

## ارزیابی عوامل موثر بر عود بیماری لنفوم هوچکین با استفاده از مدل جنگل تصادفی بقا و مقایسه آن با مدل رگرسیون کاکس

### چکیده

دریافت: ۱۴۰۱/۰۴/۰۱ ویرایش: ۱۴۰۱/۰۴/۰۸ پذیرش: ۱۴۰۱/۰۶/۲۵ آنلاین: ۱۴۰۱/۰۷/۰۱

**زمینه و هدف:** در بسیاری از تحقیقات پزشکی، رگرسیون خطرات متناسب کاکس برای بررسی توزیع بقای بیماران سرطانی، مبتنی بر گروه‌بندی‌های دموگرافیک و کلینیکی به کار برده می‌شود. هدف از این مطالعه تعیین عوامل موثر بر بقای بیماران مبتلا به لنفوم هوچکین با استفاده از روش ناپارامتری جنگل تصادفی بقا (RSF) و مقایسه آن با مدل خطرات متناسب کاکس است.

**روش بررسی:** در این مطالعه کوهورت گذشته‌نگر، پرونده‌های تمامی بیماران مبتلا به لنفوم هوچکین که در طی سال‌های ۱۳۸۱ (فروردین) تا ۱۳۹۱ (اسفند) به مرکز آنکولوژی و هماتولوژی بیمارستان شقایق اهواز مراجعه نمودند، مورد بررسی قرار گرفتند. بقای بیماران از زمان تشخیص اولیه بیماری تا عود بیماری محاسبه شد. برای ارزیابی روند بیماری، ویژگی‌های جمعیت‌شناختی و متغیرهای مربوط به بیماری (شامل مرحله بیماری، شیمی درمانی، محل درگیری لنف و غیره) از پرونده ۳۸۷ بیمار مبتلا به لنفوم‌هوچکین استخراج شد. تحلیل داده‌ها با نرم‌افزار R4.0.3 و با استفاده بسته‌های survival و RandomForestSRC انجام شد.

**یافته‌ها:** نتایج حاصل از مدل کاکس نشان می‌دهد که LDH ( $P=0/001$ ) و طبقه‌بندی کلاسیک لنفوم ( $P<0/001$ ) معنادار می‌باشند. نتایج برازش مدل RSF نشان داد که مهمترین متغیرهای موثر بر عود به ترتیب مرحله بیماری، شیمی درمانی، طبقه‌بندی کلاسیک لنفوم و هموگلوبین بودند. همچنین مدل RSF براساس شاخص‌های مناسب مدل (شاخص هماهنگی = ۸۴/۹) نسبت به مدل کاکس (شاخص هماهنگی = ۵۷/۶) عملکرد بهتری داشت.

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

**کلمات کلیدی:** مدل کاکس، لنفوم هوچکین، عود، جنگل تصادفی، بقا.

زینب عساکره<sup>۱</sup>، الهام مراغی<sup>۱</sup>، بیژن کبخایی<sup>۲</sup>، امل ساکی مالچی<sup>۳\*</sup>

۱- گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

۲- مرکز تحقیقات بیماری‌های تالاسمی و هموگلوبینوپاتی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

۳- مرکز تحقیقات درد، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

\* نویسنده مسئول: اهواز، دانشگاه علوم پزشکی جندی‌شاپور اهواز، مرکز تحقیقات درد.

تلفن: ۰۶۱-۳۳۱۱۵۳۱۱

E-mail: amalsaki@gmail.com

### مقدمه

بیماری هوچکین (Hodgkin's Disease) که به نام لنفوم هوچکین نیز شناخته می‌شود، نوعی از سرطان خون و سیستم ایمنی موسوم به سرطان لنفوم می‌باشد که ممکن است هرکدام از اجزای سیستم لنفاوی بدن از جمله غده‌های لنفاوی، مغز استخوان و یا طحال

را درگیر کند.<sup>۱</sup> گرچه بیماری هوچکین ممکن است در تمامی سنین دیده شود، شایع‌ترین سن شیوع آن در دهه سوم زندگی (۲۴-۲۰ سالگی) و پس از ۵۵ سالگی (۸۴-۸۰ سالگی) است. این بیماری در مردان نسبت به زنان شایع‌تر است.<sup>۱</sup>

در سال‌های اخیر، میزان مرگ در بیماران مبتلا به لنفوم هوچکین کاهش یافته و میزان بقای پنج ساله آن‌ها از ۴۰٪ به ۷۸٪ رسیده

نیستند را وارد مدل نمی‌کند و امکان تعدیل این متغیرها را محدود می‌کند. در نتیجه ممکن است سایر عوامل خطر بالقوه از این تحلیل خارج شوند. نقطه ضعف دیگر این است که حتی تغییرات کوچک مانند تبدیل متغیرها به دلیل اختلاف در افراز می‌توانند منجر به تولید یک درخت بی‌ثبات شوند.<sup>۸</sup> در واقع مدل‌های درختی ابزار پیش‌آگهی آسان و قابل تفسیر را ارائه می‌دهند اما نتایج بی‌ثباتی دارند، بنابراین برای ارتقای توانایی عمومی‌سازی نتایج، مدل جنگل تصادفی توسعه یافت.<sup>۹</sup>

مدل جنگل تصادفی بقا (RSF) توسط Ishwaran و همکاران برای انتخاب متغیرهای موثر بر بقا معرفی شد. همچنین در مواردی که میزان سانسور بالاست رگرسیون کاکس می‌تواند نتایج آریبی را حاصل کند و به میزان قابل توجهی بر قدرت آماری رگرسیون کاکس تاثیر بگذارد. به نظر می‌رسد جنگل‌های تصادفی کم‌تر تحت تاثیر میزان سانسور قرار می‌گیرند.<sup>۱۰، ۱۱، ۱۲، ۱۳، ۱۴</sup>

از مهم‌ترین ویژگی RSF عملکرد مناسب آن در اندازه‌گیری اهمیت متغیرها است.<sup>۱۵، ۱۶</sup> با توجه به اینکه مطالعات محدودی در زمینه تعیین عوامل تاثیرگذار بر عود لنفوم هوچکین در کشور ما وجود دارد، بنابراین هدف این مطالعه تعیین عوامل موثر بر بقای بیماران مبتلا به لنفوم هوچکین با استفاده از روش RSF و مقایسه نتایج آن با مدل کاکس است.

