



ارائه یک سیستم توصیه کننده مطلوب به منظور تشخیص سرطان سینه
با استفاده از تحلیل و طبقه‌بندی تصاویر ماموگرافی پزشکی
(Providing a Descriptive System for Diagnosis of Breast Cancer
Using the Analysis and Classification of Mammograms)

۱- گلایل قرائتی ۲- آزاده منتظری

۱- گلایل قرائتی gharaaty@ymail.com
۲- آزاده منتظری montazeri.a.85@gmail.com

چکیده

در این مقاله به منظور شناسایی نواحی توموری در تصاویر ماموگرافی پستان و همچنین طبقه‌بندی تومورهای شناسایی شده به دو دسته بدخیم و خوش خیم سیستمی، ارائه شده است. روش پیشنهادی از بلوک‌های پردازش تصویر شامل پیش‌پردازش، ناحیه‌بندی و تشخیص توده با استفاده از آستانه‌گذاری چندسطحی، حذف نواحی کاذب، استخراج ویژگی‌های مفید نظیر ماتریس هم‌رخدادی، کاهش فضای ویژگی با استفاده از الگوریتم Relief و در نهایت طبقه‌بندی نوع توده با استفاده از طبقه‌بند ماشین بردار پشتیبان تشکیل شده است. نتیجه عددی تشخیص توده توسط روش پیشنهادی براساس معیارهای ارزیابی Sensitivity و FP-rate به ترتیب به مقدار ۹۴.۶۲٪ و ۴.۰۲٪ رسیده است.

واژه‌های کلیدی: پردازش تصویر، توده در ماموگرافی، دیتاست SVM, MIAS, تومورها (خوش خیم و بدخیم)، نمودار ROC

۱- مقدمه

سرطان پستان به‌عنوان یکی از شایع‌ترین سرطان‌ها بیان شده است، بنابراین تحقیق در حوزه تشخیص تومور در تصاویر ماموگرافی سرطان پستان از اهمیت ویژه‌ای برخوردار است. این نوع از سرطان به دلیل رشد خارج از قاعده سلول‌های غیرطبیعی در بافت پستان است. در دو گروه تومورها (خوش خیم و بدخیم)، رشد سریع و زیاد سلول‌ها وجود دارد. روند ازدیاد سلول‌ها در تومورهای خوش خیم، در مرحله معینی متوقف شده ولی در تومورهای بدخیم به صورت غیرقابل مهار تا جایی ادامه می‌یابد که اگر درمان نشود، همه اندام‌های بدن را تحت تأثیر قرار می‌دهد و باعث از کار افتاد آنها می‌شود. هدف این تحقیق ارائه روشی برای کمک به تشخیص اتوماتیک، جهت تشخیص تومور و نوع بدخیم یا خوش خیم بودن آن در تصاویر ماموگرافی مربوط به بیماران است.

مهمترین نکته برای بهترین پیش‌بینی، شناسایی و تشخیص سلول‌های سرطانی است. محققان بسیاری از روش‌های تشخیصی، از جمله ماموگرافی، تصویربرداری تشدید مغناطیسی (MRI)، سونوگرافی، توموگرافی کامپیوتری، توموگرافی انتشار

پوزیترون و بیوپسی را مورد بررسی قرار داده‌اند. در سال‌های اخیر، محققان علاقه‌مندی خود را در توسعه سنسورهای بیولوژیکی برای تشخیص سرطان پستان با استفاده از نشانگرهای مختلف نشان داده‌اند. به غیر از سنسورهای بیولوژیکی و بیومارکرها، تکنیک‌های تصویربرداری میکروویو به‌عنوان ابزار تشخیصی امیدوارکننده برای تشخیص سرطان پستان مورد مطالعه قرار گرفته است. از سوی دیگر محققین با استفاده از الگوریتم‌های پیش‌پردازشی و پردازشی و با تحلیل و آنالیز تصاویر پزشکی سعی در کشف و شناسایی ویژگی‌های مطلوب نموده و به کمک روش‌های طبقه‌بندی و تشخیص الگو کمک زیادی به-تشخیص زودهنگام سرطان و اقدامات درمانی لازم نموده‌اند.

