مجله دیابت و متابولیسم ایران. آذر- دی ۱۳۹۷؛ دوره ۱۸ (شماره ۲) ۷۹-۷۱

مقاله پژوهشی

یک سیستم هوشمند پزشکیار مبتنی بر شبکه عصبی مصنوعی در تشخیص بیماری دیابت

آسیه خسروانیان[۱]، سعید آیت*[۲]

چکیده

مقدمه: تشخیص زودهنگام بیماری دیابت به‌منظور جلوگیری از عوارض و آسیب‌های ناشی از این بیماری امری حیاتی است. هدف از این مقاله طراحی یک سیستم هوشمند در دسته‌بندی افراد مبتلا به دیابت، با روش رگرسیون مبتنی بر شبکه عصبی پرسپترون چندلایه است.

روش‌ها: در این مطالعه‌ی توصیفی- تحلیلی یک سیستم هوشمند برای دسته‌بندی افراد به دو دسته سالم و مبتلا به بیماری دیابت طراحی‌شده است. سیستم طراحی‌شده با استفاده از نرم‌افزار MATLAB نسخه‌ی ۲۰۱۵ (۸/۵/۰/۱۹۷۶۱۳) شبیه‌سازی‌شده است. مجموعه داده مورد استفاده در این پژوهش، مجموعه داده معیار PID موجود در مخزن داده یادگیری ماشین دانشگاه ایروین کالیفرنیا است. این مجموعه داده شامل ۷۶۸ رکورد از زنان هندی و ۸ فاکتور تشخیصی برای بیماری دیابت است.

یافته‌ها: داده‌های این مجموعه پس از پیش‌پردازش به‌صورت تصادفی به ۲۰ دسته از کل مجموعه داده تقسیم شدند که شامل داده‌های آموزش و آزمون متفاوت بودند. در هر دسته داده از ۹۰ درصد داده‌ها برای مرحله‌ی آموزش و ۱۰ درصد باقی‌مانده برای مرحله‌ی آزمون استفاده شد. نتایج حاصل در بهترین حالت براساس شاخص‌های حساسیت، اختصاصیت، صحت و دقت در دسته‌بندی افراد، به ترتیب معادل اعداد ۰/٤٨١٥، ۰/۹۸۰٤، ۰/۸۰۷۷، ۰/۹۲۸٦ به‌دست آمد.

نتیجه‌گیری: نتایج به‌دست‌آمده برتری سیستم هوشمند طراحی‌شده در دسته‌بندی افراد به دو دسته سالم و بیمار را نسبت به سایر روش‌های پیاده‌سازی شده بر این مجموعه داده تأئید می‌کند. استفاده از شبکه عصبی پرسپترون چندلایه برای تقریب تابع، موجب افزایش دقت سیستم پیشنهادی شده است.

واژگان کلیدی: بیماری دیابت، رگرسیون، شبکه عصبی پرسپترون چندلایه، مجموعه داده PID

۱- گروه علمی مهندسی کامپیوتر، دانشکده‌ی مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

۲- گروه علمی مهندسی کامپیوتر و فناوری اطلاعات، دانشکده‌ی مهندسی کامپیوتر، دانشگاه پیام نور ، ایران

***نشانی:** اصفهان، نجف آباد، بلوار طالقانی، دانشگاه پیام نور، تلفن: ۰۳۱٤۲۷۲۷۱۲۳، فکس: ۰۳۱٤۲۷۲۷۱۲۳، پست الکترونیک: dr.ayat@pnu.ac.ir

تاریخ دریافت: ۱۳۹۷/۰۹/۱٤ **تاریخ درخواست اصلاح: ۱۳۹۷/۱۰/۲۱** **تاریخ پذیرش: ۱۳۹۷/۱۰/۳۰**

[Downloaded from journals.tums.ac.ir on 2025-01-08]

مقدمه

در دهه‌ی اخیر روش‌های هوش مصنوعی از جمله منطق فازی، شبکه عصبی مصنوعی، برنامه‌نویسی ژنتیک، ترکیب سیستم‌های عصبی و فازی و روش‌های رگرسیون در مسائل بسیاری به کار گرفته شده است [۱]. دسته‌بندی و تشخیص بیماری‌ها با کمک ابزارهای هوش مصنوعی ازجمله کاربردهای هوش مصنوعی در پزشکی است.

