مقاله پژوهشی

فصلنامهٔ بیماری‌های پستان ایران، سال هشتم، شماره اول، بهار ۱۳۹۴

کشف الگوهای پنهان در مجموعه داده‌های واقعی بیماران مبتلا به سرطان پستان با استفاده از تکنیک داده‌کاوی

علیرضا آتشی: کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی مشهد، مشهد، ایران. و گروه پژوهشی انفورماتیک سرطان، مرکز تحقیقات سرطان پستان جهاد دانشگاهی، تهران، ایران

بهزاد کیانی*: کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

ابراهیم عباسی: کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی مشهد، مشهد، ایران، گروه پژوهشی انفورماتیک سرطان، مرکز تحقیقات سرطان پستان جهاد دانشگاهی، تهران، ایران

نجمه ناظری: گروه پژوهشی انفورماتیک سرطان، مرکز تحقیقات سرطان پستان جهاد دانشگاهی، تهران، ایران

چکیده

مقدمه: روند رو به رشد سرطان پستان در سال‌های اخیر، لزوم اتکا به شیوه‌های مطمئن و جدید را برای شناسایی و کنترل این بیماری بیشتر آشکار می‌کند. داده‌کاوی یکی از این روش‌هاست که از پرطرف‌دارترین کاربردهای آن، کشف الگوهای پنهان مابین داده‌های بیماران در پایگاه‌های داده بزرگ است. در این مطالعه، پژوهشگران به بررسی و کشف الگوهای ناشناخته در یک مجموعه داده واقعی سرطان پستان می‌پردازند.

روش بررسی: به دلیل گمشدگی بالای داده‌ها در ۱۴۵ رکورد، تنها اطلاعات ۶۶۵ بیمار قابل استفاده بود. در فاز پیش‌پردازش داده‌ها، مقادیر تهی از طریق الگوریتم EM در نرم‌افزار SPSS19 تخمین زده شده است. سپس فیلدهای پیوسته تبدیل به فیلدهای گسسته شده، داده‌ها از طریق الگوریتم APRIORI تحلیل و روابط پنهان بین این داده‌ها کشف شده‌است. پس از استخراج، روابط در اختیار پزشک خبره حوزه سرطان پستان قرار گرفته تا روابط بی‌معنی حذف گردند.

یافته‌ها: تعداد ۱۰۰ رابطه انجمنی با ضریب اطمینان بالاتر از ۰/۹ توسط الگوریتم کشف شده است. پس از این که این روابط در اختیار فرد خبره قرار گرفته، تعداد ۱۰ رابطه از این روابط به لحاظ بالینی بامعنی تشخیص و گزارش شده‌اند.

نتیجه‌گیری: در این پژوهش، تعدادی از الگوهای کمتر شناخته شده و جالب توجه یک مجموعه داده واقعی استخراج گردیده‌اند. استفاده از داده‌کاوی بخصوص در داده‌های پزشکی با توجه به حجم بالای داده‌ها و وجود روابط ناشناخته فراوان بین علل بیماری‌ها، مشخصات دموگرافیک بیماران و ریزفاکتورهای خطر ابتلا به بیماری‌ها، مفید است. الگوها و مدل‌های حاصل از داده‌کاوی، در واقع فرضیات مطالعات بعدی را مشخص می‌نمایند که انجام آنها از جمله انجام انواع RCTها می‌توان این فرضیات را رد یا اثبات نمود.

واژه‌های کلیدی: سرطان پستان، داده‌کاوی، قوانین انجمنی.

* نشانی نویسنده پاسخگو: مشهد، میدان آزادی، دانشگاه علوم پزشکی مشهد، دانشکده پزشکی، دفتر گروه انفورماتیک پزشکی، بهزاد کیانی.
نشانی الکترونیک: KianiB911@mums.ac.ir

[DOR: 20.1001.1.17359406.1394.8.1.7.2]

مقدمه

سرطان پستان یکی از شایع‌ترین انواع سرطان در زنان است که تقریبا ۱۰٪ از زنان را در مراحل مختلف زندگی تحت تاثیر قرار می‌دهد (۴-۱). این سرطان، شایع‌ترین بدخیمی در میان زنان ایرانی و کانون اصلی توجهات در ایران است. در سال‌های اخیر، میزان شیوع بیماری روند رو به رشدی داشته و بررسی داده‌ها نشان می‌دهد که میزان بقای بیماران تا پنج سال پس از تشخیص، هشتاد و هشت درصد و ده سال پس از تشخیص هشتاد درصد بوده است (۲، ۳).

