

## مقایسه دقت پیش بینی شبکه های عصبی مصنوعی و رگرسیون لجستیک دو متغیره در تشخیص هم زمان بیماری فشارخون و دیابت

مهدی اداوی: گروه آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران، تهران، ایران. mehdi.adavi@yahoo.com

دکتر مسعود صالحی: مرکز تحقیقات علوم مدیریت و اقتصاد سلامت، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران، تهران، ایران. salehi74@yahoo.com

\*دکتر مسعود رودباری: گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران و مرکز تحقیقات مقاومت های میکروبی، دانشگاه علوم پزشکی ایران، تهران، ایران (نویسنده مسئول). mroudbari@yahoo.co.uk

دکتر فرشته عسگری: وزارت بهداشت درمان و آموزش پزشکی، مرکز مدیریت بیماری ها، اداره مراقبت عوامل خطر بیماری های غیرواگیر، تهران، ایران. asgarifcdc@yahoo.com

علی رافعی: وزارت بهداشت درمان و آموزش پزشکی، مرکز مدیریت بیماری ها، اداره مراقبت عوامل خطر بیماری های غیرواگیر، تهران، ایران. rafeicdc@yahoo.com

تاریخ پذیرش: ۹۲/۱۰/۲۵

تاریخ دریافت: ۹۲/۵/۴

### چکیده

**زمینه و هدف:** دیابت و فشار خون از جمله بیماری های غیر واگیر هستند که شیوع آنها برای مسئولان بهداشتی کشور بسیار مهم می باشند. هدف این مطالعه مقایسه مدل رگرسیون لجستیک (LR) دو متغیره با شبکه های عصبی مصنوعی (Artificial Neural Networks=ANN) در پیش بینی هم زمان رخداد بیماری فشارخون و دیابت می باشد.

**روش کار:** این مطالعه تحلیلی- مقطعی در سال ۱۳۹۲-۱۳۹۱ در تهران با نمونه ای ۱۲۰۰۰ نفر از بالغین انجام شد. پرسشنامه طرح شامل گزینه هایی درباره فشار خون و دیابت و عوامل خطر این بیماری ها بود. یک مدل ANN پرسپترون با دو لایه پنهان با به کارگیری نرم افزار MATLAB بر داده ها برازش داده شد. متغیرهای مطالعه سابقه دیابت و فشار خون، جنسیت، نوع روغن مصرفی، فعالیت بدنی، سابقه فامیلی، سن، سیگار کشیدن و چاقی بودند. برای مدل سازی ابتدای توام به بیماریها، از مدل LR توام در نرم افزار SAS استفاده شد. برای بررسی دقت پیش بینی های حاصل از مدل LR و ANN در ابتدای توام بیماری ها از سطح زیر منحنی ROC استفاده گردید.

**یافته ها:** متغیرهای جنس، نوع روغن مصرفی، فعالیت بدنی، سابقه فامیلی، سن، سیگاری غیرفعال و چاقی وارد مدل LR دو متغیره و ANN شدند. برای مدل LR نسبت بخت های متغیرهای فوق به ترتیب ۱/۱۴، ۰/۵۸، ۱/۸، ۱/۳۲، ۰/۳۶، ۰/۷۶ و ۰/۴۷ بدست آمد. بنابراین بخت ابتدای توام به بیماری ها در زنان (۱/۱۴)، عدم فعالیت بدنی (۱/۸) و دارای سابقه فامیلی (۱/۳۲) نسبت به سایر گروه ها بیشتر است. سطح زیر منحنی ROC برای مدل LR دو متغیره و ANN به ترتیب ۰/۷۸ (p=۰/۰۳۹) و ۰/۸۶ (p=۰/۰۴۶) حاصل شد.

**نتیجه گیری:** با توجه به نتایج و مقایسه دقت پیش بینی روش های فوق، به کارگیری ANN نسبت به مدل LR دو متغیره برای تشخیص هم زمان بیماری دیابت و فشار خون دقت بالاتری دارد.