بیماری دیابت یکی از شایع‌ترین بیماری‌ها در جهان است. به گزارش فدراسیون بین‌المللی دیابت بیش از ٤١٨ میلیون نفر در جهان مبتلا به دیابت هستند که پیش‌بینی می‌شود تا ۲۵ سال آینده این رقم به ٦۰۰ میلیون نفر خواهد رسید [۲].

بیماری دیابت زمانی به‌وجود می‌آید که بدن قادر به تولید انسولین نباشد یا هنگامی‌که بدن نمی‌تواند از انسولین تولید شده استفاده مؤثر داشته باشد. در یک تقسیم‌بندی، این بیماری به سه دسته تقسیم می‌شود. دیابت نوع یک یا دیابت وابسته به انسولین که بیشتر در کودکان دیده می‌شود. دیابت نوع دو یا دیابت غیر وابسته به انسولین که در ۹۰ تا ۹۵ درصد بیماران دیابتی مشاهده می‌شود و دیابت نوع سوم که بیشتر در زنان باردار مشاهده می‌شود که پس از بارداری به دیابت نوع دوم تبدیل می‌شود [۳].

تشخیص بیماری دیابت و یا آگاهی یافتن از احتمال بالای ابتلا به این بیماری همواره کار آسانی نیست، چراکه این بیماری علائم متعددی را بروز می‌دهد که بعضی از این علائم در سایر بیماری‌ها نیز وجود دارند؛ بنابراین پزشک برای اتخاذ یک تصمیم مناسب باید نتیجه آزمایش‌های بیمار و تصمیم‌هایی که در گذشته برای بیماران با وضعیت مشابه گرفته است را بررسی کند [٤]. تشخیص زودهنگام این بیماری و پیش‌آگهی به بیماران می‌تواند با درمان به‌موقع از عوارض و آسیب‌های ناشی از این بیماری جلوگیری کند؛ بنابراین وجود سیستم هوشمند در کمک به پزشک و پیش‌آگهی به بیماران می‌تواند مفید باشد و با تشخیص این بیماری در مراحل اولیه از عوارضی چون افزایش فشارخون، نارسایی کلیه، بیماری قلبی و در مواردی قطع عضو جلوگیری کند.

تاکنون روش‌های هوش مصنوعی مختلفی در تشخیص بیماری دیابت به‌کار گرفته‌شده است. به‌عنوان مثال در مطالعه‌ای [۵] با استفاده از روش‌های یادگیری ماشین از جمله k- نزدیک‌ترین همسایه، نایو بیز، درخت تصمیم، جنگل‌های تصادفی، ماشین بردار پشتیبان و رگرسیون لجستیک به تشخیص این بیماری پرداخته‌اند و روش‌های به‌کار گرفته شده بر ۳۰۰ نمونه‌ی تصادفی انتخاب شده از مخزن داده‌ی EHR مورد ارزیابی قرارگرفته است. در مطالعه‌ای دیگر [٦] روش‌های گروهی آدابوست و Bagging با استفاده از درختان تصمیم C4.5 و J48 برای تشخیص بیماری دیابت ارائه‌شده است و روش پیشنهادی بر مجموعه داده‌ی بیماران دیابتی کانادایی ارزیابی‌شده است. همچنین در مطالعه‌ای [۷] به مقایسه‌ی روش‌های یادگیری ماشین درخت تصمیم J48، K-NN، جنگل‌های تصادفی و SVM پرداختند و الگوریتم‌های مذکور بر مجموعه داده‌ی UCI ارزیابی‌شده است. در [۸] با استفاده از روش‌های داده‌کاوی از جمله درخت تصمیم یک مدل پیش‌بینی برای بیماری دیابت نوع دو طراحی‌شده و بر مجموعه داده‌ی جمع‌آوری شده از بیماران در شهر تبریز ارزیابی‌شده است. در مطالعه‌ای دیگر [۹] از روش بیزین و مدل‌های آماری برای پیش‌بینی دیابت نوع دو استفاده شده است و بر مجموعه داده‌ی EHR شبیه سازی شده است. محققان در یک بررسی [۱۰] به تشخیص بیماری دیابت براساس اطلاعات مستخرج از سیگنال ECG با استفاده از شبکه‌های عصبی مصنوعی پرداختند و به این نتیجه رسیدند که مدل مبتنی بر قوانین دقت بالاتری نسبت به دیگر الگوریتم‌های طبقه‌بندی داده‌کاوی دارد. دیگر روش‌های یادگیری ماشین پیاده‌سازی شده در بیماری دیابت در سایر مطالعات [۱۸-۱۱] آمده است. هدف از پژوهش حاضر طراحی و ارزیابی یک سیستم هوشمند در دسته‌بندی افراد به دودسته سالم و مبتلا به بیماری دیابت با استفاده از روش رگرسیون مبتنی بر شبکه عصبی پرسپترون چندلایه است.

