

## بررسی عوامل مرتبط با سرطان معده با رویکرد شبکه‌ی عصبی و رگرسیون لجستیک چندگانه: مطالعه‌ی مورد-شاهدی

آرمین نقی‌پور<sup>۱</sup>، میثم حسینی<sup>۲</sup>، ارسلان رحمانی<sup>۳</sup>، زهرا نقیبی‌فر<sup>۴</sup>

### مقاله پژوهشی

### چکیده

**مقدمه:** سرطان معده، پنجمین سرطان شایع و سومین عامل مرگ در جهان است. در این مطالعه عوامل مرتبط با سرطان معده با استفاده از دو مدل رگرسیون لجستیک و شبکه‌ی عصبی بررسی گردید.

**روش‌ها:** در این مطالعه‌ی مورد-شاهدی، در مجموع ۱۱۷۰ نفر وارد مطالعه شدند (۳۹۰ مورد - ۷۸۰ شاهد). ابزار جمع‌آوری اطلاعات بر اساس چک‌لیست محقق ساخته بود. نمونه‌ها به روش نمونه‌گیری در دسترس انتخاب شدند و به روش مصاحبه‌ی حضوری و تلفنی اطلاعات آن‌ها جمع‌آوری شد. قدرت برازش در مدل رگرسیون لجستیک و شبکه‌ی عصبی با استفاده از منحنی مشخصه عملکرد گیرنده (ROC) مقایسه گردید. با معرفی مدل برتر، عوامل مرتبط و معنی‌دار با ابتلا به سرطان معده گزارش شد.

**یافته‌ها:** نتایج نشان داد که دقت، حساسیت و ویژگی در شبکه عصبی به ترتیب ۰/۹۶۴، ۰/۹۳۷ و ۰/۸۱۹ درصد و در مدل رگرسیون لجستیک به ترتیب ۰/۹۵۹، ۰/۹۱۱ و ۰/۸۴۴ درصد گزارش شد. مدل شبکه‌ی عصبی نشان داد، متغیرهای سن (۰/۶۴۶)، مصرف میوه (۰/۷۱۳)، سابقه‌ی مصرف خودسرانه‌ی دارو (۰/۶۵۲)، سابقه‌ی زخم معده (۰/۷۳۴)، سابقه‌ی خانوادگی سرطان (۰/۸۵۲) و سابقه‌ی خانوادگی سرطان معده (۰/۸۳۶) از نظر آماری با بروز سرطان معده ارتباط دارند.

**نتیجه‌گیری:** با توجه به اینکه در مطالعه‌ی حاضر برازش شبکه‌ی عصبی نسبت به رگرسیون لجستیک برتر بوده و نیاز به فرضیات خاصی ندارد، لذا به محققین پیشنهاد می‌شود که مدل شبکه‌ی عصبی را می‌توان بر رگرسیون لجستیک ترجیح داد.

**واژگان کلیدی:** سرطان معده؛ شبکه‌ی عصبی؛ پرسپترون چند لایه؛ رگرسیون لجستیک

**ارجاع:** نقی‌پور آرمین، حسینی میثم، رحمانی ارسلان، نقیبی‌فر زهرا. بررسی عوامل مرتبط با سرطان معده با رویکرد شبکه‌ی عصبی و رگرسیون

لجستیک چندگانه: مطالعه‌ی مورد-شاهدی. مجله دانشکده پزشکی اصفهان ۱۴۰۱؛ ۴۰ (۶۹۳): ۸۸۹-۸۸۰

### مقدمه

ضعیف، انجام درمان‌های تسکینی نظیر مصرف خودسرانه‌ی دارو، عدم مراجعه جهت درمان دل‌دردهای مکرر و عدم شناخت عوامل پیشگویی‌کننده‌ی سرطان معده سبب افزایش بروز آن و طول عمر کوتاه‌تر بیماران شده است (۳). با توجه به مطالعات انجام شده از لحاظ جغرافیایی، سرطان معده در کشورهای آسیایی و به ویژه در ایران شایع می‌باشد. به طوری که در بین مردان رتبه‌ی اول و در بین زنان بعد از سرطان پستان و روده‌ی بزرگ رتبه‌ی سوم را دارد (۱). بنابراین با توجه به آمار نگران‌کننده‌ی بروز سرطان معده، طراحی و

سرطان معده، پنجمین سرطان شایع و سومین عامل مرگ میان افراد مبتلا به سرطان در جهان است. همچنین، یکی از چالش‌ها در نظام مدیریت و سیاست‌گذاری پزشکی، افزایش بروز سرطان معده طی سال‌های اخیر می‌باشد (۱، ۲). بر اساس گزارش سازمان بهداشت جهانی (World Health Organization) WHO در سال ۲۰۱۸، بروز سرطان معده ۱/۸٪ و نرخ مرگ آن ۹/۸ میلیون نفر بود، که نسبت به سالیان قبل، بروز آن افزایش یافته بود (۲). پیش‌آگاهی

۱- کارشناس ارشد آمار زیستی، مرکز توسعه‌ی تحقیقات بالینی، بیمارستان امام رضا (ع)، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

۲- استادیار، گروه ریاضی، پردیس بیجار، دانشگاه کردستان، سنندج، ایران

۳- استادیار، گروه ریاضی، دانشکده‌ی علوم پایه، دانشگاه کردستان، سنندج، ایران

۴- کارشناس ارشد اپیدمیولوژی، مرکز تحقیقات بیماری‌های عفونی، بیمارستان امام رضا (ع)، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

نویسنده‌ی مسؤول: میثم حسینی؛ استادیار، گروه ریاضی، پردیس بیجار، دانشگاه کردستان، سنندج، ایران

ارائه‌ی یک سیستم یکنواخت جهت بررسی عوامل مرتبط با آن دارای اهمیت ویژه‌ای است (۲).

مطالعات بسیاری جهت شناسایی عوامل مرتبط با سرطان معده در جهان انجام شده است. نتایج این مطالعات نشان دادند که عواملی دموگرافیک (جنس، سن، تحصیلات، محل سکونت و شغل)، عواملی نظیر سابقه‌ی خانوادگی سرطان، سابقه‌ی ابتلا به بیماری زمین‌های، سیگار، رژیم غذایی، مصرف الکل، عادت‌های تغذیه‌ای، وجود هلیکوباکتر، گروه خونی، مصرف میوه، مصرف سیر، مصرف سبزیجات، مصرف گوشت قرمز و زخم معده یا اثنی عشر ارتباط قوی با بروز سرطان معده دارند (۳، ۴).

در اکثر مطالعات انجام شده جهت تعیین ارتباط و شناسایی عوامل مرتبط با سرطان معده از مدل‌های آماری نظیر رگرسیون لجستیک و آزمون‌های همبستگی استفاده شده است. مدل‌های رگرسیون لجستیک و آزمون‌های همبستگی مستلزم برقراری فرضیه‌ها و محدودیت‌هایی نظیر وجود همبستگی و اثر متقابل بین متغیرها هستند. در بیشتر مطالعات تجربی برقراری فرضیه‌ها معمولاً رخ نمی‌دهد و همبستگی و اثر متقابل چندگانه بین متغیرها در نظر گرفته نمی‌شود، زیرا تشخیص همبستگی بیش از دو متغیر و اثر متقابل بین متغیرها امکان‌پذیر نمی‌باشد. از طرفی، اگر تعداد متغیرها یا حجم داده‌ها زیاد باشد محدودیت‌های این روش‌ها دوچندان می‌شود. شبکه‌ی عصبی به عنوان راه چاره‌ای برای مواجهه با چنین محدودیت‌های طراحی شده است (۵، ۶).