**کلیدواژه ها:** شبکه های عصبی مصنوعی، رگرسیون لجستیک دو متغیره، بیماری دیابت، فشارخون

### مقدمه

یکی از مسائلی که محققین در مطالعات پزشکی و بهداشتی با آن مواجه اند، پیش بینی ابتدای هم زمان به دو بیماری براساس برخی عوامل خطر مشترک است. این بیماری ها به صورت یک توزیع توام با دو یا چند متغیره پاسخ (وابسته) در نظر گرفته شده و بر اساس تعدادی متغیرهای مستقل، پیش بینی رخداد هم زمان دو بیماری صورت خواهد گرفت که در علم آمار این مدل ها را مدل های دو یا چند متغیره می نامند. زمانی که متغیرهای پاسخ هردو کمی یا کیفی باشند از روش های استاندارد آماری

مانند رگرسیون دو متغیره یا رگرسیون لجستیک دو متغیره برای مدل بندی روابط بین متغیرهای وابسته و مستقل استفاده می شود (۱-۲). از جمله کاربردهای این روش های آماری، شناسایی روابط و مدل سازی بین متغیره است و هدف اصلی این مدل سازی ها پیش بینی براساس واقعیت ها و اطلاعاتی است که از یک موضوع خاص در دسترس می باشد (۳-۶).

برای مدل سازی خطی روابط بین متغیرها در روش های کلاسیک آماری، با پیش فرض هایی مانند در نظر گرفتن توزیع نرمال برای متغیرهای پاسخ،

معرض بیماری‌های قلب و عروق، کلیوی و سکت‌های مغزی قرار می‌دهد (۱۱-۱۰). از این رو، یکی از بهترین راه‌ها برای کنترل مرگ و میر ناشی از این بیماری‌ها، پیش بینی رخداد هم‌زمان این دو بیماری به منظور پیشگیری از رخداد آنها است. هدف این پژوهش مقایسه پیش بینی شبکه‌های عصبی مصنوعی و رگرسیون لجستیک دو متغیره در تشخیص هم‌زمان رخداد دو بیماری فشارخون و دیابت می‌باشد.

### روش کار

در این پژوهش تحلیلی مقطعی که در سال تحصیلی ۹۲-۹۱ انجام شد، جامعه آماری پژوهش را بالغین ایرانی گروه سنی ۲۵ تا ۶۴ سال تشکیل می‌دادند. داده‌های مورد بررسی از پیمایش کشوری عوامل خطر بیماری‌های غیرواگیر استخراج شد که این مطالعه در سال ۱۳۸۸ توسط مرکز مدیریت بیماری‌های غیرواگیر وزارت بهداشت، درمان و آموزش پزشکی کشور با استفاده از یک پرسشنامه استاندارد و به شیوه مصاحبه حضوری جمع‌آوری شده و نمونه‌ای به حجم ۱۲۰۰۰ نفر از آن برای بررسی در این تحقیق استفاده شده است.

روش نمونه‌گیری تصادفی طبقه‌ای-خوشه‌ای با احتمال متناسب با حجم می‌باشد که تمام مراحل تعیین حجم نمونه و روش نمونه‌گیری در مرکز مربوطه انجام گرفت و بنا به درخواست محقق با موافقت مرکز، داده‌ها در اختیار محقق گذاشته شده است.

برای استفاده از شبکه عصبی مصنوعی از ۱۲۰۰۰ نمونه گرفته شده، ۸۴۰۰ نمونه (۷۰ درصد) برای آموزش و ۳۶۰۰ نمونه (۳۰ درصد) برای آزمایش شبکه عصبی و پیش‌بینی در توپولوژی شبکه نهایی مورد استفاده قرار گرفت. در مرحله آموزش، وزن‌های بین واحدهای لایه‌های ورودی، میانی و خروجی، به طور تصادفی تعیین خواهند شد. سپس شبکه با پردازش داده‌های هر واحد و ارسال آن‌ها به واحدهای بعدی، مقادیر متغیر وابسته را محاسبه می‌کند. در این مرحله مقادیر محاسبه شده متغیر وابسته با مقادیر واقعی آن‌ها مقایسه و

خطی بودن رابطه‌ی پیشنهادی و یکسان بودن واریانس خطاها مواجهه هستیم. هنگام استفاده عملی از این روش‌ها، اگر داده‌های واقعی شرایط مفروض مدل را نداشته باشند، استفاده از آن امکان پذیر نبوده یا با خطای قابل توجه همراه خواهد بود. از طرفی حساس بودن بیشترین مدل‌ها به مشاهده‌های گم‌شده و داده‌های پرت از دیگر محدودیت‌های این روش‌ها به شمار می‌آیند. بنابراین، نیاز به روش‌هایی با محدودیت‌های کمتر در این زمینه احساس می‌شود.

