مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره اول شماره اول. ص ۶۲-۵۴، ۱۳۹۳

مقاله پژوهشی

مدل‌سازی درخت تصمیم‌گیری بر اساس روش‌های داده‌کاوی و پیشنهاد یک مدل جدید در تشخیص سرطان پوست با دو روش رنگ‌آمیزی ایمونوهیستوشیمی

افشین صرّافی‌نژاد [۱،۲]*، امیرحسین سعید[۳]، عیسی‌محمد رز [٤]، علیرضا روحانی‌منش[۵]

• دریافت مقاله: ۹۳/۸/۱۴ • پذیرش مقاله: ۹۳/۹/۱۰

مقدمه: شیوه‌های نوین تشخیصی، نظیر رنگ‌آمیزی ایمونوهیستوشیمیایی در سرطان پوست به متخصصان کمک می‌کند تا با اطمینان بیشتر و در زمان کوتاه‌تر به تشخیص برسند. هدف از این مطالعه، مقایسه یک روش مبتنی بر درخت تصمیم، برای تشخیص افتراقی دو نوع سرطان پوست (سرطان سلول‌های بازال و سرطان سلول‌های سنگفرشی) بر اساس نتایج سه روش رنگ‌آمیزی می‌باشد.

روش: در حوزه داده‌کاوی و با استفاده از نرم‌افزارهای SPSS.V19 و Clementine.V12 و با روش‌های مدل‌سازی درخت تصمیم‌گیری CART و CHAID، داده‌های مربوط به ۶۰ بیمار مبتلا به سرطان پوست مربوط به کشور مالزی بررسی شد. سه روش رنگ‌آمیزی BCL2، Galectin-3(C) و Galectin-3(N) برای تشخیص، وارد مدل شدند. بهترین متغیر پیش گویی‌کننده برای ساخت درخت، شناسایی و با مدل پژوهشگر-ساخته دیگری بر مبنای مقادیر بحرانی حاصل از ROC curve analysis و نتایج پاتولوژی مقایسه شد.

نتایج: در مدل ترکیبی حاصل، میزان حساسیت و ویژگی تشخیص برای BCC به ترتیب ۸۲/۱ و ۱۰۰ درصد، برای SCC به ترتیب ۱۰۰ و ۸۲/۸ درصد، دقت کلی مدل ۹۰/۳۸ درصد، ارزش اخباری مثبت (PPV) برای تشخیص SCC و BCC به ترتیب ۸۲/۱ و ۱۰۰ درصد و نسبت درست‌نمایی مثبت (PLR) برای BCC و SCC به ترتیب ۵/۸ و ۵/۵ به دست آمد.

نتیجه‌گیری: مدل درخت تصمیم گیری بر اساس دو روش رنگ آمیزی ایمونوهیستوشیمی در سرطان پوست، می تواند در تشخیص این دو بیماری بدخیم کمک کند و زمینهٔ تحقیقات بیشتر را فراهم آورد.

کلید واژه‌ها: داده‌کاوی، درخت تصمیم‌گیری، ایمونوهیستوشیمی، BCL-2، Galectin-3، سرطان پوست

ارجاع: صرّافی‌نژاد افشین، سعید امیرحسین، رز عیسی‌محمد، روحانی‌منش علیرضا. مدل‌سازی درخت تصمیم‌گیری بر اساس روش‌های داده‌کاوی و پیشنهاد یک مدل جدید در تشخیص سرطان پوست با دو روش رنگ‌آمیزی ایمونوهیستوشیمی. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۳؛ ۱(۱): ۶۲-۵۴.

۱. پزشک عمومی، دانشجوی دکترای تخصصی رشته انفورماتیک پزشکی، گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، خراسان رضوی، ایران
۲. مرکز تحقیقات انفورماتیک پزشکی، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۳. متخصص پاتولوژی بالینی و جراحی، دانشگاه علوم پزشکی کرمان، کرمان، ایران
٤. استاد رشته پاتولوژی، دانشکده پزشکی، دانشگاه کبانگسان، کوآلالامپور، مالزی
۵. دکترای کنترل، استادیار گروه مهندسی برق، دانشگاه نیشابور، خراسان رضوی، ایران
* **نویسنده مسؤول:** کرمان، ابتدای بزرگراه هفت باغ علوی، پردیزه دانشگاه علوم پزشکی کرمان، پژوهشکده آینده پژوهی در سلامت، مرکز تحقیقات انفورماتیک پزشکی
• **شماره تماس:** ۰۹۱۳۱۴۱۴۴۰۵ • **Email:** a.sarafi@gmail.com

[Downloaded from jhbmi.com on 2025-07-20]

مقدمه

سرطان پوست به عنوان شایع‌ترین نوع سرطان در انسان [۱]، از بیماری‌های بالقوه کشنده‌ای است که تقریباً یک سوم تمامی انواع سرطان‌ها را در آمریکا تشکیل می‌دهد [۲]. تشخیص صحیح و به موقع سرطان پوست می‌تواند از مرگ و میر بیماران تا حدودی جلوگیری کند. در کشور مالزی، سرطان‌های غیر ملانوم پوست حدود ۴ درصد از تمام سرطان‌های تازه تشخیص داده شده در مردان و ۲/۷ درصد از تمام سرطان‌های تازه تشخیص در زنان را در سال ۲۰۰۳ تشکیل می‌داد [۳]. شایع‌ترین انواع سرطان پوست از نظر هیستولوژیک، سرطان سلول‌های بازال (Basal Cell Carcinoma= BCC) و سرطان سلول‌های سنگفرشی (Squamous Cell Carcinoma=SCC) می‌باشند که به ترتیب حدود ۸۰ و ۱۵ درصد انواع سرطان‌های بدخیم پوست را تشکیل می‌دهند [۴].

