مجله دانشگاه علوم پزشکی تربت حیدریه
دوره ۶. شماره ۲. تابستان ۱۳۹۷

مقاله پژوهشی

تشخیص بیماری دیابت با استفاده از شبکه عصبی مصنوعی و عصبی– فازی

ایمان ذباح[1*]، اسما اسکندی[2]، زهرا سرداری[2]، ابولفضل نوقنذی[2]

۱. گروه برق و کامپیوتر، دانشگاه تربت حیدریه، تربت حیدریه، ایران
۲. دانشگاه تربت حیدریه، تربت حیدریه، ایران

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

چکیده

زمینه و هدف: یکی از مشکلات اساسی بیماری دیابت عدم‌تشخیص به‌موقع و درمان صحیح آن است. مطالعه حاضر با هدف تشخیص بیماری دیابت با استفاده از روش مبتنی بر داده‌کاوی انجام شده است.

روش‌ها: این مطالعه از نوع تحلیلی بوده و پایگاه داده آن مشتمل بر ۷۶۸ نفر با ۸ ویژگی می‌باشد. در این پژوهش از شبکه‌های عصبی مصنوعی و عصبی– فازی جهت تشخیص بیماری دیابت و انجام محاسبات استفاده شد. تحلیل آماری با استفاده از نرم افزار SPSS 23 و برنامه نویسی در محیط نرم افزار MATLAB 2018 انجام شده است. به‌منظور حصول دقت واقعی از روش Kfold جهت تفکیک نمونه‌ها به دو دسته آموزش (Train) و آزمون (Test) استفاده گردید.

نتایج: خطای محاسبه‌شده بر اساس میانگین مربعات خطا (mean square error) در روش شبکه عصبی مصنوعی پرسپترون چندلایه (MLP) و شبکه عصبی بردار یاد گیر کوانتیزه (Learning Vector Quantization) و شبکه‌های عصبی–فازی (Nero fuzzy) به ترتیب ۹۸/۶٪ و ۹۸/۲٪ و ۹۹/۶٪ بدست آمد.

نتیجه‌گیری: با توجه به نتایج مطالعه، بنظر می رسد استفاده از مدل‌های مبتنی برداده کاوی می‌تواند بعنوان یک روش کمکی در تشخیص بیماری دیابت کارآمد باشد. اگرچه روشهای مورد مطالعه با دقت قابل قبول توانایی امکان پیش بینی بیماری دیابت را دارند اما نتایج مطالعه نشان می دهد که روش مبتنی بر عصبی فازی دقت بالاتری دارند.

کلید واژه‌ها:
دیابت، شبکه‌های عصبی مصنوعی، شبکه‌های عصبی– فازی، داده‌کاوی

تمامی حقوق نشر برای دانشگاه علوم پزشکی تربت حیدریه محفوظ است.

مقدمه

دیابت چهارمین علت مرگ‌و میر در بیشتر کشورهای توسعه‌یافته است (۱). بر اساس آمار فدراسیون جهانی دیابت (International federation of diabetes)، در سال ۲۰۱۲ بیش از ۳۷۱ میلیون نفر از مردم جهان مبتلابه دیابت بوده که هرسال نیز به آن افزوده می‌شود به‌گونه‌ای که بیش از نیمی از مبتلایان به دیابت از بیماری خود بی‌خبر هستند. برای بیماران دیابتی بیش از ٤٧٠ میلیارد دلار هزینه می‌شود این بیماری عامل مرگ ۵۰ میلیون نفر است. همچنین بر اساس سرشماری آماری که در سال ۲۰۱۳ صورت گرفته بیش از ٦ میلیون نفر ایرانی مبتلابه دیابت هستند (۲). این افزایش میزان شیوع با پیر شدن جمعیت، تغییرات سبک زندگی همراه با توسعه اقتصادی و افزایش میزان چاقی متناسب است (۳). در مطالعه‌ای که در سال ۲۰۱۰ منتشر شد نشان داد که شیوع دیابت در منطقه خاورمیانه به‌طور قابل‌توجهی تا سال ۲۰۳۰ افزایش خواهد یافت و برآورد می‌شود نرخ رشد سالانه دیابت تا سال ۲۰۳۰ در ایران بعد از پاکستان به رتبه دوم منطقه برسد (٤).

دیابت عمدتاً از طریق آزمایش‌های قند خون پلاسمای ناشتا تشخیص داده می‌شود که به سه دسته تقسیم می‌شود. دیابت نوع ۱ یا دیابت وابسته به انسولین (IDDM) بیشتر در سن پایین و کودکان دیده می‌شود. دیابت نوع ۲ یا دیابت غیر

*آدرس نویسنده مسئول: تربت حیدریه، دانشگاه آزاد اسلامی تربت حیدریه
آدرس پست الکترونیک: I.zabbah@torbath.ac.ir

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

وابسته به انسولین (NIDDM) در ۹۰٪ تا ۹۵٪ بیماران دیابتی مشاهده می‌شود (۵). در این دسته انسولین ترشح می‌شود ولی بدن در برابر مصرف انسولین از خود مقاومت نشان می‌دهد. دیابت نوع ۳ بیشتر در زنان باردار دیده می‌شود که عمدتاً پس از بارداری به دیابت نوع ۲ تبدیل می‌شود (٦). دیابت یک نگرانی فزاینده مهم بهداشت عمومی است (۷ و ۸) و باعث بروز عارضه‌هایی همچون ناراحتی‌های قلبی عروقی، نفروپاتی، نوروپاتی، رتینوپاتی و کاتاراکت و یکسری عوارض دیگر می‌شود (۹). برخی از مطالعات نشان می دهد که بیماری دیابت و عوارض آن ازجمله عوارض چشمی، قلبی، عروقی، کلیوی توسط رژیم غذایی سالم، فعالیت بدنی منظم، کنترل قند خون، کنترل فشارخون و کنترل کلسترول قابل‌پیشگیری است (۱۰). شیوع دیابت نوع ۱ و ۲ در سراسر جهان رو به افزایش است ولی سرعت افزایش دیابت نوع ۲ بیشتر از دیابت نوع ۱ است. عوامل این افزایش می‌تواند به دلیل تغییر شیوه‌ی زندگی، شیوه‌ی چاقی و کاهش میزان فعالیت بدنی باشد (۱۱). با آموزش صحیح در مورد رفتارهای خود مراقبتی می‌توان ۸۰ درصد عوارض بیماری را کاهش داد (۱۳). از طرفی خود مراقبتی علاوه بر ارتقاء کیفیت زندگی باعث کاهش هزینه‌ها می‌گردد و با پیگیری آن از عوارض حاد و مزمن بیماری‌ها پیشگیری کرد و یا بروز آن را به تعویق انداخت (١٤).