اگرچه آزمون همبستگی‌های پیشرفته یا اثرات متقابل با استفاده از مدل رگرسیون لجستیک در این مطالعه امکان‌پذیر است، نویسنندگان به دنبال این هدف بودند که دقت، حساسیت و ویژگی این دو مدل آماری را مقایسه نمایند.

### روش‌ها

مطالعه‌ی حاضر به روش مورد-شاهدی در فاصله‌ی زمانی فروردین ماه ۱۳۹۸ تا اواخر اسفند ماه ۱۴۰۰ در بیمارستان امام رضا(ع)، شهر کرمانشاه و کلینیک تشخیصی-درمانی مهدیه در شهر همدان، انجام شد. جامعه‌ی مورد مطالعه در گروه مورد، افراد مبتلا به سرطان معده بودند که به روش تشخیص قطعی پاتولوژیک تأیید شدند. حداقل حجم نمونه‌ی مورد نیاز با خطای نوع اول ۰/۰۵، توان آزمون ۸۰ درصد و نسبت شانس ابتلا به سرطان معده ۱/۵۰ در فرمول رگرسیون لجستیک دو طرفه توسط نرم‌افزار G\*power نسخه‌ی 3.1.9.2 تعداد ۳۰۴ نفر برآورد شد. جهت مقایسه و افزایش توان استنباط‌های آماری به ازای یک نفر در گروه مورد، دو نفر شاهد در نظر گرفته شد. گروه شاهد از بین مراجعین بیمارستان امام رضا(ع) و

کلینیک مهدیه که به سرطان معده مبتلا نبودند، انتخاب شدند. ابزار جمع‌آوری اطلاعات، چک‌لیست محقق‌ساخته بود. چک‌لیست شامل سؤالاتی نظیر اطلاعات دموگرافیک سن (سال)، جنس (۰ = زن، ۱ = مرد)، قد (سانتی‌متر)، وزن (کیلوگرم)، تحصیلات (۰ = بیسواد، ۱ = ابتدایی یا سیکل، ۲ = دیپلم، ۳ = دانشگاهی)، محل سکونت (۰ = روستایی، ۱ = شهری)، وضعیت تأهل (۰ = مجرد، ۱ = متأهل)، شغل (۰ = کشاورز، ۱ = کارمند، ۲ = بیکار، بازنشسته)، عادت‌های غذایی (مصرف چای (تعداد فنجان در روز، مصرف الکل (۰ = مصرف نکرده، ۱ = طی ۶ ماه مصرف نکرده، ۲ = حداقل یکبار در طی ۶ ماه مصرف کرده)، مصرف گوشت قرمز (۰ = سه بار یا کمتر از سه بار در هفته، ۱ = چهار بار یا بیشتر از چهار بار در هفته)، علاقه به غذای تند و فلفلی (۰ = خیر، ۱ = بلی)، مصرف میوه (۰ = سه بار یا کمتر از سه بار در هفته، ۱ = چهار بار یا بیشتر از چهار بار در هفته)، تعداد دفعات مصرف سیر (۰ = سه بار یا کمتر از سه بار در هفته، ۱ = چهار بار یا بیشتر از چهار بار در هفته)، مصرف خودسرانه‌ی دارو (۰ = خیر، ۱ = بلی) و مصرف سیگار (تعداد نخ سیگار در روز) و اطلاعات سابقه‌ی بیماری (مرحله‌ی بیماری (۰ = مرحله‌ی ۱، ۱ = مرحله‌ی ۲، ۲ = مرحله‌ی ۳)، گروه خونی (A = ۰، AB = ۱، B = ۲، O = ۳)، سابقه‌ی ریفلاکس معده (۰ = خیر، ۱ = بلی)، سابقه‌ی خانوادگی به انواع سرطان‌ها (۰ = خیر، ۱ = بلی) و سابقه‌ی خانوادگی به سرطان معده (۰ = خیر، ۱ = بلی)، سابقه‌ی زخم معده (۰ = خیر، ۱ = بلی)، سابقه‌ی زخم دوازده، نوع سرطان معده (۰ = آندونکارسینوم، ۱ = تومور استرومایی دستگاه گوارش)، سابقه‌ی ابتلا به هلیکوباکتر (۰ = خیر، ۱ = بلی)) بود. پژوهشگر جهت جمع‌آوری اطلاعات به پرونده‌ی بیماران مراجعه و اطلاعات ثبت شده (شامل شماره تماس و آدرس محل سکونت بیماران) از پرونده‌ی آن‌ها استخراج گردید. سپس محقق با انتخاب بیماران به روش نمونه‌گیری آسان با مراجعه‌ی حضوری به درب منازل یا بیمارستان تحت معالجه‌ی آن‌ها، اهداف مطالعه را بیان نمود و برای افرادی که مراجعه‌ی حضوری مقدور نبود به صورت تماس تلفنی، اطلاعات آن‌ها را جمع‌آوری کرد. جهت جمع‌آوری اطلاعات در گروه شاهد، چک‌لیست در اختیار بیماران بستری در مراکز درمانی گروه مورد قرار گرفت. چک‌لیست در گروه شاهد با مصاحبه‌ی حضوری توسط محقق تکمیل شد. معیار ورود به مطالعه داشتن حداقل سن ۱۸ سال و عدم ابتلا به بیماری آلزایمر بود. از سوی دیگر، افرادی که بیش از دو سال به سرطان معده مبتلا شده بودند؛ یا به تماس تلفنی پاسخ ندادند یا پرسش‌نامه را تکمیل نکردند از مطالعه خارج شدند.

تجزیه و تحلیل داده‌ها: در مطالعه‌ی حاضر پس از جمع‌آوری و

لذا از تابع انتقال رگرسیون لجستیک به عنوان فعال‌سازی استفاده گردید. قدرت مدل رگرسیون لجستیک و شبکه‌ی عصبی با مساحت زیر منحنی مشخصه عملکرد گیرنده (AUROC)، حساسیت و ویژگی سنجیده شد. تجزیه و تحلیل داده‌ها با نرم‌افزار SPSS نسخه‌ی ۲۵ (version 25, IBM Corporation, Armonk, NY) انجام شد.