در تظاهر بالینی، پارامترهای آسیب‌شناختی از جمله محل و اندازه تومور و برخی نمودهای دیگر ممکن است شبیه هم باشند که تشخیص افتراقی این سرطان با برخی ضایعات تومورال اپی‌تلیال پوست، را دشوار می‌کند و در اغلب موارد، نیاز به بررسی هیستوپاتولوژیک نمونه‌های برداشته شده از پوست بیمار می‌باشد. با توجه به اینکه هر یک از انواع سرطان‌های پوست از جمله BCC و SCC می‌توانند رفتار متاستاتیک متفاوتی داشته باشند و پتانسیل تبدیل به یک عارضه کشنده را دارند، تشخیص دقیق و به موقع آنها به ویژه از نظر انتخاب درمان مناسب و مراقبت صحیح از بیمار، اهمیت زیادی دارد [۵]. در برخی بیماران، افتراق بین دو نوع سرطان مذکور دشوار است. مثلاً برخی از انواع تومورهای non-basaloid مانند SCC ممکن است واریانت‌های basaloid داشته باشند و یا اینکه فرم انفیلتراتیو SCC گاهی به صورت basaloid نمایان می‌شود و ممکن است با basosquamous BCC قابل افتراق نباشد. همچنین desmoplastic SCC یکی از انواعی است که باید از نوع morphemic metatypical basosquamous BCC تمیز داده شود [۶].

روش ایمونوهیستوشیمی می‌تواند در تشخیص افتراقی این سرطان‌ها مؤثر باشد. بر اساس نوع هیستولوژیک و درجه‌بندی سرطان‌های غیر ملانومی پوست، اندازه‌گیری ایمونوهیستوشیمیایی BCL2 و گالکتین-۳ Galectin-3=Gal3) به عنوان دو شاخص بافتی تومور، قابل طرح می‌باشند [۷،۸]. خانواده پروتئین‌های BCL2 نقش مهمی در تنظیم آپوپتوز ناشی از یک سری محرک‌ها ایفا می‌کنند [۹]. همچنین گالکتین-۳ یکی از لکتین‌های حیوانی است که تمایل زیادی به اتصال با بتاگالاکتوزیدها دارد که منجر به عملکرد بیولوژیک متغیر سلولی نظیر cell aggregation، adhesion، proliferation، مرگ سلول و التهاب و متاستاز می‌شود [۱۰] و بیان آن در بالغین به صورت عمده به سلول‌های اپی‌تلیال و میلوئید/آمیبوئید وابسته است [۱۱].

در برخی مطالعات، شواهدی در مورد ارتباط تنگاتنگ نتایج اندازه‌گیری BCL2 و Gal3 با روش‌های رنگ‌آمیزی ایمونوهیستوشیمی در لام‌های پاتولوژی بیماران مبتلا به BCC و SCC پوست گزارش شده است [۸].

داده‌کاوی و اکتشاف دانش (Data Mining and Knowledge Discovery) یکی از رشته‌های نسبتاً جدید در حوزه علوم کامپیوتر، انفورماتیک پزشکی و آمار تخصصی است که با استفاده از روش‌های ویژه و طبقه‌بندی و تحلیل داده‌ها، روابط جدید بین متغیرها را کشف کرده و یا منجر به ساخت فرضیه‌های جدید برای پژوهش‌های بعدی می‌شود.

نیاز به تشخیص صحیح و زودرس سرطان‌های پوست که شایع‌ترین سرطان در انسان می‌باشد و اهمیت درمان درست و مبتنی بر تشخیص صحیح، حاکی از آن است که روش‌های تشخیصی جدید و علمی بایستی تا حد امکان ساده و در عین حال دقیق باشند [۲،۹].

در این مطالعه متدولوژیک، هدف پژوهشگران این است که یک مدل درخت تصمیم‌گیری ایجاد شده توسط نرم‌افزار را با مدل طراحی شده دیگری که با دخالت کاربر و با استفاده از داده‌های مربوط به بیماران مبتلا به سرطان پوست در همین مطالعه حاصل شده است را مقایسه کنند. نتایج کاربردی این تحقیق، می‌تواند برای پزشکان، به ویژه متخصصان پاتولوژی بالینی و متخصصان پوست برای تسهیل در تصمیم‌گیری برای تشخیص نوع سرطان پوست، به خصوص در موارد مبهم، کمک‌کننده باشد.