طبق آمار انستیتو ملی سرطان آمریکا در سال ۲۰۱۴، حدود ۶۷۰,۲۳۲ زن و ۳۶۰,۲ مرد مبتلا به این نوع سرطان شده، که حدود ۴۰,۰۰۰ نفر از زنان و ۴۳۰ مرد فوت کردند [۱]. در سال‌های اخیر، تحقیقات گسترده‌ای به منظور کاهش خطای تشخیص سرطان پستان و افزایش سرعت تشخیص این سرطان انجام گرفته است و مقالات مختلفی به چاپ رسیده است. نتایج این تحقیقات می‌تواند به رادیولوژیست‌ها و متخصصان در تشخیص سریع و همچنین داشتن ضریب خطای کم، کمک کنند. در سال‌های اخیر تلاش‌های بسیاری برای توسعه سیستم‌های CAD صورت گرفته است که براساس پیشرفت در زمینه‌های پردازش تصویر دیجیتال، شناسایی الگو و هوش مصنوعی است. انتظار می‌رود سیستم‌های CAD بر وابستگی به اپراتور غلبه کنند، همچنین باعث افزایش میزان تشخیص و کاهش هزینه مکمل‌های پزشکی گردند [۲,۳].

در مقاله [۴] گزارش شده است که حساسیت تشخیص بدون سیستم CAD حدود ۸۰٪ و با آن حساسیت به ۹۰٪ می‌رسد. در سال ۲۰۱۱، ساداف و همکارانش [۵] عملکرد ماموگرافی دیجیتال کامل را با ابزار CAD افزوده شده، مطالعه کردند. این مطالعه نشان داده است که ترکیب سیستم CAD با ماموگرافی منجر به شناسایی سرطان‌ها به‌عنوان میکروالکیفیت‌ها با حساسیت ۱۰۰٪ و میزان حساسیت ۸۶٪ برای سایر موارد ماموگرافی سرطان شده است.

مرحله طبقه‌بندی در تصاویر ماموگرافی، یک فرآیند استخراج داده است که برچسب‌ها یا کلاس‌ها را به گروه‌های مختلف اختصاص می‌دهد که هدف آن کشف و استخراج الگوهای پنهان از مجموعه داده‌های بزرگ با استفاده از تکنیک‌های مختلف یادگیری ماشین [۶] است. در اینجا بررسی سیستماتیکی روی روش‌های جدید ارائه شده در زمینه سیستم‌های کمک به-تشخیص برای تصاویر ماموگرافی انجام می‌گیرد.

در سال ۲۰۱۷، گاردزی و همکارانش [۷] طبقه‌بندی ماموگرام با استفاده از پراکندگی زمان پویا را پیشنهاد داده‌اند. در این روش، نواحی ROI به دو کلاس نرمال و غیرنرمال طبقه‌بندی شده است. تکنیک یادگیری ماشین مورد استفاده در این مقاله روش KNN است. نتایج ارزیابی این سیستم در این مقاله به‌صورت معیار دقت ۸۱,۹۲٪ و معیار حساسیت ۸۵,۹۲٪ گزارش شده است.

در سال ۲۰۱۶، شی و همکارانش [۸] روشی ارائه داده‌اند که در آن توده‌های پستان موجود در تصاویر ماموگرافی دیجیتالی با استفاده از الگوریتم یادگیری ماشین طبقه‌بندی می‌شوند. در این سیستم توده‌های خوش‌خیم از توده‌های بدخیم جدا می‌شوند.

