ارائه یک روش یادگیری ویژگی ترکیبی مبتنی بر الگوریتم شبیه‌سازی تبرید و برنامه‌نویسی ژنتیک
(مطالعه موردی: تشخیص بدخیمی سرطان سینه)

رسول صادقی[۱]، دانشجوی کارشناسی ارشد؛ فردین ابدالی محمدی[۲]، استادیار
۱- گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی‌ومهندسی، دانشگاه رازی، کرمانشاه، ایران، rasool.sadeghi@stu.razi.ac.ir
۲- گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی‌ومهندسی، دانشگاه رازی، کرمانشاه، ایران، fardin.abdali@razi.ac.ir

چکیده: امروزه استفاده از ابزارهای یادگیری ماشین در حوزه‌های مختلف ازجمله تشخیص بیماری‌ها در حال گسترش است. علت این امر را می‌توان عملکرد متغیر و متمایل به خطای انسان در مقابل عملکرد ثابت ابزارهای یادگیری ماشین در زمینه تشخیص و طبقه‌بندی دانست. حیاتی بودن تشخیص در حوزه‌هایی مانند پزشکی، نیاز به بهبود تشخیص با روش‌های یادگیری ماشین را توجیه می‌کند. ازجمله روش‌های افزایش دقت در این زمینه، الگوریتم‌های کاهش ویژگی و یادگیری ویژگی هستند. در این مقاله با ارائه یک روش یادگیری ویژگی، دقت روش‌های مبتنی بر یادگیری‌ماشین افزایش یافته است. روش پیشنهادی شامل سه فاز افزایش کیفیت داده، انتخاب ویژگی و یادگیری ویژگی است. در فاز اول، مقادیر ازدست‌رفته با شاخص پراکندگی میانگین و یا مد جایگزین می‌شوند در فاز دوم، الگوریتمی مبتنی بر الگوریتم شبیه‌سازی تبرید برای کاهش ویژگی و یافتن بهترین زیرمجموعه از ویژگی‌ها ارائه شده است. در فاز نهایی نیز الگوریتمی مبتنی بر الگوریتم برنامه‌نویسی ژنتیک به‌منظور یادگیری ویژگی‌های متمایزکننده ترکیبی ارائه شده است. روش پیشنهادی با استفاده از دو مجموعه‌داده استاندارد WBCD و WDBC ارزیابی شده است. نتایج به‌دست‌آمده با آخرین دستاوردها مقایسه شده است که حاکی از عملکرد بهبودیافته الگوریتم پیشنهادی است.

واژه‌های کلیدی: یادگیری ماشین، یادگیری ویژگی، برنامه‌نویسی ژنتیک، کاهش ویژگی، سرطان سینه

A Combined Feature-Learning Method Based on Simulated Annealing Algorithm and Genetic Programming (Case Study: Malignant Breast Cancer Diagnosis)

R. Sadeghi[1], MSc student; F. Abdali Mohammadi[2], Assistant Professor

[1] Department of Computer Engineering and Information Technology, Faculty of Engineering, Razi University, Kermanshah, Iran, Email: rasool.sadeghi@stu.razi.ac.ir
[2] Department of Computer Engineering and Information Technology, Faculty of Engineering, Razi University, Kermanshah, Iran, Email: fardin.abdali@razi.ac.ir

Abstract: Nowadays using machine learning tools in different areas such as disease detection is expanding. Origins of this expansion can be found in humans' unstable performance and stable performance of machine learning tools. Criticality of detection in areas such as medical proves the need for improvement in machine learning methods. feature reduction and feature learning are two ways that cause to precision increment. In this paper precision of machine learning algorithms is increased by feature learning. The proposed method contains three steps: data quality increment, feature selection, and feature learning. In the first step missing values are replaced with mean or mode (distribution index). In the second step a simulated annealing-based algorithm is presented to utilized as feature selection process and finding the best subset of features. In the final step, a genetic programming algorithms is presented to do the feature learning step. The proposed method is evaluated on two benchmark datasets (WBCD and WDBC). The results show performance improvement in machine learning algorithms in terms of precision if the proposed method used.

Keywords: Machine learning, feature learning, genetic programming, feature reduction, breast cancer

تاریخ ارسال مقاله: ۱۳۹۵/۰۸/۱۹
تاریخ اصلاح مقاله: ۱۳۹۵/۱۰/۱۱
تاریخ پذیرش مقاله: ۱۳۹۵/۱۲/۰۹
نام نویسنده مسئول: فردین ابدالی محمدی
نشانی نویسنده مسئول: ایران - کرمانشاه – باغ ابریشم - دانشگاه رازی - دانشکده فنی و مهندسی.

۱- مقدمه

دسته بزرگ و مهمی از مسائلی که می‌تواند به‌کمک کامپیوتر پاسخ داده شود، مسائل مربوط به تشخیص بیماری‌ها است. مزایای استفاده از کامپیوتر در حل چنین مسائلی را می‌توان این‌گونه برشمرد: عملکرد و دقت ثابت و بدون تأثیرپذیری از محیط، سرعت ثابت و بالا در تشخیص، عدم محدودیت جغرافیایی و عدم محدودیت زمانی در استفاده. البته لازمه این امر، ارتقاء دقت الگوریتم‌های تشخیصی است. با توجه به آمار منتشرشده از منابع معتبر [۱]، سرطان سینه یکی از بیش‌ترین موارد تشخیص داده‌شده سرطان در ایران و سایر کشورها است. برای مثال بر اساس آمار صادره از موسسه تحقیقات، درمان و آموزش سرطان در ایران [۱] در سال ۱۳۸۷، سرطان سینه با ۸۶۱۶ مورد جدید، پس از سرطان پوست، دومین نوع عمده از سرطان در ایران تشخیص داده شده است. این آمار در کشور آمریکا [۲]، در سال ۲۰۱۴ با ۲۹۵۲۴۰ مورد، بیش‌ترین بدخیمی تشخیص داده‌شده در این کشور بوده است. از آن‌جایی که یکی از بیش‌ترین بدخیمی‌های تشخیص داده‌شده در ایران و سایر کشورها سرطان سینه است؛ مبارزه با این بیماری یک تلاش همگانی و یکپارچه را می‌طلبد. یکی از راه‌های کاهش آمار مرگ‌ومیر ناشی از این بیماری اجرای طرح‌های غربالگری عمومی است که نقش به‌سزایی در تشخیص به‌موقع این بیماری و به‌تبع آن نجات بیماران دارد. با توجه به آمار بالای ابتلای به این بیماری، اجرای چنین طرح عظیمی به‌صورت دستی و بدون دخالت ابزارهای مکانیزه بسیار زمان‌بر و مستعد خطاست. به‌طور قطع، ابزارهای یادگیری‌ماشین می‌توانند تأثیر به‌سزایی در افزایش سرعت و دقت و کاهش خطا در تشخیص این بیماری داشته باشند؛ همچنین با توجه به ماهیت الگوریتمی روش‌های پیشنهادی در زمینه تشخیص بیماری، روش‌های مذکور بر روی سایر بیماری‌ها نیز قابل تعمیم هستند.