شبکه‌های عصبی مصنوعی می‌توانند یکی از مناسب‌ترین روش‌ها برای حل این مشکل باشند که نه تنها فرض اولیه‌ای را بر توزیع داده‌ها تحمیل نمی‌کند، بلکه هیچ محدودیتی نیز برای شکل تابعی رابطه بین متغیرهای مستقل و وابسته در این روش در نظر گرفته نمی‌شود. در این شبکه‌ها، رابطه تابعی توسط شبکه کشف شده و لزوماً این رابطه، یک رابطه‌ی خطی نیست. از دیگر مزایای شبکه‌های عصبی مصنوعی این است که در آن داده‌ها به صورت ضمنی پردازش می‌شوند و بر این اساس، احتمال رسیدن به جوابی صحیح حتی با حذف بخشی از لایه‌های شبکه و یا عملکرد نادرست لایه‌ها باز هم وجود دارد (۸-۷).

بیماری‌های فشارخون و دیابت دو بیماری مهم غیرواگیر هستند و فراوانی این بیماری‌ها در کشورهای صنعتی ۸۵٪، در کشورهای با درآمد متوسط ۷۰٪ و در کشورهای فقیر نزدیک به ۵۰٪ برآورد شده است (۹). با توجه به خسارت‌های بزرگ پیش آمده به دلیل مرگ و میر زودرس و یا ناتوانی ناشی از این بیماری‌ها، دولت‌ها خصوصاً در کشورهای صنعتی هزینه‌های بیشتری را برای درمان این بیماری‌های تحمل می‌کنند از این رو، پیشگیری از بیماری‌های غیرواگیر از لحاظ اقتصادی برای کشورهای جهان از اهمیت ویژه‌ای برخوردار است.

فشارخون بالا در بیماران دیابتی به خصوص دیابت نوع دو بسیار شایع است به طوری که بیش از نیمی از بیماران دیابتی نوع دو در موقع تشخیص به بیماری فشارخون نیز مبتلا هستند. همراه شدن این دو بیماری، افراد را بیشتر در

(۳۶۰۰ نمونه) استفاده شد.

نهایتاً با استفاده از شبکه عصبی مصنوعی مطلوب و مدل رگرسیون لجستیک پیش بینی ابتلای هم‌زمان به دو بیماری دیابت و فشار خون صورت گرفت. باتوجه به دو سطحی بودن متغیرهای پاسخ، ملاک مناسب برای سنجش دقت روش های پیش بینی به کارگرفته شده و تعیین میزان پیشبینی صحیح آن‌ها، استفاده از سطح زیرمنحنی مشخصه عملکرد (ROC) تعیین شد. این منحنی عبارت است از حساسیت مدل بر حسب یک منهای ویژگی آن و مساحت زیرمنحنی که عددی بین ۰ و ۱ است و به عنوان ملاکی برای سنجش توانایی پیشبینی مدل‌ها مورد استفاده قرار می‌گیرد. هرچه این عدد به ۱ نزدیک‌تر باشد، نشانه‌ی تواناتر بودن مدل برای پیش‌بینی است (۱۲).

#### یافته‌ها

در ابتدا متغیرهای مورد پژوهش و سطوح مختلف آنها معرفی گردیده و سپس توزیع سطوح مختلف مربوط به هر متغیر در جدول ۱ نشان داده شده است.

با توجه به جدول فوق، بیشترین فراوانی مشاهده شده در افراد مورد بررسی، مربوط به افرادی است که به دیابت مبتلا نبودند. نکته قابل تامل در جدول فوق ابتلای در حدود ۱۲ الی ۱۳ درصدی به هر یک از بیماری‌های دیابت و فشار خون در این نمونه است. نتایج حاصل از رگرسیون لجستیک دو متغیره برای پیش‌بینی توام بیماری‌های دیابت و فشار خون با متغیرهای مستقل در جدول ۲ ارائه شده است.