امروزه ابزارهای داده‌کاوی به‌طور گسترده در علم پزشکی مورداستفاده قرار می‌گیرد. داده‌کاوی راهی است برای تحلیل اتوماتیک داده‌ها و شناسایی الگوهای پنهان که انجام این امر به‌صورت دستی ممکن نیست (۱۵). داده‌کاوی می‌تواند در پیش‌بینی و تشخیص سریع و کم‌هزینه بیماری‌ها به‌طور مؤثری استفاده شود (١٦). ازآنجایی‌که شبکه‌های عصبی-مصنوعی به‌عنوان یک جعبه سیاه توانایی بالایی در مدل‌سازی مسائل غیرخطی دارند لذا موردتوجه بسیاری از محققین حوزه داده‌کاوی پزشکی قرارگرفته است (۱۷). در سال‌های اخیر استفاده روزافزون از داده کاوی درزمینه تشخیص پزشکی کاملاً مشهود است. به عنوان نمونه می توان به مطالعه آقای فانگ و همکاران اشاره کرد که از خوشه بندی و رگرسیون به منظور تشخیص بیماری دیابت استفاده نمودند(۲۵) . در مطالعه مشابه دیگری آقای الجوما و همکاران از همین روش رگرسیون به پیش بینی درمان دیابت دردو دسته گروه سنی جوان و پیر بر اساس نوع درمان پرداختند(۳٤). آنتونیلی و همکاران روش خوشه بندی چندسطحی را مورد استفاده قرار دادند که نتایج حاصل کمک شایانی به شناسایی مسیر درمان بیماران دیابتی کرد(۳۰). آنبانان با استفاده از درخت تصمیم و شبکه عصبی مصنوعی به پیش بینی وجود دیابت در بیماران پرداختند(۳۵). استفاده از سایر روشهای داده کاوی مانند روش SVM در تشخیص بیماری دیابت در مطالعه آقای اسلم (۳٦) و بکارگیری درخت تصمیم در مطالعه آقای گاندی و همکاران با هدف تشخیص بیماری دیابت مشاهده می شود(۳۷) . در این مطالعه سعی شده است با استفاده از یکی از متدهای نوین داده کاوی که مبتنی بر روش شبکه های مصنوعی عصبی فازی است، به بررسی و تشخیص بیماری دیابت پرداخته شود. سیستمهای فازی معمولاً از قوانین باقابلیت تفسیر زبانی که از جملات IF-THEN تشکیل‌شده‌اند، استفاده می‌کند. این سیستم‌ها به علت دارا بودن توابع عضویت با درجات تعلق دقیق، توانایی محدودی در کاهش اثر عدم قطعیت در قوانین فازی دارند و لذا می‌توانند با دقت مطلوبی به تشخیص بیماری دیابت کمک کنند.

روش‌ها

تحقیق حاضر یک مطالعه تشخیصی است که به پیش‌بینی بیماری دیابت می‌پردازد. داده‌های مورداستفاده در این مقاله از مجموعه داده‌های مربوط به دیابت تحت عنوان PID (Pima Indian diabetes) که به عنوان یک مجموعه داده استاندارد دیابت جمع‌آوری‌شده و در سایت اینترنتی دانشگاه کالیفرنیای آمریکا قابل‌دسترس است. این نمونه‌ها به‌عنوان مرجعی جامع جهت بررسی الگوریتم‌های یادگیری ماشین در بسیاری از پژوهش‌ها استفاده می‌شود و در سال ۱۹۹۸ جمع‌آوری‌شده و در مرجع (۱۹) قابل‌دسترس است. نمونه‌ها دارای ۸ شاخصه

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

مربوط به ۵۰۰ زن سالم و ۲۶۸ زن مبتلابه بیماری دیابت و با حداقل سنی ۲۱ می‌باشد که کاملاً مطابق با شاخص‌ها و استانداردهای سازمان بهداشت جهانی جمع‌آوری گردیده است. برای انجام این تحلیل، از شبکه عصبی پرسپترون چند لایه و شبکه عصبی- فازی با استفاده از نرم‌افزار MATLAB 2016 استفاده‌شده است.

با توجه به آنکه ورود داده‌ها به شکل خام باعث کاهش سرعت و دقت یادگیر می‌شود، برای پیشگیری از چنین شرایطی و به لحاظ یکسان کردن ارزش ویژگی‌ها، نرمالیزه کردن داده‌ها قبل از اعمال به شبکه انجام شده است. به‌عنوان مثال طبق جدول ۱ در مورد ویژگی گلوکز پلاسما میانگین نمونه‌ها ۱۲۰ است درحالی‌که در مورد ویژگی سابقه دیابت میانگین نمونه‌ها ۰/۴۷ است. این فاصله بین ویژگی‌ها باعث می‌شود که در صورت عدم نرمال‌سازی علاوه بر کاهش دقت یادگیرها سرعت همگرایی کاهش یابد. عمل نرمال‌سازی داده‌ها مطابق رابطه (۱) انجام گرفته است:

$$xn = \frac{x - x_{min}}{x_{max} - x_{mhn}} \quad (۱)$$

در این رابطه، X نمونه‌های نرمال نشده، X_n داده‌های نرمال شده و X_{max} و X_{min} حداقل و حداکثر ورودی می‌باشند. معمولاً دو معیار مختلف به‌منظور ارزیابی کارایی هر شبکه و توانایی آن برای پیش‌گویی دقیق استفاده می‌شود. مجذور خطا (*RMSE*) و ضریب همبستگی (R^2) که نشانگر میزان بازده شبکه می‌باشد، مطابق رابطه (۲) محاسبه می‌شوند.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N}} \quad (۲)$$

$$R2 = 1 - \frac{\sum(Y_I - \hat{Y}_I)^2}{\sum(-\bar{Y})^2} \quad (۳)$$

در این رابطه y_i نتایج مشاهداتی، $\hat{y}_i$ نتایج محاسباتی و N تعدادکل مشاهدات است. *RMES* اختلاف این دو مقدار را نشان می‌دهد. کمترین مقدار *RMES*، بالاترین صحت پیشگویی را نشان می‌دهد لذا هرچه این مقدار به صفر و R^2 به یک نزدیک‌تر شود، مدل ارائه‌شده به واقعیت نزدیک‌تر است.

