مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره چهارم شماره اول. ص۴۷-۳۹. ۱۳۹۶

تعیین مهم‌ترین عوامل مؤثر بر سرطان پوست غیرملانومایی با استفاده از الگوریتم‌های داده کاوی

فروغ‌السادات قاسم زاده[۱]، علی عرب خردمند[۲]، سروش دکلان[۳]، علیرضا شعبانی نژاد[۴]، عطا قراجه‌ای[۵]، کبری اطمینانی[۶*]

•دریافت مقاله: ۹۶/۱/۱۰ •پذیرش مقاله: ۹۶/۲/۲۶

مقدمه: در سال‌های اخیر، سرطان پوست غیرملانوما (NMSC) جزء سه سرطان شایع در ایران بوده است. مدیریت نامناسب این بیماری منجر به افزایش شیوع و هزینه‌های سربار اقتصادی شده است. تکنیک‌های داده کاوی به آنالیز داده‌های پرونده‌های بیماران و مدیریت صحیح بیماری‌ها کمک می‌نمایند. هدف این مطالعه کشف الگوها و روابط پنهان در داده‌های بیماران NMSC با استفاده از الگوریتم‌های داده کاوی می‌باشد.

روش: جامعه مورد بررسی در این مطالعه کاربردی، ۸۲۸ پرونده NMSC بود که طی سال‌های ۹۴-۸۶ به انستیتو کانسر بیمارستان امام خمینی (ره) تهران ارجاع شده بودند. متغیرهای دموگرافیک و ریسک فاکتورهای ابتلا به بیماری با استفاده از الگوریتم K-Means خوشه‌بندی شدند. همچنین از الگوریتم Apriori برای استخراج قوانین انجمنی و تعیین شاخص‌های مشترک بیماران با درجه اطمینان بالای ۰/۹ استفاده گردید.

نتایج: بیماران NMSC با توجه به متغیرهای مورد بررسی در چهار خوشه توزیع شدند و سه عامل مهم تأثیرگذار بر بیماری، BMI غیرنرمال، شغل‌های با ریسک بالا و سابقه قبلی سرطان مشخص شد. با استفاده از قوانین انجمنی هفت قانون مورد تأیید قرار گرفت و بیشترین ارتباط میان عوامل سابقه قبلی بیماری، موضع درگیر، عود و نوع سرطان پوست غیرملانوما دیده شد.

نتیجه‌گیری: این مطالعه برای اولین بار مهم‌ترین عوامل مؤثر بر NMSC را با استفاده از داده کاوی تعیین نمود. این عوامل بایستی در انجام خودآزمایی‌ها و یا آزمایش‌های غربالگری پوست در گروه‌های پرخطر مدنظر قرار بگیرند. همچنین در مطالعات آینده بایستی مشارکت عوامل فیزیولوژیک، اکولوژیک و ژنتیک در ایجاد سرطان پوست تواماً داده کاوی شوند.

کلید واژه‌ها: سرطان پوست غیر ملانومایی، داده کاوی، خوشه بندی، قوانین انجمنی، ریسک فاکتور

•ارجاع: قاسم‌زاده فروغ السادات، عرب خردمند علی، دکلان سروش، شعبانی نژاد علیرضا، قراجه‌ای عطا، اطمینانی کبری. تعیین مهم‌ترین عوامل مؤثر بر سرطان پوست غیرملانومایی با استفاده از الگوریتم‌های داده کاوی. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۶؛ ۴(۱): ۴۷-۳۹.

۱.کارشناس ارشد انفورماتیک پزشکی، گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.

۲. متخصص جراحی عمومی، دانشیار گروه جراحی سرطان، مرکز تحقیقات سرطان، انستیتو کانسر ایران، دانشگاه علوم پزشکی تهران، تهران، ایران.

۳. متخصص پوست، بیمارستان پوست رازی، گروه پوست، دانشکده علوم پزشکی تهران، تهران، ایران.

۴. کارشناس ارشد حشره شناسی، گروه گیاه پزشکی، دانشکده کشاورزی، دانشگاه صنعتی شاهرود، سمنان، ایران.

۵. متخصص جراحی دهان و فک و صورت، استادیار گروه جراحی دهان و فک و صورت، دانشکده دندانپزشکی، مرکز تحقیقات سرطان، انستیتو کانسر ایران، دانشگاه علوم پزشکی تهران، تهران، ایران.

۶. دکترای نرم افزار کامپیوتر، استادیار گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

*** نویسنده مسئول:** مشهد، دانشگاه علوم پزشکی مشهد، دانشکده پزشکی، گروه انفورماتیک پزشکی

•شماره تماس: ۰۹۱۵۵۱۱۸۳۱۲ • **Email:** EtminaniK@mums.ac.ir

مقدمه

پوست یکی از بزرگ‌ترین اندام‌های بدن از نظر سطح و وزن می‌باشد. پوست دارای دو لایه اپیدرم و درمیس می‌باشد. در زیر لایه درمیس، هیپودرم یا بافت چربی زیرپوستی قرار دارد. پوست دارای سه عملکرد اصلی حفاظت، تنظیم و احساس می‌باشد؛ بنابراین هر عاملی که پوست را تحت تأثیر قرار دهد بر این سه عملکرد اصلی نیز تأثیر می‌گذارد. عملکرد اولیه پوست ایجاد سد حفاظتی در برابر فشار و تحریکات مکانیکی، تغییرات حرارتی، میکروارگانیسم‌ها، تشعشع و مواد شیمیایی است. پوست به عنوان یک اندام تنظیمی، جوانب مختلف فیزیولوژی شامل دمای بدن را از طریق تعریق و موها و نیز تغییرات در گردش محیطی و تعادل مایعات را از طریق تعریق تعدیل می‌نماید. همچنین به عنوان منبع مهمی برای سنتز ویتامین D محسوب می‌شود. پوست حاوی شبکه گسترده‌ای از سلول‌های عصبی است که تغییرات محیطی مانند گرما، سرما، لمس و درد را شناسایی و منتقل می‌نماید [۱].