## روش بررسی

در این مطالعه کوهورت گذشته‌نگر، پرونده‌ی تمامی بیمارانی که با تشخیص لنفوم هوچکین و در بازه زمانی فروردین ماه ۱۳۸۱ تا اسفند ماه ۱۳۹۱ به مرکز آنکولوژی و هماتولوژی بیمارستان شفاي اهواز در استان خوزستان مراجعه کرده بودند، مورد بررسی قرار گرفت.

تشخیص بالینی درگیری لنف نوده‌های بیماران توسط دو پاتولوژیست مختلف مرکز مورد تایید قرار گرفت. در طی این مطالعه، پرونده‌ی بیماران بررسی شده و وضعیت نهایی بیمار از لحاظ وقوع مرگ یا عود مشخص و ثبت گردید. اطلاعات اولیه شامل متغیرهای دموگرافیک، شیمی درمانی، مرحله‌ی سرطان، لنف نودها و ارگان درگیر در زمان تشخیص و نوع درمان، مورد ارزیابی و بررسی قرار

است.<sup>۱</sup> نوع دیگری از سرطان لنفوم که به نام لنفوم غیر هوچکین شناسایی شده است، نسبت به بیماری هوچکین، بدخیم‌تر است. در حالی که لنفوم هوچکین یک نئوپلاسم لنفاوی مونوکلونال نادر با میزان بهبودی بالا است. مطالعات بیولوژیکی و بالینی این بیماری را به دو دسته مجزا تقسیم کرده است: لنفوم هوچکین کلاسیک و لنفوم هوچکین با گره لنفوسیت‌های غالب (NLP-HL). این دو نوع بیماری منجر به تفاوت در تصویر بالینی و آسیب‌شناسی بیماری خواهد شد.

لنفوم کلاسیک هوچکین تقریباً ۹۵٪ کل Hodgkin lymphoma را تشکیل می‌دهد لنفوم کلاسیک هوچکین معمولاً در غدد لنفاوی گردن ایجاد می‌شوند. لنفوم هوچکین دارای پیش‌آگهی کلی عالی با تقریباً ۸۰٪ میزان بهبود است.

درمان لنفوم هوچکین تا حد زیادی به ویژگی‌های بافت‌شناسی، مرحله بیماری و وجود یا عدم وجود عوامل پیش‌آگهی بستگی دارد. نتایج موفق درمان، با ترکیب روش‌های درمانی رادیوتراپی و شیمی درمانی به دست آمده است اما هر یک به‌تنهایی و به‌طور مستقل نیز در درمان موثر هستند.<sup>۲، ۳، ۴</sup> یکی از عوارض این بیماری عود مکرر آن است.<sup>۵</sup>

مدل رگرسیون کاکس (Cox regression) یکی از رایج‌ترین مدل‌های نیمه پارامتری است که برای تعیین عوامل موثر بر بقا به‌کار می‌رود. این مدل به‌طور گسترده‌ای برای تحلیل داده‌های بقا در حضور متغیرهای مستقل به‌کار می‌رود و نیازی به فرض معلوم بودن توزیع زمان بقا ندارد.<sup>۶</sup> نتایج مدل کاکس به برقراری فرض PH و تعیین شکل تابعی درست، وابسته است. اما در برخی از مطالعات ممکن است داده‌ها دارای حجم زیاد و شامل ساختار پیچیده همچون وجود همخطی چندگانه بین متغیرهای پیشگویی‌کننده، یا اثرات متقابل با مرتبه بالا باشند که این ساختار پیچیده می‌تواند باعث برقرار نبودن فرض خطرات متناسب کاکس باشد. در این حالت مدل کاکس برازش مناسبی نخواهد داشت و ممکن است انتخاب عوامل موثر بر بقا به درستی صورت نگیرد. بنابراین در دو دهه اخیر روش‌های جایگزین دیگری که نسبت به برقراری فرض خطرات متناسب حساس نیستند، توسعه یافته‌اند.

در بین این روش‌ها، مدل‌های مبتنی بر درخت به دلیل عدم نیاز به پیش‌فرض‌های ضروری و سادگی تفسیر نتایج، در تحلیل بقا توسعه یافتند.<sup>۷</sup> با وجود مزیت‌های فراوان مدل‌های درختی، در ایجاد درخت تصمیم، معایبی نیز وجود دارد. به‌عنوان مثال، برخی از تکنیک‌های درخت تصمیم، متغیرهای با اهمیت که معنادار

ب- قانون امتیاز لگ-رتبه (log rank score): گروه‌ها را با استفاده از آماره لگ رنک استاندارد شده افزایش می‌کند.

ج- قانون افزایش لگ رتبه تصادفی: یک افزایش تصادفی برای هر یک از  $p$  متغیری که به‌عنوان کاندید انتخاب شدند صورت می‌گیرد و متغیری که بیش‌ترین آماره لگ رنک را داشته باشد (در نقطه افزایش تصادفی خودش) برای افزایش گره استفاده می‌شود.<sup>۱۶، ۱۵</sup> که در این مقاله، تنها به قانون افزایش لگ-رتبه می‌پردازیم و به سایر افزایش‌ها فقط اشاره شده است.