ویژگی‌های مورداستفاده در مقاله پیش رو شامل مقادیری عددی است؛ که توسط متخصصین پزشکی از تصاویر ماموگرافی استخراج شده است. از طرفی عملکرد الگوریتم‌های یادگیری ماشین به‌شدت وابسته به ویژگی‌هایی است که یادگیری براساس آن‌ها انجام می‌شود [۳]؛ بنابراین استفاده از ویژگی‌های مناسب، لازمه استفاده بهینه از الگوریتم‌های یادگیری ماشین است. راهکارهایی که افزایش دقت الگوریتم‌های یادگیری ماشین را به دنبال دارد شامل کاهش ویژگی، نگاشت ویژگی، برطرف کردن همبستگی‌ها و وزن‌دار کردن ویژگی‌ها است. یکی از روش‌هایی که تمامی موارد گفته‌شده را به‌صورت ضمنی انجام می‌دهد، یادگیری ویژگی است. یادگیری ویژگی یکی از مباحث مطرح شده در حوزه یادگیری ماشین است. هدف از یادگیری ویژگی یافتن و ایجاد مجموعه‌ای از ویژگی‌ها است که منجر به اجرای بهینه و حتی سریع‌تر الگوریتم‌های یادگیری ماشین در فاز عملی خواهد شد. به بیان ساده می‌توان یادگیری ویژگی را یک تابع انتقال به‌صورت $T:\mathbf{F}\rightarrow\mathbf{NF}$ بیان کرد که در آن F بردار ویژگی‌های اولیه با مقادیر خام و NF بردار ویژگی‌هایی است که بر اساس F به دست آمده است.

برنامه‌نویسی ژنتیک (GP) درواقع شکل توسعه‌یافته‌ای از روش‌های یادگیری ماشین و الگوریتم ژنتیک (GA) است که می‌تواند در یادگیری ویژگی استفاده شود. این الگوریتم دارای وجوه تمایزی با GA نیز هست؛ ازجمله اینکه برخلاف GA، در GP نیازی به مشخص بودن دامنه‌ی پاسخ در یک محدوده خاص نیست (در GA تعداد کاراکترهای رشته نهایی از قبل مشخص است). موضوعی که سبب انعطاف بیش‌تر این روش نسبت به GA شده این است که برخلاف GA که خروجی آن یک رشته باینری است؛ خروجی GP یک رابطه ریاضی بهینه است که بیش‌ترین برازندگی را با مسئله داشته و با کم‌ترین خطای ممکن، ورودی‌های مسئله را به خروجی‌های آن نگاشت می‌کند [۴]. این ویژگی GP باعث می‌شود که خروجی آن به راحتی در نرم‌افزارها و یا حتی سخت‌افزارهای بلادرنگ قابل پیاده‌سازی باشد. در این مقاله تلاش شده است که به‌کمک الگوریتم GP، یادگیری ویژگی صورت گرفته و دقت تشخیص و طبقه‌بندی تا حد مطلوب افزایش یابد.

در این مقاله به‌منظور کاهش ویژگی‌ها نیز تلاش شده ابتدا به‌کمک الگوریتم‌های بهینه‌سازی، ابعاد مسئله کاهش یابد. چرایی این امر را می‌توان زمان‌بر بودن فاز تغییر فضای مسئله به‌کمک الگوریتم GP بر اساس زیر مجموعه‌ای از ویژگی‌ها دانست؛ چراکه در فاز مذکور، ساخت ژن بر اساس تمامی ترکیب‌های ممکن از ویژگی‌ها بررسی می‌شود. لذا ناگزیر به کنار گذاشتن ویژگی‌های بی‌ارزش پیش از فاز مذکور هستیم و ابزارهای بهینه‌سازی و به‌طور خاص الگوریتم بهینه‌سازی شبیه‌سازی تبرید، گزینه مناسبی برای این منظور خواهد بود. مزیت مذکور علاوه بر تأثیر مثبتی است که کاهش ویژگی در کاربردهای عملی و کاهش ورودی‌های مورد نیاز مسئله و به‌تبع آن کاهش هزینه‌ها دارد. کاهش ویژگی به‌گونه‌ای انجام می‌شود که به‌جای استفاده از n کمیت اندازه‌گیری‌شده برای طبقه‌بندی، از k کمیت استفاده شود به‌گونه‌ای که $k<n$ باشد. این کار به تنهایی نقش به‌سزایی در کاهش هزینه‌های محاسباتی نیز خواهد داشت؛ چراکه حجم نمونه‌برداری و آزمایش‌ها را به k/n کاهش می‌دهد. ضمن اینکه سرعت گردش کار را در فاز عملی نیز به طرز چشمگیری افزایش می‌دهد. البته لازمه آنچه گفته شد، تضمین دقت در کنار کاهش ابعاد مسئله است.

نوآوری‌های این پژوهش به‌شرح زیر است:

۱- کاهش ویژگی به‌کمک الگوریتم شبیه‌سازی تبرید.
۲- یادگیری ویژگی به‌کمک الگوریتم برنامه‌نویسی ژنتیک، که ذاتاً یک الگوریتم برازش منحنی است.
۳- بهبود عملکرد الگوریتم برنامه‌نویسی ژنتیک با تغییر تابع برازش به تابعی بر اساس ماشین بردار پشتیبان خطی.
۴- استفاده از معیارهای ارزیابی جامع برای ارزیابی روش ارائه‌شده.

در ادامه، ابتدا در بخش دوم سوابق پژوهشی مورد بررسی قرار می‌گیرد. در بخش سوم توضیحاتی در مورد الگوریتم شبیه‌سازی تبرید، GP و پیش‌پردازش داده‌ها ارائه می‌شود. در بخش چهارم، روش

مورداستفاده ارزیابی شده و در انتها نتایج به‌دست‌آمده تجزیه‌وتحلیل شده و نتیجه‌گیری‌ها ارائه می‌شود.

۲- سابقه پژوهش

تلاش برای کاهش اثر بیش‌برازش با کمک یادگیری ویژگی که در [۳] انجام شده و همچنین استفاده از یادگیری ویژگی در بازشناسی افراد بر اساس تصاویر گرفته‌شده از زوایای مختلف در [۵] نیز توانسته تأثیر مثبت یادگیری ویژگی را در بهبود الگوریتم‌های یادگیری ماشین به اثبات برساند. عملکرد مناسب الگوریتم‌های بهینه‌سازی در حذف ویژگی‌های کم‌ارزش نیز در [۱۱-۶] به اثبات رسیده است.

الگوریتم‌های یادگیری ماشین کارایی خود را در حل مسائل پزشکی ثابت کرده‌اند [۱۳-۱۲]. برای نمونه در [۱۵-۱۴] تلاش شده تا به ترتیب برای ایجاد یک طبقه‌بند مولکولی و پیش‌بینی پاسخ به درمان ضدسرطان از الگوریتم GA استفاده شود. نتایج این پژوهش‌ها کارایی این الگوریتم را در مسائل مشابه به‌وضوح تأیید می‌کند. الگوریتم‌های درخت تصمیم و شبکه عصبی نیز مزایای خود را نسبت به روش آماری رگرسیون در [۱۶] نشان داده‌اند. در [۱۷] نیز GA در کاهش ویژگی‌ها خروجی بهتری نسبت به آزمون‌های آماری داشته است. همچنین در [۱۸] الگوریتم درخت تصمیم عملکردی بهتر را نسبت به SVM نشان داده است.

تشخیص سرطانی بودن یا نبودن بافت‌ها بر اساس بیان ژنی کاری بود که در [۱۹] ارائه شد. این روش توانسته بافت سرطانی و سالم را با دقت ۹۳٪ تفکیک می‌کند.کشف رابطه بین بیان ژنی و سرطان‌های تخمدان، پروستات و ریه در [۲۰] انجام شده و روش پیشنهادی در آن توانسته با دقتی بین ۹۴٪ تا ۹۸٪ عمل طبقه‌بندی را انجام دهد. همچنین در [۲۱] تلاشی برای ایجاد قواعد تشخیص بیماری سرطان ریه صورت گرفته و سه الگوریتم Association Rules، Rough Set و Genetic Algorithm بر روی داده‌ها اجرا شده است. طبق نتایج به‌دست‌آمده الگوریتم‌های GA، AR و RS توانسته‌اند به‌گونه‌ای مشابه با قواعد تشخیص پزشکی عمل کنند. مشکل این روش‌ها نیاز به ابزارهای پیشرفته ژنتیکی و مولکولی است که پروتئین‌های بیان‌شده را اندازه‌گیری کنند.

