

## ارزیابی متغیرهای پیش آگهی در رده بندی نرخ بقای بیماران مبتلا به سرطان ریه با استفاده از مدل درخت تصمیم در استان آذربایجان غربی

مالک اباذری<sup>۱</sup>، رضا اباذری<sup>۲</sup>، قدرت اله روشنایی<sup>۳</sup>، مهدیا غلام نژاد<sup>۴</sup>، یوسف روستا<sup>۵</sup>، حسین محبوب<sup>۶\*</sup>

<sup>۱</sup> دانشجوی دوره کارشناسی ارشد آمار زیستی، گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۲</sup> باشگاه پژوهشگران و نخبگان جوان دانشگاه آزاد اسلامی واحد اردبیل، دانشگاه آزاد اسلامی، اردبیل، ایران

<sup>۳</sup> استادیار، دکترای آمار زیستی، گروه آمارزیستی و اپیدمیولوژی، مرکز تحقیقات مدل سازی بیماری های غیر واگیر، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۴</sup> استادیار، فوق تخصص بیماری های ریه، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران

<sup>۵</sup> رزیدنت داخلی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران

<sup>۶</sup> استاد، دکترای آمار زیستی، گروه آمار زیستی و اپیدمیولوژی، مرکز تحقیقات علوم بهداشتی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

\* نشانی نویسنده مسؤول: همدان، دانشگاه علوم پزشکی، مرکز تحقیقات علوم بهداشتی، دکتر حسین محبوب

E-mail: mahjub@umsha.ac.ir

وصول: ۹۳/۹/۱۱، اصلاح: ۹۳/۱۰/۲۳، پذیرش: ۹۳/۱۱/۲۸

### چکیده

**زمینه و هدف:** سرطان ریه به علت شانس پایین زنده ماندن از مهم ترین و کشنده ترین نوع سرطان محسوب می شود و بیشترین میزان مرگ و میر را در مقایسه با سرطان های سینه، پروستات و سایر سرطان های چندانگانه دارد. هدف از این مطالعه، تحلیل داده های بقای بیماران مبتلا به سرطان ریه در استان آذربایجان غربی می باشد.

**مواد و روشها:** در این مطالعه همگروهی تاریخی اطلاعات مربوط به ۳۵۵ بیمار مراجعه کننده به بیمارستان های استان آذربایجان غربی در فاصله سال های ۹۳-۱۳۸۶ مورد بررسی و وضعیت بقای بیماران تا پایان خرداد ۱۳۹۳ مورد پیگیری قرار گرفت. داده ها شامل اطلاعات دموگرافی و هیستوپاتولوژیک هستند. زمان بقای بیماران از زمان تشخیص تا مرگ یا پایان مطالعه در نظر گرفته شد. برای تحلیل داده ها و رده بندی بیماران از مدل درخت تصمیم استفاده گردید. کلیه ی تحلیل ها با استفاده از نرم افزار R صورت گرفت.

**یافته ها:** میانگین و میانه ی زمان بقا برای ۳۵۵ بیمار مبتلا به سرطان ریه به ترتیب ۱۳ ماه و ۸/۴ ماه برآورد گردید. احتمال بقای ۱ ساله ۳۹ درصد بوده و تنها ۷ درصد از بیماران، بقای ۳ ساله را تجربه کردند. مدل درخت تصمیم متغیرهای وضعیت درمان، Eastern Cooperative Oncology Group (ECOG)، وضعیت سیگاری بودن، سن تشخیص بیماری و مرحله ی تومور به عنوان متغیرهای مهم شناسایی کرد.

**نتیجه گیری:** استفاده از مدل درخت رده بندی، می تواند منجر به شناسایی ریسک فاکتورهای مهم و موثر بر بقای بیماران شود و شناسایی این عوامل در بیماران سرطانی می تواند باعث جلوگیری از پیشرفت بیماری شده و زمان بقای بیماران را افزایش دهد.