انتخاب داده‌های آموزش و آزمون به منظور تشخیص بیماری دیابت

به منظور افزایش دقت اندازه‌گیری خطا از روش kfold استفاده‌شده است. به‌این‌ترتیب که داده‌های آموزش بدون جایگذاری به k قسمت تقسیم‌شده و از k-1 نمونه به‌عنوان داده آموزش و از بقیه به‌عنوان داده اعتبار سنجی استفاده می‌شود. به عبارت دقیق‌تر پایگاه داده PID شامل ۷۶۸ نمونه است؛ که به ۵ قسمت (fold) با ابعاد ۱۶۳ نمونه تقسیم‌شده است. در هر مرحله ٤ قسمت به‌عنوان داده آموزش و ۱ قسمت برای داده آزمون مورداستفاده قرارگرفته و صحت شبکه سنجیده می‌شود. این عملیات در یک حلقه ۵ بار تکرار و در مجموع ۲۵ بار عملیات آموزش و آزمون با نمونه‌ها انجام و در پایان میانگین دقت ۲۵ بار تکرار محاسبه می‌شود.

مدل‌سازی به کمک شبکه عصبی پرسپترون چند لایه MLP و شبکه LVQ در مرحله اول و همچنین مدل‌سازی با استفاده از شبکه عصبی فازی در مرحله دوم این پژوهش استفاده شده است. یکی از کارآمدترین ساختارهای پیشنهادی برای استفاده در مدل‌سازی عصب‌های واقعی، مدل پرسپترون چندلایه می‌باشد که از یک‌لایه ورودی، یک یا چند لایه پنهان و یک‌لایه خروجی تشکیل یافته است. در این ساختار، تمام نرون‌های یک‌لایه به تمام نرون‌های لایه بعد متصل هستند. در این شبکه از دو نوع سیگنال متفاوت استفاده می‌شود. نوع اول سیگنال-هایی که بر اساس ورودی‌های هر نرون و پارامترهای وزن و تابع محرک نظیرش محاسبه می‌شوند و نوع دوم سیگنال‌های خطا که با برگشت از لایه خروجی و منشعب شدن به لایه‌های پنهانی دیگر محاسبه می‌شوند. تعداد نرون‌های لایه پنهان با سعی و خطا به دست می‌آید.

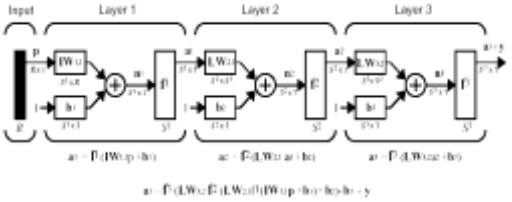
شبکه عصبی LVQ نیز به طور گسترده در حوزه کلاس بندی دیتا مورد استفاده محققین داده کاوی قرار می گیرد. این شبکه دارای دولایه رقابتی و خطی می‌باشد. لایه رقابتی دسته‌بندی کردن بردارهای ورودی را یاد گرفته و درنهایت لایه خطی کلاس‌های لایه رقابتی را به دسته‌های هدف که توسط کاربر تعیین‌شده نگاشت می‌کند. شکل ۱- الف ساختار شبکه عصبی

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

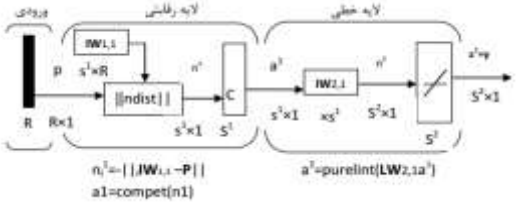
پرسپترون و شکل ۱- ب معماری شبکه عصبی LVQ مورداستفاده در این پژوهش را نشان می دهد.

از آنجایی‌که معماری شبکه عصبی پرسپترون چند لایه ازنظر تعداد نرون‌ها در لایه‌ها می‌تواند درنتیجه مدل‌سازی نقش اساسی داشته باشد، لذا حدود ۱۰۰ شبکه MLP و LVQ مورد بررسی قرار گرفت تا بهترین معماری به دست آید. ورودی‌های مدل عبارت است از: تعداد دفعات وضع حمل، غلظت گلوکوز پلاسمای خون، فشار دیاستولیک (mmHg)، ضخامت پوست (mm)، انسولین سرم دو ساعته (mu U/ml)، شاخص جرم بدن، داشتن سابقه دیابت، سن و پارامتر بیمار یا سالم بودن به‌عنوان خروجی شبکه می‌باشد.

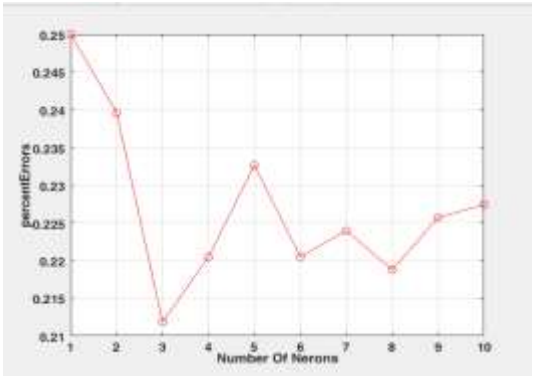
معماری‌های متفاوتی به‌منظور دستیابی به بهترین ساختار شبکه عصبی مورد آزمایش قرار گرفت. شکل ۲- الف و شکل ۲- ب به ترتیب خطای محاسبه‌شده در یک دوره از معماری‌های مختلف شبکه MLP و LVQ را نشان می‌دهد.



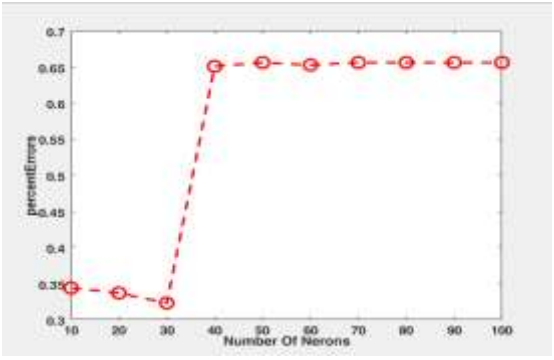
شکل ۱-الف. معماری شبکه عصبی پرسپترون ۳ لایه مورداستفاده



شکل ۱-ب. معماری شبکه عصبی LVQ مورد استفاده



شکل ۲- الف. خطای معماری های مختلف شبکه MLP به ازای تغییرات تعداد نرون در لایه مخفی



شکل ۲- ب. خطای معماری‌های مختلف شبکه عصبی LVQ به ازای تغییر در لایه رقابت

در شبکه عصبی MLP نوع آموزش از نوع پس انتشار خطا و نرخ یادگیری ۰/۰۵ و معماری شبکه ۱-i-j-۸ می‌باشد که i و j متغیر بین ۱۰ تا ۱۰۰ انتخاب شده اند.