یکی از راه‌های پیش‌بینی ابتلا به سرطان سینه استفاده از ابزاری تحت عنوان ترموگرام است. ترموگرام یک ابزار مانیتورینگ است و قادر است ابتلا به بیماری سرطان سینه را تا ۱۰ سال پیش از وقوع آن پیش‌بینی کند. داده‌های مربوط به این ابزار نیازمند تفسیر است و اگر تحلیلگر به درستی داده‌ها را تفسیر نکند ارزش و توانایی این ابزار زیر سؤال می‌رود. به‌منظور مکانیزه کردن تفسیر داده‌های ترموگرام و برای کاهش خطای انسانی در تفسیر اطلاعات آن، طرحی تحقیقاتی و عملی در [۲۲] صورت گرفته است. در این کار عده‌ای از محققان سنگاپوری از الگوریتم‌های داده‌کاوی برای استخراج قواعد تصمیم‌گیری استفاده کردند. تا به‌کمک آن بتوان نتایج ترموگرام را تفسیر کرد. آن‌ها از CLFNN (Complementary Learning Fuzzy Neural Network) استفاده کردند. CLFNN الگوریتمی است قدرتمند و یکسری قواعد تصمیم‌گیری در اختیار می‌گذارد که به‌کمک آن به راحتی می‌توان داده‌های ابزار ترموگرام را با کم‌ترین خطا تحلیل کرد. در [۲۴-۲۳] نیز تلاش‌هایی برای تشخیص خوش‌خیمی سرطان سینه به‌کمک مجموعه‌داده WBCD صورت گرفته که روش‌های پیشنهادی توانسته عملکردی با دقت ٪۹۵/۵۶ تا ٪۹۵/۶۰ داشته باشد. در [۲۶-۲۵] نیز بیماری سرطان سینه به‌کمک روش‌های یادگیری ماشین و بر اساس داده‌های استخراج‌شده از تصاویر MRI با دقت ۹۶٪ تا ۹۸٪ تشخیص داده شده است. تلاش برای پیش‌بینی بقای بیماران مبتلا به این سرطان نیز در [۲۷]، دقتی برابر با ۸۳٪ را به دنبال داشته داشته است. عملکرد حاصل از یک روش پیشنهادی در [۲۸] بر روی مجموعه‌داده WDBC شامل ۳۰ ویژگی مربوط به سرطان سینه نیز توانسته عملکردی برابر با ٪۹۷/۳۸ را به دنبال داشته باشد. پژوهش دیگری که بر روی همین مجموعه‌داده صورت گرفته است، ترکیب الگوریتم‌های ژنتیک، ازدحام ذرات و کلونی مورچه‌ها با الگوریتم ماشین بردار پشتیبان است [۲۹].

۳- روش پیشنهادی

روند کلی روش پیشنهادی به این صورت است که ابتدا کیفیت داده‌ها، با جایگزینی مقادیر ازدست‌رفته، افزایش می‌یابد؛ پس از آن با بهره گرفتن از الگورتم بهینه‌سازی شبیه‌سازی تبرید، زیرمجموعه بهینه از ویژگی‌ها به‌گونه‌ای انتخاب می‌شود که رابطه۱ برقرار باشد.

$$\mathbf{SF} = \mathrm{SA}_{\mathbf{DATA}}, \quad |\mathbf{SF}| < |\mathbf{DATA}| \qquad (۱)$$

در این رابطه SF ماتریس مقادیر زیرمجموعه ویژگی‌ها، SA تابع انتخاب ویژگی و DATA ماتریس داده‌های اولیه هستند. در گام بعدی از روش پیشنهادی، به‌کمک GP، ویژگی‌های جدیدی بر اساس ویژگی‌های انتخاب‌شده ساخته می‌شود. این گام به‌صورتی انجام می‌شود که در رابطه۲ قابل ملاحظه است.

$$\mathbf{ND} = \mathrm{GP}_{\mathbf{SF}} \qquad (۲)$$

در این رابطه ND بیانگر مجموعه ویژگی‌هایی جدیدی است که به‌کمک GP و بر اساس SF ایجاد شده است. GP نیز بیانگر تابع نگاشت فضای داده‌ها به‌کمک برنامه‌نویسی ژنتیک است. درواقع الگوریتم GP ژن‌هایی تولید می‌کند که هر ژن بیانگر یک رابطه ریاضی بر روی یک سری از ویژگی‌های ورودی بوده و هر ژن می‌تواند با اعمال روابط ریاضی بر روی تعدادی از ویژگی‌ها یک ویژگی جدید ایجاد کند و در صورت ایجاد چند ژن به‌کمک GP می‌توان چند ویژگی جدید ایجاد کرد.

روند کلی آنچه گفته شد به‌صورتی است که در الگوریتم۱ دیده می‌شود. خطوط اول و دوم از الگوریتم۱ به ترتیب بهبود کیفیت داده و کاهش ابعاد مسئله را انجام می‌دهند. همان‌طور که از خط ۱۰ الگوریتم برمی‌آید، شرط خاتمه الگوریتم، بررسی تمام زیرمجموعه‌های ممکن از ویژگی‌های کاهش‌یافته و یا رسیدن به حدی مطلوب از برازش است. در خط ۶ از الگوریتم۱ به‌کمک GP، فضای مسئله تغییر می‌یابد. چراکه

داریم: GP: **SF** → **ND**. در این میان برازش‌های به‌دست‌آمده به ازای هر زیرمجموعه از ویژگی‌های ورودی به GP در بردار fit به طول n نگه‌داری می‌شود. این کار تا محقق شدن شرط خاتمه ادامه می‌یابد. در پایان، بهترین مجموعه داده ایجادشده، با بیش‌ترین برازش، به‌عنوان داده‌ی نهایی تعیین شده و مدل بر اساس آن ایجاد می‌شود. این مجموعه داده در یکی از عناصر $D_{1\dots n}$ قرار داشته که اندیس آن از طریق خط ۱۱ از الگوریتم استخراج می‌شود.

الگوریتم۱: روند کلی روش پیشنهادی

01. $Check\ data\ quality$
02. $Data_{n\times k} \leftarrow SAFR(Data_{n\times m}).\quad k < m$
03. $fitness \leftarrow 0$
04. $fit_{1\cdot\cdot n} \leftarrow 0$
05. do
06. $\quad D_i \leftarrow GP(f_{1\cdot\cdot i})$
07. $\quad fit_i \leftarrow SVM(D_i)$
08. $\quad fitness \leftarrow fit_i$
09. $\quad i = i + 1$
10. $While\ (fitness < Expected)\ or\ (i < |f|\,)$
11. $result \leftarrow \underset{j}{\text{argmax}}(fit_j)$
12. $model \leftarrow SVM(D_{result})$

در ادامه، هر یک از بخش‌های گفته‌شده در الگوریتم۱ به تفصیل تشریح می‌شود.

