



## روش تشخیص سرطان سینه با استفاده از ماشین های بردار پشتیبان بهبود یافته توسط الگوریتم بهینه سازی شاهین هریس (HHO)

مریم سلطان محمدی<sup>۱\*</sup>، مهدی خزاعی پور<sup>۲</sup>

<sup>۱</sup> دانشجوی دکترا، دانشکده فنی و مهندسی (گروه مهندسی کامپیوتر)، واحد بیرجند، دانشگاه آزاد اسلامی، بیرجند، ایران.

[maryam\\_soltanmohammadi@yahoo.com](mailto:maryam_soltanmohammadi@yahoo.com)

<sup>۲</sup> استادیار، دانشکده فنی و مهندسی (گروه مهندسی کامپیوتر)، واحد بیرجند، دانشگاه آزاد اسلامی، بیرجند، ایران.

[mkhazaiepoor@gmail.com](mailto:mkhazaiepoor@gmail.com)

### چکیده

سرطان سینه یکی از بیماری‌های کشنده در بین زنان است و سالانه میلیون‌ها نفر در سراسر جهان بر اثر این بیماری جان خود را از دست می‌دهند. سرطان سینه اگر در مراحل اولیه تشخیص داده شود، شانس زنده بودن بیماران را افزایش خواهد داد. یکی از روش‌های تشخیص بیماری سرطان سینه، استفاده از روش‌های کشف دانش مانند یادگیری ماشین است. روش‌های یادگیری ماشین می‌توانند با تجزیه و تحلیل اطلاعات بیماران و پرونده آنها، الگوی بیماری سرطان سینه را کشف نمایند. مزیت مهم استفاده از روش‌های یادگیری ماشین برای تشخیص سرطان سینه، کاهش هزینه‌های تشخیص و کمک به تشخیص دقیق‌تر توسط پزشکان متخصص است. یکی از روش‌های تشخیص بیماری سرطان سینه استفاده از ماشین بردار پشتیبان است. ماشین بردار پشتیبان یک روش برای طبقه‌بندی نمونه‌ها با هدف کاهش ریسک عملیاتی در طبقه‌بندی است. یکی از چالش‌های مهم ماشین بردار پشتیبان، خطای خروجی مدل به دلیل عدم انتخاب بهینه پارامترهای یادگیری است. در روش پیشنهادی برای کاهش دادن خطای طبقه‌بندی افراد بدخیم و خوش خیم، از الگوریتم بهینه‌سازی شاهین استفاده شده است. نقش الگوریتم شاهین در روش پیشنهادی بهینه‌سازی پارامترهای ماشین بردار پشتیبان برای کاهش دادن خطای تشخیص بیماران بدخیم است. ارزیابی‌ها در محیط برنامه‌نویسی متلب و روی مجموعه داده ویسکانسین انجام شده است. ارزیابی‌ها نشان می‌دهد روش پیشنهادی در تشخیص سرطان سینه دارای دقتی برابر ۹۹.۳۱٪ است و از روش‌های نظیر الگوریتم بهینه‌سازی وال در تشخیص سرطان سینه دقیق‌تر است.

**واژه‌های کلیدی:** سرطان سینه، یادگیری ماشین، ماشین بردار پشتیبان، الگوریتم بهینه‌سازی شاهین هریس، بهینه‌سازی،

پیش بینی

## ۱. مقدمه

سرطان سینه<sup>۱</sup> توسط بافت‌های در حال رشد در سلول‌های سینه شناسایی می‌شود و یکی از شایع‌ترین سرطان‌ها در زنان و در سراسر جهان پس از سرطان ریه در نظر گرفته می‌شود. مطالعات نشان می‌دهد در آمریکا، هر ساله تقریباً ۳۰ درصد از موارد جدید زنان مبتلا به سرطان سینه تشخیص داده می‌شوند. میزان مرگ و میر سالانه ۱۹۰ نفر در هر ۱۰۰۰۰۰ زن است. دو نوع رایج سرطان سینه عبارتند از کارسینوم مجرای درجا<sup>۲</sup> و کارسینوم مجرای تهاجمی<sup>۳</sup> می‌باشند. موارد شناسایی شده کارسینوم مجرای درجا درصد کمی از بیماران و در حدود ۲ درصد را شامل می‌شود. علاوه بر این، کارسینوم مجرای تهاجمی بسیار شایع‌تر بوده و خطرناک‌تر است، زیرا سلول‌های سینه را در بر می‌گیرد و در این دسته میزان مرگ و میر ۱۰ نفر در هر ۱۰۰ است [۱]. علاوه بر این، کارسینوم مجرای تهاجمی در انواع مختلفی از سلول‌ها وجود دارد که تشخیص آن تقریباً غیرممکن است. سلول‌های غیرطبیعی که تومور نیز نامیده می‌شوند، دارای اشکال نامنظم و دلخواه هستند. علاوه بر این، این سلول‌های تومور به دو نوع اصلی بدخیم<sup>۴</sup> و خوش‌خیم<sup>۵</sup> طبقه‌بندی می‌شوند. در ابتدا، تومور بدخیم در سلول‌های بافت اطراف خود گسترش می‌یابد و برای سلول‌های بافت سالم مشکل ایجاد می‌کند. برخلاف اولی، دومی یک سلول بافتی غیر سرطانی است و به بافت‌های همسایه خود آسیب نمی‌رساند. حتی با تجربه‌ترین آسیب‌شناسان در تشخیص تفاوت در ساختار بافت‌های سینه از نظر نوع بدخیم و خوش‌خیم با چالش مواجه می‌باشند [۲].

مطالعات و داده‌های جهانی سرطان نشان می‌دهد که سرطان سینه پس از سرطان ریه، دومین شکل مرگبار سرطان در سراسر جهان است. در سال ۲۰۱۸، در حدود ۲ میلیون مورد جدید سرطان سینه در سراسر جهان گزارش شد که در آن ۶۲۷۰۰۰ نفر جان خود را از دست دادند. مطالعه‌ای در استرالیا نشان داد که بقای بیماران دارای سرطان سینه به شدت با اندازه تومور در زمان تشخیص مرتبط است، با اندازه کمتر از ۱۰ میلی‌متر، احتمال زنده ماندن بیمار ۹۸٪ است. یک مطالعه نشان داد که ۷۰ درصد موارد سرطان سینه زمانی شناسایی می‌شوند که اندازه تومور ۳۰ میلی‌متر باشد. سرطان سینه معمولاً زمانی که اندازه تومور حداقل ۲۰ میلی‌متر باشد در طول غربالگری قابل تشخیص است [۳]. بنابراین، تشخیص زودهنگام سرطان سینه برای تسهیل درمان زودهنگام بسیار مهم است. درمان اولیه ممکن است پس از شناسایی از طریق معاینات غربالگری مانند معاینه بالینی پستان<sup>۶</sup> [۴] و خودآزمایی پستان<sup>۷</sup> [۵] مفید باشد. معاینه بالینی پستان یک معاینه پزشکی منظم است که توسط متخصصان مراقبت‌های بهداشتی برای تشخیص ضایعات پستان انجام می‌شود، در حالی که خودآزمایی پستان توسط یک فرد برای مشاهده تغییرات فیزیکی و ظاهر سینه‌ها انجام می‌شود. عمل خودآزمایی پستان به زنان این امکان را می‌دهد که مسئولیت سلامت خود را بر عهده بگیرند و امروزه خودآزمایی پستان توسط سازمان بهداشت جهانی برای افزایش آگاهی زنان در معرض خطر توصیه می‌شود. روش‌های غربالگری، تصاویر پزشکی از سینه‌ها تولید می‌کنند. تفسیر این تصاویر معمولاً توسط متخصصان انسانی مانند رادیولوژیست‌ها و پزشکان انجام می‌شود. تحقیقات نشان می‌دهد که دقت تشخیصی پایین، به دلیل ضعف فنی و تخصص در تفسیر چنین تصاویری است. ظهور بیماری‌های مختلف و نیروی انسانی محدود، محققان و پرسنل پزشکی را ترغیب کرده است تا از فناوری به کمک رایانه برای تسهیل تشخیص مبتنی بر ترموگرافی پستان و در نتیجه به حداقل رساندن خطاها استفاده کنند [۶].

بنابراین، یک سیستم کامپیوتری که بتواند به طور خودکار ترموگرام‌ها<sup>۸</sup> را به دو دسته عادی و غیرعادی طبقه‌بندی کند، مورد نیاز است. با توجه به این نیاز، تحقیقات در جهت یافتن راه حل‌های مبتنی بر کامپیوتر برای طبقه‌بندی تصاویر پزشکی یا اطلاعات مرتبط با بیماران به طور مداوم در حال رشد بوده است. بسیاری از روش‌های تشخیص سرطان سینه به کمک رایانه

<sup>1</sup>Breast cancer

<sup>2</sup> Ductal carcinoma in situ (DCIS)

<sup>3</sup> Invasive ductal carcinoma (IDC)

<sup>4</sup> Malignant

<sup>5</sup> Benign

<sup>6</sup> Clinical-breast examination (CBE)

<sup>7</sup> Breast self-examination (BSE)

<sup>8</sup> Thermograms

برای کمک به پزشکان در تفسیر تصاویر پزشکی یا اطلاعات بیماران ایجاد شده‌اند. در طول دهه گذشته، تلاش قابل توجهی بر توسعه مدل‌های یادگیری عمیق<sup>۹</sup> و یادگیری ماشین<sup>۱۰</sup> متمرکز شده است [۷]. از آنجایی که مدل‌های یادگیری عمیق و یادگیری ماشین به صورت عمومی در دسترس هستند، می‌توان آنها را به راحتی با استفاده از شبکه‌های از پیش آموزش دیده اعمال کرد. در تشخیص سرطان پستان، بسیاری از مطالعات بر اساس یادگیری ماشین و یادگیری عمیق توسعه پیدا نموده‌اند که از جمله این روش‌ها می‌توان شبکه عصبی مصنوعی [۸]، ماشین بردار پشتیبان [۹]، درخت تصمیم‌گیری [۱۰]، جنگل تصادفی [۱۱] و شبکه عصبی کانولوشن [۱۲] غیره اشاره نمود. امروزه برای تشخیص سرطان سینه از اطلاعات و پرونده بیماران برای تحلیل الگوهای سرطان سینه بخوبی استفاده می‌شود و می‌توان از این اطلاعات با ارزش برای آموزش مدل‌های یادگیری استفاده نمود. در بسیاری از پژوهش‌ها برای کاهش دادن خطای خروجی یادگیری ماشین و یادگیری عمیق برای تشخیص سرطان سینه از روش‌های بهینه‌سازی و الگوریتم‌های فراابتکاری<sup>۱۱</sup> استفاده شده است. به عنوان مثال، از الگوریتم ژنتیک<sup>۱۲</sup> [۱۳]، الگوریتم بهینه‌سازی ذرات<sup>۱۳</sup> [۱۴] و الگوریتم کرم شب‌تاب<sup>۱۴</sup> [۱۵] برای کاهش دادن خطای یادگیری ماشین در تشخیص سرطان سینه استفاده شده است. یکی از روش‌های مهم یادگیری ماشین برای تشخیص بیماری‌ها، ماشین بردار پشتیبان است اما این روش دارای چالش‌های ذیل برای تشخیص سرطان سینه است:

- ماشین بردار پشتیبان برای آنکه خطای خروجی را در تشخیص سرطان سینه کاهش دهد نیاز به بهینه‌سازی پارامترهای یادگیری دارد.
- ماشین بردار پشتیبان بر خلاف روش‌های یادگیری عمیق فاقد فاز انتخاب ویژگی است و از این جهت مستعد افزایش خطا در تشخیص نمونه‌های بدخیم از خوش خیم است.
- برای رفع این چالش‌ها روش پیشنهادی در این مقاله برای کاهش دادن خطای تشخیص افراد بدخیم و خوش خیم ارائه می‌شود. نوآوری روش پیشنهادی در آن است پارامترهای ماشین بردار پشتیبان با الگوریتم بهینه‌سازی شاهین هریس انتخاب می‌شود. نوآوری دیگر این پژوهش در آن است که ورودی‌های ماشین بردار پشتیبان نیز با الگوریتم بهینه‌سازی شاهین هریس<sup>۱۵</sup> [۱۶]، بهینه‌سازی می‌شود. در اینجا یک نسخه باینری از الگوریتم شاهین برای انتخاب ویژگی‌های مهم نمونه‌ها ایجاد می‌شود. هدف اصلی این مقاله ارائه یک رویکرد جدید و با خطای کمینه برای تشخیص افراد بدخیم و خوش خیم بر اساس رویکردهای بهبود یافته یادگیری ماشین است. هدف دیگر مقاله تلفیق روش‌های یادگیری ماشین با روش‌های هوش گروهی برای تشخیص دقیق‌تر افراد بدخیم از خوش خیم است. سهم نویسندگان در این مقاله برای تشخیص سرطان سینه به شرح ذیل ارائه می‌گردد:
- ارائه یک نسخه باینری از الگوریتم شاهین برای انتخاب ویژگی در تشخیص سرطان سینه
- بهینه‌سازی پارامترهای ماشین بردار پشتیبان با الگوریتم شاهین
- استفاده از نسخه بهبود یافته ماشین بردار پشتیبان برای تحلیل اطلاعات و پرونده بیماران برای تشخیص الگوهای پنهان سرطان سینه

این مقاله در چند بخش سازماندهی شده است. در بخش دوم موضوعات مرتبط با الگوریتم شاهین، سرطان سینه و کارهای مرتبط در این زمینه مرور می‌شود و در بخش سوم روش پیشنهادی برای تشخیص سرطان سینه با ترکیب ماشین بردار پشتیبان و الگوریتم شاهین هریس معرفی می‌شود. در بخش چهارم روش پیشنهادی در نرم‌افزار متلب پیاده‌سازی شده و آزمایشات انجام شده با روش‌های مشابه مقایسه می‌شود. در بخش پنجم نتیجه‌گیری تحقیق و پیشنهادات آتی ارائه می‌گردد.

