مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۹۱، دوره ۸، شماره ۲: صفحات ۱۹-۱۳.
مقاله پژوهشی

ارزیابی متغیرهای پیش‌آگهی در رده‌بندی نرخ بقای بیماران مبتلا به سرطان کولورکتال با استفاده از درخت تصمیم

امل ساکی مالحی[۱]، ابراهیم حاجی زاده[۲]، سید رضا فاطمی[۳]

[۱] دانشجوی دوره دکتری، گروه آمار زیستی، دانشگاه تربیت مدرس، تهران، ایران
[۲] دانشیار، گروه آمار زیستی، دانشگاه تربیت مدرس ،تهران، ایران
[۳] فوق تخصص گوارش و کبد، مرکز تحقیقات بیماری‌های گوارش و کبد، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
نویسنده رابط: ابراهیم حاجی زاده، نشانی: تهران، جلال آل احمد، پل نصر، دانشگاه تربیت مدرس، دانشکده علوم پزشکی، گروه آمار زیستی، تلفن: ۸۲۸۸۳۸۱۰، پست الکترونیک: hajizadeh@modares.ac.ir
تاریخ دریافت: ۱۳۹۰/۶/۱۵؛ پذیرش: ۱۳۹۰/۱۱/۸

مقدمه و اهداف: جستجو برای ارزیابی و تحلیل فاکتورهای مهم و موثر در بقای بیماران یکی از مباحث کلیدی در مطالعات سرطان است. مدل درخت تصمیم روش جدیدی است که در تعیین فاکتورهای پیش آگهی بیماری و رده بندی بیمارن بر اساس زیرگروه‌های همگن استفاده می‌شود. در این روش انتخاب رده‌ها برحسب مهم‌ترین فاکتورهای پیش‌آگهی صورت می‌گیرد. هدف از این مطالعه تحلیل داده‌های بقای بیماران مبتلا به سرطان کولورکتال با استفاده از درخت تصمیم است.

روش کار: در این مطالعه، از داده‌های ۷۳۹ بیمار مبتلا به سرطان کولورکتال استفاده شده است. این داده‌ها در مرکز تحقیقات بیماری‌های گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی ثبت شده‌اند. داده‌ها شامل اطلاعات دموگرافی و هیستوپاتولوژیک هستند. پیشامد مورد نظر در این مطالعه مرگ بیماران است و زمان بقای بیماران از زمان تشخیص بیماری تا وقوع پیشامد (یا سانسور شدن) است که بر حسب ماه اندازه‌گیری شده است. برای تحلیل داده‌ها و رده‌بندی بیماران از مدل درخت تصمیم استفاده شد.

نتایج: مدل درخت تصمیم متغیرهای مرحله سرطان در زمان تشخیص (بر حسب TNM)، سن بیمار در زمان تشخیص، متغیر نوع مورفولوژی تومور و درجه سرطان را در سطح معناداری ($P<0/05$) به عنوان فاکتورهای پیش‌آگهی مهم در بقای بیماران مبتلا به سرطان کولورکتال نشان داد. همچنین بیماران بر حسب این فاکتورها به پنج زیرگروه همگن رده‌بندی شدند. مقادیر بزرگتر از ۱ معیار اندازه اختلاف (measure of separation)(SEP)، مناسبت مدل را تائید می‌کند.

نتیجه‌گیری: مدل درخت تصمیم علاوه بر ارزیابی فاکتورهای پیش‌آگهی بیماری، روشی مناسب و قدرتمند در رده‌بندی نرخ بقای بیماران است.

واژگان کلیدی: درخت تصمیم، تحلیل بقا، فاکتورهای پیش‌آگهی، زیر گروه‌های همگن، سرطان کولورکتال

مقدمه

یکی از جنبه‌های مهم در تحقیقات بالینی سرطان علاوه بر بررسی اتیولوژی، اپیدمیولوژی و ارزیابی درمان‌های مختلف، شناسایی و تشخیص فاکتورهای پیش آگهی بیماری[۱] است. در این نوع مطالعات سعی می‌شود که خط سیر بیماری برای گروهی از بیماران بوسیله این فاکتورها پیش‌بینی شود و فاکتورهای مختلف نسبتا مهم رتبه‌بندی شوند (۲-۱).