روش‌ها

مجموعه داده‌ها

در این مطالعه از مجموعه داده‌ی معیار PID: Pima Indians Diabetes موجود در مخزن داده‌ی یادگیری ماشین (UCI: University of California, Irvine) استفاده شده است [۱۹]. این مجموعه شامل ۷٦۸ رکورد جمع‌آوری‌شده از زنان هندی است که از این تعداد ۵۰۰ زن سالم و ۲٦۸ نفر مبتلا به دیابت نوع یک بودند. حداقل سن افراد مورد بررسی ۲۱ سال و حداکثر ۸۱ سال است. ۸ فاکتور تشخیصی برای هر یک از افراد در جدول ۱ نشان داده‌شده است.

[Downloaded from journals.tums.ac.ir on 2025-01-08]

جدول ۱- ریسک فاکتورهای تشخیصی بیماری دیابت

۱	تعداد دفعات وضع حمل (برای زنان)	Number of times pregnant
۲	غلظت گلوکز پلاسمای خون	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
۳	فشارخون دیاستولیک	Diastolic blood pressure (mm Hg)
٤	ضخامت پوست ماهیچه سه سر بازویی	Triceps skin fold thickness (mm)
٥	انسولین سرم دوساعته	2-Hour serum insulin (mu U/ml)
٦	نمایه‌ی توده‌ی بدنی (BMI)	Body mass index (weight in kg/(height in m)^2)
۷	داشتن سابقه‌ی دیابت	Diabetes pedigree function
۸	سن	Age (years)

رگرسیون

روش‌های رگرسیون ازجمله روش‌های یادگیری ماشین است که به‌طور گسترده در مسائل دسته‌بندی به‌کار برده شده است. رگرسیون یک تابع یادگیری است که یک داده را به یک متغیر پیش‌بینی با مقدار حقیقی نگاشت می‌کند.

به نقل از Ghazanfari و همکاران [۲۰]، در مدل رگرسیون خطی تک متغیره اگر x یک متغیر مستقل و y متغیر وابسته باشد، آنگاه معادله رگرسیون خطی تک متغیره به شکل $y = w_0 + w_1 x$ خواهد بود. اگر D، مجموعه داده‌های آموزشی یک جامعه به‌صورت $(x_1, y_1) \dots (x_D, y_D)$ باشد، ضرایب رگرسیون خطی را می‌توان براساس روش کمترین مربعات خطا به‌دست آورد (رابطه‌ی ۱). پس از تعیین مقادیر w_0 و w_1 در مشاهدات جدید با جای‌گذاری مقدار مستقل x در رابطه‌ی $y = w_0 + w_1 x$ می‌توان مقدار متناظر متغیر وابسته y را پیش‌بینی نمود.

