مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۹۳، دوره ۱۰، شماره ۳: صفحات ۸-۱.

مقاله پژوهشی

پیش‌بینی بقای بیماران مبتلا به سرطان پستان با استفاده از دو مدل رگرسیون لجستیک و شبکه عصبی مصنوعی

آذر اسدآبادی[۱]، عباس بهرامپور[۲]، علی‌اکبر حق‌دوست[۳]

[۱] کارشناس ارشد آمار زیستی، گروه آمار زیستی و اپیدمیولوژی دانشکده بهداشت و مرکز تحقیقات مدل‌سازی در سلامت دانشگاه علوم پزشکی کرمان، کرمان، ایران

[۲] استاد گروه آمار زیستی و اپیدمیولوژی دانشکده بهداشت و مرکز تحقیقات مدل‌سازی در سلامت دانشگاه علوم پزشکی کرمان، کرمان، ایران

[۳] استاد گروه آمار زیستی و اپیدمیولوژی دانشکده بهداشت و مرکز تحقیقات مدل‌سازی در سلامت دانشگاه علوم پزشکی کرمان، کرمان، ایران

نویسنده رابط: عباس بهرامپور، نشانی: دانشگاه علوم پزشکی کرمان، دانشکده بهداشت، گروه آمارزیستی و اپیدمیولوژی، کد پستی: ۷۶۱۹۸۱۳۱۵۹، پست الکترونیک: abahrampour@yahoo.com

تاریخ دریافت: ۹۲/۱۲/۲۴؛ پذیرش: ۹۳/۲/۶

مقدمه و اهداف: در سال‌های اخیر، توجه قابل ملاحظه‌ای به مدل‌های آماری برای طبقه‌بندی داده‌های پزشکی با توجه به بیماری‌های مختلف و پیامدهای آن‌ها شده است. شبکه‌های عصبی مصنوعی به دلیل عدم نیاز به پیش‌فرض با موفقیت برای تشخیص الگو و پیش‌بینی در برخی از مطالعه‌های بالینی استفاده شده‌اند. هدف از این مطالعه، مقایسه دو مدل آماری شبکه عصبی مصنوعی و رگرسیون لجستیک برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان است.

روش کار: دو مدل برای داده‌های ثبت سرطان کرمان، جنوب شرق ایران، برای پیش‌بینی بقای هر بیمار به کار برده شدند. داده‌های مورد استفاده در این مطالعه شامل ۷۱۲ بیمار مبتلا به سرطان پستان در گروه سنی ۸۵-۱۵ سال بود. برای مقایسه پیش‌بینی بقا در داده‌های یاد شده، از مدل رگرسیون لجستیک و مدل شبکه عصبی مصنوعی پرسپترون سه لایه استفاده شد. برای مقایسه و تعیین تفاوت دو مدل، از حساسیت، ویژگی، صحت پیش‌بینی و سطح زیر منحنی مشخصه عملکرد (ROC) (Receiver Operative Characteristics) استفاده گردید.

نتایج: در این مطالعه، حساسیت و ویژگی دو مدل رگرسیون لجستیک و شبکه عصبی مصنوعی به ترتیب (۰/۵۹۴ و ۰/۷۰) و (۰/۶۲۱ و ۰/۷۲۳) به دست آمد. هم‌چنین صحت و سطح زیر منحنی راک به ترتیب (۰/۶۸۸و ۰/۷۲۵) و (۰/۷۰و ۰/۷۲۵) برای دو مدل رگرسیون لجستیک و شبکه عصبی پرسپترون بود.

نتیجه‌گیری: یافته‌ها نشان می‌دهند اگرچه تفاوت کمی در دو مدل وجود دارد، اما در این‌گونه داده‌ها برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان، مدل شبکه عصبی مصنوعی ابزار مناسب‌تری می‌باشد.

واژگان‌کلیدی: رگرسیون لجستیک، شبکه عصبی مصنوعی، پیش‌بینی بقای، سرطان پستان

مقدمه

سرطان پستان شایع‌ترین سرطان در میان زنان است (۲، ۱) که در مردان نیز رخ می‌دهد (۳). با توجه به بروز و شیوع بالای سرطان پستان، مطالعه‌های تحلیلی و اپیدمیولوژیک متعددی که هدف آن‌ها شناسایی عوامل خطرساز برای این بیماری می‌باشد، انجام شده است (۴). هر چند که دانشمندان برخی از عوامل خطر که شانس ابتلاء به سرطان پستان را افزایش می‌دهند؛ می‌شناسند، اما هنوز آن‌چه که باعث بیش‌تر سرطان‌های پستان می‌شود و یا این‌که دقیقاً چه تعدادی از این عوامل خطر، باعث سرطانی شدن سلول‌ها می‌شوند را نمی‌دانند. این عوامل ویژگی‌هایی مانند سن، خطر ژنتیکی و سابقه خانوادگی می‌باشند (۱،۳).