الگوریتم اجرای RSF: ۱-  $B$  نمونه بوت استرپ از داده‌های اصلی گرفته می‌شود به طوری که هر نمونه بوت استرپ به‌طور متوسط ۳۷٪ از داده‌های out-of-bag را که اصطلاحاً به آن نمونه خارج از کیسه گفته می‌شود حذف می‌کند (در مطالعه حاضر  $B=1000$ ).

برای هر نمونه بوت استرپ یک درخت بقا رشد می‌کند. در هر گره درخت به‌طور تصادفی  $p$  متغیر به‌عنوان کاندید انتخاب می‌شود و گره با استفاده از متغیری که به‌عنوان کاندید انتخاب شده است افزایش می‌شود، به طوری که اختلاف بقای بین گره‌های درونی را به حداکثر برساند. درخت تحت این محدودیت که گره پایانی نباید کمتر از مرگ‌ومیرهای فرد باشد ( $dp > 0$ ) کاملاً رشد می‌کند. تابع خطر تجمعی (CHF) Cumulative hazard function را برای هر درخت محاسبه می‌کند و سپس میانگین این شاخص در همه درخت‌ها محاسبه می‌شود. با استفاده از داده‌های OOB، خطای پیش‌بینی برای تابع خطر تجمعی الگوریتم‌های دسته‌جمعی درختی محاسبه می‌شود.<sup>۱۵</sup>

برای ارزیابی عملکرد مدل‌های رگرسیون کاکس و RSF از شاخص خطای پیش‌بینی (که برابر یک منهای شاخص هماهنگی است) استفاده شد. این شاخص عددی بین صفر تا یک است که مقادیر کوچک نشان‌دهنده پیش‌بینی‌کنندگی خوب مدل و هر چه این مقدار کمتر از ۰/۵ باشد نشان می‌دهد مدل دارای عملکرد مناسب‌تری است یعنی با دقت بالاتری عوامل موثر بر بقا را شناسایی می‌کند. همچنین از شاخص نمره بریر (Brier Score) برای ارزیابی مدل‌ها استفاده شد. این معیار، میانگین مربع خطای پیش‌بینی و کیفیت پیش‌بینی یک مدل احتمالی را نشان می‌دهد. هرچه نمره بریر کوچک‌تر باشد پیش‌بینی دقیق‌تر است. فرم کلی نمره بریر به‌صورت زیر است:

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

گرفتند. در این مطالعه سن بالای ۱۵ سال جهت ورود به مطالعه در نظر گرفته شد. معیار خروج افراد از مطالعه ناقص بودن پرونده‌های کلینیکی آن‌ها بوده است.

در مجموع ۳۸۹ بیمار وارد مطالعه شدند. دو پرونده به‌علت اطلاعات ناقص از مطالعه خارج شدند. در نهایت از ۳۸۷ پرونده استفاده شد. بیماران پیگیری شده دارای پرونده بایگانی بیمارستان نیز بوده و آدرس و شماره تلفن برای پیگیری‌های بعدی وجود داشت. در این مطالعه سعی بر حفظ اسرار بیماران بوده و تمامی ملاحظات اخلاقی (کد اخلاق: IR.AJUMS.REC.1400.259) رعایت شده است. زمان بقای بیماران از زمان تشخیص تا اولین عود بر حسب ماه محاسبه گردید.

عوامل مؤثر بر زمان عود با استفاده از مدل کاکس و مدل RSF تعیین و مقایسه شدند. در این مدل‌سازی اهمیت ۱۱ متغیر سن، جنسیت، شیمی درمانی، LDH، تعداد گلبول‌های سفید، هموگلوبین کیفی، پلاکت، محل درگیری لنف، Histology، طبقه‌بندی کلاسیک لنفوم، مرحله بیماری مورد تحلیل و بررسی قرار گرفتند.

مدل RSF: این مدل یک روش جنگل‌های تصادفی برای تجزیه و تحلیل داده‌های بقا با سانسور از راست هستند. در این وضعیت که بقا سانسور از راست است RSF شامل زمان بقا و وضعیت سانسور می‌باشند. همچنین یک روش الگوریتم‌های دسته‌جمعی درختی می‌باشند. ساختار این الگوریتم‌ها از پایه می‌تواند مانند درختان، عملکرد پیش‌بینی را به میزان قابل‌توجهی بهبود بخشد. تصادفی‌سازی در جنگل‌های تصادفی به‌دو فرم معرفی می‌شود:

۱- به‌طور تصادفی یک نمونه بوت استرپ (Bootstrap) از داده‌ها برای رشد درخت استفاده می‌شود.

۲- در هر گره درخت به‌طور تصادفی، زیرمجموعه‌ای از متغیرهایی که به‌عنوان کاندید برای افزایش انتخاب شده‌اند، انتخاب می‌شود.

قانون‌های افزایش در جنگل‌های تصادفی بقا: قانون افزایش لگ رنک (log-rank split-rule): گروه‌هایی که بیش‌ترین مقدار آماره آزمون لگ رنک را دارند افزایش می‌کند و توسط Segal در سال ۱۹۸۸، LeBlanc and Crowley در سال ۱۹۹۳ توسعه یافت.

الف- قانون افزایش conservation-of-event: گروه‌هایی را افزایش می‌کند که گره‌های درونی نزدیک به قانون conservation-of-event پیدا کند.

**یافته‌ها**

در این مطالعه ۳۸۷ بیمار مبتلا به لنفوم هوچکین بررسی شدند. (۵۸/۷٪) ۲۲۷ بیمار مرد و بقیه زن بودند. میانه (ماکزیمم و مینیمم) سنی کل افراد حاضر (۷۸ و ۱۵) ۲۴ سال بود. در پایان مطالعه (۱/۲۷/۱) ۱۰۵ بیمار عود داشتند. اطلاعات توصیفی بیماران در جدول ۱ آمده است.