در سال ۲۰۱۷، خان و همکارانش [۹] روشی ارائه کرده‌اند که مقایسه روش‌های استخراج ویژگی مختلف گابور برای طبقه‌بندی توده در ماموگرافی انجام گرفته است. در این روش طبقه‌بندی بافت طبیعی و توده‌ای که خوش‌خیم و بدخیم است انجام شده است و در این روش از تکنیک SVM برای طبقه‌بندی استفاده شده است.

در سال ۲۰۱۲ دیهبا و همکارانش [۱۰] یک سیستم پشتیبانی بهبود یافته برای تشخیص ضایعات در تصاویر ماموگرام با استفاده از شبکه عصبی موجک بهینه‌سازی شده با الگوریتم تکامل تفاضلی ارائه کرده‌اند. این روش با معیار حساسیت ۹,۹۶٪ و SP برابر با ۹,۹۲٪ و معیار AUC برابر ۹۷,۰ گزارش شده است. این روش بر روی تصاویر دیتاست MIAS اعمال شده است.

۲- بیان مسئله

چهار عملیات پایه، به منظور تفکیک توده سرطانی، به صورت زیر مطرح می شود:

۲-۱- ساییدگی (Erosion)

عملیات کوتاه سازی یا نازک سازی را انجام می دهد. به این صورت که ساییدگی f به وسیله h ، مجموعه ای از تمام نقاط s می باشد که h ، به وسیله Z منتقل می شود، در f قرار دارد که در زیر نشان داده می شود:

$$\text{Erosion:} \quad (f \ominus h)(x, y) = \inf_{(r, a) \in H} \{f(x+r, y+s) - h(r, s)\} \quad (1)$$

۲-۲- انبساط (Dilation)

انبساط برخلاف ساییدگی، اشیاء موجود در تصویر را ضخیم کرده یا رشد می دهد. روش و بسط خاص این ضخیم کردن، تحت کنترل شکل عنصر سازنده مورد استفاده است. در شکل زیر انبساط f به وسیله عنصر سازنده h را نمایش می دهد که به صورت زیر است:

$$\text{Dilation:} \quad (f \oplus h)(x, y) = \sup_{(r, a) \in H} \{x-r, y-s + h(r, s)\} \quad (2)$$

۲-۳- بازکردن و بستن (Opening & Closing)

بازکردن، بخش های بلا تکلیف باریک را می شکند و همین طور برآمدگی های نازک را حذف می کند. با این کار به طور معمول باعث هموار شدن منحنی شیء می شود. بستن نیز باعث هموار ساختن قسمت هایی از منحنی ها می شود، اما با این تفاوت که، به طور معمول شکستگی های باریک و شکاف های نازک را می سوزاند. روزه های کوچک را حذف کرده و فاصله هایی که در منحنی وجود دارد را پر می کند. باز کردن: تعریف مجموعه f به وسیله عنصر سازنده h به صورت زیر می باشد:

$$\text{Opening: } f \circ h = (f \ominus h) \oplus h \quad (3)$$

به این صورت که بازکردن f به وسیله h ، ساییدگی f توسط h می باشد و پس از آن انبساط نتیجه، به وسیله h خواهد بود. همچنین، تعریف بستن مجموعه f به وسیله عنصر سازنده h به صورت زیر می باشد:

$$\text{Closing: } f \bullet h = (f \oplus h) \ominus h \quad (4)$$

به این صورت که بستن f به وسیله h ، با انبساط f به وسیله h برابر است و پس از آن ساییدگی نتیجه، به وسیله h خواهد بود.