درمان‌های سرطان پستان به دو نوع اصلی موضعی و سیستماتیک تقسیم می‌شوند. جراحی و رادیوتراپی، نمونه‌ای از درمان موضعی و شیمی درمانی و هورمون درمانی نمونه‌ای از روش درمانی سیستماتیک هستند. معمولاً برای نتایج بهتر، دو نوع درمان با هم استفاده می‌شوند. اگر چه سرطان پستان دومین علت عمده مرگ ناشی از سرطان در زنان است، اما میزان بقای آن زیاد می‌باشد و با تشخیص زودرس و ارایه راه‌کارهای ویژه می‌توان منجر به کاهش مراجعه دیرهنگام، ارایه درمان مؤثر به منظور افزایش بقاء، کاهش مرگ و ارتقای کیفیت زندگی بیماران شد. به این ترتیب ۹۷ درصد از زنان حداقل ۵ سال زنده می‌مانند (۱،۵). با توجه به اهمیت وشیوع سرطان پستان به عنوان دومین علت مرگ در بین بیماری‌های سرطانی در جهان، دسترسی به

[Downloaded from journals.tums.ac.ir on 2026-04-10]

مدل‌هایی که با دقت بالا بتوانند بقای این بیماران را در افراد مبتلا پیش‌بینی‌نمایند، مورد توجه است.

در سال‌های اخیر، توجه قابل ملاحظه‌ای به مدل‌های آماری برای طبقه‌بندی داده‌های پزشکی با توجه به بیماری‌های مختلف و پیامدهای آن‌ها شده است (۶). شبکه‌های عصبی مصنوعی با موفقیت برای تشخیص الگو و پیش‌بینی در زمینه‌های بالینی استفاده شده‌اند (۷). آن‌ها برای پیش‌بینی در زمینه‌های متعددی از جمله قلب و عروق، زیست‌شناسی مولکولی، پیامدهای تروما، نوزادان و انکولوژی به کار برده شده‌اند (۸).

یکی از مزایای استفاده از مدل شبکه عصبی، بررسی ارتباطات غیر خطی و اثرات متقابل پیچیده بین عوامل می‌باشد (۳و۹). این مدل بر پایه «پیش‌فرض‌ها» نیست و ارتباط بین عوامل را که برخی از تکنیک‌های آماری مانند رگرسیون چندگانه ممکن است قادر به تشخیص نباشد، ممکن می‌سازد (۸). هم‌چنین مزایای استفاده از شبکه عصبی مصنوعی برای پیش‌بینی یک متغیر پاسخ دو سطحی، شامل الزامات کم‌تر آموزش آماری، توانایی تشخیص روابط پیچیده کاملاً غیرخطی بین پاسخ و پیش‌بینی کننده‌ها، در نظر گرفتن همه اثرات متقابل ممکن بین پیش‌بینی کننده‌ها و در دسترس بودن الگوریتم‌های آموزشی متعدد می‌باشد. با وجود این، شبکه عصبی معایبی مانند ماهیت «جعبه سیاه»[1] بودن آن، بار محاسباتی بیش‌تر، تمایل برای بیش‌برازشی و ماهیت تجربی برازش مدل را دارد (۱۰).

مطالعه‌های متعددی در کشورهای مختلف برای پیش‌بینی پیامدهای مختلف در زمینه سرطان با به کار بردن شبکه‌های عصبی مصنوعی و هم‌چنین مقایسه این مدل‌ها با مدل‌های قدیمی آماری انجام شده است، که می‌توان به موارد زیر اشاره کرد:

