مقاله پژوهشی

فصلنامهٔ بیماری‌های پستان ایران، سال دوم، شماره دوم، تابستان ۱۳۸۸

معرفی یک سیستم هوشمند برای تشخیص دقیق سرطان پستان

*محمد علیپور: کارشناس مهندسی پزشکی و کارشناس ارشد الکترونیک، دانشکدهٔ فنی مهندسی دانشگاه تربیت معلم سبزوار
جواد حدادنیا: دانشیار گروه برق و کامپیوتر، دانشکدهٔ فنی مهندسی دانشگاه تربیت معلم سبزوار.

چکیده

مقدمه: تشخیص به‌موقع سرطان پستان به‌طور چشمگیری مرگ‌ومیر ناشی از آن را در جامعهٔ زنان کاهش می‌دهد. آزمایش آسپیراسیون سوزنی (FNA) روشی ساده، ارزان و غیرتهاجمی برای تشخیص دقیق و زودهنگام این سرطان است که امروزه تلاش می‌شود به‌صورت هوشمند و ماشینی انجام گیرد.

روش بررسی: مراحل ایجاد یک سیستم هوشمند برای تشخیص سرطان پستان عبارت‌اند از: ثبت تصاویر میکروسکوپیک از نمونهٔ FNA، استخراج ویژگی‌های عددی از این تصاویر، انتخاب ویژگی‌های تفکیک‌کننده و طراحی و آزمایش طبقه‌بندی‌کنندهٔ مناسب. در این تحقیق از ویژگی‌های آمادهٔ پایگاه داده WDBC که شامل ۵۶۹ نمونهٔ FNA می‌باشد، استفاده شد. برای انتخاب ویژگی روش جدیدی مبتنی بر الگوریتم بهینه‌سازی ذرات دودویی (BPSO) ارائه شد و سرانجام تلفیقی از طبقه‌بندی‌کننده‌های SVM برای کلاس‌بندی نمونه-ها به‌کار گرفته شد.

یافته‌ها: سیستم پیشنهادی با استفاده از ۲۸ ویژگی در قالب ۵ مدل SVM به دقت شناسایی ۱۰۰٪ دست یافت. این سیستم از لحاظ دقت و تعداد ویژگی‌های مورد نیاز بر سیستم‌های موجود برتری دارد.

نتیجه‌گیری: این تحقیق با ارائهٔ یک الگوریتم انتخاب ویژگی کارآمد موفق شده است دقت شناسایی سیستم‌های تشخیص سرطان پستان را بهبود دهد. این در حالی است که نسبت به سیستم‌های مشابه از تعداد کمتری ویژگی استفاده شده است. از دیگر مزیت‌های انتخاب ویژگی این است که علاوه بر تشخیص کلی، تشخیص ناهنجاری‌های ناشی از بیماری را نیز ممکن می‌سازد.

واژه‌های کلیدی: تشخیص سرطان پستان، نمونه‌برداری با سوزن ظریف (FNA)، انتخاب ویژگی، بهینه‌سازی جمعی ذرات دودویی (BPSO)، ماشین بُردار پشتیبان (SVM).

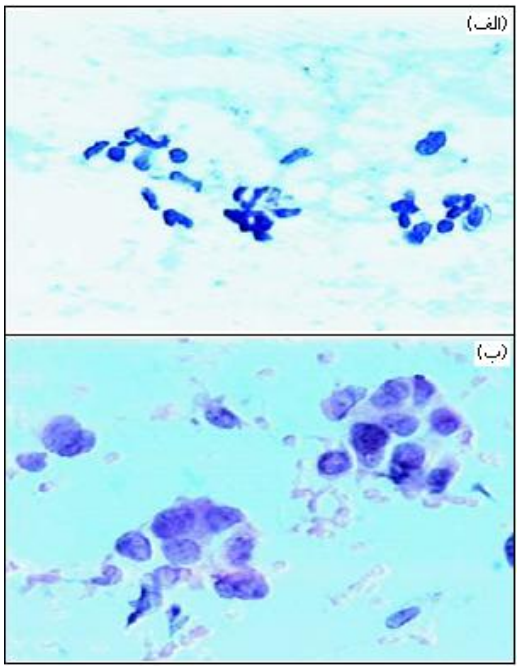
* نشانی نویسندهٔ پاسخگو: سبزوار دانشگاه تربیت معلم سبزوار، دانشکده فنی مهندسی، گروه برق و کامپیوتر، محمد علیپور
تلفن: ۰۹۱۴۴۰۰۹۰۰۵ نشانی الکترونیک: mohammad.alipoor@ymail.com

[DOR: 20.1001.1.17359406.1388.2.2.4.9]