**واژه های کلیدی:** سرطان ریه، ایران، درخت تصمیم، بقا.

**مقدمه**

سرطان، یکی از مهم‌ترین علت‌های اختلالات، مرگ‌ومیر و ناتوانی در سراسر جهان است (۱). این بیماری که یکی از بیماری‌های شایع و در حال افزایش است، حجم زیادی از تلاش‌های نظام‌های مراقبتی را به خود اختصاص می‌دهد (۲، ۳). سرطان، نوعی بیماری است که در آن، سلول‌های غیرطبیعی به شکل غیرقابل مهار تقسیم می‌شوند و قادرند به سایر بافت‌های بدن گسترش یابند (۴). سرطان در ایران بعد از بیماری‌های قلبی-عروقی و سوانح، سومین علت مرگ‌ومیر می‌باشد (۵). سرطان‌ها در ایران باعث تحمیل ۶۶۲/۲ دالی (Disability Adjusted Life Years: DALY) در هر صد هزار نفر می‌شوند و سرطان ریه، بعد از سرطان معده با ایجاد ۵۶/۳ دالی در رتبه‌ی دوم می‌باشد (۳). سرطان ریه، بیشترین میزان مرگ‌ومیر را در مقایسه با سرطان‌های سینه، پروستات و سایر سرطان‌های چندگانه (ابتلای فرد به بیش از یک سرطان) دارد (۶). سرطان ریه به علت شانس پایین زنده ماندن از مهم‌ترین و کشنده‌ترین نوع سرطان محسوب می‌گردد. به طوری که تنها ۱۳٪ از ابتلایان به سرطان ریه پنج سال پس از تشخیص زنده می‌مانند (۷، ۸). براساس تقسیم‌بندی انجمن سرطان آمریکا (American Cancer Society) دو نوع سرطان وجود دارد: سرطان ریه با یاخته‌های کوچک (Small cell lung cancer; SCLC) و سرطان ریه با یاخته‌های غیرکوچک (Non-Small cell lung cancer; NSCLC) (۹). متأسفانه چندین سال طول می‌کشد تا نشانه‌های سرطان ریه خود را بروز دهد. لذا در بیشتر مواقع این بیماری در مراحل بسیار پیشرفته تشخیص داده می‌شود (۱۰، ۱۱). میزان بقا پس از تشخیص، درمان و بررسی عوامل خطر ساز در بیماران سرطانی یکی از شاخص‌های مهم در کنترل بیماری و ارزیابی روش‌های درمانی مورد استفاده است (۱۲). در اکثر تحقیقات پزشکی از مدل نیمه پارامتری کاکس برای ارزیابی عوامل استفاده شده است (۱۳-۱۶). مدل کاکس بر فرض اساسی، بر متناسب بودن خطرات مبتنی است که

در بیشتر داده‌ها ممکن است این فرض برقرار نباشد. لذا در دو دهه‌ی اخیر، تکنیک‌های دیگری از قبیل روش‌های مبتنی بر مدل‌های درختی (Tree-based methods)، که به آن‌ها «روش‌های افراز بازگشتی» (Recursive partitioning methods) نیز گفته می‌شود، رواج یافته‌اند (۱۷). این مدل‌ها روش‌های تحلیلی قدرتمندی برای کشف ساختار داده‌ها به‌شمار می‌روند و کاربرد بسیار گسترده‌ای در علوم پزشکی دارند (۱۸). این مدل برای آنالیز اکتشافی داده‌ها با استفاده از قواعد منطقی و ایجاد رده‌بندی ساده و قابل تفسیر به‌کار می‌روند (۱۹). روش‌های مبتنی بر مدل‌های درختی در ابتدا تحت عنوان «رده‌بندی و رگرسیون درختی» توسط مورگان و سونیکویست (Morgan and Sonquist) در سال ۱۹۶۳ برای بررسی اثرات متقابل متغیرها در داده‌های علوم اجتماعی پیشنهاد شدند و سپس به‌طور جامع‌تر توسط بریمن (Breiman) در سال ۱۹۸۴ مورد بررسی قرار گرفتند (۲۰) این مدل‌ها برای اولین بار توسط اولسن (Olshen) در زمینه‌ی تحقیقات بالینی به‌کار گرفته شدند (۲۱). این مدل‌ها به دلیل عدم نیاز به پیش‌فرض‌های ضروری در روش‌های پارامتری و سادگی تفسیر نتایج، در آنالیز بقا نیز توسعه یافته و در رده‌بندی بیماران در معرض عوامل خطر و پیش‌بینی نرخ بقا، به‌کار گرفته می‌شوند (۱۸). با توجه به مطالعات اندک در زمینه رده‌بندی نرخ بقای بیماران با استفاده از مدل درخت تصمیم در ایران، هدف از این مطالعه، ارائه‌ی فاکتورهای پیش‌آگهی مهم و تعیین زیر گروه‌های همگن در بیماران مبتلا به سرطان ریه است.