رابطه‌ی ۱:

$$w_l = \frac{\sum_{i=1}^{|D|}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|}(x_i - \bar{x})^2}$$
$$w_0 = \bar{y} - w_1 \bar{x}$$

در مدل رگرسیون خطی چند متغیره تعداد متغیرهای مستقل در معادله رگرسیونی بیش از یکی است. اگر مقادیر x_i، نمونه‌های آزمایشی n بعدی باشد که برچسب آنها y_i است در این صورت معادله‌ی رگرسیونی به شکل رابطه‌ی ۲ خواهد بود.

رابطه‌ی ۲:

$$y = w_0 + w_1 x_1 + w_2 x_2$$

این معادله با استفاده از روش حداقل مربعات قابل حل است.

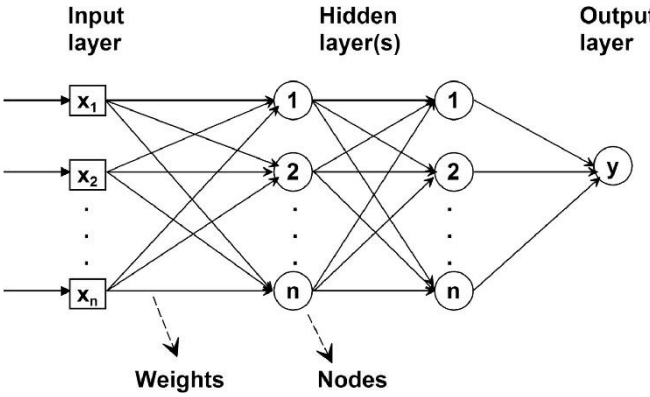
اگر وابستگی‌ها به‌صورت یک تابع چندجمله‌ای باشد، در برخی موارد می‌توان مدل غیرخطی را به یک مدل رگرسیون خطی تبدیل کرد و براساس روش حداقل مربعات مسئله را حل کرد. لازم به ذکر است که برخی از مدل‌های غیرخطی با استفاده از تغییر متغیر، به‌راحتی قابل‌تبدیل به شکل خطی نیستند. برای اینکه رگرسیون خطی بتواند متغیر طبقه‌ای را پیش‌بینی کند لازم است تعمیم یابد. در مدل‌های تعمیم‌یافته، واریانس متغیر پاسخ (y) تابعی از مقدار میانگین y است. درحالی‌که در رگرسیون خطی، واریانس y ثابت بود. مدل‌های خطی تعمیم‌یافته شامل رگرسیون لجستیک و رگرسیون پواسون هستند. مدل‌های رگرسیون لجستیک، احتمال وقوع پدیده‌هایی که تابعی خطی از یک مجموعه متغیرهای مستقل هستند را پیش‌بینی می‌کند. داده‌های شمارشی نیز از توزیع پواسون پیروی کرده و به‌طورمعمول با رگرسیون پواسون مدل‌سازی می‌شوند.

شبکه عصبی پرسپترون چندلایه

شبکه‌ی عصبی پرسپترون چندلایه (MLP: Multilayer perceptron neural network) یک شبکه‌ی عصبی جلو سو[1] است که از یک لایه‌ی ورودی، لایه‌ی پنهان و لایه‌ی خروجی تشکیل شده است [۲۱]. در MLP از روش یادگیری با نظارت با الگوریتم پس انتشار خطا برای آموزش استفاده می‌شود. ساختار شبکه عصبی MLP در شکل ۱ نشان داده شده است.

[1] Feed forward

[Downloaded from journals.tums.ac.ir on 2025-01-08]



شکل ۱- ساختار شبکه عصبی پرسپترون چندلایه [۲۲]