بیماری‌های مختلف، پوست را تحت تأثیر قرار می‌دهند که در این بین، سرطان پوست شدت بیشتری دارد [۲،۳]. به طور کلی سرطان یکی از مهم‌ترین علل مرگ‌ومیر در سراسر جهان می‌باشد و پس از بیماری‌های قلبی و عروقی دومین عامل شایع در کشورهای توسعه یافته و سومین عامل مرگ در کشورهای کمتر توسعه یافته می‌باشد. زیرا درمانی که بتواند مشکلات مربوط به این بیماری دردناک را حل کند، وجود ندارد [۴،۵]. در ایالات متحده آمریکا، اروپا و استرالیا به ترتیب در حدود ۲۵، ۲۰ و ۴۵ درصد از سرطان‌های گزارش شده در طول یک سال، سرطان‌های پوست می‌باشند [۶-۱۳]. در ایران مطالعات محدودی در مورد میزان بروز سرطان پوست صورت گرفته است. مشخص شده است که بروز این نوع سرطان ۱۰ تا ۱۵ مورد جدید در هر ۱۰۰۰۰۰ نفر می‌باشد [۱۴-۱۷]. طبق گزارش‌های ارائه شده توسط مرکز مدیریت بیماری‌های غیر واگیر اداره پیشگیری و کنترل سرطان وزارت بهداشت جمهوری اسلامی ایران، سرطان پوست غیر ملانومایی در سال ۸۸ در مردان رتبه اول و در زنان در رتبه دوم قرار داشته است و میزان بروز استاندارد شده سنی آن برای مردان ۱۸/۹۳ و زنان ۱۳/۰۹ بوده است. در سال ۸۸ در استان تهران سرطان پوست غیر ملانومایی در بین مردان (۱۴٪/۳) رتبه اول و در بین زنان (۹٪/۳) رتبه سوم را داشته است [۱۵].

سه نوع مهم سرطان پوست بر اساس نوع سلول‌هایی که دچار رشد سرطانی می‌شوند، سرطان سلول پایه BCC (Basal Cell Carcinoma)، سرطان سلول سنگفرشی SCC(Squamous Cell Carcinoma) و ملانوما (Melanoma) می‌باشند. به مجموع دو نوع سرطان اول یعنی BCC وSCC، سرطان پوست غیر ملانومایی (Non NMSC(Melanoma Skin Cancer اطلاق می‌گردد [۴].

سرطان BCC رایج‌ترین نوع سرطان پوست در افراد سفید پوستی است که در آب هوای ملایم زندگی می‌کنند و ۸۰٪ از سرطان‌های غیر ملانومایی را به خود اختصاص می‌دهد. همانند دیگر سرطان‌های پوست، قرار گرفتن در برابر اشعهUV، مهم‌ترین ریسک فاکتور برای گسترش آن می‌باشد [۱۸،۱۹].

دومین نوع رایج سرطان پوست، سرطان SCC می‌باشد که از سلوهای کراتین که اصلی‌ترین سلول‌ها در لایه اپیدرم پوست هستند، شروع می‌شود. این نوع بیماری ۲۰٪ از کل سرطان‌های غیر ملانومایی در کشورهایی مانند بریتانیا را شامل می‌شود. در مقابل BCC، SCC تمایل به رشد زیاد داشته و نه تنها رشد موضعی دارد، بلکه به غدد لنفاوی و اندام‌های دورتر از ضایعه نیز متاستاز می‌دهد. افراد با پوست لطیف چنانچه در معرض اشعهUV قرار گیرند، در تمام طول عمر خود مستعد SCC خواهند بود [۱۹].

بنابراین حفاظت از پوست بسیار حائز اهمیت بوده و کار مهم و پیچیده‌ای محسوب می‌شود. علاوه بر میزان شیوع، هزینه مالی درمان هر سرطان به طور معناداری می‌تواند بر منابع یک جامعه تأثیر بگذارد [۲۰]؛ بنابراین مردم باید بیشتر از معمول برای بیماری‌های مختلف پوست و علائم آن‌ها قدم پیشگیرانه بردارند [۲۱]، ولی متأسفانه بیشتر مردم، به اندازه‌ای که به انواع سرطان‌های دیگر توجه دارند به پوست خود توجه نمی‌کنند و هوشیار نیستند؛ بنابراین، بیمارانی که دچار سرطان پوست هستند اغلب، تا قبل از یک دوره زمانی که سرطان پیشرفت می‌کند اقدامات پزشکی مناسب انجام نمی‌دهند و این در حالی است که بیشتر سرطان‌های پوست اگر در مراحل اولیه تشخیص داده شده و درمان شوند، بهبود می‌یابند [۲۲].

حجم بالایی از داده‌ها در پرونده‌های بیماران گردآوری و ذخیره شده است و تنها افرادی که با این پرونده سروکار دارند به خوبی می‌دانند که اطلاعات بسیار مفیدی در این داده‌ها پنهان هستند. در واقع بین جمع‌آوری داده‌ها تا تفسیر آن‌ها به صورت علمی شکاف وسیعی وجود دارد. داده‌کاوی سلامت، حوزه میان رشته‌ای جدید و در حال رشدی است که با تلفیق حوزه‌های مختلف (پزشکی، پایگاه داده، آمار، یادگیری ماشین و

...) و تحلیل‌های عمیق، دانش ارزشمند نهفته در این داده‌ها را آشکار نموده و به توسعه تحقیقات پزشکی و تصمیم‌گیری‌های بالینی کمک می‌نماید [۲۳-۲۵].