مدل‌های کلاسیک به عنوان مثال رگرسیون کاکس[۲] روش مناسبی برای تعیین فاکتورهای تشخیصی مهم در پیش‌بینی بقای افراد است، اما به دلیل پیش فرض‌های موجود برای این مدل و پیچیدگی تفسیر نتایج برای محققان پزشکی، در دو دهه اخیر تکنیک‌های دیگری رواج یافته‌اند (۳،۴). از این تکنیک‌ها می‌توان به روش‌های مبتنی بر مدل‌های درختی[۳] اشاره نمود که به آن‌ها روش‌های افراز بازگشتی[۴] نیز گفته می‌شود که بر اساس مشخصات بیماران و قواعد منطقی، داده‌ها را افراز می‌کنند (۳). این مدل‌ها، روش‌های تحلیلی قدرتمند برای کشف ساختار داده‌ها هستند و کاربرد آن‌ها در علوم پزشکی بسیار وسیع است (۵). این مدل

[۱] Prognostic factors
[۲] Cox regression model
[۳] Tree- based methods
[۴] Recursive partitioning methods

برای آنالیز اکتشافی داده‌ها با استفاده از قواعد منطقی و ایجاد رده‌بندی ساده و قابل تفسیر بکار می‌رود. روش‌های مبتنی بر مدل‌های درختی در ابتدا تحت عنوان رده‌بندی و رگرسیون درختی توسط مورگان و سونیکیست[۱] در سال ۱۹۶۳ برای بررسی اثرات متقابل متغیرها در داده‌های علوم اجتماعی پیشنهاد شدند و سپس بطور جامع‌تر توسط بریمن[۲] در سال ۱۹۸۴ مورد بررسی قرار گرفتند (۱) و برای اولین بار این مدل‌ها توسط الشن[۳] در زمینه تحقیقات بالینی بکار گرفته شدند (۲). در سال‌های اخیر این مدل‌ها در آنالیز بقا نیز توسعه یافته‌اند و در رده‌بندی بیماران در معرض عوامل مختلف خطر، برای اقدامات درمانی موثر و یا پیش‌بینی نرخ بقا بدون در نظر گرفتن پیش فرض‌های لازم در روش‌های پارامتری، نقش مهمی ایفا می‌کنند. این مدل‌ها به دلیل عدم نیاز به پیش فرض‌های ضروری در روش‌های پارامتری و سادگی تفسیر نتایج آن‌ها برای محققین بالینی حائز اهمیت هستند. با توجه به مطالعات اندک در زمینه رده‌بندی نرخ بقای بیماران با استفاده از مدل درخت تصمیم در ایران، هدف از این مطالعه، ارائه فاکتورهای پیش‌آگهی مهم و تعیین زیرگروه‌های همگن بر حسب این فاکتورها در بیمارن مبتلا به سرطان کولورکتال است. ارائه این زیرگروه‌ها برای انتخاب پروتکل‌های درمانی موثر و طراحی کارآزمایی‌های بالینی برای مطالعات بعدی بسیار مفید است.

روش کار

داده‌ها از یک مطالعه توصیفی- تحلیلی در فاصله سال‌های ۱۳۷۹ تا ۱۳۸۶ بر اساس پرونده ۷۳۹ بیمار مبتلا به سرطان کولورکتال، ثبت شده در مرکز تحقیقات بیماری‌های گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی جمع‌آوری شده‌اند. افراد تحت مطالعه، بیماران مبتلا به سرطان کولورکتال از بیمارستان‌های در ارتباط با این مرکز در تهران و شهرستان‌ها هستند. شرط ورود به مطالعه برای بیماران بر اساس تشخیص آسیب شناسی در نظر گرفته شد. گزارش پاتولوژی این بیماران در بیمارستان‌های مربوطه و مرکز موجود است. مدت زمان بقای بیماران از زمان تشخیص بیماری آن‌ها تا اول مهر ماه ۱۳۸۶ بر حسب ماه تعیین گردید، که در این زمان افراد یا پیشامد مرگ را تجربه نموده و یا زنده مانده بودند (سانسور راست). از طریق تماس تلفنی مرگ و یا زنده بودن بیماران مورد تایید قرار گرفت. افرادی که دارای عود بیماری بوده‌اند در این مطالعه در نظر گرفته نشده‌اند. برای همه بیماران و بر اساس اطلاعات موجود در پرونده بیمارستانی، مشخصه‌های دموگرافیک شامل سن هنگام تشخیص، جنسیت، نژاد، وضعیت تأهل و سطح تحصیلات و ویژگی‌های بالینی شامل BMI، سابقه مصرف الکل، HNPCC، FAP، IBD، سابقه فامیلی ابتلا به سرطان، مرحله پاتولوژیک و هیستولوژیک و نوع مورفولوژی تومور، استخراج و در پایگاه اطلاعاتی مرکز سابقه ثبت گردید و در تحلیل از آن‌ها استفاده شد.

