مجله علوم پزشکی رازی دوره ۲۷، شماره ۴، تیر ۱۳۹۹ http://rjms.iums.ac.ir

مقاله پژوهشی

تشخیص توده مربوط به پستان از روی پردازش تصاویر ماموگرافی

نورالله علیزاده نوروزبولاغی: دانشکده مهندسی برق، دانشگاه صنعتی ارومیه، ارومیه ایران

یعقوب پوراسد: دانشکده مهندسی برق، دانشگاه صنعتی ارومیه، ارومیه ایران (* نویسنده مسئول) y.pourasad@uut.ac.ir

چکیده

زمینه و هدف: سرطان پستان مهم‌ترین و رایج‌ترین بیماری در بین زنان است که دومین میزان مرگ‌ومیر را بعد از سرطان ریه به خود اختصاص داده‌است. ماموگرافی دیجیتال تصویر گرفته شده با استفاده از اشعه x برای تجزیه و تحلیل، تفسیر و تشخیص می‌باشد. تشخیص خودکار سرطان پستان در تصاویر ماموگرافی یک وظیفه چالش برانگیز در بین سیستم‌های تشخیص به کمک کامپیوتر(CAD) می‌باشد.

روش کار: در این مقاله یک راهکار برای تشخیص اتوماتیک سرطان پستان ارائه شده است. راهکار ارائه شده شامل ۳ مرحله اصلی استخراج ناحیه پستان، حذف عضله پکتورال و طبقه‌بندی ویژگی‌های استخراج شده به دو دسته سرطانی و غیرسرطانی می‌باشد.

یافته‌ها: برای قطعه‌بندی از روش آستانه‌گذاری اتسو و سپس حذف عضله پکتورال با استفاده از انتخاب پیکسل دانه و الگوریتم رشد ناحیه میسر شده‌است. در مرحله بعدی ماتریس هم‌وقوعی خاکستری تصویر(GLCM)که توصیف‌کننده بافت تصویر است ایجاد شده و ۱۶ ویژگی از آن استخراج می‌شود. در نهایت طبقه‌بندی های مختلفی برای تفکیک ناحیه پستان به بافت‌های نرمال و سرطانی، آموزش داده می‌شوند. در نتایج به‌دست آمده نرخ تشخیص صحیح ۱۰۰ درصد برای شبکه عصبی و۹۶/۳ درصد برای طبقه‌بندهای درخت تصمیم گیری(C5.0,CHAID) بدست‌آمده است.

نتیجه‌گیری: اعتبار سنجی راهکار ارائه شده در این مقاله با استفاده از داده های پایگاه mini-MIAS انجام شده‌است و نتایج با کارهای قبلی انجام شده مقایسه شده‌است که نشان می‌دهد راهکار ارائه شده می‌تواند با اطمینان برای تشخیص سرطان پستان اعمال شود.

تعارض منافع: گزارش نشده است.

منبع حمایت کننده: حامی مالی نداشته است.

کلیدواژه‌ها

سیتم CAD،

سرطان پستان،

قطعه‌بندی،

استخراج ویژگی،

شبکه عصبی مصنوعی،

درخت تصمیم‌گیری

تاریخ دریافت: ۹۹/۰۱/۲۵

تاریخ چاپ: ۹۹/۰۴/۱۰

شیوه استناد به این مقاله:

Alizadeh Norouzbolaghi N, Pourasad Y. Detection and classification of breast masses using mammographically image processingd. Razi J Med Sci. 2020;27(4):60-73.

*انتشار این مقاله به‌صورت دسترسی آزاد مطابق با CC BY-NC-SA 3.0 صورت گرفته است.

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

Razi Journal of Medical Sciences. 2020;27(4):60-73. http://rjms.iums.ac.ir

Original Article

Detection and classification of breast masses using mammographically image processing

Noorollah Alizadeh Norouzbolaghi, Electrical engineering department, Urmia University of Technoology, Urmia, Iran
Yaghoub Pourasad, Electrical engineering department, Urmia University of Technoology, Urmia, Iran (* Corresponding author) y.pourasad@uut.ac.ir

Abstract

Background: Breast cancer is the most common type of cancer and the second leading cause of cancer death among American women. In Iran, the rate of breast cancer is lower than in industrialized and western countries, but with the growing trend, it is predicted that breast cancer will become one of the most common cancers in the country in the future. Mammography is currently one of the most effective and popular methods for screening and diagnosing breast cancer. Breast cancer is the primary and most common disease found in women which causes second highest rate of death after lung cancer. The digital mammogram is the X-ray of breast captured for the analysis, interpretation, and diagnosis. Automatic detection of breast cancer in mammograms is a challenging task in Computer Aided Diagnosis (CAD) techniques. This article aims to provide an automated computer diagnostic system to help diagnose early breast cancer. First, breast cancer and the survival statistics of patients with it, breast imaging techniques, and the presence of symptoms that are present in the images are signs of the disease. In the following, by introducing important and efficient methods in designing automatic diagnostic systems and its structure in order to distinguish cancerous images from non-cancerous breasts, the results obtained from this research and validations have been presented.

Methods: Breast cancer, one of the most common cancers in women, has a high mortality rate. Providing a medical assistance system for early detection of abnormalities associated with this cancer will greatly assist pathologists in identifying the causes of the disease and increase performance and accuracy in diagnosis. Studying the background of research in this field to better understand the problem and how to design this system in different ways gives us a more accurate view of this issue and also defines the design challenges of such a system. The results obtained by the methods presented in this paper are abbreviated as BMD_ML. A total of 64 mammograms, 23 cancer images containing benign and malignant masses, and 41 non-cancerous images were used to evaluate the methods used in this paper, and the results were obtained from the inputs of these images. A combination of digital image processing methods, random statistics and machine learning methods is used to perform the pre-processing, segmentation and extraction of ROI, feature extraction and classification at the lowest error rate. CAD is used in mammography screening. Mammography screening is used to detect early breast cancer. The CAD system helps diagnose lesions and classify benign and malignant tumors. This system is mostly used in the United States and the Netherlands. The first CAD system for mammography was developed during a research project at the University of Chicago. CAD systems, despite their high sensitivity, have very few features; This makes the benefits of using CAD unclear. In this report we present a methodology for breast cancer detection in digital mammograms. Proposed methodology consists of three major steps like segmentation of breast region, removal of pectoral muscle and classification of breast muscle into cancerous and normal tissues.

Keywords

CAD system,
Breast cancer,
Segmentation,
Feature extraction,
ANN,
Decision tree classifiers

Received: 13/04/2020
Published: 30/06/2020

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

Results: This article aims to provide an automated computer diagnostic system to help diagnose early breast cancer. First, breast cancer and the survival statistics of patients with it, breast imaging techniques, and the presence of symptoms that are present in the images are signs of the disease. Then, important and efficient methods in designing automatic diagnostic systems and its structure are introduced and the work done in the past is examined by researchers active in this field. Finally, the techniques used in this paper are presented in order to distinguish between cancerous and non-cancerous breast cancer images. Segmentation of breast muscle was performed by employing Otsus segmentation technique, afterwards removal of pectoral muscle is carried out by seed selection and region growing technique. In next step, Gray Level Co-occurrence Matrices (GLCM) was created form which several features were extracted. At the end, several classifiers were trained to classify breast region into normal and cancerous tissues. The proposed classifier reports classification accuracy of 100 % for ANN and 96.3 % for decision tree algorithms (C5.0 and CHAID). Proposed methodology was validated on Mini-MIAS database and results were compared with previously proposed techniques, which shows that proposed technique can be reliably apply for breast cancer detection. Classification includes the final stage of designing such a system. Machine learning techniques have good performance for classifying tissue features obtained from mammograms. Machine learning is generally divided into two categories, supervised and uncontrolled. Learning without supervision requires a large amount of data to train the network. The classification methods used in this dissertation are one of the supervised methods, so that when creating a feature vector matrix, a column is assigned to whether the data is cancerous or non-cancerous. This method both speeds up learning and compares classified data with the predetermined target value when testing them. In this paper, several machine learning methods will be used to classify methods for classifying and comparing diagnostic accuracy. In this paper, classifications are performed in SPSS modeler software.

