

## الگوی شبکه عصبی مصنوعی در تحلیل بقای بیماران سرطان معده

محمودرضا گوهری<sup>۱</sup>، پریسا مختاری<sup>۲</sup>، محمدحسین پورحسینقلی<sup>۳</sup>، اکبر بیگلریان<sup>۴\*</sup>

۱. گروه آمار زیستی، مرکز تحقیقات مدیریت بیمارستانی، دانشگاه علوم پزشکی تهران، تهران، ایران
۲. گروه آمار زیستی، دانشگاه علوم پزشکی تهران، تهران، ایران
۳. گروه آمار زیستی، مرکز تحقیقات بیماری‌های گوارش و کبد، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
۴. گروه آمار زیستی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران

نشریه پایش

سال سیزدهم شماره سوم خرداد - تیر ۱۳۹۳ صص ۲۸۵-۲۹۱

تاریخ پذیرش مقاله: ۱۳۹۲/۴/۱۹

انشر الکترونیک پیش از انتشار- ۱۸ فروردین ۹۳

### چکیده

سرطان معده در ایران در رتبه نخست شایع‌ترین سرطان برای مردان و رتبه سوم در زنان است. هدف از این مطالعه، تحلیل بقای بیماران مبتلا به سرطان معده و در ادامه پیش‌بینی مرگ این بیماران با استفاده از الگوی شبکه عصبی است. در این مطالعه طولی گذشته‌نگر تعداد ۲۴۲ بیمار که دارای تشخیص قطعی پاتولوژیک سرطان معده بودند بررسی شدند. داده‌ها توسط مرکز تحقیقات گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی در بیمارستان طالقانی تهران جمع‌آوری گردید. توابع بقا و عوامل مؤثر بر بقا به کمک روش کاپلان - مایر و الگو مخاطرات متناسب کاکس برآورد شدند. برای تحلیل به روش شبکه عصبی ابتدا داده‌ها به تصادف به دو گروه آموزش و آزمایش تقسیم شدند و همگنی زمان‌های بقای آن‌ها با آزمون رتبه لگاریتمی بررسی شد. برای انتخاب بهترین ساختار شبکه، با تغییر در مشخصه‌های شبکه الگوهای مختلف بررسی و با استفاده از سطح زیر منحنی ROC الگو مناسب انتخاب شد. برای تحلیل از نرم‌افزار SPSS نسخه ۱۹ استفاده شد. میانگین طول عمر بیماران ۲۵/۱ ماه، بقای یک‌ساله بیماران سرطان معده برابر  $0/769$  ( $0/833$ ،  $0/711$ ) و بقای سه ساله  $0/256$  ( $0/373$ ،  $0/185$ )  $CI/95$  به دست آمد. در الگوی انتخاب شده، خطای پیش‌بینی برابر  $37/7$  درصد و سطح زیر منحنی راک برابر  $0/732$  به دست آمد. پیش‌بینی صحیح برای سانسور شدن  $71/4$  درصد و پیش‌بینی صحیح مرگ بیماران نیز  $50$  درصد به دست آمد. شبکه‌ی عصبی رویکردی مناسب برای پیش‌بینی بقای بیماران مبتلا به سرطان معده است و پیش‌بینی‌های آن می‌تواند در دسته‌بندی این بیماران استفاده شود.

**کلیدواژه:** سرطان معده، شبکه عصبی مصنوعی، تحلیل بقا

\* نویسنده پاسخگو: تهران، اوین، بلوار دانشجو، پست کودکیار، دانشگاه علوم بهزیستی و توانبخشی، گروه آمار زیستی