شکل ۲- الف نشان می دهد که در شبکه MLP کمترین خطا مربوط به زمانی (Feed Forward Back propagation) است که تعداد نرون‌ها در لایه پنهان اول ۱۰ و لایه دوم ۳۰ است در این حالت میزان متوسط خطای کمترین مربعات (Mean Square Error) برای داده‌ها حدوداً برابر با ۰/۲۱ بوده است. شکل۲-ب نشان می‌دهد که تغییرات تعداد نرون‌ها در لایه رقابتی از نرون ۴۰ به بعد تفاوت چندانی در تشخیص بیماری دیابت در شبکه عصبی LVQ ندارد؛ و بهترین حالت آن در لایه رقابتی ۳۰ نرون با خطای کمتر از ۰/۳۵ است.

در حال حاضر به دلیل رویکردی که منطق فازی در محاسبات نرم دارد در بسیاری از شاخه‌های علوم اجتماعی و علوم مهندسی و پزشکی موردتوجه محققین قرارگرفته است (۲۲). اگرچه شبکه‌های عصبی مصنوعی و سیستم‌های فازی ازنظر ساختاری متفاوت هستند، اما درواقع این دو سیستم دارای ماهیت مکمل نسبت به یکدیگرند. سیستم‌های فازی، برای ایجاد نگاشت غیرخطی از ورودی به خروجی مورد استفاده قرار می‌گیرند و بر پایه مجموعه‌ای از قوانین استوارند که با این

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

قوانین بر اساس دانش بشری و با بهره‌گیری از تجربیات افراد خبره درزمینه موردنظر استنتاج می‌شوند. ویژگی مهم نگرش فازی عدم نیاز به فرمول ریاضی است، لذا در مورد طراحی سیستم‌هایی که نتوان عملکرد آن‌ها را به‌صورت مدل ریاضی بیان کرد مانند پیش‌بینی بیماری دیابت بسیار مفید و مناسب به نظر می‌رسند. به طورکلی آنچه یک سیستم فازی انجام می‌دهد، تبدیل دانش بشری در قالب مجموعه‌های فازی به یک فرمول ریاضی باقابلیت غیرخطی است .پارامترهای موجود در سیستم‌های مذکور شامل وزن‌های ارتباطی برای مقادیر ورودی، وزن‌های ارتباطی برای مقادیر بایاس و وزن‌های ارتباطی برای مقادیر خروجی با استفاده از تجربیات طراح سیستم تعیین می‌شوند. سیستم نرو فازی حاصل ترکیب شبکه عصبی با منطق فازی است که با استفاده از الگوریتم آموزش شبکه عصبی، پارامترهای سیستم فازی را تعیین می‌کند. این مدل یک سیستم فازی را در ساختاری عصبی اجرا می‌کند و برای فرایند آموزش از ترکیبی از روش‌های آموزش، شامل شیب نزولی و حداقل مربعات خطا استفاده می‌کند (۲۴-۲۵). مدل شبکه عصبی- فازی ارائه‌شده در این مطالعه شامل ۵ لایه است که اولین لایه نشان‌دهنده ورودی یعنی پارامترهای بالینی بیماری دیابت است. لایه‌های دوم و سوم نشان‌دهنده لایه‌های پنهانی هستند که شامل قواعد فازی است. لایه چهارم لایه خروجی و درنهایت لایه پنجم متغیر خروجی نهایی که دیابتی بودن فرد را مشخص می‌کند، می‌باشد. در مطالعه حاضر برای طراحی شبکه عصبی فازی از ۸ متغیر مندرج در جدول ۱ و یک متغیر خروجی استفاده‌شده است. فازی سازی (فرایند فازی کردن تمام ورودی‌ها و خروجی‌ها) و تعیین تابع تعلق در مرحله بعدی مطالعه انجام‌شده است. در این سیستم دو متغیر خروجی Ture و False برای افراد سالم و بیمار وجود دارد و الگوریتم آموزش هیبرید برای تعیین پارامترهای شبکه عصبی- فازی استفاده شده است. نحوه آموزش شبکه عصبی به‌صورت یادگیری با ناظر است. بنابراین، هدف آن است که با آموزش شبکه، توابع ناشناخته بیان‌شده توسط داده‌های آموزشی را دریافت کند و یاد بگیرد و بتواند متناسب باارزش پارامترهای ورودی، خروجی را تعیین کند. شبکه انفیس به کار گرفته‌شده در این پژوهش دارای تابع عضویت از نوع "linear" در لایه خروجی و تابع عضویت از نوع "gaussmf" در لایه ورودی می‌باشد که طبق رابطه ٤ محاسبه می‌شود. رابطه (٤):

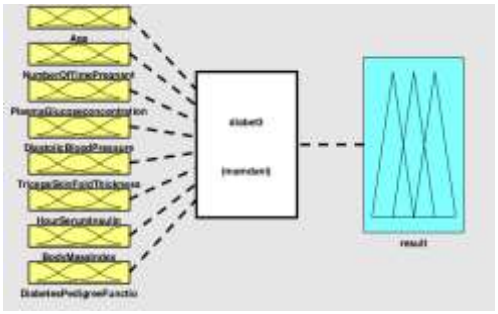
$$f(x,\sigma,c)=\exp\left(\frac{-(x-c)^2}{2\sigma^2}\right)$$

پارامترهای و c تعیین‌کننده شکل تابع می‌باشند. جهت آموزش شبکه عصبی- فازی دو روش گرادیان نزولی و روش ترکیبی وجود دارد که در این مطالعه از روش گرادیان جهت آموزش شبکه و از روش حداقل مربعات خطا جهت محاسبه دقت استفاده‌شده است. شکل ۳-الف سیستم انفیس طراحی‌شده را نشان می‌دهد. توابع عضویت برای داده‌های خروجی و ورودی تعریف می‌شوند. در این مطالعه برای هریک از ویژگی‌ها یک متغیر زبانی بین ۲ تا ٦ تابع عضویت در نظر گرفته‌شده است. برای سایر متغیرها نیز به همین ترتیب تابع عضویت ساخته‌شده است. به عنوان نمونه شکل ۳-ب متغیر شاخص توده بدنی را نشان می‌دهد زمانی که شامل ۳ تابع عضویت "کم"، "متوسط" و "زیاد" است.