Conclusion: Numerous methods for extracting features were provided in the Overview of Features Extraction section. The solution presented in this article is to use a GLCM matrix. The matrix of gray surfaces always gives rise to different combinations of the brightness of the pixels in the image. CAD systems, despite their high sensitivity, have very few features; This makes the benefits of using CAD unclear. It can concluded that CAD could not have a significant effect on cancer detection rates, but it would inadvertently increase the recall rate (ie, the FP rate); However, various studies have shown significant inconsistencies in recall effects. In the design of the CAD system, the partitioning and extraction of the feature are of special importance. It should be noted to what extent the extraction characteristics describe the segmented area. In the future, in order to increase the accuracy of classifications and commercialization of this system, flexible features can be extracted from the image, which in addition to the lack of overlap between the features can be very compatible with machine learning methods. It is also possible to classify between types of anomalies, and after diagnosing whether the image is cancerous or non-cancerous, the type of anomaly associated with it can be identified.

Conflicts of interest: None
Funding: None

Cite this article as:

Alizadeh Norouzbolaghi N, Pourasad Y. Detection and classification of breast masses using mammographically image processingd. Razi J Med Sci. 2020;27(4):60-73.

*This work is published under CC BY-NC-SA 3.0 licence.

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

مقدمه

سرطان پستان شایع‌ترین نوع سرطان و دومین عامل مرگ ناشی از سرطان در بین زنان امریکا است (۱). در ایران نرخ ابتلا به سرطان پستان نسبت به کشور های صنعتی و غربی پایین‌تر است اما با روند روبه رشد پیش‌بینی می‌شود در آینده سرطان پستان به یکی از شایع‌ترین سرطان‌های کشور تبدیل شود. در حال حاضر ماموگرافی یکی از موثرترین و محبوب‌ترین روش‌ها برای غربالگری وتشخیص سرطان پستان است. برای تشخیص این سرطان در بیماران با ریسک بالا از تصویربرداری، امواج فراصوت و تشدید مغناطیسی استفاده می‌شود. به‌تازگی از تصویر برداری امپدانس الکتریکی، تکنیک های پزشکی هسته ای، ترموگرافی (Thermography)، توموگرافی انتشار پوزیترون (Positron Emission Tomography) (PET)، تصویربرداری نوری (Optical Imaging) و توموگرافی کامپیوتری (Computer Tomography) (CT) برای غربالگری و تشخیص سرطان پستان به طور گسترده ای استفاده می‌شود (۲).

در مقاله ارائه شده (۳) به استفاده از فیلتر پایین گذر و غیر خطی میانه با پنجره ۳×۳ برای حذف خط وخش ها موجود در تصاویر ماموگرامی و نویز های روشنایی به‌وجود آمده در نزدیکی مرز پوست پستان، هنگام تصویر برداری که دارای محتوایی فرکانس بالا هستند استفاده شده است. استفاده از فیلتر میانه، موفق به حذف تمامی خط و خش‌های موجود در تصاویر ماموگرافی شده و مابقی مصنوعات و برچسب‌ها در قسمت قطعه بندی حذف شده‌اند. با اعمال فیلتر وینر بهینه شده محلی به تصویر و روش فشرده سازی هیستوگرام توانستند با استفاده از تخمین خطی، نویز سفید گوسی موجود در تصویر را حذف کرده و مقدار خطای میانگین مربعات را به حداقل برساند (۴). در برخی مطالعات به بررسی تغییر شدت روشنایی و ایجاد یک هیستوگرام جدید با توزیع یکنواخت به کمک روش متعادل‌سازی هیستوگرام پرداخته اند. از مزایای استفاده از متعادل‌سازی به کارآمدی این روش در ارتقا وضوح و کنتراست در تصاویر با کنتراست پایین اشاره شده است. در مقابل ضعف در ارتقا اطلاعات بافتی و عملکرد مطلوب فقط در تصاویر با یک سوژه تصویری را می‌توان از معایب این روش اشاره نمود (۵ و ۶).

این مقاله با هدف ارائه یک سیستم تشخیص خودکار کامپیوتری برای کمک به تشخیص زود به‌هنگام سرطان سینه ارائه شده است. ابتدا به معرفی سرطان پستان و آمار های مربوط به بقا بیماران مبتلا به آن، روش‌های تصویر برداری از پستان و معرفی علائم که وجود آنها در تصاویر نشانه‌های ابتلا به این بیماری می‌باشد، پرداخته می شود. در ادامه با معرفی روش‌های مهم و کارآمد در طراحی سیستم‌های تشخیص خودکار و ساختار آن به منظور تشخیص تصاویر سرطانی از غیر سرطانی پستان پرداخته و نتایج بدست آمده از این پژوهش و اعتبارسنجی های انجام شده ارائه شده است.

سامانه تشخیص به کمک کامپیوتر(CAD): شناسایی با کمک کامپیوتر (detection Computer-aided) (CADe)، که به آن تشخیص به کمک کامپیوتر (Computer-aided diagnosis) (CADx) نیز گفته می‌شود، سیستم‌هایی هستند که در تفسیر تصاویر پزشکی، به پزشکان یاری می‌رسانند. تکنیک‌های پرتو ایکس، MRI و سیستم‌های فراصوت، اطلاعات زیادی ایجاد می‌کنند که رادیولوژیست یا پزشک حرفه ای باید آن‌ها را در زمان کوتاهی تحلیل و ارزیابی کند. سیستم‌های CAD، تصاویر دیجیتال را به منظور تسهیل تصمیم‌گیری برای پزشک، پردازش کرده و بخش‌های مشکوک آن، مانند بیماری‌های احتمالی، را مشخص می‌کنند. به طور کلی، متخصص (معمولاً یک رادیولوژیست) مسئول تفسیر نهایی تصویر پزشکی است. هدف بعضی از سیستم‌هایCAD، شناسایی اولین نشانه‌های غیرطبیعی در انسان است که متخصص قادر به تشخیص آنها نمی‌باشد، درهم‌ریختگی ساختاری در ماموگرام می‌باشد (۷-۱۰).