داده کاوی سلامت دارای کاربردهای بسیار وسیع و در عین حال حساس و حیاتی است. با توجه به این که داده‌ها با سلامت و بهداشت انسان سروکار دارند، با ارزش‌ترین و حساس‌ترین داده‌ها برای کاوش و تحلیل هستند، تحلیل و کسب دانش از آن‌ها باید با درجه بالایی از دقت و حساسیت صورت گیرد [۲۶].

طی چندین سال اخیر مطالعات متعددی در جهت تشخیص، پیشگویی و پیشگیری بیماری‌های پوست انجام شده است. در این راستا تکنیک‌های آماری، هوش مصنوعی و الگوریتم‌های مختلف داده کاوی برای تجزیه و تحلیل اطلاعات مربوط به این بیماران استفاده شده‌اند. مطالعات داده کاوی انجام شده در مورد سرطان پوست در سه زمینه کلی تشخیص (Diagnosis)، پیشگویی (Prediction) و پیشگیری (Preventation) قابل طبقه‌بندی هستند (شکل ۱). در هر کدام از این گروه‌ها، بسته به نیاز و کارایی از الگوریتم‌های مختلف شبکه عصبی، درخت تصمیم، قوانین وابستگی و طبقه بندی استفاده شده است.

شکل ۱: سه گروه اصلی مطالعات انجام شده در زمینه سرطان پوست به کمک داده کاوی

با توجه به شیوع بالای NMSC، تأثیری که بر سال‌های مفید زندگی فرد می‌گذارد و نیز رنج عاطفی و جسمانی حاصل از آن، پیشگیری از این بیماری امری ضروری به نظر می‌رسد [۲۹] اگرچه این سرطان یکی از شایع‌ترین سرطان‌ها است، ولی درعین حال یکی از قابل پیشگیری‌ترین سرطان‌ها نیز می‌باشد [۳۱،۳۰]؛ بنابراین با گردآوری داده‌های پرونده‌های بیماران NMSC و آنالیز آن‌ها با کمک تکنیک‌های داده‌کاوی می‌توان به مدیریت صحیح بیماری از نظر تشخیص، پیشگویی و پیشگیری کمک نمود.

هدف این مطالعه گذشته‌نگر و مقطعی، کشف الگوها و روابط پنهان در داده‌های بیماران NMSC با استفاده از الگوریتم‌های داده‌کاوی می‌باشد تا بدین وسیله مهم‌ترین عوامل مؤثر بر این نوع سرطان در ایران شناسایی و به دنبال آن قدم‌های مؤثر در پیشگیری بیماری برداشته شوند.

روش بررسی

داده‌های مورد استفاده در این مطالعه شامل اطلاعات مندرج در پرونده‌های بیماران مبتلا به سرطان پوست غیر ملانوما بود که طی سال‌های ۸۶-۹۴ به مدارک پزشکی مرکز تحقیقات

سرطان بیمارستان امام خمینی (ره) تهران ارجاع شده بودند [۳۲]. متغیرهای مورد نظر شامل سن افراد، جنسیت، شهر محل زندگی، شغل،(BMI(Body Mass Index، نوع سرطان غیرملانوما (SCC, BCC)، سابقه بیماری قبلی، موضع درگیر و تعداد عود بودند که به صورت دستی از پرونده‌ها استخراج و در فایل تهیه شده با نرم افزار Excel نسخه ۲۰۱۰ وارد شدند. از ۸۶۵ پرونده مورد بررسی که با روش تصادفی ساده انتخاب شده بودند، پس از پیش پردازش، رکوردهایی که حداقل ۳ متغیر مفقود داشتند از مطالعه حذف شدند و مطالعه با ۵۵۷ رکورد ادامه یافت. به منظور پیش پردازش داده‌ها دو فاز پاک‌سازی و آماده‌سازی صورت گرفت. در فاز پاک‌سازی، بعضی فیلدهای خالی با توجه به فیلدهای دیگر پر شدند، برای مثال فیلد "موضع درگیر" با توجه به فیلد "نوع درمان" که نام قسمت جراحی شده را دست نوشته بود تکمیل شد. در مرحله خوشه‌بندی رکوردهای با یک یا دو فیلد مفقوده با استفاده از الگوریتم‌های تخمین بازسازی شدند. ناهمگونی در داده‌ها (کلمات متفاوت با یک مضمون) و اشتباهات تایپی نیز تصحیح شدند. در فاز دوم ایجاد ویژگی جدید، تبدیل مقیاس‌ها، نرمال‌سازی و تبدیل نوع ویژگی انجام شد. بدین منظور برای دستیابی به حجم کوچک‌تری از داده‌ها، دو فیلد قد و وزن در فیلد BMI با یکدیگر ادغام شدند. برای اجرای الگوریتم خوشه‌بندی، به منظور سادگی محاسبات و مقایسه بهتر داده‌ها ابتدا مقیاس‌های کیفی به کمی تغییر یافتند و پس از آن نرمال‌سازی داده‌های کمی با استفاده از فرمول ۱ انجام شد.

$$x = \frac{X - MIN}{MAX - MIN}$$

فرمول (۱)

در برخی موارد لازم بود نوع داده برای همگون‌سازی و کاهش پراکندگی تغییر یابد. تبدیل نوع برای متغیرهای سن، محل زندگی، شغل، موضع درگیر و سابقه قبلی بیماری طبق مطالعه پایه انجام شده توسط قاسم زاده و همکاران، صورت گرفت [۳۲].

