

شکل ۴-۶: حذف اثرهای مصنوعی. (a) تصویر اصلی همراه با برچسب و علائم نشانه گذاری. (b) تصویر آستانه شده با مقدار  $T=18$ . (c) انتخاب اجزای بزرگ نسبت به کل فضا. (d) تصویر سیاه و سفید پس از حذف اثرهای مصنوعی رادیوپک.

## ۴-۴ قطعه‌بندی تصویر

ماه‌یچه‌های قفسه سینه در تصاویر ماموگرام می‌تواند نتایج و فرآیند تشخیص توده و MCها را تحت تاثیر قرار دهد. تحقیقات نشان می‌دهد جهت تشخیص توده‌ها و MCها مبتنی بر رایانه باید این ماه‌یچه‌ها از تصاویر حذف شود. یافته‌های [۴۹] تاکید می‌کند که هیچ بافت توده یا MC در ناحیه ماه‌یچه‌ها وجود ندارد.

روش پیشنهادی رو به رشد ناحیه کشت<sup>۱</sup> (SRG) در این پایان‌نامه، روش مناسبی برای نیل به هدف قطعه‌بندی ماه‌یچه قفسه سینه است. این الگوریتم در ۴ گام در نرم افزار متلب به اجرا درآمده است.

### گام اول - تعیین جهت شکل سینه:

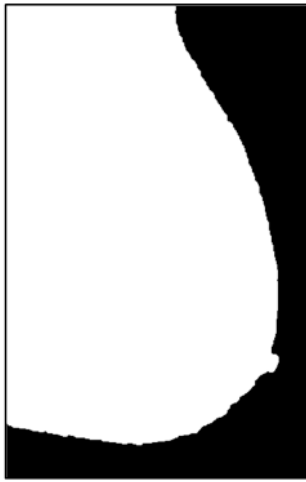
در ابتدا، برای پیاده‌سازی روش SRG لازم است جهت سینه (سمت چپ یا راست) در هر تصویر ماموگرام مشخص شود. موقعی که ماه‌یچه قفسه سینه در بالا گوشه سمت چپ قرار گرفته شکل سینه به سمت راست متمایل است (شکل ۴-۷ a).

الف) جهت فرآیند تعیین جهت شکل سینه ابتدا تصویر باینری را از چپ، راست، بالا و پایین برش می‌دهیم، به این ترتیب شکل سینه در لبه‌ها حاشیه نخواهد داشت (شکل ۴-۷ b و c).

---

<sup>1</sup> Seeded Region Growing





a



b



c



d



e



f

شکل ۴-۷: برش حاشیه‌های شکل سینه. (a) تصویر باینری شکل سینه با جهت راست. (b) برش تصویر باینری شکل a از چپ و راست. (c) برش تصویر باینری شکل a از بالا و پایین. (d) تصویر باینری شکل سینه با جهت چپ. (e) برش تصویر باینری شکل d از چپ و راست. (f) برش تصویر باینری شکل d از بالا و پایین.

(ب) پس از برش حاشیه‌های شکل سینه، مجموع پیکسل‌های شش ستون اول و بار دیگر مجموع پیکسل‌های شش ستون آخر تصاویر دودویی در رابطه‌های ۴-۴ و ۴-۵ محاسبه می‌شود. درایه‌های ماتریس متناظر تصاویر دودویی ماموگرافی در رابطه‌های ۴-۲ و ۴-۳ آمده است.

$$f = \begin{bmatrix} f(0,1) & f(0,2) & f(0,3) & f(0,4) & f(0,5) & \dots & f(0,N) \\ f(1,1) & f(1,2) & f(1,3) & f(1,4) & f(1,5) & \dots & f(1,N) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f(M,1) & f(M,2) & f(M,3) & f(M,4) & f(M,5) & \dots & f(M,N) \end{bmatrix} \quad ۲-۴$$

$$f = \begin{bmatrix} f(0,1) & \dots & f(0,N-4) & f(0,N-3) & f(0,N-2) & f(0,N-1) & f(0,N) \\ f(1,1) & \dots & f(1,N-4) & f(1,N-3) & f(1,N-2) & f(1,N-1) & f(1,N) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f(M,1) & \dots & f(M,N-4) & f(M,N-3) & f(M,N-2) & f(M,N-1) & f(M,N) \end{bmatrix} \quad ۳-۴$$

هر درایه در ماتریس رابطه‌های ۲-۴ و ۳-۴ منطبق بر یک پیکسل تصویر است که با مقدار دودویی صفر یا یک بیان می‌گردد. مجموع مقادیر شش ستون اول رابطه‌های ۴-۴ و ۵-۴ تصاویر باینری شکل ۷-۴ f، طبق رابطه‌های ۴-۴ و ۵-۴ محاسبه می‌گردد.

$$sum_{first} = \left\{ \begin{array}{l} \text{sum}(f(0,1), f(1,1), \dots, f(M, 1)) \\ + \text{sum}(f(0,2), f(1,2), \dots, f(M, 2)) \\ + \text{sum}(f(0,3), f(1,3), \dots, f(M, 3)) \\ + \text{sum}(f(0,4), f(1,4), \dots, f(M, 4)) \\ + \text{sum}(f(0,5), f(1,5), \dots, f(M, 5)) \end{array} \right\} \quad ۴-۴$$

$$sum_{last} = \left\{ \begin{array}{l} \text{sum}(f(0,N-4), f(1,N-4), \dots, f(M, N-4)) \\ + \text{sum}(f(0,N-3), f(1,N-3), \dots, f(M, N-3)) \\ + \text{sum}(f(0,N-2), f(1,N-2), \dots, f(M, N-2)) \\ + \text{sum}(f(0,N-1), f(1,N-1), \dots, f(M, N-1)) \\ + \text{sum}(f(0,N), f(1,N), \dots, f(M, N)) \end{array} \right\} \quad ۵-۴$$

ج) سپس با مقادیر صحیح<sup>۱</sup> رابطه‌های بالا و کد زیر میتوان جهت شکل سینه را تعیین کرد(شکل ۸-۴).

<sup>۱</sup>Integer

```

if (sumfirst > sumlast)
%Breast is oriented towards the right
else if (sumfirst < sumlast )
%Breast is oriented towards the left
end if

```

شکل ۴-۸: کد نوشته شده در متلب جهت تعیین جهت شکل سینه

### گام دوم- بهبود وضوح و کنتراست<sup>۱</sup> تصاویر سینه

برای قطعه‌بندی ماهیچه قفسه سینه لازم است وضوح تصویر سینه و بطور ویژه ناحیه ماهیچه قفسه سینه بهبود یابد. از آنجا که ماهیچه قفسه سینه دارای تعداد بیشتری پیکسل‌های روشن نسبت به ناحیه سینه هستند لازم است وضوح تصویر ناحیه سینه بهبود یابد تا پیکسل‌های سطح خاکستری ناحیه ماهیچه روشن‌تر شوند. با بهبود وضوح تصاویر، الگوریتم SRG بطور قابل اعتمادتری می‌تواند لبه‌های ماهیچه سینه را از شکل سینه تشخیص دهد.

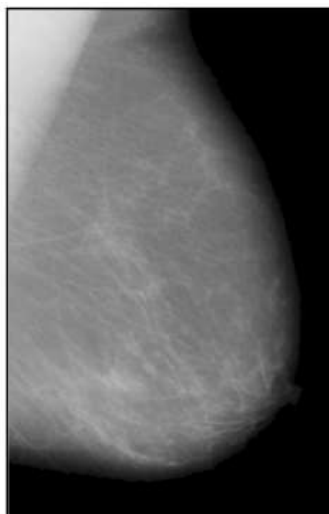
بهبود وضوح تصاویر با یافتن کرانه‌های<sup>۲</sup> کشیدگی وضوح تصویر<sup>۳</sup> انجام می‌شود از آنجا که سطح تلرانس کمیتی اسکالر است عیوب در سطح پایین و بالا اشباع می‌شود. در این تحقیق مقدار پیش فرض تلرانس  $t=[\min \max]$  در نظر گرفته می‌شود که  $\min$ ، کمترین مقدار سطح خاکستری در تصویر ماموگرام و  $\max$ ، بیشترین مقدار سطح خاکستری در تصویر ماموگرام است.

---

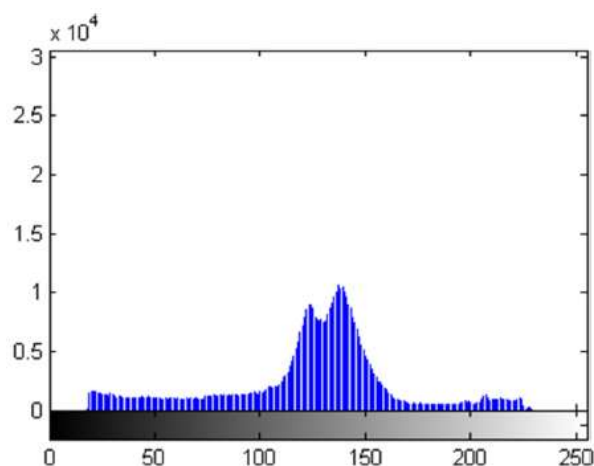
<sup>1</sup> Contrast Enhancement

<sup>2</sup> Limits

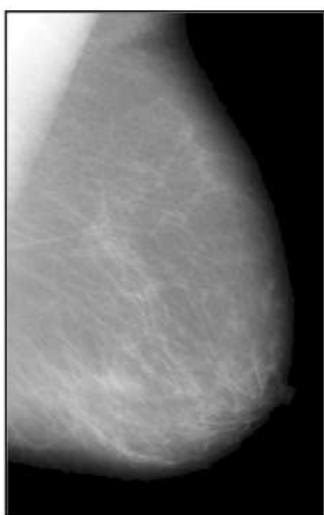
<sup>3</sup> Contrast Stretch



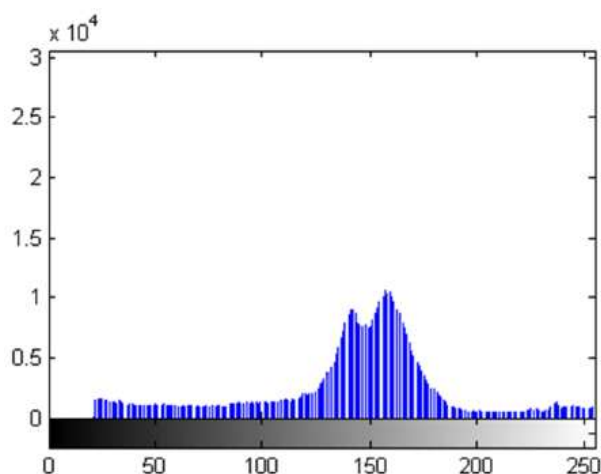
a



b



c



d

شکل ۴-۹: بهبود وضوح تصویر ماموگرام. (a) تصویر ماموگرام بعد از پیش پردازش. (b) هیستوگرام تصویر (c) وضوح

بهبود یافته تصویر الف. (d) هیستوگرام تصویر c.