تمام تومورها سرطانی نبوده و ممکن است خوش‌خیم یا بدخیم باشند. تومورهای خوش‌خیم رشد غیرطبیعی دارند ولی به ندرت مرگ‌آور هستند. با این حال، تعدادی از توده‌های خوش خیم پستان نیز می‌توانند خطر ابتلا به سرطان پستان را افزایش دهند. همچنین در برخی از زنان با سابقه نمونه‌برداری از توده‌های خوش‌خیم پستان نیز، خطر سرطان پستان افزایش یافته است. از طرف دیگر، تومورهای بدخیم جدی‌تر بوده و سرطانی محسوب می‌شوند ولی تشخیص زود هنگام این نوع از سرطان‌ها شانس درمان موفقیت‌آمیز را بالا برده است (۴). داده‌کاوی از روش‌هایی است که در تشخیص یا پیش‌بینی سرطان‌ها به‌کار می‌رود و به عنوان مثال، یکی از پرطرفدارترین رویکردهای داده‌کاوی، پیش‌بینی عود مجدد سرطان پستان است. بنابراین، رویکردهای داده‌کاوی می‌توانند با کاهش تعداد نتایج مثبت کاذب و منفی کاذب در تصمیم-گیری پزشکان برای شناسایی بهتر سرطان پستان کمک کنند (۶،۵). در نتیجه، رویکردهای جدید مانند کشف دانش از پایگاه داده (KDD[1])، که شامل تکنیک‌های داده کاوی هستند، روز به روز محبوبیت بیشتری یافته و تبدیل به یک ابزار تحقیقاتی مطلوب برای پژوهشگران علوم پزشکی شده‌اند. به کمک آنها پژوهشگران می‌توانند الگوها و روابط بین تعداد زیادی از متغیرها را شناسایی کرده و پیش‌بینی نتایج حاصل از یک بیماری با استفاده از ذخایر اطلاعاتی موجود در پایگاه‌های داده برای آنها امکان پذیر گشته است (۷) بررسی‌ها و مطالعات گوناگونی در زمینه مشکلات ناشی از پیش بینی بقای بیماران مبتلا به سرطان پستان، با استفاده از روش‌های آماری و شبکه‌های عصبی مصنوعی صورت گرفته، اما فقط تعداد کمی از مطالعات حوزه پزشکی در این حوزه برای یافتن روابط مابین داده‌ها با استفاده از روش‌های داده‌کاوی انجام شده است (۸).

