

انتخاب ویژگی پایدار و دسته‌بندی براساس ساختار سلسله مراتبی مبتنی بر چند گونه سازی ازدحام ذرات آشوبی برای داده‌های ژنتیکی

مریم یاسی*^۱، محمدحسین معطر^۲، مهدی یعقوبی^۳

^۱دانشگاه آزاد اسلامی، واحد مشهد، باشگاه پژوهشگران جوان و نخبگان، مشهد، ایران

^۲دانشگاه آزاد اسلامی واحد مشهد، گروه نرم افزار، مشهد، ایران

^۳دانشگاه آزاد اسلامی واحد مشهد، گروه هوش مصنوعی، مشهد، ایران

چکیده:

هرگونه افزایش و تکثیر غیرطبیعی سلولی را تومور گویند. سرطان زمانی اتفاق می‌افتد که رشد کنترل نشده‌ای بر روی سلولهای غیرطبیعی در یکی از اعضای بدن به وجود آید. تومورها به دو دسته خوش خیم و بدخیم تقسیم می‌شوند. با توجه به رشد روزافزون داده‌های پیرامون بشر، به کارگیری ابزارهایی برای تحلیل این نوع داده‌ها و دست‌یابی به دانش نهفته آنها کاری الزامی است. به این دلیل که بشر دیدگاه ذاتی و بصری غیرقابل درکی درباره مسائل با ابعاد بالا و ساینز بزرگ پایگاه داده‌ها دارد، از این رو استفاده از روش‌های هوشمندانه تأثیر به‌سزایی در شناخت و درک بهتر بر روی داده‌های با ابعاد بالا می‌گذارد. در این مقاله ابتدا با ادغام روش‌های رتبه‌بندی، زیرمجموعه‌ای از ویژگی‌های پایدار و متمایزکننده انتخاب می‌گردد و در گام بعدی به کمک یک سیستم فازی به شناسایی و دسته‌بندی داده‌های بیولوژیکی از لحاظ خوش خیم و بدخیم بودن آنها می‌پردازیم. سیستم فازی از نوع *TSK (Takagi-sugeno-kang)* است. برای دسته‌بندی نمونه‌ها از یک ساختار سلسله مراتبی مبتنی بر چندگونه سازی ذرات بر پایه الگوریتم ازدحام ذرات آشوبی برای بهینه‌سازی سیستم فازی استفاده می‌کنیم. نشان داده می‌شود که استفاده از نظریه آشوب در مسائل بهینه‌سازی با ایجاد تنوع در میان ذرات موجب افزایش صحت شناسایی و دسته‌بندی نمونه‌ها می‌گردد. میزان صحت شناسایی و دسته‌بندی از نظر بدخیم و خوش خیم بودن بر روی داده‌های بیولوژیکی بیش از ۹۵ درصد بوده است. بررسی دقت روش پیشنهادی بر روی دو دادگان *uci-breast cancer* و *microarray* صورت گرفته است.

واژه‌های کلیدی:

انتخاب ویژگی، الگوریتم بهینه‌سازی ازدحام ذرات، تومور، سرطان، سیستم فازی، نظریه آشوب

۱- مقدمه

کشف دانش جدید در مورد وقایع و پدیده‌ها از مجموعه داده‌های با ابعاد بالا و دسته‌بندی و پیش‌بینی رویدادهای آینده به وسیله مدل‌های هوشمندانه هدف اصلی این مقاله است. بالا بودن ابعاد یک مجموعه داده باعث افزایش فضای جستجو و کاهش قدرت تعمیم‌پذیری و پیچیدگی محاسباتی می‌شود. در این شرایط با پدیده‌ای به نام نفرین ابعاد روبه‌رو می‌شویم. از جمله مجموعه داده‌هایی که دارای ابعاد بالا هستند می‌توان به مجموعه داده‌های ریز آرایه اشاره کرد. ریز آرایه‌ها محتوای داده‌های ژنتیکی و بیولوژیکی هستند. این داده‌ها در تشخیص مجموعه‌ای از سرطانها و تومورهای گوناگون مورد استفاده قرار می‌گیرند. سرطان عبارت است از ناهماهنگی میان رشد و مرگ میر سلولها که نتیجه آن انباشته شدن بیش از حد سلولهاست. به سلولهای انباشته شده غده یا تومور می‌گویند که این تومورها یا به صورت خوش خیم و یا بدخیم هستند.

هنگامی که ژن‌ها دارای عملکرد مناسبی باشند به خوبی توانایی کنترل رشد سلولها را دارند، اما زمانی که در عملکرد آنها نوعی ناهنجاری و یا بی‌قاعدگی بروز کند، تسلط خود را بر روی رشد سلولها از دست می‌دهند. درمان موفقیت‌آمیز سرطانها بستگی به تشخیص اولیه و مراقبتهای پزشکی دارد. یکی از روش‌های درمان سرطان و تومور ژن درمانی است که طی آن با ترمیم و رفع عیب ژن، بیماری را درمان می‌کنند، اما مشکلی که با آن روبه‌رو می‌شویم، ابعاد بسیار بالای داده‌های ژنتیکی است؛ از این رو برای شناسایی ژن‌های متمایز و برتر برای دسته‌بندی نیازمند فرآیند انتخاب ویژگی هستیم.

در این مقاله قصد داریم ابتدا فرآیند انتخاب ویژگی که به صورت ادغام روش‌های رتبه‌بندی است معرفی نماییم. ویژگیهای پایدار و متمایز برای شناسایی و دسته‌بندی سرطان

و نوع تومور به کمک فرآیند انتخاب ویژگی استخراج می‌گردند و به عنوان ورودی به یک سیستم فازی داده می‌شود. با استفاده از عمل بهینه‌سازی سیستم فازی، شناسایی سرطان و نوع تومور از نظر خوش خیم و بدخیم بودن، مورد بررسی قرار می‌گیرد. یک سیستم فازی از نوع TSK ایجاد می‌نماییم که بر روی تعداد قوانین و پارامترهای این نوع سیستم، با ایجاد یک ساختار سلسله مراتبی که بر اساس چندگونه‌سازی ذرات بر پایه الگوریتم ازدحام ذرات آشوبی است، بهینه‌سازی صورت می‌گیرد. سازماندهای مطالب ارائه شده در این مقاله به صورت ذیل است:

در قسمت دوم مروری بر کارهای پیشین خواهیم داشت. در قسمت سوم ابزار و روش‌هایی همچون الگوریتم بهینه‌سازی ازدحام ذرات، سیستم فازی از نوع TSK و نظریه آشوب در مسائل بهینه‌سازی مورد بررسی قرار می‌گیرد. در قسمت چهارم روش پیشنهادی توصیف می‌گردد و در قسمت پنجم بررسی نتایج شبیه‌سازی حاصل از الگوریتم ارائه شده و مقایسه نتایج به صورت تفکیک شده بر روی پایگاه داده uci- breast cancer و microarray صورت می‌گیرد.

۲- مروری بر کارهای پیشین

انتخاب ویژگی برای بیان داده‌های ژنتیکی اغلب به نام انتخاب ژن شناخته می‌شود. هدف اول از انتخاب ژن، تعیین ارتباط ژن‌های بیماری خاص است و هدف دوم انتخاب ژن‌های متمایز برای شناسایی و دسته‌بندی الگوی بیماریهای متفاوت است.

در الگوریتم‌های انتخاب، ویژگی مبتنی بر روش‌های رتبه‌بندی است [۱] انتخاب هر ویژگی به صورت منفرد و بدون در نظر گرفتن ارتباط میان ویژگیها انجام می‌شود. دسته دیگری از الگوریتم‌های انتخاب ویژگی مبتنی بر

پیشرفت کرده و مورد استفاده قرار گرفته است. در زمینه پزشکی از سیستم‌های فازی بسیاری برای حمایت از تصمیم‌گیری ساخته شده است که در زمینه مدیریت بر روی سرطان توسعه زیادی یافته‌اند. با شناسایی نوع بیماری و سطوح مختلف آن می‌توان بر اساس متدلوژیهای به‌کاررفته در هوش مصنوعی صورت گیرد [۵].

