مجله دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران (پیاورد سلامت)
دوره ۱۴ شماره ۲ خرداد و تیر ۱۳۹۹، ۱۷۵-۱۶۳

مقاله پژوهشی

ایجاد مدل تصمیم غربال‌گری بر اساس میزان خطر ابتلا به سرطان کولورکتال

رئوف نوپور[۱]، محمد شیرخدا[۲]، شراره رستم نیاکان کلهری[۳]

دریافت مقاله : آذر ۱۳۹۸
پذیرش مقاله : فروردین ۱۳۹۹

* نویسنده مسئول :
شراره رستم نیاکان کلهری؛
دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران
Email :
sh-rniakank@sina.tums.ac.ir

چکیده

زمینه و هدف: سرطان کولورکتال از شایع‌ترین سرطان‌های دستگاه گوارش در انسان و مهم‌ترین عامل مرگ و میر در جهان محسوب می‌شود. استفاده از برنامه‌ی غربال‌گری مناسب بر اساس میزان خطر ابتلا به سرطان کولورکتال در افراد، می‌تواند در پیشگیری از این بیماری مناسب باشد. بنابراین، هدف این پژوهش طراحی مدلی برای غربال‌گری سرطان کولورکتال بر اساس ریسک فاکتورها بود تا بتوان از یک سو میزان بقای این بیماری را در سطح جامعه افزایش داد و از سوی دیگر میزان مرگ و میر کاهش یابد.

روش بررسی: ابتدا با مرور مقالات و بررسی اطلاعات موجود در پرونده بیماران، ۳۸ ریسک فاکتور شناسایی شدند، برای تعیین مهم‌ترین ریسک فاکتورها از بعد بالینی، از content validity ratio(CVR) و از نظر آماری و با توجه به مجموعه داده گردآوری شده از روش‌های آماری ضریب همبستگی اسپیرمن و تحلیل رگرسیون لجیستیک استفاده گردید. سپس از چهار الگوریتم تولید قوانین PART، J-RIP، J-48 و REP-Tree برای داده‌کاوی استفاده شد و مناسب‌ترین مدل بر اساس مقایسه میزان عملکرد الگوریتم‌ها به‌دست آمد.

یافته‌ها: پس از مقایسه عملکرد الگوریتم‌ها، الگوریتم داده‌کاوی J-48 با میزان ۰/۸۸۹ = F-Measure، عملکرد بالاتری نسبت به سایر الگوریتم‌ها داشته است.

نتیجه‌گیری: نتایج حاصل از ارزیابی عملکرد الگوریتم داده‌کاوی J-48 نشان داد که این الگوریتم می‌تواند به عنوان مناسب‌ترین مدل تعیین خطر ابتلا به سرطان کولورکتال در نظر گرفته شود.

واژه‌های کلیدی: مدل تصمیم، تشخیص زودرس، سرطان کولورکتال، ریسک فاکتور، داده‌کاوی

۱ کارشناس ارشد فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران
۲ دانشیار گروه جراحی آنکولوژی، انستیتو کانسر، مجتمع بیمارستانی امام خمینی(ره)، تهران، ایران
۳ دانشیار گروه مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

مقدمه

سرطان به رشد غیرعادی سلول‌ها اطلاق می‌شود که توانایی تهاجم یا انتشار به بافت‌های مجاور یا دوردست بدن دارند(۱). سرطان کولورکتال به رشد سرطان در کولون یا رکتوم که بخش‌هایی از روده بزرگ هستند اطلاق می‌گردد(۲). این بیماری سومین سرطان شایع در بین مردان و دومین سرطان شایع در بین جمعیت زنان را تشکیل می‌دهد(۳). همچنین سومین سرطان شایع در میان مردان و زنان آمریکایی می‌باشد(٤). عواملی مانند سن، سابقه‌ی شخصی آدنوما پولیپ، سابقه‌ی شخصی بیماری التهابی روده، سابقه‌ی خانوادگی سرطان کولورکتال یا پولیپ، عوامل ژنتیکی، عوامل محیطی، عوامل تغذیه‌ای، سیگار، چاقی و فعالیت‌های ورزشی، مصرف الکل و سایر موارد در ایجاد سرطان کولورکتال موثرند(۵). تظاهرات بالینی سرطان کولورکتال به مکان سرطان و مرحله‌ی پیشرفت آن بستگی دارد. تغییر عادات روده‌ای، دردهای عمومی یا محلی شکم، کاهش وزن بدون علت مشخص، ضعف، کمبود آهن و کم خونی از شایع‌ترین نشانه‌های سرطان کولورکتال می‌باشند(٦).

میزان بقای پنج و حتی ده ساله در بیشتر کشورهای اروپایی بالای ۵۰٪ گزارش شده است و این در حالی است که در اکثر مطالعات در کشور ایران میزان بقای پنج و ده ساله زیر ۵۰٪ گزارش شده است و علت این امر شاید وجود درمان‌های کلینیکی در سطح بالاتر در آن کشورها و یا پیگیری بیشتر مردم آن کشورها برای بهره‌برداری از مزایای درمانی باشد و از علل دیگر که میزان بقا در کشور ما نسبت به کشورهای غربی پایین است، مراجعه‌ی دیرتر بیماران و در نتیجه تشخیص بیماری در مرحله‌ی پیشرفته‌تر و نبود برنامه‌ی غربال‌گری این سرطان می‌باشد(۷). میزان پیش‌آگهی پنج ساله‌ی سرطان کولورکتال از ۳۳٪ در دهه‌ی هفتاد میلادی به ۵۵٪ در دهه‌ی نود میلادی افزایش یافته است که علت این بهبود در پیش آگهی را می‌توان تغییر در سیر طبیعی بیماری، بهبود روش‌های تغذیه‌ای و روش‌های تشخیصی دانست. همچنین میزان بقا در کشورهای غربی به دلیل روش‌های دقیق بیماریابی و جلوگیری از سیر طبیعی تومور، سیر صعودی داشته است و میزان مرگ و میر ناشی از آن رو به کاهش است(۸). میزان بقای افراد با سابقه‌ی جراحی سرطان کولورکتال در مراحل اولیه سرطان که اقدامات تشخیصی به موقع فراهم می‌شود بیش از ۸۰٪ است، درحالی‌که این میزان در افراد با مراحل پیشرفته‌ی سرطان، کمتر از ٤۰٪ است. تشخیص به موقع میزان بقا بیش از ٤۰٪ افزایش داشته است؛ بنابراین تشخیص بیماری در مراحل اولیه بسیار مهم تلقی می‌شود(۹). توسعه‌ی سریع علوم کامپیوتر منجر به خودکار شدن بسیاری از فعالیت‌ها می‌شود که در گذشته تنها به‌وسیله‌ی متخصصان انجام می‌گردید، توسعه‌ی تکنیک‌های مبتنی بر هوش مصنوعی نقش مهمی در حل مسایل پیچیده ایفا نموده است، تشخیص پزشکی به دلیل پیچیدگی جسم و روح انسان و ابهامات در تشخیص پزشکی مثال مناسبی برای نشان دادن پیچیدگی‌هاست؛ همچنین دانش موجود در بین متخصصان گوناگون نیز متفاوت می‌باشد. به‌رغم انکارپذیری دانش پزشکان در تشخیص، تکنیک‌های مبتنی بر هوش مصنوعی با کشف سریع و پایگاه دانش موثق و دقیق در کاهش هزینه در این زمینه بسیار موثر است(۱۰). هوش مصنوعی به سیستم‌هایی گفته می‌شود که می‌توانند واکنش‌هایی مشابه رفتارهای هوشمند انسانی از جمله درک شرایط پیچیده، شبیه‌سازی فرایندهای تفکری و شیوه‌های استدلالی انسانی و پاسخ موفق به آنها، یادگیری و توانایی کسب دانش و استدلال برای حل مسایل را داشته باشند(۱۱). به صورت اختصاصی یکی از کاربردهای هوش مصنوعی، داده‌کاوی است که از فرایندهای خودکار شده برای کشف اطلاعات و الگوها از میان داده‌ها استفاده می‌کند(۱۲). در حوزه‌ی پزشکی داده‌کاوی کاربردهای گوناگونی داشته است(۱۳). مثلا در پژوهش Shin و همکاران با عنوان«مدل پیش‌بینی خطر سرطان کولورکتال: مطالعه سازمان بیمه سلامت ملی کشور کره جنوبی» انجام گردیده بود، از مدل رگرسیون نسبت‌خطر کاکس برای توسعه‌ی مدل پیش‌بینی استفاده شده بود. نتایج حاصل از ارزیابی مدل نشان داده که بین موارد سرطان کولورکتال مشاهده شده و نتایج حاصل از پیش‌بینی سیستم اختلاف ناچیزی وجود دارد. همچنین قدرت تمییز مدل توسط C-Statistics اندازه‌گیری شده بود که برای مردان بین ۰/۷٦۲ تا ۰/۷۸٦ و برای زنان بین ۰/٦۷۸ تا ۰/۷٦۳ بود(١٤). نتایج حاصل از پژوهش ستاره و همکاران در سال ۱۳۹۷ با عنوان«پیش بینی بقای بیماران مبتلا به سرطان روده بزرگ در بخش پرتودرمانی بیمارستان نمازی شیراز با استفاده از روش‌های داده‌کاوی