۳-۱- بالا بردن کیفیت داده

در الگوریتم‌های یادگیری ماشین فرض بر این است که داده‌ها فاقد نویز و همگی معتبر می‌باشند و هیچ نقصی در داده‌ها متصور نیست. این در حالی است که در دنیای واقعی داده‌ها خالی از نقص نیستند؛ بنابراین می‌توان بالا بردن کیفیت داده‌ها را جزئی جدایی‌ناپذیر از یک فرایند یادگیری ماشین موفق برشمرد. به همین منظور همان‌طور که در خط ۱ از الگوریتم۱ قابل ملاحظه است، ابتدا مقادیر ازدست‌رفته با مقادیری جایگزین شده است که امید می‌رود به مقدار واقعی و صحیح نزدیک باشد. برای این منظور داده‌های نامعلوم در هر ویژگی با شاخص پراکندگی مد جایگزین می‌شوند.

۳-۲- کاهش ویژگی مبتنی بر الگوریتم شبیه‌سازی تبرید

الگوریتم پیشنهادی کاهش ویژگی که معادل خط ۲ از الگوریتم۱ است و به اختصار SAFR[1] نامیده شده است، گامی ضروری در روش پیشنهادی است. همان‌طور که گفته شد، علت این ضرورت را می‌توان این‌گونه اظهار داشت که در خطوط ۵ تا ۱۰ از الگوریتم ۱، برای تغییر فضای مسئله و ایجاد بهترین مجموعه داده، امکان ساخت ژن‌هایی بر اساس تمامی ترکیبات ویژگی‌ها بررسی می‌شود؛ بنابراین با توجه به زمانبر بودن GP، این فاز به‌شدت زمان‌بر است؛ بنابراین باید ویژگی‌های بی‌ارزش تا حد امکان از چرخه کار کنار گذاشته شوند. الگوریتم‌های بهینه‌سازی سراسری برای این منظور با توجه به اینکه تنها بخشی از فضای مسئله را مورد بررسی قرار می‌دهند، گزینه مناسبی هستند. یکی از این الگوریتم‌ها الگوریتم بهینه‌سازی شبیه‌سازی تبرید است. این الگوریتم از مباحث متالورژی الهام گرفته شده است. روشی که از آن الهام گرفته شده اساساً برای سرد کردن تدریجی فلزات به‌منظور رسیدن به بهترین آرایش و کم‌ترین انرژی است. در شبیه‌سازی تبرید هر نقطه از فضای جستجو به‌صورت یک حالت از سیستم ترمودینامیکی در نظر گرفته می‌شود. هر حالت نیز ممکن است به نقاط همسایه خود تغییر حالت دهد. این تغییر حالت اگر در راستای کاهش انرژی باشد به‌طورقطع انجام خواهد شد؛ اما اگر در راستای افزایش انرژی باشد آنگاه این تغییر حالت به‌صورت احتمالی انجام خواهد شد و این احتمال به دو عامل بستگی دارد: عامل اول اختلاف انرژی نقطه جدید با نقطه فعلی و عامل دوم دمایی است که تغییر حالت در آن صورت می‌گیرد. هرچقدر حالت جدید، انرژی بیش‌تری داشته باشد، احتمال تغییر به آن کاهش می‌یابد و هرچه دما در آن لحظه بالا باشد نیز این احتمال افزایش می‌یابد. رابطه ریاضی به‌کاررفته برای محاسبه احتمال بر اساس این دو پارامتر، می‌تواند تابع توزیع احتمال بولتزمن باشد (رابطه۳).

$$e^{\left(\frac{-\Delta E}{T}\right)} \qquad (۳)$$

در این رابطه ΔE اختلاف انرژی یا همان اختلاف هزینه است که از طریق رابطه۴ محاسبه می‌شود. پارامتر T نیز بیانگر دما است.

$$\Delta E = C_n - C_c \qquad (۴)$$

در این رابطه C_c همان هزینه پاسخ فعلی و C_n هزینه یک پاسخ جدید در الگوریتم شبیه‌سازی تبرید هستند.

در الگوریتم شبیه‌سازی تبرید، که روند کلی آن در الگوریتم۲ آمده است، در هر دما (گام) تعدادی نقاط همسایگی در اطراف پاسخ فعلی به‌صورت تصادفی ایجاد شده و پاسخ فعلی در راستای کاهش انرژی در بین این نقاط تغییر حالت می‌دهد. بررسی همسایگی‌ها در هر دما به تعداد مشخصی انجام شده و پس از این کار، دما کاهش می‌یابد و بررسی و یافتن همسایگی با انرژی کم‌تر برای دمای جدید انجام می‌شود. کاهش دما طبق رابطه۵ انجام می‌شود.

$$T_{i+1} = \alpha \times T_i\ .\ 0 < \alpha < 1 . T_0 \neq 0 \qquad (۵)$$

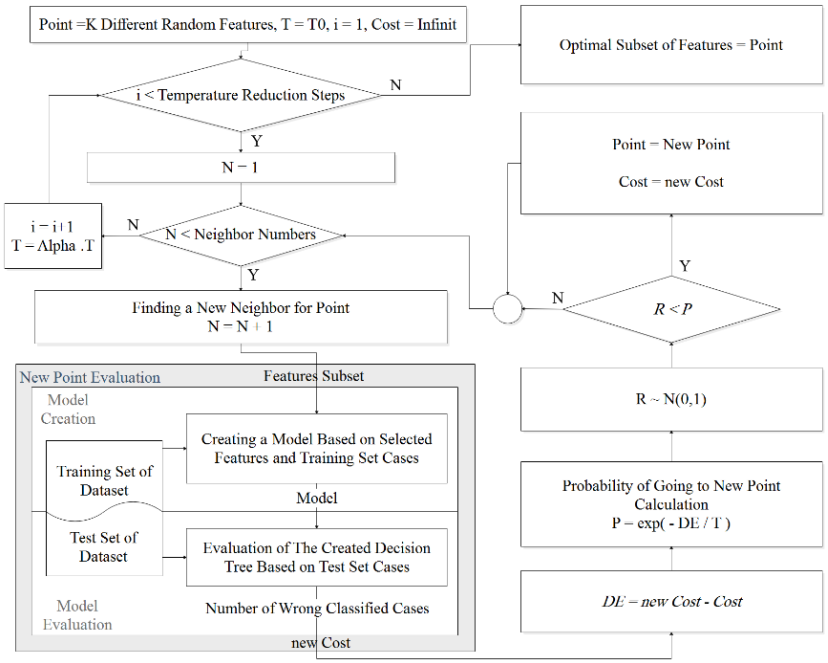
نکته مهم، این است که در طول اجرای الگوریتم و با کاهش دما، به مرور امکان تغییر به حالات با انرژی بیش‌تر، کاهش می‌یابد؛ درحالی که این احتمال در ابتدای الگوریتم بالا است. این امر سبب می‌شود که در ابتدای الگوریتم، خاصیت جستجو یا (Exploration) بالا بوده و به مرور این خاصیت کاهش یافته و خاصیت بهره‌برداری (Exploitation) افزایش یابد و این سبب همگرایی تدریجی به‌سمت نقطه بهینه خواهدشد. نمایی کلی از روش کار الگوریتم شبیه‌سازی تبرید در شکل۱ قابل مشاهده است.