لوندین و همکاران (۱۹۹۹) شبکه‌های عصبی را برای پیش‌بینی بقای ۵، ۱۰ و ۱۵ ساله‌ی بیماران مبتلا به سرطان مورد استفاده قرار دادند (۷). ماچادو (۲۰۰۱) مدل‌های شبکه عصبی مصنوعی، کاکس و رگرسیون لجستیک را برای تحلیل‌های بقا مورد استفاده قرار داد (۱۱). آریهیتو و همکاران (۲۰۰۸) مدل‌های شبکه عصبی مصنوعی، بیز، شبکه‌ی بیزی، انواع درخت تصمیم و رگرسیون لجستیک را برای پیش‌بینی بقای بیماران مبتلا به سرطان، استفاده کردند. نتایج مطالعه این پژوهشگران نشان داد که رگرسیون لجستیک بیشترین دقت را در مقایسه با سایر مدل‌ها دارد (۱۲). دارسون و همکاران (۲۰۰۵) سه مدل شبکه عصبی مصنوعی، درخت تصمیم و رگرسیون لجستیک را برای پیش‌بینی بقای بیماران مبتلا به سرطان، مورد استفاده قرار دادند. نتایج آن‌ها نشان داد که درخت تصمیم بهترین پیش‌بینی کننده با دقت ۹۳/۶ درصد، سپس شبکه عصبی با دقت ۹۱/۲ درصد و در نهایت رگرسیون لجستیک با دقت ۸۹/۲ درصد برای بقای بیماران می‌باشند (۱). جائه سونگ و همکاران (۲۰۰۵) در یک تحلیل مقایسه‌ای با استفاده از رگرسیون لجستیک و شبکه عصبی مصنوعی برای تشخیص توده‌های پستان، نشان دادند که شبکه عصبی مصنوعی عملکرد بهتری نسبت به رگرسیون لجستیک برای تشخیص دارد (۶). بوردس و همکاران (۲۰۱۰) شبکه عصبی مصنوعی و رگرسیون لجستیک را برای پیش‌بینی پیامدهای بیماران مبتلا به سرطان پستان مورد استفاده قرار دادند. نتایج مطالعه آن‌ها نشان داد که شبکه‌های عصبی در پیش‌بینی دقیق‌تر بودند (۸). کریستین مک لارن و همکاران (۲۰۰۹) شبکه عصبی مصنوعی و رگرسیون لجستیک را برای پیش‌بینی ضایعات بدخیم پستانی از ام‌آرآی مورد استفاده قرار دادند. نتایج مطالعه آن‌ها نشان داد که اجرای تشخیصی دو مدل مشابه بود (۱۰). شو تانگ چن و همکاران (۲۰۰۹) روش‌های شبکه عصبی مصنوعی، دستگاه بردار حفاظتی و رگرسیون لجستیک را برای تشخیص بدخیم یا خوش‌خیم بودن تومورهای پستان مورد استفاده قرار دادند. تحلیل ROC در مطالعه این پژوهشگران نشان داد که تفاوت معنی‌داری در اجرای این سه روش برای تشخیص تومورها وجود ندارد (۱۳). ماریانی و همکاران (۱۹۹۷) روش‌های کاکس و شبکه عصبی مصنوعی را برای برآورد فاکتورهای پیش‌آگهی در عود سرطان پستان استفاده کردند. نتایج مطالعه آن‌ها نشان داد که مقدار پیش‌بینی هر دو مدل یکسان بود (۱۴). عبدالمالکی و همکاران (۲۰۰۴) مدل‌های شبکه عصبی مصنوعی و رگرسیون لجستیک را برای پیش‌بینی پیامد بیوپسی در سرطان پستان از یافته‌های ام‌آرآی مورد استفاده قرار دادند. نتایج مطالعه این پژوهشگران نشان داد که رگرسیون لجستیک با توجه به شاخص‌های حساسیت، ویژگی، صحت پیش‌بینی، درصد پیش‌بینی‌های ناصحیح و سطح زیر منحنی راک، عملکرد بهتری نسبت به شبکه عصبی دارد (۱۵).

با توجه به این‌که رگرسیون لجستیک، یکی از رایج‌ترین مدل‌های مورد استفاده برای طبقه‌بندی و پیش‌بینی است و عموماً به عنوان یک روش در پیش‌بینی پیامدهای دو یا چند سطحی به

[1] Black box

[Downloaded from journals.tums.ac.ir on 2026-04-10]

وجود آمده است (۱۱،۱۶)، و هم‌چنین به طور گسترده در مطالعه‌های اپیدمیولوژی به علت توانایی آن برای برآورد احتمال با استفاده از یک فرمول لجستیک، استفاده می‌شود (۱۳)، مقایسه عملکرد تشخیصی این تکنیک و مدل شبکه عصبی، به دلیل تفاوت‌های ذاتی آن‌ها مورد توجه است. به طور خلاصه هنوز برتری این مدل‌ها در داده‌های متفاوت مشخص نشده است. در نتیجه، روش کلی برای انتخاب مدل در طبقه‌بندی و پیش‌بینی داده‌ها وجود ندارد. هدف از این مطالعه، مقایسه دو مدل آماری شبکه عصبی مصنوعی و رگرسیون لجستیک برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان است. منظور از بقای این است که بیمار ۵ سال پس از زمان تشخیص هنوز زنده باشد.