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

ماشین‌بردار پشتیبان و بکینگ» نشان داد که دو روش داده‌کاوی بکینگ و ماشین بردار صحت بالایی در پیش‌بینی بقای بیماران مبتلا به سرطان روده‌ی بزرگ دارند که این میزان در الگوریتم ماشین بردار بیشتر بود و نتایج این مطالعه می‌توانست کاربرد زیادی برای متخصصان بالینی درگیر در درمان بیماران مبتلا به سرطان روده بزرگ داشته باشد و مدل‌های پیشنهاد شده می‌توانستند با دقت بالایی پیامد بیماری در بیماران مبتلا به سرطان روده بزرگ را پیش‌بینی کنند(۱۵). وکیلی و همکاران در پژوهش خود با عنوان«بررسی میزان بقای بیماران مبتلا به سرطان کولورکتال در استان یزد طی سال‌های ۱۳۹۰-۱۳۸۰» به این نتیجه رسیدند که از علل اصلی کاهش میزان بقای ناشی از سرطان کولورکتال در ایران نسبت به سایر کشورها تشخیص دیررس و درمان‌های تکمیلی ناقص بعد از جراحی و فقدان برنامه‌های دقیق جهت پیگیری بیماران است. لذا اجرای برنامه‌ی منظم غربال‌گری، آموزش عمومی مردم و آموزش اختصاصی پزشکان به تشخیص زودرس سرطان کولورکتال در مراحل پایین‌تر کمک می‌کند و از طرف دیگر می‌توان با پیگیری دقیق بیماران و درمان سریع موارد عود به ارتقای بقای بیماران کمک نمود(۷). فاطمیان و همکاران با عنوان«عوامل پیش‌بینی کننده‌ی آدنوما و سرطان کولونیک پیشرفته با استفاده از داده‌کاوی» نشان دادند که روش‌های داده‌کاوی در شناخت عوامل پیش‌بینی‌کننده‌ی نئوپلاسم‌های پیشرفته کولورکتال، کاربرد مناسبی دارد(۱۶). Wen و همکاران در پژوهش خود از دو روش داده کاوی آدابوست و رگرسیون لجیستیک جهت پیش بینی سرطان کولورکتال در افراد با متاستاز کبد استفاده کردند. نتایج حاصل از ارزیابی دو الگوریتم داده‌کاوی نشان داد که رگرسیون لجیستیک عملکرد بهتری در پیش‌بینی خطر ابتلا به سرطان کولورکتال داشته است(۱۷).

تاکنون در سطح جهانی، پژوهشی که بتواند انواع متفاوتی از ریسک فاکتورهای سرطان کولورکتال را تحلیل کرده و بر اساس آن افراد را بر اساس خطر سرطان کولورکتال غربال‌گری کند، انجام نشده است. اکثر پژوهش‌ها در زمینه‌ی تشخیص و غربال‌گری سرطان کولورکتال، با استفاده از علایم و نشانه‌های آن بودند و یا ریسک فاکتورهای محدودی را بررسی کردند، بنابراین هدف از انجام این پژوهش، طراحی مدلی با استفاده از روش داده‌کاوی است تا بتوان میزان خطر سرطان کولورکتال را بر اساس انواع مختلفی از ریسک فاکتورهای این بیماری در افراد اندازه‌گیری کرد و بر اساس آن افراد پرخطر سرطان کولورکتال را غربال‌گری نمود تا پزشکان بتوانند مناسب‌ترین راهکارهای تشخیصی را در نتیجه‌ی تشخیص زودهنگام سرطان کولورکتال، به این دسته از افراد ارایه دهند. به این ترتیب از یک جهت سرطان کولورکتال در افراد قبل از پیدایش علایم و نشانه‌های آن در فرد کشف می‌گردد و از جهت دیگر میزان بقای ناشی از این بیماری در سطح جامعه افزایش می‌یابد.

روش بررسی

در این مطالعه که به‌صورت توصیفی و کاربردی بود، ابتدا به منظور آشنایی با ریسک فاکتورهای مرتبط با سرطان کولورکتال، مقالات بین سال‌های ۲۰۱٤ تا ۲۰۱۸ مرتبط با ریسک فاکتورهای سرطان کولورکتال در پایگاه‌های اطلاعاتی معتبر همچون PubMed، Google Scholar، Scopus، Willy online library و Science direct، با کلید واژه‌هایی همچون Colorectal cancer، Risk factors و Prediction، Risk determinant، Guideline، Screening و با مشورت متخصصان آشنا به ریسک فاکتورهای سرطان کولورکتال بیمارستان امام خمینی شهرستان ساری که تعداد آن‌ها ده نفر بود، کلیه ریسک فاکتورهای تعیین خطر ابتلا به سرطان کولورکتال شناسایی شدند؛ سپس اطلاعات تمامی پرونده‌های افراد مبتلا به سرطان کولورکتال طی سال‌های ۱۳۹۲ تا ۱۳۹۷ که به تعداد ۸۰۰ پرونده از افراد مبتلا به سرطان کولورکتال و افراد سالم بود از نظر وجود عناصر اطلاعاتی مرتبط با ریسک فاکتورها بررسی گردید، پس از تطبیق ریسک فاکتورهایی که از طریق جست و جو از مقالات به‌دست آمده بود و ریسک فاکتورهای موجود در پرونده‌ها، ۳۸ ریسک فاکتور به‌دست آمد. به منظور تعیین مهم‌ترین ریسک فاکتورها از دیدگاه متخصصان بالینی حوزه‌ی گوارش، از یک چک لیست به صورت طیف سه گزینه‌ای Lawshe استفاده گردید. چک لیست موردنظر شامل سه بخش«ضروری است»، «مفید است اما ضروری نیست» و «مفید نیست» بود و برای اخذ نظرسنجی بین ده متخصص گوارش بیمارستان امام خمینی شهرستان ساری که با ریسک فاکتورهای سرطان کولورکتال آشنایی داشتند، توزیع گردید. پس از اخذ نظرسنجی از متخصصان، ریسک فاکتورهای