جهت خوشه‌بندی بیماران از الگوریتم K-means استفاده شد. دسته‌بندی داده‌ها در خوشه‌های معنادار به طوری که محتویات هر خوشه ویژگی‌های مشابه و در عین حال نسبت به اشیاء دیگر در سایر خوشه‌ها غیر مشابه باشند را خوشه‌بندی می‌گویند. خوشه‌بندی به منظور کشف گروه‌بندی طبیعی درون داده‌ها به کار می‌رود و هیچ خروجی از پیش تعیین شده یا فیلد هدفی به منظور پیش‌بینی وجود ندارد. به این مدل‌ها اغلب مدل‌های یادگیری بدون نظارت گفته می‌شود. در الگوریتم k-means تعداد n مشاهده در k دسته قرار می‌گیرند به طوری که هر مشاهده در خوشه‌ای قرار می‌گیرد که کم‌ترین فاصله را با مرکز خوشه دارد [۳۳] در الگوریتم K-means وزن متغیرها یکسان در نظر گرفته شد و از معیار شباهت فاصله اقلیدسی برای خوشه‌بندی استفاده گردید.

برای بررسی ویژگی‌ها یا خصوصیاتی که با یکدیگر همراه بوده و به دنبال آن، استخراج قواعد از میان این خصوصیات، از الگوریتم Apriori استفاده شد. این الگوریتم یکی از پرکاربردترین الگوریتم‌های قوانین انجمنی (Association rules) می‌باشد. قوانین انجمنی ارتباطات جذاب و پرتکرار بین مجموعه بزرگی از داده‌ها را کشف می‌کند که این ارتباطات می‌تواند به تصمیم گیرندگان کمک کند.

برای انجام خوشه‌بندی و کشف قوانین انجمنی از نرم‌افزار Weka نسخه ۳,۶ استفاده شد.

نتایج

با استفاده از الگوریتم K-means ویژگی‌های دموگرافیک و ریسک فاکتورهای فردی و محیطی تمام بیماران بررسی و رکوردها در چهار خوشه به ترتیب با فراوانی‌های ۴۰، ۳۶، ۲۱ و ۳٪ دسته‌بندی شدند. در خوشه اول و با بیشترین مقدار فراوانی، زنان در دهه سنی ۷۰ سال و مبتلا به سرطان از نوع BCC با ۷۸٪ BMI رنج غیرنرمال، محل زندگی در عرض‌های جغرافیایی بالا، ۱۶٪ سابقه قبلی سرطان و شغل با ریسک پایین قرار داشتند. در خوشه دوم مردان در دهه سنی ۶۰ سال و مبتلا به سرطان از نوع BCC با ۶۶٪ BMI رنج غیر نرمال، محل زندگی در عرض‌های جغرافیایی بالا، ۲۶٪ سابقه قبلی سرطان و شغل با ریسک بالا قرار داشتند. در خوشه سوم مردان در دهه سنی ۶۰ سال و مبتلا به سرطان از نوع BCC با ۶۳٪ BMI رنج غیرنرمال، محل زندگی در عرض‌های جغرافیایی بالا، با سابقه قبلی سرطان و شغل با ریسک پایین قرار داشتند. در خوشه چهارم با کم‌ترین مقدار فراوانی، زنان در دهه سنی ۳۰ سال و مبتلا به سرطان از نوع BCC با ۶۹٪ BMI رنج غیرنرمال، محل زندگی در عرض‌های جغرافیایی بالا، ۱۳٪ سابقه قبلی سرطان و شغل با ریسک پایین قرار داشتند.

جزئیات خوشه بندی متغیرهای بیماران مبتلا به NMSC در جدول ۱ نشان داده شده و اثرگذارترین عامل روی سرطان پوست در هر دسته با رنگ قرمز مشخص شده است.

جدول ۱: تحلیل خوشه‌بندی داده‌های بیماران مبتلا به سرطان پوست غیرملانوما با استفاده از الگوریتم K-Means

خوشه‌ها / عوامل تأثیرگذار	خوشه اول	خوشه دوم	خوشه سوم	خوشه چهارم
فراوانی خوشه	۴۰٪	۳۶٪	۲۱٪	۳٪
دهه سنی بیماران	۷۰ سال	۶۰ سال	۶۰ سال	۳۰ سال
جنسیت	۵۶٪ زن	۹۹٪ مرد	۸۶٪ مرد	۶۹ ٪ زن
نوع سرطان غیرملانوما	BCC ۵۲٪	BCC ۵۴٪	BCC ۵۶٪	BCC ٪ ۶۹
BMI غیر نرمال	۷۸٪	۶۶٪	۶۳٪	۶۹٪
زندگی در عرض جغرافیایی پایین	۱۳٪	۱۲٪	۸٪	۱۲٪
سابقه قبلی سرطان	۱۶٪	۲۶٪	۱۰۰٪	۱۳٪
شغل با ریسک بالا	۰٪	۵۵٪	۰٪	۰٪

در مرحله استخراج قوانین انجمنی یک بار آنالیز با تمام رکوردها انجام شد و ۱۰۰۰ قانون استخراج شد. بار دیگر با حذف رکوردهای با حداقل ۳ متغیر مفقوده، آنالیز انجام شد و این بار با ۵۵۷ رکورد، تعداد ۴۶۱ قانون با درجه اطمینان بالای ۰/۹ به دست آمد. همه قوانین به دست آمده دارای معنا و مفهوم مناسبی نبودند. با انجام فیلتر، قوانین بی‌معنا و مفهوم حذف شدند و با استفاده از نظر متخصص پوست تعدادی از آن‌ها تأیید شد که در جدول ۲ گزارش شده است.