<sup>9</sup> Deep learning

<sup>10</sup> Machine learning

<sup>11</sup> Meta-heuristic algorithms

<sup>12</sup> Genetic algorithm

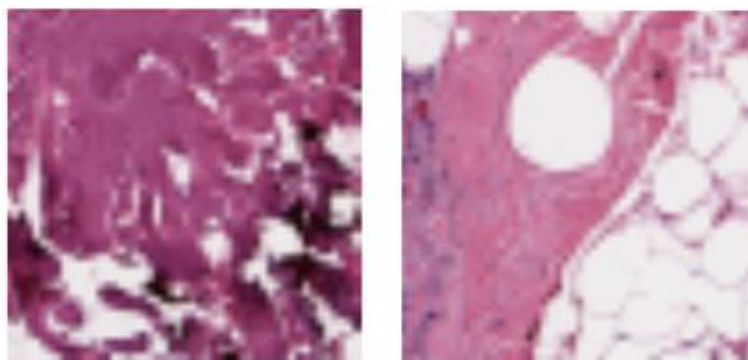
<sup>13</sup> Particle swarm optimization (PSO) algorithm

<sup>14</sup> Firefly algorithm

<sup>15</sup> Harris Hawks Optimization (HHO)

## ۲. پیشینه تحقیق

با دیجیتالی شدن گسترده پرونده‌های سلامت، سیستم‌های تشخیص بیماری به کمک رایانه که از تکنیک‌های داده کاوی و یادگیری ماشین استفاده می‌کنند، به طور فزاینده‌ای برای تشخیص انواع بیماری‌ها از جمله سرطان سینه رایج شده‌اند. سرطان سینه وضعیتی است که در آن سلول‌های سینه به طور غیرقابل کنترل تکثیر می‌شوند. طبق آمار جهانی سرطان، سرطان سینه در زنان به عنوان شایع‌ترین سرطان در جهان پیشی گرفته است و تقریباً ۱۲ درصد از کل موارد سرطان در سراسر جهان را شامل می‌شود [۱۷]. درمان سرطان پستان با تشخیص زودهنگام کمک زیادی به درمان آن می‌کند، اما اگر بیماری سرطان سینه به موقع تشخیص داده نشود، شامل اقدامات جراحی یا پزشکی شدید می‌شود. از آنجایی که درمان به مرحله پیشرفت سرطان بستگی دارد، یکی از وظایف اولیه هر پاتولوژیست شامل تجزیه و تحلیل بصری یک اسلاید هیستوپاتولوژیک رنگ‌آمیزی با هماتوکسیلین و ائوزین<sup>۱۶</sup> است. تشخیص اینکه آیا بافت در مرحله کارسینوم مجرای درجا<sup>۱۷</sup> است یا در مرحله کارسینوم مجرای تهاجمی<sup>۱۸</sup> از روی لام ممکن است. یک سرطان زمانی غیر تهاجمی است که در آن سلول‌های سرطانی در پوشش مجرای شیر مادر یافت شده‌اند اما به بافت سینه اطراف گسترش نیافته‌اند. با این حال، در حالت کارسینوم مجرای تهاجمی، سلول‌های سرطانی غیرطبیعی که شروع به تشکیل در مجاری شیر کردند، فراتر از مجاری به سایر قسمت‌های بافت پستان گسترش می‌یابند. کارسینوم مجرای تهاجمی شایع‌ترین نوع سرطان سینه است که مسئول حدود ۷۰ تا ۸۰ درصد از تمام تشخیص‌های سرطان سینه است [۱۷]. مشکل طبقه‌بندی اسلایدهای هیستوپاتولوژیک به (+) IDC و (-) IDC با استفاده از تکنیک‌های پردازش تصویر و به کمک یادگیری ماشین و یادگیری عمیق انجام می‌شود. در شکل (۱)، به ترتیب تصویر سمت چپ از نوع مثبت و تصاویر سمت راست از نوع منفی است.



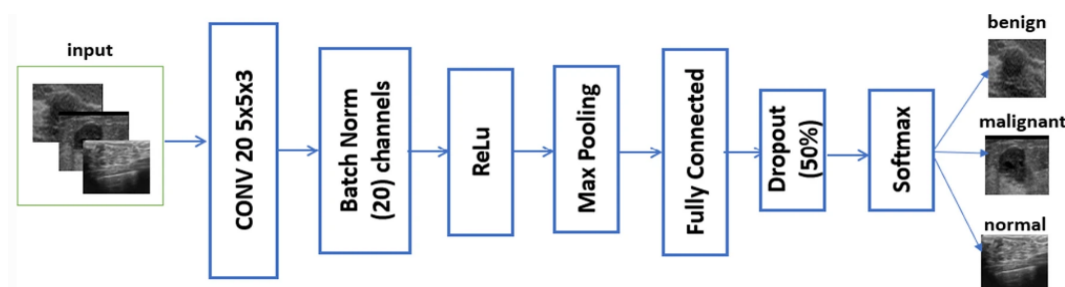
شکل ۱: تشخیص سرطان سینه با پردازش تصویر و یادگیری عمیق [۱۷]

بکارگیری روشهای یادگیری عمیق به جای یادگیری ماشین یک مزیت مهم دارد و آن، عدم نیاز روشهای یادگیری عمیق به انتخاب ویژگی می‌باشند اما روشهای یادگیری ماشین نیاز به مکانیزم انتخاب ویژگی دارند. با وجود آنکه یادگیری عمیق نسبت به یادگیری ماشین دارای دقت بیشتری است اما این دقت به قیمت پیچیدگی بیشتر و نیاز به قدرت محاسباتی بالاتر است. علاوه بر این، برای آموزش صحیح یک مدل یادگیری عمیق، نیاز به مقدار زیادی نمونه ورودی وجود دارد که گاهی اوقات ممکن است برای مشکلات واقعی در دسترس نباشد. در شکل (۲)، یک نمونه از کاربرد روشهای یادگیری عمیق برای تشخیص سرطان سینه نمایش داده شده است. در این روش از شبکه عصبی کانولوشن برای طبقه‌بندی تصاویر به دو دسته خوش خیم و بدخیم استفاده می‌شود [۱۸].

<sup>16</sup> Hematoxylin and eosin (H&E)

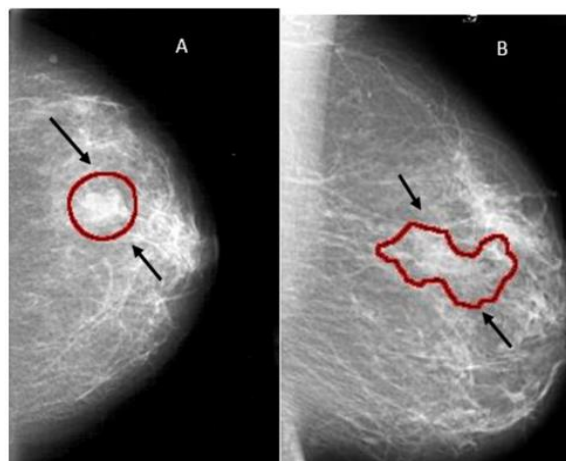
<sup>17</sup> Ductal Carcinoma In Situ (DCIS)

<sup>18</sup> Invasive Ductal Carcinoma (IDC)



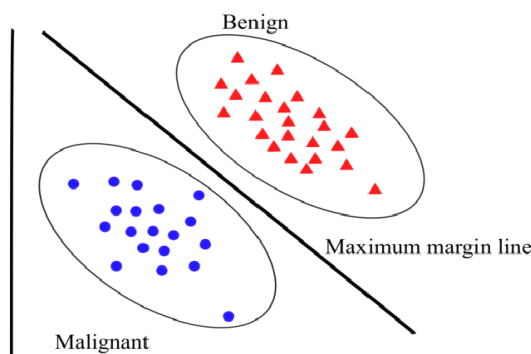
شکل ۲: طبقه بندی تصاویر بدخیم و خوش خیم سینه با یادگیری عمیق [۱۸]

در این نوع از مطالعات تلاش می‌شود تا تصاویر بر اساس ماهیتی که دارند مورد پردازش قرار گرفته شوند. در این روش مشاهده می‌شود که از شبکه CNN برای طبقه‌بندی تصاویر به سه دسته نرمال، خوش خیم و بدخیم استفاده شود. روشهای یادگیری ماشین و یادگیری عمیق در تصاویر علاوه بر طبقه‌بندی می‌تواند در ناحیه‌بندی تصاویر سینه استفاده شود. به عنوان نمونه در شکل (۳)، مشاهده می‌شود که از روشهای ناحیه‌بندی برای تشخیص ناحیه سرطان سینه و بافتهای سرطانی در دو نمونه تصویر استفاده شده است.



شکل ۳: ناحیه بندی تصاویر سینه برای تشخیص تومور در بافتهای سینه [۱۹]

روشهای پردازش تصویر برای تشخیص سرطان سینه فقط بر اساس تصاویر اقدام به تشخیص بیماری سرطان سینه می‌کنند و سایر متغیرهای مرتبط با بیماری سرطان سینه را نادیده می‌گیرند. در برخی از پژوهش‌ها از اطلاعات پرونده بیماران و تصاویر مانند مجموعه داده Wisconsin برای تشخیص بیماری سرطان سینه استفاده می‌کنند [۲۰]. تشخیص بیماران بدخیم و خوش خیم یک مسئله طبقه‌بندی است و می‌توان داده‌های مرتبط با مجموعه داده Wisconsin را با روشهای طبقه‌بندی مانند ماشین بردار پشتیبان حل نمود. در شکل (۴)، برای طبقه‌بندی افراد خوش خیم و بدخیم در ارتباط با سرطان سینه از ماشین بردار پشتیبان استفاده شده است که یک روش طبقه‌بندی کارآمد برای تشخیص بیماری‌ها است.