۲-۴- سیستم پیشنهادی جهت تشخیص توده در تصاویر ماموگرافی

جهت پردازش بهتر و یکسان سازی تصاویر موجود در پایگاه داده، عملیات نرمالیزه سازی تصویر و بهبود کنتراست تصویر انجام می شود. در بخش نرمالیزه سازی، تصویر ورودی را طبق رابطه زیر از مقدار ماکزیمم سطح خاکستری تصویر کم می کنیم تا یک نوع یکسان سازی سطوح خاکستری در تمامی تصاویر پایگاه داده انجام شود:

$$image = \max(image) - image \quad (5)$$

در ادامه برای برجسته شدن جزئیات تصویر و نواحی توده، کنتراست تصویر را توسط الگوریتم متعادل سازی هیستوگرام بهبود می دهیم. الگوریتم متعادل سازی هیستوگرام، یکی از الگوریتم های توانمند در بهبود کنتراست تصاویر است. برای یک تصویر داده شده X ، تابع احتمال توزیع PDF (هیستوگرام) به صورت زیر تعریف می شود:

$$p(X_k) = \frac{n_k}{n}, \quad k = 0, 1, \dots, L-1 \quad (6)$$

که n_k تعداد پیکسل هایی است که سطح روشنایی X_k در تصویر ورودی دارد و n تعداد کل پیکسل های تصویر ورودی است.

بر اساس تابع توزیع احتمال، تابع توزیع تجمعی (CDF) به صورت زیر تعریف می شود:

$$C(X_k) = \sum_{j=0}^k p(X_j), \quad 0 \leq C(X_k) \leq 1 \quad (7)$$

که $X_k = x$ برای $k = 0, 1, \dots, L-1$ است. متعادل سازی هیستوگرام، روشی است که تصویر ورودی را به کل دامنه پویای $(X_0$ تا $X_{L-1})$ با استفاده از CDF به عنوان تابع انتقال، نگاشت می کند.

اگر فرض کنیم یک تابع انتقال $f(x)$ بر اساس تابع توزیع تجمعی به صورت زیر تعریف می شود:

$$f(x) = X_0 + (X_{L-1} - X_0)c(x) \quad (8)$$

سپس تصویر خروجی این روش می تواند به صورت زیر بیان شود:

$$Y = f(x) \quad (9)$$

$$= \{f(X(i, j)) \mid \forall X(i, j) \in X\} \quad (10)$$

۲-۵- طبقه بندی نوع توده

برای طبقه بندی توده ها ما از طبقه بند ماشین بردار پشتیبان (SVM) استفاده کرده ایم. SVM یکی از روش های یادگیری بانظارت می باشد که از آن به منظور طبقه بندی و همچنین رگرسیون استفاده می شود. SVM به یک تابع کرنل خوب و انتخاب پارامتر C نیاز دارد.

اگر مجموعه داده‌های آموزش D شامل n عضو (نقطه) باشد، تعریف به این صورت خواهد بود:

$$D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^n, y_i = \{-1, 1\}\}_{i=1}^n \quad (11)$$

که مقدار y برابر ۱ یا -۱ است و هر، یک بردار حقیقی p بعدی می‌باشد.

هدف الگوریتم SVM، یافتن ابرصفحه جداکننده با بیشترین فاصله از نقاط حاشیه‌ای می‌باشد که نقاط با $y_i = 1$ را از نقاط با $y_i = -1$ را تفکیک کند (شکل ۵-۳).

همان‌طور که در شکل (۵-۳) مشاهده می‌شود داده‌ها شامل دو کلاس سفید و سیاه هستند که به صورت خطی نیز تفکیک می‌شوند. چندین خطوط می‌تواند وجود داشته باشد که دو کلاس سیاه و سفید را از هم جدا کند (شکل ۵-۳-الف). روش بردار پشتیبان بهترین خط یا صفحه جدا کننده کلاس را پیدا می‌کند (شکل ۵-۳-ب).