**مواد و روش‌ها**

در این مطالعه که به‌صورت کوهورت تاریخی انجام گرفته است، داده‌ها از پرونده‌ی ۳۵۵ بیمار مبتلا به سرطان ریه در فاصله‌ی سال‌های ۱۳۸۶ تا ۱۳۹۳ که به بیمارستان‌های استان آذربایجان غربی مراجعه کرده‌اند، ثبت گردید. شرط ورود به مطالعه برای بیماران، تشخیص سرطان ریه در نظر گرفته شده و گزارش پاتولوژی این

شده‌ی بقا، برای محاسبه  $G(p)$  بین دو گره، از آماره لگاریتم رتبه‌ای استفاده می‌شود. آمار لگاریتم رتبه‌ای برای یک افراز  $s$  که گره ریشه‌ای  $p$  را به گره‌های راست  $r$  و چپ  $l$  افراز می‌کند به صورت ذیل تعریف می‌شود:

$$TW = \frac{\sum_{i=1}^k w_i [a_i - E_0(A_i)]}{[\sum_{i=1}^k w_i^2 var_0(A_i)]^{1/2}}$$

اینجا  $A_i$  متغیر تصادفی متناظر با تعداد رخداد مورد نظر (مشاهدات سانسور نشده) در گره راست برای تأمین گره است.

فرض صفر که به وسیله‌ی این آماره آزمون می‌شود، یکسان بودن نرخ شکست در دو گره راست و چپ ناشی از افراز است.  $w_i$  در آن ثابت است و برای وزن دهی به گره‌های مربوط است. بنابراین معنادار شدن این آزمون به معنای انتخاب یک متغیر با نقطه‌ی برش مناسب برای افراز است. با استفاده از افراز بازگشتی مبتنی بر آماره‌ی شرح داده‌شده، 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 برحسب پاسخ مورد نظر افراز خواهند شد. فرایند افراز درخت تا جایی ادامه می‌یابد که تمام مشاهدات درون گره‌های پایانی دارای توزیع یکسانی باشند. در هر فرآیند افراز مدل‌سازی، هدف رسیدن به یک مدل بهینه است.

درخت بهینه، درختی است که از لحاظ اندازه و برآورد اشتباه، بهینه باشد. معنادار بودن نرخ بقای زیر گروه‌ها با استفاده از آزمون لگاریتم رتبه‌ای نشان‌دهنده‌ی بهینه بودن درخت و در نهایت کارایی مدل است. در روش‌های مبتنی بر مدل‌های درختی، برای رسیدن به درخت بهینه، با استفاده از تکنیک‌های هرس، درخت را هرس کرده و نهایتاً با استفاده از قاعده‌ی مناسب، درختی با مینیمم خطا و اندازه مناسب انتخاب می‌گردد. برای هرس کردن درخت تصمیم دو روش کلی پیشنهاد شده- است: روش مبتنی بر تغییرات درون گره‌ای که بر اساس

بیماران در بیمارستان‌های مربوط موجود است. مدت زمان بقای بیماران از زمان تشخیص بیماری آن‌ها تا ۱۵ خرداد ۹۳ برحسب ماه تعیین شد که در این زمان، افراد یا پیشامد مرگ را تجربه کرده (رخداد مورد نظر) و یا زنده مانده بودند (سانسور راست). از طریق تماس تلفنی، مرگ یا زنده ماندن بیماران پیگیری و تایید گردید و افرادی که مرگ آن‌ها علت‌هایی غیر از بیماری سرطان ریه داشت به عنوان سانسور منظور گردید. افرادی که دارای عود بیماری بودند در این مطالعه در نظر گرفته نشدند. برای همه بیماران بر اساس اطلاعات موجود در پرونده‌ی بیمارستانی، متغیرهای دموگرافیک همچون: سن، تشخیص بیماری، جنسیت و متغیرهای بیولوژیک شامل عملکرد ECOG، مصرف دخانیات، نوع تومور، مرحله‌ی بیماری، گسترش بیماری (Metastasis)، وضعیت درمان (بیمارانی که یک یا بیش از یک روش درمان کموتراپی، رادیوتراپی و جراحی را تجربه کرده بودند و بیمارانی که به علت عدم رضایت بیمار یا همراهان بیمار یا ناتوانی شدید جسمانی و بیماری‌های زمینه‌ای دیگر، درمان اختصاصی دریافت نکرده بودند)، استخراج و ثبت گردید.