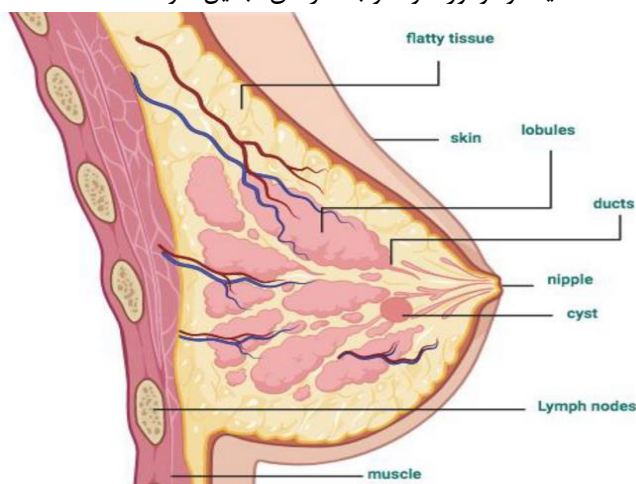
شکل ۴: طبقه بندی نمونه های مجموعه داده Wisconsin برای تشخیص بیماری سرطان سینه [۲۱]

هدف ماشین بردار پشتیبان، بکارگیری یک تفکیک کننده خطی یا غیرخطی برای طبقه‌بندی نمونه‌ها در جهت کاهش دادن ریسک طبقه‌بندی داده‌ها و نمونه‌ها است. در این شکل مشاهده می‌شود که نمونه‌ها توسط تفکیک کننده به دو کلاس بدخیم و خوش خیم طبقه‌بندی شده‌اند.

## ۱.۲. سرطان سینه

شایع‌ترین سرطان در زنان، سرطان سینه است. همانطور که توسط سازمان بهداشت جهانی ادعا شده است، سرطان سینه در ۲.۳ میلیون زن تشخیص داده شده است و ۶۸۵۰۰۰ مرگ در سراسر جهان در سال ۲۰۲۰ ثبت شده است. علاوه بر این، سازمان بهداشت جهانی پیش بینی می‌کند که تعداد بیماران جدید سرطان سینه در بیست سال آینده هفتاد درصد (۷۰ درصد) افزایش خواهد یافت. علاوه بر این، سرطان سینه پنجمین بیماری کشنده در بین انواع سرطان‌های متمایز، مانند سرطان‌های ریه، کولورکتال، کبد و معده است. طبق آمار جهانی سرطان در سال ۲۰۲۰، سرطان پستان زنان شایع‌ترین سرطان است، با موارد جدید سرطان ۲.۳ میلیون (۱۱.۷ درصد از کل موارد) در سال ۲۰۲۰. با این حال، سرطان سینه مردانه یک سرطان بسیار نادر است. بیشتر اوقات، این بیماری پنهان می‌ماند و اکثر بیماران در سطح پیشرفته بیماری در یک کشور جنوب آسیا مانند بنگلادش تشخیص داده می‌شوند. بنابراین یک استراتژی خدمات بهداشتی مقرون به صرفه مورد نیاز است، که برای بسیاری از زنان در کشورهای فقیر مقرون به صرفه خواهد بود [۲۲].

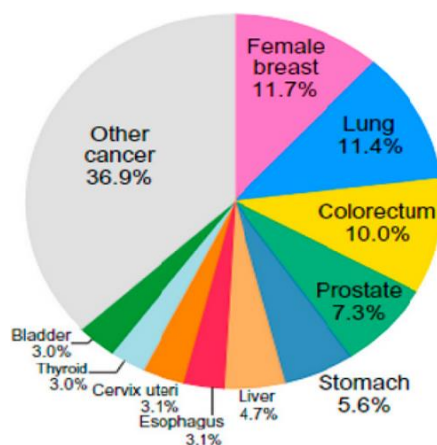
سرطان یک مانع مهم برای افزایش امید به زندگی است و یکی از علل اصلی مرگ و میر در سراسر جهان است. در سال ۲۰۱۹، سازمان جهانی بهداشت گزارش داد که اولین یا دومین دلیل اصلی مرگ زودتر از ۷۰ سالگی، سرطان در ۱۱۲ کشور از ۱۸۳ کشور بیماری سرطان است و در ۲۳ کشور دیگر رتبه سوم یا چهارم را دارد [۲۳]. سرطان باعث رشد نامنظم سلول‌ها می‌شود و اغلب بسته به قسمتی از بدن که در آن رخ می‌دهد نامگذاری می‌شود. سرطان معمولاً به سرعت در سراسر بافت‌های بدن گسترش می‌یابد. در سلول‌ها، کوچکترین واحدهای بافت‌ها و اندام‌های بدن، به عنوان مثال، در سینه شروع می‌شود. سرطان عمدتاً در نتیجه جهش، تقسیم آنارشیک و تکثیر یا تغییرات غیرطبیعی در سلول‌ها ایجاد می‌شود. سلول‌های جدید معمولاً جایگزین سلول‌های قدیمی یا آسیب دیده‌ای می‌شوند که می‌میرند. این فرآیند گاهگاه با شکست مواجه می‌شود و سلول می‌تواند به تقسیم غیرقابل کنترل یا بی‌نظم ادامه دهد و سلول‌های بیشتری شبیه به خود ایجاد کند و باعث ایجاد تومور شود. تومور به خوش خیم (غیر سرطانی) یا بدخیم (سرطانی) تقسیم می‌شود. تومورهای خوش خیم خطرناک نیستند، زیرا باعث سرطان نمی‌شوند: سلول‌های آنها نزدیک به حالت طبیعی به نظر می‌رسند، به کندی رشد می‌کنند و به بافت‌های نزدیک حمله نمی‌کنند و به سایر قسمت‌های بدن آسیب نمی‌رسانند و در مقابل، تومورهای بدخیم خطرناک هستند. اگر آنها بررسی نشوند، در نهایت از تومور اصلی فراتر رفته و به سایر اعضای بدن حمله می‌کنند. در شکل (۶)، آناتومی بافت سینه زنان را نشان می‌دهد و برخی از مناطق می‌تواند مستعد کسیت و تومور شوند و به سرطان تبدیل شوند.



شکل ۶: داخل لوب‌ها، مناطقی وجود دارد که تومورهای اپیتلیال یا کیست رشد می‌کنند [۲۴]



موارد و مرگ و میر بر اساس منطقه جهانی و نوع سرطان تقسیم می‌شوند. شکل (۷)، توزیع جهانی ابتلا برای ۱۰ نوع شایع سرطان در بین زنان در سراسر جهان در سال ۲۰۲۰ نشان می‌دهد [۲۳].



شکل ۷: درصد انواع سرطان در زنان [۲۳]

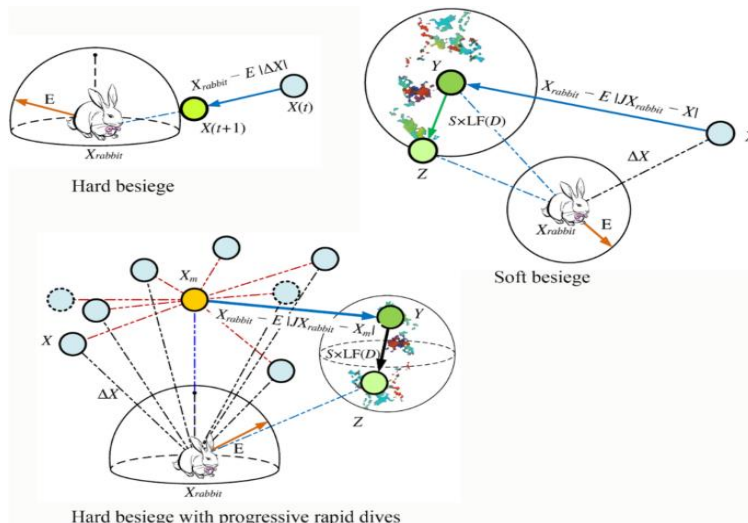
با توجه به نمودار فوق در زنان بیماری سرطان سینه در جایگاه اول قرار دارد و در جایگاه دوم سرطان ریه قرار دارد. نرخ بقای فعلی سرطان سینه در مراحل پایانی کم است و در حدود ۳۰٪ است. تشخیص زودهنگام و دقیق نقش حیاتی در بهبود پیش آگهی و افزایش میزان بقا در بیماران تا ۵۰ درصد دارد. تشخیص صحیح سرطان سینه مستلزم شناسایی صحیح هر مرحله سرطان و همچنین شناسایی دسته آن است. معمولاً برای تشخیص مؤثر سرطان سینه از تصاویر پزشکی متفاوتی نسبت به سایر روش‌های آزمایش سرطان سینه استفاده می‌شود. روش‌های تصویربرداری پزشکی مانند تصاویر هیستوپاتولوژی، تصاویر اشعه ایکس پستان (ماموگرافی)، سونوگرام (تصویربرداری اولتراسوند)، و تصویربرداری تشدید مغناطیسی به طور گسترده برای تشخیص سرطان سینه استفاده می‌شوند. در بیشتر موارد تجربه یک آسیب شناس حرفه‌ای و دانش موضوعی برای تشخیص قابل اعتماد سرطان سینه مورد نیاز است. بدون این موارد، در بیشتر موارد، تشخیص اشتباه، به خصوص در مراحل اولیه سرطان سینه رخ می‌دهد. با این حال، تشخیص سرطان سینه در مراحل اولیه بسیار مهم است. در حقیقت، بسیاری از سیستم‌های تشخیص به کمک رایانه برای کمک به پزشکان در تشخیص زودهنگام تومورهای پستان در ماموگرافی استفاده می‌شوند. این یک سیستم به آسانی در دسترس، سریع، قابل اعتماد و مقرون به صرفه برای تشخیص زودهنگام سرطان سینه است.

سیستم‌های کمک پزشکی در دهه گذشته توسعه یافته است و پیشرفت در دقت تشخیص سرطان سینه بیش از ۲۰٪ نشان داده است. این سیستم به پزشکان و رادیولوژیست ها کمک می‌کند تا با استفاده از روش‌های مختلف تصویربرداری، ناهنجاری‌ها را تشخیص دهند و میزان مرگ و میر را از ۳۰٪ به ۷۰٪ کاهش دهند. سیستم‌های خودکار تشخیص پزشکی مبتنی بر یادگیری عمیق با بررسی داده‌ها در رادیولوژی، آسیب شناسی، قلب، انکولوژی به پیشرفت قابل توجهی در زمینه پزشکی دست یافته‌اند. رویکردهای پیچیده تری در دهه‌های اخیر برای تشخیص سرطان بر اساس تکنیک‌های یادگیری ماشین استفاده شده است. یادگیری عمیق به طور گسترده‌ای به عنوان یکی از این تکنیک‌ها شناخته شده است که کارایی آن را در پیش بینی سرطان و کارایی آن در پیش آگهی را نشان می‌دهد. یادگیری ماشین دارای دقت تشخیص بالاتری برای تشخیص سرطان پستان در ماموگرافی و سونوگرافی تصویربرداری سرطان پستان است [۲۲].

## ۲.۲ الگوریتم بهینه سازی شاهین

هوش گروهی به یک نوع از الگوریتم‌های بهینه‌سازی گفته می‌شود که اعضای جمعیت از اطلاعات گروهی و موقعیت هم بهره جسته و تلاش دارند تا یک مسئله بهینه‌سازی را حل نمایند. در بیشتر الگوریتم‌های فراابتکاری با رویکرد هوش گروهی رفتار شکار گروهی مورد مدلسازی قرار گرفته به گونه‌ای که در این الگوریتم‌ها اعضای جمعیت پیرامون طعمه یا نقطه بهینه فعلی

حلقه زده و تلاش می‌نمایند تا پیرامون آن را مورد جستجو قرار داده و در فرصت مناسب به سمت شکار حمله نمایند. رفتار شکار گروهی در طبیعت در بین جانداران مختلف دیده می‌شود که برخی از آنها از رفتار حشرات، بندپایان و پستانداران الگوبرداری شده است. سیستم‌های هوش گروهی می‌تواند مشتمل بر رفتار چند جاندار باشند که با هم کار می‌نمایند که تعداد آنها می‌تواند زیاد نباشد که یک نمونه از این نوع سیستم‌ها الگوریتم بهینه‌سازی عقاب است که در سال ۲۰۱۹ ارائه و مدلسازی شده است. در این نوع جانداران به طور معمول گروه‌هایی تا ۶ پرنده برای شکار با هم مشارکت نموده و پیرامون طعمه پرواز نموده و آن طبق انواع شیرجه به سمت طعمه مانند شکل (۸)، شکار می‌نمایند.