روش

این پژوهش به عنوان یک بررسی متدولوژیک در زمینه داده‌کاوی و استفاده از روش‌های طبقه‌بندی، قابل طرح است. با استفاده از داده‌های خام مربوط به بیماران مبتلا به سرطان پوست که در یک دوره زمانی حدود هفت سال بین سال‌های ۲۰۰۲ تا ۲۰۰۹ میلادی در بخش پاتولوژی دانشکده پزشکی دانشگاه کبانگسان در شهر کوآلالامپور کشور مالزی گردآوری

[Downloaded from jhbmi.com on 2025-07-20]

شده بودند، مدل ویژه‌ای برای تشخیص دو نوع سرطان پوست طراحی شد. داده‌های مورد استفاده، مجموعه ۶۰ رکورد مربوط به بیماران مبتلا به سرطان پوست بود. در این فایل، متغیرهای مختلفی از جمله سن، جنس، نژاد، تشخیص نهایی پاتولوژیک سرطان، اندازه BCL2 و Gal3 (به دو روش هسته و سیتوپلاسم، Gal3-N و Gal3-C) و اندازه تومور ثبت شده بود. در طراحی مدل ارائه شده در این مطالعه، تنها از متغیرهای اندازه BCL2 و Gal3-N و Gal3-C و تشخیص نهایی پاتولوژیک استفاده شد.

طی مرحله پردازش اولیه داده‌ها هشت رکورد به دلایل مختلف شامل missing data (یک مورد)، صفر بودن هر سه شاخص (یک مورد) و عدم تمایز تومور (۶ مورد) از بررسی حذف و در نهایت ۵۲ رکورد وارد آنالیز شد.

برای اندازه‌گیری هر سه روش رنگ‌آمیزی، یعنی BCL2، Gal3-C و Gal3-N در تشخیص سرطان، ابتدا با استفاده از آنالیز ROC Curve در نرم‌افزار SPSS.V19 حدود بحرانی (Cut-off points) برای این روش‌ها محاسبه شد. همچنین با استفاده از نرم‌افزار Clementine.V12 و دو روش مدل‌سازی CART و CHAID و ایجاد درخت تصمیم‌گیری با در نظر گرفتن تشخیص نهایی به عنوان outcome، مدل‌ها ساخته شدند. در نهایت با استفاده از تلفیق نتایج دو مدل موجود، شاخص‌های حساسیت، ویژگی و ارزش اخباری مثبت و منفی مربوط به انواع رنگ‌آمیزی‌ها به عنوان آزمون‌های تشخیصی، محاسبه شد.

نتایج

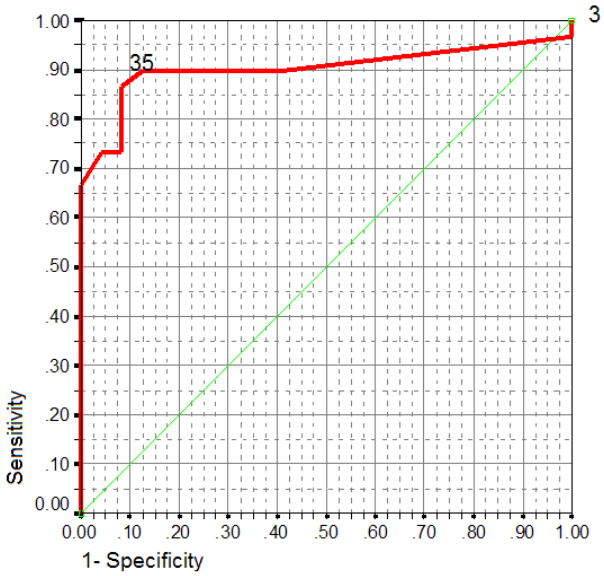
در این بررسی، داده‌های مربوط به ۶۰ بیمار مبتلا به سرطان پوست که برای اهداف پژوهشی دیگری در زمینه موضوع اصلی تحقیق گردآوری شده بود، از نظر ارتباط بین متغیرها و طراحی مدل درخت تصمیم، استفاده شد. داده‌های مورد بررسی از کشور مالزی جمع‌آوری شده و ترکیب آنها به صورت ۳۰ بیمار مبتلا به BCC و ۳۰ بیمار مبتلا به SCC بود که در نهایت ۵۲ رکورد وارد مدل تحلیل نهایی شد. میانگین (انحراف معیار±) سن بیماران ۶۷/۵۵ (۱۰/۹۵±) و از ۳۷ تا ۸۷ سال متغیر بود. بین دو گروه بیمار از نظر سن تفاوتی وجود نداشت (P=۰/۷). تفاوت نسبت دو جنس نیز در دو گروه معنی‌دار نبود (P=۰/۴). تحلیل دیگری برای متغیرهای زمینه‌ای انجام نشد.

به منظور تشخیص BCC از طریق اندازه‌گیری BCL2، با استفاده از نرم‌افزار SPSS، بهترین نقطه بحرانی با استفاده از نمودار ROC Curve و تحلیل مربوط به آن محاسبه شد که سطح ۳۵ درصد برای این روش به دست آمد. لازم به ذکر است که اندازه‌گیری BCL2 و Gal3 با شمارش درصد سلول‌هایی که رنگ گرفته‌اند انجام می‌شود و نتیجهٔ این شمارش معمولاً به درصد گزارش می‌شود. در مورد گالکتین-۳ سیتوپلاسم (Gal3-C) نیز به همین روش ولی برای تشخیص SCC از منحنی ROC استفاده شد و باز هم نقطه بحرانی ۳۵ به دست آمد. این مقدار و هر یک از رنگ‌آمیزی‌های مورد بررسی، به عنوان معیار اولیه برای تشخیص سرطان با نتیجهٔ نهایی مقایسه شدند، اما هیچ یک به تنهایی آنقدر کفایت نداشتند که بتوانند درصد بالایی از موارد را تفکیک کنند. در ضمن به دلیل اینکه Gal3-N در بررسی اولیه به صورت مستقیم ارتباط نسبتاً ضعیفی با تشخیص نهایی داشت، در تحلیل‌ها به طور مستقیم مورد استفاده قرار نگرفت.