با توجه به نتایج حاصل از برآورد ضرایب متغیرهای پژوهش می‌توان چنین استنباط نمود که سابقه فامیلی اثر معنادار بر احتمال ابتلا به بیماری دیابت و فشار خون را دارد. بر این اساس، اگر گروه مرجع عدم سابقه فامیلی در نظر گرفته شود، نسبت بخت سابقه فامیلی در ابتلای هم‌زمان به دو بیماری فوق، ۱/۳۲ برابر نسبت بخت عدم سابقه فامیلی است. یعنی بخت ابتلا توأم به بیماری دیابت و فشارخون در گروهی که سابقه

مقدارخطا محاسبه می‌شود. اگر مقدار این خطا از معیار کمترین خطا مورد نظر که از قبل در نظر گرفته شده است، بیشتر بود، شبکه به عقب برگشته و با تغییر وزن‌ها و با تکرار مراحل قبلی دوباره مقدار جدید متغیر وابسته محاسبه می‌شود تا میزان خطای مطلوب بدست آید (۶-۵).

در این مطالعه بخش اطلاعات ورودی که تعداد لایه‌ها و متغیرهای کمکی را نشان می‌دهد، شامل ۷ لایه ورودی و ۷ متغیر مستقل می‌باشد. از این رو برای طراحی شبکه‌ی عصبی مصنوعی مطلوب، ابتدا یک شبکه‌ی پرسپترون پیش‌خور سه لایه با دو لایه میانی و تابع انتقال تانژانت هایپربولیک در نظر گرفته شد (به کارگیری توپولوژی با یک لایه باعث عدم همگرایی شبکه گردید) و در مرحله‌ی بعد با اضافه کردن نرون‌ها به لایه‌ی میانی، یک شبکه پرسپترون با متغیرهای مشابه به داده‌ها برآزش داده شد. با تکرار این مرحله شبکه به شرایط توقف یادگیری که در آن باید ریشه میانگین مربعات خطای (MSE) آزمون کمتر از ۰/۰۱ باشد، دست یافت. پس از اینکه مقادیر MSE برای توپولوژی‌های مختلف شبکه عصبی مورد بررسی قرار گرفت مشاهده شد که مقدار MSE برای شبکه عصبی مصنوعی باتوپولوژی دو لایه میانی، که لایه میانی اول شامل یازده نرون و لایه میانی دوم شامل ده نرون است، کمترین مقدار را دارد. به این دلیل، شبکه عصبی مصنوعی با توپولوژی فوق‌الذکر به کار گرفته شد.

برای پیش‌بینی ابتلای هم‌زمان بیماری دیابت و فشارخون با استفاده از رگرسیون لجستیک دو متغیره نیز، ابتدا متغیرهای نوع روغن مصرفی، سابقه فامیلی، سن (بیشتر و مساوی ۵۵ و کمتر از ۵۵ سال)، جنسیت، چاقی، سیگاری غیرفعال و فعالیت بدنی به عنوان متغیرهای مستقل وارد مدل شدند. سپس با توجه به اینکه ابتلای هم‌زمان دو بیماری مورد نظر بود و مقادیر متغیرهای مستقل برای بیماری دیابت و فشارخون برای هر فرد تکرار شده‌اند، برای تحلیل و مدل‌بندی از روش برآورد معادلات برآوردی تعمیم یافته (Generalized Estimating Equations=GEE) در نرم‌افزار SAS بر روی داده‌های آزمایش شبکه

جدول ۱- توزیع متغیرهای افراد مورد مطالعه

متغیر	سطح	فراوانی	درصد
جنس	مرد	۶۰۰۵	۵۰/۰
	زن	۵۹۹۵	۵۰/۰
نوع روغن مصرفی	جامد	۴۲۸۲	۳۵/۷
	سایر	۷۷۰۳	۶۴/۳
سن	مساوی و بیشتر از ۵۵ سال	۴۷۴۰	۳۹/۵
	کمتر از ۵۵ سال	۷۲۶۰	۶۰/۵
سابقه فامیلی	دارد	۲۳۳۷	۱۹/۵
	ندارد	۹۶۶۳	۸۰/۵
سیگاری غیر فعال	بلی	۱۴۸۹	۱۲/۴
	خیر	۱۰۵۱۱	۸۷/۶
چاقی	بلی	۹۴۹۴	۷۹/۰
	خیر	۲۵۰۶	۲۱/۰
فعالیت بدنی	کم	۴۴۰۴	۳۶/۷
	متوسط	۲۹۱۳	۲۴/۳
	زیاد	۴۶۸۳	۳۹/۰
سابقه فشارخون	ابتلا	۱۵۷۲	۱۳/۱
	عدم ابتلا	۱۰۴۲۸	۸۶/۹
سابقه دیابت	ابتلا	۱۴۸۳	۱۲/۴
	عدم ابتلا	۱۰۵۱۷	۸۷/۶

