

## شناسایی عوامل مؤثر بر بقای بیماران سرطان معده دارای متاستاز با استفاده از مدل جنگل تصادفی بقا و مقایسه آن با مدل رگرسیون کاکس

ملیحه صفری<sup>۱</sup>، محمد عباسی<sup>۲</sup>، فاطمه گوهری انصاف<sup>۳</sup>، زینب برنگی<sup>۴</sup>، قدرت‌اله روشنایی<sup>۴</sup>

<sup>۱</sup>دانشجوی دکترای آمارزیستی، مرکز پژوهش دانشجویان، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۲</sup>استادیار، گروه داخلی، دانشکده پزشکی، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۳</sup>کارشناسی ارشد اپیدمیولوژی، مرکز پژوهش دانشجویان، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران

<sup>۴</sup>دانشیار، گروه آمارزیستی، دانشکده بهداشت، مرکز تحقیقات مدلسازی بیماری‌های غیرواگیر، همدان، ایران

نویسنده رابط: قدرت‌اله روشنایی، نشانی: همدان، خیابان شهید فهمیده، دانشگاه علوم پزشکی همدان، دانشکده بهداشت، گروه آمارزیستی، تلفن: ۰۸۱-۳۸۳۸۰۰۹

پست الکترونیک: roshanaei@umsha.ac.ir.Gh

تاریخ دریافت: ۹۸/۰۱/۱۲؛ پذیرش: ۹۸/۰۶/۰۹

**مقدمه و اهداف:** در تحلیل بقا استفاده از مدل کاکس برای تعیین عوامل مؤثر، نیازمند برقراری پیش‌فرض‌هایی است که عدم برقراری آن‌ها منجر به نتایج اریب می‌شود. هدف این مقاله تعیین عوامل مؤثر بر بقای بیماران مبتلا به سرطان معده دارای متاستاز با استفاده از روش ناپارامتری جنگل تصادفی بقا (RSF) و مقایسه آن با مدل کاکس است.

**روش کار:** در این مطالعه کوهورت گذشته‌نگر ۲۰۱ بیمار مبتلا به سرطان معده دارای متاستاز مراجعه‌کننده به کلینیک امام خمینی استان همدان مورد بررسی قرار گرفت. بقای بیماران از زمان تشخیص تا مرگ یا پایان مطالعه محاسبه شد. ویژگی‌های جمعیت شناختی (شامل سن و جنس) و متغیرهای مربوط به بیماری (شامل مرحله بیماری، گرید تومور، نوع درمان و ...) از پرونده بیماران استخراج شد. عوامل مؤثر با استفاده از مدل کاکس و جنگل تصادفی بقا تعیین و مقایسه شد. تحلیل داده‌ها با نرم‌افزار R3.4.3 و بسته‌های survival و RandomForestSRC انجام شد.

**یافته‌ها:** میانگین (انحراف معیار) سن تشخیص بیماران (۱۲/۹) ۶۱/۵ سال بود. بر اساس مدل کاکس تنها متغیر شیمی‌درمانی (P=0.033) بر بقا مؤثر بود. نتایج برازش مدل RSF نشان داد که متغیرهای مؤثر بر بقا به ترتیب نوع جراحی، محل متاستاز، شیمی‌درمانی، سن، گرید تومور، جراحی، تعداد لنفوم‌های درگیر، جنس و رادیوتراپی بود. همچنین مدل RSF با قاعده تقسیم لگ-رتبه بر اساس شاخص‌های مناسب مدل نسبت به مدل کاکس عملکرد بهتری داشت.

**نتیجه‌گیری:** در صورتی که تعداد متغیرها زیاد و بین متغیرها رابطه وجود داشته باشد روش RSF متغیرهای مهم و تأثیرگذار بر بقا را بدون نیاز به پیش‌فرض‌های محدودکننده با دقت بالا نسبت به مدل کاکس شناسایی می‌کند.

**واژگان کلیدی:** سرطان معده، جنگل تصادفی بقا (RSF)، متاستاز، مدل کاکس، بقا

### مقدمه

اطلاق می‌شود که در مراحل خیلی پیشرفته سرطان‌ها ایجاد می‌شود. گاهی سلول‌های سرطانی از بافت اولیه‌ای که در آن ایجاد شده‌اند جدا شده و از راه جریان خون یا سیستم لنفاوی به دیگر اعضای بدن می‌رسند و در محل جدیدی لانه‌گزینی و تکثیر شده و کانون سرطانی جدیدی می‌سازند. ادامه روند متاستاز باعث درگیری اعضای بیشتری شده و در نتیجه درمان را بسیار دشوارتر یا بعضاً غیرممکن می‌سازد. عموماً در متاستاز بیشتر ویژگی‌های سلول اولیه حفظ می‌شود و از همین راه هم می‌توان منشأ متاستاز را تشخیص داد. متاستاز سرطان طی چهار مرحله جدا شدن سلول‌های سرطانی از تومور اولیه، تخریب بستر خارج

سرطان معده و روده (gastrointestinal) یکی از رایج‌ترین سرطان‌هاست که مسبب مرگ سالانه بیش از ۱,۸ میلیون نفر در جهان است (۱). سرطان معده چهارمین سرطان رایج و دومین علت مرگ‌ومیر در بین سرطان‌ها است (۲). در ایران برخلاف سایر کشورها میزان بروز سرطان معده در حال افزایش است و این افزایش در غرب کشور فراوان است (۳). اگرچه بالاترین میزان بروز سرطان معده در استان‌های شمال و شمال غرب کشور است ولیکن سرطان معده در مناطق غرب و شمال غرب کشور دارای روند صعودی است (۴).

متاستاز به انتقال سلول‌های سرطانی به بافت‌های دیگر بدن

بیشتر بهترین مدل در شناسایی متغیرها مهم بر بقا است. نتیجه مطالعه آن‌ها نشان داد که متاستاز، سن، اندازه تومور و نوع هیستولوژی تومور بر بقا اثر دارد (۲۱).