تشخیص سرطان و نوع تومور بر اساس نتایج عکس‌برداری، کار خسته‌کننده‌ای است و تجزیه و تحلیل نتایج حاصل از عکس‌برداری توسط رادیولوژیست، ممکن است منجر به ایجاد اختلال در تشخیص و پیش‌بینی سرطان گردد. برای این منظور بر اساس فرآیند داده کاوی می‌توان الگو و دانش مشترک را از میان مجموعه داده‌ها استخراج کرد و با انتساب برچسب‌های گسسته به نمونه‌های مورد پردازش به کمک تکنیک‌های خوشه‌بندی، دسته‌بندی داده‌ها، شناسایی و تشخیص سرطان و نوع تومور را انجام داد. در زمینه شناسایی و پیش‌بینی نوع سرطان از تکنیک‌های فرآیند داده کاوی و یادگیری ماشین استفاده می‌گردد [۶].

یکی دیگر از روش‌ها برای شناسایی و آنالیز داده‌های سرطان، به‌کارگیری داده کاوی به همراه الگوریتم‌های تکاملی است [۷]. خوشه‌بندی فازی نیز جایگاه ویژه‌ای در تشخیص سرطان و نوع تومور دارد. تجزیه و تحلیل مؤلفه‌های مستقل و خوشه‌بندی فازی بر روی داده‌ها به منظور تشخیص و برچسب‌گذاری ناحیه مشکوک، می‌تواند بر اساس خوشه‌بندی فازی صورت پذیرد [۸] علاوه بر این اگر قوانین فازی موجود در یک سیستم فازی به گونه‌ای تولید گردد که در مقابل تغییرات داده‌های ورودی حساس باشد، توسط این قوانین می‌توانیم طبقه‌بندی مجموعه داده‌های ژنتیکی را انجام دهیم. به عنوان مثال در [۹] به الگوریتم بهینه‌سازی تکاملی (الگوریتم ژنتیک) و ترکیب آن با قوانین فازی پایه در جهت دسته‌بندی داده‌ها به منظور

انتخاب زیرمجموعه‌ای از ویژگیهای برتر وابسته به ارتباط میان ویژگیها هستند [۲]. مقاومت الگوریتم‌های انتخاب ویژگی مرتبط با حساسیت الگوریتم در مواجهه با شرایط مختلف اجرا، از قبیل ایجاد اختلال در داده‌های آموزشی است. اگر الگوریتم انتخاب ویژگی مقاوم نباشد، با اضافه و کم کردن نمونه‌های آموزشی در هر اجرا، نتایج متفاوتی از اجرای الگوریتم مشاهده می‌کنیم. برای بالا بردن میزان مقاومت و غیر نکردن در بهینه‌های محلی نویسنده [۳] با ادغام معیارهای گوناگون توصیف بهتری برای بیان ارتباطات ویژگیها و ساختن مدل قوی‌تر و مناسب ارائه داده است. همچنین با فرآیند بازگشت به عقب، ویژگیهای افزونه را از میان مجموعه ویژگیها حذف می‌نماید. نکته‌ای که اهمیت به سزایی دارد، میزان پایداری ویژگیهای برگزیده به وسیله الگوریتم انتخاب ویژگی است؛ زیرا اگر ناپایداری در میان ویژگیهای منتخب برقرار باشد، نمی‌توان از آنها برای آزمایش‌های بعدی استفاده نمود. یکی از قالبهایی [۴] که برای ایجاد پایداری استفاده می‌گردد، دارای سه رویکرد به شرح ذیل است:

رویکرد اول: در نظر گرفتن تنوع و تغییرات بر روی مجموعه داده‌ها، به این صورت که یک تابع انتخاب ویژگی بر روی زیرمجموعه‌هایی از داده‌ها اجرا می‌گردد که در نهایت مجموعه نتایج حاصل با یکدیگر ادغام می‌شوند.

رویکرد دوم: ایجاد تنوع در استفاده از چندین متد انتخاب ویژگی بر روی یک مجموعه داده که در نهایت مجموعه نتایج حاصل با یکدیگر ادغام می‌شوند.

رویکرد سوم: ترکیب بخش اول و دوم به این منظور که علاوه بر ایجاد تنوع و تغییرات بر روی مجموعه داده‌ها، از چندین روش انتخاب ویژگی نیز استفاده گردد.

در سالهای اخیر، روش‌های هوش مصنوعی تا حد زیادی در موارد مختلف از جمله کاربردهای پزشکی

بهبود شناسایی سرطان و نوع تومور پرداخته است. مجموعه‌ای از پیش پردازش‌ها بر روی داده‌های ورودی سیستم فازی صورت می‌گیرد تا نویز موجود در میان داده‌ها حذف گردد و با اجرای الگوریتم *k-means* بر روی داده‌ها، عمل خوشه‌بندی آنها صورت پذیرد و در نهایت با نتایج تولید شده حاصل از خوشه‌بندی، توابع عضویت بر اساس فرآیند داده‌کاوی ایجاد گردد و با تولید قوانین مورد نیاز یک سیستم فازی، ساختمان یک سیستم فازی از نوع مددانی برای شناسایی سرطان و نوع تومور ایجاد شود. میزان صحت شناسایی نوع تومور از نظر بدخیم و خوش خیم بودن در این سیستم فازی بر روی پایگاه داده *uci- breast cancer* ۹۲ درصد است [۱۰].

در مقاله [۱۱] الگوریتم ازدحام ذرات خود سازمانده توصیف شده است. یادگیری در این الگوریتم از نوع بی نظارت است. نگاهی از بردار ورودی بر روی فضایی با ابعاد کمتر صورت می‌گیرد. بدین وسیله نماینده‌ای از بردار داده‌های ورودی مانند نقشه‌های خود سازمانده در شبکه عصبی ایجاد می‌گردد. ذرات در این نقشه به وسیله تغییر سرعت مطابق با معادله به روزرسانی سرعت در الگوریتم استاندارد بهینه‌سازی انبوه ذرات نسبت به ورود داده‌ها از خود واکنش نشان می‌دهند. در این الگوریتم ذرات بر اساس تعیین همسایگی متناسب، بر روی خود سازماندهی خاصی انجام می‌دهند. میزان صحت شناسایی سرطان و نوع تومور از نظر خوش خیم یا بدخیم بودن در این الگوریتم ازدحام ذرات خود سازمانده بر روی پایگاه داده *uci- breast cancer* ۹۵ درصد گزارش شده است.

نظریه آشوب در ترکیب با سایر الگوریتم‌های فرااکتشافی بسیار موفقیت آمیز، بوده است. نقشه‌های آشوبی گوناگونی می‌توانند به جای پارامتر اصلی الگوریتم بهینه‌سازی سیستم فازی استفاده شوند، اما هر کدام از

نقشه‌های آشوبی به دلیل ماهیتی که دارا هستند، متناسب با شرایط مسئله، تأثیر متفاوتی بر روی میزان صحت شناسایی سرطان و نوع تومور از نظر خوش خیم و بدخیم بودن خواهند گذاشت و جویهای متفاوتی در خروجی حاصل خواهد شد. استفاده از یک نقشه آشوبی مناسب با شرایط مسئله موجب بالا بردن کارایی الگوریتم از جنبه‌های گوناگونی می‌گردد. بر اساس مطالعات صورت گرفته [۳۶] مشخص شده است که نقشه آشوبی *Sinusoidal* به دلیل ماهیتی که داراست نسبت به سایر نقشه‌های آشوبی عملکرد بهتری بر روی سیستم شناسایی و دسته‌بندی ارائه می‌نماید.