شکل ۸: انواع استراتژی جستجو در الگوریتم بهینه‌سازی شاهین [۲۴]

در رفتار این جانداران مشاهده می‌شود یک گروه کوچک در ابتدا برای شکار پیش قدم شده سپس در ادامه دیگر از اعضای گروه نیز برای مشارکت در شکار پیش می‌روند و در شکار گروهی با هم مشارکت و همکاری می‌نمایند. در این روش شکار همه عقاب‌ها یا راه‌حلهای مسئله در اطراف طعمه یا جواب بهینه پراکنده می‌شوند و یک پرنده آن را شکار می‌نماید. در این الگوریتم مشاهده می‌شود در ابتدا طعمه شناسایی سپس پیرامون آن احاطه شده و به آن حمله می‌گردد. در این الگوریتم هر عقاب یک راه‌حل مسئله بوده و راه‌حل بهینه فعلی را نیز به عنوان موقعیت خرگوش در نظر گرفته و عقاب‌ها به سمت آن پرواز می‌نمایند. در این الگوریتم‌ها در ابتدا عقاب‌ها فضای مسئله را جستجو نموده تا طعمه را پیدا نمایند و سپس به سمت آن حمله نمایند [۱۶].

### ۳.۲. مطالعات مرتبط

در پژوهش [۲۵]، در سال ۲۰۲۰، یک روش ناحیه‌بندی کاملاً خودکار مبتنی بر آستانه بهبود یافته و قابل آموزش برای تشخیص نواحی سرطان سینه در تصاویر ماموگرافی ارائه دادند. هدف این مطالعه پیشنهاد یک مدل تقسیم‌بندی مبتنی بر آستانه و قابل آموزش برای استخراج ROI است. یک رویکرد ناحیه‌بندی ترکیبی برای مرز ناحیه سینه و عضله سینه در تصاویر ماموگرافی بر اساس تکنیک‌های آستانه‌گذاری و یادگیری ماشین ارائه شده است. در این روش برای تخمین مرز سینه، ناحیه سینه با حذف باندهای تبدیل موجک برجسته شده است. در این روش مرز اولیه سینه از طریق تکنیک آستانه‌گذاری تعیین شده است. در این پژوهش نتایج تجربی با ناحیه‌بندی دستی بر اساس ویژگی‌های بافت مختلف مقایسه شده است. علاوه بر این، ارزیابی و مقایسه برای مرز ناحیه سینه و ناحیه‌بندی عضلات سینه‌ای به طور جداگانه انجام شده است. نتایج تجربی نشان داد که مرز ناحیه سینه و رویکرد ناحیه‌بندی عضلات سینه‌ای دارای دقتی ۹۹.۳۱ درصد برای مرز تقسیم‌بندی ناحیه سینه و دقت ۹۸.۶۴ درصد برای تقسیم‌بندی عضلات سینه‌ای است. علاوه بر این، این پژوهش همچنین حقیقت پایه را به عنوان ارزیابی شباهت همه جانبه گنجانده است. در کلینیک، این تجزیه و تحلیل ممکن است به عنوان یک پشتیبانی ارزشمند برای شناسایی سرطان پستان ارائه شود.



در پژوهش [۲۶]، در سال ۲۰۲۰، یک روش ناحیه‌بندی سرطان پستان با استفاده از تصاویر حرارتی و مبتنی بر الگوریتم گروهی سالپ<sup>۱۹</sup> ارائه دادند. مطالعات آنها نشان می‌دهد الگوریتم سالپ برای حل مسائل بهینه‌سازی تک هدفه دارای دقت بالایی است، اما از نرخ همگرایی پایین و رکود بهینه محلی در مسائل پیچیده از چالش‌های این الگوریتم است. در روش پیشنهادی، یک الگوریتم ناحیه‌بندی با استفاده از روش جابجایی سریع برای استخراج پیکسل‌هایی که پارامترهای آن توسط الگوریتم سالپ بهینه‌سازی شده‌اند، فرمول‌بندی می‌شود. روش جابجایی سریع با خوشه‌بندی پیکسل‌های تصویر حرارتی سینه، پیکسل‌های فشرده و تقریباً یکنواخت تولید می‌کند. الگوریتم سالپ بر اساس ده نقشه آشفته<sup>۲۰</sup> توسعه یافته است تا نرخ همگرایی الگوریتم سالپ را افزایش دهد. الگوریتم پیشنهادی برای تصاویر حرارتی ناحیه سینه اعمال می‌شود. نتایج آزمایشات نشان می‌دهد که الگوریتم پیشنهادی نسبت به الگوریتم سالپ دارای دقت بیشتری است. علاوه بر این، الگوریتم‌های پیشنهادی برای ناحیه‌بندی تصاویر مختلف موارد سالم و ناسالم در مقایسه با روش‌های پیشرفته، دقت بیشتری دارد.

در پژوهش [۲۷]، در سال ۲۰۲۰، مروری بر پیشرفت‌های اخیر در روش‌های تصویربرداری حرارتی و یادگیری عمیق برای تشخیص سرطان سینه شده است. با توجه به نتایج این مطالعه می‌توان گفت، توزیع حرارتی نامتقارن در ترموگرام پستان را می‌توان با استفاده از فناوری به کمک رایانه ارزیابی کرد. استفاده از این فناوری می‌تواند خطاها را به حداقل برساند. بررسی آنها نشان داده است که مدل‌های شبکه عصبی فعلی منجر به افزایش دقت طبقه‌بندی ترموگرام سرطان سینه، به ویژه در تشخیص موارد سالم و سرطانی می‌شود.

در پژوهش [۲۸]، در سال ۲۰۲۰، یک الگوریتم AdaBoost کارآمد به کمک یادگیری عمیق برای تشخیص سرطان پستان و تشخیص زودهنگام را پیشنهاد دادند. در این مطالعه، طبقه‌بندی و تخمین خطا که در یک لایه کاملاً متصل و یک لایه softmax گنجانده شده است. این پژوهش بر ترکیب این رویکردهای یادگیری ماشین با روش‌های انتخاب ویژگی‌ها و استخراج آنها از طریق ارزیابی خروجی آنها با استفاده از تکنیک‌های طبقه‌بندی و بخش‌بندی برای یافتن مناسب‌ترین رویکرد تمرکز دارد. نتایج تجربی نشان می‌دهد که این روش دارای سطح دقت بالای ۹۷.۲ درصد، حساسیت ۹۸.۳ درصد و ویژگی ۹۶.۵ درصد است.

در پژوهش [۲۹]، در سال ۲۰۲۱، یک مدل جدید یادگیری ماشین ترکیبی K-Means و مدل گوسین ترکیبی<sup>۲۱</sup> برای تشخیص سرطان سینه ارائه دادند. در رویکرد ترکیبی پیشنهادی، ویژگی‌های برجسته‌گذاری شده هر دو k-means و GMM به طور موثر برای تقسیم ناحیه یا نقاط دانه به موارد فرعی مختلف استفاده می‌شوند. ارزیابی‌ها نشان می‌دهد که رویکرد ترکیبی دارای دقت ۹۵.۵٪، نرخ خطا ۱۸.۶۴٪ و نرخ سیگنال به نویز ۱۳.۰۵ است و در مقایسه با سایر تکنیک‌های موجود دقت بیشتری دارد. روش پیشنهادی آنها نرخ پیش‌بینی بالایی را برای تکنیک تقسیم‌بندی ترکیبی مورد استفاده در تشخیص سرطان سینه نشان می‌دهد. در پژوهش [۳۰]، در سال ۲۰۲۱، مقایسه شبکه‌های عصبی کانولوشن عمیق برای ناحیه‌بندی توده‌های پستان در ماموگرافی را بررسی نمودند. در این مطالعه معماریهای یادگیری عمیق نظیر UNet، Visual Geometry Group 19 (VGG19)، linkNet-VGG19، MultiResUNet، ResNeXt، ResNet، MobileNetv2، EfficientNet، InceptionResNetV2، SeResNeXt و SEResNet برای تشخیص سرطان سینه با هم مقایسه شده است. ارزیابی‌ها نشان داد مدل‌های پیشنهادی به مقادیر Dice بالای ۹۰٪ دست یافت و بهترین معماری برای تشخیص سرطان سینه، معماری EfficientNet که دقت ۹۴.۷۵٪ و ۹۹٪ را در دو مجموعه داده مختلف به دست آورد.

در پژوهش [۳۱]، در سال ۲۰۲۲، روشهای بهینه تشخیص زودرس سرطان پستان با ناحیه‌بندی محاسباتی را مرور نمودند. مطالعات آنها نشان می‌دهد سرطان سینه شایع‌ترین سرطان در زنان ۵۹ تا ۶۹ ساله است. مطالعات نشان داده‌اند که تشخیص زودهنگام و درمان سرطان سینه شانس بقا را به میزان قابل توجهی افزایش می‌دهد. آنها همچنین نشان دادند که تشخیص زودهنگام ضایعات کوچک، پیش‌بینی را بهبود می‌بخشد و منجر به کاهش قابل توجهی در موارد مرگ می‌شود. این مطالعه بر

<sup>19</sup> Salp Swarm Algorithm

<sup>20</sup> Chaotic maps

<sup>21</sup> Gaussian mixture model (GMM)

روی تشخیص سرطان سینه و همچنین شناسایی عوامل خطر و ارزیابی آنها از سرطان پستان و همچنین تشخیص زودهنگام سرطان پستان با تجزیه و تحلیل روش‌های ماموگرافی MRI سه بعدی و تقسیم‌بندی تصاویر ماموگرافی با استفاده از یادگیری ماشین تمرکز دارد.

در پژوهش [۳۲]، در سال ۲۰۲۲، یک روش ادغام داده‌های بالینی و تصویری برای تشخیص شدت سرطان پستان با یک رویکرد سلسله‌مراتبی ارائه شده است. مزیت مطالعه تحقیقاتی آنها استفاده از هر دو ویژگی بالینی و تصویری و استخراج ویژگی‌های جدید با استفاده از تجزیه و تحلیل مؤلفه اصلی برای مرحله طبقه‌بندی است. در حالی که یک بیمار ممکن است به بیش از یک خوشه تعلق داشته باشد، نتایج تمام روش‌های طبقه‌بندی مربوطه برای بیمار مربوطه به طور مناسب ترکیب می‌شوند تا به مرحله بیمار سرطانی ختم شود. در نهایت، برای بررسی کارایی رویکرد ترکیبی پیشنهادی، از هفت مجموعه داده واقعی با داده‌های بالینی و تصویری استفاده می‌شود.

در پژوهش [۳۳]، در سال ۲۰۲۲، یک روش تشخیص سرطان سینه بر اساس تصاویر اولتراسوند و شبکه عصبی کانولوشن ارائه دادند. ارزیابی‌ها نشان داد روش Efficient-Det پیشرفت قابل توجهی در شناسایی مناطق تومور نشان داده است. در مقایسه با روش دستی (متوسط دقت: ۹۵.۳٪ و ۶۰ ثانیه در هر تصویر) و روش مهندسی ویژگی سنتی (دقت متوسط: ۹۰٪ و ۱۵ اینچ در هر تصویر)، Efficient-Det آنها دارای دقت و سرعت بیشتری است.