برای کاهش ویژگی به‌کمک الگوریتم شبیه‌سازی تبرید، یک زیرمجموعه k عضوی از بین ویژگی‌های موجود به‌گونه‌ای انتخاب می‌شود که کم‌ترین میزان طبقه‌بندی غلط را در اجرای درخت تصمیم بر روی

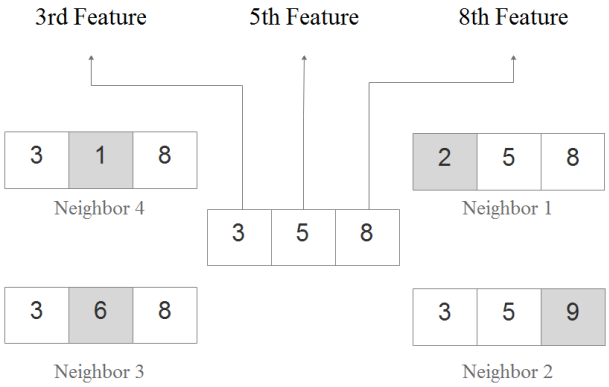
داده‌های آزمایشی داشته باشد. مدل‌سازی نقاط مربوط به الگوریتم شبیه‌سازی تبرید در مسئله به این صورت است که هر مجموعه k عضوی از ویژگی‌ها به‌عنوان یک نقطه از فضای جستجو در نظر گرفته می‌شود و نقطه همسایه به نقطه‌ای اطلاق می‌شود که در یکی از ویژگی‌ها با نقطه جاری متفاوت باشد. نمونه‌ای از حالت k=۳ در شکل۲ آمده است.

الگوریتم۲: کاهش ویژگی با الگوریتم شبیه‌سازی تبرید

$T \leftarrow t_0$
$\alpha \leftarrow \alpha_0$
$Temp \leftarrow RandPerm(n.n)$
$Train \leftarrow Temp_{1\ldots.2n/3}$
$Test \leftarrow Temp_{2n/3+1\ldots n}$
$for\ i = 1\ to\ TempRedSteps$
    $FSS \leftarrow RandPerm(m.k)$
    $TrainData \leftarrow Data_{Train\ .\ FSS}$
    $TestData \leftarrow Data_{Test\ .FSS}$
    $\text{model} \leftarrow DecisionTree(TrainData)$
    $CurrentCost \leftarrow Eval(model.TestData)$
    $for\ j = 1\ to\ Neighbors$
        $new \leftarrow newNeighbor(FSS)$
        $TrainData \leftarrow Data_{Train\ .\ new}$
        $TestData \leftarrow Data_{Test\ .new}$
        $\text{model} \leftarrow DecisionTree(TrainData)$
        $Cost \leftarrow Eval(model.TestData)$
        $\Delta E \leftarrow Cost - CurrentCost$
        $P \leftarrow e^{(-\Delta E/T)}$
        $n \leftarrow RandomN(0.1)$
        $if\ n \leq P$
        $FSS \leftarrow new$
        $CurrentCost \leftarrow Cost$
    End
    $T \leftarrow \alpha.T$
End
$return\ FSS$



شکل۱: روند کلی کاهش ویژگی با الگوریتم شبیه‌سازی تبرید

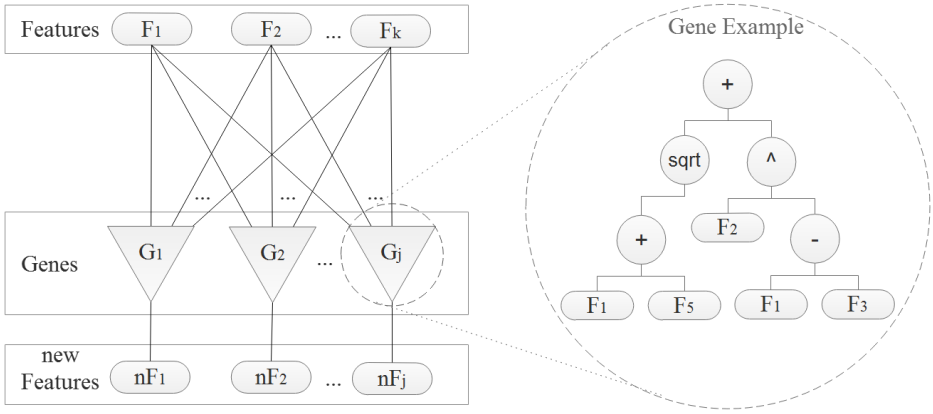


شکل۲: مدل‌سازی نقاط همسایگی در مسئله کاهش ویژگی با شبیه‌سازی تبرید

مدل‌سازی تابع هزینه نیز به این صورت است که در هر مرحله برای محاسبه هزینه‌ی مربوط به هر پاسخ الگوریتم شبیه‌سازی تبرید (که مشخص‌کننده k ویژگی است)، یک درخت تصمیم بر اساس زیرمجموعه‌ای از داده‌ها، شامل همان k ویژگی ایجاد می‌شود. مدل مذکور بر اساس درصدی از نمونه‌ها تحت عنوان داده آموزشی ایجاد می‌شود. پس از آن درخت ایجادشده بر روی بخشی از نمونه‌ها، تحت عنوان داده آزمایشی، مورد ارزیابی قرار می‌گیرد. در انتها درصد موارد طبقه‌بندی غلط در این ارزیابی به‌عنوان هزینه‌ی مربوط به پاسخ پیشنهادی الگوریتم شبیه‌سازی تبرید استفاده می‌شود. خطوط ۱۴ تا ۱۷ از الگوریتم۲ روند محاسبه هزینه را نشان می‌دهد.

۳-۳- الگوریتم یادگیری ویژگی مبتنی بر برنامه‌نویسی ژنتیک

در بخش سوم که معادل خطوط ۵ تا ۱۰ از الگوریتم۱ است، برای یادگیری ویژگی‌های جدید، یک الگوریتم برنامه‌نویسی ژنتیک ارائه شده است که فضای مسئله را از k به J تغییر می‌دهد. برای این منظور از تعداد J ژن برای یادگیری J ویژگی جدید استفاده می‌شود. روند کاری GP ارائه‌شده در شکل۳ قابل مشاهده است. درواقع اگر مجموعه ویژگی‌های اولیه به‌صورت $\vec{F} = (F_1.F_2.\ldots.F_k)$ در نظر گرفته شوند، عملکرد GP به‌صورت یک تابع GP، $\overrightarrow{nF} = GP(\vec{F})$ پیاده‌سازی شده است؛ که در آن nF بیانگر مجموعه ویژگی‌های جدید است. در انتها، در یک کار ابتکاری میزان برازش ژن‌ها با استفاده از الگوریتم SVM و بر اساس داده‌هایی انجام می‌شود که دربردارنده مقادیر ویژگی‌های جدید (nF) هستند.



شکل۳: یادگیری ویژگی با برنامه‌نویسی ژنتیک

۴- ارزیابی

به‌منظور ارزیابی، روش پیشنهادی بر روی مجموعه داده‌های استاندارد، آزمایش شده و نتایج حاصل از آن با کارهای مشابه مقایسه شد. بدین منظور در زیربخش بعدی ابتدا مجموعه‌داده مورداستفاده تشریح شده و سپس نتایج حاصل از مقایسه ارائه می‌شود. در پیاده‌سازی روش پیشنهادی از یک سیستم با پردازنده دو هسته‌ای با توان پردازشی ۲/۱ گیگاهرتز استفاده شد که دارای ۱ گیگابایت حافظه اصلی است. پارامترهای اجرای الگوریتم شبیه‌سازی تبرید و GP نیز به ترتیب به‌صورتی تنظیم شده است که به ترتیب در جداول ۱و ۲ دیده می‌شود.

پارامترهای ارزیابی نیز شامل صحت، دقت، بازیافت و F_1 هستند که تعاریف و روش محاسبه هرکدام در ادامه بیان خواهد شد.