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

حاصل جهت داده‌کاوی استخراج شد و مابقی آن از پژوهش حذف گردید. سپس داده‌های بیماران مرتبط با ریسک فاکتورها از پرونده بیماران استخراج شده و در فرم اکسل ذخیره گردیدند و مجموعه داده‌های مرتبط با افراد مبتلا به سرطان کولورکتال و افراد سالم که مشتمل بر ۸۰۰ نمونه بود، به‌دست آمد، با بررسی گایدلاین‌های مرتبط با سرطان کولورکتال از جمله گایدلاین انجمن سرطان امریکا و کنسرسیوم برک(۱۸)، از میان ۸۰۰ نمونه، ۳۳۲ نمونه‌ی آن مربوط به افراد پرخطر سرطان کولورکتال بود. افراد پرخطر، افرادی بودند که دارای ریسک فاکتورهایی مانند سابقه‌ی بیماری التهابی روده، سابقه‌ی آدنوما پولیپ، سابقه‌ی سندروم‌های ژنتیکی سرطان و سابقه‌ی خانوادگی سرطان کولورکتال زیر سن ۶۰ سال بودند و نمی‌توان سایر عوامل خطر سرطان کولورکتال را در آن بررسی کرد. مثلا فردی که دارای سابقه‌ی سندروم‌های ژنتیکی مثل توکارت بود، به خودی خود خطر بالا محسوب می‌شد و بررسی سایر عوامل خطر مانند مصرف چربی در آن بی‌فایده بود. بنابراین، به منظور بررسی سایر عوامل خطر مرتبط با سرطان کولورکتال در افراد، نمونه‌های مرتبط با ریسک فاکتورهای خطر بالا از مجموعه داده به‌دست آمده حذف گردیدند. نهایتا ٤٦٨ نمونه مورد استفاده قرارگرفتند که ٢٧٤ نمونه(۵۸/۵٪) مربوط به افراد سالم و ١٩٤ نمونه(٤١/۵٪) مربوط به افراد مبتلا به سرطان کولورکتال بودند.

پس از جمع‌آوری داده‌ها و ایجاد مجموعه داده مرتبط با سرطان کولورکتال در افراد مبتلا و سالم، ابتدا با استفاده از نرم افزار SPSS، کلیه ریسک فاکتورها و ریسک ابتلا به سرطان کولورکتال برحسب فراوانی و درصد فراوانی بررسی گردید، پس از آن با استفاده از تحلیل کوواریانس و روش اسپیرمن، معنی‌دار بودن رابطه‌ی بین هر یک از ریسک فاکتورها با ریسک ابتلا به سرطان کولورکتال با توجه به مجموعه‌ی داده جمع‌آوری و بررسی شد و ریسک فاکتورهایی که ارتباط آن‌ها با ریسک ابتلا به سرطان کولورکتال در سطح P.value<۰/۰۵، معنی‌دار بود بررسی گردیده و مابقی آن از پژوهش حذف گردید. در مرحله‌ی بعد با استفاده از تحلیل رگرسیون لجیستیک دودویی و روش Forward LR به منظور تولید مدل پیش بینی خطر ابتلا به سرطان کولورکتال استفاده گردید. در روش Forward LR برای تولید مدل پیش بینی خطر ابتلا به سرطان کولورکتال به منظور انجام غربال‌گری، از نمودار IF Term Removed برای تعیین مهم‌ترین ریسک فاکتورهای پیش بینی خطر سرطان کولورکتال استفاده گردید؛ در این نمودار با اضافه شدن هر ریسک فاکتور به مدل، میزان Log likelihood آن کاهش یافت. میزان Log likelihood ضریبی از ارزش مشاهده شده و پیش‌بینی شده مطابق رابطه‌ی(۱) می‌باشد و کاهش میزان آن با اضافه شده متغیرها نشان دهنده‌ی کارایی بهتر مدل است. در نتیجه‌ی این کار مهم‌ترین ریسک فاکتورهای مدل پیش‌بینی خطر ابتلا به سرطان کولورکتال، به منظور غربال گری استفاده شدند.

$$Log-likelihood=\sum_{i=1}^{n}(Yi(\ ln(Pi)\ )+(1-Yi)(\ln(1-Pi)) \quad (۱)$$ رابطه‌ی

با توجه به مشخص بودن ریسک ابتلا به سرطان کولورکتال از روش‌های داده‌کاوی مبتنی بر یادگیری نظارت‌پذیر و به‌منظور ایجاد قوانین در پیش‌بینی خطر ابتلا به سرطان کولورکتال به منظور غربال‌گری آن، از نرم افزار داده‌کاوی وکا با استفاده از دو درخت تصمیم J-48 و REP-Tree و در الگوریتم طبقه‌بندی Part و J-RIP، با توجه به این که این چهار الگوریتم قوانینی را در قالب ساختار IF-THEN به‌عنوان خروجی فراهم می‌کردند و تمام نمونه‌های پژوهش را طبقه‌بندی کردند، استفاده گردید و بر اساس ماتریس آشفتگی، میزان F-Measure تمامی الگوریتم‌ها مقایسه گردید و نهایتا مناسب‌ترین قوانین پیش‌بینی خطر ابتلا به سرطان کولورکتال به منظور غربال‌گری آن بر اساس الگوریتم J-48 ایجاد شد.

یافته‌ها

پس از بررسی عناصر اطلاعاتی مرتبط با ریسک فاکتورهای سرطان کولورکتال موجود در پرونده‌ی افراد مبتلا به سرطان کولورکتال و تطبیق با مقالات مرتبط با ریسک فاکتورهای سرطان کولورکتال، ۳۸ ریسک فاکتور که عناصر اطلاعاتی مرتبط با آن در پرونده موجود بود استخراج گردید؛ برخلاف مقالات دیگر که فقط برخی از ریسک فاکتورها مثلا ریسک فاکتورهای مربوط به سبک زندگی و یا عوامل تولیدمثلی در خانم‌ها را مورد بررسی قرار دادند، در این پژوهش بر اساس اطلاعات موجود در پرونده‌ها، تمامی عوامل از جمله عوامل اپیدمیولوژیک، عوامل تولیدمثلی مختص خانم‌ها، عوامل دموگرافیک و سوابق پزشکی، در کنار یکدیگر بررسی شدند. برای اخذ نظرسنجی برای اعتبار روایی ریسک فاکتورها از بعد بالینی، از ده متخصص گوارش که با انواع ریسک فاکتورهای

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

سرطان کولورکتال آشنایی داشتند، استفاده گردید. برای این منظور معیار تصمیم‌گیری Lawshe استفاده گردید. جدول ۱، کلیه ریسک فاکتورهای مرتبط با سرطان کولورکتال که از پرونده‌ها استخراج گردیدند، به همراه طیف سه گزینه ای Lawshe برای نظرسنجی از متخصصان نشان داده شده است.