روش کار

دو مدل برای داده‌های ثبت سرطان کرمان، برای پیدا کردن بقای هر بیمار به کار برده شدند و برای تعیین بقا در زمان انجام پژوهش، داده‌های ثبت سرطان و داده‌های مربوط به مرگ‌ومیر با استفاده از نام و نام‌خانوادگی، نشانی، نام پدر، و سن هر بیمار تطابق داده شدند. داده‌ها شامل ۷۱۲ بیمار مبتلا به سرطان پستان در گروه سنی ۸۵-۱۵ سال بود. این مطالعه بقا را به مدت ۵ سال پس از تشخیص بیماری پیش‌بینی می‌کند. متغیرهای مستقل مورد استفاده در این مطالعه عبارت‌اند از: سن، طول عمر پس از تشخیص بیماری تا زمان انجام پژوهش (مدت زنده بودن بیماران از سال تشخیص بیماری تا زمانی که پژوهش انجام شد)، مورفولوژی، جراحی، شیمی‌درمانی، رادیوتراپی، درجه تمایز تومور و متغیر وابسته عبارت است از بقای ۵ ساله بیماران از زمان تشخیص که به عنوان یک متغیر دو حالتی با کد صفر و یک کد گذاری شده است.

مدل شبکه عصبی مصنوعی:

شبکه‌های عصبی مصنوعی یکی از رایج‌ترین مدل‌های مورد استفاده بر اساس ساختار انسان‌شناختی هستند (۲،۱۲). آن‌ها مدل‌های ریاضی هستند که معمولاً به عنوان برنامه‌های کامپیوتری تخصصی، طراحی شده و برای بررسی مسایل پیچیده از طریق شبیه‌سازی ساختار نورون‌های به هم پیوسته مغز و مکانیسم‌های یادگیری آن اجرا می‌شوند (۲،۱۴). از این دیدگاه، شبکه‌های عصبی مصنوعی می‌توانند به عنوان شبکه‌هایی از واحدهای ساده (نورون‌ها یا گره‌ها) دیده شوند که برای یادگیری روابط بین داده‌ها آموزش می‌یابند. برخی از انواع مختلف شبکه‌های عصبی مصنوعی (چند لایه پرسپترون[۱]، تابع مبانی شعاعی[۲]، هاپفیلد[۳] و خودسازمان کوهنن[۴]) برای حل مسایل غیرخطی از طریق یادگیری پیشنهاد می‌شوند (۲،۱۴،۱۷).

به طور کلی، نورون‌ها (گره‌ها) در شبکه‌های عصبی مصنوعی به سه لایه ورودی، خروجی و مخفی (یک یا چند لایه) می‌توانند تقسیم شوند. تعداد گره‌ها در لایه ورودی مطابق با تعداد متغیرهای پیش‌بینی کننده است. تعداد گره‌ها در لایه خروجی مطابق با تعداد متغیرهای پاسخ است. تعداد گره‌های لایه مخفی معمولاً توسط روش آزمون و خطا تعیین می‌شوند (۲،۹).

با توجه به نوشته‌ها و برخی از تحلیل‌های تجربی قبلی، بیش‌تر از شبکه عصبی پرسپترون چندلایه با الگوریتم پس انتشار خطا (یک الگوریتم یادگیری تحت نظارت)[۵] برای پیش‌بینی استفاده می‌شود (۸،۱۷). شبکه عصبی پرسپترون چند لایه به عنوان تابع تقریبگر[۶] قدرتمند برای مسایل پیش‌بینی و طبقه‌بندی شناخته شده است. اجراهای تجربی پژوهشگران نیز این نکته را اثبات می‌کنند که این نوع شبکه عصبی مصنوعی نسبت به ساختارهای دیگر شبکه‌های عصبی بهتر است (۱).

ساختار شبکه در توصیف ارتباط بین گره‌ها با استفاده از الگوریتم آموزشی مناسب، برای پیدا کردن مقادیر پارامترهای شبکه (وزن‌ها و بایاس) طراحی شده است. هر نورون یک مجموع وزنی از خروجی واحدهای لایه قبلی را محاسبه می‌کند، به عبارتی واحد خروجی با توجه به یک تابع انتقال (فعالیت یا محرک) مشخص شده تولید می‌شود. تابع نورون‌ها در لایه مخفی، داوری بین ورودی و خروجی شبکه عصبی می‌باشد (۱۷, ۲).

وزن‌ها و بایاس‌های شبکه عصبی از طریق آموزش با یک روش دو مرحله‌ای تعیین می‌شوند. مرحله اول یک حالت بسیار کوتاه از الگوریتم پس انتشار خطا با یک نرخ آموزش در حد متوسط است. مرحله دوم یک اجرای طولانی‌تر از کاهش گرادیان، یک الگوریتم بسیار قوی‌تر است که به احتمال کمتری با مشکلات همگرایی

---

[۱] Multilayer perceptron
[۲] Redial basis function
[۳] Hopfield
[۴] Kohonen self organization
[۵] Supervised learning
[۶] Approximator function

[Downloaded from journals.tums.ac.ir on 2026-04-10]

نسبت به دیگر الگوریتم‌ها به دلیل استفاده از پس انتشار در مرحله اول مواجه می‌شود. در طی این فرآیند یادگیری، وزن‌ها در شبکه عصبی پرسپترون چند لایه با استفاده از برازش حداقل مربعات همراه با آموزش دو مرحله‌ای برای به حداقل رساندن تابع خطای ریشه مربع میانگین تغییر می‌کنند (۸).