### ۳. روش پیشنهادی

تشخیص بیماران بدخیم و خوش‌خیم در بیماری سرطان سینه یک مسئله از نوع طبقه‌بندی است لذا برای حل آن می‌توان از روش‌های یادگیری ماشین استفاده نمود. یکی از روش‌های کاربردی برای حل این مسئله بکارگیری روش ماشین بردار پشتیبان است. روش ماشین بردار پشتیبان برای طبقه‌بندی دقیق نمونه‌های بدخیم و خوش‌خیم نیاز به بهینه‌سازی پارامترهای مانند ضریب جریمه و خطا دارد. در این بخش یک روش پیشنهادی برای بهینه‌سازی ماشین بردار پشتیبان با استفاده از الگوریتم بهینه‌سازی شاهین انجام شده است که دارای مزایای ذیل است:

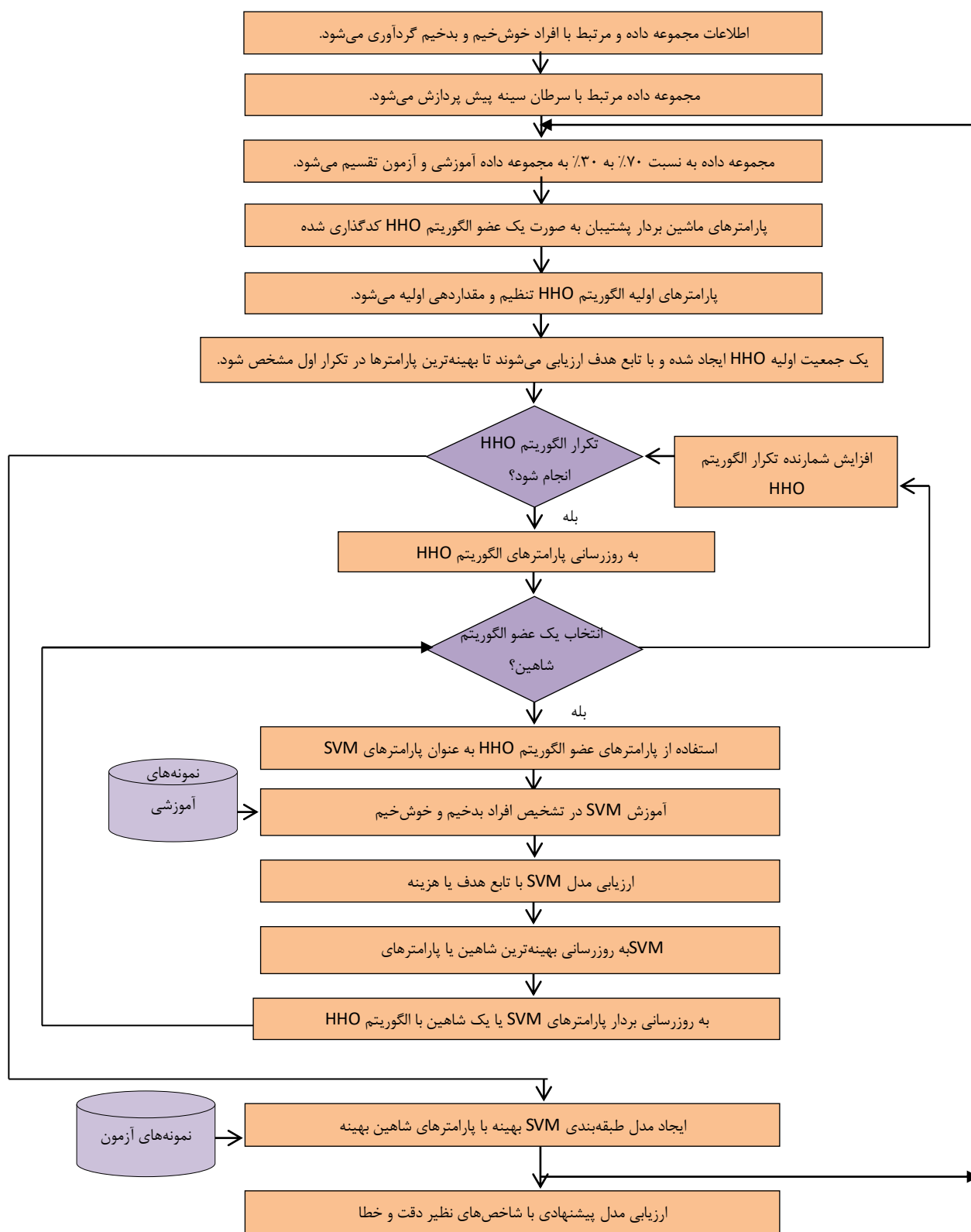
- الگوریتم بهینه‌سازی شاهین دارای انواع استراتژی از جمله شیرجه و محاصره است.
  - الگوریتم بهینه‌سازی شاهین دارای مکانیزم جستجوی محلی، سراسری و تصادفی است.
  - مدل‌سازی الگوریتم بهینه‌سازی شاهین بسیار دقیق است.
  - الگوریتم بهینه‌سازی شاهین در یک ژورنال معتبر چاپ شده و تاکنون در بسیاری از کاربردها استفاده شده است.
- در این بخش روش پیشنهادی برای تشخیص افراد بدخیم و خوش‌خیم فرموله و ارائه می‌شود.

### ۱.۳. چارچوب روش پیشنهادی

چارچوب پیشنهادی برای بهبود ماشین بردار پشتیبان با استفاده از الگوریتم بهینه‌سازی شاهین برای تشخیص سرطان سینه در شکل (۹)، نمایش داده شده است که دارای مراحل ذیل است:

- داده‌های مرتبط با بیماری سرطان سینه گردآوری شده
- داده‌های گردآوری شده مورد پیش پردازش و نرمال‌سازی قرار گرفته می‌شود و به نسبت ۷۰٪ به ۳۰٪ به داده‌های آموزشی و آزمون تقسیم می‌شوند.
- یک عضو الگوریتم شاهین به عنوان یک آرایه که پارامترهای SVM را در خود نگهداری می‌کند، کدگذاری شده
- تعدادی از آرایه‌های که پارامترهای SVM را دارند به عنوان اعضای جمعیت اولیه الگوریتم شاهین تصادفی تولید می‌شوند.
- تعدادی از آرایه‌های که پارامترهای SVM را دارند به عنوان اعضای جمعیت اولیه الگوریتم شاهین تصادفی تولید میشوند و با تابع هدف ارزیابی شده و بهینه‌ترین شاهین در تکرار اول انتخاب می‌شود.

- شمارنده تکرار الگوریتم شاهین در ابتدا روی یک تنظیم شده و پارامترهای الگوریتم شاهین مقداردهی اولیه می‌شوند.
- یک عضو از الگوریتم شاهین انتخاب شده و پارامترهای آن برای SVM استفاده شده و مدل مورد نظر با داده‌های آموزشی، آموزش داده می‌شود.
- بهینه‌ترین عضو الگوریتم شاهین در این مرحله به روزرسانی می‌شود.
- با استفاده از معادلات الگوریتم شاهین، شاهین مورد نظر به روزرسانی شده
- سایر اعضای الگوریتم شاهین نیز با معادلات الگوریتم شاهین به روزرسانی شده و در پایان این مرحله بهینه‌ترین عضو جمعیت نیز به روزرسانی می‌شود.
- در ادامه یک شمارنده به تکرار الگوریتم شاهین اضافه شده و اگر این شمارنده از مقدار بیشینه کمتر یا مساوی باشد مراحل تکرار می‌شود و در غیر اینصورت بهینه‌ترین بردار پارامتریک SVM یا یک عضو الگوریتم شاهین برای ایجاد مدل طبقه‌بندی SVM استفاده می‌شود.
- در نهایت با استفاده از داده‌های آزمون می‌توان مدل مورد نظر را ارزیابی نمود و بر حسب مقدار دقت و خطای طبقه‌بندی آن را مورد تحلیل و ارزیابی قرار داد.



شکل ۹: چارچوب روش پیشنهادی برای تشخیص سرطان سینه

## ۲.۳. پیش پردازش مجموعه داده

در بیشتر موارد مجموعه داده‌ها دارای ویژگی‌های متنوعی می‌باشند و در ویژگی دارای محدوده پایین و بالای خاص خود است. اگر محدود پایین و بالای ویژگی‌های یک مجموعه داده دارای اختلاف زیادی باشد این موضوع می‌تواند در یادگیری

ماشین اختلال ایجاد نماید. برای رفع این مشکل و قرار داده محدوده همه ویژگی‌ها در بازه  $[a, b]$  از نرمالسازی استفاده می‌شود. فرض کنید مجموعه داده سرطان سینه در بازه  $[a, b]$  مورد نرمالسازی قرار گرفته شود. برای نرمالیزه نمودن یک ویژگی مجموعه داده سرطان سینه با مقدار  $f$  از رابطه (۱)، استفاده می‌شود و  $f'$  مقدار نرمال شده ویژگی  $f$  است.

$$f' = \frac{f - \min}{\max - \min} (b - a) + a \quad (1)$$

در این معادله،  $\max \cdot \min$  به ترتیب کمینه و بیشینه یک ویژگی نظیر  $f$  در ارتباط با سرطان سینه است. می‌توان بازه نرمالسازی برای سرطان سینه را بین صفر و یک در نظر گرفت و نرمالسازی مطابق رابطه (۲)، ساده‌تر می‌شود:

$$f' = \frac{f - \min}{\max - \min} \quad (2)$$

### ۳.۳. کدینگ مسئله

فرض کنید تعدادی از بردارهای ویژگی یا الگوهای آموزشی به صورت  $\{x_1, x_2, \dots, x_i, \dots, x_N\}$  وجود دارد که هر کدام یک بردار ویژگی  $d$  بعدی بوده و دارای برچسب  $y_i$  است و  $y_i \in \{-1, +1\}$ .

هدف حل یک مسئله دسته بندی دو کلاسه به صورت بهینه است. رویکرد ماشین بردار پشتیبان بدین صورت است که در مرحله آموزش، سعی دارد مرز تصمیم‌گیری را به گونه ای انتخاب نماید که حداقل فاصله آن با هر یک از دسته های مورد نظر را بیشینه نماید. این نوع انتخاب باعث می‌شود که تصمیم‌گیری در عمل، شرایط نویزی را به خوبی تحمل نموده و همچنین پاسخ دهی مناسبی داشته باشد. این نحوه انتخاب مرز بر اساس نقاطی به نام بردارهای پشتیبان انجام می‌شود. مبنای کاری ماشین بردار پشتیبان دست بندی نمونه‌ها است، به گونه ای که یک صفحه فرعی را به منظور جداسازی نمونه‌های متفاوت از یکدیگر در نظر می‌گیرد که حداکثر فاصله را از دسته داشته باشد. در واقع، ماشین بردار پشتیبان با در نظر گرفتن نمونه‌هایی که در حاشیه دسته ها قرار دارند، به ایجاد بردارهای پشتیبان برای تمام دسته‌ها پرداخته و به بقیه نمونه‌ها توجه‌ای ندارد. فرض کنید که یک مجموعه داده در یک فضای  $X$  وجود دارد که  $x \in X$  و یک نقطه از آن در نظر گرفته می‌شود به نحوی که با تغییراتی در فضای  $Z \in Z$  انتقال یابد. یک نقطه از  $x \in X$  با نگاشتی قابل تبدیل به فضای  $z \in Z$  است که این نگاشت را با  $\phi(\cdot)$  نشان می‌دهند و به صورت  $\phi: X \rightarrow Z$  در نظر گرفته می‌شود. به عبارت دیگر یک  $z$  ای وجود دارد که برابر  $\phi(x)$  است پس به جای اینکه در فضای  $X$  ماشین بردار پشتیبان فرموله‌بندی انجام شود در فضای  $Z$  ماشین بردار پشتیبان فرموله‌بندی می‌شود. در ماشین بردار پشتیبان، فکیک کننده به صورت رابطه (۳)، فرموله می‌شود:

$$w^T x + b = 0 \Rightarrow w^T z + b = 0 \Rightarrow w^T \phi(x) + b = 0 \quad (3)$$

در این رابطه،  $W$  شیب خط تفکیک کننده و  $b$  عرض از مبدا خط تفکیک کننده است و مقدار  $w$  و  $b$  به صورت رابطه (۴) و (۵) تعیین می‌شود:

$$w = \sum \alpha_i y_i x_i \quad (4)$$

$$b = \frac{1}{|S|} \sum_i \left( y_i - \sum_j \alpha_j y_j k(x_i, x_j) \right) \quad (5)$$

در این روابط،  $k$  می‌تواند یک تابع کرنل باشد که حالت طبقه‌بندی خطی را به غیرخطی تبدیل می‌کند. تابع هدف نهایی برای ماشین بردار پشتیبان را می‌توان مانند رابطه (۶)، فرموله نمود.

$$Max = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (6)$$

در این تابع هدف که از نوع بیشینه‌یابی است شرط مسئله در رابطه (۷)، ذکر شده است و نوع کرنل هم در رابطه (۸)، فرموله شده است.

$$Max = \sum_{i=1}^n \alpha_i y_i, 0 \leq \alpha_i \leq C \quad (7)$$

$$K(x, y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right) \quad (8)$$

عملکرد ماشین بردار پشتیبان تا حد زیادی به دو پارامتر  $C$  و  $\sigma$  بستگی دارد و اگر این دو پارامتر بهینه انتخاب شوند آنگاه خطای SVM کاهش خواهد یافت. در روش پیشنهادی می توان هر عضو الگوریتم شاهین نظیر عضو  $i$  ام را به صورت رابطه (۹)، فرموله نمود:

$$X_i = \langle\langle C_i, \sigma_i \rangle\rangle \quad (9)$$

هدف آن است که یک  $X_i$  به گونه ای انتخاب شود که متوسط خطای طبقه بندی مانند رابطه (۱۰)، را کمینه نماید.

$$Error(X_i) = \frac{\sum_{i=1}^n (o_i^{real} - o_i^{est})^2}{n} \quad (10)$$

در اینجا  $o_i^{real}$  شماره کلاس واقعی از نظر بدخیم یا خوش خیم برای نمونه  $i$  و  $o_i^{est}$  مقدار پیش بینی آن نمونه است. پارامتر  $n$  تعداد نمونه ها برای ارزیابی است و  $Error(X_i)$  خطای طبقه بندی نمونه های خوش خیم و بدخیم است. یک تابع هدف دیگر می تواند از نوع دقت طبقه بندی باشد که بر خلاف خطا، بیشینه آن مد نظر است.