تلفن: ۲۲۱۸۰۱۴۶

E-mail: abiglarian@gmail.com

## مقدمه

سرطان معده در ایران در رتبه نخست شایع‌ترین سرطان برای مردان و در رتبه سوم شایع‌ترین سرطان در زنان، بعد از سرطان پستان و روده بزرگ است [۴-۱]. به علت شباهت‌های علائم این بیماری با بیماری‌های دیگر معده، در بیش‌تر موارد تشخیص سرطان معده در مراحل پیشرفته‌ی بیماری صورت می‌گیرد طوری که متأسفانه در کشور ما در ۸۰ درصد موارد این‌گونه است. بدیهی است که در این زمان اکثر روش‌های درمانی تأثیر قابل ملاحظه‌ای بر بقای بیماران ندارند [۵]. الگوهای تحلیل داده‌های بقا عمدتاً الگوی‌های ناپارامتری هستند که برآورد پارامترها توسط روش‌های ناپارامتری انجام می‌شود. برآوردگرهای کاپلان مایر، نلسن آلن برای تابع بقا و تابع خطر مهم‌ترین آماره‌های مورد استفاده در تحلیل بقا هستند [۶]. همچنین رگرسیون کاکس نیز که یکی از شناخته‌شده‌ترین الگوهای رگرسیونی است به‌عنوان یک روش رگرسیونی نیمه‌پارامتری محسوب می‌گردد [۷]. ویژگی عمده این روش‌ها آن است که برای زمان بقا توزیع مشخصی در نظر گرفته نمی‌شود و به دانستن توزیع بقا برای برآورد پارامترها نیازی نیست [۸]. به کارگیری اکثر الگوها، حتی الگوهای نیمه‌پارامتری، مستلزم در نظرگیری و برقراری پذیره‌هایی است که گاهی به آسانی برقرار نیستند و لذا استفاده از الگوهای معمول آماری را محدود می‌کند؛ از جمله می‌توان به مواردی مانند تعداد بالای متغیرهای پیش‌بین، وجود اثر متقابل بین متغیرها و همبستگی بین برخی از متغیرهای مورد بررسی اشاره کرد. یکی از راه‌های برطرف کردن چنین مشکلاتی، به کارگیری الگوی شبکه‌های عصبی مصنوعی است که به‌عنوان یک روش رگرسیون غیرخطی محسوب می‌شود. مهم‌ترین ویژگی شبکه‌های عصبی توانایی آن‌ها در پیش‌بینی و دسته‌بندی اطلاعات است. در این الگوها اطلاعات نمونه‌ها به صورت فردی و تکی از ابتدا تا انتهای تحلیل نقش دارد و اطلاعات افراد به صورت تلخیص شده وارد الگو نمی‌شود. همچنین شبکه‌های عصبی بدون پیش‌داوری در مورد شکل تابع هر پارامتر، آن را ذخیره و ارزیابی می‌کند که چنین سامانه‌ای می‌تواند نتایج بیش‌تری را از عمق داده‌ها بیرون بکشد [۹]. در کاربردهای تحلیل بقا، الگوی شبکه عصبی مصنوعی نسبت به الگو کاکس دارای این مزیت است که اگر شکل توزیع نامشخص باشد یا پذیره‌های الگو برقرار نباشند و یا به‌سختی قابل پذیرش باشند، الگو شبکه عصبی مصنوعی به‌دلیل انعطاف‌پذیری می‌تواند برای تحلیل داده‌های سانسور شده استفاده

شود؛ حالت یادشده به‌خصوص وقتی که پذیره متناسب بودن مخاطرات برقرار نباشد بسیار کارا است [۱۰]. هدف از این مطالعه تحلیل بقای بیماران سرطان معده و در ادامه پیش‌بینی مرگ این بیماران با استفاده از الگو شبکه عصبی است.