معیار Confidence مقداری عددی بین صفر و یک می باشد، که هر چه این عدد بزرگ‌تر باشد بر کیفیت قانون افزوده خواهد شد. محاسبه معیار اطمینان با استفاده از فرمول ۲ انجام شد که مدار عددی آن در این مطالعه ۰/۹ در نظر گرفته شده است.

$$Conf\ (A \rightarrow B) = \frac{SUP(A \cup B)}{SUP(A)}$$

فرمول (۲)

جدول ۲: قوانین انجمنی استخراج شده از اطلاعات بیماران مبتلا به سرطان پوست غیرملانوما

	قانون	نتیجه	درجه اطمینان (Confidence)
۱	از ۲۷۶ نفر بیمار BCC که در عرض های جغرافیایی بالا زندگی می‌کنند،	در ۲۶۹ مورد، موضع درگیر بیماری، ناحیه سر و گردن بود.	۰/۹۷
۲	از ۲۵۷ نفر بیمار با سابقه ی بیماری‌های قلبی- عروقی که سابقه سرطان دیگری ندارند،	در ۲۴۲ مورد سابقه بیماری متابولیک هم وجود نداشت.	۰/۹۴
۳	از ۲۱۰ نفر بیمار BCC که سابقه بیماری‌های قلبی- عروقی ندارند،	در۱۹۶ نفر سابقه بیماری متابولیک هم وجود نداشت.	۰/۹۳
۴	از ۲۲۲ نفر مرد که سرطان پوست در ناحیه سر و گردن دارند و سابقه سرطان دیگری ندارند،	در ۲۰۲ مورد تنها یک‌بار عود سرطان پوست وجود داشت.	۰/۹۱
۵	از ۲۴۲ بیمار که سابقه سرطان دیگر و سابقه بیماری‌های متابولیک و سابقه بیماری‌های قلبی عروقی ندارند،	در۲۲۰ مورد تنها یکبار عود سرطان پوست وجود داشت.	۰/۹۱
۶	از ۷۰ نفر بیمار مرد که در دهه ۷۰ سالگی هستند و سابقه سرطان دیگری ندارند،	در ۶۶ نفر تنها یک‌بار عود سرطان پوست وجود داشت.	۰/۹۴
۷	از ۱۳۷ بیمار مرد مبتلا به BCC در ناحیه سرو گردن هستند که سابقه سرطان دیگری هم ندارند،	در ۱۲۵ نفر آنان تنها یک‌بار عود سرطان پوست وجود داشت.	۰/۹۱

بحث و نتیجه‌گیری

مطالعات انجام شده در مورد سرطان‌های پوست در دنیا نشان می‌دهند که این سرطان در حدود ۲۰ تا ۴۰ درصد از انواع سرطان‌ها را شامل می‌شود [۶-۸،۳۴]. اگرچه شیوع سرطان پوست در ایران کمتر از کشورهای غربی است، ولی با این وجود شیوع آن رو به افزایش است و طبق آخرین گزارش‌های وزارت بهداشت این سرطان در میان زنان و مردان در کشور ما جزء سه سرطان اول بوده است. به تأخیر افتادن تشخیص و مدیریت نامناسب بیماری منجر به افزایش شیوع، هزینه‌های سربار اقتصادی و نهایتاً از دست دادن زندگی می‌شود. همه این نکات بر مشکل روبه رشد سلامت عمومی پوست تأکید دارند.

در مطالعه حاضر الگوها و روابط پنهان در داده‌های بیماران NMSC مراجعه کننده به انستیتو کانسر با استفاده از الگوریتم های K-means و Apriori مورد بررسی قرار گرفتند. با بررسی ۴ خوشه ایجاد شده، سه عامل مهم تأثیرگذار بر سرطان پوست، BMI غیرنرمال، سابقه قبلی سرطان و شغل‌های با ریسک بالا معرفی شدند. چاقی و اضافه وزن در دیگر مطالعات نیز به عنوان یکی از ریسک فاکتورها برای ابتلای به انواع سرطان‌ها محسوب شده است. همچنین نشان داده شده که با افزایش چاقی خطر ابتلا به انواع سرطان پوست نیز افزایش می یابد [۳۵] هر نوع بیماری زمینه‌ای در افراد، میزان بروز سرطان را ازدیاد می‌بخشد؛ بنابراین افرادی که به دلیل ابتلا به هر نوع

بیماری در حالت عام و سرطان در حالت خاص سیستم دفاعی ضعیف شده داشته باشند، برای ابتلا به سرطان پوست مستعدتر خواهند بود [۱۸،۱۹،۲۲] این مطلب در مطالعه حاضر نیز تأیید شد.

در مطالعات قبلی شغل و و میزان مواجهه افراد با اشعه‌ها و مواد شیمیایی زیان‌بار یک ریسک فاکتور مهم در ابتلا به سرطان پوست شناخته شده است. در آن مطالعات علت ۶۵ تا۹۰ درصد از سرطان‌های پوست، قرار گرفتن در برابر اشعه UV خورشید معرفی شده است [۷،۱۹]. مطالعه حاضر نیز شغل‌های افرادی که بیش از ۶-۷ ساعت در معرض اشعه آفتاب بودند، شغل‌های با ریسک بالا معرفی شده بودند [۳۲]، که در خوشه‌بندی انجام شده در این مطالعه نیز شغل‌های با ریسک بالا یکی از عوامل تأثیرگذار بر سرطان پوست شناخته شد.

از چهار خوشه ایجاد شده در این مطالعه، برای خوشه اول یعنی زنان ۷۰ سال، BMI غیرنرمال عامل مؤثر بر سرطان شناخته شد. دلیل احتمالی آن این است که با افزایش سن میزان چاقی در میان زنان بیشتر می‌شود [۳۶]. در خوشه‌های دوم و سوم، مردان ۶۰ سال قرار داشتند که یکی از عوامل تأثیر گذار در این خوشه‌ها شغل‌های با ریسک بالا معرفی شده بود. دلیل احتمالی آن شاید این نکته باشد که مردان میان سال و مسن نسبت به مردان جوان‌تر و زنان در طول زندگی خود مدت زمان بیشتری در معرض اشعه آفتاب قرار داشته‌اند. در آخرین خوشه زنان جوان و مبتلا به BCC قرار گرفته بودند که در آن‌ها چاقی و اضافه وزن به عنوان مهم‌ترین ریسک فاکتور سرطان پوست شناخته شد. این خوشه نشان داد که شیوع افزایش وزن در سنین پایین‌تر خانم‌ها در حال شکل‌گیری است. همچنین در این خوشه، افراد با پایین‌ترین رنج سنی قرار داشتند پس بایستی برای آموزش مراقبت از پوست و خودآزمایی این افراد اقدامات جدی‌تری صورت بگیرد.