مطالعاتی با استفاده از داده‌کاوی و بـا رویکرد پیش‌بینی در علوم پزشکی وجود دارند. بـه عنوان مثال، دلن و همکاران از شـبکه‌های عصبی مصنوعی، درخت تصمیم‌گیری و رگرسیون لجستیک برای توسعه مدل‌های پیش‌بینی سرطان پستان با تجزیه و تحلیل پایگاه‌های بزرگ داده بهره جستند. نتایج تحقیقات آنها نشان داد که الگوریتم درخت تصمیم بـرای استخراج دانش از داده‌های موجود مقدم بر سایر روش‌ها بود و نتایج به دست آمده از تحقیق، نزدیک به واقعیت بود (۹). همچنین لند و ورهگن از ماشین بردار پشـتیبان (SVM[2]) بـه تنهایی بـرای طبقه‌بندی تراکم تومور سرطان پسـتان، اسـتفاده کردند. نتایج بـه دست آمده نشان داد که SVM، روشی مناسبی بوده و نتایج به دست آمده با شواهد موجود و واقعی مطابقت داشت (۱۰). لاندین و همکارانش از مدل شـبکه‌های عصبی مصنوعی و رگرسیون لجستیک برای پیش‌بینی پنج، ده و پانزده ساله بقای بیماران مبتلا به سرطان پستان استفاده کردند. آنها ۹۵۱ بیمار مبتلا به سرطان پستان را مورد مطالعه قرار داده و اندازه تومور، وضعیت گـره‌های لنفـی، نـوع بافت، تشکیل توبول، نکروز تومور و سن را بـه عنـوان متغیرهای ورودی در نظر گرفتند. سپس به این نتیجـه رسیدند کـه درختان طبقه‌بندی و همچنین رگرسیون لجستیک برای تفسـیر بالینی بسیار آسانتر اسـت (۱۱). پندهارکر و همکاران از چندین روش داده‌کاوی برای بررسی الگوهای موجود در سرطان پستان استفاده نمودند. در این مطالعه، آنهـا نشان دادند کـه داده‌کاوی می‌تواند به عنوان یک ابزار ارزشمند در شناسایی شباهت‌ها (الگوها) در مورد سرطان پستان بـا هدف تشخیص، پیش آگهی و درمان بـه کار رود (۱۲). از طرفی، در کشور ایران هم مطالعاتی با رویکردهای داده‌کاوی در حوزه سرطان پستان انجام شده است. برای مثال طلوعی و همکاران در یک پژوهش، از سه تکنیک مختلف داده‌کاوی برای پیش‌بینی عود مجدد سرطان پستان بهره گرفتند و به مقایسه این سه تکنیک پرداختند. در نهایت هر سه راهبرد، در پیش بینی عود مجدد سرطان

[1] Knowledge Discovery and Data Mining

[2] Support Vector Machine

DOR: 20.1001.1.17359406.1394.8.1.7.2]

پستان، دقت بالایی را داشتند و SVM بیشترین دقت را در پیش‌بینی عود مجدد به دنبال داشت (۱۳). در مطالعه دیگری عظیمیان و همکاران، روش‌های داده‌کاوی را برای تشخیص بیماری سرطان پستان در زنان به‌کار گرفتند که تشخیص آنان با دقت بالایی صورت گرفت (۱۴). این مطالعات مثال‌هایی از کاربرد داده‌کاوی در علوم پزشکی برای پیش‌بینی بیماری‌ها هستند.

با عنایت به اهمیت ویژه بیماری سرطان پستان و با توجه به مزیت مضاعف روش‌های داده‌کاوی، پژوهشگران این مطالعه در صدد برآمدند تا با استفاده از شیوه‌های داده کاوی و استفاده از داده‌های موجود سرطان پستان یک پایگاه داده در کشورمان، به کشف قوانین انجمنی ناشناخته در مجموعه داده‌ها بپردازند.