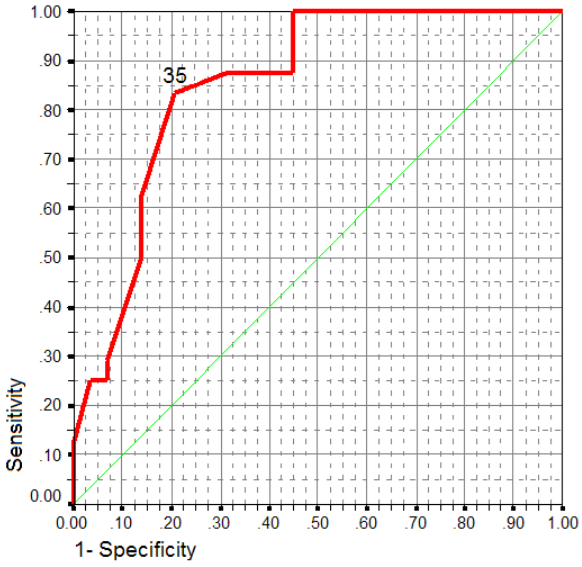
نمودارها و جدول‌های ۱ و ۲ نتایج این تحلیل‌ها را نشان می‌دهند. در جداول، BCL2≥35 و Gal3-C≥35 در واقع نشان‌دهنده متغیرهای جدیدی هستند که بر اساس نقطه بحرانی به دست آمده از ROC Analysis ایجاد شده‌اند. با توجه به اینکه ۱۱ مورد از بیماران با این متغیرهای جدید در طبقه‌بندی تفکیکی به درستی قرار نمی‌گرفتند، مدل به صورت دیگری نیز بررسی شد.

با تعریف سه متغیر گروه‌بندی‌کننده بر اساس مقدار بحرانی ۳۵، درخت تصمیم‌گیری با استفاده از دو الگوریتم CART و CHAID برای متغیرهای عددی و گروه‌بندی‌شده بازطراحی شد. در این دو روش، هر بار سطح پیشروی درخت معادل ۵ مرحله در نظر گرفته شد و با در نظر گرفتن تشخیص نهایی به عنوان متغیر outcome، سایر متغیرها وارد مدل شدند. متغیری که در ریشه درخت قرار گرفت در هر دو الگوریتم و حتی با وارد کردن تمامی متغیرهای موجود به مدل، BCL2 بود، اما یکی از محدودیت‌هایی که در مدل اول وجود داشت، ورود مجدد متغیرهای BCL2 و Gal3-C در شاخه‌های انتهایی درخت تصمیم‌گیری بود که سبب افت نسبی کیفیت مدل در پیشگویی نتیجه می‌شد. لذا با توجه به نتیجه اولیه مدل و در نظر گرفتن اختلاف عددی دو متغیر، مدل دیگری از درخت تصمیم با هفت شرط ویژه طراحی شد که این شرطها در شکل ۱ آورده شده است.

[Downloaded from jhbmi.com on 2025-07-20]



نمودار ۱: منحنی ROC برای متغیر BCL2 در تشخیص BCC



نمودار ۲: منحنی ROC برای متغیر Gal3-C در تشخیص SCC

جدول ۱: مقایسه تشخیص بر اساس BCL2≥35 به عنوان BCC بر اساس تشخیص نهایی پاتولوژیک

تشخیص BCL2≥35 / تشخیص پاتولوژیک		سرطان سلول‌های بازال (BCC)	سرطان سلول‌های سنگفرشی (SCC)	کل
سرطان سلول‌های بازال (BCC)	تعداد	۲۲	۰	۲۲
	درصد سطری	۱۰۰	۰	۱۰۰
	درصد ستونی	۷۸/۶	۰	۴۲/۳
سرطان سلول‌های سنگفرشی (SCC)	تعداد	۶	۲۴	۳۰
	درصد سطری	۲۰/۰	۸۰/۰	۱۰۰
	درصد ستونی	۲۱/۴	۱۰۰	۵۷/۷
جمع کل	تعداد	۲۸	۲۴	۵۲
	درصد سطری	۵۳/۸	۴۶/۲	۱۰۰

جدول ۲: مقایسه تشخیص بر اساس Gal3-C≥35 به عنوان SCC بر اساس تشخیص نهایی پاتولوژیک

تشخیص BCL2≥35 / تشخیص پاتولوژیک		سرطان سلول‌های بازال (BCC)	سرطان سلول‌های سنگفرشی (SCC)	کل
سرطان سلول‌های بازال (BCC)	تعداد	۲۵	۶	۳۱
	درصد سطری	۸۰/۶	۱۹/۴	۱۰۰
	درصد ستونی	۸۹/۳	۲۵/۰	۵۹/۶
سرطان سلول‌های سنگفرشی (SCC)	تعداد	۳	۱۸	۲۱
	درصد سطری	۱۴/۳	۸۵/۷	۱۰۰
	درصد ستونی	۱۰/۷	۷۵/۰	۴۰/۴
جمع کل	تعداد	۲۸	۲۴	۵۲
	درصد سطری	۵۳/۸	۴۶/۲	۱۰۰

[Downloaded from jhbmi.com on 2025-07-20]