سیگاری غیرفعال هستند ۳۲ درصد بیشتر از سایرین در معرض ابتلای هم زمان به بیماری های فوق قرار دارند.

با در نظر گرفتن مصرف روغن جامد به عنوان گروه مرجع در مقایسه با مصرف سایر انواع روغن، بخت ابتلای هم زمان به بیماری دیابت و فشار خون در افراد مصرف کننده روغن جامد ۱/۷۲ برابر بخت مصرف سایر انواع روغن می باشد. به عبارت دیگر مصرف روغن جامد بخت ابتلای هم زمان به دو بیماری فوق را ۷۲ درصد افزایش می دهد.

برای متغیر گروه سنی نسبت بخت ۰/۳۶ محاسبه شد. با توجه به اینکه گروه مرجع در این متغیر، سن بیشتر و مساوی ۵۵ سال در نظر گرفته شد، می توان گفت نسبت بخت ابتلای هم زمان به بیماری فشار خون و دیابت در سن کمتر از ۵۵ سال، کمتر از سن مساوی یا بالاتر از ۵۵ سال می باشد.

با در نظر گرفتن مردان به عنوان گروه مرجع برای جنسیت، نسبت بخت ابتلای هم زمان بیماری فشار خون و دیابت در زنان ۱/۱۴ برابر این بخت در مردان است. بنابراین بخت ابتلای هم

فامیلی ابتلا به این دو بیماری را دارند ۳۲٪ نسبت به بخت گروهی که سابقه فامیلی ابتلای به این دو بیماری را ندارند بیشتر است.

با انتخاب چاقی به عنوان گروه مرجع، نسبت بخت ابتلای هم زمان به بیماری دیابت و فشار خون در افراد غیر چاق به چاق ۰/۴۷ است. بنابراین افراد غیر چاق کمتر از افراد چاق بخت ابتلا توام به بیماری های فوق را دارا هستند.

فعالیت بدنی متغیر موثر دیگری بر ابتلای هم زمان به بیماری دیابت و فشار خون، است. با در نظر گرفتن فعالیت بدنی به عنوان گروه مرجع، بخت افراد با عدم فعالیت بدنی در ابتلای هم زمان به بیماری های دیابت و فشار خون ۱/۸ برابر بخت افراد با فعالیت بدنی است. یعنی عدم فعالیت بدنی، بخت ابتلای هم زمان به دو بیماری فوق را هشتاد درصد افزایش می دهد.

متغیر دیگری که بر ابتلا به بیماری دیابت و فشار خون تاثیر دارد، سیگاری غیرفعال است به طوری که نسبت بخت ابتلای هم زمان به دو بیماری مورد نظر در گروه افراد سیگاری غیرفعال به سایرین برابر با ۱/۳۲ است. بنابراین افرادی که