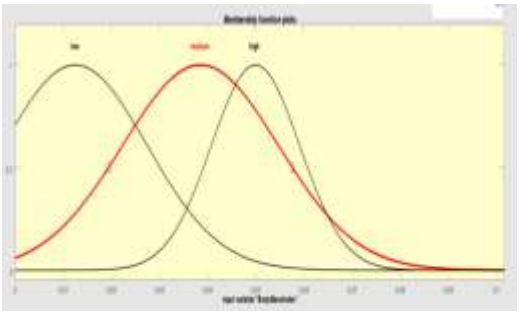
به‌منظور افزایش دقت از روش kfold استفاده شده است و همانند بخش شبکه‌های عصبی نمونه‌های آموزش به ٥ قسمت تقسیم و پس از ۱۰ بار تکرار میانگین خطا محاسبه شد. شکل ٤ متوسط خطا در هر تکرار را نشان می‌دهد.

در این قسمت سیستم استنتاج فازی- عصبی تطبیقی (ANFIS) با توجه به داده‌های مندرج در جدول ۱ طراحی شده است. در شکل ٥ (الف) خروجی واقعی و خروجی مدل پیش‌بینی کننده بیماری دیابت، (ب) خطای RMSE و (ج) هیستوگرام خطا را نشان می‌دهد.

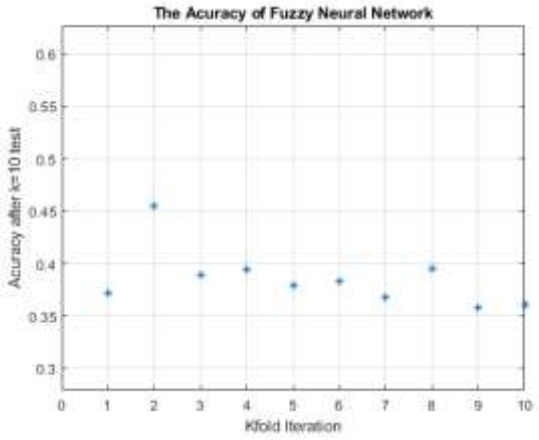
Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019



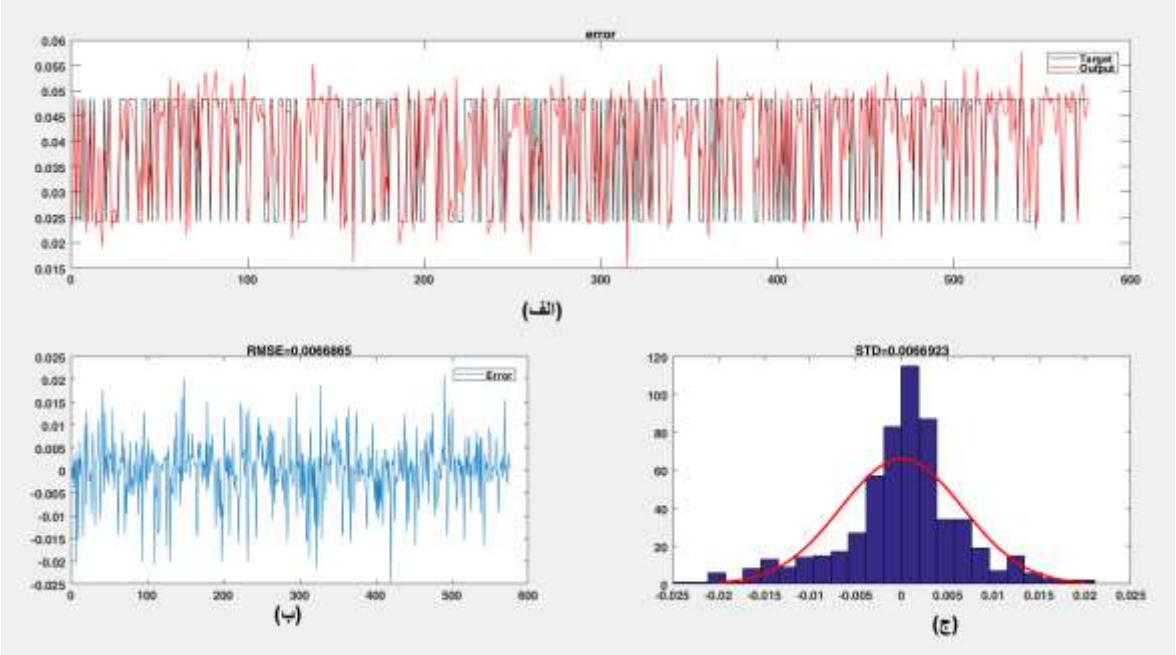
شکل ۳- الف. سیستم انفیس طراحی‌شده با ورودی های مشخص



شکل ۳- ب. تخصیص متغیر زبانی به متغیر شاخص جرم بدن



شکل ۴. میانگین خطای شبکه نروفازی در هر تکرار به روش kfold



شکل ۵. مقایسه خروجی مدل فازی ارائه‌شده با خروجی واقعی در تشخیص بیماری دیابت

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

نتایج

مطالعه حاضر بر روی مجموعه داده PID دانشگاه کالیفرنیا مشتمل بر ۸ مؤلفه بالینی مربوط به ۵۰۰ زن سالم و ۲۶۸ زن مبتلابه بیماری دیابت و با حداقل سنی ۲۱ می‌باشد، انجام‌شده است. در این پژوهش به کمک شبکه عصبی و فازی عصبی به تشخیص بیماری دیابت و مدل‌سازی و مقایسه نتایج آن پرداخته‌شده است. آماره‌های توصیفی بیماران در جدول ۱ نمایش داده شده است.