به طور قراردادی در حالت ساده اما معمول، در شبکه عصبی با تنها یک لایه مخفی و یک واحد خروجی، فرض این است که y_i نشان‌دهنده خروجی، w_{jh} وزن ارتباط بین ورودیj ام و واحد مخفی h ام ، w_{jh} وزن ارتباط بین واحد مخفی h ام و نورون خروجی، α_h و α ثابت‌های بایاس می‌باشند. همچنین فرض بر این است که f_h و f_o به ترتیب توابع محرک واحدهای لایه مخفی و خروجی هستند که می‌توانند توابعی مانند تابع لجستیک، تانژانت هایپربولیک، خطی و ... باشند. بر اساس نمادهای فوق، مدل شبکه توسط فرمول زیر ارائه می‌شود:

$$y_i = f_o\left(\alpha + \sum_{h=1}^{H} w_h f_h\left(\alpha_h + \sum_{j=1}^{p} w_{jh} x_{ij}\right)\right)$$

شبکه پیامد را با ارائه مقادیر ورودی از یک مجموعه داده‌های آموزشی، همچنین با برآورد مجموعه‌ی وزن‌ها از طریق آموزش، پیش بینی می‌کند. در طول آموزش، وزن‌های ارتباط به طور تکراری در چنین روشی برای به حداقل رساندن اختلاف بین خروجی تولید شده توسط شبکه و پیامدهای واقعی مشاهده شده، به روز می‌شوند (۱۸, ۱۷, ۱۵).

در این مطالعه، ابتدا ۶۵٪ از داده‌ها (۴۵۸ نفر) برای برازش مدل رگرسیون لجستیک و آموزش مدل شبکه عصبی مصنوعی و سپس ۳۵٪ دیگر (۲۵۴ نفر) برای تست پیش‌بینی صحیح مدل‌ها استفاده شدند. برای مدل رگرسیون لجستیک، همه متغیرها با روش اینتر در مدل مورد استفاده قرار گرفتند. برای طراحی مدل شبکه عصبی مصنوعی، یک شبکه پرسپترون سه لایه با ۱۶ نود ورودی، ۱۴ نود درلایه مخفی و ۲ نود خروجی با الگوریتم یادگیری پس انتشار خطا و بهینه‌سازی کاهش گرادیان، تابع محرک سیگموئید در لایه مخفی (میانی) و لایه خروجی، نرخ یادگیری ۰/۲ و ممنتم (اندازه حرکت) ۰/۸۵ با متغیرهای مشابه در مدل لجستیک استفاده شد که برای انتخاب این ساختار، مدل‌های مختلف شبکه عصبی سه لایه را با ۳۰ ساختار مبتنی بر ۱۵-۲ نود در لایه مخفی، اندازه‌های حرکت ۰/۹۵-۰/۸۰ ، نرخ یادگیری ۰/۰۵-۰/۵ برازش داده شدند و در نهایت با توجه به کم‌ترین خطا (مجموع مربعات خطا) مدل بهینه انتخاب شد. برای برازش مدل‌ها از نرم‌افزار SPSS نسخه ۱۸ و نرم‌افزار Stata نسخه ۱۱ استفاده شد.

با توجه به این‌که هدف اصلی در این مطالعه، مقایسه قدرت پیش‌بینی‌کنندگی مدل‌های شبکه عصبی مصنوعی و رگرسیون لجستیک برای بقای ۵ ساله بیماران مبتلا به سرطان پستان است، باید معیارهایی را در مقایسه آماری بین این دو مدل به کار برد که بتوان یکی از آن‌ها را به عنوان مدل بهینه انتخاب کرد و مورد استفاده قرار داد. هم‌چنین با توجه به متغیر پاسخ دوحالته، معیارهای مناسب در ارزیابی کیفیت مدل‌های برازش شده و تعیین توانایی‌شان برای پیش‌بینی صحیح، حساسیت، ویژگی، صحت پیش‌بینی و سطح زیر منحنی مشخصه عملکرد (ROC) می‌باشند (۱۹، ۳،۱۸).

یافته‌ها

در مطالعه حاضر، (۸۵/۷ درصد)۶۱۰ نفر که سال تشخیص بیماری آن‌ها بین سال‌های ۸۵-۱۳۷۸ بود و تا سال ۱۳۹۰ پیگیری شده‌اند، ۵ سال پس از تشخیص بیماری زنده بودند و ۱۰۲ نفر (۱۴/۳ درصد) فوت کرده‌اند. جدول شماره ۱ اطلاعات توصیفی مربوط به متغیرهای مستقل را نشان می‌دهد.