مطالعه‌ای توسط Datema و همکاران در مقایسه کاکس و RSF در بیماران مبتلا به سرطان سر و گردن انجام شد نتیجه مطالعه آن‌ها تفاوت بارزی در خطای این دو روش نشان نداد (۲۲). Manilich و همکاران از روش RSF برای تعیین عوامل مؤثر بر مرحله‌بندی بیماران سرطان کولورکتال استفاده کردند (۲۳).

Zhu و همکاران با استفاده از مدل کاکس و RSF به بررسی رابطه بین متغیرهای ژنتیکی و جمعیت شناختی بر بقای کلی (OS) و بقای عاری از بیماری (DFS) در بیماران سرطان کولورکتال پرداختند. نتیجه مطالعه آن‌ها نشان داد در صورتی که متغیرها همبستگی بالایی داشته باشند طبق شاخص‌های ارزیابی مدل، عملکرد RSF بالاتر است (۲۴). Chen و همکاران از RSF برای تعیین و اعتبارسنجی برخی از ژن‌های مؤثر بر بقای بیماران سرطان ریه پرداختند (۲۵). روشنایی و همکاران از مدل RSF برای تعیین عوامل مؤثر بر بقای پیوند کلیه در بیماران پیوندی از دهنده زنده و مقایسه کارایی آن با مدل کاکس پرداختند (۲۶).

Ingrisch و همکاران با استفاده از روش RSF به تعیین عوامل مؤثر بر بقای بیماران مبتلا به دستگاه گوارش پرداختند و نتایج حاصل از این روش را با نتایج تحلیل کاکس چندمتغیره مقایسه کردند (۲۷).

Geo و همکاران با استفاده از روش RF بیماران مبتلا به سرطان معده را بر اساس نمره خطر (risk score) مبتنی بر بقای بیماران، به سه سطح ریسک بالا، متوسط و پایین تقسیم کرده و عوامل مؤثر بر بقای هر یک از این سه گروه از بیماران را بررسی کردند. نتایج مطالعه آن‌ها نشان داد که مرحله بیماری، سن، متاستاز، اندازه تومور، محل تومور و نوع هیستولوژی تومور دارای اثر معنی‌داری بر هر یک از سطوح خطر بیماری است (۲۸).

Deng و همکاران با ترکیب روش RF و کاکس چندمتغیره، یک مدل نمره خطر برای شناسایی مهم‌ترین عوامل ژنتیکی، جمعیت شناختی و بالینی بیماران مبتلا به سرطان معده ارائه دادند. نتایج مدل کاکس نشان داد که سن، جنس، مرحله بیماری و گرید تومور و تنها یک ژن بر بقای بیماران مؤثر است ولی ترکیب مدل کاکس با RF نشان داد که علاوه بر این متغیرها، متغیر اندازه تومور و چندین ژن دیگر بر بقای بیماران مؤثر است (۲۹).

با توجه به این‌که تعداد مطالعه محدودی از روش RSF برای تعیین مهم‌ترین عوامل مؤثر بر بقای بیماران سرطان معده استفاده

سلولی، حرکت سلول‌های سرطانی، لانه‌گزینی و تکثیر سلول‌های سرطانی در محل جدید انجام می‌شود (۵، ۶).

با توجه به این‌که این بیماری عمدتاً در مراحل پیشرفته بیماری تشخیص داده می‌شود لذا در طول دو دهه گذشته، نسبت بیماران مبتلا به سرطان معده دارای متاستاز به بیش از ۴۰ درصد افزایش یافته است (۷، ۸).

تعیین عوامل مؤثر بر پیشرفت بیماری می‌تواند در پیشگیری و درمان بیماران مؤثر باشد. در عمل برای تعیین عوامل مؤثر بر بقا از مدل‌های نیمه‌پارامتری رگرسیون مخاطرات متناسب کاکس (Cox Proportional Hazards (CPH)) استفاده می‌شود. ولیکن در برخی از مطالعات داده‌ها ممکن است دارای ساختار پیچیده همچون همبستگی بالا بین کواریت‌ها یا روابط غیرخطی پیچیده با بعد بالا باشند که این ساختار پیچیده باعث شود تا فرض مخاطرات متناسب که از پیش‌فرض‌های مدل کاکس است برقرار نباشد (۹). در چنین حالتی مدل کاکس برازش مناسبی ندارد و ممکن است عوامل مؤثر بر بقا به صورت مناسب انتخاب نشوند لذا بایستی از مدل‌های جایگزین که بتوانند در صورت وجود این شرایط و عدم برقراری پیش‌فرض‌های لازم در داده‌ها عملکرد مناسب داشته باشند استفاده کرد. مدل جنگل تصادفی بقا (Random Survival Forest (RSF)) توسط Ishwaran و همکاران (۲۰۱۰) برای انتخاب متغیرهای مهم و تأثیرگذار بر بقا معرفی شد که به‌عنوان جایگزین مدل کاکس می‌تواند در برابر این مشکل توانمند باشد (۱۰).

RSF یک روش آموزش ماشین ناپارامتری است که توسط Ishwaran بر اساس جنگل تصادفی توسعه‌یافته توسط بریمن (۲۰۰۱) ارائه شد. دو مزیت عمده این روش نسبت به CPH طبیعت ناپارامتری آن و توانایی سنجش تمام اثرات یک و چندمتغیری به‌طور خودکار است و همچنین این روش قادر است تا کواریت‌های مؤثر بر بقا در یک مجموعه کواریت‌های با همبستگی بالا را شناسایی و به‌خصوص در داده‌های با بعد بالا که این مشخصه از ویژگی‌های این نوع از داده‌هاست به‌کار برد (۱۰-۱۲). اگرچه روش کاکس برای تعیین عوامل مؤثر بر بقای بیماران مبتلا به سرطان معده توسط مطالعات متعددی انجام شده است (۲۰-۱۳).