صحت: نسبت برچسب‌های درست به کل برچسب‌های زده‌شده توسط مدل؛ یعنی:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (۶)$$

دقت: نسبت برچسب‌های مثبت درست تشخیص داده‌شده به کل برچسب‌های مثبت تشخیص داده‌شده توسط مدل؛ یعنی:

$$Pr = \frac{TP}{TP+FP} \quad (۷)$$

بازیافت: نسبت تعداد موارد مثبت درست تشخیص داده‌شده به کل موارد مثبت؛ یعنی:

$$Re = \frac{TP}{TP+FN} \quad (۸)$$

معیار F1: ترکیبی است از معیارهای دقت و بازیافت که طبق رابطه x قابل محاسبه است. این معیار درواقع میانگین هارمونیک پارامترهای دقت و بازیابی است؛ یعنی:

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re} \qquad (۹)$$

جدول۱: پارامترهای الگوریتم بهینه‌سازی شبیه‌سازی تبرید

پارامتر	مقدار
بیشترین تعداد تکرار	۲۰
بیشترین تعداد همسایگی	۵
نرخ کاهش دما	۰/۰۱
دمای اولیه	۱

جدول۲: پارامترهای اولیه برنامه‌نویسی ژنتیک

پارامتر	مقدار
اندازه جمعیت	۱۰۰
تعداد نسل‌ها	۱۰۰
اندازه رقابت	۳
بیشینه عمق درخت	۵
بیشینه گره‌های درخت	بی‌نهایت
بیشینه تعداد ژن‌ها	۳
بازه اعداد ثابت	[۱۰ و۱۰-]

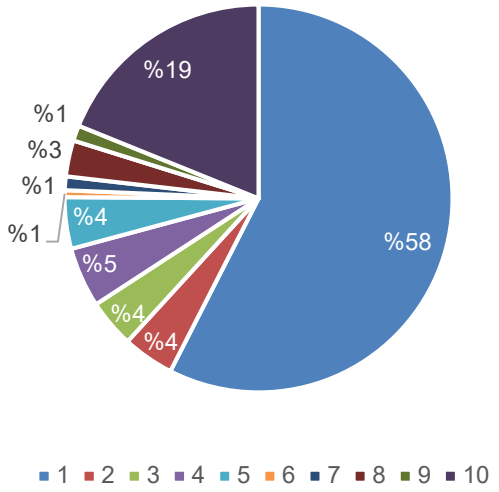
جدول۳: مقایسه تأثیر وجود فازهای مختلف در روش پیشنهادی در مجموعه‌داده WDBC

	Accuracy (%)	Feature space dimension	Time (s)
SVM	۹۷/۴۷	۳۰	۵/۷۶
SAFR+SVM	۹۷/۷۷	۱۳	۳۲۰۵/۵۶
GP+SVM	۹۷/۸۰	۸	۱۹۲۸۵/۶
SAFR+GP+SVM	۹۷/۸۹	۶	۲۲۴۹۸/۳۲

۴-۱- مجموعه‌داده

داده‌های به‌کاررفته در این مقاله شامل مجموعه‌داده استاندارد WBCD حاوی ۹ ویژگی از ۶۹۹ نمونه و مجموعه‌داده WDBC حاوی ۳۰ ویژگی از ۵۶۹ نمونه است. ارزیابی نتایج نیز بر روی هر دو مجموعه‌داده به‌صورت 10-Fold Cross Validation انجام شده است.

در مجموعه‌داده WBCD، ۱۶ مورد مقادیر ازدست‌رفته (Missing Value) وجود دارد و تمامی این ۱۶ مورد مربوط به ویژگی برهنگی هسته (bare nuclei) است؛ لذا حذف نمونه‌های دارای مقادیر ازدست‌رفته نمی‌تواند راه‌حل مناسبی باشد. همان‌طور که در شکل۴ مشاهده می‌شود، نمودار توزیع مقادیر این ویژگی، مقادیری گسسته بین ۱ تا ۱۰ را نشان می‌دهد؛ بنابراین استفاده از شاخص پراکندگی مرکزی میانگین برای جایگزینی این مقادیر، انتخاب مناسبی نیست. با توجه به اینکه بیش از نیمی از مقادیر مربوط به ویژگی مذکور دارای مقدار عددی یک هستند واضح است که بهترین انتخاب برای مقادیر ازدست‌رفته شاخص پراکندگی مد است. بر همین اساس، مقادیر ازدست‌رفته برای ویژگی مذکور با مقدار شاخص پراکندگی مد مربوط به همین ویژگی جایگزین شدند.



شکل۴: نمودار توزیع مقادیر در ویژگی برهنگی هسته در مجموعه داده WBCD

۴-۲- نتایج و مقایسه

۴-۲-۱- نتایج تأثیر مراحل مختلف روش پیشنهادی

مقایسه تأثیر وجود و یا عدم وجود هر فاز بر دقت روش پیشنهادی می‌تواند ضرورت وجود فازهای مختلف را به اثبات برساند. این امر در جدول۳ قابل مشاهده است.

همان‌طور که در جدول۳ دیده می‌شود، کاهش ویژگی به تنهایی تعداد ویژگی‌ها را به میزان چشمگیری کاهش می‌دهد و نکته جالب‌توجه این است که کاهش ویژگی صورت‌گرفته همراه با افزایش دقت است. GP نیز به تنهایی ضمن کاهش چشمگیر ویژگی‌ها افزایش دقت را به دنبال دارد؛ اما این الگوریتم به‌شدت زمان‌بر است. ترکیب کاهش ویژگی و GP نیز کاهش بیش‌ازپیش ویژگی‌ها را ضمن افزایش دقت به دنبال دارد. البته این امکان متصور است که GP به تنهایی توان رسیدن به دقت روش ترکیبی پیشنهادی را با همان تعداد ویژگی‌ها داشته باشد؛ اما این امر مستلزم سپری شدن زمان بیش‌تری خواهد بود و این زمان بیش‌تر، در اجراهای مختلف متغیر بوده و می‌تواند در مواردی بسیار طولانی باشد. این درحالی است که استفاده از فاز کاهش ویژگی زمان مورد نیاز را برای رسیدن به پاسخ مطلوب، در حد تقریباً ثابتی حفظ خواهد کرد و این زمان ثابت، نسبت به زمان مورد نیاز برای ادامه GP عموماً کم‌تر است.

۴-۲-۲- تأثیر استفاده از الگوریتم‌های مختلف در فاز نهایی

استفاده از الگوریتم‌های مختلف در فاز اصلی یادگیری ماشین نیز، که فاز نهایی روش پیشنهادی است، در جدول۴ آمده است. در این جدول نتیجه

حاصل از اجرای الگوریتم‌های مرسوم یادگیری ماشین به‌عنوان فاز نهایی در روش پیشنهادی به نمایش درآمده است. این امر استفاده از الگوریتم SVM را در روش پیشنهادی به‌وضوح توجیه می‌کند.

جدول۴: مقایسه الگوریتم‌های مرسوم یادگیری به‌عنوان فاز نهایی روش پیشنهادی در مجموعه‌داده WDBC

	Accuracy (%)	Time (ms)
Decision Tree	۹۶/۹۶	۳۷۳
KNN	۹۶/۷۰	۱۹۰
Neural Networks	۹۵/۱۴	۱۱۰۲۱
SVM	۹۷/۸۹	۵۶۰

۴-۲-۳- مقایسه نتایج با کارهای مشابه

مقایسه این نتایج با کارهایی که بر روی مجموعه‌داده WBCD در [۲۳] و [۲۴] صورت پذیرفته است افزایش دقت را ضمن کاهش چشمگیر ویژگی‌ها نشان می‌دهد. به‌گونه‌ای که تنها با استفاده از دو ویژگی، عملکرد مذکور را به دست آورده است. مقایسه نتایج حاصل از روش پیشنهادی با کارهای انجام‌شده بر روی مجموعه‌داده WBCD در جدول۵ قابل مشاهده است.