ماتریس کانفیوژن

به‌طورکلی برای بررسی میزان موفقیت و کارایی سیستم‌های یادگیری، پیش‌بینی و تشخیص از ماتریس کانفیوژن استفاده می‌شود. تحلیل‌های ماتریس کانفیوژن در دسته‌بندی و تشخیص بیماران منجر به چهار حالت مثبت حقیقی[1] (TP)، منفی حقیقی[2] (TN)، مثبت کاذب[3] (FP) و منفی کاذب[4] (FN) می‌شود. از نتایج ماتریس کانفیوژن چهار شاخص صحت[5]، دقت[6]، حساسیت[7] و اختصاصیت[8] به‌صورت زیر به دست می‌آید:

$$\text{صحت} = \frac{TP+TN}{TP+TN+FP+FN} \quad (۳)$$

$$\text{دقت} = \frac{TP}{TP+FP} \quad (۴)$$

$$\textbf{حساسیت} = \frac{TP}{TP+FN} \quad (۵)$$

$$\text{اختصاصیت} = \frac{TN}{FP+TN} \quad (۶)$$

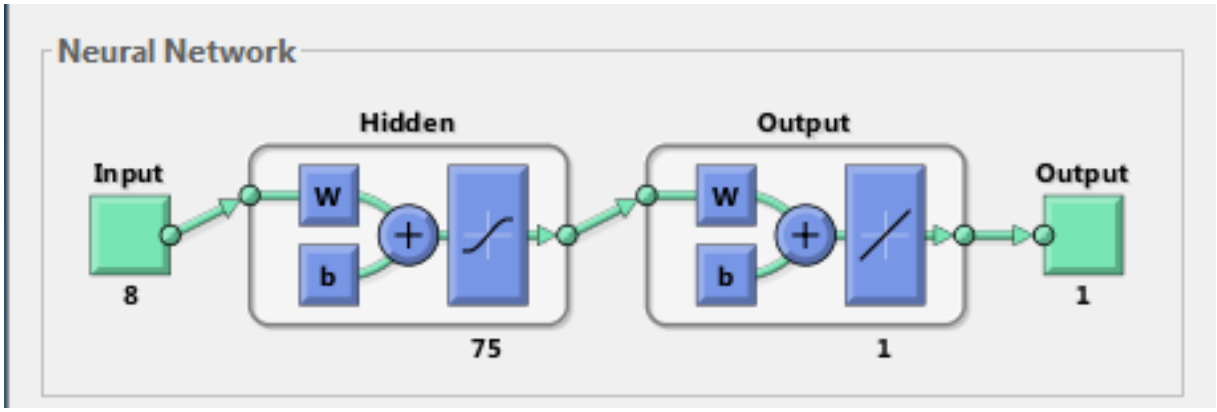
پیاده‌سازی سیستم هوشمند پیشنهادی در این مطالعه شامل سه مرحله‌ی پیش‌پردازش، آموزش و آزمون است که در نرم‌افزار MATLAB شبیه‌سازی شده است. در مرحله‌ی پیش‌پردازش، مجموعه داده PID مورد بررسی قرار گرفت و مشاهده شد این مجموعه داده دارای مقادیر گم‌شده است که در ریسک فاکتورهای تشخیصی برای این مقادیر صفر قرار داده شده است. از آنجاکه در نتایج آزمایش مقدار صفر برای بعضی فاکتورهای تشخیصی بی‌معنی است، در این مرحله، تعداد ۱۶ رکورد که دارای مقادیر گم‌شده بودند از مجموعه حذف شدند و شبیه‌سازی با ۷۵۲ داده باقی‌مانده انجام شد. پس از مرحله‌ی پیش‌پردازش، داده‌ها در هر بار شبیه‌سازی به‌طور تصادفی به دو دسته آموزش و آزمون تقسیم شدند. به‌این‌ترتیب که از ۹۰ درصد داده‌ها برای مرحله آموزش و ۱۰ درصد باقیمانده برای مرحله آزمون استفاده شد. درمجموع از ۲۰ دسته داده مختلف استخراج شده به‌صورت تصادفی از مجموعه داده PID برای شبیه‌سازی سیستم پیشنهادی استفاده شد.