تحلیل آماری

در مدل درخت تصمیم در تحلیل بقا، متغیر پاسخ مورد نظر زمان بقا است. اگر X زمان بقا باشد، C زمان سانسور شده تصادفی و Z یک بردار p متغیره از متغیرهای کمکی باشد، بنابراین بردار متغیرهای مشاهده شده بصورت زیر است (۳،۵).

$$\{(t_i, \delta_i, z_i) : i = 1,2, \dots N\} \qquad (۱)$$

T زمان بقای مشاهده شده و Δ یک تابع نشان‌گر وقوع مرگ است. فرض می‌کنیم X و C از هم مستقل باشند، در این صورت داده‌ها شامل یک نمونه تصادفی از مشاهدات مستقل هستند.

در مرحله اول آنالیز، برای برازش مدل درختی تمام افرازهای بالقوه برای هر متغیر کمکی ارزیابی می‌شوند. در نهایت متغیر کمکی و نقطه برشی برای افراز داده‌ها انتخاب خواهند شد که بیشترین کاهش در ناخالصی را ارائه دهند. منظور از ناخالصی بیان نامتجانس بودن داده‌ها در یک گره است. کاهش در ناخالصی گره p که به دو گره چپ و راست افراز می شود، به صورت زیر به دست می‌آید:

$$G(p) = R(p) - \left[R\left(l(p)\right) + R\left(r(p)\right)\right] \qquad (۲)$$

که در آن R(p) باقیمانده خطا در یک گره p است، R(r(p)) و R(l(p)) به ترتیب باقیمانده خطا در گره‌های راست و چپ ناشی از افراز گره p هستند. بیشتر تکنیک‌های افراز بازگشتی برای داده‌های سانسور شده بقا، با استفاده از آماره لگاریتم رتبه‌ای، برای محاسبه G(p) بین دو گره اجرا می‌شوند. آماره لگاریتم رتبه‌ای برای یک افراز s که گره ریشه‌ای p را به گره‌های راست r و چپ l افراز می‌کند، بصورت تعریف می‌شود (۶):

$$\emptyset_{LR}(s,p) = \frac{\sum_{i=1}^{k} w_i[a_i - E_0(A_i)]}{\left[\sum_{i=1}^{k} w_i^2 \, \mathrm{var}_0(A_i)\right]^{\frac{1}{2}}} \qquad (۳)$$

Ai متغیر تصادفی متناظر با تعداد مشاهدات سانسور نشده

[۱] Morgan and Sonquist
[۲] Breiman
[۳] Olshen

(مشاهده رخداد مورد نظر) در گره راست برای i امین گره است. فرض صفر که بوسیله این آماره آزمون می‌شود، یکسان بودن نرخ شکست در دو گره راست و چپ ناشی از افراز است. wi در آن ثابت است و برای وزن‌دهی به گره‌های مربوطه است (۷). بنابراین معنادار شدن این آزمون به معنای انتخاب یک متغیر با نقطه برش مناسب برای افراز است. امید و واریانس آماره بر اساس توزیع صفر، فوق هندسی با حاشیه‌های ثابت است که عبارتند از:

$$E_0(A_i) = \frac{m_{i1}n_{i1}}{n_i} \qquad (٤)$$

$$var_0(A_i) = \left[\frac{m_{i1}(n_i - m_{i1})}{n_i - 1}\left(\frac{n_{i1}}{n_i}\right)\left(1 - \frac{n_{i1}}{n_i}\right)\right] \qquad (٥)$$

اگر $n_{i1} = r_r(t_i)$ و $n_i = r_p(t_i)$ در نظر گرفته شوند، آنگاه $r_r(t_i)$ و $r_p(t_i)$ تعداد افراد در معرض خطر در زمان t_i در گره ریشه‌ای و گره راست هستند. δ_i نیز نشانگر وضعیت زمان بقا (سانسور شدن) است که $\delta_i = 1$ نشان دهنده وقوع رخداد مورد نظر است. بنابراین معادلات زیر به ترتیب نشان دهنده تعداد مشاهده شده و تعداد مورد انتظار مرگ در گره راست هستند.

$$\sum_{i=1}^{k} a_i = \sum_{i \in r} \delta_i \qquad (٦)$$

$$\sum_{i=1}^{k} E_0(A_i) = \sum_{i \in p, \delta_i = 1} \frac{r_r(t_i)}{r_p(t_i)} \qquad (٧)$$