### یافته‌ها

مطالعه‌ی حاضر روی ۱۱۷۰ نفر بررسی و نتایج آمار توصیفی و تحلیلی مربوط به آن گزارش شد. ۷۶۰ نفر (۶۵٪) از افراد مورد بررسی مرد و (۳۵٪) ۴۱۰ نفر زن بودند. (۳۴٪) ۲۶۵ نفر از مردان و (۳۰٪) ۱۲۵ نفر از زنان مبتلا به سرطان معده بودند. میانگین کل سن افراد مورد بررسی  $5/9 \pm 49/5$  سال و میانگین سن در گروه مورد و گروه شاهد به ترتیب  $4/9 \pm 51/7$  و  $5/9 \pm 48/4$  سال بود. از کل افراد مورد مطالعه، (۶۶٪) ۷۸۰ مورد به مصرف غذای داغ علاقه داشتند که (۳۹٪) ۳۰۸ نفر آن‌ها در گروه مورد قرار گرفتند. اطلاعات تکمیلی در جدول ۱ گزارش شده است.

پالایش داده‌ها، جهت تعیین عوامل مرتبط با سرطان معده از مدل ساده و چندگانه‌ی رگرسیون لجستیک استفاده شد. روش‌های کلاسیک نظیر رگرسیون لجستیک در مواردی که تعداد متغیرهای پیش‌بینی و اثرات متقابل بین آن‌ها زیاد باشد و یا در صورتی که همبستگی میان متغیرهای پیشگو شدید باشد، ضعیف‌تر از رویکردهای ناپارامتری عمل می‌کنند (۷، ۸). شبکه‌ی عصبی به عنوان روش رگرسیون غیرخطی اطلاعات تک به تک افراد در تمام مراحل اجرا بدون در نظر گرفتن شکل تابعی آن تحلیل در نظر می‌گیرد. اجزاء تشکیلی‌دهنده‌ی شبکه‌ی عصبی، نورون‌ها هستند که به صورت موازی عمل می‌کنند. جهت اطلاعات بیشتر درباره‌ی شبکه‌ی عصبی می‌توان به مطالعات انجام شده مراجعه نمود (۷-۹). جهت اجرا و طراحی مدل شبکه‌ی عصبی، داده‌ها به دو زیر مجموعه تقسیم شده‌اند که به صورت زیرمجموعه‌ی آموزش ۸۰٪ درصد و زیرمجموعه‌ی آزمون ۲۰٪ درصد داده‌ها می‌باشند. همچنین جهت تعیین عوامل مهم، شبکه‌ی سه لایه‌ای و هشت گره پرسپترون با ۵۰۰۰۰ نمونه از زنجیره مارکف در نظر گرفته شد. از سوی دیگر چون متغیر هدف دو حالت (گروه مورد-گروه شاهد) بود

جدول ۱. گزارش توصیفی و مقایسه‌ی عوامل خطر در دو گروه مورد و شاهد (n = ۱۱۷۰)

متغیر	سطح	کل تعداد (درصد)	گروه مورد ۳۹۰ (۳۳٪)	گروه شاهد ۷۸۰ (۶۶٪)	نتیجه‌ی آزمون آماری مقدار آماره P
اطلاعات دموگرافیک					
سن (سال)	انحراف معیار $\pm$ میانگین	۴۹/۵ $\pm$ ۵/۹	۵۱/۷ $\pm$ ۴/۹	۴۸/۴ $\pm$ ۵/۹	۰/۰۰۱
جنس	زن	۴۱۰ (۳۵٪)	۱۲۵ (۳۰٪)	۲۸۵ (۶۹٪)	۰/۱۲۹
	مرد	۷۶۰ (۶۵٪)	۲۶۵ (۳۴٪)	۴۹۵ (۶۵٪)	
شاخص توده بدن (BMI)	< ۲۴	۵۲۵ (۴۴٪)	۱۸۵ (۳۵٪)	۳۴۰ (۶۴٪)	۰/۴۵۳
	۱۴+۲۷	۵۲۳ (۴۴٪)	۱۶۷ (۳۱٪)	۳۵۶ (۶۸٪)	
	> ۲۷	۱۲۲ (۱۰٪)	۳۸ (۳۱٪)	۸۴ (۶۸٪)	
تحصیلات	بیسواد	۴۶۲ (۴۴٪)	۱۵۸ (۳۴٪)	۳۰۴ (۶۵٪)	۰/۸۰۲
	ابتدایی - سیکل	۳۲۵ (۳۱٪)	۱۰۳ (۳۱٪)	۲۲۲ (۶۸٪)	
	دیپلم	۱۷۳ (۱۶٪)	۵۴ (۳۱٪)	۱۱۹ (۶۸٪)	
	دانشگاهی	۸۸ (۸٪)	۳۱ (۳۵٪)	۵۷ (۶۴٪)	
محل سکونت	روستایی	۱۶۴ (۱۴٪)	۶۵ (۳۹٪)	۹۹ (۶۰٪)	۰/۰۶۵
	شهری	۱۰۰۶ (۸۶٪)	۳۲۵ (۳۲٪)	۶۸۱ (۶۷٪)	
وضعیت تأهل	مجرد	۱۹۶ (۱۶٪)	۶۴ (۳۲٪)	۱۳۲ (۶۷٪)	۰/۸۲۵
	متاهل	۹۷۴ (۸۳٪)	۳۲۶ (۳۳٪)	۶۴۸ (۶۶٪)	
شغل	کشاورز	۲۹۵ (۲۵٪)	۶۲ (۲۱٪)	۲۳۳ (۷۹٪)	۰/۰۰۱
	کارمند	۳۲۰ (۲۷٪)	۱۲۵ (۳۹٪)	۱۹۵ (۶۰٪)	
	بیکار	۲۸۶ (۲۴٪)	۱۲۳ (۴۳٪)	۱۶۳ (۵۷٪)	
	بازنشسته	۲۶۹ (۲۳٪)	۸۰ (۲۹٪)	۱۸۹ (۷۰٪)	
الگوی مصرف مواد غذایی					
تعداد فنجان چای در روز	انحراف معیار $\pm$ میانگین	۳/۹ $\pm$ ۲/۳	۵/۳ $\pm$ ۱/۹	۳/۲ $\pm$ ۲/۳	۰/۰۰۱
تعداد دفعات مصرف سیر	سه بار و کمتر در هفته	۵۷۴ (۴۹٪)	۲۰۴ (۳۵٪)	۳۷۰ (۶۴٪)	۰/۱۱۶
	۴ بار و بیشتر در هفته	۵۹۶ (۵۰٪)	۱۸۶ (۳۱٪)	۴۱۰ (۶۸٪)	