If	BCL2≥35 and Gal3_c<35	➔Diagnosis is: BCC
If	BCL2<35 and Gal3_c≥35	➔Diagnosis is: SCC
If	BCL2≥35 and Gal3_c≥35 and BCL2<GAL3_c	➔Diagnosis is: SCC
If	BCL2≥35 and Gal3_c≥35 and BCL2>GAL3_c and GAL3_n<35	➔Diagnosis is: BCC
If	BCL2≥35 and Gal3_c≥35 and BCL2>GAL3_c and GAL3_n≥35	➔Diagnosis is: SCC
If	BCL2<35 and Gal3_c<35 and BCL2<GAL3_c	➔Diagnosis is: SCC
If	BCL2<35 and Gal3_c<35 and BCL2>GAL3_c	➔Diagnosis is: BCC

شکل ۱: شرط‌های هفت‌گانه مدل پیشنهادی برای تشخیص دو نوع سرطان پوست

چنانچه ملاحظه می‌شود در این مدل پیشنهادی، علاوه بر دو روش رنگ‌آمیزی BCL2 و Gal3-C، از روش سوم یعنی Gal3-N که از مدل درخت تصمیم با نرم‌افزار، حذف شده بود، در دو شرط استفاده شد. با استفاده از این مدل جدید، متغیر دیگری در فایل تعریف شد که شامل تمام این شرایط بود و شاخص‌های تشخیص با استفاده از این متغیر با تشخیص نهایی، مقایسه شدند.

در نتیجهٔ طرح این مدل جدید که نتیجهٔ آن در جدول ۳ آورده شده است، شاخص‌های مربوط به آزمون به شرح جدول ۴ محاسبه شد.

جدول ۳: مقایسه تشخیص بر اساس مدل پیشنهادی و تشخیص نهایی پاتولوژیک

تشخیص با مدل پیشنهادی / تشخیص پاتولوژیک		سرطان سلول‌های بازال (BCC)	سرطان سلول‌های سنگفرشی (SCC)	کل
سرطان سلول‌های بازال (BCC)	تعداد	۲۳	۰	۲۳
	درصد سطری	۱۰۰	۰/۰	۱۰۰
	درصد ستونی	۸۲/۱	۰/۰	۴۴/۲
سرطان سلول‌های سنگفرشی (SCC)	تعداد	۵	۲۴	۲۹
	درصد سطری	۱۷/۲	۸۲/۸	۱۰۰
	درصد ستونی	۱۷/۹	۱۰۰	۵۵/۸
جمع کل	تعداد	۲۸	۲۴	۵۲
	درصد سطری	۵۳/۸	۴۶/۲	۱۰۰

جدول ٤: شاخص‌های تشخیصی مدل پیشنهادی برای تشخیص دو بیماری

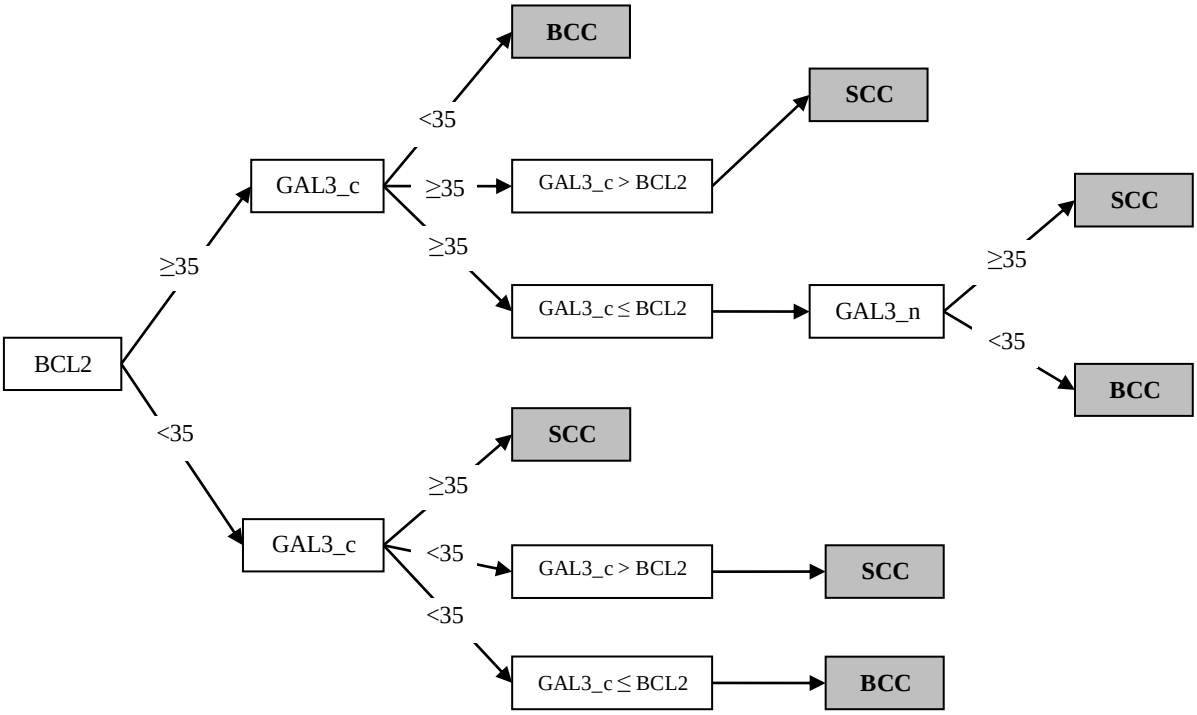
بیماری / شاخص (درصد)	حساسیت	ویژگی	دقت مدل
سرطان سلول‌های بازال (BCC)	۸۲/۱	۱۰۰	۹۰/۳۸
سرطان سلول‌های سنگفرشی (SCC)	۱۰۰	۸۲/۸	

علاوه بر حساسیت و ویژگی به دست آمده، ارزش اخباری مثبت (PPV) برای تشخیص SCC معادل ویژگی در تشخیص BCC و PPV در تشخیص BCC معادل ویژگی در تشخیص SCC می باشد.