جدول ۱. آمارهای توصیفی از جامعه آماری ۷۶۸ نفر

ویژگی‌ها	میانگین	انحراف معیار
تعداد دفعات وضع حمل	۳/۸	۳/۳۶
غلظت گلوکوز پلاسمای خون	۱۲۰/۰۴	۳۱/۹۷
فشاردیاستوک(mm Hg)	۶۸/۸	۱۹/۳۵
ضخامت پوست ماهیچه سه سربازویی(mm)	۲۰/۵۸	۱۵/۹۵
انسولین سرم دو ساعته (mu U/ml)	۷۹/۸۸	۱۱۵/۲۴
شاخص جرم بدن	۳۱/۸۹	۷/۸۸
داشتن سابقه دیابت	۰/۴۷	۰/۳۳
سن	۳۳/۱۸	۱۱/۷۶

در این مطالعه ابتدا از دو نوع شبکه عصبی LVQ و MLP جهت دست‌یابی به بالاترین درصد دقت در تشخیص بیماری دیابت استفاده شده است. دقت حاصله در بهترین معماری در شبکه MLP و LVQ به ترتیب ۹۸/۶٪ و ۹۸/۲٪ برای داده‌های آزمون محاسبه گردید. با ترکیب سیستم‌های فازی و شبکه‌های عصبی مصنوعی، قابلیت‌های یادگیری وارد مرحله جدیدی گردید.

سیستم ANFIS طراحی‌شده به همراه ۸ شاخصه بیماری، توابع عضویت، هم برای داده‌های خروجی و هم داده‌های ورودی تعریف گردید. هر شاخصه بیماری شامل ۳ تابع عضویت "کم"، "متوسط" و "زیاد" است. پس از آموزش شبکه دقت ۹۹/۶۳ حاصل گردید. نتایج حاصل در جدول ۲ نشان داده شده است.

اگرچه تفاوت چشمگیری در صحت عملکرد مدل‌های مختلف مشاهده نمی‌شود، لیکن ازآنجاکه در تشخیص بیماری مسئله جان انسان مطرح است لذا بهبود پیش‌بینی حتی به‌اندازه یک درصد، اهمیت پیدا می‌کند. در پژوهش حاضر به‌طورکلی روش داده‌کاوی نظارت‌شده مورد ارزیابی قرار گرفت. درروش با ناظر شبکه‌های عصبی مصنوعی و شبکه‌های عصبی فازی به‌منظور دسته‌بندی افراد به دودسته سالم و بیمار استفاده گردید. عملکرد یادگیرهای مختلف (یادگیر عصبی و یادگیر عصبی فازی) بر روی پایگاه داده‌ی استاندارد انجام گرفت. انتظار می‌رود تشخیص بیماری دیابت با استفاده از سیستم‌های نر و فازی عملکرد بهتری نسبت به مدل عصبی داشته باشد.

جدول ۲. مقایسه صحت عملکرد شبکه‌های عصبی و فازی عصبی

نوع مدل	صحت عملکرد
پرسپترون چند لایه	٪۹۸/۶
LVQ	٪۹۸/۲
نرو فازی	٪۹۹/۶۳

بحث

در مطالعه حاضر سعی شده تا با مدل‌سازی مبتنی بر شبکه عصبی مصنوعی و عصبی- فازی به پیش بینی بیماری دیابت پرداخته شود تا از این طریق به حذف یا کاهش خطای انسانی درروند پیش‌بینی بیماری کمک شود. این مدل‌ها پیش‌بینی دقیقی از وجود یا عدم وجود بیماری ارائه می‌دهد و بیماران در همان مراحل اول آگاه شده و اقدامات لازم را جهت کنترل این بیماری انجام دهند. به دلیل ماهیت بیماری دیابت، ممکن است بر اساس خطاهای انسانی، پزشک در بعضی موارد نتواند بیمار دیابتی و یا بیماری که شرایط مبتلا شدن به دیابت را دارد شناسایی کند. پژوهش‌های مبتنی در داده کاوی پزشکی می‌توانند به پزشکان در تشخیص بیماری‌های مختلف ازجمله تشخیص دیابت کمک کند. در این بین شبکه‌های عصبی مصنوعی و بخصوص شبکه‌های عصبی فازی قابل توجه هستند. شبکه‌های عصبی مصنوعی در بسیاری از زمینه‌های پزشکی ازجمله سرطان و بیماری‌های قلبی، پیش‌بینی بقا، تحلیل تصاویر پزشکی و... به کار گرفته شده‌اند (۲۱).

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

تا کنون روش‌های بسیاری به‌منظور تشخیص بیماری دیابت انجام‌شده است. از این میان می‌توان به کار آقای سیتی و همکاران در استفاده از RBF (۲۲) و یا کار آقای تهامی در استفاده از شبکه‌های عصبی MLP (۲۳) اشاره کرد که به ترتیب به دقت‌های ۷۳.۳۲٪ و ۷۶.۸۹٪ دست یافتند. فانگ در سال ۲۰۰۹ با استفاده از تکنیک‌های مختلف داده‌کاوی بیماران را بر اساس مبتلا بودن به دیابت، خوشه‌بندی کرده است. ویژگی‌هایی که در این مدل‌ها مهم شناخته شدند عبارت‌اند از سن، سابقه خانوادگی و وزن. دقت مدل ایجادشده با استفاده از خوشه‌بندی ۸۰ ٪ است (۲٤). هوانگ و همکاران در سال ۲۰۰۷ تحقیقی بر روی شناسایی عوامل عمده تأثیرگذار بر کنترل دیابت، با به کار بستن انتخاب ویژگی‌ها Feature Selection در سیستم مدیریت بیمار انجام دادند (۲۵).