جدول ۲- نتایج حاصل از برآورد مدل رگرسیون لجستیک دو متغیره

متغیر	سطوح	برآورد ضرایب	خطای معیار	نسبت بخت‌ها	فاصله اطمینان ۹۵٪ (ضرایب)	مقدار احتمال
				کران پایین	کران بالا	
سابقه فامیلی	دارد	۰/۱۵۹	۰/۰۲۴	۱/۳۲	۰/۱۱۲	<۰/۰۰۱
	ندارد			طبقه مرجع		
چاقی	خیر	-۰/۰۱۳	۰/۰۲۰	۰/۴۷	-۰/۰۵۳	۰/۰۴۵
	بلی			طبقه مرجع		
فعالیت بدنی	ندارد	۰/۱۲۱	۰/۰۱۹	۱/۸	-۰/۰۸۴	<۰/۰۰۱
	دارد			طبقه مرجع		
سیگاری	بلی			طبقه مرجع		
غیرفعال	خیر	-۰/۰۱۲	۰/۰۲۷	۰/۷۶	-۰/۰۶۵	۰/۰۳۵
	جامد			طبقه مرجع		
نوع روغن	سایر	-۰/۰۱۵	۰/۰۰۹	۰/۵۸	-۰/۰۳۲	۰/۰۲۹
	$\geq 55$			طبقه مرجع		
گروه سنی	<55	-۰/۰۶۳	۰/۰۱۱	۰/۳۶	-۰/۰۸۵	<۰/۰۰۱
	مرد			طبقه مرجع		
جنس	زن	۰/۰۰۱	۰/۰۱۸	۱/۱۴	-۰/۰۳۴	<۰/۰۰۱
مقدار ثابت		۱/۵۴۸	۰/۰۷۱	-----	۱/۱۸۹	<۰/۰۰۱

درپیش‌بینی عوارض بعد از عمل جراحی بیماران مبتلا به بیماری سرطانی پرداخته اند. طبق نتایج به دست آمده از این تحقیق، پیش‌بینی که توسط شبکه‌های عصبی مصنوعی حاصل شد به طور قابل توجهی بهتر از نتایج حاصل از رگرسیون لجستیک بود که با نتایج مطالعه جاری همخوانی دارد (۱۴). به عبارت دیگر می‌توان بیان کرد که شبکه‌های عصبی مصنوعی پیش‌بینی‌های بهتری از ابتلای توأم بیماری دیابت و فشارخون نسبت به رگرسیون لجستیک ارزیابی می‌دهند. گرین و همکاران در سال ۲۰۰۶ در مطالعه‌ای به منظور مقایسه پیش‌بینی شبکه‌های عصبی مصنوعی و رگرسیون لجستیک در بیماران مبتلا به سندرم کرونری در یافتند که سطح زیر منحنی ROC برای پیش‌بینی شبکه‌های عصبی مصنوعی در حدود ۸۰٪ است که اختلاف قابل توجهی با سطح زیر منحنی ROC برای پیش‌بینی رگرسیون لجستیک نشان داد (۱۲) و نتایج به دست آمده در مطالعه حاضر نیز تاییدی بر همین یافته را نشان داد.

همچنین در مطالعه جیمز و همکاران که در سال ۲۰۰۵ انجام گرفت سطح زیر منحنی ROC

زمان زنان به دو بیماری فوق، ۱۴ درصد بیش از مردان است.

بر اساس نمودار ROC، سطح زیرمنحنی برای مدل‌های رگرسیون لجستیک دو متغیره و شبکه عصبی مصنوعی به ترتیب  $0/78$  ( $p = 0/039$ ) و  $0/86$  ( $p = 0/046$ ) به دست آمد که بیانگر دقت پیش‌بینی بالاتر شبکه عصبی مصنوعی نسبت به مدل رگرسیون لجستیک در پیش‌بینی توأم بیماری‌های فشارخون و دیابت می‌باشد.

### بحث و نتیجه گیری

بیماری‌های دیابت و فشارخون، از جمله‌ی مهمترین مشکلات بهداشتی در جهان هستند که سهم قابل توجهی از منابع و امکانات بهداشتی کشورهای در حال توسعه را به خود اختصاص می‌دهند. در ایران نیز این بیماری‌ها شایع بوده و در ردیف بیماری‌های غیرواگیر قرار دارند (۱۳).

مطالعات متعددی در کشورهای مختلف در زمینه مدل‌سازی با استفاده از روش شبکه عصبی مصنوعی انجام شده است که در یکی از این مطالعات جفرسون و همکاران به مقایسه شبکه‌های عصبی مصنوعی با رگرسیون لجستیک

مذکور را کنترل نمود.

به منظور ارتقای دقت پیش بینی ابتلای همزمان به بیماری های دیابت و فشار خون بهتر است از متغیرهای مستقل دیگری مانند استرس، سیگاری مستقیم و... نیز در سایر مطالعات استفاده شود. همچنین پیشنهاد می شود که دقت پیش بینی شبکه های عصبی مصنوعی با توپولوژی های مختلف با روش های تحلیل بیزی مقایسه شوند. علاوه بر این استفاده از شبکه های عصبی مصنوعی در تحلیل های بقا و مقایسه دقت پیش بینی شبکه های عصبی مصنوعی بامدل های مختلف سری های زمانی (ARMA و ARIMA) نیز توصیه می گردد.