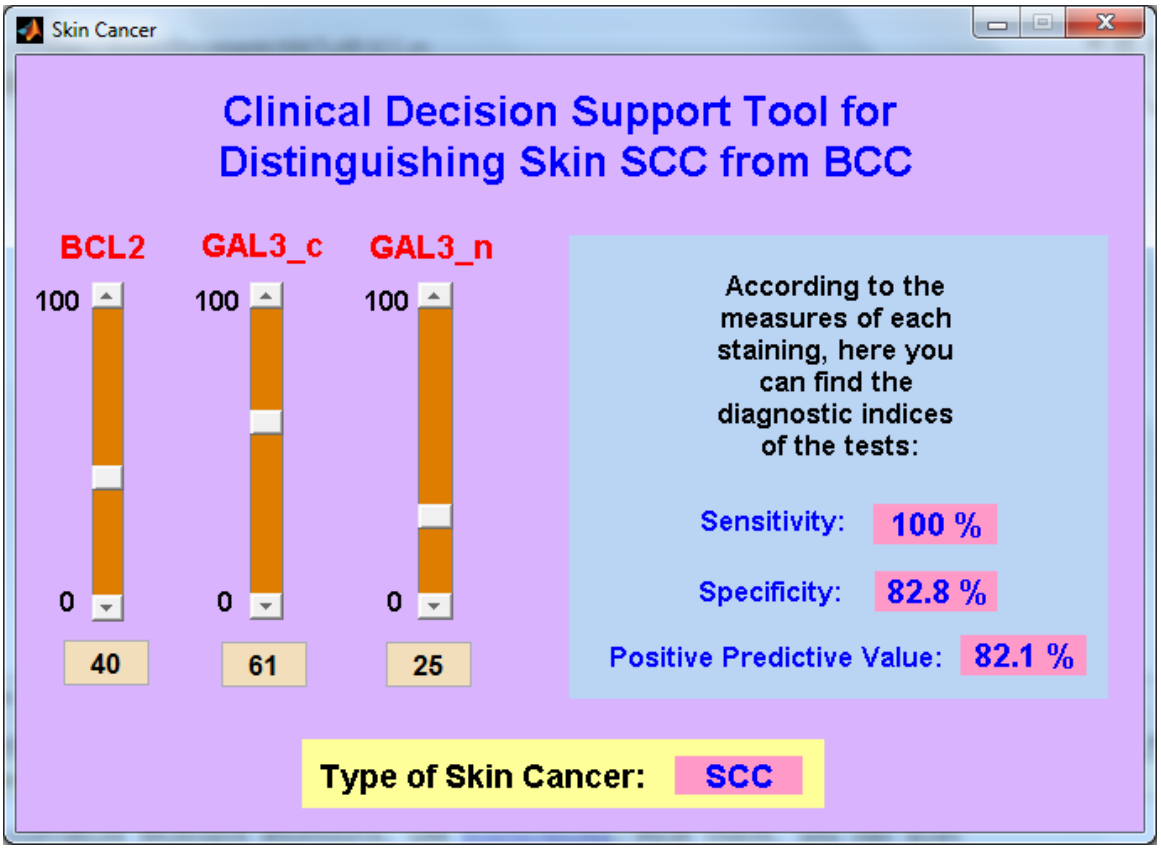
دقت کلی مدل ۹۰/۳۸ درصد محاسبه شد و نسبت درست‌نمایی مثبت (Positive Likelihood Ratio) برای تشخیص BCC و SCC با این مدل به ترتیب ۵/۸ و ۵/۵ به دست آمد. شکل ۲ درخت نهایی تصمیم‌گیری را با استفاده از این مدل نشان می‌دهد.

در نهایت برای تسهیل امر تصمیم‌گیری بر اساس نتایج آزمون‌ها، با استفاده از نرم‌افزار متلب (MATLAB)، یک نرم‌افزار ساده کاربردی طراحی و اجرا شد که قوانین به کار گرفته شده در آن از نتایج تحقیق استخراج شده و پزشک می‌تواند به سادگی با حرکت دادن دکمه‌های لغزشی، نتیجه را مشاهده کند (شکل ۳).