متغیر	سطح	کل تعداد (درصد)	گروه مورد ۳۹۰ (۳۳/۳)	گروه شاهد ۷۸۰ (۶۶/۷)	نتیجه‌ی آزمون آماری مقدار آماره P
میوه	سه بار و کمتر در هفته	۴۲۴ (۳۶/۲)	۷۵ (۱۷/۷)	۳۴۹ (۸۲/۳)	۰/۰۰۱
	۴ بار و بیشتر در هفته	۷۴۶ (۶۳/۸)	۳۱۵ (۴۲/۲)	۴۳۱ (۵۷/۸)	
گوشت قرمز	سه بار و کمتر در هفته	۶۳۱ (۵۳/۹)	۱۵۸ (۲۵/۰)	۴۷۳ (۷۵/۰)	۰/۰۰۳
	۴ بار و بیشتر در هفته	۵۳۹ (۴۶/۱)	۲۳۲ (۴۳/۰)	۳۰۷ (۵۷/۰)	
علاقه به غذای داغ داشتید؟	خیر	۳۹۰ (۳۳/۳)	۸۲ (۲۱/۰)	۳۰۸ (۷۹/۰)	۰/۰۰۱
	بلی	۷۸۰ (۶۶/۷)	۳۰۸ (۳۹/۵)	۴۷۲ (۶۰/۵)	
علاقه به غذای تند و فلفلی داشتید؟	خیر	۶۰۳ (۵۱/۵)	۲۰۷ (۳۴/۳)	۳۹۶ (۶۵/۷)	۰/۴۵۷
	بلی	۵۶۷ (۴۸/۵)	۱۸۳ (۳۲/۳)	۳۸۴ (۶۷/۷)	
آیا سابقه‌ی مصرف دارو بدون نسخه‌ی پزشک (خوددرمانی) داشتید؟	خیر	۴۴۱ (۳۷/۷)	۷۹ (۱۷/۹)	۳۶۲ (۸۲/۱)	۰/۰۰۱
	بلی	۷۲۹ (۶۲/۳)	۳۱۱ (۴۲/۷)	۴۱۸ (۵۷/۳)	
سابقه‌ی مصرف سیگار داشتید؟	خیر	۲۰۲ (۱۷/۳)	۶۲ (۳۰/۷)	۱۴۰ (۶۹/۳)	۰/۶۲۸
	۱-۹ نخ سیگار در روز	۲۳۲ (۱۹/۸)	۷۷ (۳۳/۲)	۱۵۵ (۶۶/۸)	
	۱۰-۲۰ نخ سیگار در روز	۳۰۲ (۲۵/۸)	۹۷ (۳۲/۱)	۲۰۵ (۶۷/۹)	
تعداد پاکت‌های سیگار در سال	> ۲۰ نخ سیگار در روز	۴۳۴ (۳۷/۱)	۱۵۴ (۳۵/۵)	۲۸۰ (۶۴/۵)	۰/۰۰۲
	انحراف معیار $\pm$ میانگین	۱۰۰/۸ $\pm$ ۹/۶	۱۲/۲ $\pm$ ۹/۶	۱۰۱ $\pm$ ۹/۵	
	مصرف نکرده	۶۳۲ (۵۴/۰)	۲۰۴ (۳۲/۳)	۴۲۸ (۶۷/۷)	
الکل	طی ۶ ماه مصرف نکرده	۲۸۲ (۲۴/۱)	۱۰۶ (۳۷/۶)	۱۷۶ (۶۲/۴)	۰/۲۱۱
	حداقل یکبار طی ۶ ماه مصرف کرده	۲۵۶ (۲۱/۹)	۸۰ (۳۱/۳)	۱۷۶ (۶۸/۷)	
	سوابق افراد				
سابقه‌ی خانوادگی به سرطان	بلی	۸۶۹ (۷۴/۳)	۲۴۶ (۲۸/۳)	۶۲۳ (۷۱/۷)	۰/۰۰۱
	خیر	۳۰۱ (۲۵/۷)	۱۴۴ (۴۷/۸)	۱۵۷ (۵۲/۲)	
سابقه‌ی خانوادگی به سرطان معده	بلی	۱۰۸۰ (۹۲/۳)	۳۲۹ (۳۰/۵)	۷۵۱ (۶۹/۵)	۰/۰۰۱
	خیر	۹۰ (۷/۷)	۶۱ (۶۷/۸)	۲۹ (۳۲/۲)	
سابقه‌ی ریفلاکس معده	خیر	۶۲۳ (۵۳/۲)	۱۳۷ (۲۲/۰)	۴۸۶ (۷۸/۰)	۰/۰۲۶
	بلی	۵۴۷ (۴۶/۸)	۲۵۳ (۴۶/۳)	۲۹۴ (۵۳/۷)	
سابقه‌ی زخم معده	خیر	۶۳۸ (۶۲/۱)	۸۶ (۱۳/۵)	۵۵۲ (۸۶/۵)	۰/۰۰۱
	بلی	۳۹۰ (۳۷/۹)	۳۰۴ (۷۷/۹)	۸۶ (۲۲/۱)	
گروه خونی	A	۴۶۸ (۴۰/۰)	۲۲۸ (۴۸/۷)	۲۴۰ (۵۱/۳)	۰/۰۱۱
	AB	۳۳۴ (۲۸/۵)	۷۰ (۲۱/۰)	۲۶۴ (۷۹/۰)	
	B	۲۶۹ (۲۳/۰)	۵۹ (۱۵/۱)	۲۱۰ (۷۸/۱)	
	O	۹۹ (۸/۵)	۳۳ (۳۳/۳)	۶۶ (۶۶/۷)	

\* آزمون Independent Sample T-test دو جامعه برای مقایسه‌ی میانگین متغیرهای کمی؛ آزمون Chi-square برای مقایسه‌ی متغیرهای کیفی؛ سطح معنی‌داری آزمون آماری ۰/۰۰۵ نظر گرفته شد.

از بیماران در مرحله‌ی اول، ۳۴/۶ درصد در مرحله‌ی دوم، و ۱۷/۷ درصد در مرحله‌ی سوم از مراحل پیشرفت سرطان معده تشخیص داده شدند (جدول ۲).

**گزارش رگرسیون لجستیک:** شانس متغیرها در بیماران مبتلا به سرطان معده و فاصله‌ی اطمینان ۹۵ درصدی مربوط به آن‌ها، در مدل ساده و مدل چندگانه رگرسیون لجستیک گزارش شد (جدول ۳).

**اطلاعات توصیفی بیماران مبتلا به سرطان معده:** عواملی نظیر هلیکوباکتر پیلوری، نوع و مرحله‌ی سرطان معده تنها در بیماران مبتلا به سرطان معده بررسی شد. ۸۴/۷ درصد از بیماران مبتلا به سرطان معده، سابقه‌ی ابتلا به باکتری هلیکوباکتر داشتند. بیشترین نوع سرطان معده مربوط به تومور استرومایی دستگاه گوارش بود که ۷۱/۵ درصد بیماران را تشکیل داد. بر اساس اطلاعات به دست آمده، ۴۷/۴ درصد

جدول ۲. اطلاعات توصیفی بیماران مبتلا به سرطان معده (گروه مورد)

متغیر	سطح	کل (درصد) [تعداد]
ابتلا به باکتری هلیکوباکتر	خیر	۴۹ (۱۵/۳)
	بلی	۲۷۲ (۸۴/۷)
نوع سرطان معده	آندنوکارسینوم	۱۱۱ (۲۸/۵)
	تومور استرومایی دستگاه گوارش	۲۷۹ (۷۱/۵)
مرحله‌ی تشخیص داده شده‌ی سرطان معده	مرحله ۱	۱۶۹ (۴۷/۴)
	مرحله ۲	۱۲۳ (۳۴/۶)
	مرحله ۳	۶۳ (۱۷/۷)

هفته میوه مصرف می‌کنند، شانس سرطان معده در آنان ۰/۱۹ برابر افرادی بود که ۳ بار یا کمتر از ۳ بار در هفته میوه مصرف کردند ( $P = ۰/۰۰۱$ ,  $CI = (۰/۱۰, ۰/۳۳)$ ). شانس سرطان معده در افرادی که سابقه‌ی خانوادگی سرطان معده داشتند، ۴/۳۷ بار بیشتر از افراد بدون سابقه‌ی خانوادگی سرطان معده بود. بنابراین سابقه‌ی خانوادگی سرطان معده با شانس سرطان معده از لحاظ آماری ارتباط معنی‌داری داشت ( $P = ۰/۰۱۶$ ,  $CI = (۱/۲۵, ۸/۹۱)$ ). نتایج مربوط به شانس و فاصله‌ی اطمینان ۰/۹۵ متغیرها در جدول ۳ گزارش شده است.