که  $Y_i(t)$  وضعیت رخداد رویداد برای فرد  $i$ -ام در زمان  $t$  و  $\bar{S}(t|X_i)$  احتمال بقا برای این فرد در زمان  $t$  است. ۱۹۱۴ و ۱۷۱۷ و ۱۸۱۹ برای برازش مدل RSF و مدل کاکس از پکیج‌های R software, version pec و randomForestSRC در نرم‌افزار R استفاده شده است. سطح معناداری آزمون ۰.۰۵٪ در نظر گرفته شده است.

جدول ۱: ویژگی‌های جمعیت‌شناختی بیماران مبتلا به لنفوم هوچکین

متغیر	رده	تعداد (درصد)*
جنسیت	مرد	۲۲۷(۵۸/۷)
	زن	۱۶۰(۴۱/۳)
شیمی درمانی	رژیم Stand ford V	۳۷(۹/۶)
	رژیم ABVD	۳۱۹(۸۲/۴)
هموگلوبین	کمتر از ۱۰/۵	۱۸۰(۴۶/۵)
	بیش‌تر از ۱۰/۵	۲۰۵(۵۳)
محل درگیری لنف	گردن	۲۲۴(۵۷/۹)
	زیر بغل	۳۱(۸)
Histology	کشاله ران	۲۰(۵/۲)
	سایر مکان‌ها	۶۳(۱۶/۳)
طبقه‌بندی کلاسیک لنفوم	Nodular lymphocyte predominant Hodgkin's Lymphoma	۳۲(۸/۳)
	Classical Hodgkin's Lymphoma	۲۸۸(۷۴/۴)
	Nodular sclerosing HL	۱۶۳(۴۲/۱)
	<b>Mixed – cellularity subtype</b>	۹۴(۲۴/۳)
	<b>Lymphocyte rich (Lymphocytic predominance)</b>	۲۱(۵/۴)
مرحله بیماری	<b>Lymphocyte depleted</b>	۲(۰/۵)
	II	۸۱(۲۰/۹)
	III	۸۵(۲۲)
	IV	۹۶(۲۴/۸)
	I	۲۴(۶/۲)
سن	میانه(ماکزیمم-مینیمم)	۲۴(۱۵, ۷۸)
LDH		۴۲۹(۱۳/۳۹, ۵۵۳۵)
تعداد گلبول‌های سفید		۸/۶(۰/۴, ۳۸)
پلاکت		۳۱۹(۳/۷, ۹۵۰)

\*تفاوت در جمع به‌دلیل داده‌های گمشده است.

آنالیز به‌کار رفته در این جدول صرفاً توصیفی است.

با توجه به گزارش شاخص‌های توصیفی، معناداری  $P < 0/05$  در این جدول گزارش نمی‌شود.

LDH ( $P=0/001$ ) و طبقه‌بندی کلاسیک لنفوم عوامل موثر بر عود بیماری می‌باشند. در ادامه برای تعیین اهمیت متغیرها بر عود بیماری، از مدل RSF با استفاده از روش افراز لگ-رتبه استفاده شد. این مدل با ۲۰۰ درخت (براساس کمترین میزان خطا) برازش داده شد.

میانگین (انحراف معیار) و میانه بقای این بیماران به ترتیب (۴۹/۲) (۵۸/۸ ماه و ۴۶ ماه بود. ابتدا به منظور بررسی عوامل موثر بر عود بیماری از مدل مخاطرات متناسب کاکس استفاده شد که نتایج آن در جدول ۲ آمده است. نتایج حاصل از مدل کاکس نشان می‌دهد که

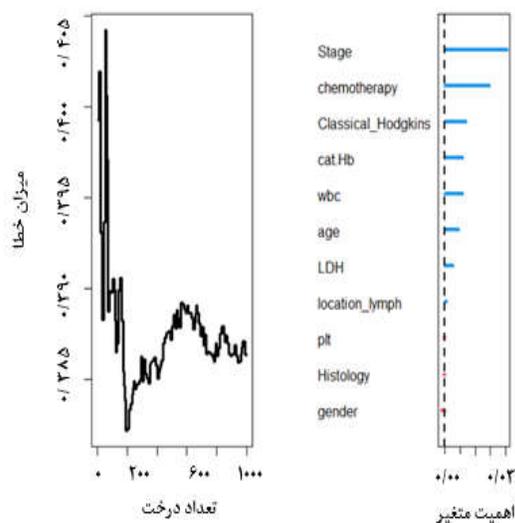
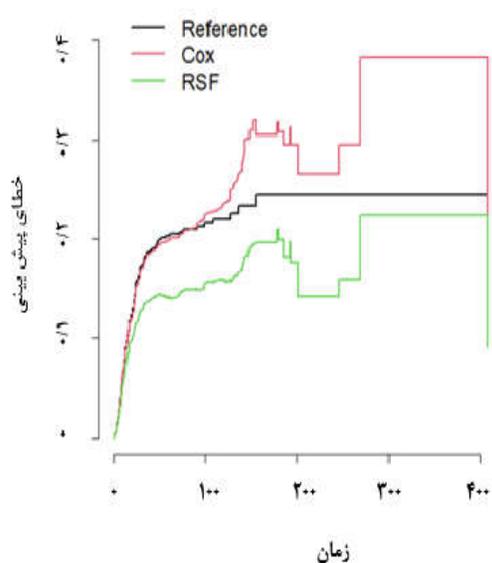
جدول ۲: نتایج برازش مدل چند متغیره کاکس