جدول ۱: متغیرهای موثر در تعیین خطر سرطان کولورکتال

ردیف	متغیر	مفید است		مفید است اما ضروری نیست		مفید نیست	
		فراوانی	فراوانی نسبی	فراوانی	فراوانی نسبی	فراوانی	فراوانی نسبی
	عوامل دموگرافیک:						
۱	سن	۹	۰/۹	۱	۰/۱	۰	۰
۲	طبقه‌ی اجتماعی	۷	۰/۷	۳	۰/۳	۰	۰
	عوامل تولیدمثلی مختص خانم‌ها:						
۳	هورمون درمانی	۹	۰/۹	۰	۰	۱	۰/۱
٤	سن یائسگی	۹	۰/۹	۱	۰/۱	۰	۰
۵	سن اولین حاملگی	۸	۰/۸	۱	۰/۱	۱	۰/۱
٦	سن اولین قاعدگی طبیعی	٦	۰/٦	۳	۰/۳	۱	۰/۱
۷	تعداد حاملگی	۹	۰/۹	۱	۰/۱	۰	۰
۸	تعداد فرزند سالم به دنیا آمده	۹	۰/۹	۱	۰/۱	۰	۰
۹	تعداد سقط	۵	۰/۵	۲	۰/۲	۳	۰/۳
۱۰	نوع یائسگی	٤	۰/٤	۱	۰/۱	۵	۰/۵
۱۱	قرص ضد بارداری	۹	۰/۹	۱	۰/۱	۰	۰
	عوامل اپیدمیولوژیک:						
۱۲	میزان فعالیت ورزشی در روز	۹	۰/۹	۱	۰/۱	۰	۰
۱۳	غذاهای لبنی	۸	۰/۸	۱	۰/۱	۱	۰/۱
۱٤	مصرف گوشت و گوشت فرآوری شده	۹	۰/۹	۱	۰/۱	۰	۰
۱۵	چای	٦	۰/٦	۲	۰/۲	۲	۰/۲
۱٦	قهوه	٤	۰/٤	۲	۰/۲	٤	۰/٤
۱۷	قرص آرام بخش	۷	۰/۷	۲	۰/۲	۱	۰/۱
۱۸	شاخص توده بدنی	۹	۰/۹	۱	۰/۱	۰	۰
۱۹	مصرف الکل	۹	۰/۹	۱	۰/۱	۰	۰
۲۰	غذاهی پرچرب و کم فیبر	۱۰	۱	۰	۰	۰	۰
۲۱	میزان مصرف انرژی روزانه	۷	۰/۷	۲	۰/۲	۱	۰/۱
۲۲	مکمل مولتی ویتامین	۹	۰/۹	۱	۰/۱	۰	۰
۲۳	مکمل ویتامین D	۹	۰/۹	۱	۰/۱	۰	۰
۲٤	مکمل آهن	۹	۰/۹	۱	۰/۱	۰	۰
۲۵	مصرف میوه و سبزی	۱۰	۱	۰	۰	۰	۰
۲٦	مصرف سیگار	۱۰	۱	۰	۰	۰	۰
۲۷	مصرف آسپرین	۹	۰/۹	۱	۰/۱	۰	۰

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

۲۸	میزان خواب	۹	۰/۹	۱	۰/۱	۰	۰
۲۹	مواد مخدر	۹	۰/۹	۱	۰/۱	۰	۰
۳۰	استاتین	۷	۰/۷	۳	۰/۳	۰	۰
۳۱	کلسیم	۹	۰/۹	۱	۰/۱	۰	۰
	سوابق پزشکی:						
۳۲	ژنتیک	۱۰	۱	۰	۰	۰	۰
۳۳	آدنوما	۱۰	۱	۰	۰	۰	۰
۳٤	کبد چرب	۹	۰/۹	۱	۰/۱	۰	۰
۳۵	سندروم‌های متابولیک	۹	۰/۹	۱	۰/۱	۰	۰
۳٦	بیماری التهابی روده	۱۰	۱	۰	۰	۰	۰
۳۷	آندوسکوپی	۹	۰/۹	۱	۰/۱	۰	۰
۳۸	سابقه‌ی سرطان‌های دیگر	۱۰	۱	۰	۰	۰	۰

جهت تعیین ریسک فاکتورهای موثر در پیش‌بینی خطر ابتلا به سرطان کولورکتال، از جدول ۲ و فرمول CVR مطابق رابطه‌ی(۲) استفاده گردید. در این رابطه، Ne تعداد متخصصانی است که به گزینه‌ی«ضروری است» پاسخ داده‌اند و N تعداد متخصصان شرکت‌کننده در نظر سنجی است.

$$CVR = \frac{N_e - N/2}{N/2}$$ رابطه‌ی(۲)

جدول ۲: حداقل مقادیر قابل قبول CVR برای تعداد متفاوتی از متخصصان

تعداد پنل متخصصان	حداقل مقدار قابل قبول CVR
۵	۰/۹۹
٦	۰/۹۹
۷	۰/۹۹
۸	۰/۷۵
۹	۰/۷۸
۱۰	۰/٦۲
۱۱	۰/۵۹
۱۲	۰/۵٦
۱۳	۰/۵٤
۱٤	۰/۵۱
۱۵	۰/٤۹
۲۰	۰/٤۲
۲۵	۰/۳۷
۳۰	۰/۳۳
۳۵	۰/۳۱
٤۰	۰/۲۹

مثلا برای ریسک فاکتور سن، میزان عددی CVR مطابق رابطه‌ی ۲،

$$CVR\ (Age) = \frac{۹ - \frac{۱۰}{۲}}{\frac{۱۰}{۲}} = ۰/۸$$

به میزان ۰/۸ به‌دست آمد. مطابق جدول ۲ از آن‌جاکه ۰/۸>۰/٦۲ می‌باشد، بنابراین اعتبار محتوایی این آیتم تایید می‌گردد.

پس از تعیین اعتبار محتوایی کلیه ریسک فاکتورها با استفاده از

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

CVR، ۲۶ ریسک فاکتور از بعد بالینی مطابق جدول ۱، به عنوان ریسک فاکتورهای موثر در پیش‌بینی خطر ابتلا به سرطان کولورکتال شناخته شدند.

پس از استخراج داده‌های افراد سالم و مبتلا به سرطان کولورکتال طبق ۲۶ ریسک فاکتور تعیین شده از پرونده و جداسازی نمونه‌های افراد پرخطر سرطان کولورکتال، ٤٦٨ نمونه به‌دست آمد که ۲۷٤ نمونه(۵۸/۵٪) مربوط به افراد سالم و ۱۹٤ نمونه(٤۱/۵٪) مربوط به افراد مبتلا به سرطان کولورکتال بود.