متغیرهایی که در مدل ساده، مقدار  $P$  آن کمتر از ۰/۲۵ بود وارد مدل چندگانه شدند. نتایج در مدل چندگانه‌ی لجستیک نشان داد که شانس سرطان معده در مردان ۱/۰۶ برابر زنان است ( $P = ۰/۸۳۶$ ,  $CI = (۰/۶۲, ۱/۱۹)$ ). همچنین نتایج حاکی از آن است که به ازای یک‌سال افزایش در سن، شانس سرطان معده ۱/۱۳ برابر افزایش می‌یابد که ارتباط معنی‌دار سن با سرطان معده را نشان داد ( $P = ۰/۰۰۱$ ,  $CI = (۱/۱۳, ۱/۱۸)$ ). به علاوه مصرف میوه یک عامل بازدارنده برای سرطان معده بوده؛ به طوری که افرادی که ۴ بار یا بیشتر از ۴ بار در

جدول ۳. نتایج آماری مدل‌های رگرسیون برای بیماران مبتلا به سرطان معده

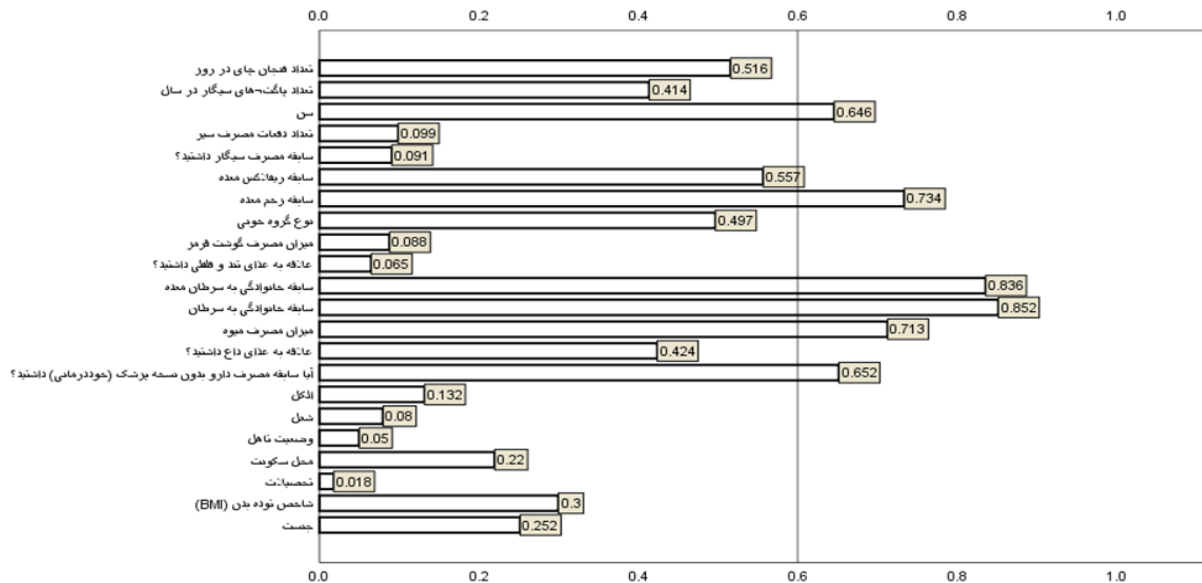
متغیر	سطح		مدل ساده رگرسیون لجستیک		مدل چندگانه رگرسیون لجستیک	
	p	فاصله‌ی اطمینان (%۹۵) شانس	p	فاصله‌ی اطمینان (%۹۵) شانس	p	فاصله‌ی اطمینان (%۹۵) شانس
اطلاعات دموگرافیک						
سن (سال)	۰/۰۰۱	۱/۱۳ (۱/۰۸, ۱/۱۸)	۰/۰۰۱	۱/۱۱ (۱/۰۸, ۱/۱۴)	۰/۰۰۱	۱/۱۳ (۱/۰۸, ۱/۱۸)
جنس						
زن	۰/۸۳۶	۱ (رفرنس)	۰/۱۳۰	۱ (رفرنس)	۰/۸۳۶	۱ (رفرنس)
مرد						
شاخص توده‌ی بدن (BMI)						
<۲۴						
۱۴+۲۷	۰/۵۹۹	۱/۲۷ (۰/۵۲, ۳/۰۹)	۰/۷۸۸	۱/۲۰ (۰/۷۹, ۱/۸۳)	۰/۵۹۹	۱/۲۷ (۰/۵۲, ۳/۰۹)
> ۲۷	۰/۸۹۶	۱/۰۶ (۰/۴۳, ۲/۶۱)	۰/۶۷۸	۱/۰۴ (۰/۶۸, ۱/۵۸)	۰/۸۹۶	۱/۰۶ (۰/۴۳, ۲/۶۱)
تحصیلات						
بیسواد						
۱ (رفرنس)						
ابتدایی - سیکل	۰/۰۸۶	۰/۴۴ (۰/۱۶, ۱/۱۲)	۰/۸۵۲	۰/۹۶ (۰/۵۹, ۱/۵۴)	۰/۰۸۶	۰/۴۴ (۰/۱۶, ۱/۱۲)
دیپلم	۰/۲۲۸	۰/۵۵ (۰/۲۱, ۱/۴۵)	۰/۵۳۰	۰/۸۵ (۰/۵۲, ۱/۴۰)	۰/۲۲۸	۰/۵۵ (۰/۲۱, ۱/۴۵)
دانشگاهی	۰/۰۵۱	۰/۳۵ (۰/۱۲, ۱/۰۱)	۰/۵۱۳	۰/۸۳ (۰/۴۸, ۱/۴۳)	۰/۰۵۱	۰/۳۵ (۰/۱۲, ۱/۰۱)
محل سکونت						
روستایی	۰/۸۵۱	۱ (رفرنس)	۰/۰۶۶	۱ (رفرنس)	۰/۸۵۱	۱ (رفرنس)
شهری						
وضعیت تأهل						
مجرد	۰/۲۱۰	۱/۰۷ (۰/۵۱, ۲/۲۸)	۰/۸۲۵	۱ (رفرنس)	۰/۲۱۰	۱ (رفرنس)
متاهل						
شغل						
کشاورز						
۱ (رفرنس)						
کارمند	۰/۱۶۰	۰/۵۷ (۰/۲۶, ۱/۲۴)	۰/۰۱۸	۰/۶۳ (۰/۴۳, ۰/۹۲)	۰/۱۶۰	۰/۵۷ (۰/۲۶, ۱/۲۴)
بیکار	۰/۰۹۴	۱/۸۲ (۰/۹۰, ۳/۶۸)	۰/۰۱۸	۱/۵۱ (۱/۰۷, ۲/۱۳)	۰/۰۹۴	۱/۸۲ (۰/۹۰, ۳/۶۸)
بازنشسته	۰/۱۵۶	۱/۶۹ (۰/۸۲, ۳/۵۲)	۰/۰۰۱	۱/۷۸ (۱/۲۵, ۲/۵۳)	۰/۱۵۶	۱/۶۹ (۰/۸۲, ۳/۵۲)