۳- معرفی ابزار و روش‌ها

۳-۱- الگوریتم بهینه‌سازی انبوه ذرات

بهینه‌سازی گروه ذرات، الگوریتمی است که از رفتار اجتماعی موجوداتی که به صورت انبوه زندگی می‌کنند، الهام گرفته است. جمعیت اولیه این الگوریتم در یک بازه مشخص و به صورت تصادفی مقدار دهی می‌شود. هر ذره در فضای *D* بعدی به عنوان یک راه حل از مسئله محسوب می‌شود. ذره *i*ام در موقعیت x_i و دارای سرعت v_i از فضای جستجو قرار دارد. معادله به‌روز نمودن سرعت و موقعیت ذره در روابط (۱) و (۲) نمایش داده شده است که جستجو بر اساس این فرمولها برای رسیدن به یک بهینه شروع می‌شود. معادلات به‌روزرسانی مکان و سرعت برای بعد *i*ام ذره *i*ام به شکل ذیل به نمایش در می‌آید:

$$v_{i+1}^d = \chi (w * v_i^d + c_1 * rand * (pbest_i^d - x_i^d) + c_2 * rand * (gbest^d - x_i^d)) \quad (1)$$

۳-۲- سیستم فازی

سیستم فازی که در اینجا مورد بررسی قرار می‌گیرد TSK از مرتبه یک است که قانون R_i به صورت ذیل نشان داده می‌شود:

$$R_i = IF X_1(K) is A_{i1}, and \dots and x_n(k) is A_{in} \quad (4)$$

$$TEHN y(K) is f_i \quad i = 1 \dots r$$

i تعداد قوانین موجود در سیستم فازی است و f_i به صورت $a_0 + \sum_{j=1}^n a_{ij}x_j$ تعریف می‌شود. x_j ورودی به سیستم فازی است و توابع عضویت از نوع گوسی به صورت ذیل تعریف می‌شود:

$$M_{ij}(x_j) = \exp\left\{-\left(\frac{x_j - m_{ij}}{\sigma_{ij}}\right)^2\right\} \quad (5)$$

موتور استنتاج عملگر AND فازی است که به وسیله حاصل ضرب جبری پیاده‌سازی می‌شود و به ازای بردار ورودی \vec{x} میزان آتش هر قانون به این صورت محاسبه می‌شود:

$$\mu_i(\vec{x}) \quad (6)$$

$$= \prod_{j=1}^n M_{ij}(x_j) = \exp\left\{-\sum_{j=1}^n \left(\frac{x_j - m_{ij}}{\sigma_{ij}}\right)^2\right\}$$

در اینجا غیر فازی ساز میانگین مراکز تعریف می‌شود:

$$y = \frac{\sum_{i=1}^r \mu_i(\vec{x}) f_i}{\sum_{i=1}^r \mu_i(\vec{x})} \quad (7)$$

هدف بهینه‌سازی پارامترهای آزاد $a_{ij}, \sigma_{ij}, m_{ij}$ و تعداد قوانین سیستم فازی است [۲۳].

که $pbest_i$ بهترین موقعیت ملاقات شده توسط ذره i ام و $gbest$ بهترین موقعیت ملاقات شده توسط کل گروه است. برای معادله (۱) سرعت تحت تاثیر دو عامل است: (۱) مکان $pbest$ هر ذره که به عنوان مشخصه درک و شناخت است (اکتشاف محلی).

(۲) $gbest$ که مشخصه گروهی بودن است. (اکتشاف عمومی) و توسط v_t محدود شده است.

$rand$ بیانگر تصادفی در بازه $[0, 1]$ است و c_1 و c_2 ضرایب تعادل میان مؤلفه‌های شناختی و اجتماعی می‌باشند.

$$\varphi = \varphi_1 + \varphi_2 \cdot \varphi \quad (2)$$

$$\chi = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|} > 4$$

$$x_{i+1}^d = x_i^d + v_{t+1}^d$$

مشکل الگوریتم گروه ذرات بهینه‌سازی استاندارد، نبود یک مکانیزم برای کنترل اندازه سرعت ذرات است. نبود چنین مکانیزمی خطر انشعاب و انفجار (خارج از محدوده) را افزایش می‌دهد. به منظور حل شدن این مسئله مشخصه‌ای به نام w معرفی شده است که رابطه (۱) به صورت ذیل تعریف می‌شود:

$$w = w_{max} - (w_{max} - w_{min}) * \frac{iter}{iter \max} \quad (3)$$

w_{max} و w_{min} به ترتیب میزان بیشینه و کمینه اینرسی وزنی و $iter$ تکرار الگوریتم است و بالأخره $iter \max$ تعداد تکرار الگوریتم است. بهترین مقدار برای این پارامتر 0.4 تا 0.9 است و مقدار این پارامتر با گذشت زمان کاهش می‌یابد. پارامتر χ به عنوان کاهش دهنده سرعت ذرات، مقداردهی می‌شود. معادلات فوق مشخص می‌کند که خط سیر هر ذره تحت تاثیر سرعت قبلی ذره و موقعیت $pbest$ و $gbest$ است. [22].

۳-۳- آشوب و مسائل بهینه‌سازی

آشوب شاخه‌ای از حوزه‌های تحقیقات غیرخطی دینامیک است که به طور گسترده مورد مطالعه قرار گرفته است. موارد کاربردی زیادی در سیستم‌های واقعی، چه ساخت انسان و چه طبیعی، با استفاده از روش‌های نوین در حوزه غیرخطی دینامیک مورد تحقیق قرار گرفته‌اند. در نهایت، رفتار آشوب یک رفتار دقیق از سیستم‌های غیرخطی است که به ظاهر اتفاقی به نظر می‌رسد. هر چند این ظاهر اتفاقی، منشأ اتفاقی ندارد و به کلی نتیجه‌ای از یک پردازش قطعی و تعریف شده است. ریشه لغوی آشوب به کلمه رومی کائوس بر می‌گردد. آشوب در لغت به معنی در هم ریختگی، آشفتگی و بی‌نظمی است و مترادف آن در مکانیک تلاطم است. اغلب مسائل در زمینه‌های متنوع، می‌توانند در قالب یک مسئله بهینه‌سازی عمومی فرموله شوند که جستجو برای یافتن بهینه سراسری برای اغلب آنها امری بسیار مهم است. با این حال بر خلاف بهینه محلی که توسط رفتار محلی تابع نظیر گرادیان و هسیان آن توصیف شده است. یک بهینه سراسری توسط هیچ وضعیت و شرایط ریاضی توصیف نشده است؛ بنابراین مسائل بهینه‌سازی سراسری از دیدگاه محاسبات به عنوان یک چالش در حوزه ریاضیات باقی می‌ماند. در سالهای اخیر نظریه‌ها و کاربردهای دینامیکی غیرخطی، مخصوصاً نظریه آشوب، توجه بسیاری را در زمینه‌های مختلف به خود جلب نموده است. به عنوان نمونه می‌توان کنترل آشوبی و همگام‌سازی آشوبی را نام برد و در زمینه دیگری می‌توان به کاربرد بالقوه آشوب در مسائل متنوع بهینه‌سازی اشاره کرد. در حال حاضر بر روی کاربرد آشوب در مسائل بهینه‌سازی، سه گروه تحقیق می‌کنند. اولین گروه بهینه‌سازی سراسری توسط شبکه عصبی آشوبی پیشنهاد کرده‌اند [۲۴، ۲۵] که به همراه باز پخت آشوبی برای یافتن راه حل‌های بهینه سراسری

در مسائل ترکیبی نظیر فروشنده دوره گرد و زمان‌بندی نگهداری استفاده نموده‌اند. دومین گروه حل بهینه سراسری را به کمک الگوریتم بهینه‌سازی آشوبی مورد جستجو قرار داده‌اند. این دسته از الگوریتم‌ها عموماً ماهیت دنباله‌های آشوبی نظیر شبه تصادفی بودن را برای یافتن راه حل بهینه سراسری مورد استفاده قرار داده‌اند [۲۶، ۲۷]. گروه سوم از ناپایداریه‌های محاسباتی برای حل معادلات غیرخطی در مسائل بهینه‌سازی استفاده نموده‌اند [۲۸، ۲۹]. به طور کلی وجود ویژگی‌های شبه تصادفی همه جهت، میل به محدودیت و در نهایت بی‌نظمی در متغیرهای آشوب سبب شده است تا استفاده از این نظریه در الگوریتم‌های بهینه‌سازی نتایج خوبی را به دنبال داشته و از آنها در مسائل مختلف بهینه‌سازی به منظور جستجو برای یافتن راه حل بهینه سراسری استفاده می‌شود [۳۰].