همانگونه که شکل ۴-۹ نشان می‌دهد هیستوگرام تصویر بهبود یافته نسبت به هیستوگرام تصویر اصلی کشیده‌تر است که به این معناست که پیکسل‌های روشن‌تر افزایش یافته‌اند.

**گام سوم** - ناحیه کشت شده رو به رشد (SRG)

پس از تعیین جهت سینه در گام اول و بهبود وضوح تصویر سینه در گام دوم، روش رو به رشد ناحیه کشت (SRG) [۵۰] جهت قطعه‌بندی ماهیچه قفسه سینه اعمال می‌گردد. به منظور اعمال SRG لازم

است یک کشت درون ماهیچه سینه تصویر ماموگرام قرار گیرد.

با استفاده از نتایج گام‌های اول و دوم، اگر شکل سینه جهت راست داشته باشد یک کشت در بالا گوشه سمت چپ (ماهیچه سینه) تصویر شکل ۴-۱۰a مطابق پیکسل  $f(M,5)$  رابطه ۴-۲ قرار می‌گیرد. بطور مشابه، اگر شکل سینه به سمت چپ متمایل باشد یک کشت در بالا گوشه سمت راست تصویر شکل ۴-۱۰d مطابق پیکسل  $f(M,N-4)$  رابطه ۴-۳ قرار می‌گیرد. در فرآیند SRG:

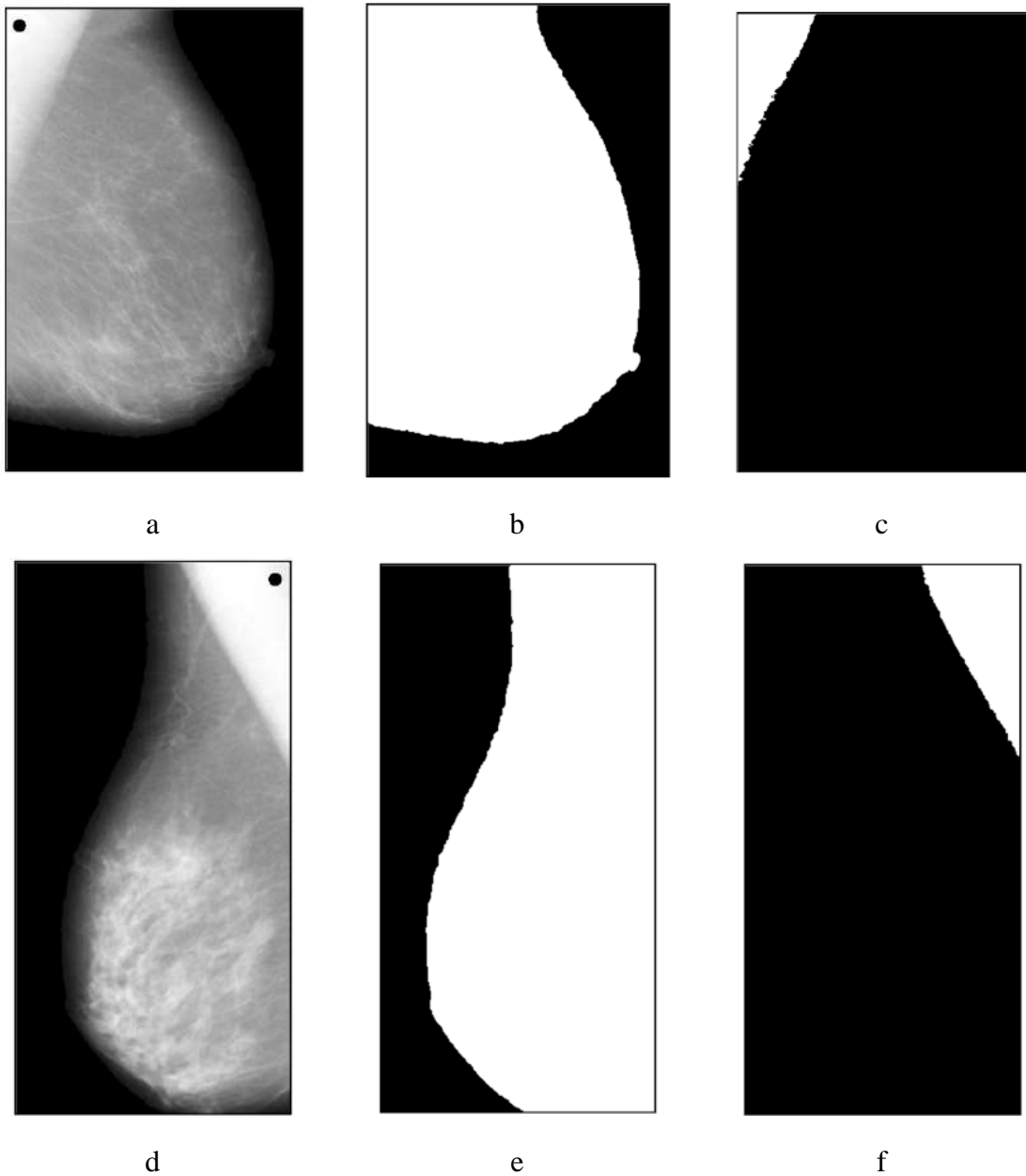
الف) ناحیه بطور تکراری با مقایسه تمام پیکسل‌های مجاور تخصیص نیافته رشد میابد.

ب) تفاوت بین مقدار پیکسل مورد علاقه و میانگین ناحیه به عنوان معیار همگنی استفاده می‌شود.

ج) پیکسل با کمترین مقدار اختلاف به ناحیه مربوطه اختصاص میابد.

د) فرآیند زمانی متوقف می‌شود که اختلاف مقدار بین میانگین ناحیه و پیکسل جدید بزرگتر از مقدار آستانه (بیشترین مقدار اختلاف) باشد.

با استفاده از روش سعی و خطا با بررسی عملکرد قطعه‌بندی تصاویر ماموگرافی برای تمام مقادیر ممکن آستانه ۰ تا ۲۵۵، آستانه مناسب SRG برای تمام تصاویر ماموگرام  $S=18$  تعیین می‌شود. آستانه SRG معمولاً به بیشترین مقدار فاصله در برگزیده تمام تصاویر ماموگرام جهت قطعه‌بندی قابل اعتماد ماهیچه قفسه سینه از شکل سینه ارجاع داده می‌شود. نتیجه قطعه‌بندی ماهیچه قفسه سینه بعد از فرآیند SRG، یک تصویر باینری خواهد بود که در شکل ۴-۱۰c نشان داده شده است.



شکل ۴-۱۰: بخش‌بندی ماهیچه سینه با روش SRG. (a) تصویر بهبود یافته سینه با جهت راست. (b) تصویر باینری شکل a. (c) ماهیچه قطعه‌بندی شده شکل a با روش SRG. (d) تصویر بهبود یافته سینه با جهت چپ. (e) تصویر باینری شکل d. (f) ماهیچه قطعه‌بندی شده شکل d با روش SRG.

سپس تصویر دودویی شکل ۴-۱۰c از تصاویر دودویی سینه متناظرشان در شکل ۴-۱۰b و e تفریق خواهد شد. در نهایت تصویر دودویی شکل سینه باقی خواهد ماند (شکل ۴-۱۱). این کار در نرم افزار متلب با تابع `Imsubtract` قابل اجراست.



a



b

شکل ۴-۱۱: حذف ماهیچه سینه‌ای ناحیه شکل سینه. (a) تفریق تصویر دودویی شکل ۴-۱۰c از شکل ۴-۱۰b. (b)

تفریق تصویر دودویی شکل ۴-۱۰f از شکل ۴-۱۰e.

در بعضی از نمونه‌ها ممکن است بعد از فرآیند قطعه‌بندی اجزای کوچکی مانند بخش‌هایی از ماهیچه نزدیک سینه و یا لبه‌های ماهیچه باقی بماند. این اجزای کوچک باینری که در قطعه‌بندی تصاویر نویز خوانده می‌شوند. عملیات ریخت‌شناسی<sup>۱</sup> از گام ۴ تا ۶ بخش ۳-۲ بر روی تصویر باینری شکل ۴-۱۱ جهت کاهش نویز و اعوجاج، پر کردن حفره‌ها و حذف اجزای کوچک اعمال می‌شود. تصویر باینری بعد از اعمال عملیات ریخت‌شناسی در شکل ۴-۱۲ نشان داده شده است.

با مقایسه شکل ۴-۱۱ و شکل ۴-۱۲ مشاهده می‌شود بعد از اعمال عملیات ریخت‌شناسی بر روی

تصویر باینری، مرز بین ناحیه شکل سینه و ماهیچه سینه هموارتر<sup>۲</sup> شده است.

<sup>۱</sup>Morphological Operation

<sup>۲</sup>Smother



a



b

شکل ۴-۱۲: تصویر دودویی قطعه‌بندی شده سینه. (a) شکل سینه با جهت راست بعد از عملیات ریخت شناسی. (b)

شکل سینه با جهت چپ بعد از عملیات ریخت شناسی.

#### گام چهارم - هموار کردن مرز ماهیچه سینه بخش‌بندی شده

از آنجا که مرز بین ناحیه شکل سینه و ماهیچه قطعه‌بندی شده، لبه‌های ناهموار دارد (شکل

۴-۱۲)، یک روش ساده و موثر را جهت هموارتر کردن این لبه‌ها اعمال خواهیم نمود.