با استفاده از افراز بازگشتی مبتنی بر آماره شرح داده شده، در نهایت گره‌هایی از افراد با ویژگی‌های مشابه (همگن) تشکیل خواهند شد و گره‌ها دارای حداکثر تفاوت ممکن در نرخ شکست (احتمال بقا) خواهند بود. به عبارت دیگر افراد به زیرگروه‌های همگن با حداکثر تفاوت بر حسب پاسخ مورد نظر افراز خواهند شد. فرآیند افراز درخت تا جائی ادامه می‌یابد که تمام مشاهدات درون گره‌های پایانی دارای توزیع یکسانی باشند. اما از آنجایی که در هر فرآیند مدل‌سازی رسیدن به یک مدل بهینه مدنظر است، بنابراین در این روش نیز برازش یک درخت بهینه جزء اهداف مورد نظر می‌باشد. درخت بهینه، درختی است که از لحاظ اندازه و برآورد رده‌بندی اشتباه، بهینه باشد. در این مدل‌ها با استفاده از تکنیک‌های مرسوم، درخت را هرس کرده و در نهایت از بین زیردرخت‌های هرس شده با استفاده از قاعده‌ایی مناسب درختی با مینیمم خطا، هزینه و اندازه مناسب انتخاب می‌گردد. برای هرس درخت تصمیم دو روش کلی پیشنهاد شده است. روش مبتنی بر تغییرات درون گره‌ای که بر اساس تکنیک هزینه– پیچیدگی درخت است (۸) و روش مبتنی بر اختلاف بین گره‌ها که بر اساس تکنیک افراز- پیچیدگی است (۸،۹). منظور از عبارت پیچیدگی در دو تکنیک اندازه درخت است که همان تعداد گره‌های پایانی درخت می‌باشد. برای بررسی توانایی پیش‌بینی مدل، نسبت خطر و یا خطرنسبی در زیرگروه‌ها با هم مقایسه می‌شوند. برای انجام این مقایسه از یک معیار اندازه‌گیری به عنوان اندازه اختلاف[۱] استفاده می‌شود:

$$\mathrm{SEP} = \exp\left[\sum_{i=1}^{5} \frac{n_i}{n} \left|\hat{\beta}_j\right|\right] \qquad (٨)$$

n_j تعداد بیماران در زیرگروه j ام است و $\hat{\beta}_j$ برآورد لگاریتم نسبت خطر یا لگاریتم خطر نسبی بیماران در معرض خطر در زیرگروه j ام نسبت به یک طبقه مرجع است. مقادیر بزرگتر از ۱ این معیار نشان دهنده مناسبت مدل است. زیرگروه پنجم به عنوان گروه مرجع در نظر گرفته شده است. برای تجزیه و تحلیل داده‌ها از نرم افزار R استفاده شده است.