ادهم و همکاران با استفاده از مدل RSF به شناسایی عوامل مؤثر بر بقای ۱۲۷ بیمار سرطان معده پرداختند. نتیجه مطالعه آن‌ها نشان دارد مدل RSF با روش تقسیم لگ-رنگ در بین روش‌های تقسیم متفاوت دارای خطای پیش‌بینی کمتر و شاخص هماهنگی

کرده‌اند لذا هدف این مطالعه تعیین عوامل مؤثر بر بقای بیماران مبتلا به سرطان معده دارای متاستاز با استفاده از روش RSF و مقایسه نتایج آن با مدل کاکس است.

## روش کار

در این مطالعه همگروهی گذشته‌نگر اطلاعات ۲۰۱ بیمار مبتلا به سرطان معده دارای متاستاز که از تاریخ ۱۳۸۲-۱۳۹۶ به کلینیک امام خمینی استان همدان مراجعه داشتند مورد بررسی قرار گرفت. بقای بیماران از زمان تشخیص تا زمان مرگ یا پایان مطالعه محاسبه شد. عوامل مؤثر بر بقا با استفاده از مدل کاکس و جنگل تصادفی بقا تعیین و مقایسه شد. اهمیت ۱۲ متغیر جنس، سن تشخیص، گرید تومور، نوع تومور، مرحله بیماری، محل متاستاز، محل تومور، شیمی‌درمانی، جراحی، نوع جراحی و رادیوتراپی مورد بررسی قرار گرفت. ابتدا اهمیت متغیرها با به‌کارگیری مدل جنگل تصادفی بقا با قاعده تقسیم تصادفی، لگ-رتبه اسکور و لگ-رتبه انجام شد و روش تقسیم با کمترین خطا به‌عنوان روش تقسیم مناسب انتخاب شد. به‌طور کلی الگوریتم جنگل تصادفی بقا (RSF) به‌صورت زیر است:

۱- به تعداد B نمونه بوت استرپ از داده‌های اصلی انتخاب می‌شود. در هر نمونه بوت استرپ حدود یک‌سوم از داده خارج می‌ماند که نمونه خارج از کیسه (OBB) گفته می‌شود. در این مطالعه به تعداد ۱۰۰۰ نمونه بوت استرپ از داده‌های اصلی انتخاب شد که در هر نمونه بوت استرپ، ۶۷۰ نمونه برای آموزش و بقیه برای آزمایش و تعیین خطای پیش‌بینی استفاده شد.

۲- برای هر نمونه بوت استرپ یک درخت بقا رشد داده شد. در هر گره از درخت به‌طور تصادفی به تعداد ۴ پیشگو (کواریت) برای تقسیم انتخاب شد. متغیری برای تقسیم انتخاب می‌شود که بیشترین تفاوت در بقای دو گره دختر تشکیل شده ایجاد کند.

۳- درخت تا اندازه کامل رشد می‌کند. آخرین گره، گره نهایی نام دارد. گره نهایی نباید کمتر از  $d_o > 0$  باشد ( $d_o$  نشان‌دهنده رخداد مورد نظر است). شرط توقف درخت در این مطالعه تعداد ۴ مرگ در هر گره نهایی تعیین شد.

۴- برای هر درخت تابع تجمعی خطر (CHF) محاسبه و سپس میانگین این CHFها، کلی را گزارش

می‌کند.

۵- با استفاده از داده‌های OOB خطای پیش‌بینی محاسبه شد.

همچنین از مدل رگرسیون کاکس برای تعیین عوامل مؤثر بر مخاطره استفاده شد و نتایج این دو مدل مقایسه شدند. به‌منظور ارزیابی مدل‌ها از شاخص خطای پیش‌بینی (که برابر یک منهای شاخص هماهنگی است) استفاده شد. این شاخص عددی بین صفر تا ۱ است که مقادیر کوچک نشان از پیش‌بینی‌کنندگی خوب مدل و هر چه مقدار این شاخص کمتر از ۰/۵ باشد نشان می‌دهد عملکرد مدل مطلوب‌تر است یعنی در شناسایی عوامل مؤثر بر بقا با دقت بالاتری عمل می‌کند. همچنین از شاخص نمره بریر (Brier Score) برای ارزیابی مدل‌ها استفاده شد. هرچه نمره بری‌یر کوچک‌تر باشد دلیلی بر مناسب بودن مدل پیش‌بینی است. روش دیگر برای ارزیابی کارایی مدل‌های نمره بری‌یر (Brier Score) است که فرم کلی آن به‌صورت زیر است:

$$BS(t, \hat{S}) = E(Y_i(t) - \hat{S}(t|X_i))^2$$

که  $Y_i(t)$  وضعیت بقای مشاهده‌شده فرد در زمان t است و  $\hat{S}(t|X_i)$  احتمالات بقا پیش‌بینی‌شده در زمان t برای فرد i با متغیرهای پیش‌بین Xi است. مقدار نمره بریر بین ۰ و ۱ است و مقادیر کوچک‌تر نمره بری‌یر، نشان‌دهنده پیش‌بینی دقیق‌تر است. برای برازش مدل RSF و مدل COX از پکیج‌های `survival`، `RandomForestSRC` و `pec` در نرم‌افزار R نسخه ۳،۴،۳ استفاده شده است. سطح معنی‌داری ۵ درصد در نظر گرفته شده است.

## یافته‌ها

در این مطالعه ۲۰۱ بیمار مبتلا به سرطان معده دارای متاستاز بررسی شدند. ۱۵۱ بیمار (۷۵/۱٪) مرد بودند. میانگین (انحراف معیار) سن تشخیص بیماران (۱۲/۶) ۶۱/۵ سال بود. در پایان مطالعه ۹۹ بیمار (۴۹/۲٪) فوت شدند. اطلاعات توصیفی بیماران در جدول ۱ آمده است.