مدل چندگانه رگرسیون لجستیک		مدل ساده رگرسیون لجستیک		سطح	متغیر
p	(فاصله‌ی اطمینان ۹۵٪) شانس	p	(فاصله‌ی اطمینان ۹۵٪) شانس		
					الگوی مصرف مواد غذایی
۰/۰۰۱	۱/۶۲ (۱/۴۳، ۱/۸۳)	۰/۰۰۱	۱/۵۱ (۱/۴۲، ۱/۶۰)		تعداد فنجان چای در روز
۰/۰۸۲	۱ (رفرنس)	۰/۱۱۶	۱ (رفرنس)	سه بار و کمتر در هفته	تعداد دفعات مصرف سیر
	۱/۵۷ (۰/۹۴، ۲/۶۲)		۰/۸۲ (۰/۶۴، ۱/۰۵)	۴ بار و بیشتر در هفته	
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	سه بار و کمتر در هفته	میوه
	۰/۱۹ (۰/۱۰، ۰/۳۳)		۳/۴۰ (۲/۵۵، ۴/۵۳)	۴ بار و بیشتر در هفته	
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	سه بار و کمتر در هفته	گوشت قرمز
	۰/۲۶ (۰/۱۶، ۰/۴۴)		۲/۲۶ (۱/۷۶، ۲/۸۹)	۴ بار و بیشتر در هفته	
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	خیر	علاقه به غذای داغ داشتید؟
	۰/۳۲ (۰/۱۷، ۰/۵۸)		۲/۴۵ (۱/۸۵، ۳/۲۵)	بلی	
۰/۲۹۶	۱ (رفرنس)	۰/۴۵۷	۱ (رفرنس)	خیر	علاقه به غذای تند و فلفلی داشتید؟
	۱/۳۰ (۰/۷۸، ۲/۱۲)		۰/۹۱ (۰/۷۲، ۱/۱۶)	بلی	
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	خیر	آیا سابقه‌ی مصرف دارو بدون نسخه‌ی پزشک (خوددرمانی) داشتید؟
	۰/۳۰ (۰/۱۸، ۰/۵۲)		(۲/۵۷، ۴/۵۳)	بلی	سابقه‌ی مصرف سیگار داشتید؟
	۱ (رفرنس)		۱ (رفرنس)	خیر	
۰/۱۱۸	۲/۲۵ (۰/۶، ۸/۱/۲۲)	۰/۲۳۵	۰/۸۱ (۰/۵۶، ۱/۱۵)	۱-۹ نخ سیگار در روز	
۰/۶۷۰	۱/۱۹ (۰/۵۳، ۲/۷۱)	۰/۵۵۳	۰/۹۰ (۰/۶۴، ۱/۲۶)	۱۰-۲۰ نخ سیگار در روز	
۰/۸۵۵	۱/۰۷ (۰/۵۳، ۲/۱۶)	۰/۳۴۴	۰/۸۶ (۰/۶۳، ۱/۱۷)	> ۲۰ نخ سیگار در روز	
۰/۱۰۰	۱/۰۳ (۰/۹۹، ۱/۰۷)	۰/۰۰۱	۱/۰۲ (۱/۰۱، ۱/۰۴)		تعداد پاکت‌های سیگار در سال
۰/۹۳۵	۱ (رفرنس)		۱ (رفرنس)	مصرف نکرده	الکل
	۰/۹۷ (۰/۴۸، ۱/۹۷)	۰/۷۶۶	۱/۰۵ (۰/۷۷، ۱/۴۳)	طی ۶ ماه مصرف نکرده	
۰/۶۰۵	۱/۲۲ (۰/۵۸، ۲/۵۸)	۰/۱۲۳	۱/۳۲ (۰/۹۳، ۱/۸۹)	حداقل یکبار طی ۶ ماه مصرف کرده	
					سوابق افراد
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	خیر	سابقه‌ی خانوادگی به سرطان
	۳/۲۱ (۱/۷۱، ۶/۰۴)		۲/۳۲ (۱/۷۷، ۳/۰۴)	بلی	
۰/۰۱۶	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	خیر	سابقه‌ی خانوادگی به سرطان معده
	۳/۳۴ (۱/۲۵، ۸/۹۱)		۴/۸۰ (۳/۰۳، ۷/۶۱)	بلی	
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	خیر	سابقه‌ی ریفلاکس معده
	۴/۳۷ (۲/۶۱، ۷/۳۳)		۳/۰۵ (۲/۳۷، ۳/۹۳)	بلی	
۰/۰۰۱	۱ (رفرنس)	۰/۰۰۱	۱ (رفرنس)	خیر	سابقه‌ی زخم معده
	۳۶/۱۶ (۲۰/۱۱، ۶۵/۰۳)		۲۲/۶۹ (۱۶/۳۱، ۳۱/۵۶)	بلی	
	۱ (رفرنس)		۱ (رفرنس)	A	گروه خونی
۰/۱۵۸	۲/۰۳ (۰/۷۶، ۵/۴۴)	۰/۰۰۶	۱/۹۰ (۱/۲۱، ۲/۹۹)	AB	
۰/۳۰۶	۰/۵۸ (۰/۲۰، ۱/۶۵)	۰/۰۱۲	۰/۵۳ (۰/۳۲، ۰/۸۷)	B	
۰/۴۶۲	۰/۶۷ (۰/۲۳، ۱/۹۳)	۰/۰۲۶	۰/۵۶ (۰/۳۴، ۰/۹۳)	O	

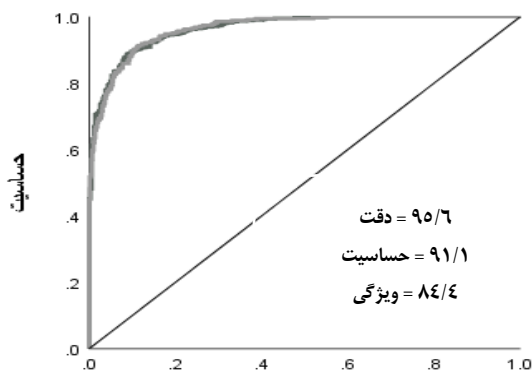
\* سطح معنی‌داری آزمون آماری ۰/۰۵ در نظر گرفته شد.

مدل در گروه آموزش ۱۸۳/۸۰ و در گروه آزمون ۳۸/۱۲ بود. همچنین درصد پیشگویی نادرست در گروه آموزش، ۱۰/۸ درصد و در گروه آزمون، ۷/۸ درصد گزارش شد.