تعداد موارد واقعی و تعداد موارد پیش‌بینی شده توسط دو مدل در جدول شماره ۲ آورده شده است و در نهایت، جدول شماره ۳ شاخص‌های مقایسه دو مدل را نشان می‌دهد.

با توجه به تعداد پیش‌بینی‌های انجام شده توسط دو مدل رگرسیون لجستیک و شبکه عصبی مصنوعی و تعداد مشاهدات واقعی (جدول شماره ۲)، هم‌چنین معیارهای مقایسه مدل‌ها (جدول شماره ۳)، در تمامی معیارهای مقایسه به جزء منحنی ROC، مقادیر مربوط به شبکه عصبی مصنوعی بیشتر از رگرسیون لجستیک می‌باشد.

[Downloaded from journals.tums.ac.ir on 2026-04-10]

جدول شماره ۱- متغیرهای مستقل مورد استفاده در دو مدل شبکه عصبی مصنوعی و رگرسیون لجستیک یا تاثیرگذار در بقای سرطان پستان

متغیر	مقدار
میانگین±انحراف معیار طول عمر از زمان تشخیص سرطان تا زمان انجام پژوهش	۶/۴۵ ± ۲/۳۳
دریافت شیمی‌درمانی	
بله	۱۶۲(٪۲۲/۸)
خیر	۵۵۰(٪۷۷/۲)
دریافت جراحی	
بله	۳۹۸(٪۵۵/۹)
خیر	۳۱۴(٪۴۴/۱)
دریافت رادیوتراپی	
بله	۳۶۹(٪۵۱/۸)
خیر	۳۴۳(٪۴۸/۲)
درجه تمایز تومور	
خوب	۸۹(٪۱۲/۵)
متوسط	۳۴۰(٪۴۷/۸)
ضعیف	۲۸۳(٪۳۹/۷)
مورفولوژی	
نئوپلاسم، بدخیم	۳۲۶(٪۴۵/۸)
کارسینوم	۲۳ (٪۳/۲)
کارسینوم مجرایی	۳۴۴(٪۴۸/۳)
کارسینوم مدولی	۸ (٪۱/۱)
کارسینوم لوبولی	۱۱ (٪۱/۵)
میانگین±انحراف معیار سن	۴۸/۰۳ ± ۱۲/۲۷

جدول شماره ۲- مشاهدات واقعی و موارد پیش‌بینی شده توسط دو مدل شبکه عصبی مصنوعی و رگرسیون لجستیک

	وضعیت واقعی	
پیش‌بینی مدل	مرگ	بقا
رگرسیون لجستیک		
مرگ	۲۲	۶۴
بقا	۱۵	۱۵۳
شبکه عصبی مصنوعی		
مرگ	۲۳	۶۰
بقا	۱۴	۱۵۷

[Downloaded from journals.tums.ac.ir on 2026-04-10]

جدول شماره ۳- معیارهای مقایسه پیش‌بینی دو مدل شبکه عصبی مصنوعی و رگرسیون لجستیک

مدل	حساسیت	ویژگی	صحت پیش‌بینی	ROC
رگرسیون لجستیک	۰/۵۹۴	۰/۷۰	۰/۶۸۸	۰/۷۲۵
شبکه عصبی	۰/۶۲۱	۰/۷۲۳	۰/۷۰	۰/۷۲۵

بحث

سرطان پستان یک مسأله مهم اپیدمیوژیک با گسترش جهانی و یکی از مهم‌ترین علل مرگ و میر و در واقع دومین علت مرگ بر اثر سرطان‌ها در زنان می‌باشد. در حال حاضر این سرطان، شایع‌ترین بدخیمی تشخیص داده شده در تمام طول عمر زنان و علت اصلی منتهی به مرگ در سنین ۷۹-۴۰ می‌باشد که نقشی مشابه بیماری‌های قلبی- عروقی در مردان بازی می‌کند. در کشورهای غربی این بیماری بیش‌تر در سنین بالای ۵۰ سال دیده می‌شود، اما در کشور ما بیماران جوان‌تر هستند و در بسیاری از موارد به علت عدم آگاهی از علایم بیماری، در مراحل پیشرفته‌تری مراجعه می‌کنند. از این نظر، تشخیص زودرس این بیماری از اهمیت ویژه‌ای در میزان بقای بیماران برخوردار است (۲۰).