مقدمه

سرطان پستان نوعی سرطان با نرخ مرگ‌ومیر بالا در میان زنان است به‌طوری‌که شایع‌ترین دلیل مرگ در جامعهٔ زنان می‌باشد[۱]. تشخیص به‌موقع سرطان پستان (حداکثر ۵ سال پس از اولین تقسیم سلولی سرطانی) شانس زنده ماندن بیمار را از ۵۶٪ به بیش از ۸۶٪ افزایش می‌دهد[۲]. بنابراین وجود یک سیستم دقیق و مطمئن برای تشخیص به‌موقع خوش‌خیم یا بدخیم بودن تومور پستان ضروری به‌نظر می‌رسد[۳]. به‌طور مرسوم تشخیص این سرطان از طریق نمونه‌برداری به کمک جراحی[1] انجام می‌شود که در بین روش‌های موجود، این روش بالاترین دقت تشخیص را داراست، اما روشی تهاجمی، زمان‌بر و گران می‌باشد[۴]. آزمایش آسپیراسیون سوزنی[2] (FNA) روش دیگری است که معایب روش قبلی را ندارد و با کمک تکنیک‌های یادگیری ماشینی، می‌تواند سیستمی کارآمد را برای تشخیص سرطان پستان فراهم کند. آزمایش FNA شامل استخراج مایع از بافت پستان و معاینهٔ بصری این نمونه در زیر میکروسکوپ می‌باشد[۵] که تقریباً بدون عوارض جانبی و دارای دقت بالا و به صورت سرپایی قابل انجام است[۶].

شکل ۱ دو تصویر گرفته شده از آزمایش FNA پستان را نشان می‌دهد. در تشخیص هوشمند، نخست این تصویر به یک تصویر خاکستری تبدیل می‌شود (اطلاعات رنگی تصویر مورد نیاز نیست) و سپس یک برنامهٔ کامپیوتری با تکنیک‌های پردازش تصویر، مرز هستهٔ سلول‌ها را در این تصویر میکروسکوپیک مشخص می‌کند و ویژگی‌هایی همچون شعاع، بافت (واریانس سطح خاکستری پیکسل‌های داخل هسته)، محیط، مساحت، فشردگی (مربع محیط تقسیم بر مساحت)، همواری[3] (میانگین تفاضل طول خطوط شعاعی مجاور)، تقعر، تقارن و بُعد فراکتالی (بُعد فراکتالی مرز هسته، حاصل از تقریب خط ساحلی[4] [۷]) را برای هر هسته محاسبه می‌کند. سرانجام میانگین، خطای استاندارد و میانگین سه تا از بزرگ‌ترین مقادیر به‌دست آمده، محاسبه می‌شود و به این ترتیب ۳۰ ویژگی با مقدار حقیقی برای هر نمونه به‌دست می‌آید. پایگاه داده WDBC[5] یک مجموعهٔ داده شامل نتایج کمّی آزمایش FNA (تهیه شده به شرح فوق) است که برروی ۵۶۹ بیمار در بیمارستان Wisconsin انجام شده است[۸].



شکل ۱: تصاویر گرفته شده از نمونهٔ آزمایش FNA برای تومورهای خوش‌خیم (الف) و بدخیم (ب) سرطان پستان

تحقیقات زیادی در رابطه با تشخیص سرطان پستان به کمک تکنیک‌های یادگیری ماشینی صورت گرفته است که در این میان به مطالعات انجام شده روی پایگاه دادهٔ WDBC می‌پردازیم. به‌این‌ترتیب به‌آسانی می‌توان کارآمدی سیستم پیشنهادی را با سیستم‌های قبلی مشابه مقایسه کرد. گروهی از محققان با استفاده از روش انتخاب پروتوتایپ مبتنی بر بُردار ماشین پشتیبان[6] (SVM) برای قواعد نزدیک‌ترین همسایه به دقت ۹۵/۶۱٪ دست یافتند [۹]. بنا به گزارشی دیگر، به‌کارگیری طبقه‌بندی‌کنندهٔ[7] SVM بدون انتخاب ویژگی به دقت شناسایی ۹۷٪ منجر شده است[۵]. در یک کار جدیدتر محققان کاربرد یک تکنیک کاهش بُعد، موسوم به Isomap را به همراه طبقه‌بندی‌کنندهٔ مبتنی بر SVM آزمودند[۲]. آن‌ها دریافتند که با وجود کاهش چشمگیر ابعاد دادگان

[1] Surgical Biopsy
[2] Fine Needle Aspiration
[3] Smoothness
[4] Coastline Approximation
[5] Wisconsin Database for Breast Cancer
[6] Support Vector Machine
[7] Classifier

WDBC توسط Isomap، دقت شناسایی همچنان بالا باقی می‌ماند و از این نظر سیستم پیشنهادی آن‌ها بر PCA[8] برتری دارد. در این تحقیق برای ترکیب SVM و Isomap، دقت شناسایی ۹۷/۳٪ گزارش شد. شبکه‌های عصبی احتمالی[9] [۱۰و۱۱]، طبقه‌بندی‌کنندهٔ SVM با هسته[10] RBF[11] [۱۲] و طبقه‌بندی‌کنندهٔ فازی [۱۳] نیز برای حل مسئلهٔ سرطان پستان Wisconsin استفاده شده‌اند و در بهترین حالت دقت شناسایی ۹۹/۳۱٪ به‌دست آمده است.