در بررسی قوانین انجمنی بیشترین ارتباط میان عود سرطان پوست، سابقه قبلی بیماری موضع درگیر و نوع سرطان پوست غیرملانوما دیده شد. هر سرطانی دارای عوارضی می‌باشد، یکی از عوارض بالینی برای سرطان‌ها تأثیر آن بر روی سیستم قلبی عروقی و سیستم تنفس بدن می‌باشد [۳۷] از طرف دیگر یکی از عوامل تشدید کننده سرطان پوست، اختلالات متابولیک مثل دیابت یا بیماری‌های تیروئید معرفی شده است. این مطالب نشان از این موضوع می‌توانند داشته باشند که میان عود سرطان پوست و بیماری‌های متابولیک و قلبی–عروقی ارتباطی وجود دارد [۳۸]. در مطالعه اولیه انجام شده توسط قاسم‌زاده و همکاران نیز نشان داده شد که میان نوع سرطان غیرملانوما و موضع درگیر اختلاف معناداری از نظر آماری وجود دارد [۳۲].

مطالعات قبلی استفاده از داده کاوی در زمینه سرطان پوست، بیشتر مربوط به پیش‌بینی و تشخیص بیماری بودند و گزارش-های منتشر شده از این مطالعات در رابطه با دقت الگوریتم‌ها و کاربرد آن‌ها می‌باشد. پژوهش‌های انجام شده در رابطه با فراوانی ویژگی‌های بیماران مبتلا به سرطان پوست غیرملانوما با استفاده از روش‌های آماری انجام شده بودند و در حوزه دسته بندی ویژگی‌های بیماران با استفاده از الگوریتم‌های داده کاوی مطالعه‌ای در دسترس نبود.

مطالعه حاضر با بهره‌گیری از الگوریتم‌های K-means و Apriori ارتباط بین ویژگی‌ها و داده‌های بیماران مبتلا به سرطان پوست را مورد بررسی قرار داد. در مطالعات مشابه در زمینه سرطان پوست از الگوریتم‌های خوشه‌بندی و قوانین انجمنی با اهداف مختلف استفاده شده است.

در مطالعه‌ای که توسط Ahmed و همکاران برای پیشگویی سرطان پوست انجام شده بود. در آن مطالعه، داده‌های ۲۰۰ نفر از بیماران مبتلا به سرطان پوست و غیر سرطانی طبق عوامل خطر سرطان پوست (سن، جنسیت، وراثت، فعالیت در فضای باز، کار کردن در صنایع، رنگ پوست، سابقه در کودکی، آزمایش‌های سلامت قبلی، مصرف داروهای ضد حساسیت، سیگار کشیدن، عادات غذایی، چاقی، عوامل ژنتیکی، محیط، مصرف الکل، رادیو تراپی و در معرض مواد شیمیایی بودن) از مراکز تشخیصی مختلف دریافت و اطلاعات تکراری و مفقود آن‌ها پیش‌پردازش شده بودند. بعد از آن، از الگوریتم K-means clustering برای دسته‌بندی داده‌ها استفاده شده بود تا داده‌های مربوط و غیر مربوط با سرطان پوست تفکیک و الگوهای تکرار شونده مهم با استفاده از الگوریتم ابداعی MAFIA حاصل شوند. از الگوهای مهم به دست آمده سیستمی پیاده‌سازی شده بود که سطح ریسک سرطان را به آسانی و مقرون به صرفه از نظر زمان و هزینه پیشگویی می-نمود. این سیستم به صورت نرم‌افزار برخط در دسترس عموم قرار دارد تا همه افراد بتوانند سطح ریسک سرطان پوست خود را بررسی کنند [۴].

Thangaraju و Deepa سیستمی برای پیشگویی ریسک سرطان ملانومای پوست پیاده‌سازی کردند. هدف اصلی آن‌ها، ساخت سیستمی بود که مورد استفاده همه افراد قرار بگیرد و افراد بتوانند سطح خطر ابتلا به سرطان پوست را برای خود

بررسی نمایند. داده‌های بیماران سرطانی و غیر سرطانی با توجه به ریسک فاکتورها از مراکز تشخیصی مختلف جمع‌آوری شده بودند. الگوریتم K-means clustering برای جداسازی دسته بیماران از افراد غیر بیمار استفاده شده بود و در نهایت الگوریتم MAFIA برای ایجاد الگوهای پرتکرار مهم به کار رفته بود. سیستم طراحی شده علاوه بر تعیین سطح خطر برای فرد، پیشنهاداتی نیز به او می‌داد که این پیشنهادات نسبت به مطالعات موجود، راحت‌تر، کم هزینه‌تر و از نظر زمان مقرون به صرفه‌تر بودند و این پیشنهادات برای پیشگیری از سرطان داشتند [۳۹].

در مطالعه Swami و Houtsma از الگوریتم SETM برای یافتن رابطه در میان مشخصات متفاوت عکس‌های سرطان پوست استفاده شده بود. در آن مطالعه ابتدا کاندیدی ایجاد و مجموعه آیتم‌های پرتکرار یافت شده بودند. در نهایت، با استفاده از این مجموعه آیتم‌ها، قوانین وابستگی مهم و کاربردی برای تشخیص و مطالعه سرطان پوست ایجاد شدند [۴۰].

اعمال روش‌های داده کاوی بر روی داده‌های پزشکی می‌تواند به عنوان سیستم‌های تصمیم یار، در تصمیم‌گیری برای انتخاب نوع درمان و یا تشخیص بیماری‌ها، به متخصصان کمک نماید. این همان فلسفه داده‌کاوی است که با شناخت درست از گذشته، آینده را پیش‌بینی می‌نماید.