میانگین (انحراف معیار) و میانه بقای این بیماران به ترتیب (۲۹/۲ (۳/۷) ماه و ۱۴ ماه بود و میزان بقای یک‌ساله، سه‌ساله و پنج‌ساله این بیماران به ترتیب ۰/۵۷، ۰/۲۵ و ۰/۱۶ بوده است. به‌منظور بررسی تأثیر متغیرها در حضور سایر متغیرها از مدل مخاطرات متناسب کاکس استفاده شد لازم به ذکر است فرض متناسب بودن مخاطرات به‌صورت کلی برقرار بود لذا این مدل به

کاکس بر اساس شاخص‌های موردنظر بود. در ادامه متغیرهای مهم بر اساس RSF باقاعده تقسیم لگ-رتبه که در بین قواعد تقسیم عملکرد بهتری داشت برای شناسایی متغیرهای مهم استفاده شد که نتایج در نمودار ۱ آمده است. نمودار ۱ نشان می‌دهد که با استفاده از روش RSF باقاعده تقسیم لگ-رتبه، متغیرهای مهم و تأثیرگذار بر بقای بیماران مبتلا به سرطان معده دارای متاستاز به ترتیب نوع جراحی، محل متاستاز، شیمی‌درمانی، سن، گرید تومور، جراحی، تعداد لنفوم‌های درگیر، جنس و رادیوتراپی شناسایی شد ولیکن مدل کاکس موفق به شناسایی این متغیرها نبود.

داده‌ها برازش شد که نتایج در جدول ۲ آمده است. بر اساس مدل کاکس تنها متغیر شیمی‌درمانی تأثیر معنی‌داری بر بقای بیماران داشت. در ادامه برای تعیین اهمیت متغیرها، از مدل RSF با استفاده از هر سه روش تقسیم استفاده شد که جدول ۳ میزان خطای پیش‌بینی، نمره بریر و شاخص هماهنگی را برای هر یک از روش نشان می‌دهد. طبق جدول ۳ در همه شاخص‌ها روش RSF باقاعده تقسیم لگ-رتبه با توجه به اینکه دارای بالاترین شاخص هماهنگی و کمترین مقدار نمره بریر است، لذا در بین روش‌های موجود دارای بالاترین دقت و کمترین خطا در پیش‌بینی بقا و شناسایی مهم‌ترین عوامل مؤثر بر آن است. همچنین این روش دارای عملکرد بهتری از مدل

جدول شماره ۱- ویژگی‌های جمعیت شناختی بیماران سرطان معده دارای متاستاز در استان همدان

| متغیر        | سطوح              | تعداد(درصد) |
|--------------|-------------------|-------------|
| جنس          | مرد               | ۱۵۱(۷۵/۱)   |
|              | زن                | ۵۰(۲۴/۹)    |
| سن تشخیص     | زیر ۵۰ سال        | ۳۴(۱۶/۹)    |
|              | ۷۵-۵۱             | ۱۴۴(۷۱/۶)   |
|              | بالای ۷۵ سال      | ۲۳(۱۱/۴)    |
| گرید تومور   | ۱                 | ۱۵(۷/۵)     |
|              | ۲                 | ۳۶(۲۱/۹)    |
|              | ۳                 | ۴۶(۲۲/۹)    |
|              | ۴                 | ۴(۲)        |
|              | نامشخص            | ۱۰۰(۴۴/۳)   |
| نوع تومور    | آدنوکارسینوما     | ۱۸۰(۸۹/۶)   |
|              | سایر              | ۲۱(۱۰/۴)    |
| جراحی        | ندارد             | ۱۲۶(۶۲/۷)   |
|              | دارد              | ۷۵(۳۷/۳)    |
| شیمی‌درمانی  | بله               | ۱۷۵(۸۷/۱)   |
|              | خیر               | ۲۶(۱۲/۹)    |
| رادیوتراپی   | بله               | ۲۴(۱۱/۹)    |
|              | خیر               | ۱۷۷(۸۸/۱)   |
|              | Total             | ۲۷(۳۶)      |
|              | subtotal          | ۱۴(۱۸/۷)    |
| نوع جراحی    | Partial or Distal | ۲۴(۳۲)      |
|              | unknown           | ۱۰(۱۳/۳)    |
| مرحله بیماری | ۳                 | ۳(۱/۵)      |
|              | ۴                 | ۱۹۸(۹۸/۵)   |
|              | کبد               | ۷۹(۳۹/۳)    |

شناسایی عوامل مؤثر بر بقای بیماران سرطان معده دارای متاستاز با استفاده از مدل جنگل تصادفی بقا و مقایسه آن با مدل رگرسیون کاکس / ۳۴۷