بدین منظور لازم است وتر در شکل ۴-۱۳ با دو نقطه  $(x_1, y_1)$  و  $(x_2, y_2)$  نسبت به نقطه  $(0, 0)$

تعیین شود این نقاط درایه‌های ماتریس متناظر تصویر دیجیتال ماموگرام است که در فرمت (سطر و

ستون) بیان می‌گردد. الگوریتم پیشنهادی برای یافتن دو نقطه  $(x_1, y_1)$  و  $(x_2, y_2)$  برای شکل سینه-

هایی با جهت راست و چپ به بدین ترتیب خواهد بود:

شکل سینه با جهت راست

الف) برای تعیین نقطه  $(x_1, y_1)$  در شکل ۴-۱۲، سطر اول تصویر دودویی شکل مذکور

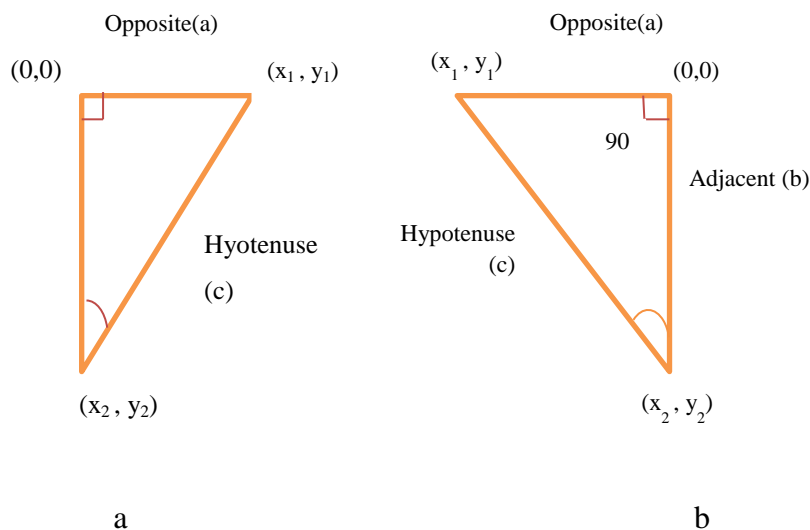
$(f(0,1), f(0,2), \dots, f(0,N))$  در رابطه ۴-۲ برای یافتن اولین "۱" بررسی می‌شود زیرا "۰" بیانگر

پیکسل‌های سیاه در تصویر است. اولین درایه با مقدار "۱" متناظر با نقطه  $(x_1, y_1)$  در شکل ۴-۱۳،



خواهد بود.

ب) بطور مشابه، برای تعیین نقطه  $(x_2, y_2)$  در شکل ۴-۱۲، ستون اول تصویر دودویی شکل مذکور  $(f(0,1), f(2,1), \dots, f(M,1))$  در رابطه ۴-۲ برای یافتن اولین "۱" بررسی می‌شود. اولین درایه با مقدار "۱" متناظر با نقطه  $(x_2, y_2)$  در شکل ۴-۱۳، خواهد بود.



شکل ۴-۱۳: (a) ماهیچه سینه در شکل. که به شکل مثلث قائم الزاویه دیده شده برای سینه با جهت راست. (b) ماهیچه سینه در شکل. که به شکل مثلث قائم الزاویه دیده شده برای سینه با جهت چپ.

شکل سینه با جهت چپ

الف) برای تعیین نقطه  $(x_1, y_1)$  در شکل ۴-۱۲، سطر اول تصویر دودویی شکل مذکور  $(f(0,1), \dots, f(0, N-1), f(0, N))$  در رابطه ۴-۳ برای یافتن اولین "۱" بررسی می‌شود زیرا "۰" بیانگر پیکسل‌های سیاه در تصویر است. اولین درایه با مقدار "۱" متناظر با نقطه  $(x_1, y_1)$  در شکل ۴-۱۳، خواهد بود.

ب) بطور مشابه، برای تعیین نقطه  $(x_2, y_2)$  در شکل ۴-۱۲، ستون آخر تصویر دودویی شکل مذکور  $(f(0, N), f(1, N), \dots, f(M, N))$  در رابطه ۴-۳ برای یافتن اولین "۱" بررسی می‌شود. اولین درایه با مقدار "۱" متناظر با نقطه  $(x_2, y_2)$  در شکل ۴-۱۳، خواهد بود.

بعد از مشخص شدن دو نقطه  $(x_1, y_1)$  و  $(x_2, y_2)$  از مثلث قائم الزاویه در شکل ۴-۱۳، مقدار آفست<sup>۱</sup>  $r$  برای این دو نقطه تعیین می‌گردد، که  $r$  یک مقدار صحیح در محدوده  $0 < r < 30$  می‌باشد. برای تصویر سینه با جهت راست، تعیین مقدار آفست  $r$ ، دو نقطه  $(x_1, y_1)$  و  $(x_2, y_2)$  را به  $(x_1 - r, y_1)$  و  $(x_2, y_2 + r)$  نگاشت می‌دهد. به طور مشابه برای تصویر سینه با جهت چپ، تعیین مقدار آفست  $r$ ، دو نقطه  $(x_1, y_1)$  و  $(x_2, y_2)$  را به  $(x_1 + r, y_1)$  و  $(x_2, y_2 + r)$  نگاشت می‌دهد. سپس معادله خط وتر مثلث قائم الزاویه شکل ۴-۱۳، که بیانگر مرز بین ناحیه سینه و ماهیچه قطعه‌بندی شده است طبق رابطه ۴-۶ زیر بیان می‌گردد:

$$y = mx + c \quad ۴-۶$$

که  $m$  شیب خط و  $c$  مقدار عرض از مبدا است.

$$m = (y_1 - x_1) / (y_2 - x_2) \quad ۴-۷$$

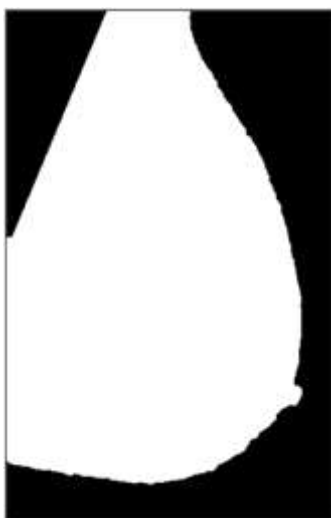
یافتن خط رابطه ۴-۶ برای تصاویر ماموگرام از طریق نرم افزار متلب اجرا شده است.

به منظور هموارتر شدن مرز بین ناحیه شکل سینه و ماهیچه قطعه‌بندی شده، برای تصاویر دودویی ماموگرام سینه با جهت راست، مقادیر پیکسل‌های سمت چپ خط محاسبه شده به صفر تغییر می‌یابد (زمینه سیاه) (شکل ۴-۱۴a).

به طور مشابه، برای تصاویر دودویی ماموگرام سینه با جهت چپ، مقادیر پیکسل‌های سمت راست خط محاسبه شده به صفر تغییر می‌یابد (زمینه سیاه) (شکل ۴-۱۴b).

---

<sup>1</sup> Offset



a



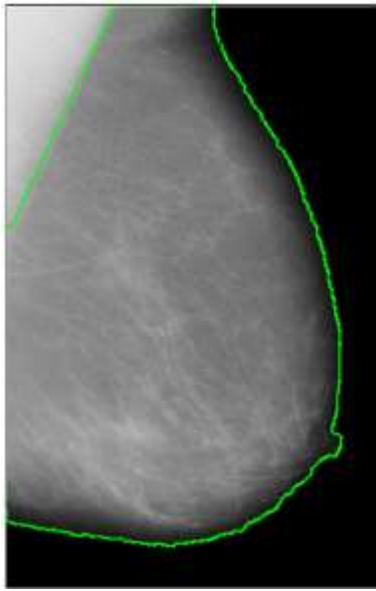
b

شکل ۴-۱۴: (a) شکل سینه با جهت راست بعد از اعمال خط مستقیم ماهیچه سینه. (b) شکل سینه با جهت چپ بعد از اعمال خط مستقیم ماهیچه سینه.

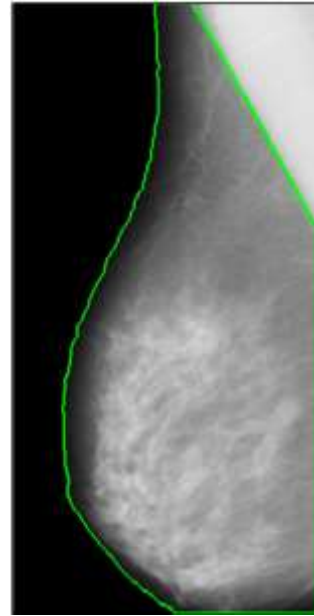
در اینجا نیز عملیات ریخت‌شناسی فرسایش و انبساط بر روی تصویر دودویی شکل ۴-۱۴ برای حذف اجزای کوچک بوجود آمده پس از اعمال خط مستقیم، بکار گرفته می‌شود.

تصویر قطعه‌بندی شده ماموگرام پس از حذف ماهیچه سینه با ضرب تصویر دودویی شکل ۴-۱۴ در تصویر ماموگرام اصلی بدست می‌آید (شکل ۴-۱۵).

هیستوگرام شکل ۴-۱۶ نشان می‌دهد که حذف ماهیچه‌های سینه از تصاویر ماموگرام، تعداد پیکسل‌های روشن‌تر را نسبت به تصویر اصلی کاهش می‌دهد.