مطالعات اخیر نشان می‌دهد طبقه‌بندی‌کننده‌های مبتنی بر هسته به سایر روش‌های موجود از قبیل طبقه‌بندی‌کننده‌های خطی، مبتنی بر موجک و شبکه‌های عصبی برتری دارد[۱۴]. از طرفی مقایسهٔ عملکرد هسته‌های مختلف SVM در رابطه با مسئلهٔ سرطان پستان نیز قبلاً با اجرای اعتبارسنجی عرضی ۱۰-دسته‌ای[12] برای طبقه‌بندی‌کنندهٔ SVM با هسته‌های خطی، چندجمله‌ای و RBF انجام شده است [۱۵]. به این ترتیب ۳۰ مدل SVM به‌دست آمده‌است که ۵ مدل برتر انتخاب و در یک سیستم واحد براساس قانون اکثریت برای تعیین خروجی نهایی ترکیب شده‌اند که دقت این روش ۹۹/۲۹٪ گزارش شده است. در این مطالعات [۱۵و۱۶] مشاهده شده است که هستهٔ چندجمله‌ای عملکرد بهتری دارد، لذا در این مقاله نیز این هسته ترجیح داده می‌شود.

یک سیستم شناسایی الگوی متداول شامل ۴ بخش است: استخراج ویژگی، انتخاب ویژگی، طراحی و آموزش طبقه‌بندی‌کننده و سرانجام آزمایش آن. در این مقاله از ویژگی‌های آمادهٔ پایگاه داده WDBC استفاده شده است. بسیاری از سیستم‌های یادگیری ماشینی به‌دلیل عدم توانایی در حذف ویژگی‌های نامطلوب و زاید دقت پایینی دارند،[۱۶] لذا در گام بعدی به‌منظور انتخاب بهترین ویژگی‌ها و دست‌یابی به بالاترین دقت، یک الگوریتم کارآمد و جدید برای انتخاب ویژگی پیشنهاد شده است. در این الگوریتم تکنیک بهینه‌سازی جمعی ذرات دودویی[13] (BPSO) برای انتخاب ویژگی و طبقه‌بندی‌کنندهٔ SVM برای کلاس‌بندی نمونه‌ها استفاده شده است.

روش بررسی

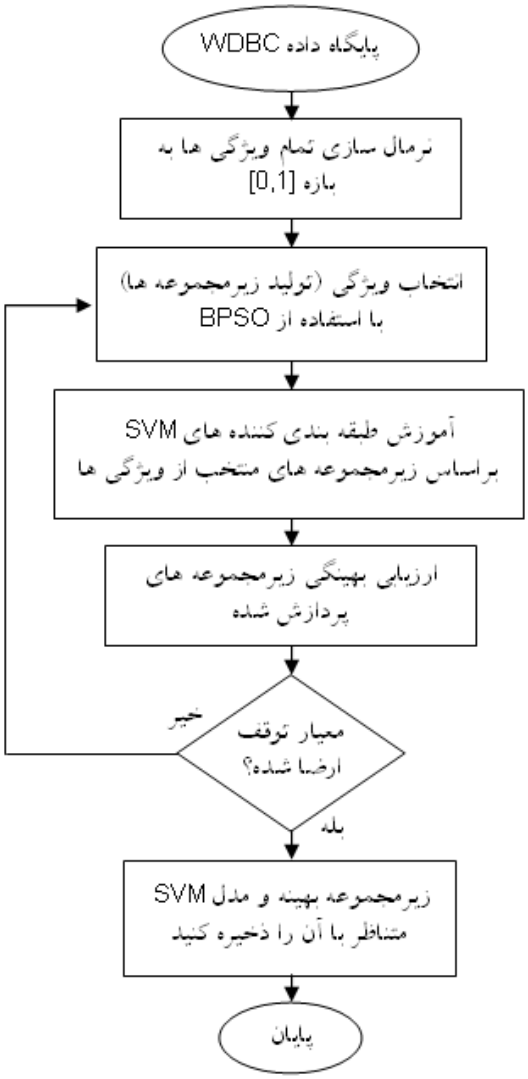
پس از مطالعهٔ مختصر روی سیستم‌های موجود، سیستم پیشنهادی به گونه‌ای که در شکل ۲ نمایان شده‌است، سازماندهی گردید. به‌عنوان یک حقیقت پذیرفته شده، نرمال‌سازی پایگاه داده عموماً به کسب نتایج بهتری منجر می‌شود. پس از نرمال‌سازی داده‌ها به بازه [۱،۰]، روش انتخاب ویژگی مبتنی بر BPSO پیاده شد. قبل از پرداختن به نتایج، بحث کامل راجع به اجزای سیستم پیشنهادی ضروری است. این بحث را در دو زیر بخش انجام می‌دهیم:

الف) انتخاب ویژگی

انتخاب ویژگی یک مسئلهٔ چالش برانگیز در بسیاری از حوزه‌ها از قبیل یادگیری ماشینی، شناسایی الگو و پردازش سیگنال است. مسئلهٔ انتخاب ویژگی در واقع برگزیدن ویژگی‌هایی است که حداکثر توان را در پیشگویی خروجی دارا باشند[۱۷]. اگر چند بُردار ویژگی n بُعدی را در نظر بگیریم، کار انتخاب ویژگی را می‌توان به‌صورت جستجو برای یک زیرمجموعهٔ بهینه از ویژگی‌ها در میان 2^n زیرمجموعهٔ ممکن توصیف کرد. تعریف این‌که زیرمجموعهٔ بهینه چه می‌تواند باشد، به مسئله‌ای که قصد داریم حل کنیم، وابسته است[۱۷]. الگوریتم‌های انتخاب ویژگی بسته به روند ارزیابی آن‌ها به دو دستهٔ عمده تقسیم می‌شوند. اگر انتخاب ویژگی مستقل از هرگونه الگوریتم یادگیری انجام شود (یعنی به‌صورت یک پیش پردازندهٔ کاملاً مجزا)، آن را روش فیلتر یا حلقهٔ باز گویند. در این مورد ویژگی‌های نامطلوب قبل از استنتاج دور ریخته می‌شوند. اما، اگر روند ارزیابی با یک الگوریتم طبقه‌بندی در ارتباط باشد، روش انتخاب ویژگی را Wrapper یا حلقهٔ بسته می‌نامند. این روش، جستجو در فضای زیرمجموعه‌ها را براساس تخمین دقت ناشی از انتخاب یک زیرمجموعهٔ خاص تحت شرایط الگوریتم طبقه‌بندی مورد استفاده (به عنوان معیاری از بهینگی آن زیرمجموعه) انجام می‌دهد[۱۸]. الگوریتم‌های دستهٔ دوم معمولاً نتایج بهتری به‌دست می‌دهند. مهم‌ترین بخش هر

[8] Principal Component Analysis
[9] Probabilistic Neural Networks
[10] Kernel
[11] Radial Basis Function
[12] 10-fold Cross-validation
[13] Binary Particle Swarm Optimization

روش انتخاب ویژگی حلقهٔ بسته، الگوریتم جستجویی است که در آن به کار رفته است. طی دههٔ گذشته محققان روی الگوریتم‌های جستجوی تکاملی مثل الگوریتم ژنتیک [۱۹و۲۰]، ACO[14] [۲۱،۱۸،۱۷] و ترکیب ACO/PSO [۲۲] تمرکز کرده‌اند.



شکل ۲: گردش کار سیستم پیشنهادی در یک نگاه

ب- بهینه‌سازی جمعی ذرات دودویی

مفهوم بهینه‌سازی جمعی ذرات[15] (PSO) با الهام از رفتار و حرکت دسته‌جمعی زنبورها، ماهی‌ها و پرندگان توسط کندی و ابرهارت در سال ۱۹۹۵ ارائه شد[۲۳]. در واقع، PSO یک روش قدرتمند و تصادفی[16] برای محاسبات تکاملی[17] بر مبنای جابه‌جایی هوشمند دستهٔ حیوانات در جستجوی غذا می‌باشد که روزبه‌روز رواج بیشتری پیدا می‌کند زیرا در مقایسه با الگوریتم ژنتیک به‌ویژه در مسائل شامل متغیرهای طراحی پیوسته، در یافتن جواب‌های بهینهٔ سراسری بازده بالاتری دارد[۲۴]. برخی از مزیت-های PSO شامل سهولت پیاده‌سازی و عدم نیاز به اطلاعات گرادیان می‌باشد[۲۵]. این الگوریتم از روش مرسوم محاسبات تکاملی پیروی می‌کند: الف: با جمعیتی تصادفی از جواب‌های ممکن آغاز می‌شود ب: با به‌روزرسانی نسل‌ها، جستجو برای جواب بهینه را انجام می‌دهد ج: ارزشیابی جمعیت براساس نسل‌های گذشته صورت می‌گیرد. در PSO جواب‌های احتمالی (ذرات) در فضای جواب مسئله به دنبال ذرات بهینهٔ فعلی جابه‌جا می‌شوند. این جابه‌جایی تحت تأثیر یک تابع برازندگی که کیفیت هر یک از ذرات را ارزیابی می‌کند، صورت می‌گیرد. فرض کنید فضای جستجو D-بُعدی باشد، در این صورت موقعیت i-امین ذره از دسته به شکل یک بردار D-بعُدی قابل نمایش است: $X_i = (x_{i1}, x_{i2} \ldots x_{iD})$. سرعت این ذره (تغییرات موقعیت) نیز با یک بُردار D-بُعدی دیگر قابل نمایش است: $V_i = (v_{i1}, v_{i2} \ldots v_{iD})$.