روش شناسی و ساختار شناسی سیستم CAD: CAD اساساً از بازشناسی الگوی به شدت پیچیده استفاده می‌کند. تصاویر پرتو ایکس و سایر تصاویر برای یافتن ساختارهای مشکوک اسکن می‌شوند. به طور معمول برای بهینه کردن الگوریتم، چندین هزار تصویر مورد نیاز است. داده‌های تصاویر دیجیتال با فرمت دایکام (Dicom) در سرور CAD کپی شده و طی چندین مرحله، آماده‌سازی و تحلیل می‌شوند. عموما سیستم CAD از ساختار مشخصی برای تحلیل و آنالیز تصویر

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

استفاده میکند که میتوان آنها را در چند دسته کلی به همراه دسته بندی جزیی به فرم زیر ارائه کرد:

۱. پیش پردازش (Preprocessing):

- کاهش درست نماها (اشکالات در تصاویر)
- کاهش نویز تصویر
- نرمال‌سازی (هماهنگ سازی) کیفیت تصاویر (بهبود کنتراست) برای از بین بردن شرایط پایه ای مختلف تصویر مانند تفاوت پارامترهای پرتودهی
- فیلتر کردن

۲. قطعه‌بندی (segmentation):

- شناسایی ساختارهای مختلف در تصویر مانند برچسب ها ومصنوعات، رگ‌های خونی و ضایعات احتمالی و تطبیق با پایگاه داده آنوتیمیکی.
- نمونه برداری از سطوح خاکستری در حجم دلخواه

۳. تحلیل ساختار (Region of Interest) ROI/ ناحیه دلخواه:

ویژگی‌های خاص هر یک از نواحی شناخته شده به طور جداگانه تحلیل می‌شود که این ویژگی‌ها از قرار زیر است:

- میزان فشرده بودن
- شکل، اندازه و محل قرارگیری
- مراجعه به ساختارهاROI/های نزدیک
- بررسی میانگین سطوح خاکستری داخل یکROI
- متناسب کردن سطوح خاکستری با مرز ساختارهای داخلROI

۴. ارزیابی/ طبقه‌بندی:

پس از اینکه ساختار مورد تحلیل قرار گرفت، هر یک از ROI ها به صورت جداگانه برای بررسی احتمال (True positive) (TP) ارزیابی می‌شوند. در ادامه نمونه‌هایی از الگوریتم‌های طبقه‌بندی ذکر خواهد شد:

- قوانین نزدیکترین همسایه (به عنوان مثال، k امین همسایه نزدیک (KNN)
- طبقه‌بندی حداقل فاصله (Minimum distance classifier)
- طبقه‌بندی آبشاری (Cascading classifier)
- شبکه عصبی مصنوعی (ANN)
- شبکه تابع پایه شعاعی (RBF)
- ماشین بردار پشتیبانی (SVM)
- تجزیه و تحلیل اجزای اصلی (PCA)

در صورتی که ساختارهای شناسایی شده به حد آستانه مشخصی برسند، در تصویر علامت گذاری خواهند شد. این علامت گذاری‌ها، بسته به نوع سیستم CAD، به طور دائم یا موقت است. مزیت بعدی این است که تنها علامت گذاری‌های تاییدشده توسط رادیولوژیست ذخیره خواهد شد. علامت گذاری‌های غلط نباید ذخیره شود، زیرا معاینات بعدی را دشوار می‌کند.

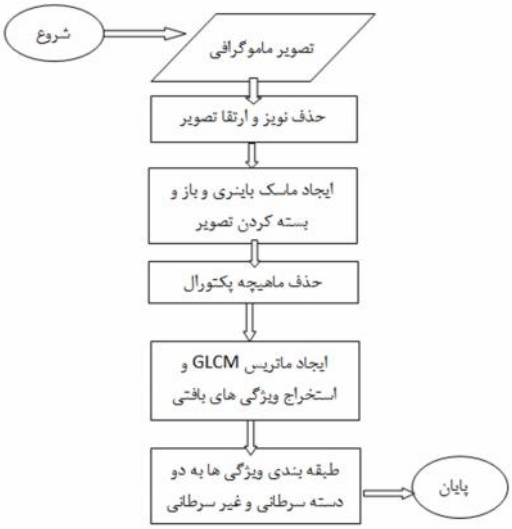
حساسیت، دقت و ویژگی سیستم CAD: یستم‌های CAD، به دنبال علامت گذاری ساختارهای مشکوک هستند. در حال حاضر این سیستم‌ها قادر به شناسایی ۱۰۰٪ تغییرات پاتولوژیک نیستند. نرخ ضربه (حساسیت) و دقت، بسته به نوع سیستم و کاربرد آن می‌تواند بیش از ۹۰٪ باشد (۱۱). به تشخیص‌های درست، مثبت واقعی (True Positive: TP) و به علامت گذاری نادرست بخش‌های سالم، مثبت کاذب (False Positive: FP)گفته می‌شود. هرچه میزانFP ها کمتر باشد، نشان دهنده بالاتر بودن ویژگی است. کم بودن ویژگی، مقبولیت سیستم CAD را کاهش می‌دهد، زیرا کاربر باید همه این تشخیص‌های نادرست را پیدا کند.

کاربرد CAD در تشخیص سرطان پستان: از CAD در غربالگری ماموگرافی استفاده می‌شود. غربالگری ماموگرافی برای تشخیص زودهنگام سرطان پستان مورد استفاده قرار می‌گیرد. سیستم CAD به تشخیص ضایعه‌ها و طبقه‌بندی تومورهای خوش‌خیم و بدخیم کمک می‌کند. این سیستم بیشتر در آمریکا و هلند به کار گرفته شده‌است. اولین سیستم CAD برای ماموگرافی، طی یک پروژه تحقیقاتی در دانشگاه شیکاگو ایجاد شد.. سیستم‌های CAD، علی‌رغم برخورداری از حساسیت بالا، ویژگی بسیار کمی دارند؛ این امر موجب نامشخص بودن مزایای استفاده از CAD می‌شود. در سال ۲۰۰۸، یک بررسی اصولی بر روی غربال گری ماموگرافی با استفاده از CAD به این نتیجه دست یافت که CAD نمی‌تواند اثر معنی داری بر نرخ تشخیص سرطان داشته باشد، اما نرخ recall (یعنی نرخ FP) را به صورت ناخواسته افزایش می‌دهد؛ با این حال، در مطالعات مختلف، ناهمگونی قابل توجهی را در تأثیر بر recall نشان داد(۱۲).

مراحل انجام CAD: یک سیستم CAD از ۲ قسمت

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

شکل ۱- مراحل کلی طی شده در یک سیستم CAD



شکل ۲- فلوچارت مراحل طراحی سیستم تشخیص سرطان پستان در روش پیشنهادی

سخت‌افزاری و نرم‌افزاری تشکیل یافته است. قسمت سخت افزار جهت دیجیتال سازی تصاویر دریافت شده می‌باشد و قسمت نرم‌افزاری با انجام الگوریتم‌های، پردازش تصویر و یادگیری ماشین به منظور آنالیز تصویر دیجیتال(ماموگرام) طراحی شده است. بیش ترین ناهنجاری‌های مربوط به پستان شامل توده ها و میکروکلسیفیکاسیون ها می‌باشند که در اکثر موارد وجود این ناهنجاری‌ها میتواند موجب سرطان سینه شود (۱۳-۱۵). همانطور که در شکل ۱ فلوچارت مربوط به ساختار و مراحل کلی طی شده در سیستم CAD مشاهده می‌شود، یک سیستم تمام خودکار تشخیص به کمک کامپیوتر برای آنالیز این ناهنجاری‌ها به سه مرحله کلی، پیش‌پردازش، تشخیص و طبقه‌بندی که هر کدام از این قسمت‌ها با استفاده از روش‌های مختلفی صورت گرفته‌اند.