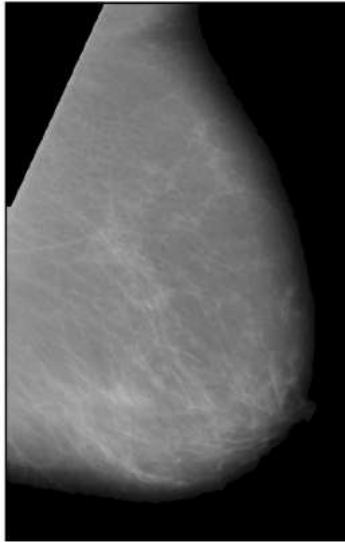


a

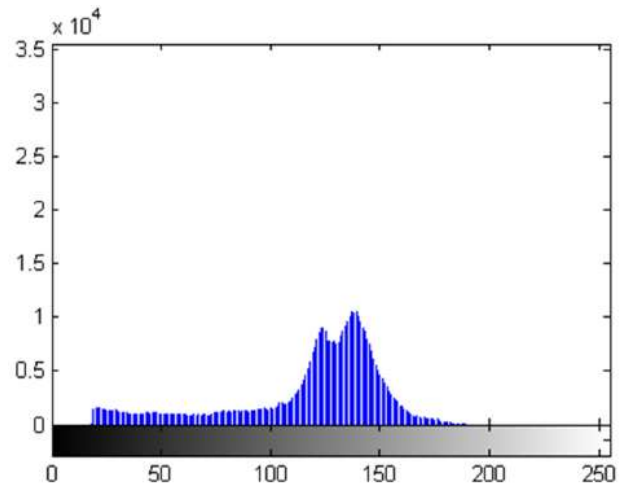


b

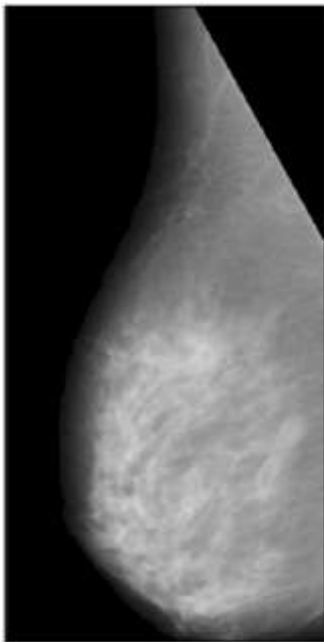
شکل ۴-۱۵: (a) قطعه‌بندی ماهیچه سینه از تصاویر ماموگرام با جهت راست. (b) قطعه‌بندی ماهیچه سینه از تصاویر ماموگرام با جهت چپ.



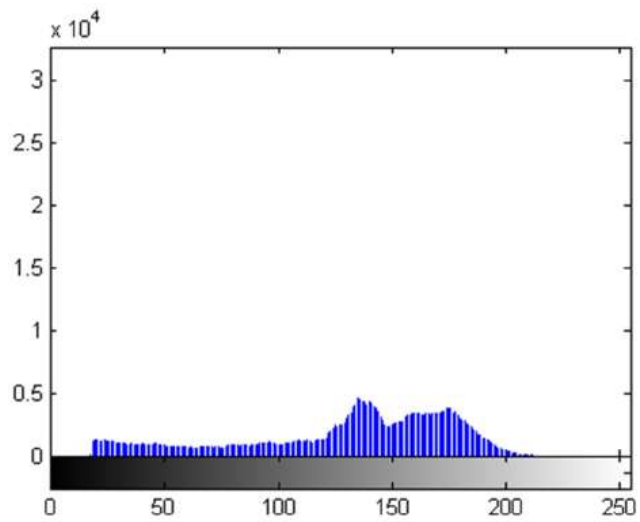
a



B



c



D

شکل ۴-۱۶: (a) تصویر سیاه و سفید ماموگرام با جهت راست پس از حذف ماهیچه سینه. (b) هیستوگرام متناظر تصویر

(c). (c) تصویر سیاه و سفید ماموگرام با جهت چپ پس از حذف ماهیچه سینه. (d) هیستوگرام متناظر تصویر (c).

## ۴-۵ استخراج ویژگی‌های بافت

### ۴-۵-۱ انتخاب ناحیه مورد توجه (ROI)

ویژگی‌ها نمی‌توانند بطور مستقیم از تصاویر ماموگرام بدست آمده در شکل ۴-۱۶ محاسبه شوند زیرا آنها نتایج شناسایی را تحت تاثیر قرار خواهند داد. بنابراین لازم است ویژگی‌ها تنها از نواحی ناهنجاری سینه، به استثنای سایر بخش‌های بی اهمیت بافت سینه محاسبه گردند. به این منظور از تصاویر قطعه‌بندی شده در شکل ۴-۱۶ جهت استخراج ناحیه‌های مورد علاقه (ROI)<sup>۱</sup> یا ناحیه‌های ناهنجار<sup>۲</sup> (نمونه‌های خوش خیم و بدخیم) استفاده می‌شود.

اطلاعات بدست آمده برای نمونه‌های خوش خیم و بدخیم از پایگاه داده معرفی شده به شرح زیر بیان می‌گردد:

۱- مرکز محل ناهنجاری (خوش خیم و بدخیم) در تصاویر با مختصات  $(x,y)$ .

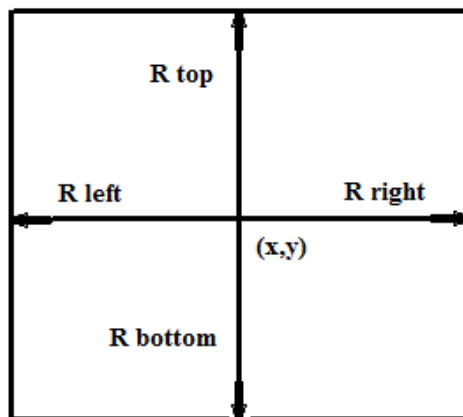
۲- شعاع تقریبی پیکسل‌ها از محیط دایره ناحیه ناهنجاری.

در این پایان‌نامه برای استخراج ناحیه ROI از محل مرکز ناهنجاری استفاده می‌شود که ROI یک ناهنجاری (خوش خیم یا بدخیم) در تصاویر ماموگرافی تعریف می‌شود. از آنجا که ویژگی‌های بافتی GLCM که می‌توانند برای تصاویر محاسبه شوند در قالب ماتریس دو بعدی با  $N$  سطر و  $M$  ستون معرفی می‌شود، ناحیه ROI در قالب یک مربع انتخاب می‌شود.

---

<sup>1</sup> Region Of Interest

<sup>2</sup> Abnormal



شکل ۴-۱۷: استخراج مربع ROI از نمونه‌ها با اندازه‌های متفاوت

استخراج ناحیه‌های ROI از تصاویر قطعه‌بندی ماموگرام شکل ۴-۱۶ با استفاده از شکل ۴-۱۷ انجام می‌گیرد، این گونه که مرکز محل ناهنجاری با مختصات  $(x,y)$  در شکل ۴-۱۷ پیکسل مبدا در نظر گرفته می‌شود. در این پایان‌نامه، برای تخمین اندازه ناهنجاری‌های ماموگرافی با استفاده از یک مربع ROI (شکل ۴-۱۷) بسیار وابسته به در نظر گرفتن قطر ناهنجاری است. برای بدست آمدن بهترین اندازه ROI، قطرهای متفاوت امتحان می‌شود.

با بررسی تصاویر پایگاه داده مورد نظر، حداقل و حداکثر قطر محیط دایره برحسب پیکسل برای تمام ناهنجاری‌های خوش‌خیم و بدخیم به ترتیب ۴۸ و ۱۳۰ پیکسل بدست آمد. با استفاده از این اطلاعات رایج‌ترین اندازه‌های ROI برای پایگاه داده تعیین می‌شود:  $96 \times 96$  pixels,  $110 \times 110$  pixels,  $48 \times 48$  pixels,  $64 \times 64$  pixels,  $128 \times 128$  pixels,  $136 \times 136$  pixels,  $148 \times 148$  pixels

مختصات  $(x,y)$  از مرکز ناهنجاری با  $(0,0)$  بیان می‌شود. به منظور استخراج ناحیه ROI، لازم است مختصات  $(x,y)$  از پایگاه داده با مقادیر آفست بیان شود.

همان گونه که در شکل ۴-۱۷ دیده می‌شود مقادیر آفست بیانگر مربع ROI،  $r_{top}$ ,  $r_{bottom}$ ,  $r_{left}$ ,  $r_{right}$  هستند. مقادیر مزبور برای استخراج ناحیه ROI با اندازه  $48 \times 48$  pixels به شرح زیر است:

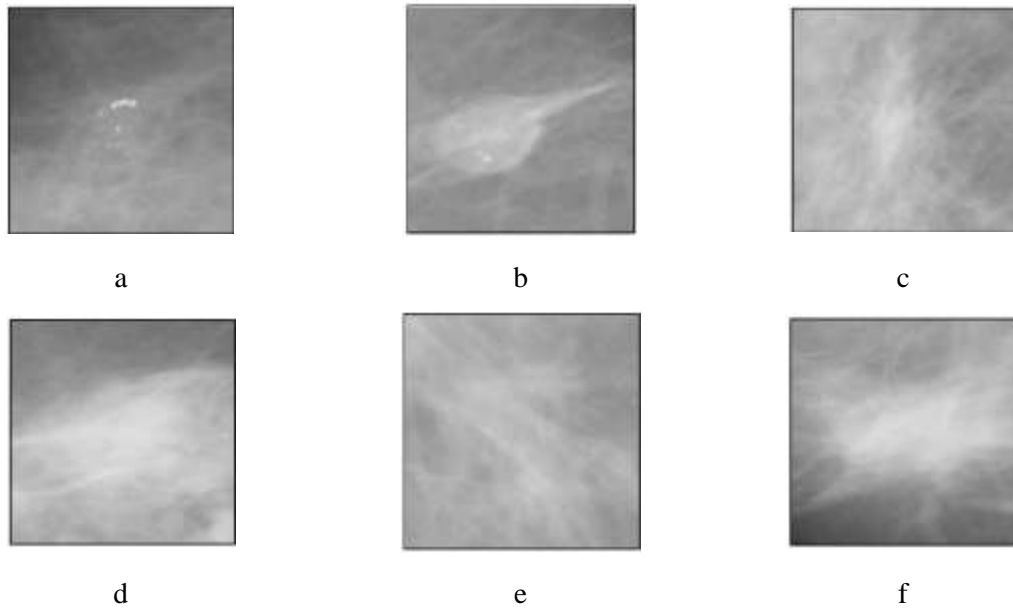
$$r_{top} = (x, y + 23), r_{bottom} = (x, y - 24), r_{left} = (x - 24, y), r_{right} = (x + 24, y)$$

به طور مشابه برای استخراج ناحیه ROI با اندازه  $64 \times 64$  pixels داریم:

$$r_{\text{top}}=(x,y+31), r_{\text{bottom}}=(x,y-32), r_{\text{left}}=(x-32,y), r_{\text{right}}=(x+31,y)$$

همین فرآیند برای سایر اندازه‌های ROI اعمال می‌شود. شکل ۴-۱۸ و شکل ۴-۱۹ نواحی ROI

خوش خیم و بدخیم با روش شکل ۴-۱۷ از تصاویر ماموگرام را نشان می‌دهد.