نتایج حاصل از تحلیل همبستگی هر یک از ریسک فاکتورها با ریسک ابتلا به سرطان کولورکتال با استفاده از روش اسپیرمن نشان داد که متغیرهای میزان مصرف چربی، مصرف گوشت قرمز و فرآوری شده، میزان فعالیت ورزشی، مصرف میوه و سبزی، مصرف سیگار، مصرف قرص آهن، مصرف قرص آسپرین، مصرف قرص ضدبارداری، مصرف الکل و مصرف مواد مخدر در پژوهش، همبستگی معنی‌داری را با ریسک ابتلا به سرطان کولورکتال در سطح P.value>۰/۰۵ نشان دادند و مابقی متغیرها از پژوهش حذف گردیدند، مثلا برای ریسک فاکتور مصرف گوشت قرمز و فرآوری شده، نتایج حاصل از تحلیل همبستگی آن مطابق جدول ۳ به‌دست آمد.

جدول ۳: نتایج حاصل از تحلیل همبستگی مصرف گوشت قرمز با ریسک ابتلا به سرطان کولورکتال

میزان همبستگی بین متغیر و ریسک ابتلا به سرطان کولورکتال به صورت ماتریس همبستگی	میزان همبستگی بین متغیر ریسک ابتلا به سرطان کولورکتال به صورت ماتریس همبستگی	مشخصات همبستگی	نام متغیر	نوع همبستگی
۰/۲۲۳	۱/۰۰۰	ضریب همبستگی	ریسک ابتلا به سرطان کولورکتال	اسپیرمن
۰/۰۰۰	۰	میزان (P.value)		
٤٦٨	٤٦٨	تعداد نمونه		
۱/۰۰۰	۰/۲۲۳	ضریب همبستگی	مصرف گوشت قرمز	
۰	۰/۰۰۰	میزان (P.value)		
٤٦٨	٤٦٨	تعداد نمونه		

طبق جدول ۳، رابطه این ریسک فاکتور در سطح کمتر از P.value<۰/۰۵ با ریسک ابتلا به سرطان کولورکتال معنادار بود، بنابراین به‌عنوان ریسک فاکتور موثر در پیش‌بینی خطر ابتلا به سرطان کولورکتال به منظور غربال‌گری در نظر گرفته شد.

جدول شماره ٤، نتایج حاصل از تحلیل رگرسیون لجیستیک دودویی با روش Forward LR و ارزیابی کارایی مدل با استفاده از مدل IF Term Removed را به‌منظور ایجاد مدل پیش بینی خطر ابتلا به سرطان کولورکتال در گام هفتم نشان می‌دهد.

جدول ۴: نتایج حاصل از مدل رگرسیون لجیستیک دودویی پیش‌بینی خطر سرطان کولورکتال در گام هفتم مدل‌سازی

Model IF Term Removed

	Variable	Model Log Likelihood	Change in Likelihood	df	Sig of change
Step 7	Smoking (Day)*Smoking(Years)	-۱٤۰/۳۳۷	۱۱/۰۲۸	۱	۰/۰۰۱
	High-fat	-۱۳۷/۱۸۰	٤/٦۳٦	۱	۰/۰۳۱
	Red-meat	-۱۳٦/۸٦۸	٤/۰۱۰	۱	۰/۰٤٥
	Fruit & Veg	-۱٤۰/٥۳۸	۱۱/۳٥۱	۱	۰/۰۰۱
	Exercise	-۱٤۲/۱۷٥	۱٤/٦۲٦	۱	۰/۰۰۰
	Aspirin(Day/2)*Aspirin(Years)	-۱٤۰/۳٤۰	۱۰/۹٥٥	۱	۰/۰۰۱
	Genetic(After age 60)	-۱۳٦/۹٦۹	٤/۲۱۳	۱	۰/۰٤۰

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

مطابق جدول ٤، هفت ریسک فاکتور مصرف چربی حیوانی، مصرف گوشت قرمز، مصرف میوه و سبزی، مصرف قرص آسپرین، سابقه‌ی ژنتیکی سرطان کولورکتال(بالای ٦٠ سال)، میزان فعالیت ورزشی و مصرف سیگار به‌عنوان مهم‌ترین ریسک فاکتورها جهت ایجاد مدل پیش‌بینی خطر ابتلا به سرطان کولورکتال به منظور غربال‌گری آن می‌باشند.

نتایج حاصل از ارزیابی چهار الگوریتم داده کاوی Part، J-RIP، J-48 و REP-Tree بر اساس ماتریس آشفتگی همراه با ١٠٪ Cross-validation (جدول ٥) و تعیین میزان F-Measure حاصل از آن‌ها در جدول ٦ بیان گردیده است.

جدول ٥: ماتریس آشفتگی حاصل از چهار الگوریتم داده کاوی

نوع الگوریتم داده کاوی	TP	FP	FN	TN
PART	٢٤٤	٣٠	٥٥	١٣٩
J-RIP	٢٣٠	٤٤	٦٨	١٢٦
J-48	٢٦٣	١١	٤٠	١٥٤
REP-Tree	٢٤٧	٢٧	٩٨	٩٦

در ماتریس آشفتگی، TP معرف تعداد نمونه‌هایی است که به درستی توسط مدل به‌عنوان درست طبقه‌بندی شده‌اند. در الگوریتم Part، ٢٤٤ نمونه از افراد سالم، به‌درستی توسط مدل، دسته‌بندی شده بودند؛ بنابراین میزان True Positive(TP) حاصل از الگوریتم Part مقدار ٢٤٤ به‌دست آمده بود. False Positive(FP) معرف تعداد نمونه‌هایی است که به اشتباه به‌عنوان درست طبقه‌بندی شده‌اند، در الگوریتم Part، ٣٠ نمونه از افراد سالم، به‌عنوان مبتلا دسته‌بندی شده‌اند، بنابراین میزان FP آن مقدار ٣٠ به‌دست آمده بود. False Negative(FN) معرف تعداد نمونه‌هایی است که به اشتباه به‌عنوان نادرست طبقه‌بندی شده‌اند، در الگوریتم Part، ٥٥ نمونه از افراد مبتلا به سرطان کولورکتال، به‌عنوان سالم دسته‌بندی شده‌اند، بنابراین میزان FN آن مقدار ٥٥ به‌دست آمده بود. True Negative(TN) معرف داده‌هایی است که به‌درستی به‌عنوان غلط طبقه‌بندی شده‌اند؛ در الگوریتم Part، ١٣٩ نمونه از افراد مبتلا به سرطان کولورکتال، به‌عنوان مبتلا دسته‌بندی شده‌اند، بنابراین میزان TN آن مقدار ١٣٩ به‌دست آمده بود. همچنین میزان F-Measure حاصل از ماتریس آشفتگی حاصل از الگوریتم طبقه‌بندی Part ٨١/٦٪ به‌دست آمده بود.

میزان TP، FP، FN و TN حاصل از ماتریس آشفتگی الگوریتم JRIP به‌ترتیب ٢٣٠، ٤٤، ٦٨ و ١٢٦ به‌دست آمده بود، بنابراین میزان F-Measure حاصل از ماتریس آشفتگی این الگوریتم به میزان ٧٥/٨٪ به‌دست آمده بود.