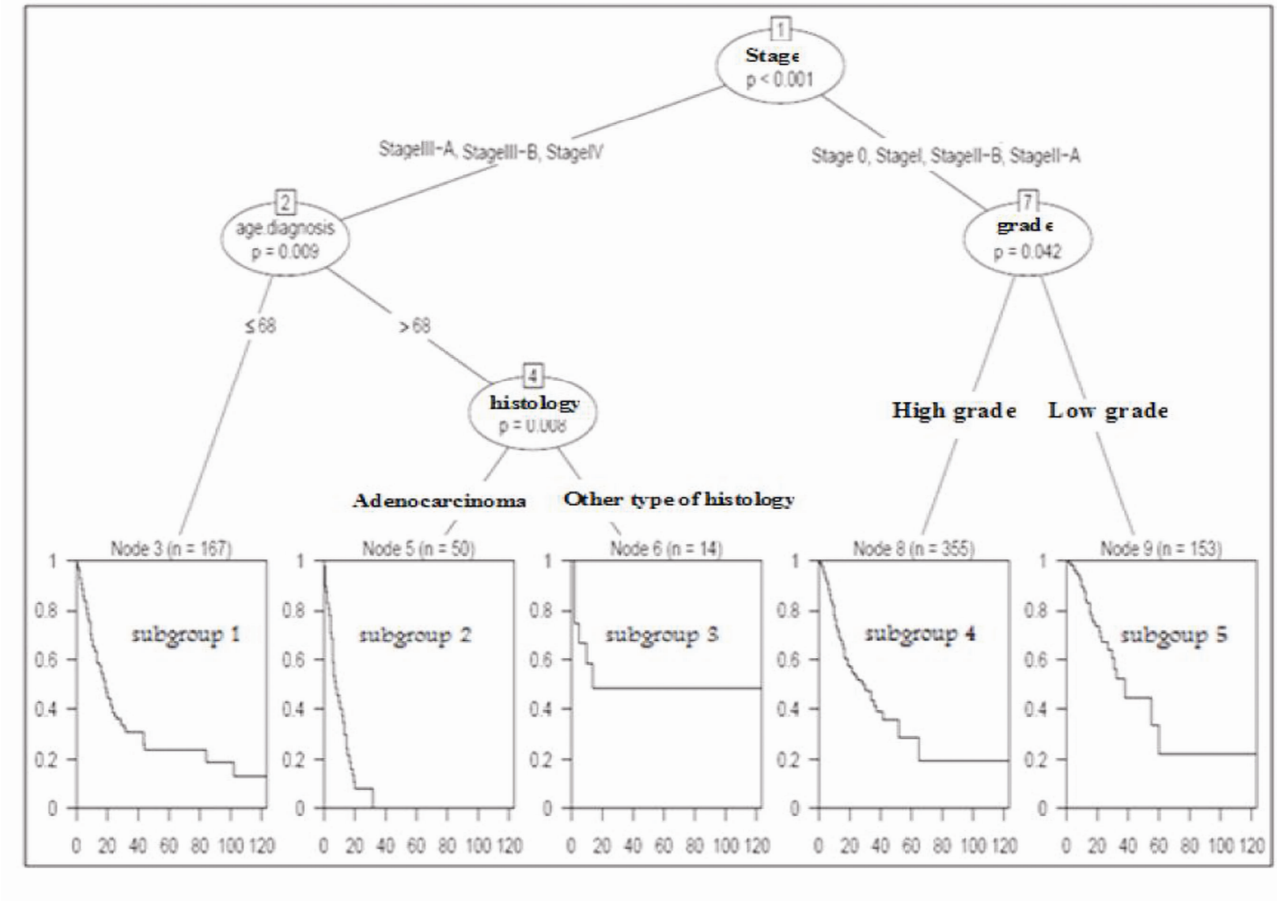
یافته‌ها

۷۳۹ بیمار مبتلا به سرطان کولورکتال ثبت شده در مرکز تحقیقات گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی، به کمک مدل مورد بحث، مورد تحلیل قرار گرفتند. از این تعداد بیمار ۵۲۶ نفر (۷۱/۲٪) مرد و ۲۱۳ (۲۸/۸٪) زن بودند. سن تشخیص بیماری بین ۸۸-۲۰ سال با میانگین (± انحراف معیار)، (۱۲/۸۵) ±۵۹/۷ می‌باشد. زمان بقای بیماران از زمان تشخیص بیماری تا زمان وقوع پیشامد مورد نظر (مرگ) یا سانسور شدن (پایان مطالعه) بر حسب ماه در نظر گرفته شده است. میانگین و میانه بقا ± (انحراف معیار) به ترتیب برابر (۳/۴)± ۴۲/۴۶ و (۲/۷)±۲۲/۸ هستند و احتمال بقای ۵ ساله بیماران برابر ۶۳/۳٪ است. در برازش مدل در اولین مرحله مهم‌ترین متغیری که برای افراز داده‌ها انتخاب می‌شود، متغیر مرحله تومور در زمان تشخیص سرطان (بر حسب TNM) است. بوسیله این افراز ۳۱٪ از بیماران در Stage III-A , III-B, IV رده‌بندی می‌شوند. با تکرار الگوریتم برای گره‌های حاصل از افراز مرحله اول، سن تشخیص بیماری با P =۰/۰۰۹ به عنوان مهم‌ترین متغیر بعدی در رده‌بندی بیماران انتخاب می‌شود. انتخاب نقطه افراز ۶۸ سال برای سن تشخیص بیماری به عنوان بهترین نقطه برش، براساس یافتن بیشترین

[۱] Measure of separation

اختلاف بین بقای دو زیرگروه و در نتیجه افراز بیماران به دو زیرگروه همگن انتخاب شده است. سپس متغیرهای نوع مرفولوژی تومور و درجه تومور (grade) به عنوان دیگر متغیرهای پیش‌آگهی شناسایی شدند. شکل شماره ۱ مدل درخت تصمیم و زیرگروه‌های افراز شده بیماران به همراه مهم‌ترین متغیرها به عنوان متغیرهای پیش‌آگهی بیماری را نشان می‌دهد. میانگین و میانه بقا در زیرگروه‌های حاصل از مدل درخت تصمیم در جدول شماره ۱ گزارش شده است. زیرگروه پنجم یعنی بیماران با Stage 0,I,II-A,II-B و درجه پایین تومور (G1, G2) دارای بیشترین میانگین و میانه بقا هستند. در گروه سوم به دلیل کم بودن حجم نمونه (۱۴ نفر در این زیرگروه هستند) و بزرگ‌تر بودن تعداد داده‌های سانسور شده نسبت به رخدادها (۸ داده سانسور شده)، امکان برآورد خطای معیار وجود ندارد.