متغیر	ضریب رگرسیون	خطای معیار	نسبت مخاطره	فاصله اطمینان نسبت مخاطره		P*
				حد پایین	حد بالا	
سن	۰/۰۰۵۲	۰/۰۰۹۷	۱/۰۰۵	۰/۹۸۶	۱/۰۲۴	۰/۵۹
پلاکت	۰/۰۰۰۸	۰/۰۰۰۷	۱/۰۰۱	۰/۹۹۹	۱/۰۰۲	۰/۲۹
LDH	۰/۰۰۰۷	۰/۰۰۰۲	۱/۰۰۱	۱/۰۰۰۳	۱/۰۰۱۲	۰/۰۰۱
WBC	-۰/۰۶۵	۰/۰۳۵	۰/۹۳۷	۰/۸۷۵	۱/۰۰۳	۰/۰۶
جنسیت	-۰/۴۷۶	۰/۳۳۳	۰/۶۲	۰/۳۲۳	۱/۱۹۳	۰/۱۵
	---	---	---	---	---	---
شیمی درمانی	۰/۱۰۵	۰/۵۹۰	۱/۱۱۱	۰/۳۴۹	۳/۵۳۴	۰/۸۶
	---	---	---	---	---	---
هموگلوبین	۰/۳۲۶	۰/۳۹۹	۱/۳۸۵	۰/۶۳۳	۳/۰۳۲	۰/۴۱
	---	---	---	---	---	---
	۰/۶۸۶	۰/۶۷۳	۱/۹۸۷	۰/۵۳۱	۷/۴۳۹	۰/۳۱
محل درگیری	-۰/۰۹۵	۰/۸۰۴	۰/۹۰۹	۰/۱۸۸	۴/۳۹۲	۰/۹۱
لنف	۰/۷۳۶	۰/۳۸۵	۲/۰۸۸	۰/۹۸۲	۴/۴۳۹	۰/۰۶
	---	---	---	---	---	---
	-۰/۵۰۴۰	۰/۴۱۹۶	۰/۶۰۴۱	۰/۲۶۵۴	۱/۳۷۴۹	۰/۲۳۰
هیستولوژی	---	---	---	---	---	---
	-۰/۱۹۹	۰/۳۸۲	۰/۸۱۹	۰/۳۸۸	۱/۷۳	۰/۶۰
طبقه‌بندی	۰/۴۹۹	۰/۵۳۹	۱/۶۵	۰/۵۷۳	۴/۷۳۸	۰/۳۵
کلاسیک لنفوم	-۳۴/۹۵۸	۱/۰۷۸	۶/۵۸ <sup>-18</sup>	۷/۹۵ <sup>-14</sup>	۵/۴۴ <sup>-12</sup>	۰/۰۰۰۱
	---	---	---	---	---	---
مرحله بیماری	۰/۲۵۰	۰/۵۴۹	۱/۲۸۵	۰/۴۳۷	۳/۷۷۵	۰/۶۵
	۰/۶۷۲	۰/۴۸۳	۱/۹۵۹	۰/۷۵۹	۵/۰۴۹	۰/۱۶
	۰/۸۷۹	۰/۶۱۱	۲/۴۱	۰/۷۲۸	۷/۹۸۱	۰/۱۵
	---	---	---	---	---	---

\*آزمون Wald

نسبت به مدل RSF است. (نمودار ۲). در جدول ۳ میزان خطای پیش‌بینی، نمره بریر و شاخص هماهنگی را برای مقایسه عملکرد دو مدل کاکس و RSF نشان می‌دهد که همه شاخص‌های مدل RSF با قاعده افراز لگ-رتبه بهتر از مدل کاکس است. با توجه به اینکه مدل RSF دارای بالاترین شاخص هماهنگی و کمترین مقدار نمره بریر است، بنابراین این روش دارای بالاترین دقت و کمترین خطا در پیش‌بینی بقا و شناسایی مهم‌ترین عوامل موثر بر عود بیماری است.

براساس نتایج مدل RSF، مرحله بیماری، شیمی درمانی، طبقه‌بندی کلاسیک لنفوم، هموگلوبین بالای ۱۰/۵، تعداد گلوبول‌های سفید، سن تشخیص و LDH مهمترین متغیرهای تاثیرگذار بر عود بیماری شناسایی شدند (نمودار ۱). برای مقایسه عملکرد دو مدل کاکس و RSF در تعیین عوامل تاثیرگذار بر عود بیماری ابتدا خطای پیش‌بینی مدل در طی زمان مورد مطالعه ارزیابی شد. نتایج این مقایسه بیانگر بالا بودن خطای پیش‌بینی مدل کاکس



نمودار ۲: خطای پیش‌بینی براساس روش RSF با روش افراز لگ-رتبه. منحنی مدل کاکس نرخ خطای پیش‌بینی بالاتری نسبت به جنگل تصادفی بقا نشان می‌دهد.

نمودار ۱: نرخ خطا و ترتیب اهمیت متغیرها براساس روش RSF با روش افراز لگ-رتبه. جنگل تصادفی بقا با تعداد ۲۰۰ درخت کمترین مقدار نرخ خطای پیش‌بینی را نشان می‌دهد. متغیر مرحله بیماری و شیمی درمانی دارای بیشترین اهمیت در بین متغیرها هستند.

جدول ۳: مقایسه عملکرد مدل‌های کاکس و RSF با قاعده افراز لگ-رتبه

مدل	شاخص هماهنگی* (C-index)	نرخ خطای پیش‌بینی** (predictor error rate)	نمره بریر** (IBS)
Cox	۵۷/۶	۰/۱۳	۰/۲۸۶
RSF با قاعده افراز لگ-رتبه	۸۴/۹	۰/۰۹	۰/۱۷۵

\*نمره شاخص هماهنگی بیشتر بیانگر بهتر بودن مدل  
\*\*نرخ خطای پیش‌بینی و نمره بریر کمتر بیانگر بهتر بودن مدل

## بحث

انتظار نیست) به خوبی عمل نمی کند، بنابراین عدم معناداری متغیرها در این مطالعه بر خلاف مطالعات مشابه ممکن است به عدم وجود شرایط استفاده از این مدل (پیش فرض خطرات متناسب) مربوط باشد که اینگریچ نیز در مقاله خود به این موضوع اشاره کرده است.<sup>۲۳،۲۲،۲۱</sup> در این مطالعه مدل RSF نسبت به مدل کاکس عملکرد بهتری در تحلیل داده‌ها نشان داد. Miao و همکاران با استفاده از دو پایگاه داده جمعیت نارسایی قلبی و پایگاه داده آریمی‌های قلبی نشان دادند که مدل RSF نسبت به مدل کاکس کارایی بالایی دارد که با نتایج این مطالعه همخوانی دارد.<sup>۲۸</sup>