شکل ۴-۱۸: نواحی ROI از ناهنجاری‌های خوش خیم تصاویر ماموگرام. a) کلسیفیکیشن b) توده مشخص<sup>۱</sup> c) توده

سوزنی شکل<sup>۲</sup> d) توده III-defined e) اعواج در معماری<sup>۳</sup> f) توده نامتقارن<sup>۴</sup>

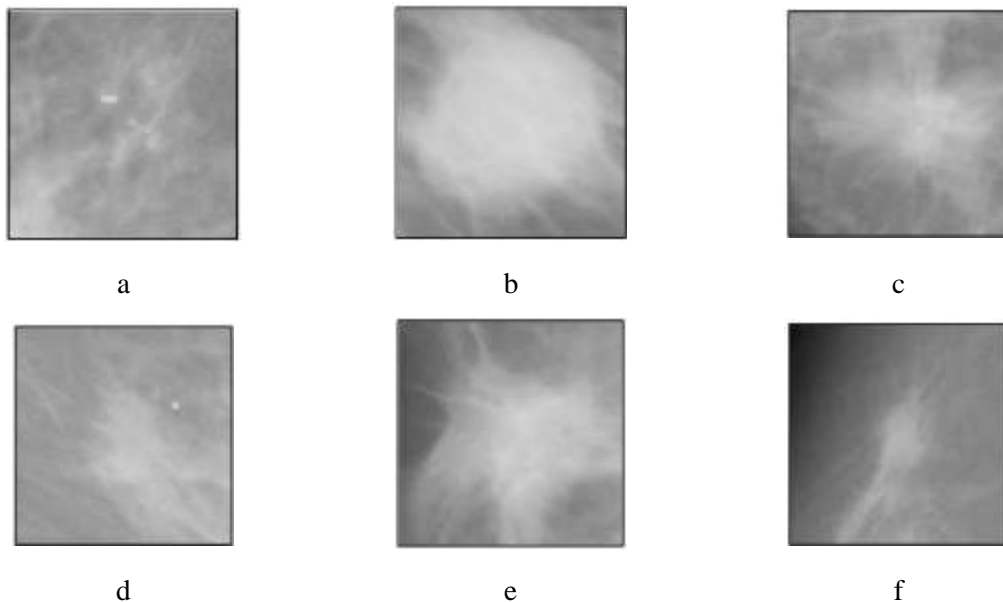
<sup>1</sup> Circumscribe

<sup>2</sup> Spiculate

<sup>3</sup> Architectural Distortion

<sup>4</sup> Asymmetrical





شکل ۴-۱۹: نواحی ROI از ناهنجاری‌های بدخیم تصاویر ماموگرام. (a) کلسیفیکیشن (b) توده مشخص (c) توده سوزنی

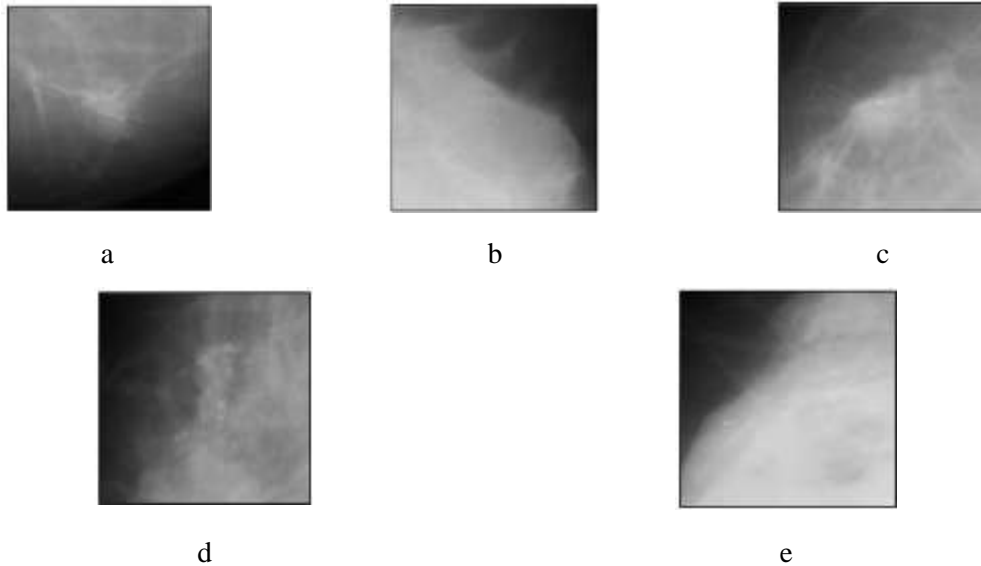
(شکل d) توده III-defined (e) اعوجاج در معماری (f) توده نامتقارن

نواحی ROI با نقاط روشن و یا خوشه‌های کوچک در شکل ۴-۱۸ و شکل ۴-۱۹، MCها و قالب

توده‌ها را نشان می‌دهد.

#### ۴-۵-۱-۱ لزوم پردازش تصاویر ماموگرافی

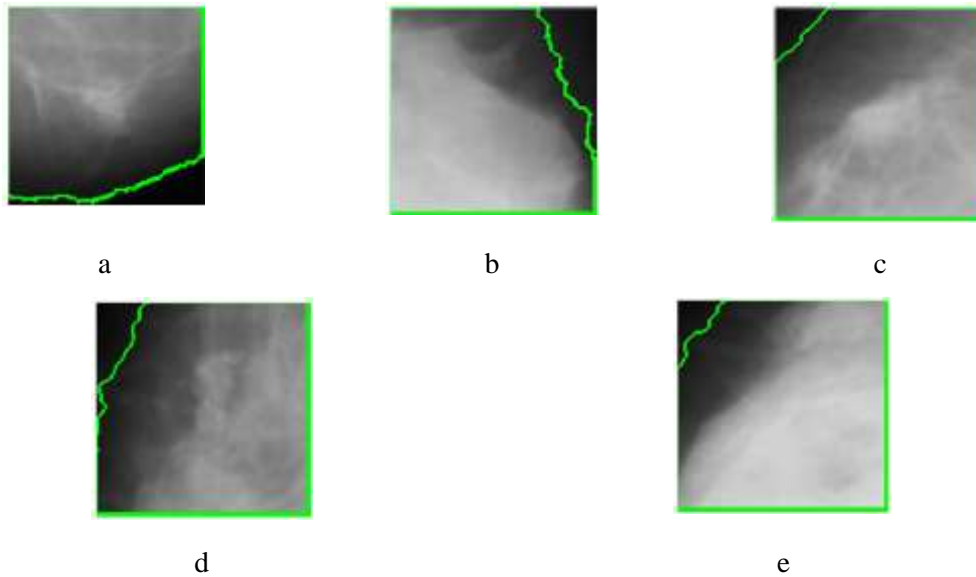
یکی از مهمترین نگرانی‌ها در حین استخراج ROI اینست که بخش کوچکی از ناحیه ROI شامل ناحیه زمینه (پیکسل‌های سیاه) است. این معمولا در تصاویر ماموگرافی زمانی رخ می‌دهد که MCها یا توده‌های خوش‌خیم و بدخیم نزدیک لبه‌های شکل سینه بخش‌بندی شده، یافت می‌شود (شکل ۴-۲۰).



شکل ۴-۲۰: نواحی ROI بدون بخش‌بندی ناحیه زمینه

به منظور استخراج ویژگی‌های بافتی بهینه از تصاویر ROI (شکل ۴-۲۰) با استفاده از GLCM، لازم است پیکسل‌های زمینه در ناحیه ROI حذف شده و همچنین شکل سینه بخش‌بندی گردد. مفاهیم نظری و الگوریتم‌های این دو مهم در بخش‌های قبل توضیح داده شد.

در این تحقیق برای استخراج ویژگی‌های بافتی، تنها از شکل سینه بخش‌بندی شده در ROI نشان داده شده در شکل ۴-۲۱ استفاده شده است. ROI‌های در شکل ۴-۲۱ نشان می‌دهد که ناحیه زمینه (پیکسل‌هایی با مقدار صفر) در شکل سینه از فرآیند استخراج ویژگی‌های بافتی مستثنی شده-اند.



شکل ۴-۲۱: نواحی ROI شکل ۴-۲۰ با بخش‌بندی ناحیه زمینه

#### ۲-۵-۴ استخراج ویژگی‌های بافتی

برای استخراج ویژگی‌های بافتی خوش‌خیم و بدخیم از نواحی ROI شکل ۴-۱۸ و شکل ۴-۱۹ استفاده شده است. همانگونه که در بخش ۳-۳-۳-۱ آمد ماتریس GLCM رایج‌ترین و موفق‌ترین روش برای تحلیل بافت تصاویر ماموگرام شناخته شده است. ابزار پردازش تصاویر نرم افزار متلب قابلیت تحلیل این ویژگی‌های بافتی را دارد. تابع graycoprops در متلب توصیف‌گرهای بافتی آماری از ماتریس GLCM به نام‌های کنتراست<sup>۱</sup>، همبستگی<sup>۲</sup>، انرژی<sup>۳</sup> و یکنواختی<sup>۴</sup> را محاسبه می‌کند.

در مجموع ۱۴ توصیف‌گر بافت در این تحقیق تعریف می‌شود که در جدول ۳-۱ آورده شد [۲۴]. پارامترهای محاسباتی ماتریس GLCM در ادامه بیان خواهد شد.

- تعداد سطوح خاکستری

برای استخراج ویژگی‌های بافتی GLCM، شش سطح خاکستری ۸، ۱۶، ۳۲، ۶۴، ۱۲۸

---

<sup>1</sup> Contrast  
<sup>2</sup> Correlation  
<sup>3</sup> Energy  
<sup>4</sup> Homogeneity

و ۲۵۶ استفاده می‌شود.

- فاصله بین پیکسل‌ها

در این پایان‌نامه، به منظور بیان فاصله بین پیکسل مورد علاقه و پیکسل‌های همسایه در هر ROI،  $d=1$  انتخاب می‌گردد.

- زاویه

در این پایان‌نامه، هشت جهت  $0^\circ$ ،  $45^\circ$ ،  $90^\circ$ ،  $135^\circ$ ،  $180^\circ$ ،  $135^\circ$ ،  $90^\circ$ ،  $45^\circ$ ،  $0^\circ$  برای هر ROI محاسبه می‌گردد.

مجموع ویژگی‌های بافتی GLCM محاسبه شده با استفاده از پارامترهای GLCM که بیان گردید در جدول ۴-۱ آورده شده است. برای هر نمونه ROI، ۶۷۲ ویژگی با استفاده از ۱۴ واصف‌گر بافت محاسبه گردید.