مواد و روش‌ها

این مطالعه از نوع مطالعات گذشته‌نگر است. داده‌های این پژوهش از مرکز تحقیقات سرطان دانشگاه شهید بهشتی دریافت شدند. این داده‌ها شامل داده‌های ۸۰۹ زن بزرگسال مبتلا به سرطان پستان سطح یک تا پنج و دارای هیجده ویژگی برای هر بیمار بودند. جمع‌آوری این داده‌ها از ابتدای سال ۱۳۸۸ تا پایان شهریورماه ۱۳۹۰ و از بیماران مراجعه‌کننده به مرکز تحقیقات سرطان دانشگاه علوم پزشکی شهید بهشتی انجام گرفته است. با توجه به گمشدگی بسیار زیاد داده‌ها در بین ۱۴۵ رکورد، تنها اطلاعات مربوط به ۶۶۵ بیمار قابل‌استفاده بودند. در واقع، تمام رکوردهای مربوط به بیماران با بیش از ۱۵٪ گمشدگی حذف گردیدند. بنابراین در این پژوهش، از تعداد ۶۶۵ رکورد مربوط به افرادی که مبتلا به سرطان پستان بوده‌اند استفاده شده است. با توجه به اینکه تعدادی از فیلدهای موجود در رکوردهای باقیمانده دارای مقادیر تهی بوده‌اند به عنوان یکی از فازهای پیش‌پردازش و آماده‌سازی داده‌ها، این مقادیر از طریق الگوریتم EM[3] و با استفاده از نرم‌افزار SPSS ورژن ۲۰ تخمین زده شده‌اند. پس از آن داده‌های پیوسته به مقادیر گسسته تبدیل شده‌اند. تمامی مقادیر عددی به مقادیر گسسته تبدیل شده‌اند. این گسسته‌سازی به معنی خوشه‌بندی مقادیر پیوسته در گروه‌های متفاوت توسط نرم‌افزار جهت ورود داده‌ها به الگوریتم می‌باشد. سپس، این رکوردها با استفاده از الگوریتم داده‌کاوی APRIORI از طریق ابزار داده کاوی WEKA آنالیز شده‌اند. پس از اجرای این الگوریتم بر روی داده‌ها تعداد 1206 رابطه انجمنی با ضریب اطمینان بالاتر از ۰/۹ به دست آمده است که به شکل تصادفی ۱۰۰ رابطه از این روابط انتخاب شدند. برای اینکه روابط بی‌معنی و زاید حذف گردند تمامی این روابط در اختیار فرد خبره در حوزه سرطان پستان قرار گرفته است تا روابط معنی‌دار به لحاظ پزشکی را مشخص نماید. به این خاطر، این روابط به زبان روزمره برای پزشکان تعریف شده و اندکی پردازش در برخی روابط (برای مثال تبدیل به عکس نقیض گزاره‌ای) برای درک بهتر از گزاره مورد نظر اعمال گردید. در نهایت ۱۰ گزاره از این نتایج به عنوان نتایج قابل توجه از لحاظ بالینی توسط پزشک انتخاب و گزارش گردیدند.

یافته‌ها

با توجه به داده‌های موجود بیماران، تمامی این بیماان زن و بزرگسال بوده‌اند که بیشتر آنان (۷۹٪) در بازه سنی ۳۰ تا۵۰ سال قرار گرفته و مابقی (۲۱٪) بیش از پنجاه سال داشته‌اند. تمامی این بیماران تحت درمان بوده یا پروسه درمانی خود را به پایان رسانده‌اند (مرگ یا سلامت). تعداد ۱۰۰ قانون انجمنی با ضریب اطمینان بیشتر از ۰/۹ به دست آمده است که این قوانین در اختیار پزشک با تجربه در حوزه داده‌کاوی سرطان پستان قرار گرفته است. پس از بررسی، قوانین زیر به عنوان قوانین معنی دار به لحاظ بالینی در نظر گرفته شده است. نتایج حاصله پس از ترجمه به زبان روزمره از قرار زیر بودند. (توجه می‌کنیم که این قوانین تنها قوانین انتخابی از بین ۱۰۰ رابطه اول بوده‌اند) سایر مشخصات روابط انتخاب شده نیز در جدول ۱ آمده‌است.

۱. بیماران با تومورسایز بالا کمتر عود سرطان داشته‌اند.
۲. بیماران متاهل و یائسه تومورهایی با سایز بالاتر داشته‌اند.
۳. بیماران غیرسیگاری تومورهای کوچکتری داشته‌اند.
۴. بیماران غیرسیگاری تومورهای کوچکتری داشته‌اند.
۵. بیماران با فاکتور HER9 منفی تومورهای بزرگی داشته‌اند.

[3] Expectation Maximization Algorithm

[DOR: 20.1001.1.17359406.1394.8.1.7.2]

۶. بیماران با فاکتور P53 منفی تومورهای کوچکتری داشته‌اند.
۷. بیماران متاهل سایز تومور کوچکتری داشته‌اند.
۸. اندازه تومور بیماران با ER و PR مثبت، کوچکتر بوده‌است.

جدول خروجی نرم‌افزار WEKA یا روابط منتخب توسط پزشک به زبان ماشین قبل از اعمال پردازش زبانی به همراه جزئیات در ضمیمه ۱ به پیوست آمده است.