در این مطالعه براساس مدل RSF مرحله بیماری و سن به ترتیب از لحاظ اهمیت متغیر دارای رتبه اول و ششم بودند هر چند مدل کاکس این متغیرها را به عنوان عوامل پیش‌آگهی عود بیماری شناسایی نکرد. در بسیاری از مطالعات که مبتنی بر مدل کاکس بودند متغیر سن و مرحله بیماری بر بقای بیماران موثر بودند.<sup>۳۱،۳۲،۳۳،۳۴</sup> اگرچه در برخی مطالعات با افزایش سن، مخاطره مرگ نیز افزایش می‌یابد اما این متغیر تاثیر معناداری بر عود بیماری نداشته است.<sup>۳۰،۳۱،۳۲،۳۳،۳۴،۳۵</sup>

دلیل تفاوت نتایج در سایر مطالعات می‌تواند به متغیرهای گوناگون و تعداد متغیرهای مورد بررسی در هر یک از مطالعات و ساختار آن‌ها در مطالعات مختلف بستگی داشته باشد. اما در اکثر مطالعات مبتنی بر مدل کاکس، عدم معناداری این متغیر می‌تواند به دلیل عدم برقراری پیش‌فرض‌های مدل و همبستگی بین متغیرهای مستقل باشد. با این حال چون این بیماری معمولاً در سنین بالا شناسایی می‌شود بنابراین تاثیر معنادار این متغیر بر بقا دور از انتظار نیست.

در این مطالعه عملکرد مدل RSF بهتر از مدل کاکس بود که این نتیجه در برخی مطالعات مربوط به بقای سایر بیماری‌ها این موضوع تایید شده است.<sup>۳۱</sup> ولی در مطالعه حاضر مدل RSF با قاعده افراز لگ-رنگ نسبت به مدل کاکس عملکرد بهتری داشت به طوری که میزان شاخص هماهنگی این دو مدل به ترتیب برابر  $۸۴/۹\%$  و  $۵۷/۶\%$  به دست آمد، در مطالعه Ingrisich شاخص هماهنگی مدل RSF با قاعده افراز لگ-رنگ و مدل کاکس به ترتیب برابر  $۶۵/۷\%$  و  $۶۵/۲\%$  به دست آمد.<sup>۱۹</sup> در مطالعه Safari و همکاران در شناسایی عوامل موثر بر بقای افراد مبتلا به سرطان معده، عوامل شناسایی شده با استفاده از RSF بیشتر از مدل کاکس بود و عملکرد مدل RSF نسبت به مدل کاکس دارای خطای پیش بین و نمره بریر کمتر و شاخص هماهنگی بالاتری بود.<sup>۳۷</sup>

عوامل موثر بر زمان تا رخداد یک رویداد (مرگ، عود بیماری و تولد اولین فرزند) یکی از مهم‌ترین پیامدهای مورد علاقه در مطالعات بالینی است. آنالیز بقا روشی آماری برای ساختار اینگونه داده‌ها است. در این مطالعه ۳۸۷ بیمار مبتلا به لنفوم هوچکین که به مرکز آنکولوژی و هماتولوژی بیمارستان شغای اهواز مراجعه نمودند، مورد بررسی قرار گرفتند. میانه (ماکزیمم و مینیمم) سنی کل افراد حاضر  $۷۸ (۳, ۲۴)$  سال و  $۵۸/۷\%$  بیماران مرد بودند. میانه بقای این بیماران ۴۶ ماه بود که مشابه با برخی از مطالعات بوده است.<sup>۲۰،۲۱،۲۲،۲۳</sup> براساس نتایج حاصل از برازش مدل خطرات متناسب کاکس، متغیرهای LDH و طبقه‌بندی کلاسیک لنفوم عوامل موثر بر عود بیماری بودند، در حالی که براساس مدل RSF مهم‌ترین متغیرهای موثر بر عود بیماری به ترتیب مرحله بیماری، شیمی درمانی، طبقه‌بندی کلاسیک لنفوم، هموگلوبین، تعداد گلبول‌های سفید، سن، LDH، محل درگیری لنف بودند.

نتایج این مطالعه با برخی از مطالعات که زمان بقای کلی بیماران مبتلا به لنفوم هوچکین را بررسی نموده بودند، همخوانی داشت. در مطالعه گروس (Gross) و همکاران، نوع بیماری و رژیم شیمی درمانی بر بقا اثر معنادار داشتند.<sup>۳۳</sup>

در مطالعه Shenoy و همکاران، نژاد سیاه پوست و آسیایی، نژاد سفید پوست و آسیایی، آسیایی‌ها و سیاه‌پوستان با میانگین سنی ۳۸ سال در مقایسه با سفیدپوستان با میانگین سنی ۴۲ سال به عنوان عوامل موثر بر بقای کلی شناسایی شدند.<sup>۲۴</sup>

همچنین در برخی از مطالعات مرحله بیماری و سن از مهم‌ترین عامل پیش‌آگهی لنفوم هوچکین معرفی شده‌اند، نوع شیمی درمانی، لوسمی میلوئید حاد/سندرم‌های میلودیسه‌پلاستیکی و نئوپلاسم‌های بدخیم ثانویه نیز بر بقای بیماران مبتلا به لنفوم هوچکین موثر بودند.<sup>۲۵،۲۶</sup>

