

چکیده

امروزه میکروآرایه‌ها به عنوان یکی از پرکاربردترین ابزارهای مورد استفاده در حل مسائل زیستی شناخته شده‌اند. با استفاده از این ابزار قدرتمند امکان تحلیل و شناخت بهتر رویدادهای زیستی پیچیده‌ای مانند سرطان فراهم شده است. اما ابعاد بالای داده‌های میکروآرایه و ماهیت نویزی و مبهم آن موجب ناکارآمدی الگوریتم‌های پردازشگر و در نتیجه ارائه نتایج ناسازگار در مطالعات مختلف شده است. ناپایداری عملکرد الگوریتم‌ها در مجموعه داده‌های مستقل و معرفی مجموعه ژن‌های شاخص متنوع به منظور پیشگویی بازگشت سرطان سینه در سالیان اخیر یکی از نمونه‌های این مشکل است که باعث بی اعتباری تفسیر شاخص‌های زیستی قبلی شده است. یکی از عوامل مهم این ناپایداری که کمتر مورد توجه قرار گرفته الگوریتم‌های مورد استفاده در تحلیل این داده‌های ویژه است، که در این پایان نامه مورد بررسی قرار گرفته است.

در این پایان نامه سه روش جدید در راستای استخراج ژن‌های شاخص مناسب به منظور پیشگویی بازگشت سرطان سینه ارائه شده است. ابتدا سعی شده تا با انتخاب مناسب از بین الگوریتم‌های موجود برای انتخاب ویژگی و طبقه‌بندی، ژن‌های شاخص با صحت پیشگویی بالا و پایداری مناسب در مجموعه داده‌های مستقل استخراج شوند. در روش اول از ترکیب روش استخراج ویژگی مجموعه راف و مدل طبقه‌بند فازی-عصبی برای مقابله با ماهیت نویزی و مبهم داده‌های میکروآرایه استفاده شده است. در روش دوم یک راه کار جدید به منظور استفاده از اطلاعات شبکه‌های تعامل پروتئینی در ترکیب توام با روش‌های انتخاب ویژگی تک متغیره و جاسازی شده به کار گرفته شده است. با استفاده از این روش صحت طبقه‌بندی نمونه‌ها به نحو مطلوبی افزایش یافت. روش آخر یک تابع امتیاز دهی جدید در حوزه اطلاعات متقابل برای رتبه‌بندی ژن‌ها در چارچوب روش‌های انتخاب ویژگی چند متغیره است. در این تابع امتیاز دهی با کمک اطلاعات متقابل توام وزن بیشتری به وابستگی بین ژن‌ها در جهت هدف پیشگویی اختصاص یافته و همچنین کاهش افزونگی اطلاعات با کمک گرفتن از شبکه تعامل پروتئینی انجام گرفته است.

ژن‌های شاخص انتخاب شده توسط روش‌های پیشنهادی از هفت مجموعه داده میکروآرایه، شامل ۱۵۳۸ نمونه تومور سرطان سینه، از نظر شاخص پایداری و صحت طبقه‌بندی مورد ارزیابی قرار گرفته‌اند. ۵۸ تا ۹۲ درصد همپوشانی بین مجموعه ژن‌های شاخص انتخاب شده از هفت مجموعه داده و نیز صحت طبقه‌بندی مناسب در مقایسه با سایر روش‌های مشابه نشان دهنده عملکرد بسیار مناسب روش سوم پیشنهاد شده است. با تکیه بر پایداری مجموعه نهایی ژن‌های معرفی شده و همچنین انطباق آنها با ژن‌های شناخته شده دخیل در سرطان می‌توان آنها را با قابلیت اطمینان بالا به عنوان شاخص‌های زیستی عمومی مربوط به بازگشت سرطان سینه معرفی کرد.

کلید واژه‌ها

انتخاب ژن، اطلاعات متقابل، بازگشت سرطان سینه، پایداری ژن‌های شاخص، شبکه‌های تعامل پروتئینی