[Downloaded from jhbmi.com on 2025-07-20]



شکل ۲: مدل نهایی درخت تصمیم‌گیری برای تشخیص BCC و SCC پوست با استفاده از نتایج سه روش رنگ آمیزی



شکل ۳: نمایی از نرم‌افزار طراحی شده برای کمک به تصمیم‌گیری

[Downloaded from jhbmi.com on 2025-07-20]

بحث و نتیجه‌گیری

پژوهش حاضر که با هدف تعیین ارزش مدل پیشنهادی مبتنی بر نتایج رنگ‌آمیزی‌های ایمونوهیستوشیمی در دو نوع سرطان پوست شامل BCC و SCC طراحی شده، نشان داد که با استفاده از نتایج رنگ‌آمیزی می‌توان با دقت، حساسیت و ویژگی بالایی دو نوع بیماری سرطان پوست را از یکدیگر متمایز کرد. این مدل با تشخیص نهایی پاتولوژی برای SCC تطبیق ۱۰۰ درصد داشت، یعنی در صورتی که با مدل مذکور یک بیمار مبتلا به BCC شناخته شود، قطعاً وی به همین بیماری مبتلا است و تشخیص SCC به اشتباه برای وی منظور نخواهد شد.

در پژوهش حاضر که بر اساس داده‌های مربوط به یک مطالعهٔ پژوهشی دیگر طرح‌ریزی شده است، مدلی پیشنهاد شده که علاوه بر سادگی درک و قابلیت استفاده آن برای متخصصان رشته‌های مرتبط بالینی، در صورت تأیید در مطالعات مشابه، می‌توان در آزمایشگاه‌های تخصصی پاتولوژی نیز به عنوان یک استاندارد طلایی در بررسی صحت و سقم سایر روش‌های تشخیصی به کار گرفته شود.

در پژوهش‌های دیگری که در آنها از روش‌های رنگ‌آمیزی ویژهٔ پژوهش حاضر استفاده شده است، به ندرت به مدل‌های تشخیصی بر اساس این روش‌ها اشاره شده و تنها شواهدی در زمینه ارتباط بین تشخیص و میزان رنگ‌پذیری سلول‌های سرطانی در هر یک از روش‌های سه‌گانهٔ مورد اشاره وجود دارد [۸]. در یک پژوهش دیگر، Galectin-3 به عنوان یک هدف درمانی جدید برای برخی از بیماری‌های پوست مطرح گردیده، اما به عنوان یک تعیین‌کننده تشخیص، از آن یاد نشده است [۱۲].

در مطالعهٔ حاضر، این موضوع از جهت دیگر، یعنی ارتباط بین نتیجهٔ رنگ‌آمیزی با احتمال وجود هر یک از دو نوع سرطان BCC و SCC مطرح شده و به همین دلیل، مدل جدید می‌تواند پس از انجام مطالعات اصلاحی دیگر به عنوان یک روش آزمون تشخیصی مطرح گردد. چرا که در نتیجهٔ به دست آمده از این مدل جدید، علاوه بر پوشش تمامی بیماران و عدم وجود موارد استثنا، تصمیم‌گیری در مورد تشخیص نهایی توسط متخصصان بالینی به سادگی میسر می‌باشد. ضمن اینکه شاخص‌های تشخیصی آن از جمله حساسیت و ارزش اخباری مثبت در مورد نوع SCC که پیش آگهی و سیر بدتری نسبت به BCC دارد نیز از ارزش بالایی برخوردار است.

توضیحات ذکر شده در بخش نتایج نشان می‌دهد که تنها با دانستن سه مقدار مربوط به سه نوع رنگ‌آمیزی و مقایسهٔ آنها با عدد ۳۵ و با یکدیگر، می‌توان به سادگی به تشخیص نهایی (بخش خاکستری رنگ در تصویر ۴) رسید و بیش از ۹۰ درصد به صحت این تشخیص اطمینان داشت.

چنانچه پیشتر نیز ذکر شد، اگر با استفاده از این مدل، فردی مبتلا به BCC تشخیص داده شود، تشخیص ۱۰۰ درصد صحیح و اگر کسی مبتلا به SCC تشخیص داده شود، احتمال بیماری در وی حدود ۸۳ درصد است. از طرف دیگر اگر ۱۰۰ بیمار مبتلا به SCC با این مدل بررسی شوند، تشخیص همگی آنها SCC خواهد بود و هیچ بیماری به اشتباه BCC تشخیص داده نمی‌شود و اگر ۱۰۰ بیمار BCC نیز با همین مدل بررسی شوند حدود ۸۲ درصد آنها به درستی در گروه BCC قرار خواهند گرفت. ضمن اینکه دقت بالای مدل (۹۰/۳۸) و مقادیر نسبت درست‌نمایی مثبت به دست آمده، در مجموع نشان‌دهنده توان بالای مدل در تشخیص صحیح بیماری است.

در حقیقت به دلیل ماهیت بیماری، متخصصان بالینی ترجیح می‌دهند در تشخیص سرطان پوست، تا حد امکان بیماران مبتلا به SCC به اشتباه BCC تشخیص داده نشوند که در مدل حاضر این احتمال به صفر رسیده است.

به دلیل اهمیت خاص تشخیص صحیح این دو بیماری و برای بررسی دقیق‌تر مدل پیشنهادی، به دانش‌پژوهان حوزه پاتولوژی بالینی پیشنهاد می‌شود با مشارکت متخصصان انفورماتیک پزشکی یا آمار، در مطالعات دیگر با طرح‌های دوسویه کور و استفاده از داده‌های بیشتر، به پژوهش در این باره بپردازند.