در این مطالعه با توجه به دسته‌بندی و مشخص شدن عوامل تأثیرگذار روی بیماران می‌توان برای گروه‌های پرخطر راهکارهای پیشگیری کننده از جمله ترویج خودآزمایی‌های پوستی یا انجام آزمایش‌های غربالگری توصیه کرد که بتوان شیوع این بیماری و به دنبال آن هزینه‌های سربار بر جامعه سلامت را کاهش داد. در مطالعات آینده بایستی خوشه‌بندی‌ها بر اساس سایر عوامل و ویژگی‌های دخیل در سرطان پوست از جمله سوابق ژنتیکی، رنگ و نوع پوست همچنین تعداد خال و یا حساسیت‌های پوستی انجام شوند تا مشارکت عوامل فیزیولوژیک، اکولوژیک و ژنتیک در ایجاد سرطان پوست تواماً مورد ارزیابی قرار بگیرند.

References

1. Tortora G J, Grabowski SR, Tortora GJ, Roesch B. Principles of anatomy and physiology. 8th ed. U.S: Wiley; 1996.

2. Jesmin T, Ahmed K, Rahman MZ, Miah MB. Brain cancer risk prediction tool using data mining. International Journal of Computer Applications 2013;61(12): 22-7.

3. American cancer society. Skin Cancer Preventation and Early Detection. [cited 2012 Jun 16]. Available from:https://dermdetect.com/wp-content/uploads/skin_cancer_prevention.pdf

4. Ahmed K, Jesmin T, Rahman MZ. Early Prevention and Detection of Skin Cancer Risk using Data Mining. International Journal of Computer Applications 2013;62(4):1-6.

5. Priyanga A, Prakasam S. Effectiveness of Data Mining - based Cancer Prediction system (DMBCPS). International Journal of Computer Applications 2013; 83(10):11-17.

6. Mackie RM, Quinn AG. Non melanoma skin cancer and other epidermal skin tumors. In: Burns T, Breathnach S, Cox N, editors. Textbook of Dermatology. 7th ed. USA: Blackwell Scientific; 2004. p. 1-50.

7. Goldsmith LA, Katz SI, Gilchrest BA, Paller AS, Leffell DJ, Wolff K. Goldsmith LA. Fitzpatrick's Dermatology in General Medicine. In: Carucci JA, Leffell DJ, Pettersen JS. Basal cell carcinoma. Fitzpatrick's dermatology in general medicine. 2008. p. 1036-42.

8. Goldsmith LA, Katz SI, Gilchrest BA, Paller AS, Leffell DJ, Wolff K. Goldsmith LA. Fitzpatrick's Dermatology in General Medicine. In: Grossman D, Leffell DJ. Squamous cell carcinoma. 2008. p. 1028-36.

9. Alam M. Fitzpatrick's dermatology in general medicine. Arch Dermatol 2004;140(3):372.

10. Jemal A, Devesa SS, Hartge P, Tucker MA. Recent trends in cutaneous melanoma incidence among whites in the United States. J Natl Cancer Inst 2001;93(9):678-83.

11. Giles GG, Marks R, Foley P. Incidence of non-melanocytic skin cancer treated in Australia. Br Med J (Clin Res Ed) 1988;296(6614):13-7.

12. Alam M, Ratner D. Cutaneous squamous-cell carcinoma. N Engl J Med 2001;344(13):975-83.

13. Kennedy C, Bajdik CD. Descriptive epidemiology of skin cancer on Aruba: 1980-1995. Int J Dermatol 2001;40(3):169-74.

14. Noorbala MT, Kafaie P. Analysis of 15 years of skin cancer in central Iran (Yazd). Dermatol Online J 2007;13(4):1.

15. Nourbala MT. Skin cancer in Yazd. Iranian Journal of Dermatology 2007;10(1):13-9.

16. Iraji F, Arbabi N, Asilian A, Siadat AH, Keshavarz J. Incidence of non-melanoma skin cancers in Isfahan. Iranian Journal of Dermatology 2007;9(38):330-34.

17. Aghajani H, Etemad K, Goya MM, Ramezani R, Modirian M, Nadali F. Iranian Annual Cancer Registration Report 2008-2009. Ministry Of Health & Medical Education Health Deputy Center for Disease Control and Prevention, Cancer Control Office: Tandis; 2011.