میزان TP، FP، FN و TN حاصل از ماتریس آشفتگی الگوریتم J-48 به‌ترتیب ٢٦٣، ١١، ٤٠ و ١٥٤ به‌دست آمده بود، بنابراین میزان F-Measure حاصل از ماتریس آشفتگی این الگوریتم به میزان ٨٨/٩٪ به‌دست آمده بود.

میزان TP، FP، FN و TN حاصل از ماتریس آشفتگی الگوریتم REP-Tree به‌ترتیب ٢٤٧، ٢٧، ٩٨ و ٩٦ به‌دست آمده بود؛ بنابراین میزان F-Measure حاصل از ماتریس آشفتگی این الگوریتم به‌ترتیب میزان ٧١/٨٪ به‌دست آمده بود.

جدول ٦: میزان F-measure حاصل از چهار الگوریتم داده‌کاوی

Algorithm	Part	J-RIP	J-48	REP-Tree
F-measure	٠/٨١٦	٠/٧٥٨	٠/٨٨٩	٠/٧١٨

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

نتایج حاصل از جدول ٦ نشان داد که الگوریتم درخت تصمیم J-48 میزان کارایی بالاتری را نسبت به سایر الگوریتم‌های داده‌کاوی در تولید قوانین تعیین خطر سرطان کولورکتال به‌منظور غربال‌گری داشته است.

نمایی از درخت تصمیم J-48 در شکل ۱ نشان داده شده است.

```
J48 pruned tree
------------------

smoking-(day) = 0
|   Aspirin-(P.days/2) = 0
|   |   exercise = 0
|   |   |   red-meat = <1: 0 (18.0/6.0)
|   |   |   red-meat = 1-2: 1 (33.0/14.0)
|   |   |   red-meat = 2-3: 1 (17.0/2.0)
|   |   |   red-meat = >3: 1 (0.0)
|   |   exercise = <1
|   |   |   genetics = 0: 0 (60.0/14.0)
|   |   |   genetics = 1
|   |   |   |   red-meat = <1: 0 (4.0/2.0)
|   |   |   |   red-meat = 1-2: 1 (3.0/1.0)
|   |   |   |   red-meat = 2-3: 0 (4.0/1.0)
|   |   |   |   red-meat = >3: 0 (0.0)
|   |   |   genetics = 2: 1 (1.0)
|   |   |   genetics = 3: 1 (2.0)
|   |   exercise = 1-2: 0 (23.0/6.0)
|   |   exercise = >2: 0 (3.0/1.0)
```

شکل ۱: نمایی از درخت تصمیم J-48

در اینجا به تفسیر دو قانون بر آمده از این درخت می‌پردازیم:

قانون ۱: IF (Smoking-(day) ==0 and Aspirin-(Days/2) ==0 and exercise==0 and red-meat<1 then 0).

قانون ۲: IF (Smoking-(day) ==0 and Aspirin-(Days/2) ==0 and exercise==0 and red-meat == 1-2 then 1).

الگوریتم J-48 ۱۸ نمونه را در قانون شماره ۱ طبقه‌بندی کرده که ٦ نمونه به اشتباه طبقه‌بندی شده، بنابراین ۱۲ نمونه به درستی طبقه‌بندی شده‌اند. در این قانون، افرادی که در روز سیگار مصرف نمی‌کنند، قرص آسپیرین مصرف نمی‌کنند، فعالیت ورزشی ندارند و میانگین مصرف گوشت قرمز و فرآوری شده آن‌ها در روز کمتر از یک وعده‌ی پنجاه گرمی است، در گروه افراد کم خطر از نظر ابتلای سرطان کولورکتال که الگوی افراد مصون از سرطان کولورکتال می‌باشد، طبقه‌بندی شده‌اند. تفسیر قانون شماره ۲ که ۱۹ نمونه به‌درستی در آن طبقه‌بندی شده بود، بدین‌صورت بود: افرادی که در روز سیگار مصرف نمی‌کنند، قرص آسپیرین مصرف نمی‌کنند، فعالیت ورزشی ندارند و میانگین مصرف گوشت قرمز و فرآوری شده آن‌ها در روز بین یک تا دو وعده‌ی پنجاه گرمی است، در گروه افراد پر خطر سرطان کولورکتال، طبق الگوی ابتلا به سرطان کولورکتال موجود در درخت تصمیم، طبقه بندی شده‌اند.

بحث

سرطان کولورکتال که یکی از مهم‌ترین عوامل مرگ و میر در جهان است و بیش از ۹٪ از سرطان‌ها را شامل می‌شود، سومین سرطان شایع و چهارمین علت مرگ و میر در جهان می‌باشد(۵). به‌رغم وجود روش‌های درمانی فراوان برای این بیماری، میزان بقای افراد در مراحل پیشرفته پایین است(۱۹). در طول سال‌های اخیر پیش‌آگهی افزایش یافته است که علت بهبود در این پیش‌آگهی را می‌توان در سیر طبیعی بیماری، تغییر عادات تغذیه‌ای و پیشرفت در روش‌های تشخیصی دانست(۸). در پژوهش حاضر، مدل تصمیم جهت تولید قوانین در قالب ساختار