بهترین برازندگی به‌دست آمده برای i-امین ذره (بهینهٔ شخصی) و موقعیت متناظر با آن به ترتیب با $pbest_i$ و $xpbest_i$ نمایش داده می‌شوند. برازندگی و موقعیت بهترین ذره در دسته (بهینه سراسری) نیز باید به حافظهٔ الگوریتم سپرده شود (gbest, xgbest). ثابت شده است که استفاده از وزن لختی w عملکرد الگوریتم را بهتر می‌کند[۲۶]. در یک سیستم PSO با احتساب وزن لختی به‌روزرسانی ذرات با معادلات زیر انجام می‌شود[۲۷]:

$$v_{id}^{k+1} = wv_{id}^{k} + c_1 r_1 (xpbest_{id} - x_{id}) + c_2 r_2 (xgbest_{id} - x_{id}) \quad (1)$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \quad (2)$$

که در آن $d = 1, 2, \ldots, D$; $i = 1, 2, \ldots, N$ و N جمعیت دسته و c_1 و c_2 ضرایبی مثبت هستند. ضرایب r_1 و r_2 اعداد تصادفی هستند که به‌طور یکنواخت در بازه [۱،۰] توزیع شده‌اند و $k = 1, 2, \ldots$ تعداد تکرار را

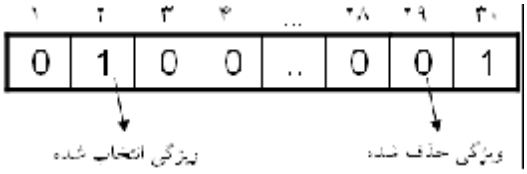
[14] Ant Colony Optimization
[15] Particle Swarm Optimization
[16] Stochastic
[17] Evolutionary Computations

مشخص می‌کند. روابط از نظر ابعادی صحیح هستند زیرا فرض می‌شود هر تکرار الگوریتم یک ثانیه طول می‌کشد($\Delta t = 1s$).

نسخهٔ باینری PSO توسط کندی و ابرهارت در ۱۹۹۷ ارائه شد [۲۸] و برخلاف PSO استاندارد، قادر به بهینه‌سازی در فضاهای گسسته می‌باشد. با در نظر گرفتن یک رشتهٔ دودویی برای هر ذره (شکل ۳)، BPSO شروع می‌شود. در این رشته، ۰ نمایانگر حذف ویژگی و ۱ نمایانگر انتخاب آن ویژگی می‌باشد. تنها تفاوتی که نسبت به PSO استاندارد وجود دارد، معادلهٔ به‌روزرسانی موقعیت ذرات است که به شکل زیر اصلاح می‌گردد:

$$S(v_{id}^{k+1}) = \frac{1}{1+\exp(-v_{id}^{k+1})} \qquad (3)$$

$$\text{If } S(v_{id}^{k+1}) > \text{rand} \text{ then } x_{id}^{k+1} = 1 \text{ else } x_{id}^{k+1} = 0 \qquad (4)$$



شکل ۳: یک رشته دو دویی برای هر ذره

که در آن *rand* یک عدد تصادفی است که به‌طور یکنواخت در بازه [۱،۰] توزیع شده‌است. مبتکران BPSO برای جلوگیری از اشباع تابع سیگموئید، محدود کردن سرعت در بازهٔ [۴،۴-] را توصیه می‌کنند[۲۹]. سیستم پیشنهادی با استفاده از BPSO، زیرمجموعه‌ای از ویژگی‌ها را که به بهترین نتیجه منجر می‌شود، برمی‌گزینند. دیگر مسئلهٔ چالش برانگیز، تشخیص بهینگی زیرمجموعه است. در روش‌های غیرفراگیر همواره باید تعادلی بین کوچک بودن یا مطلوب بودن زیرمجموعهٔ منتخب برقرار باشد[۱۸]. در این مسئله به‌نظر می‌رسد دقت شناسایی مهم‌تر از کوچک بودن زیرمجموعهٔ منتخب است، اگرچه بین دو مجموعه که به دقت یکسانی منجر می‌شوند، مجموعهٔ کوچک‌تر ترجیح داده می‌شود. بنابراین ما تابع برازندگی زیر را پیشنهاد می‌کنیم:

$$Fitness = \alpha.Accuracy + \beta.\frac{|n|-|S|}{|n|} \qquad (5)$$

که در آن $|n|$ تعداد کل ویژگی‌ها و $|S|$ تعداد ویژگی‌های منتخب است. جملهٔ اول ضریبی از دقت شناسایی[18] و جملهٔ دوم ضریبی از نرخ کاهش ویژگی‌هاست. مجموع ضرایب α و β را ثابت (و برابر ۱۰۰) در نظر می‌گیریم. حال با توجه به اهمیتی که برای دقت شناسایی و کمتر بودن تعداد ویژگی‌های مورد استفاده در تشخیص قائلیم، ضرایب α و β را تنظیم می‌کنیم. مسلماً در این مسئله دقت شناسایی خیلی مهم‌تر است و به تبع آن α بسیار بزرگ‌تر از β خواهد بود. پارامترهای BPSO مطابق جدول ۱ تنظیم شده‌اند. بدیهی است که BPSO با چند زیرمجموعهٔ تصادفی شروع به‌کار می‌کند و درواقع، سعی دارد تعدادی زیرمجموعه را که حاوی بیشترین و مفیدترین اطلاعات هستند، به ما معرفی کند. بُردارهای ویژگی کم‌بُعد (کمتر از ۳۰) معرفی شده توسط BPSO برای آموزش طبقه‌بندی‌کنندهٔ SVM به‌کار برده می‌شوند. هر زیرمجموعه از ویژگی‌ها براساس دو معیار ارزیابی می‌شود: دقت طبقه‌بندی‌کنندهٔ SVM متناظر و عدد اصلی بُردار ویژگی (تعداد ویژگی‌های منتخب). به این ترتیب BPSO جستجو میان 2^{30} زیرمجموعهٔ ممکن را برای یافتن جواب بهینه انجام می‌دهد. این در حالی است که رسیدن به حداکثر تعداد تکرار به عنوان شرط خاتمهٔ الگوریتم درنظرگرفته شده است.

جدول ۱: نحوهٔ تنظیم پارامترهای PSO

جمعیت ذرات	۲۰
C_1	۱
C_2	۲۰
ω (نزولی)	۱-۰
$V_{max}=-V_{min}$	۴
حداکثر تعداد تکرار	۶۰
α	۹۹
β	۱

یافته‌های تجربی

برای اثبات قابلیت‌های سیستم پیشنهادی، فرآیند اعتبارسنجی عرضی ۱۰-دسته‌ای بر روی پایگاه دادهٔ WDBC (شامل ۳۵۷ نمونهٔ خوش‌خیم و ۲۱۲ نمونهٔ بدخیم) انجام گرفت. ۵۶۹ نمونهٔ موجود به ۱۰ قسمت تقسیم شدند که هرقسمت دارای ۵۷ نمونه شامل ۳۶ نمونهٔ خوش‌خیم و ۲۱ نمونهٔ بدخیم است، به استثنای

[18] Accuracy

قسمت دهم که شامل ۳۳ نمونهٔ خوش‌خیم و ۲۳ نمونهٔ بدخیم می‌باشد. در اعتبارسنجی عرضی ۱۰-دسته‌ای ۹ قسمت از داده‌ها برای آموزش[19] و قسمت باقیمانده برای آزمایش به‌کار می‌روند. این فرآیند آموزش و آزمایش ۱۰ بار به‌صورت چرخشی انجام می‌شود به‌طوری‌که هر بار یک قسمت متفاوت برای آزمایش کنار گذاشته می‌شود. نتایج حاصل از این بخش در جدول ۲ آورده شده است. لازم به ذکر است که پیاده‌سازی تمام مراحل کار به کمک برنامه‌نویسی در محیط MATLAB7 انجام شده است. ۵ مدل برتر از نظر دقت شناسایی (مشخص شده در جدول ۲) انتخاب شدند و در یک سیستم یکپارچه قرار گرفتند و سپس کل ۵۶۹ نمونهٔ حاضر در پایگاه داده از این سیستم یکپارچهٔ عبور داده شدند و خروجی نهایی سیستم با قانون اکثریت[20] (رأی‌گیری) تعیین گردید. سیستم یکپارچهٔ پیشنهادی به‌دقت شناسایی ۱۰۰٪ دست یافت. نتایج به‌دست‌آمده در جدول ۳ گردآوری شده است. با دست‌یابی به چنین دقت شناسایی بالایی، نیازی به محاسبهٔ حساسیت، قطعیت و ویژگی آزمایش نیست، چون بدیهی است که همگی ۱۰۰٪ هستند.