### تقدیر و تشکر

این مقاله حاصل طرح تحقیقاتی تحت عنوان مقایسه دقت پیش بینی شبکه های عصبی مصنوعی و رگرسیون لجستیک دو متغیره در تشخیص توأم بیماری دیابت و فشار خون، مصوب دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران در سال ۱۳۹۱ و با شماره ۱۹۳۰۳ می باشد که با حمایت دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران اجرا شده است.

### منابع

1. Regan M, Catalano P. Likelihood models for clustered binary and continuous outcomes Application to Developmental toxicology. *Biometrics*, 1999. 55: p. 760-768.
2. Sammel MD, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1997. 59: p. 667-678.
3. Kutner MH, Nachtsheim CJ, And Neter J. *Applied linear regression models*, 4th ed. New York: McGraw-Hill/Irwin; 2004.
4. Jobson DJ. *Applied Multivariate Data Analysis: Categorical and Multivariate Methods*, New York: Springer, 1992.
5. Lindsey JK, and Jones B. choosing among generalized linear models applied to medical data. *Stat Med* 1998; 17: 59-68.
6. Kay JW, Titterington DM, (eds). *Statistics and neural networks: Advanced at the interface*, Oxford:

محاسبه شده به ترتیب مقادیر ۰/۷۵، ۰/۸۷ برای رگرسیون لجستیک و شبکه های عصبی مصنوعی بود که حاکی از دقت بیشتر شبکه های عصبی مصنوعی در پیش بینی مرگ و میر نسبت به رگرسیون لجستیک دارد (۱۵).

قابل ذکر است که در مطالعاتی که به آنها اشاره شد حالت تک متغیره مد نظر بوده است، در حالیکه در مطالعه حاضر حالت دو متغیره ابتلای توأم به دو بیماری دیابت و فشار خون مورد مطالعه قرار گرفت که نهایتاً با توجه به سطح زیر منحنی ROC نشان داده شد که مدل شبکه های عصبی مصنوعی نسبت به مدل رگرسیون لجستیک دو متغیره مناسب تر است.

مطالعاتی هم در این زمینه در داخل کشور انجام گرفته است که می توان به مطالعه سدهی و همکاران اشاره نمود. در این مطالعه وضعیت ابتلا یا عدم ابتلا به سندرم متابولیک مورد بررسی قرار گرفت و مدل رگرسیون لجستیک وضعیت ۰/۷۲/۴، مدل تحلیل ممیزی وضعیت ۰/۶۶/۷، مدل شبکه ی عصبی (۱:۸:۱۵) وضعیت ۰/۷۳/۶ و در نهایت مدل شبکه عصبی (۱:۱۰:۱۵) وضعیت ۰/۸۷/۴ افراد را به درستی پیش بینی کردند (۱۶) که این نتیجه نیز با یافته های تحقیق جاری همخوانی دارد.

در مطالعه ی دیگر انجام شده توسط بیگلریان و همکاران بر روی بیماران مراجعه کننده با تشخیص قطعی سرطان معده که در بخش گوارش بیمارستان طالقانی که تحت عمل جراحی گرفته بودند، این نتایج به دست آمد که سطح زیر منحنی ROC برای شبکه های عصبی مصنوعی و رگرسیون کاکس به ترتیب برابر ۸۲/۶ درصد و ۷۵/۴ درصد بود. در نتیجه شبکه های عصبی مصنوعی نسبت به مدل رگرسیون کاکس پیش بینی های بهتری ارائه داد (۱۷).

در پایان می توان نتیجه گیری کرد که شبکه عصبی مصنوعی تعیین شده در این مطالعه دارای دقت پیش بینی بیشتری نسبت به رگرسیون لجستیک دو متغیره است که با استفاده از این روش می توان بهترین توپولوژی را جهت پیش بینی ابتلای توأم به این دو بیماری ارائه نمود و با اتخاذ روش های پیشگیرانه مناسب بیماری های

Oxford University Press, 1999.