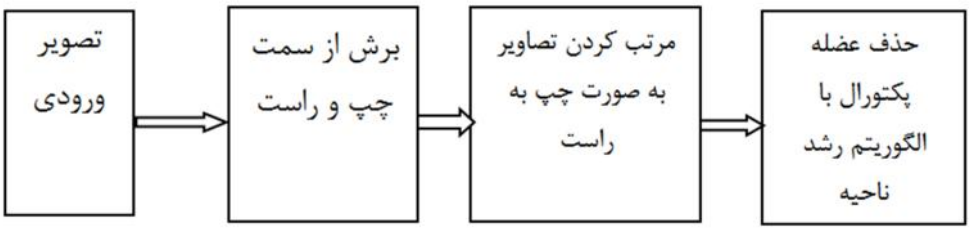
روش کار

سرطان پستان به عنوان یکی از رایج‌ترین سرطان‌ها در میان زنان میزان مرگ‌ومیر زیادی را به خود اختصاص می‌دهد. ارائه یک سیستم کمک پزشکی برای تشخیص زود به‌هنگام ناهنجاری های مربوط به این سرطان کمک شایانی به متخصصان آسیب‌شناسی در شناسایی عوامل این بیماری می‌کند و باعث بالا رفتن عملکرد و دقت در تشخیص می‌شود. مطالعه پیشینه تحقیق در این موضوع به فهم بهتر مساله و نحوه طراحی این سیستم به روش‌های مختلف دید دقیق تری از این مساله به ما می‌دهد و همچنین چالش‌های طراحی چنین سیستمی مشخص می‌شود. در این بخش از تلفیق روش‌های پردازش تصاویر دیجیتال، آماره‌های تصادفی و روش‌های یادگیری ماشین استفاده میشود تا قسمت‌های پیش‌پردازش، قطعه بندی و استخراج ROI، استخراج ویژگی و طبقه بندی را با کمترین نرخ خطا انجام شود. شکل ۲ فلوچارت طراحی این سیستم مشاهده می‌شود.

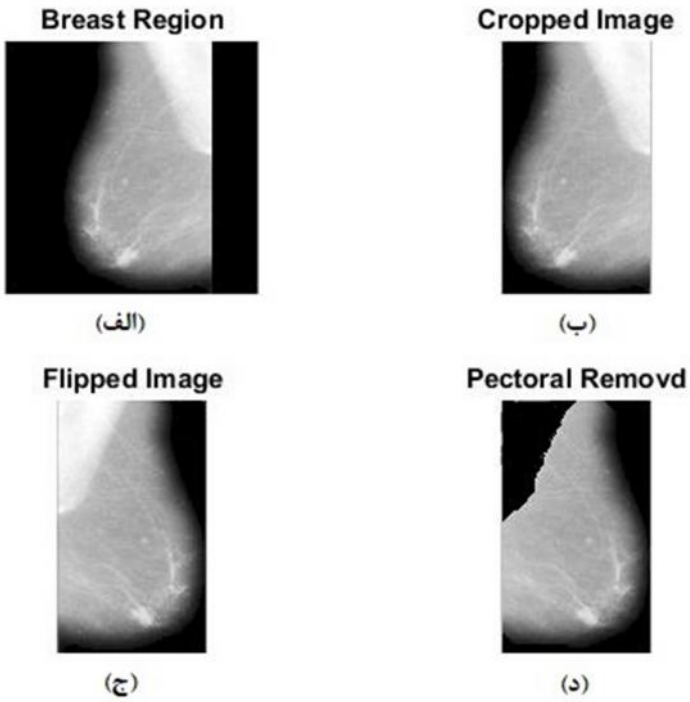
یافته‌ها

حذف ماهیچه پکتورال: ماهیچه پکتورال همانطور که در شکل ۳-الف مشاهده می‌شود به صورت یک ناحیه روشن در قسمت بالایی سمت راست یا چپ تصویر نمایان است. ماهیچه پکتورال دارای سطح روشنایی نزدیک به بافت سرطانی است. بنابراین حذف این عضله برای تشخیص صحیح سرطان سینه بسیار مهم است. درحالی که حذف این عضله کار ساده ای نیست چون در بعضی از ماموگرام‌ها به دلیل متراکم بودن بافت پستان عملا هیچ مرزی برای این ماهیچه نمی‌توان تعیین کرد. برای حذف این ماهیچه طبق فلوچارت

[Downloaded from rjms.iums.ac.ir on 2025-07-17]



شکل۳- فلوچارت مراحل حذف عضله پکتورال



شکل ٤- مراحل حذف ماهیچه پکتورال از ناحیه پستان الف-تصویر اصلی از راست به چپ، ب- برش اتوماتیک تصویر از راست و چپ، ج-تصویر دوران داده شده حول محور x، د-حذف عضله پکتورال

شکل ۴، از روش رشد ناحیه با استفاده از یک پیکسل دانه (یک حلقه‌ای) استفاده شده است. روش کار به این صورت است که ابتدا تصویر را به صورت خودکار از چپ و راست با استفاده از حلقه ها تکرار همانند شکل‌های ۴-ب و ج برش می‌دهیم و اگر تصویر از راست به چپ بود آن‌را حول محور x ۱۸۰ درجه دوران می‌دهیم تا عضله پکتورال بالا سمت چپ باشد. حال مختصات یک پیکسل دانه را از این ماهیچه انتخاب میکنیم تا با استفاده از اپراتورهای مورفولوژی ریاضی شروع به مقایسه پیکسل‌های همسایه غیرمستقیم می‌کند تا به طور تکراری رشدکند .تفاوت بین مقدار شدت پیکسل و میانگین منطقه به عنوان یک اندازه از شباهت استفاده می‌شود.پیکسل‌ها در محدوده پارامتر تحمل در اطراف شدت روشنایی متوسط منطقه به این ناحیه در حال رشد اختصاص داده می‌شوند. این فرآیند زمانی متوقف می شود که مناطق تکامل خود را از تکرار به پیکسل بعدی متوقف کند.

استخراج ویژگی: پس از حذف ماهیچه پکتورال ویژگی‌هایی برای مشخصه بندی و توصیف بافت سرطانی از بافت های غیر سرطانی استخراج می‌شود. روش های متعددی برای استخراج ویژگی در قسمت مروری بر استخراج ویژگی ارائه شد. راه کار ارائه شده در این مقاله استفاده از یک ماتریس GLCM است. ماتریس هم‌رخداد سطوح خاکستری، رخداد ترکیب‌های

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

متفاوتی از شدت روشنایی پیکسل‌ها را در تصویر می‌دهد. همچنین توابع GLCM با استفاده از اینکه چگونه جفت پیکسل‌هایی با مقادیر خاص و در یک رابطه فضایی مشخص در یک تصویر اتفاق می‌افتند بافت یک تصویر را مشخص می‌کند. ماتریس GLCM با نگاشت سطوح روشنایی تصویر به ۸ سطح یک ماتریس ۸×۸ می‌باشد. در این پایان‌نامه ماتریس GLCM تصاویر سرطانی و غیر سرطانی در ۴ زاویه: ۰، ۴۵، ۹۰و ۱۳۵ درجه با فاصله بین پیکسل‌ها d=1 تشکیل شده و سپس ۴ ویژگی آماری: انرژی، کنتراست، همبستگی و همگن‌بودن از هر GLCM استخراج می‌شود در نتیجه ۱۶ ویژگی از هر تصویر استخراج می‌شود. ویژگی های استخراجی بصورت اعدادی بین صفر و یک می‌باشند. به‌عنوان مثال، ماتریس‌های GLCM و بردار ویژگی‌های استخراج شده از تصویر ۴ به شرح زیر می‌باشد.