## مواد و روش کار

در این مطالعه‌ی گذشته‌نگر تعداد ۲۴۲ بیمار که دارای تشخیص قطعی پاتولوژیک سرطان معده بودند وارد مطالعه شدند. پس از بررسی پرونده‌ها و ثبت اطلاعات بالینی بیمار، آخرین وضعیت سلامت بیماران به‌صورت تلفنی پیگیری و زمان بقای آن‌ها ثبت گردید. داده‌ها توسط مرکز تحقیقات گوارش و کبد دانشگاه علوم پزشکی شهید بهشتی در بیمارستان طالقانی در سال‌های ۱۳۸۵-۱۳۸۱ جمع‌آوری گردید. پس از پالایش داده‌ها تحلیل اولیه داده‌ها به کمک روش کاپلان - مایر و الگو مخاطرات متناسب کاکس انجام شد. برای تحلیل نهایی داده‌ها از الگو شبکه عصبی مصنوعی استفاده شد. شبکه‌های عصبی از لایه‌هایی شامل اجزای ساده پردازشگری به نام نورون تشکیل شده است که به صورت موازی با هم عمل می‌کنند. لایه اول، لایه ورودی است که می‌تواند پارامترهای آماری یا مؤلفه‌های حاصل از تبدیلات ریاضی روی توابع باشند. لایه دوم، لایه یا لایه‌های میانی هستند که اساس ساختار یک شبکه را تشکیل می‌دهند. کار اصلی این لایه، استخراج اطلاعات دسته‌بندی از داده‌های موجود است. لایه آخر یا لایه خروجی می‌تواند توسط یک یا چند عنصر پردازشگر که خروجی آن نشان‌دهنده دسته‌بندی نهایی است، شناخته شود. هر لایه دارای وزنی است که بیان‌گر میزان تأثیر دو نورون بر یکدیگر است. تخمین پارامترها (وزن‌ها) توسط قوانین یادگیری صورت می‌پذیرد. این الگوریتم به این صورت عمل می‌کند که در خلال مرحله آموزش، الگوهای ورودی به‌گونه‌ای به شبکه ارائه می‌شوند که هر الگوی ورودی لایه به لایه به سمت جلو حرکت می‌کند تا یک الگوی خروجی محاسبه شود. این الگوی خروجی، محاسبه و سپس با الگویی مطلوب یا دلخواه مقایسه می‌شود و در ادامه مقدار خطا تعیین می‌گردد. برای تحلیل با روش شبکه عصبی ابتدا داده‌ها به تصادف به دو مجموعه‌ی آموزشی (یادگیری) و آزمایشی (اعتبارسنجی) تقسیم شدند. برای این منظور، ۳۰ درصد از داده‌ها برای اعتبارسنجی و ۷۰ درصد از داده‌ها برای آموزش در نظر گرفته شدند. برای بررسی همگن بودن دو گروه، از آزمون رتبه لگاریتمی



جدول ۱: توزیع فراوانی مشخصه‌های بیماران مبتلا به سرطان معده

| جنسیت                         | تعداد | درصد |
|-------------------------------|-------|------|
| مرد                           | ۱۷۶   | ۷۲/۷ |
| زن                            | ۶۶    | ۲۷/۳ |
| گروه‌های سنی                  |       |      |
| کمتر از ۵۰ سال                | ۴۴    | ۲۵/۴ |
| ۵۰ سال و بالاتر               | ۱۲۹   | ۷۴/۶ |
| شاخص توده بدنی                |       |      |
| کمتر از ۲۰                    | ۶۳    | ۲۶/۰ |
| ۲۰ تا ۲۵                      | ۸۴    | ۳۴/۷ |
| ۲۵ تا ۳۰                      | ۱۴    | ۵/۸  |
| ۳۰ و بالاتر                   | ۶     | ۲/۵  |
| سابقه فامیلی                  |       |      |
| ندارد                         | ۱۸۱   | ۷۴/۸ |
| دارد                          | ۶۱    | ۲۵/۲ |
| مصرف زیاد نمک یا مصرف دخانیات |       |      |
| دارد                          | ۱۰۸   | ۴۵/۸ |
| ندارد                         | ۱۲۸   | ۵۴/۲ |
| اندازه تومور                  |       |      |
| T1                            | ۲۷    | ۱۱/۲ |
| T2                            | ۳۷    | ۱۵/۹ |
| T3                            | ۹۸    | ۴۲/۲ |
| T4                            | ۷۰    | ۳۰/۲ |
| نامعلوم                       | ۱۰    | ۴/۱  |
| تعداد غدد لنف درگیر           |       |      |
| N1                            | ۶۵    | ۲۶/۹ |
| N2                            | ۱۳۷   | ۵۶/۶ |
| N3                            | ۲۴    | ۹/۹  |
| نامعلوم                       | ۱۶    | ۶/۶  |
| مرحله بیماری                  |       |      |
| I                             | ۱۴    | ۵/۸  |
| II                            | ۲۸    | ۱۱/۶ |
| III                           | ۸۶    | ۳۵/۵ |
| IV                            | ۹۶    | ۳۹/۷ |
| نامعلوم                       | ۱۸    | ۷/۴  |