هر ابر صفحه می‌تواند به صورت مجموعه‌ای از نقاط باشد، در صورتی که شرط زیر را شامل شود:

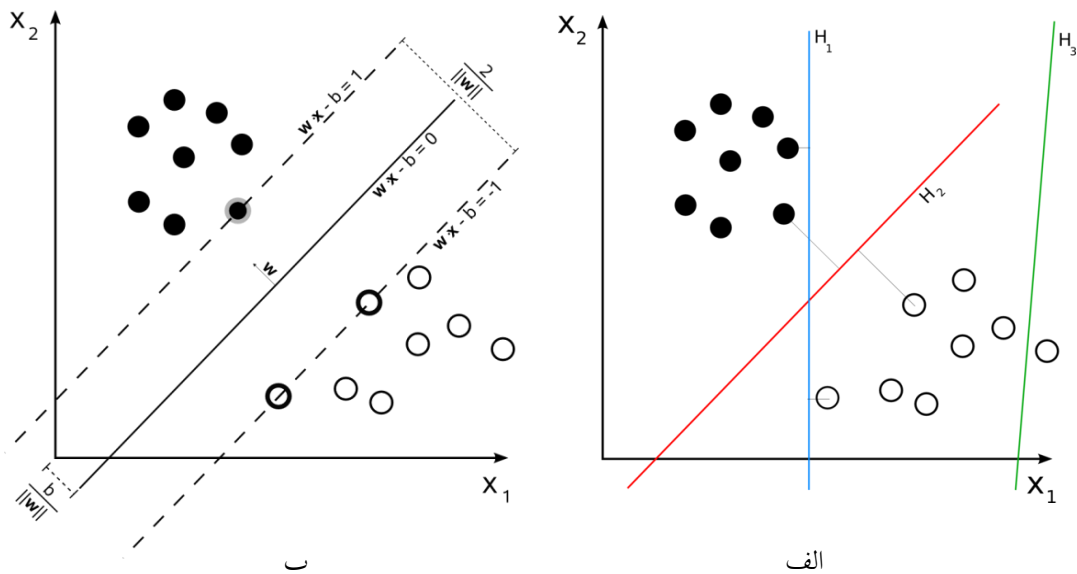
$$w \cdot x - b = 0 \quad (11)$$

علامت. نشانه ضرب و نشانه بردار نرمال می‌باشد، که به ابر صفحه عمود شده است.

هدف SVM آن است که w و b را به طوری انتخاب کند که بیشترین فاصله میان ابر صفحه‌های موازی که داده‌ها را از یکدیگر جدا می‌کنند، ایجاد کنند.

این ابر صفحه‌ها به وسیله دو رابطه زیر توصیف می‌شوند:

شکل ۱: نقش SVM در بدست آوردن ابر صفحه جدا کننده بهینه



$$w \cdot x - b = 1 \quad (12)$$

$$w \cdot x - b = -1 \quad (13)$$

اگر داده‌های آموزشی جدایی‌پذیر خطی باشند، می‌توانیم دو ابر صفحه را در حاشیه نقاط به طوری که هیچ‌گونه نقطه مشترکی نداشته باشند، در نظر بگیریم و بعد، فاصله آنها را ماکزیم کنیم. اگر به کمک هندسه، فاصله این دو صفحه را محاسبه کنیم، مقدار بدست می‌آید.

هدف در SVM این است که فاصله بین ابر صفحه‌ها ماکزیم باشد.

بنابراین باید برای افزایش فاصله مینیمم شود. با هدف این که از ورود نقاط، به حاشیه جلوگیری کنیم، این شرایط نیز اضافه می‌شود:

$$\begin{cases} w \cdot x_i - b \geq 1 & \text{for } x_i \in \text{Class1} \\ w \cdot x_i - b \leq -1 & \text{for } x_i \in \text{Class2} \end{cases} \quad (14)$$

این شرایط می‌تواند به صورت زیر نوشته شود:

$$y_i(w \cdot x_i - b) \geq 1 \quad \text{for } i = 1 \dots n \quad (15)$$

با کنار هم قرار دادن این دو شرط یک مسئله بهینه‌سازی بدست می‌آید:

$$\begin{cases} y_i(w \cdot x_i - b) \geq 1 & \text{for } i = 1 \dots n \\ \text{mi minimize } \left\{ \frac{1}{2} \|w\|^2 \right\} \end{cases} \quad (16)$$