جدول ۲: نتایج اعتبارسنجی عرضی ۱۰-دسته‌ای سیستم ترکیبی یکپارچه

شمارهٔ مدل	|s|	ویژگی‌های منتخب	دقت
۱	۷	۲،۶،۱۰،۱۹،۲۱،۲۵،۲۷	۹۶/۴۹
۲	۸	۱۱،۱۲،۱۵،۱۶،۱۷،۲۱،۲۵،۲۹	۹۲/۹۸
۳	۸	۵،۱۲،۱۶،۱۸،۲۰،۲۴،۲۷،۲۸	۹۴/۷۴
۴	۷	۵،۷،۹،۱۲،۱۴،۲۲،۳۰	۸۹/۴۷
۵	۸	۳،۵،۱۲،۱۷،۲۲،۲۵،۲۷	۹۸/۲۴
۶	۸	۳،۵،۱۲،۱۵،۱۸،۲۲،۲۶،۲۸	۹۴/۷۳
۷	۸	۳،۱۳،۱۴،۱۷،۲۲،۲۵،۲۷،۲۹	۸۹/۴۷
۸	۸	۵،۹،۱۲،۱۷،۱۹،۲۳،۲۵،۲۷	۹۴/۷۳
۹	۷	۲،۹،۱۴،۲۲،۲۴،۲۵،۳۰	۸۷/۷۱
۱۰	۷	۴،۱۰،۱۵،۱۸،۲۱،۲۲،۲۵	۹۴/۶۴
*		دقت میانگین	۹۳/۳۲

جدول ۳: نتایج ۵ مدل SVM برتر و سیستم ترکیبی یکپارچه

شمارهٔ مدل	دقت
۱	۹۹/۸۲
۳	۹۹/۶۴
۵	۹۹/۴۷
۶	۹۹/۴۷
۸	۹۹/۴۷
میانگین	۹۹/۵۷
سیستم یکپارچه	۱۰۰

انتخاب ویژگی مفهومی کلیدی در شناسایی الگو است که تا به امروز در تحقیقات مرتبط با سرطان پستان مورد توجه قرار نگرفته‌است. در مقایسه با تحقیقات انجام شده در این زمینه، افزودن مرحلهٔ انتخاب ویژگی و نیز ابداع روشی جدید برای انجام آن از نوآوری‌های این‌کار محسوب می‌شود. این مقاله با ارائهٔ یک الگوریتم انتخاب ویژگی کارآمد موفق شده است دقت شناسایی سیستم‌های تشخیص سرطان پستان را بهبود دهد. این در حالی است که نسبت به سیستم‌های مشابه از تعداد کمتری ویژگی استفاده شده است. از دیگر مزیت‌های انتخاب ویژگی این است که می‌توان با بهره‌گیری از آن به قدرت تفکیک هریک از ۳۰ ویژگی ذکر شده در مقدمه پی برد که از نظر پزشکی بسیار حائز اهمیت است، زیرا برخلاف هوش مصنوعی علاوه بر تشخیص کلی در پزشکی، تشخیص نابهنجاری‌های ناشی از بیماری نیز مورد توجه است که در این‌جا با استفاده از انتخاب ویژگی و در نتیجه استخراج اطلاعات بیشتر، این امر میسر شده است. البته در مسئلهٔ حل شده به دلیل عدم آگاهی از نحوهٔ چینش ویژگی‌ها در پایگاه داده WDBC اظهار نظر دقیق راجع به ناهنجاری‌های سلولی ناشی از سرطان پستان امکان‌پذیر نیست.

بحث و نتیجه گیری

این مقاله سیستم جدیدی برای حل مسئلهٔ سرطان پستان WDBC (به‌عنوان یک مسئلهٔ شناسایی الگوی تشخیصی[21]) ارائه می‌کند. سیستم پیشنهادی شامل

[19] Training
[20] Majority Rule
[21] Diagnostic Pattern Recognition

انتخاب ویژگی مبتنی بر BPSO و طبقه‌بندی‌کنندهٔ SVM، به دقت شناسایی ۱۰۰٪ منجر شد. لذا به وضوح بر کارهای قبلی انجام شده در این حوزه برتری دارد. همچنین بهره‌گیری از BPSO، خود ایدهٔ جدیدی به شمار می‌رود. مزیت بهره‌گیری از انتخاب ویژگی، دست‌یابی به شناسایی کامل با استفاده از کوچک‌ترین زیرمجموعه از ویژگی‌ها می‌باشد. سیستم پیشنهادی با استفاده از ۲۸ ویژگی (به جای ۳۰ ویژگی)، در قالب ۵ مدل SVM به شناسایی کامل دست یافت. چنین سیستم دقیقی امید برای عملی کردن تشخیص هوشمند سرطان پستان را بیشتر می‌کند. به‌علاوه استفاده از تعداد کمتری از ویژگی‌ها سرعت کارکرد سیستم‌های برخط را نیز افزایش می‌دهد.

به‌طورکلی، به‌کارگیری این روش انتخاب ویژگی در مسائل شناسایی الگوی پزشکی می‌تواند افق جدیدی در کشف عوامل و نشانه‌های بیماری‌ها ایجاد کند. با این روش قادر خواهیم بود علاوه بر دست‌یابی به دقت بالا در تشخیص بیماری، عوامل اصلی متمایزکنندهٔ نمونه‌های سالم و بیمار را نیز شناسایی کنیم.

منابع:

INNS-ENNS Int. Joint Conf. Neural Networks 2000; 3: 3655-8.

12. Mu T, Nandi A. Breast cancer detection from FNA using SVM with different parameter tuning systems and SOM-RBF classifier. J Franklin Institute 2007; 344: 285-311.