$$GLCM\ 0^\circ = \begin{bmatrix} 331268 & 1275 & 36 & 853 & 62 & 2 & 0 & 0 \\ 1275 & 75646 & 1455 & 0 & 0 & 0 & 0 & 0 \\ 36 & 1455 & 62198 & 1545 & 8 & 0 & 0 & 0 \\ 853 & 0 & 1545 & 101770 & 4613 & 59 & 0 & 0 \\ 62 & 0 & 8 & 4613 & 585728 & 6462 & 1 & 0 \\ 2 & 0 & 0 & 59 & 6462 & 51710 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 & 5 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$GLCM\ 45^\circ = \begin{bmatrix} 329552 & 1851 & 57 & 920 & 127 & 4 & 0 & 0 \\ 1851 & 74018 & 2280 & 0 & 0 & 0 & 0 & 0 \\ 57 & 2280 & 60370 & 2526 & 9 & 0 & 0 & 0 \\ 920 & 0 & 2526 & 98132 & 7203 & 59 & 0 & 0 \\ 127 & 0 & 9 & 7203 & 577814 & 11721 & 0 & 0 \\ 4 & 0 & 0 & 59 & 11721 & 46440 & 14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 14 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$GLCM\ 90^\circ = \begin{bmatrix} 333312 & 1086 & 21 & 67 & 66 & 5 & 0 & 0 \\ 1086 & 75580 & 1483 & 0 & 0 & 0 & 0 & 0 \\ 21 & 1483 & 62056 & 1682 & 0 & 0 & 0 & 0 \\ 67 & 0 & 1682 & 102222 & 4869 & 0 & 0 & 0 \\ 66 & 0 & 0 & 4869 & 583482 & 8456 & 1 & 0 \\ 5 & 0 & 0 & 0 & 8456 & 49764 & 13 & 0 \\ 0 & 0 & 0 & 0 & 1 & 13 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$GLCM\ 135^\circ = \begin{bmatrix} 329836 & 1564 & 57 & 920 & 127 & 7 & 0 & 0 \\ 1564 & 74620 & 1965 & 0 & 0 & 0 & 0 & 0 \\ 57 & 1965 & 61150 & 2063 & 7 & 0 & 0 & 0 \\ 920 & 0 & 2063 & 99096 & 6703 & 58 & 0 & 0 \\ 127 & 0 & 7 & 6703 & 580724 & 9312 & 1 & 0 \\ 7 & 0 & 0 & 58 & 9312 & 48846 & 15 & 0 \\ 0 & 0 & 0 & 0 & 1 & 15 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$Feature\ Vector = \begin{bmatrix} Feature\ Type \\ GLCM\ 0 \\ GLCM\ 45 \\ GLCM\ 90 \\ GLCM\ 135 \end{bmatrix} = \begin{bmatrix} Contrast & Correlation & Energy & Homogenity \\ 0.0394 & 0.9938 & 0.3087 & 0.9864 \\ 0.0588 & 0.9908 & 0.3016 & 0.9779 \\ 0.0313 & 0.9951 & 0.3074 & 0.9856 \\ 0.0525 & 0.9918 & 0.3043 & 0.9811 \end{bmatrix}$$

طبقه‌بندی: طبقه‌بندی آخرین مرحله از طراحی چنین سیستمی را شامل می‌شود. تکنیک‌های یادگیری ماشین برای طبقه‌بندی ویژگی‌های بافتی بدست آمده از ماموگرام‌ها عملکرد مطلوبی دارند. یادگیری ماشین به طور کلی به دو دسته با نظارت و بدون نظارت تقسیم می‌شوند. در یادگیری بدون نظارت نیاز به حجم داده فراوان جهت آموزش شبکه می‌باشد. روش های طبقه‌بندی مورد استفاده در این پایان‌نامه جزو روش‌های با نظارت می‌باشند، به طوری که به هنگام درست کردن ماتریس بردار ویژگی، یک ستون به سرطانی یا غیر سرطانی بودن داده‌ها اختصاص می‌یابد. این روش هم سرعت یادگیری را بالا می‌برد و هم به هنگام آزمودن آنها داده‌های طبقه‌بندی شده را با مقدار هدف از پیش تعیین شده مقایسه می‌کند. در این مقاله به منظور بررسی روش‌های طبقه بندی و مقایسه صحت تشخیص از چندین روش یادگیری ماشین برای طبقه‌بندی استفاده خواهد شد. در این مقاله طبقه‌بندی‌ها در نرم افزار SPSS modeler انجام گرفته است.

در این بخش به بررسی نتایج بدست آمده از مراحل مختلف کار پرداخته می‌شود و نتایج کارهای انجام شده توسط محققان دیگر نیز برای مقایسه با کار انجام شده در این پژوهش ارائه شده است (۳، ۵ و ۱۵). نتایج بدست آمده توسط روش‌های ارائه شده در این مقاله با نام اختصاری BMD_ML (Breast Mass Detection by Machine Learning) به تحریر آمده است. برای ارزیابی روش‌های انجام شده در این مقاله از مجموع ۶۴ تصویر ماموگرافی، ۲۳ تصویر سرطانی و حاوی توده‌های خوش‌خیم و بدخیم و ۴۱ مورد تصاویر غیر سرطانی استفاده شده است و نتایج بدست آمده حاصل از ورودی این تصاویر می‌باشد.

نتایج قطعه‌بندی ناحیه پستان: آستانه‌گذاری انجام شده در روش BMD_ML با آستانه بدست آمده از روش اتسو انجام گرفته در این پایان‌نامه از آستانه گذاری اتسو ۱۶ کلاسه استفاده شده که کوچکترین

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

جدول ۱– نتایج بدست آمده از قطعه بندی به روش آستانه گذاری اتسو

نتایج به درصد	تعداد تصاویر	دسته بندی
۹۸/۴۳	۶۳	خوب
۱/۶۷	۱	قابل قبول

جدول ۲– نتایج قطعه بندی با استفاده از آستانه گذاری اتسو در (۴)

نتایج به درصد	تعداد تصاویر	دسته بندی
۹۶/۸۹	۳۱۲	خوب
۲/۴۸	۸	قابل قبول
۰/۶۲	۲	غیر قابل قبول

جدول ۳– نتایج قطعه بندی با استفاده از روش آستانه‌گذاری اتسو در (۵)

کلاس بندی تشخیص داده شده		کلاس بندی اصلی
تصویر سالم	تصویر سرطانی	
۰۶	۲۹	تصویر سرطانی
۱۲	۰۳	تصویر سالم

سطح بدست آمده از این ۱۶ کلاس برای آستانه گذاری استفاده شده است. هدف از انجام این آستانه گذاری حذف نویز، علامت‌ها و برچسب‌ها می‌باشد که نتایج بدست آمده از قطعه‌بندی را می‌توان در ۲ گروه:

خوب: این دسته شامل تصاویر ماموگرافی است که برچسب‌ها و مصنوعات کاملا حذف شده و مرز پستان بطور کامل استخراج شده است.

قابل قبول: این دسته شامل تصاویر ماموگرافی است که برچسب ها و مصنوعات کاملا حذف شده اما خطای قطعه بندی هنوز بزرگ است.

