



طبقه بندی سرطان مبتنی بر بیان ژن از طریق انتخاب ویژگی با طبقه بندی KNN و SVM

سمیه بلوچ محمدرضاخانی*^۱، مسلم بحرینی نژاد^۲

۱- دانشجوی کارشناسی ارشد مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر - موسسه آموزش عالی بهمنیار کرمان-ایران
Somaye.balloch1991@gmail.com

۲- دانشجوی کارشناسی ارشد مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر - موسسه آموزش عالی بهمنیار کرمان-ایران
m.b.nezhad1020@gmail.com

چکیده

امروزه سرطان ها یکی از مهمترین عوامل مرگ ومیر انسان ها محسوب می شوند که فناوری ریز آرایه نقش مهمی در تشخیص، درمان و طبقه بندی انواع بافت های سرطانی دارد. وجود ژن های زیاد و تعداد اندک نمونه ها باعث ایجاد مشکلاتی در طراحی طبقه بندها شده است. بنابراین کاهش ابعاد و انتخاب تعداد کمی از ژن ها با دقت پیش بینی بالا برای طبقه بندی سرطان از چالش های مهم تحلیل داده های ریز آرایه است. در این مقاله مطالعه تاثیر روش های انتخاب ویژگی با استفاده از روش های فیلتر در دقت و خطای طبقه بندی نظارت شده سرطان را ارائه می کند. مقایسه بین روش های مختلف فیشر، T-آمار، SNR و ReliefF با استفاده از مجموعه داده های مختلف سرطان شامل سرطان خون، پروستات و روده بزرگ ارزیابی شده است. نتایج طبقه بندی با استفاده از طبقه بندی K نزدیک ترین همسایه (KNN) و ماشین بردار پشتیبان (SVM) که ترکیبی بین روش های SNR و طبقه بندی SVM با بالاترین دقت را نشان می دهد.

واژگان کلیدی: ماشین بردار پشتیبان، K نزدیک ترین همسایه، بیان ژن، طبقه بندی، انتخاب ویژگی

مقدمه

فناوری ریز آرایه در سال 1996 متولد و با عناوین آرایه های DNA، تراشه های ژنی، تراشه های DNA و تراشه های زیستی نامگذاری شده است. فناوری ریز آرایه یکی از آخرین پیشرفت ها در زمینه زیست شناسی ملکولی است که اجازه نظارت بر بیان هزاران ژن را به صورت همزمان تنها در یک آزمایش هیبریداسیون می دهد. علاوه بر پتانسیل علمی این فناوری در مطالعه بنیادین بیان ژن؛ یعنی تنظیم و تعاملات ژن ها، کاربردهای مهمی در پژوهش های دارویی و کلینیکی دارد. برای مثال، با مقایسه بیان ژن در سلول های سالم و ناسالم، ریز آرایه می تواند در شناسایی ژن های ناسالم برای داروهای درمانی یا ارزیابی تاثیر آن ها استفاده شود. ریز آرایه دارای هزاران نقطه بوده، هر یک از این نقاط حاوی دنباله های مختلف شناخته شده DNA، به نام نشانگر هستند. این نقاط روی یک اسلاید شیشه ای توسط یک arrayer رباتیک چاپ می شوند. توسعه در بیوتکنولوژی زیست شناسی ملکولی اطلاعات موجود در ژن ها را با ارائه میکرو آرایه ابزار تشخیصی

* Corresponding author: ایران-کرمان-موسسه آموزش عالی بهمنیار-دانشکده کامپیوتر

Email: Somaye.balloch1991@gmail.com