بحث

مطالعه فعلی در سطحی کوچک و با هدف آشنایی جامعه علمی کشور بخصوص جامعه پزشکی با یکی از شیوه‌های استخراج دانش از داده‌های پزشکی صورت پذیرفت. در این پژوهش، تعدادی از الگوهای کمتر شناخته شده و جالب توجه یک مجموعه داده واقعی استخراج گردیده‌اند. البته معنی‌دار بودن این قوانین توسط پزشک متخصص در زمینه سرطان پستان تایید شده است اما به‌هرحال هرکدام از این قوانین باید به زبان محاوره‌ای و ترجیحاً به زبان پزشکان ترجمه و ارایه شوند و نهایتاً سازنده فرضیات مطالعات بعدی باشند. حسن اصلی این‌گونه مطالعات داده-کاوی این است که همراه با نتیجه مطالعه، احتمال وقوع و تعداد شاهد برای هر نتیجه‌گیری هم ارایه می‌گردد که این امر از لحاظ مستند بودن نتیجه‌گیری و اعتبار آن با توجه به تصمیم‌گیری مبتنی بر شاهد بسیار حایز اهمیت است. مطالعه‌ای از این دست تاکنون در حوزه سرطان در کشور ما انجام نشده است اما در نقاط مختلف دنیا این‌گونه مطالعات (البته با الگوریتم‌های متفاوتی از داده‌کاوی) بیشتر به چشم می‌خورد. در مقام مقایسه نتایج این مطالعه با مطالعات دیگر باید گفت که در مطالعاتی که قوانین انجمنی را بررسی می‌کنند نتایج آنچنان متفاوت به دست می‌آیند که عملا از لحاظ آماری قابل مقایسه نیستند اما از لحاظ ضریب اطمینان، میزان حداقل ۰/۹ اطمینان بسیار بسیار مناسب و بالاتر از بسیاری از مقالات دیگر این حوزه در سرطان‌ها است.

در مورد آن دسته از مطالعات حوزه پزشکی که معمولا بدون فرض اولیه (مانند مطالعات داده‌کاوی) انجام میشوند معمولا مقایسه نتایج با سایر مطالعات، بخصوص مطالعاتی که مبتنی بر داده‌های سایر پایگاه‌های داده انجام می‌پذیرد تا حد زیادی مشکل است، زیرا مطالعاتی که بر پایه داده‌های متفاوت انجام شده باشد احتمالا با توجه به نوع داده‌ها، تکنیک‌ها و روش‌های متفاوتی نیز اتخاذ می‌نمایند. در واقع، انتخاب نوع روش کار بسیار وابسته به نوع داده‌ها و ویژگی‌های منتخب بیماران می‌باشد. اما در زمینه سطح آموزش سیستم و نتایج آزمون‌ها و نیز دقت مدل‌های ایجاد شده، مقایسه نتایج این مطالعات حایز اهمیت است. در مقایسه با نتیجه مطالعه لند و ورهگن در به‌کارگیری SVM بر روی داده‌های سرطان پستان علی‌رغم وجود داده‌های بیشتر (چه از جهت تعداد رکورد و چه از جهت تعداد ویژگی‌ها) نتایج مطالعه فعلی میزان اطمینان بیشتری دارد. شاید دلیل این امر، انتخاب نتایج با ضریب اطمینان بالا (حداقل ۰/۹) از جانب پژوهشگران باشد که این خود به دلیل نوع تکنیک انتخاب شده و نیز بررسی متخصص بالینی است. از طرفی، با توجه به این مطلب که نتایج این مطالعه هم از لحاظ آماری (دادهای) و هم از لحاظ کلینیکی (توسط پزشک) اعتباربخشی شده، نتایج این مطالعه برای کاربری در محیط بالینی دارای اعتبار بیشتری است (۱۰).