در این مطالعه، دو مدل رگرسیون لجستیک و شبکه عصبی مصنوعی برای پیش‌بینی بقای ۵ ساله بیماران مبتلا به سرطان پستان طراحی شدند. حساسیت و ویژگی برای دو مدل رگرسیون لجستیک و شبکه عصبی مصنوعی به ترتیب (۰/۵۹۴ و ۰/۷۰) و (۰/۶۲۱ و ۰/۷۲۳) به دست آمد. دقت و سطح زیر منحنی راک به ترتیب (۰/۶۸۸و ۰/۷۲۵) و (۰/۷۰و ۰/۷۲۵) در مدل لجستیک و شبکه عصبی بود. یافته‌ها نشان می‌دهند با این که اختلاف کمی بین عملکرد دو مدل برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان وجود دارد، هم‌چنین سطح زیر منحنی ROC در دو مدل یکسان است اما در شاخص‌های حساسیت، ویژگی و صحت پیش‌بینی، مقادیر مدل شبکه عصبی بیشتر از مدل رگرسیون لجستیک می‌باشد. در مطالعه‌ای که در ژاپن توسط آریهیتو و همکاران برای پیش‌بینی بقای ۵ ساله بیماران مبتلا به سرطان پستان با مقایسه ۷ مدل رگرسیون لجستیک، شبکه عصبی مصنوعی، شبکه بیز و انواع درخت تصمیم انجام شد، دقت مدل رگرسیون لجستیک از دیگر مدل‌ها بیش‌تر بود (۱۲). در مطالعه‌ای که در آمریکا توسط دارسون و همکاران برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان انجام شد، مدل درختی در مقایسه با دو مدل شبکه عصبی و رگرسیون لجستیک دقت بیشتری داشت (۱). نتایج مطالعه حاضر با این مطالعه‌ها هم‌خوانی ندارد، و با توجه به مقادیر، شبکه عصبی عملکرد بهتری نسبت به رگرسیون لجستیک دارد که این می‌تواند به دلیل تفاوت در انتخاب متغیرهای تأثیرگذار در پیش‌بینی بقای بیماران مبتلا به سرطان، در این مطالعات با مطالعه حاضر باشد.

از آن‌جا که دو مدل تفاوت کمی در عملکرد برای پیش‌بینی بقای بیماران داشتند و با توجه به این که توزیع پارامترهای شبکه عصبی مشخص نیست و امکان استنباط آماری برای پارامترها وجود ندارد، نمی‌توان اهمیت هر متغیر را در پیش‌بینی بقا مشخص کرد در نتیجه، استفاده از رگرسیون لجستیک در حالتی که اهمیت متغیرها مد نظر است؛ مناسب‌تر می‌باشد، اما اگر هدف مطالعه مانند مطالعه حاضر، پیش‌بینی یا طبقه‌بندی باشد، شبکه عصبی مصنوعی ابزار مناسب‌تری است.

نتیجه‌گیری

یافته‌ها نشان می‌دهند مدل شبکه عصبی مصنوعی ابزار مناسب‌تری برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان در این گونه داده‌ها می‌باشد و بدین‌وسیله با تشخیص زودرس بیماری در سنین پایین‌تر و در درجه‌های ابتدایی بیماری و انجام درمان‌های لازم، می‌توان بقای این بیماران را افزایش داد.

تشکر و قدردانی

پژوهشگران، همکاری مرکز تحقیقات مدل‌سازی در سلامت و معاونت پژوهشی دانشگاه علوم پزشکی کرمان را در تصویب و تأمین اعتبار این پژوهش، شایسته قدردانی می‌دانند.

[Downloaded from journals.tums.ac.ir on 2026-04-10]