### ۴.۳. بهینه سازی با الگوریتم شاهین

در الگوریتم های بهینه سازی شاهین، هر شاهین یا پارامترهای SVM می تواند با توجه به مرکز تجمع سایر شاهین ها تصمیم گیری نماید و نقطه ثقل آنها را تعیین نماید. با تکرار متوالی الگوریتم بهینه سازی شاهین موقعیت شاهین ها و وش یا جواب بهینه فعلی مرتباً به روز شده و در نهایت و در تکرار آخر موقعیت طعمه به عنوان جواب بهینه استخراج می شود. در اینجا مجموعه ای از رفتار وجود دارد که شاهین ها یا راه حل های مسئله انجام می دهند تا بتواند فضای مسئله را مورد جستجو قرار دهند و این رفتار در واقع جستجو پیرامون نقطه ثقل راه حل ها و نقطه بهینه و سایر فضاهای تصادفی است. الگوریتم بهینه سازی شاهین دارای رفتار پیچیده و دارای مدلسازی دقیق است به گونه ای که در این الگوریتم تعداد زیادی رابطه وجود دارد و فضای مسئله به کمک روابط و معادله های مختلف مورد جستجو قرار گرفته می شود و از این جهت می توان انتظار داشت که الگوریتم مورد نظر همزمان دارای جستجوی اکتشافی (سراسری) و بهره ورداری (محلی) است. در روش پیشنهادی هر بردار  $\langle\langle C_i, \sigma_i \rangle\rangle$  نظیر  $i$  در تکرار  $t$  با  $X_i(t)$  نمایش داده می شود. در روش پیشنهادی تعداد بردار  $\langle\langle C_i, \sigma_i \rangle\rangle$  یا جمعیت الگوریتم شاهین که  $N$  نمایش داده می شود در هر مرحله وجود دارد. هر بردار پارامترهای SVM یا  $\langle\langle C_i, \sigma_i \rangle\rangle$  که متوسط خطای تشخیص افراد بدخیم و خوش خیم سرطان سین را کمینه تر نماید به عنوان بردار بهینه  $\langle\langle C_i, \sigma_i \rangle\rangle$  در SVM در نظر گرفته می شود بهینه ترین پارامترهای SVM با  $X_{rabbit}(t)$  نمایش داده می شود. برای به روزرسانی بردارهای  $\langle\langle C_i, \sigma_i \rangle\rangle$  در بیشتر موارد نیاز است متوسط مقدار بردارهای  $\langle\langle C_i, \sigma_i \rangle\rangle$  نیز محاسبه شود که از رابطه (۱۱)، استفاده می شود:

$$X_M(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \quad (11)$$

در این معادله،  $(t)$  نقطه ثقل جمعیت شاهین ها یا بردارهای پارامتریک SVM است و  $N$  تعداد بردارهای پارامتریک  $\langle\langle C_i, \sigma_i \rangle\rangle$  است. برای به روزرسانی بردارهای پارامتریک  $\langle\langle C_i, \sigma_i \rangle\rangle$  با استفاده از رفتار تصادفی می توان از رابطه (۱۲)، در الگوریتم بهینه سازی شاهین استفاده نمود:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 \cdot X(t)| & rand \geq 0.5 \\ (X_{rabbit}(t) - X_M(t)) - r_3 (LB + r_4 (UB - LB)) & rand < 0.5 \end{cases} \quad (12)$$



در این معادله،  $X(t)$  موقعیت فعلی یک شاهین یا یک بردار پارامتریک  $\langle\langle C_i, \sigma_i \rangle\rangle$  در تکرار فعلی یا  $t$ ،  $X(t+1)$  موقعیت یک شاهین یا بردار پارامتریک  $\langle\langle C_i, \sigma_i \rangle\rangle$  در تکرار جدید،  $X_{rabbit}(t)$  موقعیت بهینه‌ترین بردار پارامتریک ماشین بردار پشتیبان است.  $X_{rabbit}(t)$  یک موقعیت تصادفی از بردارهای پارامتریک SVM در فضای مسئله،  $X_M(t)$  نقطه ثقل جمعیت شاهین‌ها،  $x_1, x_2, x_3$  و  $x_4$  اعدادی تصادفی در بازه صفر و یک،  $LB$  و  $UB$  به ترتیب محدوده پایین و بالای راه-حلها در فضای مسئله است. در الگوریتم بهینه‌سازی شاهین، ضریب انرژی میزان حمله یا نزدیک شدن شاهین‌ها به جواب بهینه را نشان می‌دهد و مانند رابطه (۱۳)، فرموله شده و بر حسب تکرار الگوریتم مرتباً در حال کاهش است.

$$E = 2E_0(1 - \frac{t}{T}) \quad (13)$$

در این معادله، مقدار پارامتر  $E$  مرتباً از دو به عدد صفر کاهش پیدا می‌نماید و در این معادله  $E_0$  مقدار انرژی اولیه،  $t$  شماره تکرار فعلی و  $T$  حداکثر تکرار الگوریتم شاهین می‌باشد. در الگوریتم‌های بهینه‌سازی شاهین، مکانیزم محاصره آرام یا نرم<sup>۲۲</sup> مطابق معادله (۱۴) و (۱۵)، بردارهای  $\langle\langle C_i, \sigma_i \rangle\rangle$  را به روزرسانی می‌نماید:

$$X(t+1) = \Delta X(t) - E|J \cdot X_{rabbit}(t) - X(t)| \quad (14)$$

$$\Delta X(t) = X_{rabbit}(t) - X(t) \quad (15)$$

در این معادلات،  $J$  ضریب فرار است و یک مقدار تصادفی بین صفر و دو است. در الگوریتم‌های بهینه‌سازی شاهین، هر راه‌حل یا بردار پارامتر SVM می‌تواند مستقیماً خود را تا نزدیکی بردار بهینه  $\langle\langle C_i, \sigma_i \rangle\rangle$  برسانده و به سمت آن شیرجه رود که مدلسازی این رفتار را می‌توان در رابطه (۱۶)، ارایه نمود:

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \quad (16)$$

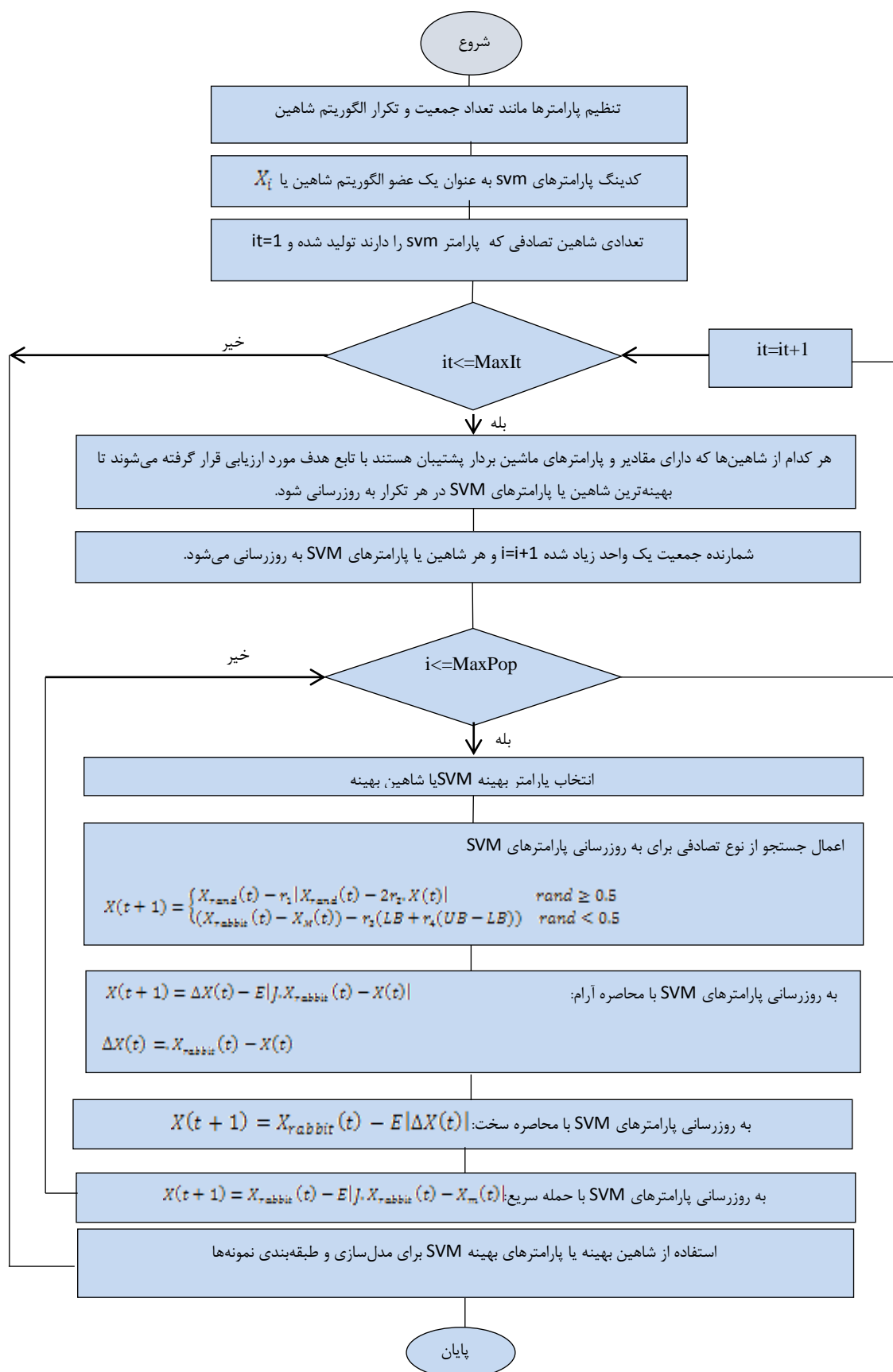
با رفتارهای متنوع الگوریتم شاهین می‌توان بهینه‌ترین بردار پارامترهای SVM را محاسبه نمود و از آن برای به روزرسانی پارامترهای ماشین بردار پشتیبان استفاده نمود تا در نهایت خطای طبقه‌بندی SVM را در تشخیص بیماران بدخیم و خوش-خیم کمینه نمود.

## ۹. فلوچارت روش پیشنهادی

فلوچارت روش پیشنهادی برای تشخیص افراد خوش‌خیم و بدخیم در ارتباط با سرطان سینه در شکل (۱۰)، نمایش داده شده است و با توجه به فلوچارت روش پیشنهادی دارای مراحل ذیل است:

- ایجاد یک جمعیت اولیه از پارامترهای SVM در قالب جمعیت الگوریتم شاهین
- ارزیابی جمعیت شاهین یا پارامترهای SVM با تابع هدف
- انتخاب بهینه‌ترین شاهین در هر تکرار که مقادیر بهینه پارامترهای SVM را دارد.
- تکرار مراحل الگوریتم شاهین مانند محاصره آرام، شیرجه سریع و رفتارهای تصادفی

<sup>22</sup>Soft besiege



شکل ۱۰: فلوچارت روش پیشنهادی برای تشخیص سرطان سینه

#### ۴. تجزیه و تحلیل

روش پیشنهادی برای طبقه‌بندی افراد بدخیم و خوش‌خیم از محیط نرم‌افزاری متلب R2018b استفاده می‌کند و برای ارزیابی روش پیشنهادی هر آزمایش به تعداد مشخص تکرار شده سپس توسط نتایج آن به عنوان خروجی نهایی در نظر گرفته می‌شود. در این بخش روش پیشنهادی روی شاخص‌های مختلف مورد تحلیل و ارزیابی قرار گرفته و نتایج با روشهای مشابه مقایسه می‌شود.