جدول ۴-۱: ویژگی‌های بافتی GLCM استخراج شده از نمونه ROI

No.	GLCM Parameters	No. of Texture feature
1.	Texture Descriptors	14
2.	GLCM Direction ( $0^\circ$ , $45^\circ$ , $90^\circ$ , $135^\circ$ , $180^\circ$ , $135^\circ$ , $90^\circ$ , $45^\circ$ )	8
3.	GLCM Distance between pixels ( $d=1$ )	1
4.	Number of GLCM grey (quantization) levels: 8, 16, 32, 64, 128, 254	6
	<b>Texture feature values calculated for each ROI</b>	<b>672</b>

#### ۳-۵-۴ انتخاب ویژگی‌های بهینه با استفاده الگوریتم دودویی PSO

از آنجا که تعداد ویژگی‌ها زیاد است لازم است برای کارآیی بالاتر طبقه‌بند ویژگی‌های مؤثر انتخاب و بقیه حذف شوند. در این پایان‌نامه، برای انتخاب ویژگی‌های بهینه از الگوریتم دودویی بهینه ازدحام ذرات یا PSO استفاده شده است.

ابتدا با استفاده از توابع موجود در نرم افزار متلب ماتریسی تصادفی شامل فقط اعداد ۰ و ۱ را تولید می‌نماییم که صفر نشانه عدم انتخاب ویژگی و یک نشان دهنده انتخاب ویژگی متناظر می‌باشد. این ماتریس به عنوان جمعیت اولیه در الگوریتم PSO استفاده می‌شود. تعداد سطر این ماتریس برابر تعداد طول بردار ویژگی و تعداد ستون‌ها آن برابر تعداد الگوهای انتخابی می‌باشد. این تعداد نمی‌تواند بزرگ در نظر گرفته شود زیرا باعث می‌شود الگوریتم زمان زیادی را برای رسیدن به پاسخ بهینه صرف کند و همچنین نمی‌توان خیلی کوچک در نظر گرفت تا تنوع الگوی انتخابی اولیه کم نشود. این مقدار دست برنامه‌نویس بوده و عموماً سعی می‌شود بنابر مقدار زمانی که برای محاسبه مقدار برازندگی هر الگو صرف می‌شود این تعداد انتخاب شود. از این به بعد این مقدار را بزرگی توده<sup>۱</sup> می‌نامیم (شکل ۴-۲۲).

1	1	0	0	1	0	...
1	0	1	1	1	0	
0	1	0	1	1	1	
1	0	1	1	0	0	
0	1	0	0	0	1	
0	1	1	0	1	1	
.						
.						

شکل ۴-۲۲: ماتریسی تصادفی برای انتخاب ویژگی

در این ماتریس هر ستون به عنوان یک الگو انتخاب ویژگی مورد استفاده قرار می‌گیرد. با استفاده از این الگوها کل داده‌های آموزش ورودی انتخاب و سپس مقدار برازندگی<sup>۲</sup> برای هر الگو بدست می‌آید، یعنی اگر مقدار Swarm Size=100 باشد طول بردار برازندگی ما نیز برابر ۱۰۰ خواهد بود (شکل ۴-۲۳).

<sup>۱</sup> Swarm Size

<sup>۲</sup> Fitness Value

ماتریس داده‌های انتخاب شده بر اساس الگوی

الگوی انتخابی

ماتریس داده‌های ورودی

انتخابی

12	2	42	50	27	70	20	20	14	19	1
13	0	36	64	48	71	10	13	12	89	0
2	19	43	34	73	94	19	3	22	52	1
40	44	23	28	93	73	34	36	38	31	0
31	38	16	75	24	23	41	66	39	24	1
89	76	15	46	10	44	53	73	42	63	0
17	50	51	46	9	39	62	90	53	17	1
0	18	29	43	34	36	77	81	87	97	0

شکل ۴-۲۳: نحوه انتخاب ویژگی بر اساس الگوی مورد نظر

پس از بدست آوردن مقدار برازندگی برای هر الگو مقدار بهترین شخصی  $P\_best$  و بهترین سراسری  $G\_best$  را بدست می‌آوریم و در حالت دقیقتر از  $L\_best$  استفاده می‌نماییم و سپس ماتریس اولیه سرعت را نیز که به اندازه ماتریس جمعیت اولیه می‌باشد ایجاد نموده در این مورد ما مقدار ماکزیمم و مینیمم ماتریس را به بازه (۰ و ۱) محدود نموده‌ایم.

#### ۱-۳-۵-۴ تابع برازندگی

از تابع برازندگی<sup>۱</sup> در الگوریتم‌های بهینه سازی برای هدایت کل الگوریتم به پاسخ بهینه کل استفاده می‌شود. برای این مورد از توابع مختلفی تا به امروز استفاده شده است. در اینجا از طبقه‌بند SVM کتابخانه LIBSVM برای محاسبه تابع برازندگی استفاده نموده‌ایم. بدین گونه که برای هر الگوی انتخابی با استفاده از طبقه‌بند SVM مقدار تشخیص محاسبه شد و سپس ترکیبی از مقدار نرخ

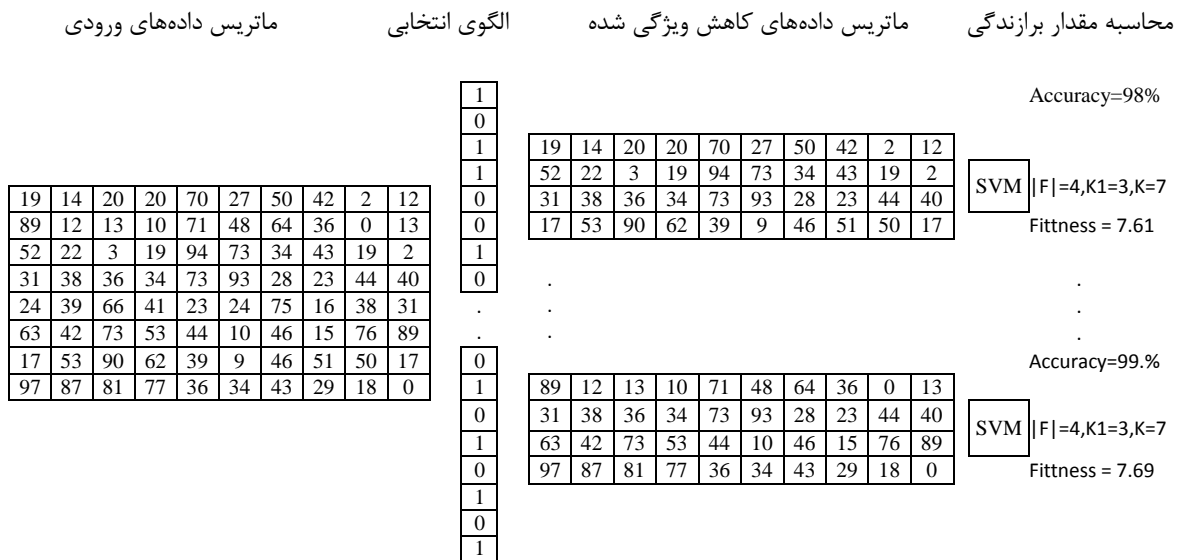
<sup>۱</sup> Fitness function

تشخیص و تعداد ویژگی را بر اساس رابطه ۴-۸ برای مقدار برازندگی استفاده نمودیم.

$$\text{Fitness}(i) = \frac{k_1}{|F|} + k_2 \times \text{Accuracy} \quad ۸-۴$$

در این رابطه مقدار  $K_1$  و  $K_2$  اعداد ثابت اختیاری و  $|F|$  برابر تعداد ویژگی‌های انتخابی است. در

شکل ۴-۲۴ مراحل محاسبه مقدار برازندگی نشان داده شده است.



شکل ۴-۲۴: محاسبه مقدار برازندگی بر اساس الگوها

در ادامه الگویی که بهترین مقدار برازندگی را کسب کرده را به عنوان الگوی بهینه سراسری  $G_{best}$  و کل الگوها را به عنوان بهینه شخصی  $P_{best}$  برداشته و برای تکرار بعدی مورد استفاده قرار می‌گیرند. در تکرار بعد مقدار  $P_{best}$  بدست آمده در تکرار قبلی به عنوان جمعیت اولیه یا در اینجا الگوی انتخابی اولیه مورد استفاده قرار می‌گیرد. سپس الگوها بر اساس رابطه‌های ۳-۷ و ۳-۸ به روزرسانی شده و در این مرحله کل الگوها بر اساس سرعت و پارامترهای اینرسی،  $C_1$  و  $C_2$  انتخابی به سمت الگوی بهتر یعنی  $G_{best}$  حرکت می‌کنند.

همان گونه که بیان شد ما برای انتخاب ویژگی از الگوریتم BPSO استفاده نمودیم در این الگوریتم اساس کار به این صورت است که پس از به روزرسانی سرعت قبلی مقدار سرعت را با استفاده از تابع سیگموئید بین ۰ و ۱ محدود نموده و سپس جمعیتی بر اساس این سرعت به روز شده با درایه‌های صفر و

یک ایجاد می‌نماییم. در هر تکرار پس از بدست آوردن  $P\_best$  و مقدار برازندگی، مقدار برازندگی جدید را با برازندگی قبلی مقایسه می‌نماییم برای هر درایه از بردار برازندگی اگر برازندگی بدست آمده بزرگتر از درایه متناظر قبلی خود باشد الگوی مربوط به آن درایه را از ماتریس الگوی  $P\_best$  جدید انتخاب نموده و با الگوی متناظر خود در  $P\_best$  قبلی جایگزین می‌نماییم. همچنین پارامتر SVM جدید متناظر با آدرس این الگو را نیز انتخاب و با پارامتر متناظرش تعویض می‌نماییم. این مراحل تا زمانی که شرط توقف برقرار نشده ادامه می‌یابد. امکان دارد الگوریتم ویژگی‌ها را بیش از حد مورد نیاز کاهش دهد و در نتیجه کاهش قابل توجهی در نرخ تشخیص داشته باشیم که بایستی با انتخاب دقیق مقادیر  $K_1$  و  $K_2$  در تابع برازندگی و شرط توقف به این مورد توجه کنیم. در نهایت پس از پایان الگوریتم و بدست آوردن بهترین الگوی انتخابی از داده‌های ورودی، در مرحله آزمایش، داده‌ها را با این الگو انتخاب نموده و سپس عمل کلاسه‌بندی را با ویژگی‌های کاهش یافته بدست ادامه می‌دهیم. در این بررسی پارامترهای مربوط به PSO بر اساس جدول ۴-۲ می‌باشد.