## تحلیل‌های آماری

در روش‌های مبتنی بر مدل‌های درختی، برای برازش مدل، تمام افرازهای ممکن برای هر متغیر کمکی ارزیابی می‌شوند. در نهایت متغیر کمکی و نقطه‌ی برشی برای افراز داده‌ها انتخاب خواهد شد که بیشترین کاهش در ناخالصی را ارایه دهند. منظور از ناخالصی، بیان ناهمگنی داده‌ها در یک گره است. کاهش ناخالصی در گره  $p$  که به دو گره چپ و راست افراز می‌شود به صورت ذیل به دست می‌آید:

$$G(p) = R(p) - [R(l(p)) + R(r(p))]$$

که در آن  $R(p)$  باقیمانده‌ی خطا در یک گره است،

$R(r(p))$  باقیمانده‌ی خطا در گره راست و

$R(l(p))$  باقیمانده‌ی خطا در گره چپ، ناشی از افراز گره  $p$  هستند.

در بیشتر تکنیک‌های افراز بازگشتی برای داده‌های سانسور

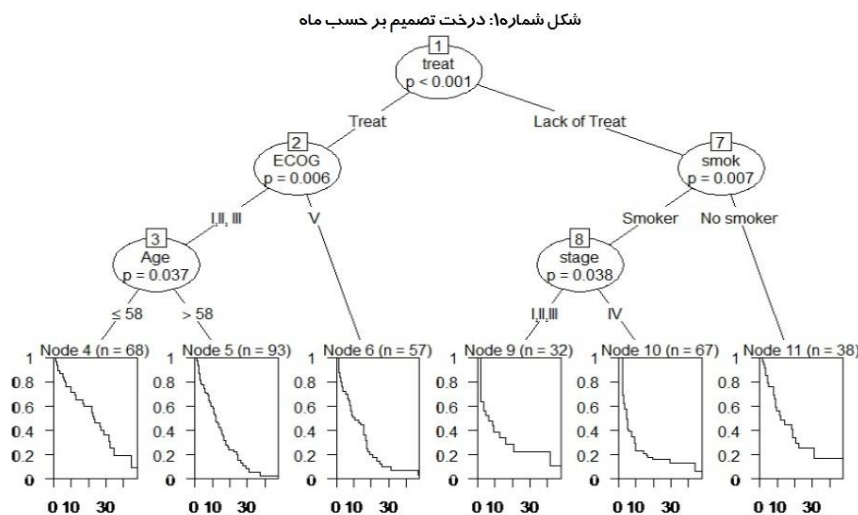
اساله، ۲ساله و ۳ساله به ترتیب ۳۹ درصد، ۱۸ درصد و ۰/۰۷ درصد به دست آمد. در برازش مدل در اولین مرحله مهم‌ترین متغیری که برای افراز داده‌ها انتخاب می‌شود، متغیر وضعیت درمان بیماران با  $P < 0.001$  می‌باشد. با تکرار الگوریتم برای گره‌های حاصل از افراز مرحله‌ی اول، متغیر ECOG با  $P = 0.006$  به‌عنوان مهم‌ترین متغیر بعدی برای بیماران که در گره چپ (دریافت درمان) و متغیر وضعیت سیگاری بودن با  $P = 0.007$  برای بیماران که در گره راست (عدم دریافت درمان) حاصل از افراز در مرحله‌ی اول هستند، انتخاب می‌شوند. متغیرهای سن در زمان تشخیص و مرحله‌ی تومور به‌عنوان متغیرهای مهم دیگر شناسایی شدند. شکل ۱ مدل درخت تصمیم و زیرگروه‌های افراز شده‌ی بیماران به همراه مهم‌ترین متغیرهای انتخاب شده را نشان می‌دهد. در جدول شماره‌ی ۱ میانگین و میانه بقا در زیر گروه‌های حاصل از مدل درخت تصمیم آورده شده است. براین اساس بیماران که درمان دریافت کرده‌بودند و دارای مرحله‌ی I, II, III از متغیر ECOG و بالای ۵۸ سال بودند، دارای بالاترین بقا (با میانه بقای ۲۲،۲۶ و میانگین بقای ۱۹،۲۷) محسوب می‌شدند و بیماران که در رده‌ی عدم دریافت درمان، سیگاری و دارای مرحله‌ی تومور IV بودند کمترین بقا را (با میانه بقای ۲ و میانگین بقای ۳،۸۲) تجربه کرده‌بودند. نتیجه‌ی آزمون لگاریتم رتبه‌ای برای مقایسه نرخ بقا در شش زیر گروه، بیانگر اختلاف معنادار بین این زیر گروه‌ها هست ( $P < 0.001$ ) که نشان دهنده‌ی کارایی مدل