همچنین در مطالعه Elhaei و همکاران با بررسی عوامل پیش‌آگهی در عود بیماری لنفوم هوچکین، هموگلوبین کم‌تر از  $۱۰/۵$  یکی از عوامل موثر بر عود بیماری شناسایی شد.<sup>۲۷</sup>

نکته حایز اهمیت که معمولاً در مطالعات کمتر به آن دقت می‌شود این است که با توجه به این که مدل خطرات متناسب کاکس، در صورت انحراف از فرض خطرات متناسب (در مطالعات بقا دور از

کند بنابراین از مدل RSF که از دقت بالایی برخوردار است استفاده می‌شود. در واقع این مدل برای تحقیقات پزشکی که امروزه با حجم و بعد عظیمی از داده‌ها مواجه است، مناسب است. از خصوصیات دیگر این مدل سادگی در تفسیر و درک نتایج آن برای افراد غیرمتخصصین آماری در علوم مختلف از جمله علوم پزشکی است. سیاسگزرایی: این مقاله حاصل (بخشی از) پایان‌نامه تحت عنوان "تحلیل بقا با استفاده از درختان رگرسیون تجمعی و جنگل تصادفی بقا" در مقطع کارشناسی ارشد در سال ۱۴۰۰ و کد U-00130 می‌باشد که با حمایت دانشگاه علوم پزشکی جندی شاپور اهواز اجرا شده است.

همچنین در مطالعه Roshanaei و همکاران در تعیین عوامل موثر بر بقای پیوند کلیه با استفاده از مدل RSF، این مدل نسبت به روش Kaplan-Meier و مدل کاکس دارای عملکرد بهتری در شناسایی عوامل موثر و پیش‌بینی بقا افراد نشان داد.<sup>۳۷</sup> از محدودیت‌های مطالعه حاضر می‌توان به نقص در پرونده‌های پزشکی بیماران، تغییر شماره تماس آن‌ها و عدم دسترسی به برخی از بیماران اشاره کرد. نتیجه‌گیری: در صورتی که تعداد متغیرها زیاد و بین متغیرها رابطه وجود داشته باشد مدل کاکس به‌خوبی نمی‌تواند متغیرها را شناسایی

## References

- Kaufman D, Longo L. Hodgkin's disease in: E.Lenhord R, T.Osteen R, Gansler T. Clinical oncology. 2nd ed. United states of America: Churchill livingstone; 2000 Chapter 90. P. 2620- 48.
- Lees C, Keane C, Gandhi MK, Gunawardana J. Biology and therapy of primary mediastinal B-cell lymphoma: current status and future directions. *Br J Haematol* 2019;185(1):25-41.
- Amraee A, Evazi MR, Shakeri M, Roozbeh N, Ghazanfarpour M, Ghorbani M, et al. Efficacy of nivolumab as checkpoint inhibitor drug on survival rate of patients with relapsed/refractory classical Hodgkin lymphoma: a meta-analysis of prospective clinical study. *Clin Transl Oncol* 2019;21(8):1093-103.
- Metzger ML, Mauz-Körholz C. Epidemiology, outcome, targeted agents and immunotherapy in adolescent and young adult non-Hodgkin and Hodgkin lymphoma. *Br J Haematol* 2019;185(6):1142-157.
- Kaseb H, Babiker HM. Hodgkin Lymphoma. StatPearls [Internet]. 2020 Nov 21.
- Cox David R. Regression models and life-tables In Breakthroughs in Statistics. Springer; 1992.
- 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.
- Kang Y, McHugh MD, Chittams J, Bowles KH. Utilizing Home Healthcare Electronic Health Records for Telehomecare Patients With Heart Failure: A Decision Tree Approach to Detect Associations With Rehospitalizations. *Comput Inform Nurs* 2016;34(4):175-82.
- Yosefian I, Farkhani EM, Baneshi MR. Application of Random Forest Survival Models to Increase Generalizability of Decision Trees: A Case Study in Acute Myocardial Infarction. *Comput Math Methods Med* 2015;2015:576413.
- Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. *J Am Stat Assoc* 2010;105(489):205-17.
- Saha S. Survival analysis with Bayesian additive regression trees and its application. 2017.
- Nasejje JB, Mwambi H, Dheda K, Lesosky M. A comparison of the conditional inference survival forest model to random survival forests based on a simulation study as well as on two applications with time-to-event data. *BMC Med Res Methodol* 2017;17(1):115.
- Kleinbaum DG, Klein M. Survival analysis: a self-learning text: Springer; 2012.
- Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat* 2008;2(3):841-60.
- Ishwaran H, Kogalur UB. Random survival forests for R. *R News* 2007;7(2):25-31.
- Niknam M, Mahmoudi M, Andonian L, Akaberi A. Random forests analysis: A modern statistical method for screening in high-dimensional studies and its application in a population-based genetic association study. *J North Khorasan Univ Med Sci* 2012;3(5):93-101.
- Brier GW. Verification of forecasts expressed in terms of probability. *Mon weather Rev* 1950;78(1):1-3.
- Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biom J* 2006;48(6):1029-40.
- Ingrisch M, Schöppe F, Paprottka K, Fabritius M, Strobl FF, De Toni EN, et al. Prediction of 90Y Radioembolization Outcome from Pretherapeutic Factors with Random Survival Forests. *J Nucl Med* 2018;59(5):769-73.
- Roshanaei G, Kazemnejad A, Sadighi S. Survival estimating following recurrence in gastric cancer patients and its relative factors. *Koomesh* 2011;12(3):223-8.
- 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(1):129-34.
- 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. *J North Khorasan Univ Med Sci* 2015;6(4):689-96.
- Gross TG, Hale GA, He W, Camitta BM, Sanders JE, Cairo MS, et al. Hematopoietic stem cell transplantation for refractory or recurrent non-Hodgkin lymphoma in children and adolescents. *Biol Blood Marrow Transplant* 2010;16(2):223-30.
- Shenoy P, Maggioncald A, Malik N, Flowers CR. Incidence patterns and outcomes for hodgkin lymphoma patients in the United States. *Adv Hematol* 2011;2011:725219.
- Tzankov A, Meier C, Hirschmann P, Went P, Pileri SA, Dirnhofer S. Correlation of high numbers of intratumoral FOXP3+ regulatory T cells with improved survival in germinal center-like diffuse large B-cell lymphoma, follicular lymphoma and classical Hodgkin's lymphoma. *Haematologica* 2008;93(2):193-200.
- Eichenauer DA, Becker I, Monsef I, Chadwick N, De Sanctis V, Federico M, et al. Secondary malignant neoplasms, progression-free survival and overall survival in patients treated for Hodgkin lymphoma: a systematic review and meta-analysis of randomized clinical trials. *Haematologica* 2017;102(10):1748-57.
- Elhaei A, Malehi AS, Seghatoleslami M. Evaluation of Prognostic Variables on Time to Relapse In Patients with Hodgkin's Lymphoma Using Classified Cox Model. *Depiction of Health* 2019;10(3):227-34.