جدول ۳: مقایسه متغیرهای توضیحی در دو گروه آموزشی و آزمایشی

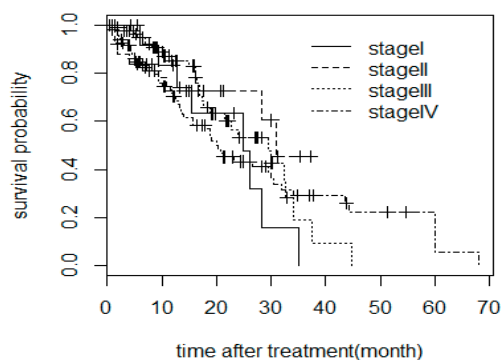
| مقدار احتمال | گروه آموزش (n = ۱۷۳)<br>تعداد (درصد) | گروه پیش‌بینی (n = ۶۹)<br>تعداد (درصد) | جنسیت           |
|--------------|--------------------------------------|----------------------------------------|-----------------|
| ۰/۲۹۳        | ۱۲۸ (۷۴/۰)                           | ۴۸ (۶۹/۶)                              | مرد             |
| ۰/۱۷۸        | ۴۵ (۲۶/۰)                            | ۲۱ (۳۰/۴)                              | زن              |
| ۰/۱۷۸        | ۴۴ (۲۵/۴)                            | ۱۳ (۱۸/۸)                              | گروه‌های سنی    |
| ۰/۱۷۸        | ۱۲۹ (۷۴/۶)                           | ۵۶ (۸۱/۲)                              | کمتر از ۵۰ سال  |
| ۰/۱۷۸        | ۱۲۹ (۷۴/۶)                           | ۵۶ (۸۱/۲)                              | ۵۰ سال و بالاتر |
| ۰/۱۷۸        | ۶۲ (۳۵/۸)                            | ۲۷ (۳۹/۱)                              | شاخص توده بدنی  |
| ۰/۱۷۸        | ۸۹ (۵۱/۴)                            | ۳۳ (۴۷/۸)                              | کمتر از ۲۰      |
| ۰/۱۷۸        | ۱۶ (۹/۲)                             | ۸ (۱۱/۶)                               | ۲۰ تا ۲۵        |
| ۰/۱۷۸        | ۶ (۳/۵)                              | ۱ (۱/۴)                                | ۲۵ تا ۳۰        |
| ۰/۱۷۸        | ۱۳۳ (۷۶/۹)                           | ۴۸ (۶۹/۶)                              | سابقه فامیلی    |
| ۰/۱۷۸        | ۴۰ (۲۳/۱)                            | ۲۱ (۳۰/۴)                              | ندارد           |
| ۰/۱۷۸        | ۷۹ (۴۶/۷)                            | ۲۹ (۴۳/۳)                              | دارد            |
| ۰/۱۷۸        | ۹۰ (۵۳/۳)                            | ۳۸ (۵۵/۱)                              | عادات پرخطر     |
| ۰/۱۷۸        | ۷۹ (۴۶/۷)                            | ۲۹ (۴۳/۳)                              | دارد            |
| ۰/۱۷۸        | ۹۰ (۵۳/۳)                            | ۳۸ (۵۵/۱)                              | ندارد           |