تکنیک هزینه - پیچیدگی درخت است و روش مبتنی بر اختلاف بین گره‌ها که بر اساس تکنیک افراز - پیچیدگی است (۱۹، ۲۰). در این مطالعه برای هرس کردن درخت تصمیم از روش هزینه - پیچیدگی که پیش‌فرض بسته نرم‌افزاری مورد استفاده می‌باشد، استفاده گردیده است. برای تحلیل اطلاعات، از آزمون‌های تی مستقل و لگاریتم رتبه‌ای و برای برازش مدل درخت بقا از نرم‌افزار R و بسته‌ی نرم‌افزاری partykit استفاده شد.

## یافته‌ها

تعداد بیماران مورد بررسی، ۳۵۵ نفر بود که ۲۴۰ نفر فوت کرده و ۱۱۵ نفر سانسور (censoring) شده بودند. ۷۲/۱ درصد افراد مبتلا (۲۵۶ نفر)، مرد و ۲۷/۹ درصد افراد مبتلا (۹۹ نفر)، زن بودند (نسبت مردها به زن‌ها برابر ۲/۶ می‌باشد). میانگین سن تشخیص افراد مبتلا  $63/5 \pm 13/5$  سال تعیین شد (با دامنه‌ی ۹۱-۱۱ سال). میانگین سن تشخیص سرطان برای مردان  $65 \pm 12/4$  سال و برای زنان  $59/4 \pm 16/5$  سال به دست آمد که میانگین سن تشخیص سرطان برای مردان و زنان از نظر آماری، متفاوت بود ( $P = 0/002$ ). ۴۹ درصد از مردان و ۶۲ درصد از زنان مبتلا به سرطان ریه در سنین زیر ۶۵ سال و ۲۰/۴ درصد مردان و ۱۴/۴ درصد زنان در رده‌ی سنی بالای ۷۵ سال قرار داشتند. میانگین و میانه‌ی زمان بقا برای بیماران به ترتیب، ۱۳ ماه و ۸/۴ ماه حاصل شد.

با استفاده از روش کاپلان مایر میزان بقای



جدول شماره ۱: برآورد میانگین بقا، میانه بقا و خطای استاندارد در زیرگروه‌های حاصل از مدل (برحسب ماه)

زیرگروه	برآورد میانگین بقا	خطای استاندارد	برآورد میانه بقا	خطای استاندارد
۱	۱۹/۲۷	۱/۹۴	۲۲/۲۶	۴/۵۱
۲	۱۳/۸۱	۱/۵۲	۱۴/۶	-/۹۹
۳	۱۰/۵۳	۱/۴۸	۸/۱	۱/۱۲
۴	۶/۱۲	۴/۷۶	۴/۱	-/۶۹
۵	۳/۸۲	-/۷۴	۲	-/۳۶
۶	۱۰/۲۳	۱/۹۱	۱۰/۳۳	۱/۵۳

می‌باشد.

ما نیز این مقوله به‌عنوان متغیر مهم شناسایی شد، ولی مطالعه‌ی شو و همکاران (۲۵) با این موضوع همخوانی ندارند و وضعیت سیگار، به‌عنوان متغیر تأثیرگذاری معرفی نشده است.

بر پایه‌ی مطالعه ما میانگین سن تشخیص بیماری همخوانی داشت (۶، ۲۶). در این مطالعه سن تشخیص بیماری مانند مطالعات (۶، ۲۷) متغیر مهمی تشخیص داده شد. انتخاب نقطه‌ی افراز ۵۸ سال برای سن تشخیص بیماری، بر اساس یافتن بیشترین اختلاف بین میزان بقای دو زیرگروه می‌باشد.