**برازش مدل با رویکرد شبکه عصبی:** با انجام برازش مدل بر اساس رویکرد شبکه‌ی عصبی با ۷ لایه‌ی ورودی، ۴ لایه‌ی میانی و یک لایه‌ی خارجی، خطای آنتروپی محاسبه گردید. خطای آنتروپی



شکل ۱. گزارش میزان اهمیت عوامل مرتبط با سرطان معده از نظر مدل شبکه‌ی عصبی مصنوعی



ویژگی - ۱

شکل ۳. منحنی راک در مدل برازش شده‌ی رگرسیون لجستیک

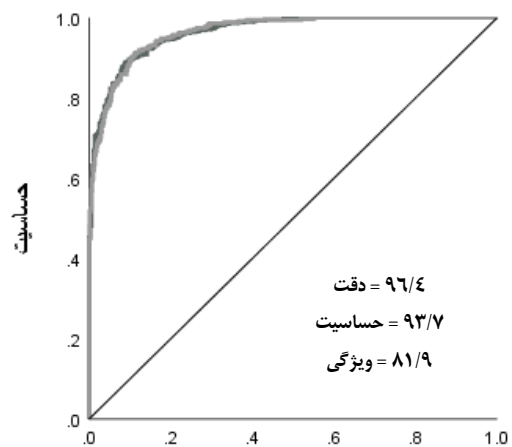
نتایج حاصل از برازش مدل شبکه‌ی عصبی در شکل ۱ نشان می‌دهد که ۶ عامل شامل سابقه‌ی خانوادگی به سرطان (۰/۸۵۲)، سابقه‌ی خانوادگی به سرطان معده (۰/۸۳۶)، سابقه‌ی زخم معده (۰/۷۳۴)، مصرف میوه (۰/۷۱۳)، سابقه‌ی مصرف داروی خودسر (۰/۶۵۲) و سن (۰/۶۴۶) از مهم‌ترین عوامل مرتبط با سرطان معده بودند. سایر عوامل مرتبط با سرطان معده کمتر از ۰/۶۰۰ اهمیت داشتند. لذا در مطالعه‌ی حاضر نقش متغیرها با اهمیت کمتر از ۰/۵۰۰ در پیش‌گویی سرطان معده حائز اهمیت نبودند.

**منحنی راک:** جهت مقایسه‌ی مدل رگرسیون لجستیک با مدل شبکه‌ی عصبی از منحنی راک استفاده شد. نتایج حاصل در شکل ۲ و شکل ۳ گزارش گردید.

در شبکه‌ی عصبی دقت، حساسیت و ویژگی به ترتیب ۹۶/۴، ۹۳/۷ و ۸۱/۹ درصد و در مدل برازش شده‌ی رگرسیون لجستیک به ترتیب ۹۵/۹، ۹۱/۱ و ۸۴/۴ درصد گزارش شد (شکل ۲، ۳).

### بحث

در دهه‌ی اخیر دستاوردهای پیچیده نظیر شبکه‌های عصبی مصنوعی جهت پیش‌بینی پیامدهای مختلف در حوزه‌ی علوم پزشکی و بهداشت داشته است. مطالعات متعددی در این زمینه در حال انجام یا خاتمه یافته است (۷). در مطالعه‌ی حاضر، مدل شبکه‌ی عصبی طراحی و مورد استفاده قرار گرفت که یکی از روش‌های کاربردی در بررسی عوامل خطر سرطان معده و پیش‌بینی آن است. نتایج مطالعه‌ی حاضر حاکی از آن بود که مدل شبکه‌ای عصبی، دقت و حساسیت بالاتر نسبت به مدل رگرسیون لجستیک دارد. همچنین سطح زیر منحنی



ویژگی - ۱

شکل ۲. منحنی راک در مدل برازش شده‌ی شبکه‌ی عصبی

سرطان، سابقه‌ی خانوادگی به سرطان معده، سابقه‌ی ریفلاکس معده و سابقه‌ی زخم معده در برازش مدل چندگانه‌ی رگرسیون لجستیک معنی‌دار بودند. متغیرهای تعداد پاکت‌های سیگار در سال، گوشت قرمز، تعداد فنجان چای در روز و علاقه به غذای داغ در مدل چندگانه‌ی رگرسیون لجستیک معنی‌دار بود اما در مدل شبکه‌ی عصبی، متغیرهای تأثیرگذاری نبودند. این که برخی از متغیرها در مدل رگرسیون لجستیک معنی‌دار هستند اما در مدل شبکه‌ی عصبی، معنی‌دار نبودند ممکن است به این دلیل باشد که مدل شبکه‌ی عصبی، شرایط تعامل بین متغیرها را در نظر می‌گیرد.

### نتیجه‌گیری

نتایج مطالعه‌ی حاضر نشان داد که سطح زیرمنحنی راک، دقت و حساسیت در مدل شبکه‌ی عصبی از مدل رگرسیون لجستیک بیشتر بود. لذا به پژوهشگران پیشنهاد می‌شود که مدل شبکه‌ی عصبی، جایگزین رگرسیون لجستیک شود. همچنین بر اساس مدل شبکه‌ی عصبی، ۶ عامل مهم مرتبط با سرطان معده معرفی شد. این نتیجه می‌تواند به مدیران و سیاست‌گذاران سلامت در جهت افزایش آگاهی مردم و کاهش بروز سرطان معده مفید باشد.

### تشکر و قدردانی

کمیته‌ی اخلاق دانشگاه علوم پزشکی کرمانشاه این مطالعه را با کد IR.KUMS.REC.1400.171 تأیید کرد. اهداف مطالعه برای نمونه‌ها توضیح داده شد. رضایت کتبی آگاهانه از همه‌ی نمونه‌ها اخذ گردید. به تمامی افراد اطمینان داده شد که اطلاعات آن‌ها محرمانه و تحلیل‌ها به صورت گروهی گزارش می‌شوند. این مطالعه توسط دانشگاه کردستان با شماره گزنت ۰۰/۱۱/۱۴۱۴۱ تأمین مالی شده است. نویسندگان از حمایت‌های مالی معاونت پژوهشی دانشگاه کردستان کمال تشکر را دارند.

راک مدل شبکه‌ی عصبی بیشتر از رگرسیون لجستیک بود. لذا مدل شبکه‌ی عصبی از برازش بهتری برخوردار بود.

در مطالعه‌ی Hirasawa که توسط همکاران در ژاپن انجام شد نتایج نشان داد که حساسیت مدل شبکه‌ی عصبی بالاتر از رگرسیون لجستیک بود. همچنین قدرت بیشتری جهت بررسی و شناسایی عوامل خطر با سرطان معده داشت (۱۰). مطالعه‌ی مشابهی که در کره انجام شد، نشان داد آنالیز شبکه‌ی عصبی دارای اثربخشی بالایی بود (۱۱).