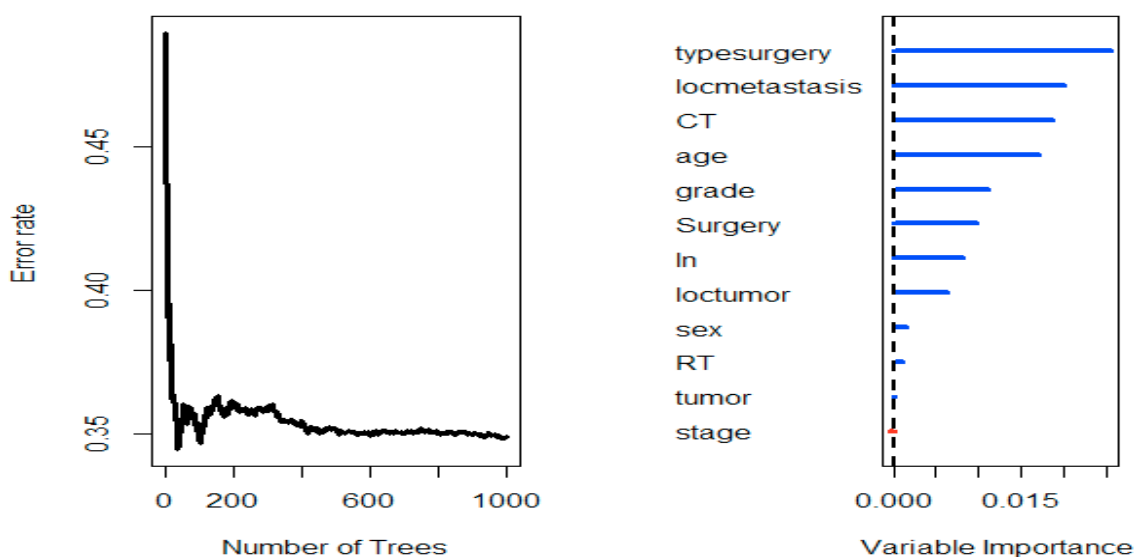
|           |         |           |
|-----------|---------|-----------|
| ۴۳(۲۱/۴)  | سایر    |           |
| ۷۹ (۳۹/۳) | نامشخص  |           |
| ۲۳(۱۱/۴)  | cardia  |           |
| ۳۰(۱۴/۹)  | Body    |           |
| ۲۱(۱۰/۴)  | Antrum  | محل تومور |
| ۳۴(۱۶/۹)  | Other   |           |
| ۹۳(۴۶/۳)  | unknown |           |

جدول شماره ۲- نتایج برازش مدل چند متغیره مخاطرات متناسب کاکس بر عوامل موردبررسی

| متغیر             | ضریب رگرسیون | مقدار احتمال نسبت مخاطره |          | فاصله اطمینان نسبت مخاطره (۰/۹۵) |          |
|-------------------|--------------|--------------------------|----------|----------------------------------|----------|
|                   |              | حد بالا                  | حد پایین | حد بالا                          | حد پایین |
| سن                | -۰/۰۰۱       | ۰/۸۶۸                    | ۰/۹۹۹    | ۰/۹۸۳                            | ۱/۰۱     |
| جنسیت             | ۰/۲۴۱        | ۰/۲۸۱                    | ۱/۲۷۳    | ۰/۸۲۱                            | ۱/۹۷     |
| نوع تومور         | -۰/۰۴۳       | ۰/۹۹۱                    | ۰/۹۵۷    | ۰/۲                              | ۳/۷      |
| محل تومور         | -۰/۰۲۵       | ۰/۴۶۰                    | ۰/۹۷۶    | ۰/۹۱۴                            | ۱/۰۴     |
| گرید              | ۰/۲۳۷        | ۰/۱۵۷                    | ۰/۷۸۹    | ۰/۵۶۸                            | ۱/۱      |
| تعداد لنفوم درگیر | ۰/۰۱۰        | ۰/۷۷۴                    | ۱/۰۱۰    | ۰/۹۴۳                            | ۱/۰۸۱    |
| محل متاستاز       | ۰/۰۰۴        | ۰/۹۲                     | ۱/۰۰۴    | ۰/۹۱۷                            | ۱/۱      |
| مرحله بیماری      | ۰/۶۶۹        | -۰/۵۶۸                   | ۱/۹۵۳    | ۰/۱۹۶                            | ۴/۴۵     |
| شیمی درمانی       | -۰/۵۵۷       | ۰/۰۳۳                    | ۰/۵۷۳    | ۰/۳۴۴                            | ۰/۹۵۵    |
| رادیوتراپی        | -۰/۶۵۱       | ۰/۰۶۸                    | ۰/۵۲۲    | ۰/۲۵۹                            | ۱/۰۵     |
| جراحی             | -۰/۳۸۳       | ۰/۰۷۲                    | ۰/۶۸۲    | ۰/۴۴۹                            | ۱/۰۴     |
| نوع جراحی         | -۰/۱۱۰       | ۰/۵۶۲                    | ۰/۸۹     | ۰/۶۱۷                            | ۱/۳۰     |

جدول شماره ۳ - مقایسه مدل‌های کاکس و RSF با قواعد تقسیم مختلف

| روش                             | شاخص هماهنگی (C-index) | نرخ خطای پیش‌بینی (predictor error rate) | نمره بریر [(۵۹);time=۰IBS] |
|---------------------------------|------------------------|--|----------------------------|
| کاکس                            | ۶۳/۳                   | ۳۶/۷                                     | ۰/۲۰۳                      |
| RSF باقاعده تقسیم تصادفی        | ۶۴                     | ۳۶                                       | ۰/۱۷۵                      |
| RSF باقاعده تقسیم لگ-رتبه اسکور | ۶۵                     | ۳۵                                       | ۰/۱۷۰                      |
| RSF باقاعده تقسیم لگ-رتبه       | ۶۵/۱                   | ۳۴/۹                                     | ۰/۱۶۶                      |



نمودار شماره ۱ - نرخ خطا و متغیرهای مهم بر اساس روش RSF با روش تقسیم لگ-رتبه

## بحث

مطالعه یزدانی و همکاران سن، جراحی و متاستاز مؤثر بودند (۲۰). در مطالعه Ingrisch و همکاران نوع و اندازه تومور بر بقا تأثیر معنی‌دار داشتند (۲۷) و در مطالعه Deng و همکاران متغیرهای جنس، سن، مرحله تومور و گرید تومور بر بقا مؤثر است (۲۹). با توجه به این‌که در صورت وجود ارتباط بین کواریت‌ها و انحراف از فرض متناسب بودن مخاطرات که در مطالعات بقا دور از انتظار نیست روش کاکس در این شرایط به‌طور مناسب عمل نمی‌کند (۹، ۱۲) لذا عدم معنی‌داری متغیرها در مطالعه حاضر برخلاف مطالعات مشابه ممکن است مربوط به نبود شرایط استفاده از این مدل باشد که Ingrisch نیز در مقاله خود به این موضوع اشاره کرده است (۲۷).