مطالعه فابیان و همکاران در سال ۲۰۰۵ با عنوان مقایسه شبکه‌های عصبی مصنوعی و رگرسیون لجستیک برای مرگ‌ومیر بیماران مبتلابه سپس انجام گرفت. در این مطالعه سطح زیر منحنی راک برای رگرسیون لوژستیک و شبکه‌های عصبی مصنوعی به ترتیب مقادیر ۱۷/۷۵٪ و ۸۲/۸۷٪ به دست آمد که این مطالعه نشان از دقت بیشتر شبکه‌های عصبی مصنوعی نسبت به رگرسیون لجستیک دارد (۲٦). برفه فرزانه (۱۳۹٤) با مطالعه بر روی ۱۳٤۲۳ نفر از شرکت‌کنندگان بالای ۲۵ سال که هیچ‌کدام دیابت کنترل‌شده‌ای نداشتند و با استفاده از مدل شبکه عصبی مصنوعی سه لایه با ساختار ۵۳:۲۰:۲ مدلی ارائه دادند که به‌دقت سطح زیر منحنی راک ۷۲/۷٪ و صحت پیش‌بینی آموزش ۹۲ درصد و صحت پیش‌بینی آزمون ٦/۹۱٪ دست‌یافت. آنان نتیجه گرفتند که با توجه به عدم نیاز مدل شبکه عصبی مصنوعی به‌پیش فرض‌های معمول روش‌های کلاسیک آماری و صحت پیش‌بینی، شبکه‌های عصبی از آماری بالاتر است (۲۷). اسما الجراح از درخت تصمیم برای تشخیص بیماری دیابت نوع دو استفاده کرده است. با به کار بردن طبقه‌بندی الگوریتم درخت تصمیم j٤٨ بروی داده‌ها در نرم‌افزار، Weka درخت تصمیم تولید شده است (۲۸). آنتونلی و همکاران در سال ۲۰۱۳ یک چارچوب تجزیه‌وتحلیل مبتنی بر خوشه‌بندی چند سطحی برای شناسایی مسیرهای درمان و معاینه بیماران دیابتی ارائه کرده‌اند. پایگاه داده موردمطالعه شامل ۸۵٦ رکورد از بیماران یکی از بیمارستان‌های انگلستان بوده و روش پیشنهادی در شناسایی گروه‌های بیماران با تاریخچه بیماری مشابه و افزایش شدت عوارض آن‌ها به‌خوبی عمل کرده است (۲۹). در سال ۲۰۰۹ سانتی و همکاران با استفاده از روش ماشین بردار پشتیبان توانستند با دقت ۳۸٪ بیماری دیابت نوع دوم را در افراد تشخیص دهند. داده‌های آن‌ها مربوط به ۷٦۸ بیمار می‌شد و از هشت متغیر ازجمله فشارخون افراد و میزان انسولین تزریقی برای پیش‌بینی خود استفاده نمودند (۳۰).

در برخی دیگر از مطالعات مدل‌سازی به‌وسیله ترکیب شبکه‌های عصبی و فازی انجام‌شده است که دراین‌بین می‌توان به پژوهش محمد فیوضی و همکاران با ترکیب روش سیستم فازی، درخت تصمیم(D-T)، الگوریتم‌های کاوشی(BPSO)، الگوریتم‌های طبقه‌بندی (SVM)، عصبی- فازی(ANFIS) به‌دقت ۹۵/۱٪ رسیده‌اند (۵). آنچه در مطالعه حاضر مشهود است و نقطه قوت آن نسبت به مطالعات مشابه است، استفاده از روش Kfold در تفکیک نمونه‌های آموزش و آزمون می‌باشد. در بسیاری از تحقیقات مشابه در فرایند مدل‌سازی، مدلی ارائه‌شده است که صرفاً حاصل بهترین عملکرد سیستم در مرحله آموزش بوده است. لیکن در این مطالعه تلاش شده، ضمن مدل‌سازی بیماری دیابت و طراحی مدل پیش‌بینی کننده مناسب، میانگین خطاها با توجه به تفکیک نمونه‌های آموزش و تست حداقل در ۵۰ بار جداسازی آن‌ها از یکدیگر حاصل شود. به‌این‌ترتیب دقت‌های به‌دست‌آمده نسبت به دقت‌های حاصله در مطالعات مشابه ضمن بالاتر بودن به واقعیت نزدیک‌تر است. چراکه میانگین خطای شبکه‌ها ثبت‌شده است.

نتیجه‌گیری

بر اساس نتایج حاصله از جدول ۲، دقت ارائه‌شده در این پژوهش در مقایسه با سایر پژوهش‌های انجام‌شده بالا و

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

قابل‌قبول بوده و می‌تواند در طراحی مدل‌های مناسب جهت پیشگویی امکان ابتلای افراد به بیماری های دیابت استفاده شود. همچنین می‌تواند در برنامه‌های غربالگری جهت شناسایی افراد در معرض خطر استفاده شود. همچنین ازآنجایی‌که داده‌های مورداستفاده در این مطالعه استاندارد بوده و در بسیاری از مطالعات داده‌کاوی بیماری دیابت مورداستفاده قرار می‌گیرد، الگوریتم مورداستفاده در این پژوهش می‌تواند به‌عنوان مبنای کار مطالعات آتی خصوصاً در مورد داده‌های بومی در پژوهش‌های آینده موردتوجه قرار گیرد. استفاده از روش‌های نوین داده‌کاوی می‌تواند به تشخیص سریع‌تر و دقیق‌تر بیماری دیابت کمک کند. این سیستم‌ها می‌توانند همانند یک دستیار پزشک در خدمت پزشک متخصص جهت تشخیص بیماری قرار بگیرد.

تشکر و قدردانی

این مقاله برگرفته از طرح مصوب شورای پژوهشی دانشگاه تربت حیدریه می باشد.

تضاد منافع

در این پژوهش هیچ گونه تعارض منافعی توسط نویسندگان گزارش نشده است.

References

1. Nazarzadeh M, Bidel Z, Sanjari Moghaddam A. Meta-analysis of diabetes mellitus and risk of hip fractures small study effect. Osteoporos Int (2016) 27: 229.

2. Janahmadi Z, Nekooeian AA, Mozafari M. Hydroalcoholic extract of Allium eriophyllum leaves attenuates cardiac impairment in rats with simultaneous type 2 diabetes and renal hypertension. Research in pharmaceutical sciences. 2015; 10(2):125.

3. American Diabetes A. Economic costs of diabetes in the U.S. in 2012. Diabetes Care. 2013;36(4):1033-46.

4. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010;87(1):4-14.

5. Naim Abadi, Mohammad Reza Ahmadi Chamachar, Noshas Tahami, Ehsan Rabbani, Hossein. Diagnosis of diabetes using the 14th Iranian Student Conference on Electrical Engineering, Kermanshah, Iran, Kermanshah University of Technology, pp. 251-258.persian

6. American Diabetes Association, Diabetes, Basics www.diabetes.org/diabetes-basics, (08/29/2012).

7. Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract. 2011; 94(3):311-21.

8. Guariguata L. By the numbers: new estimates from the IDF Diabetes Atlas Update for 2012. Diabetes Res Clin Pract. 2012; 98(3):524-5.

9. Ahmadi A, Hasanzadeh J, Rahimi M, Lashgari L. Factors af –fecting the quality of life in patients with type 2 diabetes Chahar Mahal Bakhtiari. J North Khorasan Univ Med Sci. 2011;3(1):13-7.

10. Shi L, Shu XO, Li H, Cai H, Liu Q, Zheng W, et al. Physical activity, smoking, and alcohol consumption in association with incidence of type 2 diabetes among middle-aged and lderly Chinese men. PLoS One. 2013;8(11).