#### ۱.۴. مجموعه داده

برای انجام شبیه‌سازی و تجزیه و تحلیل روش پیشنهادی در تشخیص سرطان سینه از مجموعه داده‌های ویسکانسین [۳۴]، استفاده می‌شود. پایگاه داده دارای مجموعه داده‌های مانند WBC، WDBC و WPBC است که به ترتیب دارای ۹، ۳۱ و ۳۲ ویژگی است و از طرفی تعداد نمونه‌های آن به ترتیب برابر ۶۹۹، ۴۶۹ و ۱۹۵ است و هر نمونه می‌تواند در دو دسته بدخیم و خوش‌خیم قرار گرفته شود.

#### ۲.۴. معیارهای سنجش و ارزیابی

جهت ارزیابی روش پیشنهادی و مقایسه آن با روش‌های تشخیص بیماری معیار دقت، حساسیت، تشخیص و صحت بکار گرفته شده است. مقدار درصد دقت در بهترین و بدترین حالت به ترتیب ۰ و ۱۰۰ می‌باشد و نزدیکی به عدد ۱۰۰ نشان دهنده دقت مناسب و خوب الگوریتم پیشنهادی است. در رابطه (۱۷)، (۱۸)، (۱۹) و (۲۰) به ترتیب نحوه محاسبه دقت، حساسیت، تشخیص و صحت را نشان داده است:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (17)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (18)$$

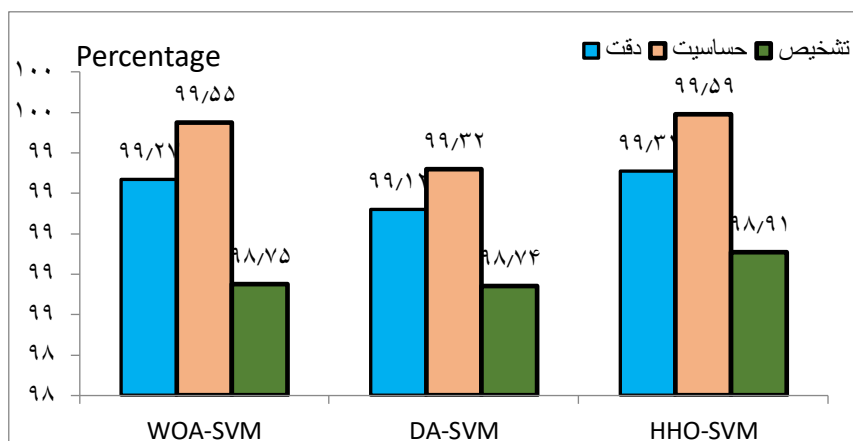
$$Specificity = \frac{TN}{(TN + FP)} \quad (19)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (20)$$

TP، افراد بیماری که به درستی بیمار تشخیص داده شده، TN، افراد سالم که به درستی سالم تشخیص داده شده، FP، افرادی که به غلط بیمار تشخیص داده شده و FN افرادی که به غلط سالم تشخیص داده شده است.

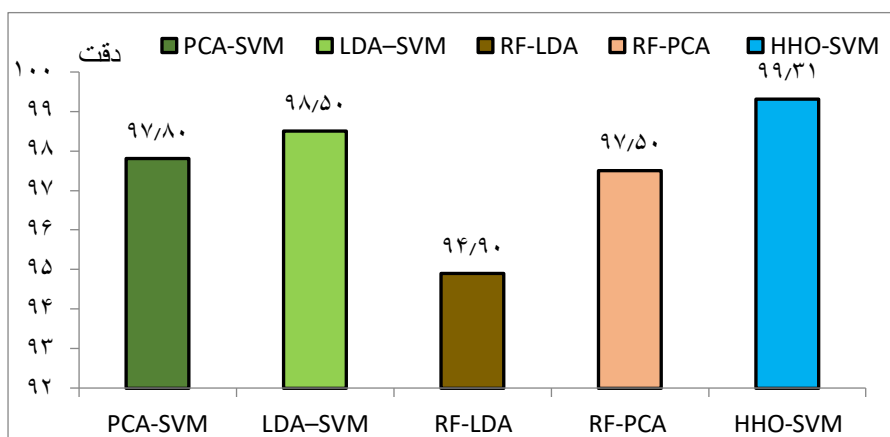
#### ۳.۴. نتایج و مقایسه ها

پیاده‌سازی روش پیشنهادی روی مجموعه داده WBCD نشان می‌دهد که دقت، حساسیت، تشخیص و صحت روش پیشنهادی برای تشخیص سرطان سینه به ترتیب برابر ۹۹.۳۱٪، ۹۹.۵۹٪، ۹۸.۹۱٪، ۹۸.۸۵٪ است. برای مقایسه می‌توان نتایج روش پیشنهادی را با مقاله [۳۵]، که در سال ۲۰۲۲ ارائه شده است روی این مجموعه داده انجام داد. در نمودار شکل (۱۱)، روش پیشنهادی با روش WOA-SVM و DA-SVM مورد مقایسه قرار گرفته شده است.

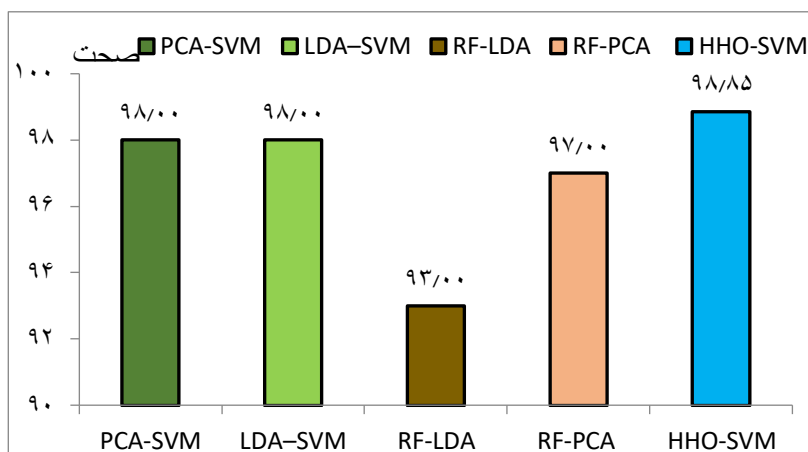


شکل ۱۱: مقایسه دقت، حساسیت، صحت روش پیشنهادی با روشهای مبتنی بر SVM و بهینه‌سازی پارامتر

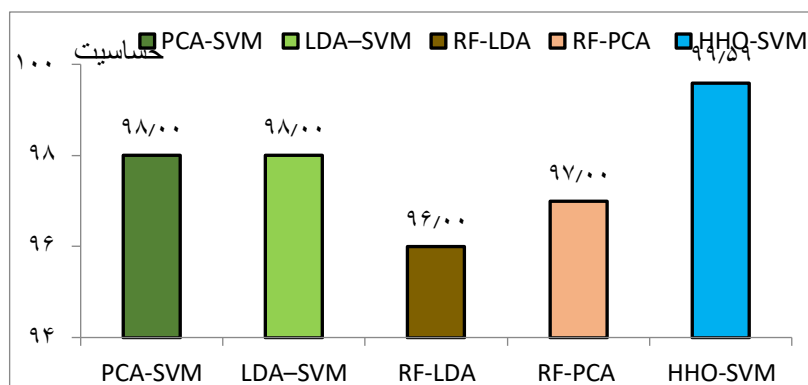
آزمایشات نشان می‌دهد دقت طبقه‌بندی روش WOA-SVM و DA-SVM به ترتیب برابر ۹۹.۲۷٪ و ۹۹.۱۲٪ است و این در حالی است که دقت روش پیشنهادی برابر ۹۹.۳۱٪ است. ارزیابی‌ها نشان می‌دهد حساسیت روش WOA-SVM و DA-SVM به ترتیب برابر ۹۹.۵۵٪، ۹۹.۳۲٪ و حساسیت روش پیشنهادی برابر ۹۹.۵۹٪ است و حساسیت روش پیشنهادی از دو روش WOA-SVM و DA-SVM بیشتر است. ارزیابی‌ها نشان داد تشخیص روش WOA-SVM و DA-SVM به ترتیب برابر ۹۸.۷۵٪ و ۹۸.۷۴٪ است و تشخیص روش پیشنهادی برای شناسایی سرطان سینه برابر ۹۸.۹۱٪ است. در پژوهش [۳۶]، در سال ۲۰۲۲ برای تشخیص سرطان سینه از یک ماشین بردار پشتیبان استفاده شده است و نتایج روش پیشنهادی نیز با آن مقایسه شده است. در نمودار شکل (۱۲)، (۱۳) و (۱۴) شاخص دقت، صحت و حساسیت روش پیشنهادی در مجموعه داده WBCD با روشهای PCA-SVM، LDA-SVM، RF-LDA و RF-PCA مورد مقایسه قرار گرفته شده است.



شکل ۱۲: مقایسه دقت روش پیشنهادی با روشهای مبتنی بر SVM و انتخاب ویژگی



شکل ۱۳: مقایسه صحت روش پیشنهادی با روشهای مبتنی بر SVM و انتخاب ویژگی



شکل ۱۴: مقایسه حساسیت روش پیشنهادی با روشهای مبتنی بر SVM و انتخاب ویژگی

ارزیابی‌ها نشان می‌دهد دقت روشهای PCA-SVM، LDA-SVM، RF-LDA و RF-PCA و روش پیشنهادی به ترتیب ۹۷.۸۰٪، ۹۸.۵۰٪، ۹۴.۹۰٪، ۹۷.۵۰٪ و ۹۹.۳۱٪ است. روش پیشنهادی دارای دقت بیشتری نسبت به روشهای PCA-SVM، LDA-SVM، RF-LDA و RF-PCA در تشخیص سرطان سینه است. ارزیابی‌ها نشان می‌دهد شاخص صحت در روش PCA-SVM، LDA-SVM، RF-LDA و RF-PCA و روش پیشنهادی به ترتیب برابر ۹۸٪، ۹۸٪، ۹۳٪، ۹۷٪ و ۹۸.۸۵٪ است و روش پیشنهادی دارای شاخص حساسیت بیشتری است. در شاخص صحت نیز مشاهده می‌شود صحت روش پیشنهادی از روشهای PCA-SVM، LDA-SVM، RF-LDA و RF-PCA در تشخیص سرطان سینه بیشتر است.