جدول ۴: پارامترهای برازش الگو در الگو شبکه عصبی برای داده‌های سرطان معده

| معماری (خروجی لمبانی / ورودی) | نرخ یادگیری | اندازه حرکت | مجموع مربعات خطا | سطح زیر منحنی راک | درصد پیش‌بینی نادرست |
|-------------------------------|-------------|-------------|------------------|-------------------|----------------------|
| ۱ / ۱۳ / ۳ / ۲                | ۰/۳         | ۰/۹۵        | ۱۱/۳۷            | ۰/۷۱۴             | ۳۵/۲                 |
| ۲ / ۱۳ / ۴ / ۲                | ۰/۳         | ۰/۹۵        | ۱۴/۳۳            | ۰/۶۶۹             | ۳۷/۷                 |
| ۳ / ۱۳ / ۵ / ۲                | ۰/۳         | ۰/۸         | ۱۴/۵۲            | ۰/۶۷۰             | ۴۲/۶                 |
| ۴ / ۱۳ / ۶ / ۲                | ۰/۲         | ۰/۸         | ۱۴/۲۸            | ۰/۷۳۲             | ۳۷/۷                 |
| ۵ / ۱۳ / ۷ / ۲                | ۰/۲         | ۰/۹۵        | ۱۴/۰۵            | ۰/۷۲۳             | ۴۱/۰                 |
| ۶ / ۱۳ / ۸ / ۲                | ۰/۴         | ۰/۹۵        | ۱۴/۳۶            | ۰/۶۹۱             | ۳۷/۷                 |
| ۷ / ۱۳ / ۹ / ۲                | ۰/۲         | ۰/۸۵        | ۱۴/۰۸            | ۰/۶۶۳             | ۳۲/۸                 |
| ۸ / ۱۳ / ۱۰ / ۲               | ۰/۳         | ۰/۸         | ۱۴/۷۹            | ۰/۶۸۶             | ۳۷/۷                 |

\* معماری انتخاب شده

شکل ۱. منحنی بقای بیماران سرطان معده به تفکیک مرحله بیماری با استفاده از کاپلان-مایر



جدول ۲: آزمون رتبه لگاریتم برای مقایسه بقای بیماران مبتلا به سرطان معده بر حسب ویژگی‌های بیماران در دو گروه آموزشی و پیش‌بینی

| ویژگی‌های بیماران                           | آماره آزمون | درجه آزادی | مقدار احتمال * |
|---------------------------------------------|-------------|------------|----------------|
| جنسیت                                       | ۰/۲         | ۱          | ۰/۶۲۲          |
| گروه‌های سنی                                | ۴           | ۱          | ۰/۰۴۵          |
| شاخص توده بدنی                              | ۱۴/۹        | ۳          | ۰/۰۰۱          |
| سابقه فامیلی                                | ۰/۴۲        | ۱          | ۰/۵۴۱          |
| عادات پرخطر                                 | ۰/۴۴        | ۱          | ۰/۵۲۷          |
| (مانند مصرف زیاد نمک یا مصرف دخانیات و ...) |             |            |                |
| اندازه تومور                                | ۸           | ۳          | ۰/۰۳۷          |
| تعداد لنف درگیر                             | ۲/۲         | ۲          | ۰/۳۳۳          |