نقشه آشوبی Sinusoidal با توجه به تحقیقات صورت گرفته [۳۰] و با توجه به شرایط مسئله مطرح شده، دارای عملکرد بهتری نسبت به سایر نقشه‌های آشوبی در شرایط مسئله دارا است. این نقشه آشوبی بر اساس رابطه (۸) تعیین می‌گردد. اگر مقدار $a=2.3$ و $x_0=0.7$ از رابطه‌ای ساده شده (۹) استفاده می‌گردد [۳۱].

$$x_{k+1} = ax_k^2 \sin(\pi x_k^2) \quad (8)$$

$$x_{k+1} = \sin(\pi x_k) \quad (9)$$

۴- روش پیشنهادی

گام اول در روش پیشنهادی، انتخاب ویژگی‌ای متمایز و مقاوم به وسیله ادغام روش‌های رتبه‌بندی است. پس از آن شناسایی سرطان و نوع تومور به وسیله یک ساختار سلسله مراتبی مبتنی بر چند گونه‌سازی الگوریتم بهینه‌سازی ازدحام ذرات آشوبی صورت خواهد گرفت. در سطح بالا از

کند. با وجود این که هر روشی سعی می‌کند بتواند بهترین ویژگیها را انتخاب کند، اما با توجه به وسعت جوابهای ممکن و این که این مجموعه‌های جواب به صورت توانی با N افزایش پیدا می‌کنند، پیدا کردن جواب بهینه مشکل برای N های متوسط و بزرگ بسیار پر هزینه است. در روش فرآیند انتخاب ویژگی که برگرفته از تکنیک فیلتر مبتنی بر رتبه‌بندی ویژگیهاست، بالاترین رتبه از ویژگیها که بر اساس مفاهیم آماری بیانگر اطلاعات کافی و قابلیت تمایز برای بالا بردن کارایی دسته‌بندی کننده هستند، برمی‌گزینند. در این روش از چندین روش انتخاب ویژگی بر حسب معیارهای خاصی که دارند استفاده می‌گردد. نتیجه حاصل از هر روش که شامل زیرمجموعه‌ای از ویژگیهای برتر هستند با هم ترکیب می‌شوند. با توجه به این که امتیاز هر ویژگی منحصر به فرد است، ترکیب بر اساس یک روش وزن دار صورت می‌پذیرد [۳]. در نهایت زیرمجموعه‌ای از ویژگیها حاصل از ادغام چندین روش رتبه‌بندی مختلف به عنوان خروجی ارائه می‌گردد. مجموعه روش‌های رتبه‌بندی مورد استفاده در جدول (۱) ارائه شده‌اند. در موارد فوق پارامتر x_i مقادیر ویژگی i ، میانگین، σ_{x_i} واریانس و ویژگی مذکور در کل نمونه‌ها است. پارامتر c برچسب کلاس است و پارامترهای n_1 و n_2 بیانگر تعداد نمونه‌های متعلق به هر ویژگی خاص که دارای یک برچسب کلاس مشترک هستند.

به ازای خروجی اجرای هر یک از روش‌های رتبه‌بندی، برداری از امتیاز ویژگیها تولید می‌گردد. امتیاز هر ویژگی بر اساس رابطه (۱۰) نرمالیز می‌شود:

$$\hat{u}_i = \frac{u_i - u_{imin}}{u_{imax} - u_{imin}} \quad (10)$$

u_{imax} و u_{imin} کمترین و بیشترین مقدار در بردار امتیاز است. در نهایت امتیاز هر ویژگی مبتنی بر وزن‌های میانگین

ساختار سلسله مراتبی، تولید پارامترهای سیستم فازی و بررسی کارایی آن انجام می‌شود و در سطح پایین موقعیت و سرعت ذرات به‌روزرسانی می‌گردد. روند نمای روش پیشنهادی در شکل (۱) ارائه شده است:



شکل (۱): روند نمای روش پیشنهادی

۱-۴- انتخاب ویژگی

روش‌های مختلف انتخاب ویژگی، تلاش می‌کنند تا از میان 2^N زیرمجموعه کاندید که N تعداد ویژگیها است، بهترین زیرمجموعه را پیدا کنند. در تمام این روش‌ها بر اساس کاربرد و نوع تعریف، زیرمجموعه‌ای به عنوان جواب انتخاب می‌شود که بتواند مقدار یک تابع ارزیابی را بهینه

$$u = \frac{1}{m} \sum_{i=1}^m u_i$$

(۱۱) وزن دار در رابطه (۱۱) محاسبه می‌گردد. مجموع امتیاز حاصل از هر ویژگی بر تعداد روش‌های رتبه‌بندی ادغام شد، m تقسیم می‌شود. عمل ادغام روش‌های رتبه‌بندی باعث تولید ویژگی‌های پایدار و متمایز می‌گردد:

جدول (۱): مجموعه روش‌های رتبه‌بندی

منبع	متریک	نام
[12]	$S = \text{Info}(X) - \text{Info}_x(X)$ $\text{Info}(X) = -\sum_{i=1}^k P(c_i, X) \times \log(P(c_i, X))$ $\text{Info}_x(X) = -\sum_{i=1}^v \frac{ V_i }{ X } \times \text{Info}(V_i)$ <p>K-number of classes v-number of individual values of a genesx V_i-the set of instances whose values in gene x equal x_i V_i-number of samples in V_i, $X = n$</p>	Information gain
[13]	$\text{ADC}(C, i) = \frac{\text{MI}(C, x_i)}{H(C)}$ $H(C) = -\sum_c (P(C=c) \times \log P(C=c))$ $H(X_i) = -\sum_x (P(X_i=x) \times \log P(X_i=x))$ $\text{MI}(C, X_i) = H(C)H(X_i) - H(C, X_i)$	Asymmetric Dependency Coefficient
[14]	$\text{FR}(x) = \frac{(\bar{x}_1 - \bar{x}_2)^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2}$	Fisher ratio
[15]	$S = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{x_1} + \sigma_{x_2}}$	T-test
[16]	$S = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_x}$	Z-score
[17]	$S = \bar{x}_1 - \bar{x}_2$	Fold-change difference
[18]	$S = \frac{\bar{x}_1}{\bar{x}_2}$	Fold-change ratio
[19]	$S = \sum_{j=1}^k R_j, K = \min(n_1, n_2)$	Wilconxon rank sum
[20]	$S = \left(\prod_{j=1}^n R_j \right)^{1/n}$	Rank product
[21]	$S = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_{x_1}^2/n_1 + \sigma_{x_2}^2/n_2}}$	Welch-t-test

می‌کنیم:

$$m_{1j} = x_j(0) \text{ and } \sigma_{1j} = \sigma_{init}, j = \dots n \quad (13)$$

در ابتدا مقدار σ_{init} از قبل به صورت پیش فرض تعیین می‌شود. مقدار دهی گونه اول به صورت ذیل می‌باشد:

$$\begin{aligned} \vec{s}_1^q &= [m_{11} + \Delta m_{11}^q, \sigma_{11} + \Delta \sigma_{11}^q, \dots, m_{1n} \\ &\quad + \Delta m_{1n}^q, \sigma_{1n} \\ &\quad + \Delta \sigma_{1n}^q, a_{10}^q, \dots, a_{1n}^q] \quad q \\ &= 1, \dots, p_s \end{aligned} \quad (14)$$

که Δm_{1j}^q و $\Delta \sigma_{1j}^q$ به صورت تصادفی از بازه $[-0.1, 0.1]$

انتخاب می‌شود و a_{1j}^q به صورت تصادفی از مقدار خروجی

سیستم فازی انتخاب می‌گردد.