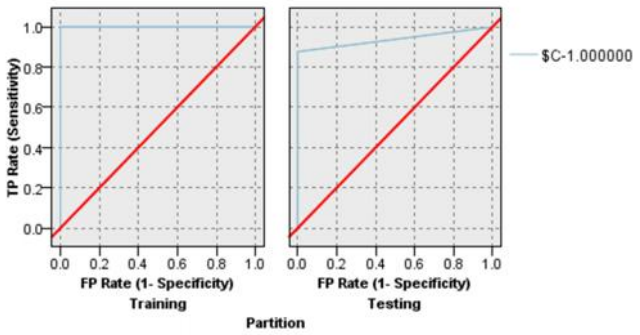
دسته‌بندی کرد و نتایج آن را می‌توان در جدول ۱ مشاهده کرد. جدول ۲ نیز نتایج بدست آمده از روش اتسو می‌باشد (۳). در جدول ۳ نیز می‌توان نتایج قطعه بندی بدست آمده از (۵) وجود دارد. مشاهده می‌شود الگوریتم استفاده شده توانایی بالای در تعیین مقدار آستانه و استخراج ناحیه پستان دارد.

بحث و نتیجه‌گیری

نتایج قطعه بندی با استفاده از پارامترهای Accuracy، Sensitivity و Specificity معین می‌شود

جدول ٤– نتایج طبقه بندی C5.0

٪	تست	٪	یادگیری	بخش
۹۶/۳	۲۶	۱۰۰	۳۶	درست
۳/۷	۱	۰	۰	غلط
	۹۶,۲۹۶		Accuracy	



شکل ۵– نمودارROC مربوط به طبقه بندی C5.0

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

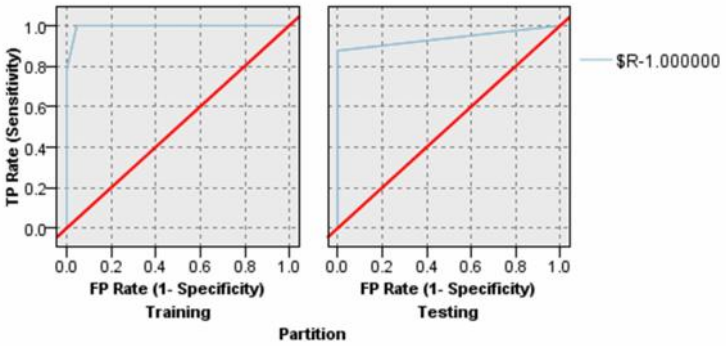
که نتایج Accuracy را به صورت می‌توان بیان کرد و نتایج Sensitivity و Specificity را با نمودار ROC بیان کرد. جدول۴ نتایج بدست آمده از این طبقه بندی را نشان می‌دهد و شکل ۵ نمودار ROC یادگیری و آزمایش مربوط به این طبقه بندی را نشان می‌دهد که به نرخ تشخیص صحیح ۹۶/۳ دست یافته است.

سورندیران و همکاران (۹) طبقه بندی ماموگرام‌های بدست آمده از پایگاه DDSM را با یک روش هیبریدی مورد استفاده از C5.0 و شبکه عصبی استفاده کردند و به نرخ طبقه بندی ۹۳/۳۳٪ رسیدند؛ جدول ۵ نتایج بدست آمده از این طبقه بندی را نشان میدهد و شکل ۶ نمودار ROC یادگیری و آزمایش مربوط به این طبقه بندی را نشان می‌دهد که به نرخ تشخیص صحیح ۹۶/۳ دست یافته است.

جدول ۷ نتایج بدست آمده از این طبقه بندی را نشان می‌دهد و شکل ۷ نمودار ROC یادگیری و آزمایش مربوط به این طبقه بندی را نشان می‌دهد که به نرخ تشخیص صحیح ۸۴/۴۸ دست یافته است.

جدول ٦– نتایج طبقه بندی CHAID

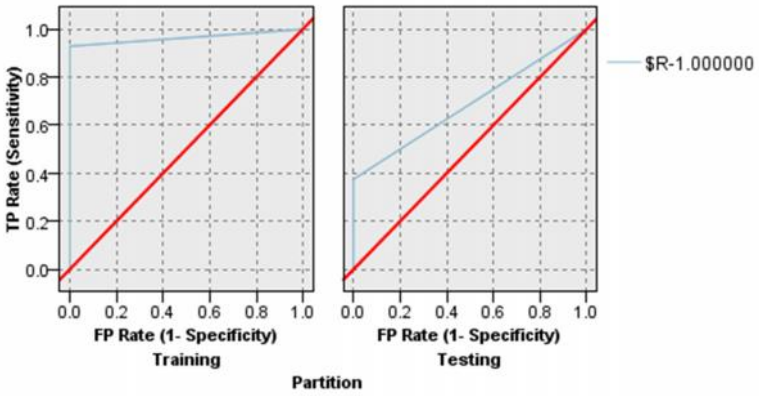
بخش	یادگیری	٪	تست	٪
درست	۳۵	۹۷/۲۲	۲۶	۹۶/۳
غلط	۱	۲/۷۸	۱	۳/۷
Accuracy		۹۶,۲۹۶		



شکل٦– نمودارROC مربوط به طبقه بندی CHAID

جدول ۷– نتایج طبقه بندی C&R

بخش	یادگیری	٪	تست	٪
درست	۳۵	۹۷/۲۲	۲۲	۸۴/۴۸
غلط	۱	۲/۷۸	۵	۱۵/۵۲
Accuracy		۸۴/۴۸		



شکل ۷– نمودارROC مربوط به طبقه بندی C&R

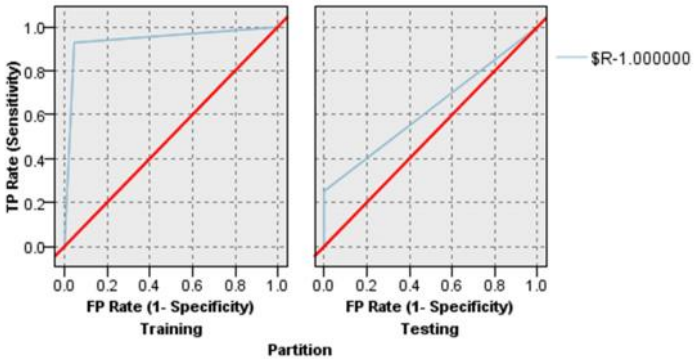
[Downloaded from rjms.iums.ac.ir on 2025-07-17]

جدول ۸ نتایج بدست آمده از این طبقه بندی را نشان می‌دهد و شکل ۸ نمودار ROC یادگیری و آزمایش مربوط به این طبقه بندی را نشان می‌دهد که یه نرخ تشخیص صحیح ۷۷/۷۸ دست یافته است.