## ۵. نتیجه گیری و پیشنهادات آتی

مطالعات نشان می‌دهد سرطان سینه با تعداد زیادی مرگ جدید در سال ۲۰۲۰، پس از سرطان ریه، علت اصلی مرگ و میر زنان در جهان بوده است. سرطان سینه نوعی سرطان است که از بافت پستان منشأ می‌گیرد، به طور کلی از لایه داخلی مجرای شیر یا لوبول‌هایی که شیر را به مجرای شیر می‌رسانند. سلول‌های سرطانی از سلول‌های طبیعی به دلیل تغییر یا جهش اسید دئوکسی ریبونوکلیک (DNA) و اسید ریبونوکلیک (RNA) به وجود می‌آیند. این تغییرات یا جهش‌ها ممکن است به طور خود به خود در نتیجه افزایش آنتروپی رخ دهند یا ممکن است توسط عوامل دیگر ایجاد شوند. به عنوان مثال، تابش الکترومغناطیسی (اشعه ایکس، مایکروویو، اشعه ماوراء بنفش، اشعه گاما، و غیره)، تابش هسته‌ای، باکتری‌ها، ویروس‌ها، قارچ‌ها، انگل‌ها، مواد شیمیایی موجود در هوا، گرما، غذا، آب، رادیکال‌های آزاد، مکانیکی آسیب در سطح سلول، تکامل و پیری در

سرطان سینه نقش دارند. به طور کلی، خوش خیم و بدخیم دو دسته از تومورهای مرتبط با سرطان سینه هستند. اگرچه خوش خیم تهدید کننده زندگی و سرطانی نیست، اما ممکن است شانس ابتلا به سرطان سینه را افزایش دهد. در مقابل، تومورهای بدخیم هشدار دهنده تر و سرطانی تر هستند. نگرانی اصلی درمان سرطان سینه با پیش‌بینی دقیق محل سرطان و طبقه‌بندی نوع سرطان برای تعیین نحوه درمان سرطان آغاز می‌شود. با این حال، پیش‌بینی نوع سرطان سینه یکی از مشکلات کلاسیک در تحقیقات مرتبط با سلامت است. طبقه‌بندی دقیق سرطان سینه به تشخیص زودهنگام، تشخیص، درمان و در صورت امکان ریشه کن کردن کامل آن منجر می‌شود. علاوه بر این، طبقه‌بندی دقیق تومورهای خوش خیم می‌تواند از انجام درمان‌های غیر ضروری بیماران جلوگیری کند. در این مقاله برای طبقه‌بندی نمونه‌های خوش خیم و بدخیم از ماشین بردار پشتیبان بهبود یافته استفاده شده است. در روش پیشنهادی برای کاهش دادن خطای تشخیص بیماری سرطان سینه از الگوریتم بهینه‌سازی شاهین جهت بهینه‌سازی پارامترهای ماشین بردار پشتیبان استفاده شده است. آزمایشات نشان داد روش پیشنهادی نسبت به روش‌های بهینه‌سازی پارامتر ماشین بردار پشتیبان نظیر WOA-SVM و DA-SVM دقت بیشتری در تشخیص نمونه‌های بدخیم دارد. روش پیشنهادی همچنین از روش‌های انتخاب ویژگی در ترکیب با ماشین بردار پشتیبان در تشخیص سرطان سینه از جمله PCA-SVM، LDA-SVM، RF-LDA و RF-PCA دقیق‌تر است. مزیت مهم روش پیشنهادی دقت بالای آن در تشخیص سرطان سینه است و چالش آن زمان قابل توجه برای اجرای الگوریتم پیشنهادی است. در پژوهش آتی از یادگیری عمیق در ترکیب با روش‌های هوش گروهی و همچنین بکارگیری فاز انتخاب ویژگی برای تشخیص سرطان سینه استفاده می‌شود.

#### ۶. تشکر و قدردانی

با سپاس فراوان از استاد گرامی جناب آقای دکتر مهدی خزاعی پور برای اینکه دید جدیدی برای توسعه کارهای تحقیقاتی و پژوهشی در زمینه پزشکی و مطالعات سودمند در این زمینه برای اینجانب فراهم نمودند.

#### ۷. منابع و مراجع

1. Wang, X., Ahmad, I., Javeed, D., Zaidi, S. A., Alotaibi, F. M., Ghoneim, M. E., ... & Eldin, E. T. (2022). Intelligent Hybrid Deep Learning Model for Breast Cancer Detection. *Electronics*, 11(17), 2767.
2. Akbulut, S., Cicek, I. B., & Colak, C. (2022). Classification of Breast Cancer on the Strength of Potential Risk Factors with Boosting Models: A Public Health Informatics Application. *Medical Bulletin of Haseki/Haseki Tip Bulteni*, 60(3).
3. Roslidar, R., Rahman, A., Muharar, R., Syahputra, M. R., Arnia, F., Syukri, M., ... & Munadi, K. (2020). A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access*, 8, 116176-116194.
4. Mango, V. L., Olasehinde, O., Omisore, A. D., Wuraola, F. O., Famurewa, O. C., Sevilimedu, V., ... & Kingham, T. P. (2022). The iBreastExam versus clinical breast examination for breast evaluation in high risk and symptomatic Nigerian women: a prospective study. *The Lancet Global Health*, 10(4), e555-e563.
5. Amegbedzi, R. A., Komesuor, J., Amu, H., & Tarkang, E. E. (2022). Factors Influencing the Practice of Breast Self-Examination among Female Tertiary Students in Ho, Ghana. *Advances in Public Health*, 2022.
6. Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, 4(1), 49-56.



7. Yu, X., Zhou, Q., Wang, S., & Zhang, Y. D. (2022). A systematic survey of deep learning in breast cancer. *International Journal of Intelligent Systems*, 37(1), 152-216.
8. Bourouis, S., Band, S. S., Mosavi, A., Agrawal, S., & Hamdi, M. (2022). Meta-Heuristic Algorithm-Tuned Neural Network for Breast Cancer Diagnosis Using Ultrasound Images. *Frontiers in Oncology*, 12, 834028.
9. Alqudah, A., & Alqudah, A. M. (2022). Sliding window based support vector machine system for classification of breast cancer using histopathological microscopic images. *IETE Journal of Research*, 68(1), 59-67.
10. Abbas, S., Jalil, Z., Javed, A. R., Batool, I., Khan, M. Z., Noorwali, A., ... & Akbar, A. (2021). BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Computer Science*, 7, e390.
11. Huang, Z., & Chen, D. (2021). A breast cancer diagnosis method based on VIM feature selection and hierarchical clustering random forest algorithm. *IEEE Access*, 10, 3284-3293.
12. Gonçalves, C. B., Souza, J. R., & Fernandes, H. (2022). CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images. *Computers in Biology and Medicine*, 142, 105205.
13. Jijitha, S., & Amudha, T. (2021). Breast cancer prognosis using machine learning techniques and genetic algorithm: experiment on six different datasets. In *Evolutionary Computing and Mobile Sustainable Networks* (pp. 703-711). Springer, Singapore.
14. Pravesjit, S., Longpradit, P., Kantawong, K., Pengchata, R., & Oul, N. (2021, August). A Hybrid PSO with Rao Algorithm for Classification of Wisconsin Breast Cancer Dataset. In *2021 2nd International Conference on Big Data Analytics and Practices (IBDAP)* (pp. 68-71). IEEE.
15. Mazen, F., AbulSeoud, R. A., & Gody, A. M. (2016). Genetic algorithm and firefly algorithm in a hybrid approach for breast cancer diagnosis. *International Journal of Computer Trends and Technology*, 32(2), 62-68.
16. Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future generation computer systems*, 97, 849-872.
17. Roy, S. D., Das, S., Kar, D., Schwenker, F., & Sarkar, R. (2021). Computer aided breast cancer detection using ensembling of texture and statistical image features. *Sensors*, 21(11), 3628.
18. Masud, M., Eldin Rashed, A. E., & Hossain, M. S. (2020). Convolutional neural network-based models for diagnosis of breast cancer. *Neural Computing and Applications*, 1-12.
19. Ragab, D. A., Sharkas, M., Marshall, S., & Ren, J. (2019). Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, 7, e6201.
20. Mohammad, W. T., Teete, R., Al-Aaraj, H., Rubbai, Y. S. Y., & Arabyat, M. M. (2022). Diagnosis of Breast Cancer Pathology on the Wisconsin Dataset with the Help of Data Mining Classification and Clustering Techniques. *Applied Bionics and Biomechanics*, 2022.
21. Obaid, O. I., Mohammed, M. A., Ghani, M. K. A., Mostafa, A., & Taha, F. (2018). Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer. *International Journal of Engineering & Technology*, 7(4.36), 160-166.
22. Mridha, M. F., Hamid, M., Monowar, M. M., Keya, A. J., Ohi, A. Q., Islam, M., & Kim, J. M. (2021). A Comprehensive Survey on Deep-Learning-Based Breast Cancer Diagnosis. *Cancers*, 13(23), 6116.
23. Mobark, N., Hamad, S., & Rida, S. Z. (2022). CoroNet: Deep Neural Network-Based End-to-End Training for Breast Cancer Diagnosis. *Applied Sciences*, 12(14), 7080.

24. Zhang, H., Nguyen, H., Bui, X. N., Pradhan, B., Asteris, P. G., Costache, R., & Aryal, J. (2021). A generalized artificial intelligence model for estimating the friction angle of clays in evaluating slope stability using a deep neural network and Harris Hawks optimization algorithm. *Engineering with Computers*, 1-14.
25. Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. *Ieee Access*, 8, 203097-203116.
26. Ibrahim, A., Mohammed, S., Ali, H. A., & Hussein, S. E. (2020). Breast cancer segmentation from thermal images based on chaotic salp swarm algorithm. *IEEE Access*, 8, 122121-122134.
27. Roslidar, R., Rahman, A., Muharar, R., Syahputra, M. R., Arnia, F., Syukri, M., ... & Munadi, K. (2020). A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access*, 8, 116176-116194.
28. Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., & Fan, J. (2020). Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access*, 8, 96946-96954.
29. Jebarani, P. E., Umadevi, N., Dang, H., & Pomplun, M. (2021). A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection. *IEEE Access*, 9, 146153-146162.
30. Anaya-Isaza, A., Mera-Jiménez, L., Cabrera-Chavarro, J. M., Guachi-Guachi, L., Peluffo-Ordóñez, D., & Rios-Patiño, J. I. (2021). Comparison of Current Deep Convolutional Neural Networks for the Segmentation of Breast Masses in Mammograms. *IEEE Access*, 9, 152206-152225.
31. Saeed, S., Jhanjhi, N. Z., Naqvi, M., Humyun, M., Ahmad, M., & Gaur, L. (2022). Optimized Breast Cancer Premature Detection Method With Computational Segmentation: A Systematic Review Mapping. *Approaches and Applications of Deep Learning in Virtual Medical Care*, 24-51.
32. Rahimi Rise, Z., Mahootchi, M., & Ahmadi, A. (2022). Fusing clinical and image data for detecting the severity of breast cancer by a novel hierarchical approach. *Journal of Experimental & Theoretical Artificial Intelligence*, 1-24.
33. Du, R., Chen, Y., Li, T., Shi, L., Fei, Z., & Li, Y. (2022). Discrimination of Breast Cancer Based on Ultrasound Images and Convolutional Neural Network. *Journal of oncology*, 2022.
34. [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(Prognostic\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(Prognostic))
35. Elkorany, A. S., Marey, M., Almustafa, K. M., & Elsharkawy, Z. F. (2022). Breast Cancer Diagnosis Using Support Vector Machines Optimized by Whale Optimization and Dragonfly Algorithms. *IEEE Access*, 10, 69688-69699.
36. Egwom, O. J., Hassan, M., Tanimu, J. J., Hamada, M., & Ogar, O. M. (2022). An LDA-SVM Machine Learning Model for Breast Cancer Classification. *BioMedInformatics*, 2(3), 345-358.

## Breast cancer detection method using support vector machines improved by Harris Hawks optimization algorithm (HHO)

Maryam Soltan Mohammadi(\*), Mehdi khazaiepoor

1. Ph.D. student, Faculty of Technology and Engineering (Department of Computer Engineering), Birjand Branch, Islamic Azad University, Birjand, Iran.

[maryam\\_soltanmohammadi@yahoo.com](mailto:maryam_soltanmohammadi@yahoo.com)

2. Assistant Professor, Technical and Engineering Faculty (Department of Computer Engineering), Birjand Branch, Islamic Azad University, Birjand, Iran.

[mkhazaiepoor@gmail.com](mailto:mkhazaiepoor@gmail.com)

**Abstract—** Breast cancer is one of the deadly diseases among women and every year millions of people around the world die due to this disease. If breast cancer is detected in the early stages, the chances of survival will increase. One of the methods of breast cancer diagnosis is the use of knowledge discovery methods such as machine learning. Machine learning methods can discover the pattern of breast cancer by analyzing the information of patients and their records. The important advantage of using machine learning methods to diagnose breast cancer is to reduce diagnosis costs and help more accurate diagnosis by specialist doctors. One of the methods of breast cancer diagnosis is the use of a support vector machine. Support vector machine is a method for classifying samples with the aim of reducing operational risk in classification. One of the important challenges of the support vector machine is the output error of the model due to the lack of optimal selection of the learning parameters. In the proposed method to reduce the classification error of malignant and benign people, Harris Hawks's optimization algorithm has been used. The role of Harris Hawks's algorithm in the proposed method is to optimize the parameters of the support vector machine to reduce the diagnosis error of malignant patients. The evaluations have been done in the MATLAB programming environment and on the Wisconsin dataset. The evaluations show that the proposed method in breast cancer diagnosis has an accuracy of 99.31% and is more accurate than methods such as Wall's optimization algorithm in breast cancer diagnosis.

**Keywords:** Breast cancer, machine learning, support vector machine optimization, Harris Hawks optimization algorithm, optimization, prediction.