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

IF-THEN استفاده از روش داده‌کاوی و الگوریتم درخت تصمیم استفاده گردید. در پژوهش Wu و همکاران با عنوان«توسعه خدمات غربال‌گری استفاده از درخت تصمیم ضایعات کولورکتال با استفاده از گوشی هوشمند» از سه درخت تصمیم برای تولید قوانین جهت افتراق ضایعات کولورکتال استفاده گردید. نتایج حاصل از پژوهش نشان داد که حداکثر دقت درخت تصمیم ۸۸٪ بوده است(۲۰). در پژوهشی که توسط Robertson و همکاران با عنوان«ارزیابی خطر ابتلا به سرطان کولورکتال در افراد با خونریزی رکتال» انجام شده بود، علایم و نشانه‌های بیمار از قبیل نوع رنگ خون مدفوع، وجود هموروئید و نشانه‌هایی مانند کمبود وزن و سابقه‌ی سندروم روده تحریک‌پذیر و سابقه‌ی بیماری دیورتیکولیت در بیماران مورد ارزیابی قرارگرفتند. برای پیش‌بینی خطر ابتلا به سرطان کولورکتال از روش تحلیل رگرسیون لجیستیک دودویی استفاده گردیده بود(۲۱). در پژوهش حاضر که برای پیش‌بینی خطر ابتلا به سرطان کولورکتال بر روی افراد انجام گردیده به جای استفاده از علایم و نشانه‌های سرطان، ریسک فاکتورهای سرطان کولورکتال برای پیش‌بینی خطر ابتلا به سرطان کولورکتال بررسی گردید. اصولا در سرطان‌هایی از قبیل کولورکتال وقتی علایم و نشانه‌ها بروز می‌کنند که سرطان در بدن فرد به میزان زیادی پیشرفت کرده است؛ بنابراین بررسی ریسک فاکتورهای سرطان کولورکتال بسیار کارامدتر از علایم و نشانه‌های آن می‌باشد. با توجه به این نکته که تشخیص زودرس سرطان کولورکتال می‌تواند ۹۰٪ در افزایش بقای این گونه از بیماران کمک‌کننده باشد(۱۸). Battista و همکاران از شبکه‌ی عصبی برای تشخیص سرطان کولورکتال استفاده کردند. ۱۷ سطح سرمی موجود در آزمایش خون به‌عنوان بیومارکر مهم سرطان کولورکتال شناخته شدند و شبکه‌ی عصبی قادر بود با استفاده از این بیومارکرها که عنوان لایه ورودی شبکه در نظر گرفته می‌شدند، وضعیت افراد را از نظر سالم بودن و یا ابتلا به سرطان کولورکتال شناسایی کند(۲۲). در پژوهش حاضر مانند پژوهش Battista از داده‌های افراد سالم و مبتلا به سرطان کولورکتال استفاده کرده است؛ اما به‌منظور پیش‌بینی خطر افراد از نظر ابتلا به سرطان کولورکتال پژوهش را گذشته‌نگر در آورده که در آن ویژگی افراد سالم همان ویژگی گروه کم خطر و ویژگی افراد سرطانی به‌عنوان ویژگی افراد پرخطر در نظر گرفته شد. در پژوهش حاضر برخلاف پژوهش باتیستا برای داده‌کاوی از الگوریتم‌های طبقه‌بندی و درخت تصمیم استفاده گردیده که از مزیت آن نسبت به شبکه عصبی تولید قوانین تعیین خطر ابتلا به سرطان کولورکتال به‌منظور غربال‌گری آن به‌عنوان خروجی بود. Anderson و همکاران در پژوهش خود برای پیش‌بینی خطر سرطان کولورکتال، سه الگوریتم داده‌کاوی را موردمقایسه قرار دادند و الگوریتمی را که کارایی بالاتری از خود نشان داده بود، به‌عنوان مناسب‌ترین مدل انتخاب کرد(۲۳). در پژوهش حاضر نیز مدل نهایی تعیین خطر ابتلا به سرطان کولورکتال به‌منظور غربال‌گری آن، از مقایسه‌ی کارایی چهار الگوریتم داده‌کاوی و انتخاب الگوریتم با کارایی بالاتر به‌دست آمد. در نتیجه، این پژوهش با ایجاد مدلی مبتنی بر داده‌کاوی میزان خطر سرطان کولورکتال را در افراد ارزیابی کرده و بر اساس آن افراد را با توجه به نوع ریسک طبقه‌بندی می‌کند تا پزشکان بتوانند بر اساس آن به غربال‌گری بیماران بپردازند و تصمیمات بالینی صحیح‌تری بگیرند. در گایدلاین‌های مختلف تعیین خطر سرطان کولورکتال(۱۸)، اغلب به عوامل ارثی و شخصی سرطان کولورکتال اشاره شده است؛ در پژوهش حاضر از عوامل خطر دموگرافیک، اپیدمیولوژیک، تولیدمثلی و سوابق پزشکی استفاده شده است؛ وقتی‌که تمامی این عوامل با هم بررسی شوند، می‌توانند به تصمیم‌گیری بهتر پزشکان کمک کنند؛ بنابراین در پژوهش حاضر تمامی این عوامل مورد بررسی قرار گرفته است و مدل‌های تصمیم و دانش محور بر اساس کلیه این عوامل ایجاد گردیده است. در این پژوهش نمونه‌های مربوط به ریسک فاکتورهای پرخطر سرطان کولورکتال خصوصا در ارتباط با افراد مبتلا به سرطان کولورکتال بسیار فراوان بودند و از پژوهش حذف گردیدند، بنابراین نمونه‌های بسیار فراوانی از پژوهش حذف گردیدند، در نتیجه برای داده‌کاوی از تعداد نمونه‌های تقریبا کمی استفاده گردید که از جمله محدودیت‌های این پژوهش محسوب می‌گردید. بسیاری از اطلاعات مرتبط با ریسک فاکتورها که در پرونده وجود داشتند به صورت کیفی بودند، بنابراین ممکن بود تا حدی بر عملکرد سیستم تاثیر بگذارند.

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

از محدودیت‌های دیگر این پژوهش، تعداد ریسک فاکتورهای تعیین خطر ابتلا به سرطان کولورکتال بود، البته با وجود متخصصان آگاه و با تجربه در حوزه‌ی گوارش و کولوکتال، مهم‌ترین ریسک فاکتورها انتخاب شدند اما باز هم ممکن است ریسک فاکتوری در پژوهش مورد بررسی قرار نگرفته باشد و این امر بر عملکرد الگوریتم‌های داده‌کاوی تاثیر گذاشته باشد.

نتیجه‌گیری

سرطان کولورکتال از جمله سرطان‌هایی است که علایم و نشانه‌های آن زمانی بروز می‌کند که در بدن فرد بسیار پیشرفت کرده و میزان بقای بیماران در این مرحله بسیار کاهش پیدا کند؛ بنابراین تشخیص زودهنگام این بیماری در مراحل اولیه و به تبع آن پیشگیری موثر، بسیار مهم تلقی می‌شود. کاربرد الگوریتم‌های داده‌کاوی در زمینه‌ی تشخیص و پیش‌بینی بیماری‌ها در حیطه‌ی پزشکی بسیار وسیع است. بنابراین در این پژوهش از الگوریتم‌های داده‌کاوی در زمینه‌ی پیش‌بینی خطر سرطان کولورکتال در افراد استفاده گردید. نتایج حاصل از مقایسه‌ی عملکرد الگوریتم‌های داده کاوی نشان داد که الگوریتم J-48 با میزان F-Measure ۸۸/۹٪، نسبت به سایر الگوریتم‌های داده کاوی بالاترین میزان کارایی را داشته است؛ بنابراین انتظار می‌رود که به‌عنوان مدلی مناسب برای تعیین خطر ابتلا به سرطان کولورکتال در نظر گرفته شود. پیشنهاد می‌شود که تولید مدل تصمیم تعیین خطر ابتلا به سرطان کولورکتال با تعداد نمونه‌های بیشتری انجام شود. همچنین طبق نتایج حاصل از رگرسیون لجیستیک، هفت متغیر: میزان فعالیت ورزشی در روز، میزان مصرف میوه و سبزی، میزان مصرف چربی در روز، میزان مصرف گوشت در روز، سابقه‌ی ژنتیک، مصرف قرص آسپرین و مصرف سیگار، به‌عنوان مهم‌ترین ریسک فاکتورها در نظر گرفته شدند؛ بنابراین پیشنهاد می‌شود که با تغییر سبک زندگی افراد میزان خطر این بیماری را در جامعه کاهش داد. سرطان کولورکتال از جمله بیماری‌هایی است که ریسک فاکتورهای بسیار فراوانی دارد؛ بنابراین پیشنهاد می‌شود تا حد امکان از مهم‌ترین ریسک فاکتورهای تعیین خطر سرطان کولورکتال با مشورت چندین متخصص آگاه و باتجربه استفاده گردد تا از یک سو تمام ریسک فاکتورهای مهمی که در تعیین خطر سرطان کولورکتال نقش دارند در نظر گرفته شوند و از سوی دیگر عملکرد الگوریتم‌های داده کاوی به دلیل افزونه شدن بیش از حد حجم مجموعه داده افت نکند .