در مطالعه لاندین (۱۱) علی‌رغم متفاوت بودن ویژگی‌های وارد شده در مطالعه ذکر می شود که تکنیک‌های درخت تصمیم و رگرسیون برای تفسیر پزشکان و سایر متخصصین بالینی مناسبتر است که این مطالعه نتایج پژوهش مذکور را حداقل در مقام مقایسه با قوانین انجمنی تایید می‌نماید. به دست آوردن قوانین انجمنی در مرحله ابتدایی و نیز در مرحله استخراج نتایج وترجمه به زبان پزشکان لزوما نیاز به وجود متخصصین داده‌کاوی مسلط به زبان بالینی دارد، اما به دلیل رواج و حتی قابلیت مشاهده روند تکنیک‌ها، رگرسیون و بخصوص درخت تصمیم از نظر متخصصین بالینی ساده‌تر و قابل‌ارزیابی‌تر است. بنابراین پژوهشگران استفاده از نتایج قوانین انجمنی را تنها زمانی پیشنهاد می‌نمایند که فرد خبره در حوزه داده‌کاوی و مسلط به زبان بالینی آن را در اختیار قرار دهد. بعلاوه، این مطالعه نتایج و میزان دقت مطالعه پندهارکر و همکاران و نیز مطالعه بومی عظیمیان و همکاران را تا حد زیادی تایید می‌کند (۱۴،۱۲). نکته‌ای که مجددا باید یادآوری شود توجه به مطالعات ذکر شده داخلی است که بیشتر هدف مدل‌سازی مهندسی را بر مجموعه داده‌های پزشکی دنبال می‌کنند و هدفی بالینی از این مدل‌سازی نداشته‌اند. این مساله از این جهت حایز

DOR: 20.1001.1.17359406.1394.8.1.7.2]

اهمیت است که گاهی مطالعه‌ای از لحاظ آماری به نتایجی می‌رسد که از لحاظ بالینی بی‌ارزش و غیرقابل استفاده‌اند اما از آن جهت اجرایی هستند که قابلیت سیستم‌ها و مدل‌های مختلف آموزش مدل‌های مهندسی را برروی داده‌های مختلف نشان می‌دهند (۱۴،۱۳) یکی از محدودیت‌هایی که در این پژوهش مطرح بود مقادیر زیاد داده‌های از دست رفته بوده که همان‌گونه که در روش اجرای تحقیق توضیح داده شده، این مقادیر توسط الگوریتم EM تخمین زده شده است. تخمین این مقادیر ممکن است تا حدی نتایج مطالعه را تحت تاثیر قرار داده باشد. محدودیت دیگر، تعداد کم بیماران برای ایجاد مدل است که باید در مطالعات بعدی این مدل با تعداد بیشتری از بیماران توسعه یابد. نکته حایز اهمیت دیگر در این مطالعه انتخاب تعداد کمی از نتایج است. مسلم است که اگر این متد برای تمامی قوانین ارایه شده توسط ماشین لحاظ می‌گردید، مسلما نتایج بسیار زیادتر و قابل ملاحظه-تری در دست بود که البته انجام چنین مطالعاتی در سطح بالاتر به علاقمندان و پژوهشگران پیشنهاد می‌شود. نقطه قوت این مطالعه، استفاده از یک مجموعه داده واقعی بیماران است که معمولا کمتر در مطالعات مشابه صورت می‌گیرد. زیرا معمولا در کشور ما، پژوهشگرانی که به مطالعاتی از این دست می‌پردازند بیش از آنکه جنبه پزشکی آن را در نظر بگیرند به دنبال تست مدل‌ها و متدهای آن در حوزه مهندسی هستند و بیشتر از مجموعه داده‌هایی بهره می‌برند که ماهیت واقعی آنها کمتر و یا غیربومی هستند.

نتیجه‌گیری

در این مطالعه تعدادی از الگوهای پنهان یک مجموعه واقعی از داده های بیماران سرطان پستان استخراج گردید. به هرحال همانند تمامی مطالعات داده‌کاوی، نتایج این پژوهش برای پایگاه داده مورد مطالعه (و نه سایر پایگاه‌ها مگر به شرط توسعه) معتبر است. نتایج این مطالعه تنها درصورتی قابل تعمیم هستند که در چندین پایگاه داده دیگر و به همین ترتیب آزموده شوند. اما مطالعاتی از این دست، جهت ایجاد فرضیه برای مطالعات بعدی می‌توانند استفاده و بررسی شوند و از طرفی به دلیل کم هزینه بودن و سرعت بالای انجام فرآیند از بسیار به صرفه خواهند بود. همچنین، چنین مطالعاتی برای بررسی نوع بیماران مراجعه کنند و نیز تخصیص مناسب منابع به مدیران مراکز بهداشتی نیز پیشنهاد می‌گردد.