[Downloaded from jhbmi.com on 2025-07-20]

References

1. Cassarino DS, Derienzo DP, Barr RJ. Cutaneous squamous cell carcinoma: a comprehensive clinicopathologic classification. Part one. J Cutan Pathol. 2006; 33(3):191-206.
2. Jerant AF, Johnson JT, Sheridan CD, Caffrey TJ. Early detection and treatment of skin cancer. Am Fam Physician. 2000; 62(2):357-68, 375-6, 381-2.
3. Lim GCC, Rampal S, Halimah Y. Cancer Incidence in Peninsular Malaysia, 2003-2005. The Third Report of the National Cancer Registry. National Cancer Registry: Malaysia, Kuala Lumpur; 2008.
4. Walling HW, Fosko SW, Geraminejad PA, Whitaker DC, Arpey CJ. Aggressive basal cell carcinoma: presentation, pathogenesis, and management. Cancer Metastasis Rev. 2004; 23(3-4):389-402.
5. Sanders DSA, Carr RA. The use of immunohistochemistry in the differential diagnosis of common epithelial tumours of the skin. Current Diagnostic Pathology. 2007; 13(4): 237-51
6. Elder DE, Elenitsas R, Johnson Jr BL, Murphy GF, Xu X. Lever's histopathology of the skin. 9th ed. Philadelphia: Lippincott Williams and Wilkins; 2005.
7. Cherpelis BS, Marcusen C, Lang PG. Prognostic factors for metastasis in squamous cell carcinoma of the skin. Dermatol Surg. 2002; 28(3):268-73.
8. Puizina-Ivic N, Sapunar D, Marasovi D, Miri L. An Overview of Bcl-2 Expression in Histopathological Variants of Basal Cell Carcinoma, Squamous Cell Carcinoma, Actinic Keratosis and Seborrheic Keratosis. Collegium Antropologicum. 2008; 32(2):61-5.
9. Rogers HW. Coldiron BM. A relative value unit–based cost comparison of treatment modalities for nonmelanoma skin cancer: Effect of the loss of the Mohs multiple surgery reduction exemption. Journal of the American Academy of Dermatology. 2009; 61(1):96-103.
10. Kapucuoglua N, Basakb PY, Bircana S, Serta S, Akkaya VB. Immunohistochemical galectin-3 expression in non-melanoma skin cancers. Pathology - Research and Practice. 2009; 205(2):97-103.
11. Dumic J, Dabelic S, Flogel M. Galectin-3: an open-ended story. Biochim Biophys Acta. 2006; 1760(4):616-35.
12. Larsen L, Chen HY, Saegusa J, Liu FT. Galectin-3 and the skin. J Dermatol Sci. 2011; 64(2):85-91.

[Downloaded from jhbmi.com on 2025-07-20]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2014; 1(1): 54-62

Modeling a Data Mining Decision Tree and Propose a New Model for the Diagnosis of Skin Cancer by Immunohistochemical Staining Methods

Afshin Sarafi Nejad*[1,2], Amirhossein Saeid[3], Isa Mohammed Rose[4], Alireza Rowhanimanesh[5]

• **Received:** 5 Nov, 2014 • **Accepted:** 1 Dec, 2014

Introduction: New diagnostic methods like immunohistochemistry staining in skin cancer can help the physicians to have more accurate diagnosis. The purpose of this study was to compare a method based on decision tree for differential diagnosis of two kind of skin cancer (Basal cell cancer and Squamous cell cancer) based on the results of staining methods.

Method: Sixty skin cancer patients' data from Malaysia were assessed by two methods of decision tree, CART and CHAID, in data mining and using Clementine 12 and SPSS 19. The results of three staining methods including B-cell lymphoma-2 antibody (BCL2), Galectin-3 (Cytoplasm), and Galectin-3 (Nucleus) were analyzed. The best predictive model for decision tree induction was compared with another researcher-made model based on critical values resulted from Receiver Operating Characteristic (ROC) curve analysis.

Results: In final synthetic model, the sensitivity and specificity for Basal Cell Carcinoma (BCC) were 82.1% and 100%, and for Squamous Cell Carcinoma (SCC) were 100% and 82.8%, respectively. The overall accuracy of the model was 90.38% and the positive predictive values (PPV) for SCC and BCC were 82.1% and 100%, and the positive likelihood ratios (PLR) were 5.8 and 5.5 respectively.

Conclusion: The decision tree model based on two methods of immunohistochemistry staining in skin cancer, can help in the diagnosis of these malignant disease and provide further studies.

Key words: Data mining, Decision tree, Immunohistochemistry, BCL2, Galectin-3, Skin cancer

• **Citation:** Sarafi Nejad A, Saeid AH, Rose I.M, Rowhanimanesh AR. Modeling a Data Mining Decision Tree and Propose a New Model for the Diagnosis of Skin Cancer by Immunohistochemical Staining Methods. Journal of Health and Biomedical Informatics 2014; 1(1): 54-62.

1. MD, MSc, Ph.D Candidate of Medical Informatics,Department of medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Khorasan Razavi, Mashhad, Iran
2. Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
3. MD, Clinical and Surgical Pathologist, Kerman University of Medical Sciences, Kerman, Iran
4. MD, Professor of Pathology, Department of Pathology, School of Medicine, University Kebangsaan Malaysia, Kuala Lumpur, Malaysia
5. Ph.D, Control Engineering, Department of Electrical Engineering, University of Neyshabur, Khorasan Razavi, Neyshabur, Iran.

***Correspondence**: Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Haft Bagh Alavi Highway, Kerman, Iran
• **Tel:** 09131414405 • **Email:** a.sarafi@gmail.com

[Downloaded from jhbmi.com on 2025-07-20]