جدول۵: مقایسه روش پیشنهادی با سایر روش‌های انجام‌شده در مجموعه‌داده WBCD

	Methods	Accuracy (%)
Karabatak [23]	AR+ ANN	۹۵/۶۰
Subashini [24]	SVM+RBFNN	۹۵/۵۶
Proposed	SAFR+GP+ SVM	۹۵/۸۵

در [۲۳] که بر روی مجموعه‌داده WBCD صورت گرفته است. ابتدا به‌کمک AR ابعاد مجموعه‌داده کاهش یافته است؛ به‌صورتی که تعداد ویژگی‌ها از ۹ ویژگی به ۴ ویژگی رسیده است. در ادامه، مجموعه‌داده حاصل از کاهش ویژگی به الگوریتم شبکه عصبی داده شد و نتایج به روش 3-Fold Cross Validation ارزیابی شده است. نتایج نشان می‌دهد روش پیشنهادی آن‌ها، عمل طبقه‌بندی را با دقت٪۹۵/۶ انجام داده است. در [۲۴] نیز که بر روی همین مجموعه‌داده، شامل ۹ ویژگی، صورت گرفته است؛ نتایج بیانگر این حقیقت است که RBFNN و SVM در مسئله مذکور به ترتیب دقتی برابر با ٪۹۵/۵۶ و ٪۹۲/۱۳ داشته‌اند.

روش ارائه‌شده در [۲۸] به‌منظور افزایش دقت طبقه‌بندی، بر روی مجموعه داده WDBC مورد ارزیابی قرار گرفته است. بر همین اساس، روش پیشنهادی مقاله پیش رو با [۲۸] مقایسه شده است. اعمال روش پیشنهادی بر روی داده‌های WDBC، شامل ۳۰ ویژگی مربوط به سرطان سینه، و ارزیابی نتایج به‌صورت 10-Fold Cross Validation منجر به تشخیص بدخیمی و خوش‌خیمی با دقت ٪۹۷/۸۹ شده است که در مقایسه با روش پیشنهادی در [۲۸] دقت بالاتری است و این افزایش دقت، ضمن کاهش ویژگی‌ها دقیقاً به اندازه کاری است که در [۲۸] صورت گرفته است. در [۲۸] عمل طبقه‌بندی با استفاده از ۶ ویژگی انجام شده است. روش ابداعی در [۲۸] به این شکل بوده است که ابتدا به‌کمک الگوریتم K-means یک سری الگوی پنهان در بین داده‌ها یافته شده و به‌عنوان یک سری ویژگی جدید به مجموعه‌داده اضافه شده است. پس از آن مجموعه‌داده حاصل، به‌عنوان داده‌ی یادگیری به الگوریتم SVM داده شده و نتایج نیز به روش 10-Fold Cross Validation مورد ارزیابی قرار گرفته است. نتایج گویای این حقیقت است که این روش ابداعی توانسته دقتی ٪۹۷/۳۸ را به دنبال داشته باشد. مقایسه نتایج حاصل از روش پیشنهادی مقاله پیش رو بر روی مجموعه‌داده WDBC با آنچه که بر روی همین مجموعه‌داده در [۲۸] و [۲۹] انجام شده است در جدول۶ قابل مشاهده است.

جدول۶: مقایسه روش پیشنهادی با سایر روش‌های انجام‌شده بر روی مجموعه‌داده WDBC

	Feature space dimension	Accuracy (%)
ACO-SVM [29]	۱۵	۹۵/۹۶
GA-SVM [29]	۱۸	۹۷/۱۹
PSO-SVM [29]	۱۷	۹۷/۳۷
K-SVM [28]	۶	۹۷/۳۸
Proposed	۶	۹۷/۸۹

۴-۲-۴- ارزیابی کلی نتایج روش پیشنهادی

نتایج ارزیابی روش پیشنهادی بر روی مجموعه‌داده WDBC در جدول۷ قابل مشاهده است.

جدول۷: ارزیابی آزمایش روش پیشنهادی روی مجموعه‌داده WDBC

	Acc	Re	Pr	F1
Benign	۹۵/۸۵	۹۸/۷۰	۹۷/۴۴	۹۸/۰۶
Malignant	۹۵/۸۵	۹۶/۹۲	۹۸/۴۳	۹۷/۶۷

نتایج حاصل از اعمال روش پیشنهادی بر روی مجموعه‌داده WBCD شامل ۹ ویژگی که مرتبط با بیماری سرطان سینه است نیز در جدول۸ قابل ملاحظه است.

جدول۸: ارزیابی آزمایش روش پیشنهادی روی مجموعه‌داده WBCD

	Acc	Re	Pr	F1
Benign	۹۵/۸۵	۹۶/۵۱	۹۷/۱۴	۹۶/۸۲
Malignant	۹۵/۸۵	۹۴/۶۱	۹۳/۴۴	۹۴/۰۲

۵- بحث

فلسفه روش پیشنهادی بر دو پایه استوار است. اول اینکه بالا بودن تعداد ویژگی‌ها نه‌تنها لزوماً منجر به دقت بالا نمی‌شود؛ بلکه در برخی موارد منجر به افت دقت نیز خواهد شد؛ بنابراین کاهش ویژگی در چنین

مواردی می‌تواند دقت را نیز افزایش دهد. دلیل این امر، وجود ویژگی‌های بی فایده در کنار ویژگی‌های مخرب است [۳۰]؛ بنابراین کاهش ویژگی می‌تواند با حذف چنین ویژگی‌هایی افزایش دقت طبقه‌بندی را به دنبال داشته باشد. دوم این که عملیات یادگیری ویژگی بر روی داده‌های اولیه و خام، می‌تواند منجر به ایجاد ویژگی‌های سطح بالا و متمایزکننده جدید شود؛ به‌گونه‌ای که ویژگی‌های خروجی الگوریتم یادگیری ویژگی، دقت بالاتری برای طبقه‌بندی داشته باشد. استفاده از الگوریتم GP نشان داد که این الگوریتم توانایی بالایی در یادگیری ویژگی دارد. گرچه در بطن این الگوریتم، عمل کاهش ویژگی نیز نهفته است؛ اما کاهش ویژگی اولیه (که در این پژوهش با الگوریتم بهینه‌سازی شبیه‌سازی تبرید انجام شده است) می‌تواند منجر به تسریع و بهبود الگوریتم GP شود.

نکته قابل ذکر، کاربرد روش ارائه‌شده در امور بالینی است. مجموعه‌داده‌های استفاده‌شده در این پژوهش، توسط مراکز بالینی جمع‌آوری شده است و طبقه‌بندی خوش‌خیمی و بدخیمی نیز توسط متخصصین مربوطه انجام شده است. از آن‌جایی که روش ارائه‌شده یک روش یادگیری تحت نظارت است، نتایج به‌دست‌آمده نشان می‌دهد که مدل حاصل از روش ارائه‌شده، توانسته با دقت بالایی به عملکرد متخصصین نزدیک شود.

نکته دیگری که در این پژوهش باید به درستی مورد بحث قرار گیرد، بالا بودن زمان یادگیری است. باید توجه داشت که این زمان، صرف یادگیری در فاز آفلاین خواهد شد؛ لذا در صورت استفاده عملی، تفاوتی در سرعت سیستم تشخیصی مبتنی بر روش پیشنهادی، نسبت به سایر روش‌ها وجود نخواهد داشت.