در این مطالعه بر اساس مدل RSF متغیر نوع درمان (جراحی و شیمی‌درمانی و رادیوتراپی) متغیر مؤثر بر بقا بود. در برخی مطالعات انجام‌شده نیز درمان به‌عنوان متغیر مؤثر شناسایی شد (۳۰، ۲۰، ۱۸، ۱۶، ۱۵، ۱۳). اگرچه در برخی مطالعات این متغیر اثر معنی‌داری بر بقا نداشت (۳۲). درحالی‌که در مطالعه حاضر بر اساس مدل RSF، مهم‌ترین متغیرها به ترتیب محل جراحی، محل متاستاز، شیمی‌درمانی، سن، گرید تومور، نوع جراحی، تعداد لنفوم‌های درگیر، جنس و رادیوتراپی شناسایی شدند. در مطالعه ادهم بر اساس مدل RSF متغیرهای متاستاز، سن و اندازه تومور به‌عنوان مهم‌ترین متغیر شناخته شدند و نوع درمان جزو متغیرهای مهم نبود علت این تفاوت ممکن است به این دلیل باشد که در این مطالعات نسبت متفاوتی از افراد از درمان‌های مختلف استفاده کرده‌اند و ممکن است میزان و کیفیت استفاده از

در این مطالعه ۲۰۱ بیمار سرطان معده دارای متاستاز مراجعه‌کننده به کلینیک امام خمینی استان همدان مورد بررسی قرار گرفت. میانگین (انحراف معیار) سن تشخیص بیماران (۱۲/۶) ۶۱/۵ سال و بیش از ۷۵٪ بیماران مرد بودند. میانگین سن در برخی مطالعات از مطالعه حاضر بیشتر (۳۲-۳۰، ۲۷، ۲۰، ۱۴، ۱۳) و برخی مطالعات پایین‌تر از این مطالعه بود (۳۳-۱۸، ۱۵).

میان‌ه بقای این بیماران ۱۴ ماه بود که مشابه با برخی از مطالعات (۳۰، ۱۴) ولیکن از برخی مطالعات انجام‌گرفته در ایران پایین‌تر (۳۱، ۳۰، ۱۸-۱۷، ۱۵، ۱۳) و از برخی مطالعات بالاتر بوده است (۳۳، ۲۷، ۲۱، ۱۶). علت متفاوت بودن میزان بقا در این مطالعات ممکن است به دلیل مشخصات بیماران مورد مطالعه و ویژگی‌های مربوط به بیماری و وضعیت درمان تکمیلی آن‌ها باشد چون مطالعات موردبررسی در مناطق جغرافیایی مختلف و با امکانات متفاوت انجام شده است. مثلاً در مطالعه حاضر تقریباً تمام بیماران در مرحله پیشرفته بیماری تشخیص داده‌شده‌اند لذا این میزان تفاوت در میان بقا قابل توجیه است.

در این مطالعه در روش کاکس تنها متغیر شیمی‌درمانی بر بقا مؤثر بود. در مطالعه قربانی و همکاران، نوع درمان و متاستاز و سن بر بقا اثر معنی‌دار داشت (۱۳)، در مطالعه روشنایی و همکاران سن، نوع درمان بر بقا اثر معنی‌دار داشتند (۱۸)، در مطالعه قربانی و همکاران متاستاز و نوع درمان مؤثر شناسایی شدند (۱۴)، در

مدل با توجه به مقدار شاخص هماهنگی  $0.70/3$  به عنوان مدل مناسب انتخاب شد (۲۱). نتایج این مطالعات با مطالعه حاضر در خصوص نوع مدل انتخابی یکسان است و با توجه به متغیرهای مؤثر که توسط این مدل شناسایی می‌شود تا حدودی می‌تواند به پزشکان در ارزیابی‌های تشخیصی و پیشگیرانه کمک‌کننده باشد. از محدودیت‌های مطالعه حاضر مشابه اکثر مطالعات مشاهده‌ای که از پرونده بیماران اطلاعات استخراج می‌شود می‌توان به عدم ثبت دقیق اطلاعات پاتولوژیک در پرونده بیماران و عدم دسترسی به متغیرهای مهم مرتبط دانست. از دیگر محدودیت‌های این مطالعه می‌توان به نقص در پرونده‌های پزشکی بیماران، تغییر شماره تماس بیماران و عدم دسترسی به برخی از بیماران و در موارد معدودی به همکاری ضعیف بعضی از بستگان در دادن اطلاعات بیماران اشاره نمود.

### نتیجه‌گیری

در صورتی که تعداد متغیرها زیاد و بین متغیرها رابطه وجود داشته باشد مدل کاکس در شناسایی متغیرها چندان موفق نیست لذا در این حالت روش RSF متغیرهای مهم و تأثیرگذار بر بقا را با دقت مناسب شناسایی می‌کند و این روش دارای مکانیسم ساده و سراسر است برای پیش‌بینی بقا در داده‌های با بعد بالا است.

### تشکر و قدردانی

این مقاله برگرفته از طرح مصوب شورای پژوهشی دانشگاه علوم پزشکی در تاریخ ۹۶/۳/۱۶ به شماره طرح ۹۶۰۳۱۶۱۷۵۰ با شناسه اختصاصی کمیته اخلاق IR.UMSHA.REC.1396.218 است که بدین‌وسیله از زحمات معاونت محترم تحقیقات و فناوری استان همدان قدردانی و تشکر می‌گردد.