\* آزمون رتبه لگاریتمی

## بحث و نتیجه گیری

این مطالعه بر روی داده‌های بقای بیماران مبتلا به سرطان معده انجام شد که در آن ضمن بررسی بقای بیماران از الگو شبکه‌ی عصبی مصنوعی برای پیش‌بینی استفاده شد. در دو دهه‌ی اخیر رهیافت‌های پیچیده مانند شبکه‌های عصبی مصنوعی برای پیش‌بینی پیامدهای مختلف در حوزه‌ی علوم پزشکی و بهداشت از جمله سرطان، به‌کار گرفته شد و مطالعات متعددی در این زمینه انجام شده است. امیری و همکاران، روش‌های کاپلان - مایر، مخاطرات متناسب کاکس و شبکه‌ی عصبی مصنوعی سلسله‌مراتبی را برای ارزیابی بقای بیماران مبتلا به سرطان معده مورد استفاده قرار دادند و اختلاف معنی‌دار در روند تغییرات احتمالات بقا مشاهده نکردند [۱۱]. بیگلریان و همکاران برای پیش‌بینی بقای پس از جراحی بیماران مبتلا به سرطان معده، از الگو شبکه عصبی و رگرسیون کاکس استفاده نموده و وضعیت نهایی بیماران را در شبکه عصبی صحیح‌تر از پیش‌بینی رگرسیون کاکس معرفی نمودند [۱]. چین و همکاران، روش‌های رگرسیون لجستیک، درخت تصمیم و شبکه‌ی عصبی را برای پیش‌بینی عوارض پس از جراحی بیماران مبتلا به سرطان معده استفاده کردند. نتایج مطالعه‌ی این پژوهش‌گران نشان داد که شبکه‌ی عصبی مصنوعی در مقایسه با رگرسیون لجستیک و درخت تصمیم رهیافت بهتری برای پیش‌بینی عوارض پس از جراحی بیماران مبتلا به سرطان معده است [۱۲]. بلشویلر و همکاران با به‌کارگیری شبکه‌ی عصبی مصنوعی صحت پیش‌بینی ۷۹ درصدی را برای بیماران مبتلا به سرطان معده برآورد نمودند [۱۳]. صرف‌نظر از اجرای متغیر شبکه‌ی عصبی، معماری به‌کار گرفته شده برای شبکه نیز در پیش‌بینی تأثیر دارد [۱۴]. در مطالعه حاضر میزان بقای یک و سه ساله این بیماران به ترتیب برابر ۷۶/۹ درصد و ۲۵/۶ درصد به‌دست آمد که از بقای یک‌ساله مطالعات در شیراز که بقای یک‌ساله ۵۴ درصد [۱۵]، در مازندران ۳۳ درصد [۱۶] و در اردبیل ۲۶ درصد [۱۷] گزارش شده، بالاتر است.

در تهران بقای سه‌ساله بیماران سرطان معده ۳۱ درصد برآورد شده است که از مطالعه حاضر بالاتر است [۱۸]. بقای پنج ساله در این مطالعه برابر ۴ درصد برآورد گردید در حالی که میزان بقای ۵ ساله در کشورهای پیشرفته از جمله آمریکا ۳۷ درصد، سوییس ۲۲ درصد، فرانسه ۳۰ درصد، چین ۳۰ درصد، ژاپن ۳۵ درصد گزارش شده است. در اکثر کشورها بقای پنج‌ساله بین ۱۰ تا ۳۰ درصد است [۱۹-۲۳]. یک علت برآورد پایین بقای پنج‌ساله در این مطالعه کم بودن حجم نمونه و ضعف برآوردگر کاپلان مایر در این موارد است که هر یک از پیشامدها در زمان‌های انتهایی وزن بالاتری از مرگ‌های زمان‌های ابتدایی می‌گیرد و برآورد صحیحی را در زمان‌های انتهایی ارائه نمی‌دهد. در این پژوهش، نتایج الگو شبکه‌ی عصبی مصنوعی پرسپترون با یک لایه‌ی پنهان در پیش‌بینی بقای بیماران مبتلا به سرطان معده، مورد استفاده قرار گرفت. یافته‌های پژوهش حاضر نشان داد که شبکه‌ی عصبی رهیافت مناسبی در پیش‌بینی بقای بیماران مبتلا به سرطان معده است.