با توجه به رابطه (۱۵) به ازای داده ورودی سیستم فازی $\vec{x}(k)$ بررسی می‌کنیم آیا نیازی به تولید قانون جدید است یا خیر؟

$$I = \arg \max_{1 \leq i \leq r} \mu_i(\vec{x}(k)) \quad (15)$$

میزان آتش داده ورودی را برای قوانین موجود محاسبه می‌کنیم. به کمک رابطه (۱۵) مشخص می‌شود که کدام قانون به ازای داده ورودی بیشتر تطابق داشته یا میزان آتش بیشتری دارد. حالا اگر رابطه $\mu_i \leq \mu_{th}$ که μ_{th} مقداری است که از پیش تعیین شده برقرار باشد، قانون جدید تولید می‌گردد. با توجه به روابط زیر پارامترهای آزاد قانون تولید شده برای مقداردهی ذرات شکل می‌گیرد:

$$m_{(r+1)i} = x_i(K) \quad i = 1, \dots, n \quad (16)$$

$$\sigma_{(r+1)i} = \alpha \sum_{j=1}^n \frac{(x_j(k) - m_{1j})^2}{\sigma_{1j}^2} \quad (17)$$

۲-۴- ساختار سلسله مراتبی برای چندگونه سازی ذرات بر پایه الگوریتم ازدحام ذرات آشوبی

این ساختار سلسله مراتبی شامل دو سطح است که سطح بالا شامل چندین گونه است. هر گونه بر اساس صفات مشترک ذرات تشکیل می‌شود. همچنین در سطح بالا چگونگی نمایش دادن پارامترهای سیستم فازی با استفاده از ذرات و بررسی کارایی آنها صورت می‌گیرد. در سطح پایین تر به روز رسانی موقعیت ذرات و سرعت آنها مورد ارزیابی قرار می‌گیرد که این فرآیند به ازای هر قانون انجام می‌شود.

۱-۲-۴- سطح بالادر ساختار سلسله مراتبی

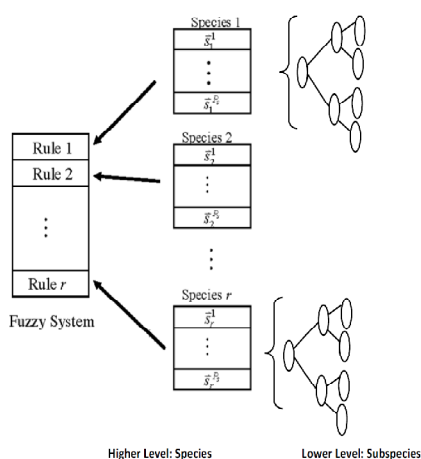
به ازای هر قانون تولید شده یک گونه به آن قانون نسبت داده می‌شود و مجموعه‌ای از ذرات که متعلق به آن گونه هستند مقدار دهی می‌شوند و هر ذره نماینده یک جواب برای پارامترهای آزاد آن قانونی است که گونه این مجموعه ذرات به آن نسبت داده شده، می‌باشد. مقدار دهی هر ذره به این صورت است که هر ذره متعلق به کدام گونه است و شماره هر ذره در هر گونه باید مشخص باشد و پارامترهایی که باید بهینه شود به ذره نسبت داده می‌شود. این پارامترها مقدار دهی اولیه می‌شوند.

$$\vec{s}_i = [m_{i1}, \sigma_{i1}, \dots, m_{in}, \sigma_{in}, a_{i1}, \dots, a_{in}] \in R^{3n+1} \quad (12)$$

با توجه به مواردی که ذکر شده است، ذرات در یک فضای ۳ بعدی قرار می‌گیرند. در ابتدا قوانین به صورت دینامیک تولید می‌شود و به ازای تولید هر قانون یک گونه شکل می‌گیرد، با شکل گیری هر گونه باید ذرات تشکیل دهنده آن مقدار دهی شود در حالت اولیه با ورود داده ورودی \vec{x} پارامترهای آزاد سیستم فازی را مقدار دهی

روز رسانی سرعت و موقعیت هر ذره بهترین ذره‌ای که انتخاب می‌گردد، پدر آن ذره در ساختار سلسله مراتبی درخت دودویی است. تنها بار محاسباتی در سطح پایین تشکیل این درخت دودویی است و در نهایت با دانستن پدر هر ذره مطابق با رابطه (۱) عمل به روز رسانی سرعت ذرات بر طبق این همسایگی صورت می‌گیرد.

استفاده از این نوع همسایگی در سطح پایین ساختار سلسله مراتبی به جای به کارگیری بهترین ذره در کل جامعه در معادله به روز رسانی سرعت، باعث افزایش سرعت همگرایی الگوریتم می‌گردد. در الگوریتم استاندارد بهینه‌سازی ازدحام ذرات به روز رسانی سرعت ذرات وابسته بهترین تجربه شخصی ذره جاری و بهترین ذره در اجتماع است. اگر موقعیت بهترین ذره در اجتماع مربوط به یکی از نقاط محلی در مسئله باشد، بقیه ذرات نیز به سوی آن جذب می‌شوند و در نقطه محلی به دائم می‌افتند؛ اما در روش پیشنهادی برای به روز رسانی سرعت و موقعیت هر ذره، بهترین ذره‌ای که انتخاب می‌گردد، پدر آن ذره در ساختار سلسله مراتبی درخت دودویی است و مانع از گیر کردن در بهینه‌های محلی در فضای مسئله می‌شود.



شکل (۲): ساختار سلسله مراتبی [۳۲]

$$i = 1, \dots, n \quad (18)$$

$$\vec{s}_{(r+1)}^q [m_{(r+1)1} + \Delta m_{(r+1)1}^q \sigma_{(r+1)1} + \Delta \sigma_{(r+1)1}^q, \dots, m_{(r+1)n} + \Delta m_{(r+1)n}^q \sigma_{(r+1)n} + \Delta \sigma_{(r+1)n}^q, a_{(r+1)0}, \dots, a_{(r+1)n}]$$

۲-۲-۴- ارزیابی کارایی ذرات

بعد از شکل‌گیری سیستم فازی از هر قانون (گونه) یک ذره انتخاب می‌شود و کارایی ذره مورد نظر بر اساس RMSE میان آن چیزی که مطلوب ما است و آن چیزی که واقعی است محاسبه می‌گردد. به این صورت که برای ارزیابی ذرات هر قانون (گونه)، هر ذره از گونه مورد نظر با بهترین ذرات مربوط به قوانین دیگر یک سیستم فازی تشکیل می‌دهند و خروجی حاصل با خروجی مطلوب مقایسه می‌شود.

$$FS1: [\bar{s}_1^g \dots \bar{s}_{i-1}^g \bar{s}_i^1 \bar{s}_{i+1}^g \dots \bar{s}_r^g] \quad (19)$$

$$FSPS: [\bar{s}_1^g \dots \bar{s}_{i-1}^g \bar{s}_i^{PS} \bar{s}_{i+1}^g \dots \bar{s}_r^g]$$

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (20)$$

از میان FS-1 تا FS-PS یک گونه آن ذره‌ای که کمترین مقدار (RMSE) را دارد، به عنوان بهترین ذره برای گونه‌ای که اختصاص دارد، انتخاب می‌گردد. ارزیابی کارایی به ازای هر قانون و برای تمام ذرات هر گونه باید محاسبه گردد.