7. Anderson J.A. An Introduction to Neural Networks. Cambridge, MA: MIT Press, 1995: p. 795-851.

8. Sadeghian S. [The knowledge of hospitalized patients about major risk factors of IHD in university hospitals of Tehran]. Daneshvar Medicine, 2001: p. 55-60.(Persian)

9. Fausett L. Fundamentals of Neural Networks Architectures, Algorithms and Applications. Prentice Hall, 1994.

10. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. The Lancet, 1995. 346(8982): 1075-1079.

11. Mitchell T. Machine Learning. McGraw Hill, 1997.

12. Green M, Bjork J, Forberg J, Ekelund U, Edenbrandt L, Ohlsson M. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artificial intelligence in medicine, 2006. 38(3): 305-318.

13. Preventing Chronic Diseases a Vital Investment. Cataloguing-in-Publication Data (World Health Organization).

14. Jefferson MF, Pendleton N, Lucas SB, Horan MA. Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with non-small cell lung carcinoma. Cancer, 1997. 79(7): 1338-42.

15. Jaimes F, Farbiarz J, Alvarez D, Martinez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. Crit Care, 2005. 9(2): R150-R156.

16. Sedehi M, Mehrabi Y, Kazemnejad A, Hadaegh F. [Comparison of Artificial Neural Network, Logistic Regression and Discriminant Analysis Methods in Prediction of Metabolic Syndrome]. Iranian Journal of Endocrinology & Metabolism 2010. Vol 11:39-646. (Persian)

17. Biglarian A, Hajizadeh E, Kazemnejad A. [Comparison of artificial neural network and Cox regression models in survival prediction of gastric cancer patients]. Koomesh, 2010. 11(3): 215-220.(Persian)

## The comparison of the predictive precision of artificial neural networks and bivariate logistic regression in diagnosis of patients with hypertension and diabetes

**Mehdi Adavi**, Msc in Biostatistics, Iran University of Medical Sciences, Tehran. Iran. mehdi.adavi@yahoo.com

**Masoud Salehi**, Assistant professor of Biostatistics, Iran University of Medical Sciences, Tehran. Iran. salehi74@yahoo.com

**\*Masoud Roudbari**, Associate Professor of Biostatistics, Iran University of Medical Sciences, Tehran, Iran (\*Corresponding author). mroudbari@yahoo.co.uk.

**Fereshteh Asgari**, MD, MPH, Senior Expert of Non-Communicable Diseases Risk Factors, Center for Non-Communicable Diseases Control and Prevention, Deputy for Health, Ministry of Health and Medical Education, Tehran, Iran. asgarifcdc@yahoo.com

**Ali Rafei**, MSc of Biostatistics, Research assistant, National Health Registries Program, Deputy for Research, Ministry of Health and Medical education, Tehran, Iran. rafeicdc@yahoo.com

### Abstract

**Background:** Diabetes and hypertension are from important non-communicable diseases in the world and their prevalence are very important for health authorities. The objective of this study was to compare the predictive precision of joint logistic regression (LR) and artificial neural network (ANN) in concurrent diagnosis of diabetes and hypertension.

**Methods:** This cross-sectional study was performed on 12000 Iranian people in 2013. The study questionnaire included some items on hypertension and diabetes and their risk factors. A perceptron ANN with two hidden layers was applied to data. The variables in the study were diabetes, hypertension, gender, type of cooking oil, physical activity, family history, age, obesity and passive smokers. To build a joint LR model, and ANN, SAS 9.2 and Matlab software were used. The ROC curve was used to find the higher accuracy model for predicting diabetes and hypertension.

**Results:** The variables of gender, type of cooking oil, physical activity, family history, age, passive smokers and obesity entered to the LR model and ANN. The odds ratio of affliction to both diabetes and hypertension is high in females, user of solid oil, people with no physical activity, with positive family history, age of equal to or higher than 55, passive smokers and obesity. The area under ROC curve for LR model and ANN were 0.78 ( $p=0.039$ ) and 0.86 ( $p=0.046$ ) respectively.

**Conclusion:** The best model for concurrent affliction to hypertension and diabetes is ANN which has higher accuracy than the joint LR model.

**Keywords:** Artificial neural network, Joint logistic regression, Diabetes, Hypertension.