می‌توان عبارت پیشین را به وسیله ضرایب نامنفی لاگرانژ به شکل زیر بیان نمود که در آن ضرایب لاگرانژ می‌باشند. همچنین نتیجه عددی تشخیص توده توسط روش پیشنهادی بر اساس معیارهای ارزیابی در جدول زیر ذکر شده‌اند. نتایج نشان‌دهنده بهبود مناسبی در تشخیص نواحی تومور در تصاویر ماموگرافی است:

جدول ۱: نتایج شناسایی توده توسط روش پیشنهادی

روش	Sensitivity	FP-rate
روش پیشنهادی	٪۹۴.۶۲	٪۴.۲

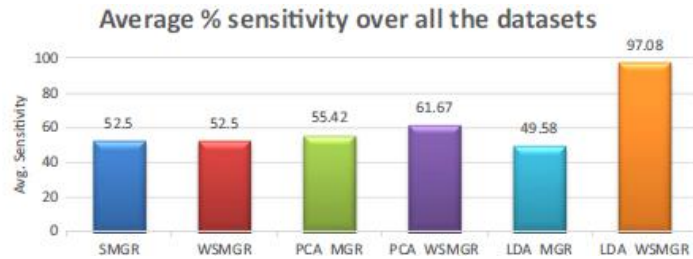
همان‌طور که در جدول مشاهده می‌شود، نتایج حاصل از شبیه‌سازی به ازای هر دو معیار ارزیابی بهبود مناسبی داشته که بیانگر کارایی روش پیشنهادی است.

همچنین نتایج طبقه‌بندی توده به دو نوع بدخیم و نرمال در جدول زیر ذکر شده است:

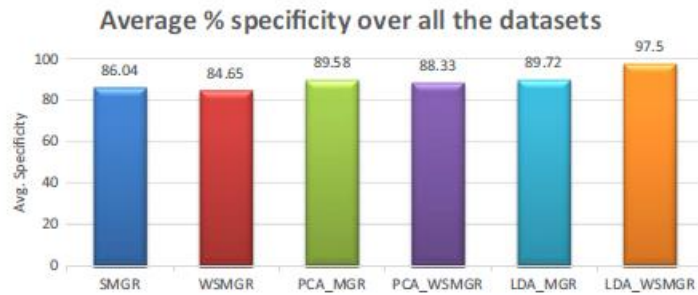
جدول ۲: نتایج طبقه‌بندی توده توسط روش پیشنهادی و مقایسه با سایر روش‌ها

روش	Accuracy
روش مقاله [۱۲]	۹۰.۷۲٪
روش مقاله [۱۳]	۹۳.۱۷٪
روش پیشنهادی	۹۵.۶۵٪

شکل ۲: نتیجه حاصل از این مقاله براساس معیار Sensitivity [۹]

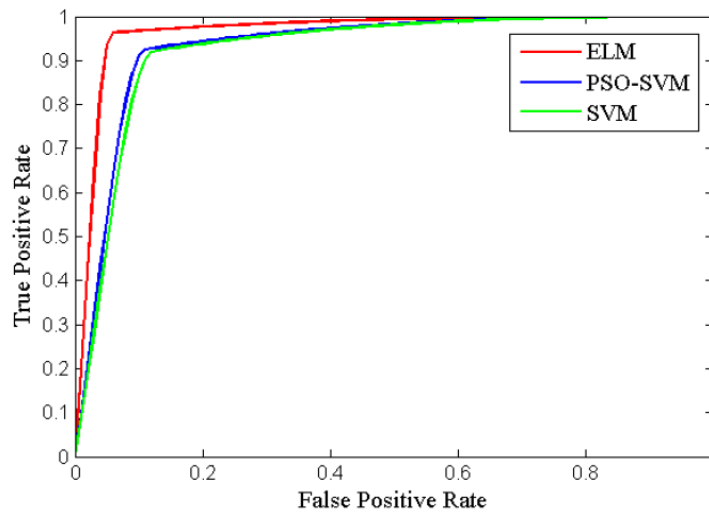


شکل ۳: نتیجه حاصل از این مقاله براساس معیار Sensitivity [۹]