۳-۲-۴- سطح پایین در ساختار سلسله مراتبی

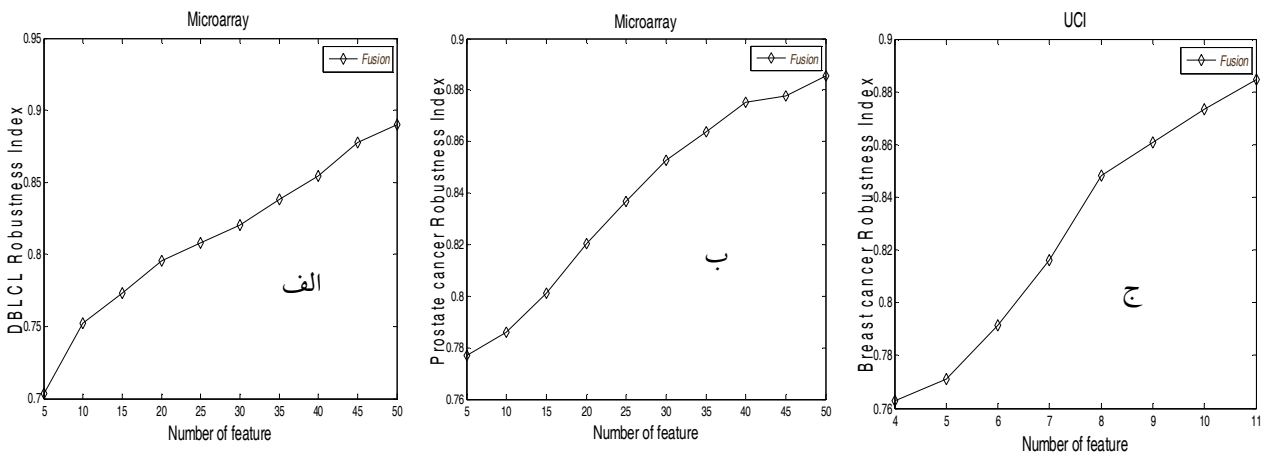
در ساختار سلسله مراتبی در سطح پایین، همسایگی که استفاده می‌گردد از نوع درخت دودویی است؛ به این نحو که کارایی ذرات هر گونه محاسبه و مرتب می‌شوند و در یک درخت دودویی چیده می‌شوند (شکل ۲). بعد از آن که ذرات در یک درخت دودویی چیده شدند، هر ذره یکی از گره‌های تشکیل دهنده درخت دودویی است و برای به

۴-۲-۴- ساختار سلسله مراتبی مبتنی بر چندگونه سازی ذرات بر پایه الگوریتم ازدحام ذرات آشوبی

در ساختار سلسله مراتبی که ذکر شده، سرعت ذرات بر اساس رابطه (۱) به روزرسانی می‌گردد. پارامتر c_1 و c_2 مقدار ثابت یک را اختیار می‌کنند و مقادیر rand یک مقدار تصادفی در بازه (۰،۱) است. نقشه‌های آشوبی می‌توانند جایگزین این مقدار تصادفی شوند. رابطه (۸) یا (۹) جایگزین مقدار پارامتر rand رابطه (۱) می‌شود؛ از این رو استفاده از نقشه آشوب Sinusoidal به جای مقادیر تصادفی باعث افزایش تنوع در میان ذرات جمعیت و همچنین موجب گیر نکردن در بهینه‌های محلی و افزایش قدرت

دسترسی برای پیدا کردن بهینه سراسری در فضای جستجو خواهند شد. جایگذاری صورت گرفته در رابطه (۲۱) ارائه شده است [۳۲]:

$$\begin{aligned} \text{rand} &= ax_k^2 \sin(\pi x_k^2) \\ \text{rand} &= \sin(\pi x_k) \\ v_{i+1}^d &= \chi \left(w * v_i^d + c_1 * \text{rand} * (pbest_i^d - x_i^d) \right. \\ &\quad \left. * \text{rand} * (gbest^d - x_i^d) \right) \end{aligned} \quad (21)$$



شکل (۳): مقاومت بر مقاومت بر روی مجموعه دادگان (الف) سرطان سینه، (ب) پروستات، (ج) بافت لنفاوی سلول B بزرگ

آموزش و آزمایش است. خلاصه‌ای از اطلاعات ۳ پایگاه داده معرفی شده در جدول (۲) مشاهده می‌شود.

جدول (۲): پایگاه داده‌های ارزیابی

منبع	تعداد ویژگی‌ها	تعداد نمونه‌ها	نام پایگاه داده
Uci[33]	۱۱	۷۰۰	سرطان سینه
Microarray[34]	۱۱۲۲۶	۸۳	سرطان پروستات
Microarray[35]	۲۳۰۹	۷۲	سرطان بافت لنفاوی سلول B بزرگ

۵- نتایج پیاده‌سازی

۵-۱- معرفی دادگان

به منظور ارزیابی روش‌های ارائه شده، آزمایش‌های صورت گرفته به وسیله نرم‌افزار Matlab بر روی ۳ پایگاه داده بیان‌کننده ژن مربوط به مجموعه داده‌های ژنتیکی انجام شده است. مراحل آزمایش روش ارائه شده شامل دو فاز

۲-۵- معیارهای ارزیابی

معیارهای مورد استفاده برای ارزیابی روش کار ارائه شده شامل بررسی مقاومت و پایداری برای فرآیند انتخاب ویژگی و دقت دسته‌بندی مربوط به پیش‌بینی نمونه‌های ۳ پایگاه داده معرفی شده، در فاز آزمایش است.

- دقت دسته‌بندی

دقت دسته‌بندی، اصلی‌ترین معیار ارزیابی مربوط به دسته‌بندی و پیش‌بینی نمونه‌ها در فاز آزمایش است. در جدول (۳) اطلاعات مربوط به سازماندهی نمونه‌های هر کلاس که به درستی و یا نادرستی دسته‌بندی شده‌اند، تعریف می‌شوند. در این

تحقیق کلاس اکثریت به عنوان کلاس منفی و کلاس اقلیت به عنوان کلاس مثبت نامبرده شده است. با توجه به این که کارایی دسته‌بندی مسائل دو کلاسه بر اساس جدول (۳) محاسبه می‌گردد، دقت دسته‌بندی در معادله (۲۲) تعریف می‌شود:

$$curacy = \frac{TP + TN}{TP + TN + FN + FP} \quad (22)$$

جدول (۳): معرفی معیارهای دسته‌بندی

پیش‌بینی نمونه‌های	پیش‌بینی نمونه‌های
کلاس مثبت	کلاس منفی
کلاس مثبت	منفی نادرست (TP)
کلاس مثبت	منفی درست (FN)
کلاس منفی	مثبت درست (FP)
کلاس منفی	مثبت نادرست (TN)

- مقاومت

محاسبه معیار مقاومت به این صورت است که، فرض کنیم m تعداد دسته داده با تکرارهای مختلف از مجموعه داده اصلی باشد و از هر دسته داده یک زیرمجموعه از ویژگی‌های متمایز انتخاب شده باشد. برای اندازه‌گیری کلی و دقیق‌تر شباهت هر مجموعه با مجموعه دیگر از رابطه (۲۳)