مرحله‌ی تومور دیگر متغیر مهم در این مطالعه با استفاده از مدل درخت تصمیم بود. در مطالعه‌ی کاواگوچی و همکاران نیز این متغیر عامل مؤثری بر میزان بقای بیماران تشخیص داده شده است (۲۴). با این‌که جنسیت یکی از فاکتورهای مؤثر در بقاست، اما مطالعه‌ی ما این متغیر را مهم نشان نداد. هرچند با مطالعه‌ی زهیر و همکاران همخوانی داشت (۶)، اما برخی مطالعات نیز نشان داد که بقا برای زنان بیشتر از مردان است (۲۷، ۲۸).

در این مطالعه نوع تومور متغیر با اهمیتی تشخیص داده نشد که این مورد، اگرچه با مطالعات داخلی همخوانی داشت (۶، ۲۹)، اما با مطالعات سایر کشورها متفاوت بود. مطالعات در سایر کشورها نشان می‌دهد نوع تومور متغیر تأثیرگذاری بر بقای بیماران می‌باشد (۸، ۳۰).

باتوجه به نتایج این مطالعه، استفاده از مدل درخت رده‌بندی، می‌تواند منجر به شناسایی ریسک فاکتورهای مهم و مؤثر بر بقای بیماران شود و شناسایی

## بحث

مدل‌های رده‌بندی درختی، هیچ نوع پیش‌فرض‌ها و محدودیت‌هایی را که برای مدل‌های پارامتری و نیمه پارامتری برقرار است، شامل نمی‌شوند. مدل‌های رده‌بندی درختی علاوه بر شناسایی و معرفی متغیرهای مهم، زیرگروه‌های همگنی از بیماران را برای تحقیقات بالینی بعدی ارائه می‌دهد. ارائه‌ی این زیرگروه‌ها برای انتخاب پروتکل‌های درمانی، مؤثر و در طراحی کارآزمایی بالینی برای مطالعات بعدی بسیار مفید است.

یافته‌های حاصل از مدل درخت تصمیم، متغیر وضعیت درمان را به‌عنوان مهم‌ترین فاکتور پیش‌آگهی معرفی می‌کند. مطالعات مختلفی در این زمینه وجود دارد که مهم و مؤثر بودن متغیر وضعیت درمان بر بقای بیماران را تأیید می‌کند که از میان آن‌ها می‌توان به مطالعه‌ی زهیر و همکاران در سال ۲۰۱۲ و مطالعه‌ی برگمنس و همکاران در سال ۲۰۰۵ اشاره کرد (۶، ۲۲).

در ارزیابی فاکتورهای پاتولوژیک، عملکردی براساس سیستم Eastern Cooperative Oncology Group (ECOG)، به‌عنوان دیگر متغیر مهم شناسایی شد. به‌گونه‌ای که کمترین میزان بقا برای بیمارانی که در رده‌ی دریافت درمان قرار داشتند، با میزان ۸٫۱ ماه مربوط به مرحله‌ی V از متغیر ECOG بود.

در مطالعات میتسودومی و همکاران (۲۳)، کاواگوچی و همکاران (۲۴) وضعیت سیگاری بودن یا نبودن، متغیر تأثیرگذاری بر زمان بقا می‌باشد. در مطالعه‌ی

جمع‌آوری اطلاعات مورد نیاز این تحقیق، همکاری کرده‌اند، تشکر و قدردانی نمایند. این مقاله برگرفته از پایان‌نامه‌ی دوره‌ی کارشناسی ارشد گروه آمار زیستی و اپیدمیولوژی دانشگاه علوم پزشکی همدان می‌باشد.

این عوامل در بیماران سرطانی می‌تواند باعث جلوگیری از پیشرفت بیماری شده و زمان بقای بیماران را افزایش دهد.