جدول ۹ نتایج بدست آمده از این طبقه بندی را نشان می‌دهد و شکل ۹ نمودار ROC یادگیری و

جدول ۸– نتایج طبقه بندی Quest

بخش	یادگیری	٪	تست	٪
درست	۳۴	۹۴/۴۴	۲۱	۷۷/۷۸
غلط	۲	۵/۵۶	۶	۲۲/۲۲
Accuracy		۷۷,۷۸		



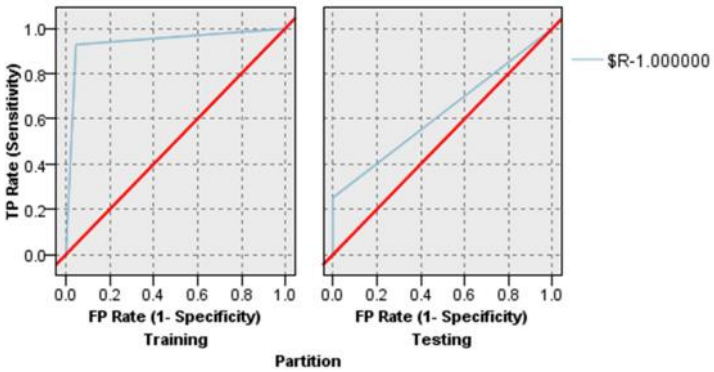
شکل ۸– نمودارROC مربوط به طبقه بندی Quest

جدول ۹– نتایج طبقه بندی SVM

بخش	یادگیری	٪	تست	٪
درست	۴۰	۹۰/۹۱	۱۷	۸۹/۴۷
غلط	۴	۹/۰۹	۲	۱۰/۵۳
Accuracy		۸۹/۴۷		

جدول ۱۰– نتایج طبقه بندی SVM (۳)

بخش	یادگیری	٪	تست	٪
درست	۸۲		۳۳	
غلط	۹۲		۵۴	
Accuracy		۹۶		



شکل ۹– نمودارROC مربوط به طبقه بندی SVM

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

آزمایش مربوط به این طبقه بندی را نشان می‌دهد که به نرخ تشخیص صحیح ۸۹/۴۷ دست یافته است.

جدول ۱۰ نتایج طبقه بندی ویژگی‌های آماری (۳) را نشان می‌دهد.

جدول ۱۱ نتایج بدست آمده از این طبقه بندی را نشان می‌دهد و شکل ۱۰ نمودار ROC یادگیری و آزمایش مربوط به این طبقه بندی را نشان می‌دهد که به نرخ تشخیص صحیح ۱۰۰درصد دست یافته است.

مطالعه‌ای دیگر از یک شبکه عصبی مصنوعی سه لایه (ANN) با هفت ویژگی برای طبقه‌بندی مناطق مشخص شده استفاده کردند و نتایجی مطابق با جدول ۱۲ بدست آوردند (۱۴).

جدول ۱۳ نتایج بدست آمده از این طبقه بندی را نشان می‌دهد و شکل ۱۱ نمودار ROC یادگیری و آزمایش مربوط به این طبقه بندی را نشان می‌دهد که به نرخ تشخیص صحیح ۸۹/۴۷درصد دست یافته است.

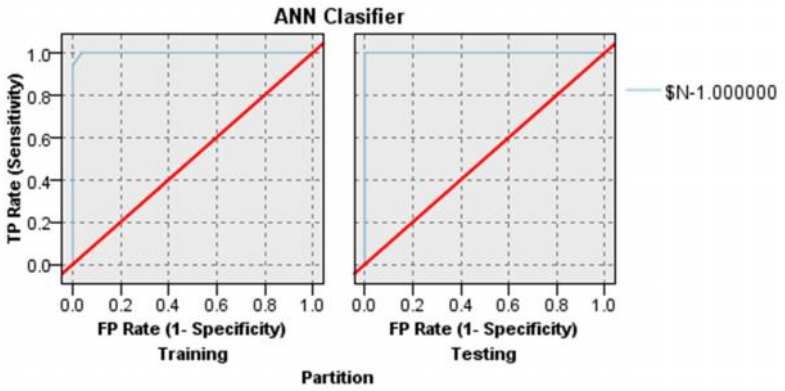
سونار و همکاران (۱۲) برای تصاویر ماموگرافی بدست آمده از پایگاه MIASبا استفاده از KNN طبقه بندی انجام دادند و نتایج آن به شرح جدول ۱۴ می‌باشد.

جدول ۱۱– نتایج طبقه بندی ANN

بخش	یادگیری	٪	تست	٪
درست	۴۲	۹۵/۴۵	۱۹	۱۰۰
غلط	۲	۴/۵۵	۰	۰
Accuracy		۱۰۰		

جدول ۱۲– نتایج طبقه بندی ANN اسلام و همکاران

طبقه بندی	حساست	خصوصیت
درصد	۸۳/۸۷	۹۰/۹۱



شکل ۱۰– نمودار ROC مربوط به طبقه بندی ANN

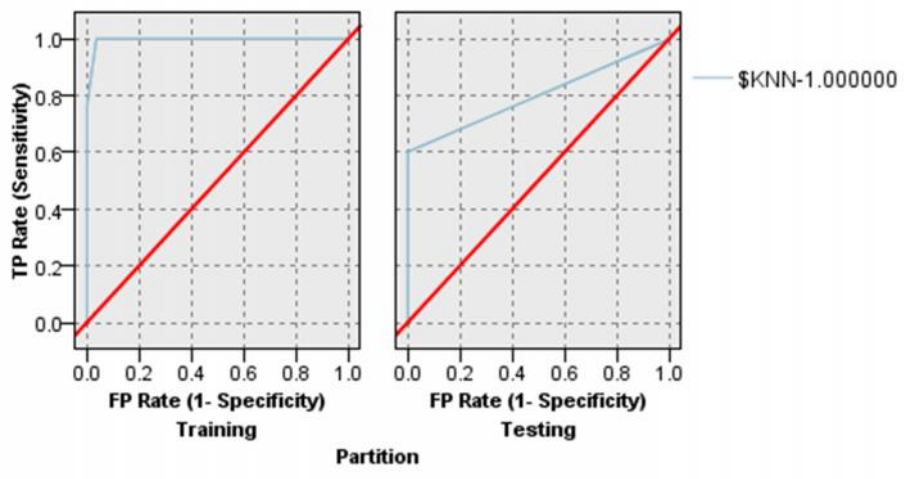
جدول ۱۳– نتایج طبقه بندی KNN برای K=3

بخش	یادگیری	٪	تست	٪
درست	۴۰	۹۰/۹۱	۱۷	۸۹/۴۷
غلط	۴	۹/۰۹	۲	۱۰/۵۳
Accuracy		۸۹/۴۷		

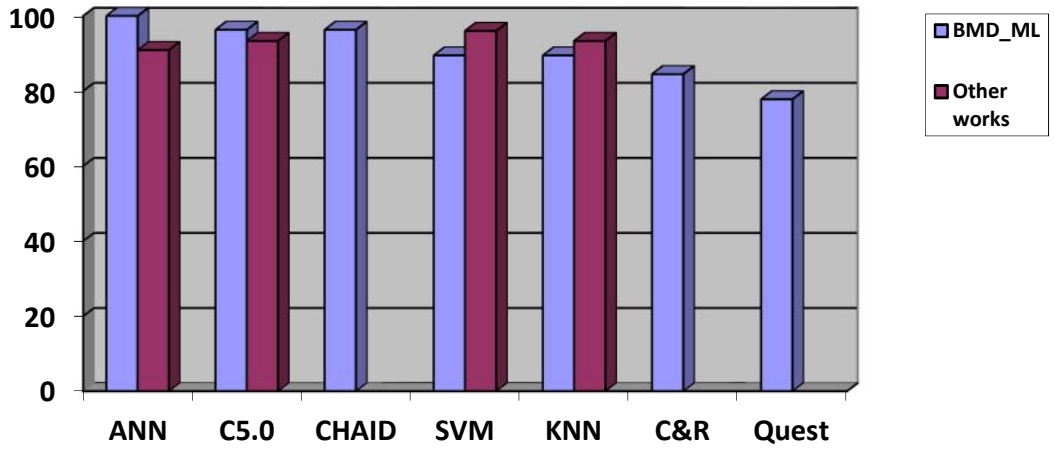
جدول ۱۴– نتایج طبقه بندی KNN (۱۴)