جدول ۴-۲: پارامترهای انتخاب شده برای PSO

تعداد ذرات	۶۷۲
ضریب $C_1$ و $C_2$	۲
ضریب $K_1$	۳
ضریب $K_2$	۶
تعداد تکرار	۷۰

#### ۴-۵-۴ نرمال‌سازی ویژگی

قبل از اعمال ویژگی‌ها به کلاسه‌بند لازم است ویژگی‌ها در بازه صفر تا یک نرمال‌سازی<sup>۱</sup> شوند (رابطه ۴-۹).

<sup>۱</sup> Normalization



$$NF(x) = \frac{F(x) - \min(F(x))}{\max(F(x)) - \min(F(x))}$$

۹-۴

در رابطه مزبور  $F(x)$  مقدار ویژگی‌ها می‌باشد.

#### ۵-۵-۴ کلاسه‌بندی

به دلیل اینکه دو کلاس بیمار و سالم داریم در ابتدا به سراغ استفاده از SVM می‌رویم اما به جهت مقایسه، نتایج بدست آمده از شبکه عصبی را نیز منظور می‌کنیم. همانگونه که بیان شد از هر دو پایگاه داده، ۷۰٪ تصاویر برای آموزش<sup>۱</sup> سیستم و ۳۰٪ برای آزمایش<sup>۲</sup> به طور تصادفی انتخاب شده‌اند. از ۱۲۰ تصویر ماموگرافی از که هر یک پایگاه داده DDSM و MIAS به صورت تصادفی گرفته شد، تعداد ۱۳۰ عدد ROI شامل خوشه‌های MC استخراج گردید.

نتایج حاصل از دو کلاسه بند شبکه عصبی و SVM روی پایگاه داده DDSM و MIAS با توجه به انتخاب ویژگی‌های بهینه در جدول ۳-۴ و جدول ۴-۴ نشان داده شده است. نتایج نشان می‌دهد نرخ دقت این سیستم‌ها بین ۸۲٪ تا ۹۶٪ است.

جدول ۳-۴: نتایج حاصل از کلاسه‌بندی‌های مختلف روی پایگاه داده DDSM

روش انتخاب ویژگی بهینه	شبکه عصبی	SVM
--	۸۵,۸۷	۸۸,۸۸
PSO	۹۰,۵	۹۳,۵

<sup>۱</sup> Train

<sup>۲</sup> Test

جدول ۴-۴: نتایج حاصل از کلاس‌بندی‌های مختلف روی پایگاه داده MIAS

روش انتخاب	شبکه عصبی	SVM
ویژگی بهینه	۸۲,۶۱	۸۷,۸۷
PSO	۹۱,۳	۹۶,۵

در فصل سوم برخی از معیارهای کارایی سیستم‌های تشخیص بیماری گفته شد. این معیارها معمولاً در دسته‌بندی‌های دو کلاسه انجام می‌شود و نشان دهنده قدرت و کارایی سیستم مورد بحث می‌باشد. در جدول ۴-۵ و جدول ۴-۶ با استفاده از PSO و کلاس‌بندی‌های مختلف مقایسه این روش‌ها و معیارهای کارایی سیستم آورده شده است.

جدول ۴-۵: پارامترهای کارایی سیستم با دسته‌بندی‌های مختلف بدون انتخاب ویژگی‌های بهینه

پایگاه داده	دسته‌بند	Sensitivity	Specificity	Precision	BR	Accuracy	Error
DDSM	SVM	۰,۷	۰,۹۵	۰,۸۸	۰,۸۳	۰,۸۸	۰,۱۴
	ANN	۰,۸۸	۰,۹۵	۰,۸۸	۰,۹۲	۰,۸۵	۰,۰۷
MIAS	SVM	۰,۵۳	۰,۹۲	۰,۷۳	۰,۷۳	۰,۸۷	۰,۱۱۹
	ANN	۰,۶	۰,۹۵	۰,۸۲	۰,۷۸	۰,۸۲	۰,۱۵

جدول ۴-۶: پارامترهای کارایی سیستم با دسته‌بندی‌های مختلف با انتخاب ویژگی‌های بهینه با روش PSO

پایگاه داده	دسته‌بند	Sensitivity	Specificity	Precision	BR	Accuracy	Error
DDSM	SVM	۰,۶۳	۰,۹۴	۰,۸۳	۰,۷۹	۰,۹۰	۰,۱۵
	ANN	۰,۷۱	۰,۹۵	۰,۸۳	۰,۸۳	۰,۹۳	۰,۱۱
MIAS	SVM	۰,۸	۱	۱	۰,۹	۰,۹۱	۰,۱۱
	ANN	۰,۸	۱	۱	۰,۹	۰,۹۶	۰,۱۱

فصل پنجم:

نتیجہ گیری و پیشنهادات

## ۵ فصل پنجم

### ۱-۵ نتیجه گیری

در این پایان نامه به مسئله تشخیص یک نوع سرطان سینه بنام MCها پرداخته شد. در حالت کلی یک سیستم بازشناسی الگو، یا تشخیص (چهره، اثر انگشت، رگ انگشت، رگ دست، بیماری، ...)، شامل مراحل پیش پردازش، استخراج ویژگی و کلاسه بندی است. در این خصوص مروری بر کارهای گذشته انجام شد.

برای ارزیابی کارایی سیستم، ابتدا پایگاه داده را به طور مفصل در فصل ۴ شرح داده ایم. روش های به کار گرفته شده را روی تمامی داده های این پایگاه داده امتحان کرده ایم. نتایج به صورت کامل در فصل چهارم آورده شده است.

در این پایان نامه دو روش اتوماتیک جهت تفکیک افراد دارای اختلال و افراد سالم ارائه گردید. روش پیشنهادی اول، بعد از پیش پردازش تصاویر با استخراج ویژگی های بافت تصاویر، اقدام به کلاسه بندی افراد می کنیم. در پیشنهاد دوم بعد از پیش پردازش به سراغ انتخاب ویژگی ها با روش BPSO می رویم. این روش در بسیاری از روش های تشخیص از روی بافت تصاویر مثل تشخیص چهره یا تشخیص رگ انگشت، با پیچیدگی محاسباتی کمی که دارد ابزاری بسیار قدرتمند است و ویژگی های بسیار خوبی از تصاویر در بدترین شرایط محیطی و روشنایی استخراج می کند. در مرحله آخر، در روش دوم از کلاسه بند ماشین بردار پشتیبان استفاده می کنیم که یک تفکیک کننده قوی می باشد.

به طور کلی بهترین نتیجه ای که روش اول می تواند برای تمییز و تشخیص داده های ماموگرافی افراد سالم از ناسالم ارائه کند بین  $82\%$  تا  $88\%$  می باشد و بهترین نتیجه ای که روش دوم می تواند برای تشخیص داده های ماموگرافی افراد سالم از ناسالم ارائه کند بین  $90\%$  تا  $96\%$  می باشد.

به طور کلی دستاوردهای این تحقیق را می توان به صورت زیر خلاصه نمود:

۱. استفاده از بافت تصاویر ماموگرافی در یک بعد، برای استخراج ویژگی‌ها و برخورد با این تصاویر مانند تصاویر تشخیص چهره یا تشخیص رگ انگشت که در خور توجه است.
۲. این تحقیق به صورت پایان‌نامه، نتایج قابل قبولی برای ارائه به پزشکان و متخصصان ارائه می‌دهد. در واقع در این تحقیق، روش جدیدی برای طبقه‌بندی و تشخیص MCها ارائه شده است. در این روش، دسته بندی بر اساس اطلاعات ریخت شناسی با استفاده از داده‌های تصویر برداری اشعه x انجام شده است.
۳. فهمیدن و تشخیص به موقع این اختلال برای بانوان در سنین پایین‌تر بسیار مقرون به صرفه است. پزشکان می‌توانند با دانستن اختلال در این بانوان، مشکل را تحت کنترل گیرند و از عواقب بدی که برخی از داروها و محرک‌ها در آینده در سنین بالاتر در این افراد می‌گذارند با تشخیص زود هنگام اختلال و کنترل داروها و دز موثر داروها توسط پزشکان و متخصصان و همچنین مساعدت خانواده‌ها جلوگیری کرد.

## ۲-۵ پیشنهادها

پیشنهاداتی که بعد از انجام این کار به آن می‌توان توجه داشت به قرار زیر است:

۱. همانطور که در این تحقیق به آن اشاره شد، با توجه به نتایجی که بر اساس اطلاعات ریخت شناسی از این کار حاصل شده می‌توان به کمک پزشک یا تیمی از پزشکان این نواحی را مشخص نمود و سینه را بخش‌بندی کرد و به طور خاص روی بخش‌های درگیر سینه در این اختلال پرداخت. این کار باعث کم شدن حجم محاسبات می‌شود. همچنین پیچیدگی محاسباتی و حجم حافظه اشغال شده توسط سیستم پایین می‌آید و کارایی سیستم برای محاسبات بالا می‌رود.
۲. در حل این مسئله دو کلاس وجود دارد باید سیستم آنرا تشخیص دهد اما می‌توان کلاس دیگری برای آن تعریف کرد که این کار در کارهای آتی قرار می‌گیرد. در واقع در این کار می‌-

توان کلاس سومی تعریف کرد که نه کلاس افراد بیمار باشد و نه کلاس افراد سالم باشد. این کلاس، کلاس افراد مشکوک می‌باشد. در واقع بدترین حالت در تشخیص زمانی است که سیستم فرد سالمی را بیمار تشخیص دهد. با تعریف این کلاس سوم، سیستم فردی را که به بیمار یا سالم بودن آن شک دارد، با این روش رد<sup>۱</sup> می‌کند و در کلاس سوم قرار می‌دهد. مزیت این روش این است که ممکن است فرد مورد نظر در کلاس سوم دچار اختلالات دیگر باشد و توسط متخصصان مورد معاینات بالینی قرار گیرد.