ضمیمه ۱: روابط (قوانین)منتخب توسط پزشک به زبان ماشین قبل از اعمال پردازش زبانی

قانون منتخب	مشخصات قانون
sizeoftumorcent='(-inf-45.3]' 426 ==> relapse=No 426	<conf:(1)> lift:(1.01) lev:(0) [2] conv:(2.56)
sizeoftumorcent='(-inf-45.3]' 273 ==> mencestatus=Menstruation maritalstatus=Married 273	<conf:(1)> lift:(1.01) lev:(0) [1] conv:(1.64)
sizeoftumorcent='(-inf-45.3]' 265 ==> ER=Yes PR=Yes relapse=No 265	<conf:(1)> lift:(1.01) lev:(0) [1] conv:(1.59)
smoking= No ==> relapse=No surgury=B 251	<conf:(1)> lift:(1.01) lev:(0) [1] conv:(1.51)
smoking= No ==> sizeoftumorcent='(-inf-2.2]' 242	<conf:(1)> lift:(1.01) lev:(0) [1] conv:(1.46)
sizeoftumorcent='(-inf-45.3]' 238 ==> P53=No relapse=No 238	<conf:(1)> lift:(1.01) lev:(0) [1] conv:(1.43)
sizeoftumorcent='(-inf-45.3]' 251 ==> ER=Yes PR=Yes no.oflymlhatictumor='(-inf-2.2]' 252	<conf:(1)> lift:(1) lev:(0) [0] conv:(0.76)
sizeoftumorcent='(-inf-45.3]' 238 ==> ER=No 239	<conf:(1)> lift:(1) lev:(0) [0] conv:(0.72)
sizeoftumorcent='(-inf-45.3]' 391 ==> ER=Yes PR=Yes 393	<conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.79)
sizeoftumorcent='(-inf-45.3]' 530 ==> maritalstatus=Married 533	<conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.8)

[DOR: 20.1001.1.17359406.1394.8.1.7.2]

References

1. Hortobagyi GN, de la Garza Salazar J, Pritchard K, et al. The global breast cancer burden: variations in epidemiology and survival. Clinical Breast Cancer 2005; 6(5):391-401.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: A Cancer Journal for Clinicians 2005; 55(2):74-108.
3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: Cancer Journal for Clinicians 2011; 61(2):69-90.
4. National Breast Cancer Foundation. Early Detection [Internet]. National Breast Cancer Foundation, Inc. 2012. Available from: http://www.nationalbreastcancer.org/early-detection-of-breast-cancer. page
5. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144–152, Pittsburgh, PA, 1992. ACM Press.
6. Calle J. Breast cancer facts and figures 2003-2004 [Internet]. American Cancer Society: 2004 [last accessed: Jan. 2010]. Available from: http://www.cancer.org/. page
7. Jiang P, Liu XS. Big data mining yields novel insights on cancer. Nat Genet 2015; 47(2):103-4.
8. Karabatak M, Cevdet M. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 2009; 36: 3465-9.
9. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. J. Artificial Intelligence in Medicine 2010; 34: 113-27.
10. Land WH, Verheggen EA. Multiclass Primal Support Vector Machines For Breast Density Classification. Int J Comput Biol Drug Des 2009; 2(1):21-57.
11. Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. Oncology 1999; 57(4): 281-6.
12. Pendharkar PC, Rodger JA, Yaverbaum GJ, Herman N, Benner M. Associations statistical, mathematical and neural approaches for mining breast cancer patterns. Expert Systems with Applications 1999; 17:223-32.
13. Tolouei-Sh a, Pourebrahimi A, Ebrahimi M, Ghasem-A L. Predicting Breast Cancer Relapse Using Three DataMining Techniques. Iranian Journal of Breast Diseases 2012; 5(4):23-34.
14. Azimian F, Tadaion-T Gh, Jalali M. Breast Cancer Detection Using Data Mining Techniques. In proc. 4th Iranian Data Mining Conference 2010.

DOR: 20.1001.1.17359406.1394.8.1.7.2]