13. Fuentes-Uriarte J, Garcia M, Castillo O. Comparative study of fuzzy methods in breast cancer diagnosis. Annual Meeting of the NAFIPS 2008: 1-5.

14. Yang Lwei Y, Nishikwa R, Jiang Y. A study on several machine learning methods for classification of malignant and benign clustered micro calcification. IEEE Trans Med Imaging 2005; 24(3): 371-80.

15. Sewak M, Vaidya P, Chan C, Duan Z. SVM approach to breast cancer classification. Proc. 2nd Int. Multi-Symp. Computer and Computational Sciences 2007: 32-7.

16. Anagnostopoilosi I, Anagnostopoilosi C, Vergados D, Rouskas A, Kormentzas G. The Wisconsin Breast Cancer Problem: Diagnosis and TTR/DFS time prognosis using probabilistic and generalized regression information classifiers. J Oncol Reports 2006; 15: 975-81.

17. Jensen R. Combining rough and fuzzy sets for feature selection. Ph.D. Thesis, School of informatics, Univ. Edinburgh, 2005.

18. Kanan H, Faez K, Hosseinzadeh M. Face recognition system using ant colony optimization-based selected features. Proc. IEEE Symp. Computational Intelligence in Security and Defense Applications 2007: 57-62.

19. Siedlecki W, Sklansky J. A note on genetic algorithms for large-scale feature selection. J Pattern Recognition Letters 1989; 10(5): 335-47.

20. Bon A, Ogier J, Razali A, Yasin A. Feature selection in beltline Moulding process using

References:

1. Media Centre for World Health Organization. Fact sheet No. 297: Cancer, WHO: 2007.

2. Zhaohui L, Xiaoming W, Shengwen G, Binggang Y. Diagnosis of breast cancer tumor based on manifold learning and support vector machine. Proc. IEEE Int. Conf. Information and Automation 2008: 703-7.

۳. Litigate J. Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. J Clin Cancer Res 2004; 10: 2725-37.

4. Mangasarian O, Nick Street W, Wolberg W. Breast cancer diagnosis and prognosis via linear programming. J Operations Res 1995; 43: 570-7.

5. Maglogiannis I, Zafiropoulos E, Anagnostopoulos I. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. J Appl Intell 2007; 30(1): 24-36.

6. Khooei AR, Mehrabi Bahar M, Ghaemi M, Mirshahi M. Sensitivity and specificity of CNB in diagnosis of breast masses. Iranian journal of Basic Medical sciences 1384; 8(2): 6-100.

7. Mandelbrot B. The Fractal Geometry of Nature. Freeman and Company: New York, 1997.

8.http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

9. Li Y, Hu Z, Cai Y, Zhang W. Support vector based prototype selection method for nearest neighbor rules. In: Lecture Notes in Computer Science 3610. Springer: Berlin/Heidelberg 2005: 528-35.

10. Wang Y, Wan F. Breast cancer diagnosis via support vector machines. Proc. 25th Chinese Control Conf 2006: 1853-6.

11. Yang Z, Lu W, Yu D, Yu R. Detecting false benign in breast cancer diagnosis. Proc. IEEE-

[DOR: 20.1001.1.17359406.1388.2.2.4.9]

genetic algorithm. J Appl Sci Res 2008; 4(7): 783-92.
21. Liu B, Abbass H, McKay B. Classification rule discovery with ant colony optimization. IEEE Computational Intelligence Bulletin 2004; 3(1): 31-5.
22. Meng Y. A swarm intelligence based algorithm for proteomic pattern detection of ovarian cancer. IEEE Symp. Computational Intelligence and Bioinformatics and Computational Biology (CIBCB) 2006: 1-7.
23. Kennedy J, Eberhart R. Particle swarm optimization. Proc. IEEE Int. conf. Neural Networks 1995: 1942-8.
24. Nakamura M. A multi-objective particle swarm optimization incorporating design sensitivities. 11th AIAA/ISSMO multidisciplinary Analysis and Optimization Conf 2006: 70-5.
25. Chen Y. A local linear wavelet neural network. Proc. 5th world Cong. Intelligent Control and Automation 2004: 15-9.
26. Shi Y, Eberhart R. A modified particle swarm optimizer. Proc. IEEE Int. Conf. Evolutionary Computation 1998: 69-73.
27. Eberhart R, Shi Y. Comparing inertia weights and constriction factors in particle swarm optimization. Proc. IEEE Cong. Evolutionary Computations 2000: 84-8.
28. Kennedy J, Eberhart R. A discrete binary version of the particle swarm algorithm. Proc. Conf. Systems, Man, and Cybernetics 1997: 4104-9.
29. Franken N, Engelbrecht A. Investigating binary PSO parameter influence on the knights cover problem. IEEE Cong. Evolutionary Computations 2005; 1: 282-9.