منابع

1. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine, 2005; 34: 113-27.
2. Chou SM, Lee TS, Shao YE, Chen IF. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 2004. 27: 133-42.
3. Jerez-Aragones J, Gomez-Ruiz J, Ramos-Jimenez G, Munoz-Perez J, Alba-Conejo E . A combined neural network and decision trees model for prognosis of breast cancer relapse. Artificial intelligence in medicine, 2003. 27: 45-63.
4. Ronco AL. Use of artificial neural networks in modeling associations of discriminant factors: towards an intelligent selective breast cancer screening. Artificial intelligence in medicine, 1999. 16: 299-309.
5. .http : //www.mums.ac.ir/ Breast Cancer 2012.
6. Song JH, Venkatesh SS, Conant EA, Arger PH, Sehgal CM. Comparative Analysis of Logistic Regression and Artificial Neural Network for Computer-Aided Diagnosis of Breast Masses. Academic Radiology, 2005, 12: 487-495.
7. Lundin M, Lundin J, Burke H, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. Oncology. 2000; 57: 281-86.
8. Lisboa P, Bourdes V, Bonnevay S, Defrance R, Perol D, Chabaud S. Comparison of artificial neural network with logistic regression as classification models for variable selection for prediction of breast cancer patient outcomes. Advances in Artificial Neural Systems. Volume 2010, Article ID 309841, 10 pages.
9. Franco L, Jerez JM, Alba E. Artificial neural networks and prognosis in medicine. Survival analysis in breast cancer patients. European Symposium on Artificial Neural Networks, ISBN 2-930307-05-6, 2005: 27-29.
10. McLaren CE, Chen WP, Nie K, Su MY. Prediction of Malignant Breast Lesions from MRI Features: A Comparison of Artificial Neural Network and Logistic Regression Techniques. Academic radiology, 2009. 16:842-51.
11. Ohno-Machado L. Modeling medical prognosis :survival analysis techniques. Journal of biomedical informatics. 2001; 34: 428-39.
12. Endo A, Shibata T, Tanaka H. Comparison of seven algorithms to predict breast cancer survival. Survival. 2008; 30: 11-16.
13. Chen ST, Hsiao YH, Huang YL, Kuo SJ, Tseng HS, Wu HK, Chen DR. Comparative Analysis of Logistic Regression, Support Vector Machine and Artificial Neural Network for the Differential Diagnosis of Benign and Malignant Solid Breast Tumors by the Use of Three-Dimensional Power Doppler Imaging. Korean Journal of Radiology. 2009, 10.
14. Mariani L, Coradini D, Biganzoli E, Borac chi P, Marubini E, Pilotti S. Prognostic factors for metachronous contralateral breast cancer: a comparison of the linear Cox regression model and its artificial neural network extension. Breast cancer research and treatment. 1997; 44: 167-78.
15. Abdolmaleki P, Yarmohammadi M, Gity M. Comparison of logistic regression and neural network models in predicting the outcome of biopsy in breast cancer from MRI findings. Iranian Journal of Radiation Research. 2004; 1: 217-28.
16. Kayri M, Cokluk O. Using Multinomial Logistic Regression Analysis In Artificial Neural Network: An Application. 2010.
17. Sordo M. Introduction to neural networks in healthcare. For Open Clinical: Knowledge Management for Medical Care. Harvard, October 2002.
18. Escano L, Sanz G, Lorente F, Fernando A, Ugarriza J. Logistic Regression Versus Neural Networks For Medical Data. Monograflas del Seminario Matematico Garcla de Galdeano. 2006, 33: 245–52.
19. Bryce TJ, Dewhirst MW, Floyd CE, Hars V, Brizel DM. Artificial neural network model of survival in patients treated with irradiation with and without concurrent chemotherapy for advanced carcinoma of the head and neck. International Journal of Radiation Oncology, 1998, 41: 339-45.
20. http : //www.pezeshk.us/ Warning Signs of Breast Cancer, 2012.

[Downloaded from journals.tums.ac.ir on 2026-04-10]

Iranian Journal of Epidemiology 2014; 10(3): 1-8.

Original Article

Prediction of Breast Cancer Survival by Logistic Regression and Artificial Neural Network Models

Asadabadi A[1], Bahrampour A[2], Haghdoost AA[3]

1- MSc, Department of Biostatistics & Epidemiology, School of Health, Kerman University of Medical Sciences, Iran
2- Professor, PhD, Modeling of Health Research Center, Department of Biostatistics & Epidemiology, School of Health, Kerman University of Medical Sciences, Iran
3- Associate Professor, PhD, Modeling of Health Research Center, Department of Biostatistics & Epidemiology, School of Health, Kerman University of Medical Sciences, Iran
Corresponding author: Bahrampour A, abahrampour@yahoo.com

Background and Objectives: recent years, considerable attention has been paid to statistical models for classification of medical data according to various diseases and their outcomes. Artificial neural networks have been successfully used for pattern recognition and prediction since they are not based on prior assumptions in clinical studies. This study compared two statistical models, artificial neural network and logistic regression, to predict the survival of patients with breast cancer.

Methods: Two models were applied on cancer registry data, Kerman, southeast of Iran, to predict survival. The data of 712 breast cancer patients in the age group 15 to 85 years was used in this study. The logistic regression and three-layer perceptron neural network models were compared in terms of predicting the survival. Sensitivity, specificity, prediction accuracy, and the area under ROC curve were used for comparing the two models.

Results: In this study, the sensitivity and specificity of logistic regression and artificial neural network models were (0.594, 0.70) and (0.621, 0.723), respectively. Prediction accuracy and the area under ROC curve for two models were (0.688, 0.725) and (0.70, 0.725), respectively.

Conclusion: Although there were insignificant differences in the performance of the two models for predicting the survival of the patients with breast cancer, the corresponding results of artificial neural network were more appropriate for predicting survival in such data.

Keywords: Logistic Regression, Artificial Neural Network, Survival, Brest Cancer

[Downloaded from journals.tums.ac.ir on 2026-04-10]