18. National Library of Australia Cataloguing–in–Publication data: Lifestyle and cancer: knowledge, attitudes and behavior in NSW 2009 SHPN (CI) 120203: Cancer Institute NSW; 2012.
19. Rajpar S, Marsden J. ABC of skin cancer. UK: John Wiley & Sons; 2009.
20. Housman TS, Feldman SR, Williford PM, Fleischer AB, Goldman ND, Acostamadiedo JM, et al. Skin cancer is among the most costly of all cancers to treat for the Medicare population. J Am Acad Dermatol 2003;48(3):425-9.
21. Chang CL, Chen CH. Applying decision tree and neural network to increase quality of dermatologic diagnosis. Expert Syst Appl 2009;36(2):4035-41.
22. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005;19(2):64-72.
23. Berka P, Rauch J, Zighed DA. Data Mining and Medical Knowledge Management: Cases and Applications. 1th ed. USA: Medical Information Science Reference; 2009.
24. Zhu L, Wu B, Cao C. Introduction to medical data mining. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2003;20(3):559-62.
25. Cios KJ. Medical Data Mining and Knowledge Discovery. Physica-Verlag Heidelberg; 2000.
26. Han J, Kamber M, Pei J. Data preprocessing. Data mining: concepts and techniques. San Francisco: Morgan Kaufmann; 2006.
27. Duch W, Grabczewski K, Adamczak R, Grudzinski K, Hippe ZS. Rules for melanoma skin cancer diagnosis; 2001. Available from: http://www.fizyka.umk.pl/ftp/pub/papers/kmk/01-melanoma-Kosyr.pdf.
28. Nahar J, Tickle KS, Ali AB, Chen YP. Significant cancer prevention factor extraction: an association rule discovery approach. J Med Syst 2011;35(3):353-67.
29. Allahverdipoor H. Passing through traditional health education towards theory-oriented health education. Health Promotion and Education Magazine 2005;1(3):75-9.
30. Geller AC, Swetter SM, Brooks K, Demierre MF, Yaroch AL. Screening, early detection, and trends for melanoma: current status (2000-2006) and future direction. J Am Acad Dermatol 2007;57(4):555-72.
31. Thomas DB. Sun Awareness and Skin Cancer Prevention in the Teen Population: Using a School Based Approach In Teaching Adolescent Self-Health: Citeseer; 2006. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.580.8583&rep=rep1&type=pdf
32. Ghasemzadeh F, Etminani K, Arab-Kheradmand A, Hosseini Moini SB. A retrospective study on non melanoma skin cancer in Cancer Institute, Imam Khomeini Medical Center, Tehran, Iran. Journal of Dermatology and Cosmetic 2017; 8(1):9-21. Persian.
33. Mann AK, Kaur N. Survey Paper on Clustering Techniques. International Journal of Science, Engineering and Technology Research 2013; 2(4): 803-6.
34. Thompson JF, Scolyer RA, Kefford RF. Cutaneous melanoma. The Lancet 2005;365(9460):687-701.
35. Dennis LK, Lowe JB, Lynch CF, Alavanja MC. Cutaneous melanoma and obesity in the Agricultural Health Study. Ann Epidemiol 2008;18(3):214-21.
36. Sarshar N, Khajavi A. The prevalence of obesity in females of 15-65 years of age in Gonabad, Iran. Horizon Med Sci 2006; 12(3) :38-43. Persian
37. Wikipedia. Cancers. Physiotherapy in cancer; 2017 [cited 2017 Jun 25]. Available from: https://fa.wikipedia.org/wiki/%D9%BE%D8%B3%D9%88%D8%B1%DB%8C%D8%A7%D8%B2%DB%8C%D8%B3
38. Latifi Z. Skin cancer caused by sunshine. 2011. Tebyan. [cited 2017 Jun 25] Available from: https://article.tebyan.net/172806/.
39. Thangaraju P, Deepa B. A case study on perclusion and discovery of skin melanoma risk using clustering techniques. International Journal of Advanced Research in Electronics and Communication Engineering 2014;3(7):723-7.
40. Houtsma A, Swami M. Set-Oriented Mining for Association Rules in Relational Databases. Proceedings of the Eleventh International Conference on Data Engineering; 1995 Mar 6-10; Taipei, Taiwan: IEEE; 1995. p. 25-33.

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2017; 4(1): 39-47

Determination of the Most Important Factors Affecting Non-Melanoma Skin Cancer Using Data Mining Algorithms

Ghasemzadeh Foroughossadat[1], Arab-Kheradmand Ali[2], Daklan Soroush[3], Shabaninezhad Alireza[4], Garajei Ata[5], Etminani Kobra[6]*

• **Received:** 30 Mar, 2017 • **Accepted:** 16 Apr, 2017

Introduction: Non-melanoma skin cancer (NMSC) has recently been one of the three most common cancers in Iran. Inappropriate management of the disease has led to an increase in the prevalence and overhead costs. Data mining techniques are helpful in the analysis of patient records and accurate management of diseases. This study aimed to find hidden patterns and relationships in the data of NMSC patients using data mining algorithms.

Methods: In this applied study, study population consisted of medical records of 828 NMSC patients referred to the Cancer Institute of Imam Khomeini Hospital in Tehran during 2006-2015. Demographic variables and NMSC risk factors were clustered using K-Means algorithm. Apriori algorithm was applied as well for extraction of association rules and determination of patient's common information with a confidence of ≥ 0.9.

Results: According to the studied variables, NMSC patients were classified in four clusters and three important factors influencing the disease were identified as abnormal BMI, high risk occupations and previous history of cancer. Seven rules were approved by association rules and the highest associations were found between the past history of the disease, the involved site, the relapse, and the type of NMSC.

Conclusion: For the first time, this study could highlight the most important factors affecting NMSC using data mining methods. These factors should be considered either in self examination or screening skin tests in high-risk groups. In future studies, the contribution of physiological, ecological and genetic factors in the development of skin cancer should be jointly investigated as well.

Keywords: Non-Melanoma Skin Cancer, Data Mining, Clustering, Association rules, Risk Factor

• **Citation:** Ghasemzadeh F, Arab-Kheradmand A, Daklan S, Shabaninezhad A, Garajei A, Etminani K. Determination of the Most Important Factors Affecting Non-Melanoma Skin Cancer Using Data Mining Algorithms. Journal of Health and Biomedical Informatics 2017; 4(1): 39-47.

1. M.Sc. in Medical Informatics, Medical Informatics Dept., Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
2. Associate Professor of Plastic & Reconstruction Surgery, Surgical Oncology Dept., Cancer Institute, Imam Khomeini Hospital Complex, Tehran University of Medical Science, Tehran, Iran.
3. Dermatologist , Razi Skin Hospital, Dermatology Dept., Tehran University of Medical Sciences, Tehran, Iran.
4. M.Sc. in Entomology, Plant Protection Dept., Faculty of Agriculture, Shahrood University of Technology, Semnan, Iran.
5. Assistant Professor of Oral and Maxillofacial Surgery Dept., School of Dentistry, Cancer Institute, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran.
6. Ph.D. in Software Engineering, Assistant Professor of Medical Informatics Dept., Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
***Correspondence:** Medical Informatics Dept., Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
• **Tel:** 09155118312 • **Email:** EtminaniK@mums.ac.ir