نمودار ROC حاصل از روش پیشنهادی این مقاله نیز در شکل زیر نشان داده شده است:

شکل ۳: نمودار ROC





۳- نتیجه گیری

در این مقاله به منظور شناسایی نواحی توموری در تصاویر ماموگرافی پستان و همچنین طبقه بندی تومورهای شناسایی شده به دو دسته بدخیم و خوش خیم سیستمی پیشنهاد شده است که به تفصیل، جزئیات آن بیان شده است. سیستم پیشنهادی مورد نظر شامل بخش‌های پیش پردازش، ناحیه بندی و تشخیص توده، حذف نواحی کاذب، استخراج ویژگی از نواحی مورد نظر، کاهش فضای ویژگی و در نهایت طبقه بندی نوع توده می باشد.

داده‌های به کار گرفته شده برای مراحل مختلف شبیه سازی روش پیشنهادی و بررسی نتایج، تصاویر مربوط به پایگاه داده MIAS می باشد.

نتیجه عددی تشخیص توده توسط روش پیشنهادی، براساس معیارهای ارزیابی Sensitivity و FP-rate به ترتیب به مقدار ۹۴.۶۲٪ و ۴.۲٪ رسیده است. این نتایج نشان دهنده دقت مناسبی در تشخیص نواحی توموری در تصاویر ماموگرافی پستان از دیتاست MIAS می باشد.

همچنین دقت حاصل از طبقه بندی توده به دو نوع بدخیم و نرمال به ازای روش پیشنهادی به مقدار ۹۵.۶۵٪ رسیده است که نسبت به دیگر مقالات جدید موجود در این زمینه، بهبود داشته است.



مراجع

- [1] National Cancer Institute (NCI). Available: <http://www.cancer.gov/cancertopics>
- [2] I. Leichter, et al., 2000. "Optimizing parameters for computer-aided diagnosis of microcalcifications at mammography". Academic radiology.
- [3] J. Tang, , et al, 2009. "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances". IEEE Transactions on Information Technology in Biomedicine.
- [4] A. Horsch, A. Hapfelmeier, and M. Elter, 2011. "Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies". International journal of computer assisted radiology and surgery. Springer.
- [5] A. Sadaf, et al, 2011. "Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers". European journal of radiology.
- [6] A. Jalalian, et al, 2013. "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review". Clinical imaging.
- [7] S.J.S. Gardezi, et al, 2018. "Mammogram classification using dynamic time warping". Multimedia Tools and Applications. Springer.
- [8] S. Khan, et al, 2017. "A comparison of different Gabor feature extraction approaches for mass classification in mammography". Multimedia Tools and Applications. Springer.
- [9] W. Xie, Y. Li, and Y. Ma. 2016. "Breast mass classification in digital mammography based on extreme learning machine". Neurocomputing.
- [10] F. Burling-Claridge, M. Iqbal, and M. Zhang, 2016. "Evolutionary algorithms for classification of mammographic densities using local binary patterns and statistical features. in Evolutionary Computation (CEC)", IEEE.
- [11] Dheeba, J. and S., 2012. "Tamil Selvi, An Improved Decision Support System for Detection of Lesions in Mammograms Using Differential Evolution Optimized Wavelet Neural Network. Journal of Medical Systems". Springer.
- [12] W. Xie, Y. Li, and Y. Ma. 2016. "Breast mass classification in digital mammography based on extreme learning machine". Neurocomputing.
- [13] F. Burling-Claridge, M. Iqbal, and M. Zhang, 2016. "Evolutionary algorithms for classification of mammographic densities using local binary patterns and statistical features. in Evolutionary Computation (CEC)", IEEE.