تشکر و قدردانی

این مقاله بخشی از پایان‌نامه مقطع کارشناسی‌ارشد در رشته‌ی فناوری اطلاعات سلامت، دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران با شماره ۹۸/ف/۲۸۰/۳ و شناسه‌ی اخلاق IR.TUMS.SPH.REC.1398.191 می‌باشد.

پژوهشگر بر خود لازم می‌داند از پزشکان و کارکنان بخش مدارک پزشکی بیمارستان امام خمینی و کلینیک طوبی شهرستان ساری به پاس همکاری صمیمانه و شکیبایی در جمع‌آوری داده‌ها و تکمیل پرسش‌نامه قدردانی و تشکر نماید.

منابع

1. National Cancer Institute. Large intestine in NCI dictionary of Cancer terms. Available at: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/large-intestine?redirect=true. 2011.

2. National Cancer Institute. What is Cancer? Available at: https://www.cancer.gov/about-cancer/understanding/what-is-cancer. 2015.

3. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi REM & Corcione F. Worldwide burden of colorectal cancer: A review. Updates in Surgery 2016; 68(1): 7-11.

4. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RG, Barzi A, et al. Colorectal Cancer statistics, 2017. CA: A Cancer Journal for Clinicians 2017; 67(3): 177-93.

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

5. Haggar FA & Boushey RP. Colorectal Cancer epidemiology: Incidence, mortality, survival, and risk factors. Clinics in Colorectal Surgery 2009; 22(4): 191-7.

6. Buccafusca G, Ilaria P, Antonino T, Sebastiano RG & Paolo T. Early colorectal cancer: Diagnosis, treatment and survivorship care. Critical Reviews in Oncology/Hematology 2019; 136(1): 20-30.

7. Vakili M, Aghakoochak A, Pirdehghan A, Shiryazdy M & Saffarmoghadam A. The survival rate of patients with colorectal cancer in Yazd during 2001-2011. Journal of Shahid Sadoughi University of Medical Sciences 2014; 22(3): 1187-95[Article in Persian].

8. Roudbari M, Abbasiasl M, Brafei F, Gohari MR & Khodabakhshi R. Survival analysis of colorectal cancer patients and its prognostic factors using cox regression. Razi Journal of Medical Sciences (RJMS) 2015; 22(130): 21-8[Article in Persian].

9. Zhang YL, Zhang ZS, Wu BP & Zhou DY. Early diagnosis for colorectal cancer in China. World Journal of Gastroenterology 2002; 8(1): 21-5.

10. Innocent PR, John RI & Garibaldi JM. Fuzzy methods for medical diagnosis. Applied Artificial Intelligence 2004; 19(1): 69-98.

11. Russell SJ & Norvig P. Artificial intelligence: A modern approach. New Jersey: Prentice Hall; 2009: 34-59.

12. Borana J. Applications of artificial intelligence & associated technologies. Available at: https://pdfs.semanticscholar.org/d5b0/61e6565ce421b4b0b7d56296e882085dc308.pdf. 2016.

13. Ghorbani R & Ghousi R. Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. International Journal of Data Network Science 2019; 3(2): 47-70.

14. Shin A, Joo J, Yang HR, Bak J, Park Y, Kim J, et al. Risk prediction model for Colorectal cancer: National health insurance corporation study, Korea. PloS One 2014; 9(2): e88099.

15. Setareh S, Zahiri Esfahani M, Zare Bandamiri M, Raeesi A & Abbasi R. Using data mining for survival prediction in patients with colon cancer. Iranian Journal of Epidemiology 2018; 14(1): 19-29[Article in Persian].

16. Fatemian AS, Abdolvand N, Salimzadeh H & Delavari A. Predictive factors of advanced colonic adenomas and cancer using data mining. Middle East Journal of Digestive Diseases 2019; 11(4): 192-8.

17. Wen J, Zhang X, Xu Y, Li Z & Liu L. Comparison of AdaBoost and logistic regression for detecting colorectal cancer patients with synchronous liver metastasis, Singapore: International Conference on Biomedical and Pharmaceutical Engineering (IEEE), 2009.

18. Burke C & Van RS. Colorectal cancer screening: Making sense of the different guidelines. Cleveland Clinic Journal of Medicine 1999; 66(5): 303-11.

19. Mustafa M, Menon J, Muniandy R, Illzam E, Shah M & Sharifa A. Colorectal cancer: Pathogenesis, management and prevention. (IOSR) Journal of Dental and Medical Sciences 2016; 15(5): 94-100.

20. Wu HC, Chang CJ, Lin CC, Tsai MC, Chang CC & Tseng MH. Developing screening services for colorectal cancer on android smartphones. Telemedicine Journal E-Health 2014; 20(8): 687-95.

21. Robertson R, Campbell C, Weller DP, Elton R, Mant D, Primrose J, et al. Predicting colorectal cancer risk in patients with rectal bleeding. British Journal of General Practice 2006; 56(531): 763-7.

22. Battista A, Battista RA, Battista F, Cinquanta L, Iovane G, Corbisieri M, et al. Development of a new mathematical tool for early colorectal cancer diagnosis and its possible use in mass screening. Journal of Interdisciplinary Mathematics 2019; 22(6): 811-35.

23. Anderson B, Hardin JM, Alexander DD, Meleth S, Grizzle WE & Manne U. Comparison of the predictive qualities of Three prognostic models of Colorectal cancer. Frontiers in Bioscience 2010; 2(1): 849-56.

[Downloaded from payavard.tums.ac.ir on 2024-07-21]

Decision Making Model for Colorectal Cancer Screening Based on Risk Determination Method

Raoof Nopour[1] (M.S.) - Mohammad Shirkhoda[2] (M.D.) - Sharareh Rostam Niakan Kalhori[3] (Ph.D.)

1 Master of Science in Health Information Technology, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

2 Associate Professor, Department of Cancer Surgery, Cancer Institute, Imam Khomeini Hospital Complex, Tehran, Iran

3 Associate Professor, Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

Abstract

Received: Nov 2019

Accepted: Mar 2020

Background and Aim: Colorectal cancer is one of the most common gastrointestinal cancers among human beings and the most important cause of death in the world. Based on the risk of colorectal cancer for individuals, using an appropriate screening program can help to prevent the disease. Therefore, the purpose of this study was to design a model for screening colorectal cancer based on risk factors to increase the survival rate of the disease on the one hand and to reduce the mortality rate on the other.

Materials and Methods: By reviewing articles and patients' records, 38 risk factors were detected. To determine the most important risk factors clinically, CVR(content validity ratio) was used; and considering the collected data, Spearman correlation coefficient and logistic regression analysis were applied for statistical analyses. Then, four algorithms -- J-48, J-RIP, PART and REP-Tree -- were used for data mining and rule generation. Finally, the most common model was obtained based on comparing the performance of the algorithms.

Results: After comparing the performance of algorithms, the J-48 algorithm with an F-Measure of 0.889 was found to be better than the others.

Conclusion: The results of evaluating J-48 data mining algorithm performance showed that this algorithm could be considered as the most appropriate model for colorectal cancer risk prediction.

Keywords: Decision Model, Early Detection, Colorectal Cancer, Risk Factor, Data Mining

*** Corresponding Author:**
Rostam Niakan Kalhori SH
Email :
sh-rniakank@sina.tums.ac.ir

[Downloaded from payavard.tums.ac.ir on 2024-07-21]