نتیجه آزمون مقایسه نرخ بقا در پنج زیرگروه، بیان گر اختلاف معنادار بین این زیرگروه‌ها است (P<۰/۰۵). معیار SEP نیز برابر ۲/۳ است. بنابراین اختلاف معنادار بین زیرگروه‌ها و مقدار بزرگ‌تر از ۱ معیار SEP کارایی مدل تایید شد.



نمودار شماره ۱- درخت تصمیم

جدول شماره ۱- برآورد میانگین، میانه و خطای استاندارد در پنج زیرگروه حاصل از مدل

گروه	برآورد میانگین	خطای استاندارد	برآورد میانه	خطای استاندارد
۱	۳۷/۸۸۱	۴/۷۲۱	۱۸/۸	۱/۷۶۱
۲	۱۰/۴۹۵	۱/۳۴۹	۷/۴	۲/۲۱۰
۳	۲۱/۲۴۷	۴/۷۰۲	۱۳/۸	.
۴	۳۸/۰۲۵	۴/۸۴۹	۲۷/۵	۴/۰۴۷
۵	۵۱/۵۹۴	۹/۱۴۶	۳۷/۸	۵/۶۹۶

بحث

در سال‌های اخیر مطالعات اپیدمیولوژیک بیانگر سرعت رو به رشد نرخ ابتلا به سرطان کولورکتال در ایران هستند (۱۱-۱۰). بنابراین به عنوان یک مسئله مهم در بهداشت عمومی و تحقیقات سرطان کشور مطرح است. از آنجایی که ارزیابی فاکتورهای پیش‌آگهی و رده‌بندی بیماران در هر بیماری به خصوص در بیماری سرطان کمک شایانی به محققان بالینی ارائه می‌دهد، این مطالعه به تحلیل داده‌های سرطان کولورکتال پرداخته است. در اغلب این گونه مطالعات از مدل‌های کلاسیک پارامتری و نیمه پارامتری استفاده می‌شود. اما این مدل‌ها علاوه بر نیاز به برقراری پیش‌فرض‌های لازم، قادر به تعیین یک رده‌بندی درست از بیماران نیستند. مقبولیت روش‌های مبتنی بر مدل‌های درختی در زمینه‌های علوم پزشکی ناشی از نیاز محققین بالینی برای تعریف ضوابط رده‌بندی مشخص، برای تشخیص و پیش‌بینی بیماری است (۱۲). در واقع ایده اساسی این مدل‌ها تشکیل زیرگروه‌هایی از بیماران است، که این زیرگروه‌ها با در نظر گرفتن پیشامد مورد نظر همگن باشند (۴،۱۳). با تعیین زیرگروه‌ها می‌توان کارآزمایی‌های بالینی[۱] برای مطالعات بعدی را نیز طراحی کرد. به این ترتیب که با تشکیل زیرگروه‌های متفاوت از بیماران، بیماران با شرایط مناسب برای ورود به مطالعات بالینی با هدف درمانی نیز مشخص می‌شوند (۳). یافته‌های حاصل این مدل با استفاده از تحلیل درخت تصمیم، متغیر مرحله سرطان را به عنوان یکی از فاکتورهای پیش‌آگهی مهم نشان می‌دهد. مطالعات مختلفی در این راستا وجود دارند که این موضوع را تایید می‌کنند که از بین آن‌ها می‌توان به مطالعه شایانفر و همکاران در سال ۸۸ و مطالعه مقیمی دهکردی و همکاران در سال ۸۷ در ایران اشاره کرد. (۱۸-۱۴) در واقع اهمیت این متغیر به این دلیل است که این متغیر شامل اطلاعاتی در مورد توصیف سرطان در بدن است؛ اینکه سرطان تا چه حد به دیواره روده، غدد لنفاوی و یا اندام‌های دیگر سرایت کرده باشد. سن در زمان تشخیص بیماری نیز متغیری است که در برخی مقالات به معنادار بودن تاثیر آن در زمان بقا اشاره شده است (۲۰-۱۸). اما مطالعاتی نیز وجود دارند که به عدم تأثیر سن زمان تشخیص در بقای بیماران اشاره دارند؛ از جمله این مطالعات مطالعه صفایی و همکاران در سال ۸۸ است (۱۷،۱۵). البته شاید این نتایج متناقض به دلیل نقاط برش مختلف سن در مطالعات مختلف و یا به دلیل ناحیه درگیر سرطان (کولون یا رکتال) باشد. در مطالعات آخوند و همکاران و اصغری و همکاران در سال‌های ۹۰ و ۸۹ که نواحی درگیر به طور مجزا ارزیابی شده‌اند، سن زمان تشخیص بیماری یکی از عوامل پیش‌آگهی در سرطان کولون شناسایی شده است (۲۲-۲۱). درجه سرطان (grade) یکی دیگر از عواملی است که بقای بیماران را تحت تأثیر قرار می دهد. این عامل نشان دهنده شباهت سلول‌های تومور به سلول‌های نرمال است و هر چه این سلول‌ها تفاوت بیشتری با سلول‌های نرمال داشته باشند، سرعت پخش سلول‌های سرطانی بیشتر است و در مطالعات بسیاری این متغیر به عنوان یک متغیر موثر در بقای بیماران اشاره شده است. مطالعه پارک[۲] و همکارانش یکی از این مطالعات مورد اشاره است (۱۸-۱۷،۲۰،۱۵). در یافته‌های این مطالعه نوع مورفولوژی تومور به عنوان یکی دیگر از عوامل پیش آگهی سرطان کولورکتال معرفی شده است. در این زمینه نیز مطالعات متفاوتی با نتایج متفاوت وجود دارد (۲۵-۲۲،۲۰،۱۵). بر اساس یافته‌های بیان شده در این مطالعه افراد به پنج رده تقسیم شدند که درون هر رده افراد نسبتا همگن وجود دارند.