استفاده می‌گردد. پارامترهای S_0 و S_i دو زیرمجموعه از ویژگی‌های منتخب هستند:

$$JC_i(k) = \frac{|S_i \cap S_0| + SC_i}{k} \quad (23)$$

با در نظر گرفتن پیوستگی ویژگی‌های انتخاب شده در مجموعه‌های S_0 و S_i مقدار پارامتر SC_i برابر با حاصل جمع قدر مطلق بیشترین مقدار وابستگی میان ویژگی‌های منتخب است. $|S_i \cap S_0|$ معادل با تعداد ویژگی‌های مشترک در میان دو مجموعه انتخاب ویژگی است. K برابر با اندازه مجموعه S_0 و S_i است. حال میزان مقاومت الگوریتم انتخاب ویژگی به وسیله معادله (۲۴) محاسبه می‌شود. تغییرات اندیس مقاومت بین بازه $[0 - 1]$ است.

$$\overline{JC}(k) = \frac{\sum_{i=1}^m JC_i(k)}{m} \quad (24)$$

۳-۵- ارزیابی مقاومت فرآیند انتخاب

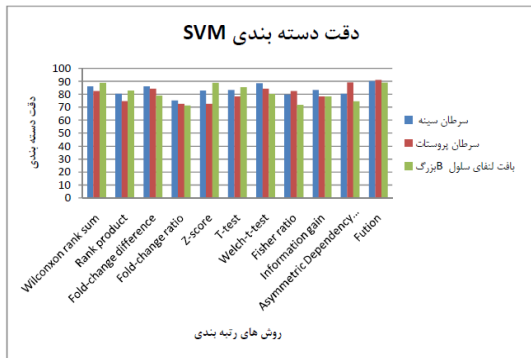
ویژگی

بهترین ژن‌های متمایز و برتر انتخاب شده مبتنی بر روش ادغام روش‌های رتبه‌بندی برای هر یک از پایگاه داده‌ها در جدول (۴) ارائه می‌گردد:

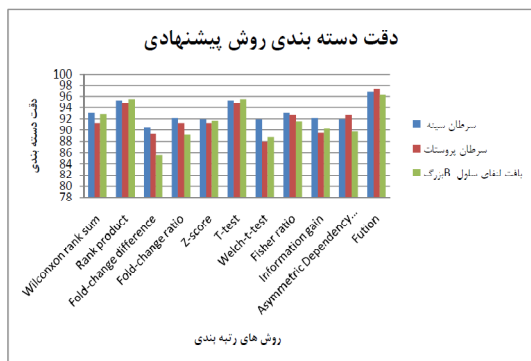
جدول (۴): پایگاه داده‌های ارزیابی

نام پایگاه داده	تعداد نمونه‌ها	تعداد ویژگی‌ها
سرطان سینه	۷۰۰	۴
سرطان پروستات	۸۳	۵
سرطان بافت لنفاوی سلول B بزرگ	۷۲	۵

همان طور که در شکل (۳) مشاهده می‌نمایم، میزان تغییر نمودار مقاومت روش پیشنهادی که بر اساس رابطه (۲۴) محاسبه گردیده است، به ازای افزایش تعداد



شکل (۵): میزان صحت شناسایی سرطان و نوع تومور به وسیله دسته‌بندی کننده نزدیک‌ترین همسایه SVM



شکل (۶): میزان صحت شناسایی سرطان و نوع تومور به وسیله دسته‌بندی کننده روش پیشنهادی

در شکل‌های (۴) و (۵) دقت دسته‌بندی برای روش ادغام روش‌های رتبه‌بندی مستقل از نوع دسته‌بندی کننده مبتنی بر روش نزدیک‌ترین همسایه و ماشین بردار پشتیبان بر روی مجموعه داده‌گان ژنتیکی در بازه [۰.۸۸، ۰.۹۱] تغییر می‌کند، که نسبت به سایر روش‌های رتبه‌بندی دقت بالاتری را ارائه می‌نماید.

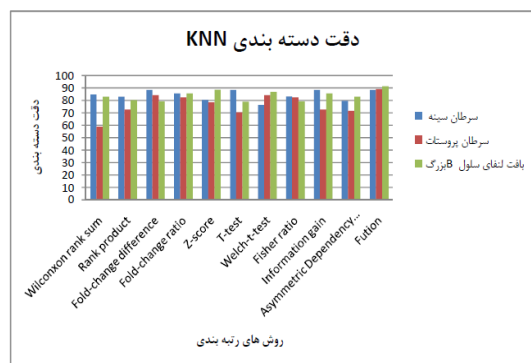
دقت دسته‌بندی برای روش پیشنهادی در شکل (۶) بیش از ۹۵ درصد است. اطلاعات به نمایش گذاشته شده بیان کننده کارایی روش پیشنهادی برای انتخاب ویژگی و روش پیشنهادی به دلیل بهینه‌سازی پارامترهای سیستم فازی به وسیله الگوریتم بهینه‌سازی ازدحام ذرات آشوبی است. این سیستم فازی بهینه شده به عنوان یک دسته‌بندی کننده،

ویژگی‌های منتخب در بازه [۰.۷۰، ۰.۸۸] می‌کند. با افزایش تعداد ویژگی منتخب مقاومت روش پیشنهادی بیشتر می‌شود. محدود بودن این تغییرات در ازای اختلال بر روی مجموعه داده‌گان در فاز آموزش نشان دهنده پایداری روش پیشنهادی در فاز مربوط به فرآیند انتخاب ویژگی است.

۴-۵ نتایج بررسی دقت دسته‌بندی

نتایج میزان صحت شناسایی و نوع تومور مبتنی بر روش‌های دسته‌بندی کننده نزدیک‌ترین همسایگی (KNN)، ماشین بردار پشتیبان (SVM) و روش پیشنهادی به صورت اجرای منحصر به فرد هر روش رتبه‌بندی و ادغام روش‌های رتبه‌بندی برای انتخاب ویژگی پایدار و متمایز به ترتیب، در شکل‌های (۴)، (۵) و (۶) ارائه شده است.

با توجه به اطلاعات شکل‌های (۴)، (۵) و (۶) دقت دسته‌بندی روش پیشنهادی در مقابل دسته‌بندی کننده نزدیک‌ترین همسایه و ماشین بردار پشتیبان، به ازای برترین تعداد ویژگی‌های منتخب شده در حالاتی که توابع رتبه‌بندی منحصر به فرد یا به صورت ادغام شده اجرا می‌گردند، بیشتر است.



شکل (۴): میزان صحت شناسایی سرطان و نوع تومور به وسیله دسته‌بندی کننده نزدیک‌ترین همسایه KNN

عملکرد بهتری نسبت به دسته‌بندی کننده نزدیکترین همسایه و ماشین بردار پشتیبان دارد. از این رو به کارگیری ادغام روش‌های رتبه‌بندی به عنوان فرآیند انتخاب ویژگی‌های پایدار و متمایز برای دسته‌بندی و استفاده از ساختار سلسله مراتبی مبتنی بر چندگونه سازی ذرات بر پایه الگوریتم ازدحام ذرات آشوبی به عنوان دسته‌بندی کننده در کنار هم می‌تواند تأثیر بسزایی بر روی پیش‌بینی تشخیص سرطان و نوع تومور داده‌های ژنتیکی با ابعاد بالا دارد.