## سهم نویسندگان

محمودرضا گوهری: همکاری در اجرا، جمع‌بندی و نگارش مقاله  
پریسا مختاری: بررسی پیشینه و جمع‌آوری مطالب مرتبط با بحث  
محمدحسین پورحسینقلی: در اختیارگزاری و پالایش داده‌ها  
اکبر بیگلریان: انجام تحلیل، جمع‌بندی و ویراستاری مقاله

## تشکر و قدردانی

این مقاله حاصل از طرح تحقیقاتی به شماره ۹۱-۰۳-۱۳۶-۱۹۱۸۸ مصوب دانشگاه علوم پزشکی تهران است که با حمایت دانشگاه علوم پزشکی تهران در سال ۱۳۹۱ اجرا شده است. از کلیه همکارانی که در گردآوری داده‌ها همکاری داشته‌اند سپاس‌گزاری و قدردانی می‌گردد.

## منابع

1. Biglarian A, Hajizadeh E, Kazemnejad A, Zali M. Postoperative Survival Prediction in Patients with Gastric Cancer. *Medical Daneshvar* 2009;16:55-62 [Persian]
2. Sajadi A, Nouraei M, Mohagheghi MA. Cancer Occurrence in Iran in 2002, an International Perspective. *Asian Pacific Journal of Cancer Prevention* 2005;6:359-63
3. Babaei M, Mousavi S, Toussy J. Cancer occurrence in old age: results of a population-based cancer registry in Semnan, Iran. *Asian Pacific Journal of Cancer Prevention* 2006; 7: 191-94
4. Ansari R, Mahdavinia M, Sadjadi A, Nouraei M, Kamangar F, Bishehsari F, et al. Incidence and age distribution of colorectal cancer in Iran: results of a population-based cancer registry. *Cancer Letter* 2006; 240: 143-47
5. Malekzadeh R, Derakhshan M, Malekzadeh Z. Gastric Cancer in Iran: epidemiology and risk factors. *Archives of Iranian Medicine* 2009;12:576-83
6. Kleinbaum DG, Klein M. *Survival analysis, a self-learning text*. 1<sup>st</sup> Edition, Springer: New York 2005
7. Therneau tM, Grambsch. PM. *Modeling survival data:extending the Cox model*. 1<sup>st</sup> Edition, Springer: New York 2000
8. Collett David. *Modelling survival data in medical research*. 2<sup>st</sup> Edition, Chapman & Hall: London, 2003
9. Warner B, Misra M. *Understanding Neural Networks as Statistical tools*. *The American Statistician* 1996;50:284-93
10. Ohno-Machado L. A comparison of Cox proportional hazards and artificial neural network models for medical prognosis. *Computers in Biology & Medicine* 1997; 27: 55-65
11. Amiri Z, Mohammad K, Mahmoudi M, Zeraati H, Fotouhi A. Assessment of gastric cancer survival: using an artificial hierarchical neural network. *Pakistan Journal of Biological Sciences* 2008;11:1076-84
12. Chien C, Lee Y, Ma T, Lin Y, Lee T, Wang W, et al. The application of artificial neural networks and decision tree model in predicting post-operative complication for gastric cancer patients. *Hepatogastroenterology* 2008; 55:1140-5
13. Bollschweiler E, Mönig S, Hensler K, Baldus S, Maruyama K, Hölscher A. Artificial neural network for prediction of lymph node metastases in gastric cancer: a phase II diagnostic study. *Annals of Surgical Oncology* 2004;11:506-11
14. Azen S, Xiang A, Lapuerta P, Ryutov A, Buckley J. Comparison of predictive accuracy of neural network methods and cox regression for censored survival data. *Computational Statistics & Data Analysis* 2000; 34: 243-57
15. Moghimi Dehkordi B, Rajaeefard A, Tabatabaee H, Zeighami B, Dafaee A, Tabeie Z. Modeling survival analysis in gastric cancer patients using proportional hazards model of Cox. *Iranian Journal of Epidemiology* 2007;3:24-19
16. Esmaielee H. Comparing survival rate and its characteristics of esophagus and stomach cancer in Mazandran province [MSc dissertation]. Tehran: Tarbiat Modarres University, School of Medical Sciences; 1992 [Persian]
17. Yazdanbod A, Samadi F, Malekzadeh R, Babaei M, Iranparvar M, Azami A. Four-year survival rate of upper gastrointestinal cancer patients in Ardabil Province. *Ardabil University of Medical Sciences Journal* 2005;5:184-0 [Persian]
18. Zeraati H, Mahmoudi M, Mohammad M, Kazemnejad A, Mohagheghi M. Postoperative survival in gastric cancer patients and its related factors. *Journal of School of Public Health and Institute of Public Health Research* 2005; 3:21-30 [Persian]
19. Ahmadloo N, Omidvari S, Mosalaei A. Results of post-operative radiotherapy in patients with high risk Gastric cancer. *Journal of Medical Research* 2003;1:49-3 [Persian]
20. Hosseini S, Mousavian AS, Moghimi M, Fallah R. Delay in diagnosis and treatment of gastric cancer from the beginning of symptoms to surgery. *Medical Journal of Tabriz University of Medical Sciences* 2009;29:39-42 [Persian]
21. Coleman M, Gatta G, Verdecchia A. EURO CARE Working Group: EURO CARE-3 summary: cancer survival in Europe at the end of the 20th century. *Annals of Oncology* 2003;14 128-49
22. Deasi A, Pareek M, Nightingale P, Fielding J. Improving outcomes in gastric cancer over 20 years. *Gastric Cancer* 2004;7:196-201; discussion 201-203
23. Landry J, Tepper J, Wood W. Patterns of failure following curative resection of gastric cancer carcinoma. *International Journal of Radiation Oncology Biology Physics* 1990;19:1357-62