نتیجه‌گیری

مدل‌های مبتنی بر درخت تصمیم نیاز به برقراری هیچ نوع پیش‌فرضی برای مدل‌سازی ندارند و به همین دلیل محدودیت مدل‌های پارامتری و شبه‌پارامتری را دارا نیستند. در این مطالعه مدل درختی علاوه بر معرفی متغیرهای مهم در پیش‌آگهی بیماری، زیرگروه‌های همگن از بیماران را برای تحلیل‌های بالینی بعدی ارائه می‌دهد. تحلیل این زیرگروه‌ها برای مطالعات کارآزمایی و پروتکل‌های درمانی برای محققان بالینی بسیار حائز اهمیت هستند. اما یکی از محدودیت‌های مدل‌های مبتنی بر درخت تصمیم همانند اغلب مدل‌های آماری، عملکرد ضعیف مدل در صورت وجود همخطی بین متغیرهاست. در این صورت بهترین متغیر و نقطه برش برای افراز به خوبی تعیین نخواهد شد.

تشکر و قدردانی

مراتب تشکر و قدردانی خود را نسبت به مرکز تحقیقات گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی و دانشگاه تربیت مدرس ابراز می‌داریم.

[۱] Clinical trial

[۲] Park

منابع

1. Negassa A, Ciami A, Abrahamowicz M, Shapiro S, Boivin J. Tree Structured prognostic classification for censored survival data: Validation of computationally inexpensive model selection criteria. Statistical and computation and simulation. 2000; 67: 289-317.
2. Breiman L, Friedman JH, OlShen RA and Ston CJ. Classification and Regression Trees.California, A Division of Wadsworth Inc 1984.
3. LeBlanc M. Handbook of Statistics in Clinical Oncology. In J Crowley (ed.), Tree-Based Methods for Prognostic Stratiification, New York, Basel, Marcel Dekker Inc 2001; 457–72.
4. Ulm K, Nekarda H, Gerein Panf Berger U. Handbook of Statistics in Clinical Oncology. In J Crowley (ed.), Statistical Methods to Identify Prognostic Factor, New York, Basel, Marcel Dekker Inc. 2001 379–95.
5. Banerjee M, Noone A M. Advances in the Biomedical Sciences. In Biswas A. and et al. (ed), New Jersey, John Wiley& Sons, Inc. 2008; 265-85.
6. Bacchetti P, segal M. Survival trees with time-dependent covariates: Application to Estimating Changes in the Incubation Period of AIDs. Biostatistics.1994; 39: 1-20.
7. Segal M. Regression trees for Censored data. Biometrics. 1988; 44: 35-48.
8. Zhang H. Recursive Partitioning in the Health Sciences. U.S.A, Springer 1999.
9. Negassa A, Ciami A, Abrahamowicz M, Shapiro S, Boivin J. Tree Structured subgroup analysis for censored survival data: Validation of computationally inexpensive model selection criteria. Statistics and computing 2005; 15: 231-9.
10. Kalavi B. Colorectal cancer and its epidemiological aspects in Iran (2004). Turk J Gastroenterol 2005; 16: 248-249.
11. Shafayan B, Keyhani M. Epidemiological evaluation of colorectal cancer. Acta Medica Iranica 2003; 41: 156- 160.
12. Noon AM, Banerjee M. Computational Methods in Biomedical Research. In Chow S C, Jones B, Liu P J and Peace K (ed), New York, Chapman& Hall, 2008; 77-101.
13. Schumacher M, Holl¨ander N, Schwarzer G, Sauerbrei W. Handbook of Statistics in Clinical Oncology. In J Crowley (ed.), Prognostic Factor Studies, New York, Basel, Marcel Dekker Inc 2001; 321–78.
14. Shayanfar N, Shahzadi SZ. Immunohistochemical assessment of neuroendocrine differentiation in colorectal carcinomas and Its relation with age, sex and grade plus stage. Iranian Journal of Pathology 2009; 4: 167-71.
15. Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. International Journal Colorectal Disease. 2008; 23: 683- 8.
16. Oh H-S, Chung H-J, Kim H-K, Choi J-S. Differences in overall survival when colorectal cancer patients are stratified into new TNM staging strategy. Cancer Research Treatment 2007; 39: 61-4.
17. Safaee A, Moghimi-dehkordi B, Fatemi S, Ghiasi S, Zali M. Pathology and Prognosis of Colorectal Cancer. Iranian Journal of Cancer Prevention 2009; 2: 137-41.
18. Laohavinij S, Maneechavakajorn J. Prognostic factors for survival in colorectal cancer patients. Journal of Medical Association Thailand. 2010; 93: 1156-66.