۶- نتیجه‌گیری

بر کسی پوشیده نیست که GP الگوریتمی مناسب برای برازش منحنی و رگرسیون است اما همان‌طور که در این مقاله اثبات شد، الگوریتم مذکور قابلیت استفاده در مسائل یادگیری ویژگی را نیز به‌خوبی دارد. اگرچه در این مقاله ابتدا تعداد ویژگی‌ها به‌کمک الگوریتم شبیه‌سازی تبرید کاهش یافت و این امر بیانگر قدرت الگوریتم‌های فراابتکاری و به‌طور خاص الگوریتم شبیه‌سازی تبرید در کاهش ویژگی است؛ اما نکته جالب‌توجه در این پژوهش این است که ژن‌های ایجادشده توسط الگوریتم GP تنها شامل تعدادی از ویژگی‌های کاهش‌یافته است و این حکایت از این حقیقت دارد که الگوریتم GP، عمل کاهش ویژگی را نیز به‌صورت ضمنی انجام می‌دهد و این به معنی کاهش بیش‌ازپیش ابعاد مسئله است. البته در صورت حذف فاز کاهش ویژگی، رسیدن به نتیجه مطلوب با استفاده از الگوریتم GP زمان بیش‌تری را می‌طلبد. با توجه به زمان‌بر بودن GP، فاز کاهش ویژگی جزئی لازم و جدایی‌ناپذیر در روش پیشنهادی است. نتایج این مقاله بیانگر این حقیقت است که الگوریتم GP می‌تواند به‌منظور یادگیری ویژگی در مسائلی به کار برده شود که تعداد ویژگی‌های اولیه بسیار زیادی دارند.

مراجع

[1] "Institute for Research, Education and Treatment of Cancer", www.ncii.ir, [Online Access: Feb. 22, 2015].

[2] "Breastcancer.org", www.breastcancer.org, [Online Access: Feb. 22, 2015].

[3] S. Kim, Y. Choi, M. Lee, "Deep learning with support vector data description," *Neurocomputing*, vol. 165, no. 1, pp. 111-117, Oct. 2015.

[4] A.H. Gandomi, D. Mohammadzadeh, J.L. Pérez-Ordóˆ nez, A.H. Alavi, "Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups," *Applied Soft Computing*, vol. 19, no. 1, pp. 112-120, Feb. 2014.

[5] Sh. Ding, L. Lin, G. Wang, H. Chao, " Deep Feature Learning with Relative Distance Comparison for Person Re-identification," *Pattern Recognition*, vol. 48, no. 1,pp. 2993-3003, Oct. 2015.

[6] E. Emary, H.M. Zawbaa, A.E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, no. 1, pp. 371-381, Jan. 2016.

[7] I. Zyout, J. Czajkowska, M. Grzegorzek, "Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 95-107, Feb. 2015.

[8] H. Min, R. Weijie, "Global mutual information-based feature selection approach using single-objective and multi objective optimization," Neurocomputing, vol. 168, pp. 47-54, Nov. 2015.

[9] K. Shima, E. Mahdi, " Feature selection using multimodal optimization techniques," *Neurocomputing*, vol. 171, pp. 586-597, Jan. 2015.

[10] M. Parham, R. Mehrdad, "Integration of graph clustering with ant colony optimization for feature selection," *Knowledge-Based Systems*, vol. 84, pp. 144-161, Aug. 2015.

[11] N.K. Sreeja, A. Sankar, "Pattern Matching based Classification using Ant Colony Optimization based Feature Selection," *Applied Soft Computing*, vol. 31, pp. 91-102, Jun. 2015.

[12] A.T. Mohammad Reza, H. Seyyed Abed, N.S. Mohammad Bagher, "Evaluation of Visual Selective Attention by Event Related Potential Analysis in Brain Activity," *Tabriz Journal of Electrical Eng*, vol. 46, no. 1, pp. 13-24, 2016.

[13] B. Morteza, P. Hosein, "Epilepsy Detection Based on Optimization of Fused Hartley Transform Feature with Hybrid Model of MLP and GA using Memetic Learning Strategy," *Tabriz Journal of Electrical Eng*, vol. 45, no. 4, pp. 51-67, 2015.

[14] J. Yu, J. Yu, A.A. Almal, S.M. Dhanasekaran, D.Ghosh, W.P. Worzelz, A.M. Chinnaiyan, "Feature Selection and Molecular Classification of Cancer Using Genetic Programming," *Neoplasia*, vol. 9, no. 4, pp. 292-303, April. 2007

[15] F. Archetti, I. Giordani, L. Vanneschi, "Genetic programming for anticancer therapeutic response prediction using the NCI-60 dataset," *Computers & Operations Research*, vol. 37, no. 1, pp. 1395-1405, Mar. 2009.

[16] G. Phillips-Wren, P. Sharkey, S. Morss Dy, "Mining lung cancer patient data to assess healthcare resource utilization," *Expert Systems with Applications*, vol. 35, no. 1, pp. 1611-1619, Aug. 2008.

[17] L. Li, H. Tang, Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, R.A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," *Artificial Intelligence In Medicin*, vol. 32, no. 1, pp. 71-83, Mar. 2004.

[18] S.C. Shah, A. Kusiak, M.A. O'Donnell, "Patient-recognition data-mining model for BCG plus interferon immunotherapy bladder cancer treatment," *Computers in Biology and Medicine*, vol. 36, no. 1, pp. 634-655, Mar. 2005.

[19] K.R. Seeja, "Feature selection based on closed frequent itemset mining: A case study on SAGE data classification," *Neurocomputing*, vol. 151, pp. 1027-1032, Mar. 2015.

[20] S. Shah, . Kusiak, "Cancer gene search with data-mining and genetic algorithms," *Computers in Biology and Medicine*, vol. 37, no. 1, pp. 251-261, Jan. 2007.

[21] Y. Qiang, Y. Guo, X. Li, Q. Wang, H. Chen, D. Cuic, "The Diagnostic Rules of Peripheral Lung Cancer PreliminaryStudy

Based on Data Mining Technique," *Journal of Nanjing Medical University*, vol. 21, no. 3, pp. 190195, Nov. 2006.

[22] T.Z. Tan, C. Quek, G.S. Ng, E.Y.K. Ng, "A novel cognitive interpretation of breast cancer thermography with complementary learning fuzzy neural memory structure," *Expert Systems with Applications*, vol. 33, no. 1, pp. 652-666, Jun. 2007.

[23] M. Karabatak, M.C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Systems with Applications*, vol. 36, no. 1, pp. 3465-3469, Feb. 2009.

[24] T.S. Subashini, V. Ramalingam, S. Palanivel, "Breast mass classification based on cytological patterns using RBFNN and SVM," *Expert Systems with Applications*, vol. 36, no. 1, pp. 5284-5290, Jun. 2009.

[25] M. Kowa, P. Filipczuk, A. Obuchowicz, J. Korbicz, R. Monczak, "Computer-aided diagnosis of breast cancer based onfine needle biopsy microscopic images," *Computers in Biology and Medicine*, vol. 43, no. 1, pp. 1563-1572, Aug. 2013.

[26] A.E.Hassaniena, H.M. Moftahb,A.T. Azar, M. Shoman, "MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier," *Applied Soft Computing*, vol. 14, no. 1, pp. 62-71, Aug. 2013.

[27] P.J. García-Laencina, P.H. Abreu, M.H. Abreu, N. Afonoso, "Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values," *Computers in Biology and Medicine*, vol. 59, no. 1, pp. 125-133, Feb. 2015.

[28] B. Zheng, S.W. Yoon, S.S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Systems with Applications*, vol. 41, no. 1, pp. 1476-1482, March. 2014.

[29] Prasad. Y, Biswas. k, Jain. c, "Svm classifier based feature selection using ga, aco and pso for sirna design," *In Proceedings of the first international conference on advances in swarm intelligence*, pp. 307-314, Beijing, China, June. 2010.

[30] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh et all, " Application of high-dimensional feature selection: evaluation for genomic prediction in man," *Scientific Reports*, vol. 5, pp. 1-12, May. 2015.

زیرنویس‌ها

۱ Simulated Annealing Feature Reduction