صحت تشخیص	۹۳/۳۳
حساسیت	۹۳/۳۳
ویژگی	۹۳/۳۳

[Downloaded from rjms.iums.ac.ir on 2025-07-17]



شکل ۱۱– نمودار ROC مربوط به طبقه بندی KNN



شکل ۱۲– مقایسه طبقه‌بندی‌های BMD_ML با سایر روش‌های مطالعه شده

نتایج بدست آمده از سیستم BMD_ML نشان می‌دهند که مطابق نمودار شکل ۱۲ رابطه فضایی بین ویژگی‌های بافتی استخراج شده به راحتی می‌تواند با طبقه بندی شبکه عصبی و درخت‌های تصمیم گیری C5.0 و CHAID از هم تفکیک شوند و در نتیجه تصاویر سرطانی از غیر سرطانی جدا شوند. نتایج ضعیف طبقه‌بندی‌های C&R وQuest مناسب طبقه بندی با ویژگی‌های استخراجی نبوده و به همین خاطر عملکرد نامطلوبی دارند.

این مقاله با هدف ارائه یک سیستم تشخیص خودکار کامپیوتری برای کمک به تشخیص زود به‌هنگام سرطان سینه ارائه شده است. ابتدا به معرفی سرطان پستان و آمارهای مربوط به بقا بیماران مبتلا به آن، روش‌های تصویر برداری از پستان و معرفی علائم که وجود آنها در تصاویر نشانه‌های ابتلا به این بیماری می‌باشد، پرداخته می شود. سپس روش‌های مهم و کارآمد در طراحی سیستم‌های تشخیص خودکار و ساختار آن معرفی شده و کارهای انجام شده در گذشته توسط محققین فعال در این حوزه کاری بررسی می‌گردد. نهایتا تکنیک های استفاده شده در این مقاله به منظور تشخیص تصاویر سرطانی از غیر سرطانی پستان ارائه شده و نتایج بدست آمده از این پژوهش و اعتبارسنجی های انجام شده ارائه شده‌است. از CAD در غربالگری ماموگرافی استفاده می‌شود. غربالگری ماموگرافی برای تشخیص زودهنگام سرطان پستان مورد استفاده قرار می‌گیرد. سیستم CAD به تشخیص ضایعه‌ها و طبقه‌بندی تومورهای

[Downloaded from rjms.iums.ac.ir on 2025-07-17]

خوش‌خیم و بدخیم کمک می‌کند. این سیستم بیشتر در آمریکا و هلند به کار گرفته شده‌است. اولین سیستم CAD برای ماموگرافی، طی یک پروژه تحقیقاتی در دانشگاه شیکاگو ایجاد شد. سیستم‌های CAD، علی‌رغم برخورداری از حساسیت بالا، ویژگی بسیار کمی دارند؛ این امر موجب نامشخص بودن مزایای استفاده از CAD می‌شود. در سال ۲۰۰۸، یک بررسی اصولی بر روی غربال گری ماموگرافی با استفاده از CAD به این نتیجه دست یافت که CAD نمی‌تواند اثر معنی داری بر نرخ تشخیص سرطان داشته باشد، اما نرخ recall (یعنی نرخ FP) را به صورت ناخواسته افزایش می‌دهد؛ با این حال، در مطالعات مختلف، ناهمگونی قابل توجهی را در تأثیر بر recall نشان داد. در طراحی سیستم CAD قسمت قطعه بندی و استخراج ویژگی از اهمیت ویژه ای برخور دارند بایستی توجه شود که ویژگی های استخراجی تا چه میزان توصیف کننده ناحیه قطعه بندی شده می‌باشد. در آینده می‌توان به منظور بالا رفتن دقت طبقه بندی‌ها و تجاری سازی این سیستم ویژگی‌هایی منعطف از تصویر استخراج کرد که علاوه بر اینکه هم‌پوشانی بین ویژگی‌ها وجود نداشته باشد سازگاری زیادی با روش‌های یادگیری ماشین داشته باشد. همچنین میتوان طبقه بندی بین انواع ناهنجاری‌ها انجام داد و بعد از تشخیص سرطانی یا غیر سرطانی بودن تصویر بتوان نوع ناهنجاری مربوط به آن را هم تشخیص داد.

References

1. Sree SV, Acharya EY, Faust O. Breast imaging: A survey. World J Clin Oncol. 2011;2:171-179.
2. Qayyum A, Basit A. Automatic breast segmentation and cancer detection via SVM in mammograms. International conference on emerging technologies (ICET). 2016:1-6.
3. Cheng HD, Shi XJ. A simple and effective histogram equalization approach to image enhancement. Digit Signal Process. 2014;14:158-170.
4. Singh N, Veenadhari S. Breast Cancer Segmentation Using Global Thresholding and Region Merging. J Clin Oncol. 2018;12:297-292.
5. Baker JA, Rosen EL, Lo JY, Gimenez EI, Walsh R, Soo MS. Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. Am J Roentgenol. 2003;4:1083-1088.
6. Taylor PH, Potts WW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. Eur J Cancer. 2008;6:798-807.
7. Jamshidinaeini Y, Akbari ME, Abdollahi M, Ajami M, Davoodi SH. Vitamin D status and risk of breast cancer in Iranian women: a case–control study. J Amer Coll Nut. 2016 2;35(7):639-46.
8. Kazemian E, Amouzegar A, Akbari ME, Moradi N, Gharibzadeh S, Jamshidi-Naeini Y, et al. Vitamin D receptor gene polymorphisms affecting changes in visceral fat, waist circumference and lipid profile in breast cancer survivors supplemented with vitamin D3. Lip Heal Dis. 2019 1;18(1):161.
9. Sundaram KM, Sasikala D, Rani PA. A study on preprocessing a mammogram image using Adaptive Median Filter. Int J Innov Res Sci Eng Technol. 2014;3:10333-10337.
10. Ojo J. Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram. IOSR J Comput Eng. 2014;16:6-9.
11. Qiam SB, Sadagian VA. Classifying Benign and Malignant Masses Using Statistical Measures. ICTACT J Image Video Process. 2016;2:319-326.
12. Sonar P, Bhosle U, Choudhury C. Mammography classification using modified hybrid SVM-KNN. Proceedings of IEEE International Conference on Signal Processing and Communication, ICSPC 2017. 2018;3:305-311.
13. Safdar Gardezi S, Elazab A, Lei B, Wang T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. J Med Internet Res. 2019;21(7):e14464.
14. Hongyu W, Jun F, Qirong B, Feihong L, Min Z, Yu R, et al. Breast Mass Detection in Digital Mammogram Based on Gestalt Psychology. Comp Med Biomed Inf Anal. 2018;4015613:1-13.
15. AlFayez F, El-Soud M, Gaber T. Thermogram Breast Cancer Detection: A Comparative Study of Two Machine Learning Techniques. Appl Sci. 2020;10:551.

[Downloaded from rjms.iums.ac.ir on 2025-07-17]