19. Molaei M, Mansoori BK, Ghiasi S, Khatami F, Attarian H, et al. Colorectal cancer in Iran: immunohistochemical profilesof four mismatch repair proteins. International Journal Colorectal Disease.2010; 25: 63 9.
20. Park YJ, Park KJ, Park J-G, Lee KU, Choe KJ, et al. Prognostic Factors in 2230 Korean Colorectal Cancer Patients: Analysis of Consecutively Operated Cases. WORLD Journal of SURGERY 1999; 23: 6.
21. Asghari Jafarabadi M,Hajizade E,Kazemnejad A,Fatemi SR. Comparison the Role of BMI, Pathologic Stage and Hereditary Related Factors on Survival between Colon and Rectal Cancers: Frailty Competing Risks Model. Iranian Journal of Epidemiology 2010; 6: 35-49. Akhoond MR, Kazemnejad A, Hajizade E, Fatemi SR, Motlagh A. Investigation of Influential Factors Affecting Survival Rate of Patients with Colorectal Cancer using Copula Function. Iranian Journal of Epidemiology 2011; 6: 40-9.
22. Kalavi B. Colorectal cancer and its epidemiological aspects in Iran (2004). Turk J Gastroenterol. 2005; 16: 248-9.
23. Pahlavan PS, Kanthan R. The epidemiology and clinical findings of colorectal cancer in Iran. Journal Gastrointest Liver Disease. 2006; 15: 15-19.
24. Chew MH, Koh PK, Ng KH. Improved survival in an Asian cohort of young colorectal cancer patients: an analysis of 523 patients from a single institution. International Journal Colorectal Disease 2009; 24: 1075–83.

Iranian Journal of Epidemiology 2012; 8(2): 13-19.

Original Article

Evaluation of Prognostic Variables for Classifying the Survival In Colorectal Patients using The Decision Tree

Saki Malehi A[1], Hajizade E[2], Fatemi R[3]

1- Biostatistics Department, Tarbiat Modares University, Tehran, Iran
2- Biostatistics Department, Tarbiat Modares University, Tehran, Iran
3- Research Center of Gastroenterology and Liver Diseases, Shahid Beheshti Medical University, Tehran, Iran.
Corresponding author: Hajizadeh E., hajizade@modares.ac.ir

Background & Objectives: Identifying the important influential factors is a great challenge in oncology studies. Decision tree is one of methods that could be used to evaluate the prognostic factors and classifying the patients' homogeneously. This method identifies the main prognostic factors and then determines the subgroups of patients based on those prognostic factors. The aim of this study was to assess the prognostic factors and homogeneous subgroups of colorectal patient through survival tree.

Methods: Data collected from an observational of 739 colorectal patients registered in the cancer registry affiliated to the center of Research Center of Gastroenterology and Liver Disease (RCGLD), Shahid Beheshti Medical University, Tehran, Iran. Death was the interested event and the survival time was calculated from date of diagnosis until occurrence of event (or censoring) in months. Finally we used decision tree based method for classifying and analyzing the data.

Results: Based on our result, decision tree identified four covariates as important prognostic factors in 0.05 significant levels: general stage of cancer, age of diagnosis, histology of tumor and morphology type of tumor. Also patients based on these prognostic factors divided into five homogeneous subgroups. The greater values of measure of separation (SEP) criterion support the appropriateness of this model for such the data.

Conclusion: Decision tree is powerful and intuitive method. It has a key feature that in addition to evaluate the prognostic factors, provides the homogeneous subgroups for future analysis.

Keywords: Decision tree, Survival analysis, Prognostic factors, Homogeneous subgroups, Colorectal cancer