---

<sup>1</sup> Rejection approach

- [1] “<https://www.wikipedia.org/>.” .
- [2] R. Siegel, D. Naishadham, and A. Jemal, “Cancer Statistics , 2012,” *Cancer J. Clin.*, vol. 62, no. 1, pp. 10–29, 2012.
- [3] B. Boyle, P. and Levin, “World Cancer Report 2008 International Agency for Research on Cancer, Lyon, France.”
- [4] “<http://crc.tums.ac.ir/Default.aspx?tabid=266>.” .
- [5] I. Harirchi, S. Azary, A. Montazeri, S. M. Mousavi, Z. Sedighi, G. Keshtmand, Z. Zarinkolah, and S. M. A. Khalifeh, “Literacy and Breast Cancer Prevention : a Population-Based Study from Iran,” *Asian Pacific J. Cancer Prev.*, vol. 13, pp. 3927–3930, 2012.
- [6] A. Mohd Khuzi, R. Besar, W. Wan Zaki, and N. Ahmad, “Identification of masses in digital mammogram using gray level co-occurrence matrices,” *Biomed. Imaging Interv. J.*, vol. 5, no. 3, p. e17, Jul. 2009.
- [7] S. Timp and N. Karssemeijer, “A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography .,” *Med. Phys*, vol. 31, pp. 958–971, 2004.
- [8] M. Bozek, J. Delac, K. and Grgic, “Computer-Aided Detection and Diagnosis of Breast Abnormalities in Digital Mammography,” *Proc. 50th Int. Symp. ELMAR-2008, Zadar, Croat.*, pp. 45–52, 2008.
- [9] J. Tang, S. Member, R. M. Rangayyan, J. Xu, and I. El Naqa, “Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.
- [10] J. Bozek, M. Mustra, K. Delac, and M. Grgic, “A Survey of Image Processing Algorithms in Digital Mammography,” *Pattern Recognit.*, pp. 631–657, 2009.
- [11] A. O. Malagelada, “Automatic Mass Segmentation in Mammographic Images,” *G I R O N A U n i v e r s i t y , P H D T e s i s , 2 0 0 7 .*
- [12] S. van Engeland, “Detection of Mass Lesions in Mammograms by Using Multiple Views,” *Radboud Universiteit Nijmegen, PHD Tesis, 2006.*

- [13] C. Sciences, "Detection of Microcalcification Clusters in Digitized X-ray Mammograms Using Unsharp Masking and Image Statistics," *Turkish J. Electr. Eng. Comput. Sci.*, pp. 2048–2061, 2013.
- [14] I. Zyout, "Classification of Clustered Microcalcifications in Mammograms using Particle Swarm Optimization and Least-Squares Support Vector Machine.," *Int. J. Comput. Appl.*, vol. 59, 2012.
- [15] ف. تاج نیا، "استفاده از روشهای آماری و آنالیز Wavelet جهت تشخیص تومورهای سرطانی در تصاویر ماموگرافی،" دانشکده مهندسی برق و الکترونیک، دانشگاه صنعتی شاهرود، پایان نامه ارشد، ۱۳۹۰.
- [16] K. Geethal and K. T. A. Kishore, "New Particle Swarm Optimization for Feature Selection and Classification of Microcalcifications in Mammograms Method Anlyl," *IEEE-International Conf. Signal Process. Netw. Madras Inst. Technol. Anna Univ. Chennai India*, vol. 4, no. 6, pp. 458–463, 2008.
- [17] S. S. Basha and K. S. Prasad, "Automatic Detection of Breast Cancer Mass in Mammograms using Morphological Operators and fuzzy C – Means Clustering," *J. Theor. Appl. Inf. Technol.*, pp. 704–709, 2009.
- [18] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*, vol. 2. Gatesmark Publishing Knoxville, 2009.
- [19] V. Märgner and H. El Abed, *Guide to OCR for arabic scripts*. Springer, 2012.
- [20] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 24, no. 7, pp. 971–987, 2002.
- [21] C. Chen, *Handbook of pattern recognition and computer vision*. World Scientific, 2009.
- [22] J. M. Sanches, "Tissue Characterization by Image Analysis for Diagnostic Purposes," *Course Notes*, p. 25, 2012.
- [23] M. Lyra, S. Lyra, B. Kostakis, C. Georgosopoulos, and K. Skouroliakou, "Digital Mammography texture analysis by computer assisted image processing," *Proc. IEEE Int. Work. Imaging*, pp. 12–15, 2008.
- [24] I. Haralick, R.M. Shanmugam, K. and Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.
- [25] Haralick and R. M. K, "Statistical and Structural Approaches to Texture," *Proc. IEEE*, vol. 67, no. 5, 1979.
- [26] A. M. Sabu and D. N. Ponraj, "Textural Features Based Breast Cancer Detection : A Survey," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 3, no. 9, pp. 1329–1334, 2012.



- [27] M. Pietikäinen, *Computer vision using local binary patterns*, vol. 40. Springer, 2011.
- [28] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [29] G. Zhao and M. Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 29, no. 6, pp. 915–928, 2007.
- [30] J. Kennedy and R. Eberhart, “Particle Swarm Optimization,” *Proceedings., IEEE Int. Conf. Neural Networks, Perth, WA.*, vol. 4, pp. 1942–1948, 1995.
- [31] R. Poli, J. Kennedy, and T. Blackwell, “Particle swarm optimization,” *Swarm Intell. Sci.*, vol. 1, no. 1, pp. 33–57, Aug. 2007.
- [32] A. Tandan, R. Raja, and Y. Chouhan, “Image Segmentation Based on Particle Swarm Optimization Technique,” *Int. J. Sci. Technol. Res.*, vol. 3, no. 2, pp. 257–260, 2014.
- [33] M. Jakubcová, P. Máca, and P. Pech, “A Comparison of Selected Modifications of the Particle Swarm Optimization Algorithm,” *J. Appl. Math.*, vol. 2014, pp. 1–11, 2014.
- [34] J. Kennedy and R. C. Eberhart, “A Discrete Binary Version of the Particle Swarm Algorithm,” *Present. Present. IEEE Int. Conf. Syst. Man, Cybern.*, pp. 4–8, 1997.
- [35] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [36] M. Martínez-Ramón, J. L. Rojo-Álvarez, G. Camps-Valls, J. Muñoz-Marí, Á. Navia-Vázquez, E. Soria-Olivas, and A. R. Figueiras-Vidal, “Support vector machines for nonlinear kernel ARMA system identification,” *Neural Networks, IEEE Trans.*, vol. 17, no. 6, pp. 1617–1622, 2006.
- [37] S. H. Duda, B. Pusich, G. Richter, P. Landwehr, V. L. Oliva, A. Tielbeek, B. Wiesinger, J. B. Hak, H. Tielemans, and G. Ziemer, “Sirolimus-eluting stents for the treatment of obstructive superficial femoral artery disease six-month results,” *Circulation*, vol. 106, no. 12, pp. 1505–1509, 2002.
- [38] Q.-A. Tran, X. Li, and H. Duan, “Efficient performance estimate for one-class support vector machine,” *Pattern Recognit. Lett.*, vol. 26, no. 8, pp. 1174–1182, 2005.
- [39] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

- [40] M. T. Mahmoudi, N. Forouzideh, C. Lucas, and F. Taghiyareh, "Artificial neural network weights optimization based on imperialist competitive algorithm," in *7th International conference on computer science and information technologies (CSIT'09), Yerevan, 2009*, pp. 244–247.
- [41] S. S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall Englewood Cliffs, NJ, 2007.
- [42] N. M. Abdel Razek, W. a. Yousef, and W. a. Mustafa, "Microcalcification detection with and without CAD system (LIBCAD): A comparative study," *Egypt. J. Radiol. Nucl. Med.*, vol. 44, no. 2, pp. 397–404, Jun. 2013.
- [43] A. Oliver, A. Torrent, X. Lladó, M. Tortajada, L. Tortajada, M. Sentís, J. Freixenet, and R. Zwigelaar, "Automatic Microcalcification and Cluster Detection for Digital and Digitised Mammograms," *Knowledge-Based Syst.*, vol. 28, pp. 68–75, Apr. 2012.
- [44] "<http://marathon.csee.usf.edu/Mammography/Database.html>."
- [45] "<http://peipa.essex.ac.uk/info/mias.html>."
- [46] A. M. Aboshosha, M. Ragaa, and B. A. Mashad, "Improved Auto Detection of Calcification in Digital Mammogram Relying on Multi-Stage CAD Technique," *Eng. Sci.*, vol. 22, no. 2, pp. 171–184, 2011.
- [47] K. Czaplicka and J. Włodarczyk, "Automatic Breast line and Pectoral Muscle Segmentation," *Comput. Sci.*, vol. 20, pp. 195–209, 2012.
- [48] R. D. Yapa and K. Harada, "Breast Skin-Line Estimation and Breast Segmentation in Mammograms using," *Int. J. Medical, Heal. Pharm. Eng.*, vol. 1, no. 5, pp. 217–225, 2007.
- [49] N. Nicolaou, S. Petroudi, J. Georgiou, M. Polycarpou, and M. Brady, "Digital Mammography: Towards Pectoral Muscle Removal via Independent Component Analysis," *Proc. 4th IET Int. Conf. Adv. Medical, Signal Inf. Process.*, pp. 1–4, 2008.
- [50] R. Adams and L. Bischof, "Seeded Region growing.pdf," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 641–647, 1994.

## **Abstract**

Breast cancer is the most common cancer and the second leading cause of cancer deaths among women. Currently mammography is the most popular method for detection of breast cancer. Masses and Microcalcifications are most abnormalities that may appear in breast cancer. Breast abnormalities are presented with a wide range of features. Studies show that 15-30% of breast cancer may be ignored by the radiologists. With the recent advances in the field of digital image processing has occurred, radiologists have the opportunity to reduce the error rate. To help radiologist computer aid detection (CAD) systems have been developed.

A recognition system generally consists of preprocessing, feature extraction and classification. In this thesis two methods have been proposed to implement this system and also they have been compared to existing methods.

In the both proposed method after preprocessing mammography images, with extract the texture characteristics of the co-occurrence matrix (GLCM), to classify individuals with support vector machine classifiers (SVM) and neural networks adapted. Finally, the second proposed method, the selection of optimal features using particle swarm optimization algorithm (PSO) and support vector machine classifier, the best result in the detection of MC obtain from mammography images.



**University of Shahrood**

**Faculty of Electrical Engineering**

**Detection of Microcalcifications in Mammograms using Particle  
Swarm Optimization**

**Mehdi Akbari**

**Supervisor:**

**Dr. Ali Solimani**

**Adviser:**

**Dr. Hosein Khosravi**

**February 2015**