## تشکر و تقدیر

نویسندگان مقاله بر خود لازم می‌دانند از کلیه‌ی کارکنان بیمارستان‌های استان آذربایجان غربی که در

## References

- Haghshenas M, Golini-Moghaddam T, Rafiei A, Emadeian O, Shykhpour A, Ashrafi GH. Prevalence and type distribution of high-risk human papillomavirus in patients with cervical cancer: a population-based study. *Infect Agent Cancer*. 2013;8(1):20.
- Bollschweiler E, Wolfgarten E, Nowroth T, Rosendahl U, Mönig SP, Hölscher AH. Vitamin intake and risk of subtypes of esophageal cancer in Germany. *J Cancer Res Clin Oncol*. 2002;128(10):575-80.
- Ghanbari N, Golestan Jahromi F, Bani Hashem K, Moazen M. The Effect of Stress Inoculation Training (SIT) on Depression and Anxiety in Patients with prostate Cancer. *Journal of Sabzevar University of Medical Sciences*. *J Sabzevar Univ Med Sci*. 2014;21(5):735-42. [Persian]
- Hamta A, Ghazzafi S. The study of *Thymus vulgaris* cytotoxicity effects on breast cancer cell's line. 3. 2015; 21 (1) :122-30. [Persian]
- Mousavi SM, Gouya MM, Ramazani R, Davanlou M, Hajsadeghi N, Seddighi Z. Cancer incidence and mortality in Iran. *Ann Oncol*. 2009;20(3):556-63.
- Zahir ST, Mirtalebi M. Survival of patients with lung cancer, Yazd, Iran. *AsianPac J Cancer Prev*. 2012;13(9):4387-91.
- Özlu T, Bulbul Y. Smoking and lung cancer. *Tüberküloz ve Toraks Dergisi*. 2005;53(2):200-9.
- Haraguchi S, Koizumi K, Mikami I, Junichi O, Iijima Y, Ibi T, et al. Clinicopathological characteristics and prognosis of non-small cell lung cancer patients associated with a family history of lung cancer. *Int J Med Sci*. 2012;9(1):68.
- Macdonald F, Ford C, Casson A. *Molecular biology of cancer*: Taylor & Francis; 2004.
- Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J. *Harrison's principles of internal medicine*: McGraw-Hill Medical New York; 2008.
- Silvestri GA, Alberg AJ, Ravenel J. The changing epidemiology of lung cancer with a focus on screening. *BMJ*. 2009;451-4.
- Kleinbaum DG, Klein M. *Survival analysis: a self-learning text*: Springer; 2005.
- Schabath MB, Thompson ZJ, Gray JE. Temporal trends in demographics and overall survival of non-small cell lung cancer patients at the Moffitt Cancer Center from 1986 to 2008. *Cancer Control*. 2014; 21(1): 51-6.
- Santoro IL, Ramos RP, Franceschini J, Jamnik S, Fernandes ALG. Non-small cell lung cancer in never smokers: a clinical entity to be identified. *Clinics*. 2011;66(11):1873-7.
- Grose D, Devereux G, Brown L, Jones R, Sharma D, Selby C, et al. Simple and objective prediction of survival in patients with lung cancer: staging the host systemic inflammatory response. *Lung Cancer International*. 2014;2014.
- Goksel T, Eser S, Guclu SZ, Karadag M, Cilli A, Ozlu T, et al. Prognostic factors affecting survival in cases with lung cancer [A lung cancer mapping project in Turkey (LCMPT)]. *European Respiratory Journal*. 2013;42(Suppl 57):P2920.
- Schumacher M, Holla'nder N, Schwarzer G. Prognostic factors studies, in Crowley J (ed): *Handbook of Statistics in Clinical Oncology*. New-York, NY, Marcel Dekker. 2001: 321-78.
- Saki Malehi A, Hajizadeh E, Fatemi R. Evaluation of Prognostic Variables for Classifying the Survival In Colorectal Patients using The Decision Tree. *Iran J Epidemiol*. 2012; 8 (2) :13-9. [Persian]
- Zhang H, Singer B. *Recursive partitioning in the health sciences*: Springer; 1999.
- Negassa A, Ciampi A, Abrahamowicz M, Shapiro S, Boivin J-F. Tree-structured subgroup analysis for censored survival data: validation of computationally inexpensive model selection criteria. *Statistics and computing*. 2005;15(3):231-9.
- Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and regression trees*: CRC press; 1984.