نتیجه‌گیری

هدف از این مقاله شناسایی و پیش‌بینی سرطان و نوع تومور برای داده‌های ژنتیکی با ابعاد بالاست که در دو بخش قرار می‌گیرد. بخش اول شامل انتخاب ویژگی‌های پایدار و متمایز به وسیله ادغام نتایج حاصل از روش‌های رتبه‌بندی است. بخش دوم شامل شناسایی سرطان و نوع تومور بر اساس متدولوژی‌های هوشمندانه بوده است. روش پیشنهادی بهینه‌سازی سیستم فازی بر مبنای یک ساختار سلسله مراتبی بر اساس چندگونه سازی ذرات با به کارگیری الگوریتم ازدحام ذرات آشوبی است و عمل بهینه‌سازی هم بر روی

مراجع

1. Cosmin Iazar, J. T. (2012). "A Survey On Filter Techniques For Feature Selection In Gene Expression Microarray Analysis" s. IEEE/ACM Transactions On Computational Biology And Bioinformatics, 14.
2. Ruiz .R, J.C. Riquelme, J.S.Aguilar-Ruize, M.Garcia-Torres (2012). "Fast Feature Selection Aimed At High-Dimensional Data Via Hybrid-Sequential-Ranked Searches". Expert Systems With Applications , 9.
3. Mao, F. Y. (2012). "Robust Feature Selection For Microarray Data Based On Multicriterion Fusio". Ieee/Acm Transactions on Computational Biology And Bioinformatics, 13.
4. Waelawada, T. M. (2012). "A Review Of The Stability Of Feature Selection Techniques For Bioinformatics Data". (P. 8). Las Vegas, Nevada, Usa: IEEE.
5. Ahmed Abou Elfetouh Saleh, S. E. (2011). "A Fuzzy Decision Support System for Management of Breast Cancer". International Journal of Advanced Computer Science and Applications , 7.
6. Shelly Gupta, D. K. (2011). "Data Mining Classification Techniques Applied Forbreast Cancer Diagnosis And Prognosis". Indian Journal Of Computer Science And Engineering (Ijcse) , 8.
7. Wei-Pin, C. D.-M. (2007). "Comparison of Three Data Mining Techniques with Genetic Algorithm in the Analysis of Breast Cancer Data". First Authors Last Name , 26
8. Abu-Amara, I. A.-Q. (2008). "A Computer-Aided Diagnosis System for Breast Cancer Using Independent Component Analysis and Fuzzy Classifier". modelling and simulation in engineering , 10

تعداد قوانین سیستم فازی از نوع TSK و هم بر روی پارامترهای آزاد صورت گیرد. در این فرآیند با تولید قوانین، گونه‌هایی تشکیل می‌گردد و ذرات داخل هر گونه، ایجاد و مقداردهی می‌شوند آنگاه بر اساس کارایی که ذرات دارند مرتب می‌گردند و در یک درخت دودویی چیده می‌شوند. برای به‌روزرسانی معادله سرعت و موقعیت برای هر ذره پدر آن در ساختار درختی بهترین همسایگی برای ذره فعلی است و از طرفی عمل به‌روزرسانی معادله سرعت به وسیله جایگزینی نقشه آشوبی Sinusoidal به جای پارامتر تصادفی rand صورت می‌پذیرد که باعث افزایش صحت پیش‌بینی در فاز آزمایش برای شناسایی سرطان و نوع تومور برای داده‌های ژنتیکی با ابعاد بالا می‌گردد. میزان صحت شناسایی و دسته‌بندی بر روی داده‌های بیولوژیکی بیش از ۹۵ درصد است و از نظر دسته‌بندی خوش خیم یا بدخیم بودن، دارای قابلیت بالاتری برای تشخیص سرطان و نوع تومور دارا است.

9. Nakashima, G. S. (2010). "Hybrid cost-sensitive fuzzy classification for breast cancer diagnosis". 32nd Annual International Conference of the IEEE EMBS (p. 4). Buenos Aires, Argentina, : ieeec"
10. M.yassi, M.yaghoobi (2012). "Creating fuzzy rule based on clustering for the detection of breast cancer".(p. 6).Iran:12th Iranian Conference on Fuzzy System(ICFS2012)
11. M. O'Neill, A. Brabazon. (2006), "Self-organizing swarm (SOSwarm): a particle swarm algorithm for unsupervised Learning", in: proceedings of IEEE Congress on Evolutionary Computation, July 16–21, Vancouver, BC, Canada, pp. 634–639.
12. L.-Y. Chuang et al. (2009). "A Two-Stage Feature Selection Method for Gene Expression Data," OMICS: J. Integrative Biology, vol. 13, pp. 127-137,
13. D.V. Shridhar, E.B. Bartlett, And R.C. Seagrave.(1998) "Information Theoretic Subset Selection For Neural Network Models," Computer And Chemical Eng., Vol. 22, Nos. 4-5, Pp. 613-626.
14. Fisher, R., (1936). "The Use Of Multiple Measurements In Taxonomic Problems." Annals Eugen. 7, 179–188.
15. J.G. Thomas Et Al.(2001). "An Efficient And Robust Statistical Modeling Approach To Discover Differentially Expressed Genes Using Genomic Expression Profiles," Genome Research, Vol. 11, No. 7, Pp. 1227-1236.
16. T.R. Golub Et Al.(1999). "Molecular Classification Of Cancer: Class Discovery And Class Prediction" By Gene Expression Monitorin Science, Vol. 286, No. 5439, Pp. 531-537.
17. M.K. Kerr, M. Martin, and G.A. Churchill. (2000). "Analysis Of Variance for Gene Expression Microarray Data," J. Computational Biology, Vol. 7, No. 6, Pp. 819-837.
18. H. Tao Et Al.(1999). " Functional Genomics: Expression Analysis Of Escherichia Coli Growing On Minimal And Rich Media," J. Bacteriology, Vol. 181, Pp. 6425-6440.
19. L. Deng Et Al.(2004). "A Rank Sum Test Method For Informative Gene Discovery," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery And Data Mining, Pp.410-41.
20. R. Breitling Et Al.(2004). "Rank Products: A Simple, Yet Powerful, New Method To Detect Differentially Regulated Genes In Replicated Microarray Experiments," FEBS Letters, Vol. 573, Nos. 1-3. Pp. 83-92.
21. S. Dudoit Et Al.(2002). "Statistical Methods For Identifying Differentially Expressed Genes In Replicated Cdna Microarray Experiments," Statistica Sinica, vol. 12, pp. 111-139.
22. Kennedy, J. and Eberhart, R.C., (1995), "Particle swarm optimization", Proc. IEEE Int. Conf. on N.N., pp. 1942-1948
23. T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Trans.
24. Arora J. S, Elwakeil O. A.(1995) "Global Optimization Methods For Engineering Application", A Review.Struct Optim, Vol. 9, Pp. 137-59.
25. T. Kowk, K. Smith,(1999) "A unified framework for chaotic neural network approaches to combinatorial optimization", IEEE Trans Neural Networks, vol. 10, no. 4, pp. 978-81.
26. B. Li, W. S. Jiang (1998), "Optimizing complex function by chaos search", Cybern Syst , vol. 29, no. 4, pp. 409-19
27. P. Zhang, C. He, X. Zhang, Y. Qiao,(2001) "The survey and future development on global optimization algorithms in engineering", In: Extended abstracts of the fourth world congress of structural and multidisciplinary optimization (WCSMO-4), Dalian, China, pp. 328-9,
28. V. Jovanovic , (2000) "Chaotic descent method and fractal conjecture", Int J Numer Methods Eng , vol. 48, no. 1, pp. 137-52.
29. V. Jovanovic, K. Kazerounian ,(2000) "Optimal design using chaotic descent method", ASME J Mech Des, vol. 122, no. 3, pp. 265-70.
30. L. Chen, K. Aihara,(1995) "Chaotic simulated annealing by a neural network model with transient chaos", Neural Networks, vol. 8, no. 6, pp. 915-30.
31. Ott E, Grebogi C , Yorke J A (1990). "Controlling Chaos", Phys Rev Lett , Vol. 64, No. 11, Pp. 1196-9.
32. M.yassi, M.yaghoobi(2013). "Distinguishing and clustering breast cancer according to hierarchical structures based on multispecies particle swarm optimization".(p. 6).Iran:11th Iranian Conference on Intelligence systems(ICIS)

33. <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer/>
34. <http://www-genome.wi.mit.edu/cgi-bin/cancer/datasets.cgi>
35. <http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi>
36. M.yassi,A.yassi,M.yaghoobi(2014). “**Distinguishing and clustering breast cancer according to hierarchical structures based on chaotic multispecies particle swarm optimization**”.(p. 6).Iran:12th Iranian Conference on Intelligence systems(ICIS2014)