28. Miao F, Cai Y-P, Zhang Y-T, Li C-Y, editors. Is random survival forest an alternative to Cox proportional model on predicting cardiovascular disease? 6TH European conference of the international federation for medical and biological engineering; 2015: Springer.
29. Roshanaei G, Kazemnejad A, Sadighi S. Survival estimating following recurrence in gastric cancer patients and its relative factors. *Koomesh* 2011;12(3):223-8.
30. Roshanaei G, Sadighi S, Safari M, Faradmal J. Estimated survival time in gastric cancer patients and its associated factors. *Koomesh* 2012;14(1).
31. Roshanaei G, Safari M, Baghestani AR, Sadighi S. Assessment of the survival risk factors in patients with gastric cancer in Cancer Institute of Imam Khomeini Hospital between 2003-2007. *J Adv Med Biomed Res* 2012;20(80):40-50.
32. Yazdani J, Sadeghi S, Janbabaei G, Haghghi F. Applying survival analysis to estimate survival time in gastric cancer patients. *J Mazandaran Univ Med Sci* 2011;21(85):28-36.
33. 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 Mazandaran Univ Med Sci* 2014;24(119):11-8.
34. 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 Mazandaran Univ Med Sci* 2010;20(79):2-7.
35. Barfei F, Abbasi M, Khodabakhshi R, Gohari MR. Survival analysis of patients with adenocarcinoma gastric cancer in Fayazkhsh hospital, Tehran. *Razi J Med Sci* 2014;21(123):1-9.
36. Roshanaei G, Omid T, Faradmal J, Safari M, Poorolajal J. Determining affected factors on survival of kidney transplant in living donor patients using a random survival forest. *Koomesh J* 2018;20(3):517-23.
37. Safari M, Abbasi M, Gohari Ensaf F, Berangi Z, Roshanaei G. Identification of factors affecting metastatic gastric cancer patients' survival using the random survival forest and comparison with cox regression model. *Iran J Epidemiol* 2020;15(4):343-51.

## Evaluation of risk factors of recurrence of hodgkin's lymphoma using random survival forest and comparison with cox regression model

### Abstract

Received: 22 Jun. 2022 Revised: 29 Jun. 2022 Accepted: 16 Sep. 2022 Available online: 23 Sep. 2022

Zeinab Asakereh M.Sc.<sup>1</sup>  
Elham Maraghi Ph.D.<sup>1</sup>  
Bijan Keikhaei M.D.<sup>2</sup>  
Amal Saki Malehi Ph.D.<sup>1,3\*</sup>

1- Department of Biostatistic and Epidemiology, Faculty of Health Science, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

2- Thalassemia & Hemoglobinopathy Research Center, Health Research Institute, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

3- Pain Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

\* Corresponding author: Pain Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.  
Tel: +98-61-33115311  
E-mail: amalsaki@gmail.com

**Background:** In many studies, Cox regression was used to assess the important factors that affect the survival of cancer patients based on demographic and clinical variables. The aim of this study was to determine the factors affecting the survival of patients with Hodgkin's lymphoma using the random survival forest (RSF) method and compare it with the Cox model.

**Methods:** In this retrospective cohort study, all patients with Hodgkin's lymphoma who were referred to the Oncology and Hematology Center of Ahvaz Shafa Hospital from March 2000 to February 2010 were included. The survival time was calculated from diagnosis to the first recurrence event date (based on month). To assess the process of the disease, demographic characteristics and disease-related variables (including disease stage, chemotherapy, site of lymph involvement, etc.) were extracted from the records of 387 patients with Hodgkin's lymphoma. To investigate the prognostic factors that affect the recurrence of disease the Cox model and RSF were implemented. Moreover, their performance based on the C-index, IBS, and predictor error rate of the two models were compared. Data analysis was implemented by using R4.0.3 software (survival and RandomForestSRC packages).

**Results:** The results of the Cox model showed that LDH ( $P=0.001$ ) and classical lymphoma classification ( $P<0.001$ ) were associated with an increased risk of relapse in patients. However, the results of the RSF model showed that the important variables affecting the recurrence of disease were the stage of disease, chemotherapy, classical lymphoma classification, and hemoglobin, respectively. Also, the RSF model showed a higher (c-index=84.9) than the Cox model (c-index=57.6). Furthermore, the RSF model revealed a lower error rate predictor (0.09) and IBS index (0.175) than the Cox model. So, RSF has performed better than the Cox model in determining prognostic factors based on the suitability indicators of the model.

**Conclusion:** The RSF has high accuracy than the Cox model when there is a high number of predictors and there is collinearity. It can also identify the important variables that affect the patient's survival.

**Keywords:** cox model, hodgkin's lymphoma, recurrence, random forest, survival.

Copyright © 2022 Asakereh et al. Tehran University of Medical Sciences. Published by Tehran University of Medical Sciences.



This work is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.