22. Berghmans T, Paesmans M, Meert AP, Mascaux C, Lothaire P, Lafitte JJ, et al. Survival improvement in resectable non-small cell lung cancer with (neo)adjuvant chemotherapy: results of a meta-analysis of the literature. *Lung Cancer*. 2005;49(1):13-23.
23. Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol*. 2010;11(2):121-8.
24. Kawaguchi T, Takada M, Kubo A, Matsumura A, Fukai S, Tamura A, et al. Gender, histology, and time of diagnosis are important factors for prognosis: analysis of 1499 never-smokers with advanced non-small cell lung cancer in Japan. *J Thorac Oncol*. 2010;5(7):1011-7.
25. Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, et al. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring *ALK* gene rearrangement: a retrospective analysis. *Lancet Oncol*. 2011;12(11):1004-12.
26. Arrieta Ó, Núñez-Valencia C, Reynoso-Erazo L, Alvarado S, Flores-Estrada D, Angulo LP, et al. Health-related quality of life in patients with lung cancer: validation of the Mexican-Spanish version and association with prognosis of the EORTC QLQ-LC13 questionnaire. *Lung Cancer*. 2012;77(1):205-11.
27. Puri V, Garg N, Engelhardt EE, Kreisel D, Crabtree TD, Meyers BF, et al. Tumor location is not an independent prognostic factor in early stage non-small cell lung cancer. *Ann Thorac Surg*. 2010;89(4):1053-9.
28. Battafarano RJ, Piccirillo JF, Meyers BF, Hsu H-S, Guthrie TJ, Cooper JD, et al. Impact of comorbidity on survival after surgical resection in patients with stage I non-small cell lung cancer. *J Thorac Cardiovasc Surg*. 2002;123(2): 280-7.
29. Hajmanoochehri F, Mohammadi N, Zohal MA, Sodagar A, Ebtehaj M. Epidemiological and clinicopathological characteristics of lung cancer in a teaching hospital in Iran. *Asian Pac J Cancer P*. 2014;15(6):2495.
30. Jiang X, de Groh M, Liu S, Liang H, Morrison H. Rising incidence of adenocarcinoma of the lung in Canada. *Lung Cancer*. 2012;78(1):16-22.

# Evaluation of Prognostic Variables for Classifying the Survival In lung cancer Patients using The Decision Tree

*Malek Abazari,*

Department of Epidemiology and Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

*Reza Abazari,*

Young Researchers and Elite Club, Ardabil Branch, Islamic Azad University, Ardabil, Iran.

*Ghodratollah Roshanaei,*

Assistant professor, Modeling of Noncommunicable diseases Research center, Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

*Mahdia Gholamnejad,*

Pulmonology Department, Urmia University of Medical Sciences, Urmia, Iran.

*Yousef Roosta,*

Resident of Internal Medicine, Urmia University of Medical Sciences, Urmia, Iran.

\* *Hossein Mahjub ,*

Research Center for Health Sciences and Department of Biostatistics & Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

Received:02/12/2014, Revised:13/01/2015, Accepted:17/02/2015

## Correspond Author:

Hossein Mahjub ,  
School of Public Health, Hamadan  
University of Medical Sciences,  
Hamadan, Iran.  
E-mail: mahjub@umsha.ac.ir

## Abstract

**Background & Objectives:** Lung cancer due to low chance of survival is the most important and the most common fatal cancer and has the highest mortality rates compared to breast and prostate cancer and other multiple cancers. The purpose of this study is to analyze survival data of patients having lung cancer in the province of West Azerbaijan.

**Materials & Methods:** The information of this study is relevant to 355 patient's referred to hospitals existing in West Azerbaijan province between 1386-93. The situation of the patients was followed up until the end of the Khordad 1393. The data are compromised from Histopathologic and demographic information. For analyzing the data and classifying the patients, the model of decision tree was used. All analysis and synthesis were performed using the software of R.

**Results:** The mean and the median time of survival for 355 patients having lung cancer were estimated 13 and 4.8 months, respectively. The probable of 1-year surviving was about 39% and only 7% of patients had experienced 3-year survival. The model of decision tree of state therapy variables, Eastern Cooperative Oncology Group (ECOG), smoking status, age of diagnosis and tumor stage were identified as important variables.

**Conclusion:** The model of classification tree can identify the important and effective risk factors on the survival of the patients and identifying these factors in the cancer patients can prevent progression of the disease and increase survival time of patients.

**Keywords:** Lung Neoplasm; Decision Tree; Survival; Iran.