درمان‌ها همان‌طور که آمار ارائه شده نشان می‌دهد متفاوت باشد. در این مطالعه بر اساس مدل RSF سن از لحاظ اهمیت دارای رتبه چهارم بود هرچند مدل کاکس این متغیر را بر بقا مؤثر شناسایی نکرد. مطالعات زیادی که مبتنی بر مدل کاکس بودند متغیر سن را بر بقای بیماران مؤثر دانسته‌اند (۲۰، ۱۸-۱۶، ۱۳) اگرچه برخی مطالعات به این نتیجه رسیده‌اند که با افزایش سن، مخاطره افزایش یافته ولیکن این متغیر تأثیر معنی‌داری بر بقا نداشته است (۳۰، ۱۵، ۱۴-۳۲). در مطالعه ادهم و Ingrisich متغیر سن دارای رتبه دوم از لحاظ اهمیت بر بقا بوده است. دلیل تفاوت نتایج مطالعات مختلف می‌تواند به متغیرهای گوناگون و تعداد متغیرهای مورد بررسی در هر یک از مطالعات و ساختار متغیرها در مطالعات مختلف باشد. ولیکن در اکثر مطالعات مبتنی بر مدل کاکس، عدم معنی‌داری این متغیر می‌تواند به دلیل ارتباط این متغیر با سایر متغیرها باشد. همان‌طور که می‌دانیم سن تشخیص ممکن است با مرحله بیماری، اندازه تومور و ... رابطه داشته باشد که در صورت معنی‌داری هر یک از آن‌ها در مدل ممکن است نیازی به حضور متغیر سن نباشد و این متغیر به‌عنوان متغیر معنی‌دار تشخیص داده نشود) با این حال چون این بیماری معمولاً در سنین بالا شناسایی می‌شود لذا تأثیر معنی‌دار این متغیر بر بقا دور از انتظار نیست.

عملکرد مدل RSF در این مطالعه بهتر از مدل کاکس بود اگرچه در برخی مطالعات مربوط به بقای سایر بیماری‌ها این موضوع تأیید نشده است (۲۶) ولیکن در این مطالعه مدل RSF با قاعده تقسیم لگ-رتبه بهترین عملکرد را در بین مدل‌های RSF با قواعد تقسیم مختلف داشت به طوری که میزان شاخص هماهنگی این مدل و مدل کاکس در این مطالعه به ترتیب برابر  $0.65/1$  و  $0.63/3$  به دست آمد، در مطالعه Ingrisich شاخص هماهنگی مدل RSF با قاعده تقسیم لگ-رتبه و مدل کاکس به ترتیب برابر  $0.65/7$  و  $0.65/2$  به دست آمد (۲۷). در مطالعه ادهم و همکاران هم این

### منابع

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. a cancer journal for clinicians. 2014; 64: 9-29.
2. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin D. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127: 2893-917.
3. Rahimi F, Heidari M. Time Trend Analysis of Stomach Cancer Incidence in the West of Iran. journal of health & Development. 2012; 1: 100-11.
4. Enayatrads M, Salehiniya H. Trends in Gastric Cancer Incidence in Iran. J Mazandaran Univ Med Sci. 2014; 24: 8-16.
5. King R. J. B., Robins M. Cancer biology. 3rd ed. Harlow: Prentice Hal; 2006.
6. Ruddon RW. Cancer biology. 4th ed. Oxford University Press: Oxford; 2007.
7. Bernards N, Creemers G, Nieuwenhuijzen G, Bosscha K, Pruijt J, Lemmens V. No improvement in median survival for patients with metastatic gastric cancer despite increased use of chemotherapy. Ann Oncol. 2013; 24: 3056-60.
8. Thomassen I, Van Gestel Y, Van Ramshorst B, Luyer M, Bosscha K, Nienhuijs S, et al. Peritoneal carcinomatosis of gastric origin: a population-based study on incidence, survival and risk factors. Int J Cancer. 2014; 134: 622-8.
9. Kleinbaum D, Klein M. Survival Analysis: A Self-Learning