پژوهشی که روی ۳۱۷۰ فرد مبتلا به سرطان معده انجام شد، نشان داد که دقت، حساسیت و ویژگی مدل شبکه‌ی عصبی برای آنالیز داده‌ها بیشتر است (۱۲). مطالعه‌ی که در ژونگشان چین انجام گرفت حاکی از دقت و حساسیت مدل شبکه‌ی عصبی بالاتر از سایر روش‌های تشخیصی بود (۱۳).

مطالعه‌ی Biglarian و همکاران در رابطه با سرطان معده انجام شد، نشان داد که دقت پیش‌بینی متغیرها برای ایجاد سرطان معده در شبکه‌ی عصبی، بالاتر از مدل رگرسیون وایبل بود (۸). اگرچه مدل‌های یادگیری عمیق به طور گسترده برای وظایف مختلف پیش‌آگهی استفاده شدند و عملکرد پیش‌بینی‌کنندگی برتری نسبت به مدل‌های آماری کلاسیک داشتند، اما محدودیت‌هایی در مورد توضیح‌پذیری آن‌ها وجود دارد، که مانع از کاربرد گسترده‌تر مدل‌های یادگیری عمیق برای استفاده‌ی بالینی و تحقیقات شده است.

نتایج حاصل از برازش شبکه‌ی عصبی نشان داد که عواملی نظیر سابقه‌ی خانوادگی به سرطان، سابقه‌ی خانوادگی به سرطان معده، سابقه‌ی زخم معده، مصرف میوه، سابقه‌ی مصرف خودسرانه‌ی دارو، سن و سابقه‌ی ریفلاکس معده بیشترین ارتباط را با سرطان معده داشت. از طرفی، عواملی نظیر سن، تعداد فنجان چای در روز، مصرف میوه، گوشت قرمز، علاقه به غذای داغ، مصرف داروی خودسرانه، تعداد پاکت‌های سیگار در سال، سابقه‌ی خانوادگی به

### References

- Naghi Pour A, Moghimbeigi A, Shirmohamadi N, Soltanian A, Khazaei S, Nick Ceiar S. Geographical distribution of breast cancer in Hamadan Province during 2008-2015 using bayesian method [in Persian]. *Iran J Epidemiol* 2022; 17(4): 362-71.
- Ishak NS, Abdul Rahman H, Lee SHF, Lu SK, Naing L. Incidence, Survival and Prognostic Factors of Oesophagogastric Cancer. *J Gastrointest Cancer* 2022; 53(1): 130-43.
- Maron SB, Catenacci DVT. Novel Targeted Therapies for Esophagogastric Cancer. *Surg Oncol Clin N Am* 2017; 26(2): 293-312.
- Desai G, Schelske-Santos M, Nazario CM, Rosario-Rosado RV, Mansilla-Rivera I, Ramírez-Marrero F, et al. Onion and Garlic Intake and Breast Cancer, a Case-Control Study in Puerto Rico. *Nutr Cancer* 2020; 72(5): 791-800.
- Allahyari E, Hanachi P, Mirmoosavi SJ, Ferns GA, Bahrami A, Ghayour-Mobarhan M. Association between Cardiometabolic risk factor and responsiveness to vitamin D supplementation: a new approach using artificial neural network analysis. *BMC Nutr* 2021; 7(1): 7.
- Li Y, Rao S, Mamouei M, Salimi-Khorshidi G, Canoy D, Hassaine A, et al. Risk factor identification for incident heart failure using neural network distillation and variable selection. *arXiv preprint arXiv 2021: 2102. 12936.*



7. Gohari MR, Mokhtari P, Pourhoseingholi MA, Biglarian A. Artificial neural network in survival analysis of gastric cancer patients [in Persian]. *Payesh* 2014; 13(3): 285-91.
8. Biglarian A, Hajizadeh E, Kazemnejad A, Zayeri F. Determining of prognostic factors in gastric cancer patients using artificial neural networks. *Asian Pac J Cancer Prev* 2010; 11(2): 533-6.
9. Mohammadpour RA, Esmaeili MH, Ghaemian A, Esmaeili J. Application of Artificial Neural Network for Assessing Coronary Artery Disease [in Persian]. *J Mazandaran Univ Med Sci* 2012; 21(86): 9-17.
10. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018; 21(4): 653-60.
11. Yoon HJ, Kim JH. Lesion-based convolutional neural network in diagnosis of early gastric cancer. *Clin Endosc* 2020; 53(2): 127-31.
12. Wu L, Zhou W, Wan X, Zhang J, Shen L, Hu S, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. *Endoscopy* 2019; 51(6): 522-31.
13. Zhu Y, Wang QC, Xu MD, Zhang Z, Cheng J, Zhong YS, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* 2019; 89(4): 806-15.

## Investigating the Factors Associated with Gastric Cancer by Neural Network Approach and Multiple Logistic Regression: A Case-Control

Armin Naghipour<sup>1</sup>, Meysam Hosseini<sup>2</sup>, Arsalan Rahmani<sup>3</sup>, Zahra Naghibifar<sup>4</sup>

### Original Article

#### Abstract

**Background:** Stomach cancer is the fifth most common disease and the third cause of death in the world. Therefore, in order to prevent and reduce the incidence of stomach cancer, factors related to logistic regression and neural network models were investigated.

**Methods:** In this study, a survey was conducted on 1,170 people as (n = 390) cases and (n = 780) controls. The data collection tool was based on the researcher's checklist. The samples were selected by available sampling method and their information was collected by face-to-face and telephone interviews. The fitting power in the logistic regression model and neural network was compared with receiver function characteristic curve (AUROC), sensitivity and specificity. By introducing the superior model, significant and related factors with stomach cancer were reported.

**Findings:** The results showed that the accuracy, sensitivity and specificity of the neural network were 96.4%, 93.7% and 81.9%, respectively. But the accuracy, sensitivity and specificity of the logistic regression model were reported as 95.9%, 91.1% and 84.4%, respectively. The neural network model indicates the variables of age (0.646), fruit consumption (0.713), history of self-medication (0.652), history of gastric ulcer (0.734), family history of cancer (0.852) and Family history of stomach cancer (0.836) were associated with the incidence of stomach cancer.

**Conclusion:** Considering that in the present study, the fit of the neural network was superior to logistic regression and it does not need any special assumptions, so it is suggested to the researchers that the neural network model can be preferred over logistic regression.

**Keywords:** Stomach cancers; Gastric cancer; Neural network; Multilayer perceptron; Logistic regression

**Citation:** Naghipour A, Hosseini M, Rahmani A, Naghibifar Z. **Investigating the Factors Associated with Gastric Cancer by Neural Network Approach and Multiple Logistic Regression: A Case-Control.** J Isfahan Med Sch 2023; 40(693): 880-9.

1- MSc of Biostatistics, Clinical Research Development Center, Imam Reza Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran

2- Assistant Professor, Department of Mathematics, Campus of Bijar, University of Kurdistan, Sanandaj, Iran

3- Assistant Professor, Department of Mathematics, University of Kurdistan, Sanandaj, Iran

4- MSc of Epidemiology, Infectious Diseases Research Center, Imam Reza Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran

**Corresponding Author:** Meysam Hosseini, Assistant Professor, Department of Mathematics, Campus of Bijar, University of Kurdistan, Sanandaj, Iran; Email: me.hosseini@uok.ac.ir