11. Corriere M, Rooparinesingh N, Kalyani RR. Epidemiology of diabetes and diabetes complications in the elderly: an emerging public health burden. Curr Diab Rep. 2013;13(6):805-13

12. Arastoo A, Ghassemzadeh R, Nasseh H, Kamali M, Rahimi FA, Arzaghi M, et al. [Factors Affecting Quality of life in Elderly Diabetic Residents of the kahrizak Geriatric Nursing Home of Tehran]. Iran J Endocrin Metab. 2012;14(1):18-24.

13. Shirazi M, Anoosheh M, Sabohi F. Barriers of diabetes self-care education: viewpoint of patients and nurses. J Diabet Nurs. 2014;2(2):63-76.

14. Bidi F, Hassanipour K, Ranjbarzadeh A, Arab KA. Effectiveness of Educational Program on Knowledge, Attitude, Self Care and Life Style in patients with type II diabetes. J Sabzevar Uni Med Sci.2013;19(4):336-44.

15. Elsappagh S, Elmogy M, Riad AM. A fuzzy ontology oriented case based reasoning framework for semantic diabetes diagnosis. Artif Intell Med2015;14:92-5.

16. Baronepel O, Heymann AD, Friedman N, Kaplan G. Development of an unsupportive social interaction scale for patients with diabetes. Patient Prefer Adherence 2015;23;9:1033-41.

17. Zabbah I, Hassaanzadeh M, kohjani Z. The Effect of Continuous Parameters on the Diagnosis

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

of Coronary Artery Disease Using Artificial Neural Networks. Journal of Torbat Heydariyeh University of Medical Sciences. 2017; 4 (4):29-39

18. Zadeh, L.A. (1965), "Fuzzy Sets, Information and Control", Vol. 8, pp. 338-353.

19. Newman,D.J.Hettich, S.Blake, C.L.S. & Merz,C.J. 1007. UCI repository of machine learning database, Irvine, CA: University of California, Dept. of Information and Computer Science.archive.ics.uci.edu /ml/datasets/Pima +Indians Diabetes.

20. John, R. and Coupland,S. Type-2 Fuzzy Logic: A Historical View, Computational Intelligence Magazine, IEEE, 2007; 2(1), 57-62.

21.I. Zabbah, S. E. Yasrebi Naeini, Z. Ramazanpoor, and K. Sahragard, The Diagnosis of Thyroid Diseases Using Combinati on of Neural Networks through Hierarchical Method, 2017;4(1), 21–31.

22. A Heydari, d. Tavakoli, and P Fakharian, Approximation of Special Sheet Values Using Artificial Neural Network, Modeling in Engineering, 2013;11(35), 49-62.[Persian]

23. j Ahadian and F Behrouzi, Application of the ANFIS Adaptive System to Estimate the Risk Soil Consolidation Potential, Modeling in Engineering,1959; 14(45),17-31.[persain]

24. Biglarian A, Bakhshi E, Gohari MR, Khodabakhshi R. Artificial neural network for prediction of distant metastasis in colorectal cancer. Asian Pac J Cancer Prev,2013; 27(9),13-30.

25. Fang X. Are you becoming a diabetic A data mining approach. In Fuzzy Systems and Knowledge Discovery. FSKD'09. Sixth International Conference on IEEE, vol. 5, 2009; pp. 18-22.

26. Huang Y. McCullagh P. Black N. & Harper R.Feature selection and classification model construction on type2 diabetic patients' data. Artificial intelligence in medicine,2007; 41(3), 251-262.

27. Fabian Jaimes, Diego Alvarez, Martinez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. Critical Care.2005;9(2):150-6.

28. Burfei Farzaneh, Salehi Masoud, Najafi Iraj. Anticipating Diabetes Using An Artificial Neural Network. Razi Medical Journal. 1394; 22 (135): 29-37

29. Al Jarullah, Asma A. Decision tree discovery for the diagnosis of type II diabetes. International Conference on Innovations in Information Technology (IIT), 2011; pp. 303-307.

30. Antonelli D. Baralis E. Bruno G.Cerquitelli T.Chiusano S. & Mahoto N. Analysis of diabetic patients through their examination history. Expert Systems with Applications,2013; 40(11), 4672–78.

31. Santi WP.A new smooth support vector machine and its applications in diabetes disease diagnosis. J Comput Sci 2009; 5(1), 3-8.

32. Breault, Joseph L. Colin R. Goodall, and Peter J.Fos. Data mining a diabetic data warehouse. Artificial Intelligence in Medicine, 2002; 37-54.

33.Patil B. M. R. C. J. and Durga T. Association rul e for classification of type-2 diabetic patients. In Second International, Conference on achine Learning and Computing (ICMLC), 2010; 330-334.

Journal of Torbat Heydariyeh University of Medical Sciences
2018, Volume 6, No.2:10-20.

Original article

Diagnosis of Diabetes using Artificial Neural Network and Neuro-Fuzzy approach

Iman Zabbah[1]*, Asma Eskandari [2], Zahra Sardari [2], Abolfazl Noghandi[2]

1. Departement fo Electrical and Computer Engineering, University of Torbat Heydarieh, Iran
2. University of Torbat Heydariyeh, Torbat Heydarieh, Iran

***Corresponding author*: I.zabbah@torbath.ac.ir**

Downloaded from jms.thums.ac.ir at 8:27 IRDT on Sunday June 16th 2019

Keywords:
Diabetes,
Artificial Neural Networks,
Fuzzy Neural networks
Data mining

©2018 Torbat Heydariyeh University of Medical Sciences. All rights reserved.

Abstract

Background & Aim*:* A main problem in diabetes is its timely and accurate diagnosis. This study aimed at diagnosing diabetes using data mining methods.

***Methods*:** The present study is an analytical investigation including 768 individuals with 8 attributes. Artificial neural networks and fuzzy neural networks were used to diagnose the diabetes. To achieve a real accuracy, the Kfold method was used to divide samples into training and test groups.

***Results*:** The mean square errors in multilayer perceptron network (MLP), learning vector quantization and Nero fuzzy networks were 98.6%, 98.2% and 99.6%, respectively.

***Conclusion*:** According to the results of this study, , data mining method can be effective in diagnosing diabetes. In this regard, both used methods are useful; however, higher precision was obtained following the use of Neuro-Fuzzy approach.

How to Cite this Article: Zabbah I, Eskandari A, Sardari Z, Noghandi A. Diagnosis of Diabetes using Artificial Neural Network and Neuro-Fuzzy approach. Journal of Torbat Heydariyeh University of Medical Sciences. 2018;6(2):10-20.