- Text, Third ed. New York: Springer; 2012.
10. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High Dimensional Variable Selection for Survival Data. *Journal of the American Statistical Association* 2010; 105: 205–17.
  11. Breiman L. Random forests. *Machine Learning* 2001; 45: 5-32.
  12. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *The Annals of Applied Statistics*. 2008; 2: 841–60.
  13. Ghorbani- Gholiabad S, Yazdani-Cherati J, Jan -Babai G, Shabankhani B. Survival of Patients with Gastric Cancer in Mazandaran Province 2007-2012. *J MazandaranUniv Med Sci*. 2013; 23: 43-50.
  14. Ghorbani-Gholiabad S, Yazdani-Charati J, Jan -Babaie G. Evaluation of Parametric and Semi-parametric Models in Survival Analysis of Patients with Gastric Cancer. *J MazandaranUniv Med Sci*. 2014; 24: 11-8.
  15. Roshanaei G, Ghannad M, Asghari -Jafarabadi M, Faradmaj J, Sadighi S. Prognostic impact of risk factors in patients with gastric cancer in Iran. *Asian Pac J Cancer Prev*. 2011; 12: 3005-8.
  16. Roshanaei G, Kazemnejad A, Sadighi S. Survival estimating following recurrence in gastric cancer patients and its relative factors. *Koomesh*. 2011; 12: 223-8.
  17. Roshanaei G, Sadighi S, Safari M, Faradmaj J. Estimated survival time in gastric cancer patients and its associated factors. *Koomesh*. 2012; 14: 47-54.
  18. Roshanaei G, Safari M, Baghestani A, Sadighi S. Assessment of the survival risk factors in patients with gastric cancer in Cancer Institute of Imam Khomeini Hospital between 2003-2007. *ZUMS Journal*. 2012; 20: 40-50.
  19. Yazdani Charati J, Zare S, Ghorbanpour E, Shabankhani B. Demographic and Geographical Pattern of Mortality Rate from Stomach Cancer and Related Factors in Mazandaran Province From 2001 to 2005. *J MazandaranUniv Med Sci*. 2010; 20: 2-7.
  20. Yazdani J, Sadeghi S, Janbabaei G, Haghghi F. Applying Survival Analysis to Estimate Survival Time in Gastric Cancer Patients. *J MazandaranUniv Med Sci*. 2011; 21: 28-36.
  21. Adham D, Abbasgholizadeh N, Abazari M. Prognostic Factors for Survival in Patients with Gastric Cancer using a Random Survival Forest. *Asian Pac J Cancer Prev*. 2017; 18: 129-34.
  22. Datema FR, Moya A, Krause P, Baeck T, Willmes L, Langeveld T, et al. Novel Head and Neck Cancer Survival Analysis Approach: Random Survival Forests Versus Cox Proportional Hazards Regression. *Head and Neck* 2012; 34: 50-8.
  23. Manilich EA, Kiran RP, Radivoyevitch T, Lavery I, Fazio VW, Remzi FH. A Novel Data-Driven Prognostic Model for Staging of ColorectalCancer *J Am Coll Surg*. 2011; 213: 579–88.
  24. Zhu X, Tian X, Yu C, Shen C, Yan T, Hong J, et al. A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol Cancer*. 2016; 15.
  25. Chen G, Kim S, Taylor JM, Wang Z, Lee O, Ramnath N, et al. Development and validation of a qRT-PCR classifier for lung cancer prognosis. *J Thorac Oncol*. 2011; 6: 1481-7.
  26. Roshanaei G, Omid T, Faradmaj J, Safari M, Poorolajal J. Determination of the effective factors on the survival of kidney transplants in transplant patients from the live donors using the RSF method. *koomesh*. 2018; 3: 1-9.
  27. Ingrisich M, Schöppe F, Paprottka K, Fabritius M, Strobl FF, De Toni EN, Ilhan H, Todica A, Michl M, Paprottka PM. Prediction of 90Y Radioembolization Outcome from Pretherapeutic Factors with Random Survival Forests. *J Nucl Med*. 2018; 59: 769-773.
  28. Deng X, Xiao Q, Liu F, Zheng. A gene expression-based risk model reveals prognosis of gastric cancer. *PeerJ*. 2018 ;6:e4204. doi: 10.7717/peerj.4204. eCollection 2018.
  29. Guo J, Chen S, Chen Y, Li S, Xu D. Combination of CRP and NLR: a better predictor of postoperative survival in patients with gastric cancer .*Cancer Manag Res*. 2018; 10: 315–21.
  30. Barfei F, Abbasi M, Khodabakhshi R, Gohari MR. Survival analysis of patients with adenocarcinoma gastric cancer in Fayazkhsh hospital, Tehran. *RJMS*. 2014; 21: 1-9.
  31. Baeradeh NA, Lotfi MH, Fallahzadeh H, Kargar S, Salman-Roghani H. Survival Rate of Patients with Stomach Cancer and its Effective Factors in Yazd Province. *JCHR*. 2015; 3: 278-87.
  32. Veisani Y, Delpisheh A, Sayehmiri K, Rahimi E. Survival rate estimation and its associated factors in patients with stomach cancer in Tohid Hospital in Sanandaj. *SJKU* 2013; 18: 1-8.
  33. Abad M, Gangy R, Sharifian E, Nikdel R, Jafarzadeh M, Jafarzadeh F. Epidemiologic distribution of cancer in a 10-year study: Retrospective review of hospital records and pathology centers of North Khorasan Province from 2003 to 2012. *Journal of North Khorasan University* 2014; 6: 689-96.



# Identification of Factors Affecting Metastatic Gastric Cancer Patients' Survival Using the Random Survival Forest and Comparison with Cox Regression Model

Safari M<sup>1</sup>, Abbasi M<sup>2</sup>, Gohari Ensaf F<sup>3</sup>, Berangi Z<sup>3</sup>, Roshanaei Gh<sup>4</sup>

1- Candidate of PhD, Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

2- Assistant Professor, Department of Internal Medicine, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran

3- MSc of Epidemiology, Students Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

4- Associate Professor, Department of Biostatistics, Modeling of Noncommunicable Disease Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

**Corresponding author:** Roshanaei GH, roshanaei@umsha.ac.ir.Gh

(Received 1 April 2019; Accepted 31 August 2019)

**Background and Objectives:** In survival analysis, using the Cox model to determine the effective factors requires the assumptions whose failure of leads to biased results. The aim of this paper was to determine the factors affecting the survival of metastatic gastric cancer patients using the non-parametric method of Randomized Survival Forest (RSF) model and to compare its result with the Cox model.

**Methods:** In this retrospective cohort study, 201 patients with metastatic gastric cancer were evaluated in Hamadan Province. Patient survival was calculated from diagnosis to death or end of study. Demographic characteristics (such as gender and age) and clinical variables (including stage, tumor size, etc.) were extracted from the patient records. Factors affecting survival were determined using the Cox model and RSF. Data analysis was performed using the R3.4.3 software and RandomForestSRC and survival packages.

**Results:** The mean (SD) age of patients was 61.5 (12.9) years old. The Cox model showed that chemotherapy ( $p=0.033$ ) was effective in survival, and the results of fitting the RSF model showed that the most important variables affecting survival were type of surgery, location of metastasis, chemotherapy, age, tumor grade, surgery, number of involved lymph nodes, sex and radiotherapy. Based on the model appropriateness, the RSF model with log-rank split rule had a better performance compared to the Cox model.

**Conclusion:** If the number of variables is high and there is a relationship between the variables, the RSF method identifies the important and effective variables on survival with high accuracy without requiring restrictive assumptions compared to the Cox model.

**Keywords:** Gastric cancer, Random survival forest (RSF), Metastasis, Cox model, Survival