## ABSTRACT

### Artificial Neural Network in survival analysis of gastric cancer patients

Mahmood Reza Gohari<sup>1</sup>, Parisa Mokhtari<sup>2</sup>, Mohammad Amin Pourhoseingholi<sup>3</sup>, Akbar Biglarian<sup>4\*</sup>

1. Department of Biostatistics, Hospital management Research Center, Tehran University of Medical Sciences (TUMS), Tehran, Iran

2. Department of Biostatistics, Tehran University of Medical Sciences (TUMS), Tehran, Iran

3. Department of Health System Research, Research Center of Gastroenterology and Liver Disease, Shahid Beheshti university of Medical Sciences, Tehran, Iran

4. Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences (USWRS), Tehran, Iran

Payesh 2014; 13: 285-291

Accepted for publication: 10 Julay 2013

[EPub a head of print-7 April 2014]

**Objective (s):** Gastric cancer (GC) is the most prevalent cancer in men and third for women. The aim of this study was to analyze the survival of GC patients and then the prediction of death in these patients using a neural network model.

**Methods:** This was a historical cohort study with data selected from 232 registered gastric cancer patients who underwent surgery between 2002 and 2007 at a referral center for gastrointestinal cancers, in Tehran, Iran. Prognostic factors of survival using Kaplan-Meier and Cox proportional hazards models were estimated. To analyze the data using the neural network, the data were divided randomly into two groups, training and testing sets, and homogeneous survival times were tested using log rank test. To select the best network structure, the characteristics of network models modified and the best model was selected using ROC analysis. The analysis was performed using SPSS version 19.

**Results:** The median survival time of GC patients was 25.1 months. One and three year survival times of GC patients were 0.769 (CI 95%: 0.711, 0.833) and 0.256 (CI 95%: 0.185, 0.373), respectively. The prediction error was 37.7% and the area under the ROC was 0.732 at final model. True prediction for censorship and death were 71.4% and 50%, respectively.

**Conclusion:** Neural network is an appropriate approach for predicting the survival of GC patients and its predictions can be used to classify these patients.

**Key Words:** Gastric Cancer, Neural Network, Survival Analysis

\* Corresponding author: University of Social Welfare and Rehabilitation Sciences (USWRS), Tehran, Iran

Tel: 22180146

E-mail: abiglarian@gmail.com