در مرحله‌ی آموزش از ۸ فاکتور تشخیصی نشان داده شده در جدول ۱ به‌عنوان ورودی سیستم استفاده شد. برای این منظور ماتریسی شامل ۸ سطر (۸ فاکتور تشخیصی) و به تعداد داده‌های تصادفی انتخاب شده ستون، برای مرحله‌ی آموزش و ماتریس دیگری با دو سطر (سالم=۰ و بیمار=۱) و به تعداد داده‌های تصادفی انتخاب‌شده ستون، به‌عنوان ماتریس هدف به کار گرفته شد. از ۱۰ درصد باقیمانده با ساختاری مشابه مرحله‌ی آموزش برای مرحله آزمون استفاده شد. ورودی‌ها با ساختاری که شرح داده شد برای تقریب تابع هدف به‌عنوان یک روش رگرسیون مورداستفاده قرار گرفت. مطالعات نشان می‌دهد شبکه‌های عصبی قادر به‌تقریب هر تابع عملیاتی می‌باشند. در میان شبکه‌های عصبی مختلف، شبکه‌ی عصبی پرسپترون چندلایه با ساختار پس انتشار خطا و تابع انتقال سیگموئید در لایه‌های مخفی و تابع انتقال خطی در لایه‌ی خروجی در ارتباط با کاربردهای رگرسیون از کارایی مناسبی برخوردار است؛ بنابراین پس از بارگذاری داده‌های ورودی یک

[1] True Positive
[2] True Negative
[3] False Positive
[4] False Negative
[5] Accuracy
[6] precision
[7] Sensitivity
[8] Specificity

[Downloaded from journals.tums.ac.ir on 2025-01-08]

شبکه با ساختاری که شرح داده شد، همان‌طور که در شکل ۲ مشاهده می‌شود ایجاد شد. تعداد نرون‌های لایه‌ی مخفی با روش سعی و خطا ۷۵ به‌دست آمد. استفاده از این تعداد نرون در لایه مخفی اگرچه حجم محاسبات را بالا می‌برد اما قابلیت شبکه را برای حل مسئله مورد نظر افزایش می‌دهد. سیستم طراحی‌شده با داده‌های آموزشی و به‌کارگیری روش cross-validation در ۲۰ دسته آموزش داده شد و با داده‌های آزمون مربوط به هر دسته، مورد آزمایش قرار گرفت. در ادامه به‌منظور گسسته‌سازی مقدار تابع هدف پیش‌بینی‌شده از حد آستانه استاندارد ۰/۵ استفاده شد و نتایج خروجی به دودسته سالم و بیمار دسته‌بندی شد.



شکل ۲- ساختار شبکه عصبی MLP پیاده‌سازی شده در نرم‌افزار MATLAB

یافته‌ها

سیستم پیشنهادی با استفاده از داده‌های مرحله‌ی آموزش، در هر یک از ۲۰ دسته به‌صورت جداگانه آموزش داده شد سپس داده‌های مرحله‌ی آزمون مربوط به هر دسته به سیستم شبیه‌سازی‌شده اعمال گردید. خروجی سیستم، دسته‌بندی افراد به دو دسته سالم و مبتلا به بیماری دیابت بود. جدول ۲ خروجی سیستم پیشنهادی برای هریک از دسته‌ها براساس شاخص‌های حساسیت، اختصاصیت، دقت و صحت نشان می‌دهد. همان‌طور که در جدول ۲ مشاهده می‌شود سیستم پیشنهادی، در داده‌های دسته‌ی هفدهم ازنظر شاخص دقت عملکرد مناسبی داشته است. در مقایسه نتایج این مقاله با سایر روش‌های پیاده‌سازی شده بر این مجموعه داده، شاخص دقت از جمله شاخص‌هایی بوده است که مورد توجه قرارگرفته است و شاخص‌های حساسیت و اختصاصیت در سایر مقالات کمتر مورد توجه بوده است. لذا برتری یک دسته براساس شاخص دقت توسط نویسندگان این مقاله در نظر گرفته‌شده است.

[Downloaded from journals.tums.ac.ir on 2025-01-08]

جدول ۲- شاخص‌های عملکردی سیستم پیشنهادی در هر دسته داده

	تعداد داده‌های مرحله آموزش	تعداد داده‌های مرحله آزمون	حساسیت	اختصاصیت	دقت	صحت
دسته اول	٦٧٠	٨٢	٠/٥٠٠٠	٠/٩٥٤٥	٠/٩٠٤٨	٠/٧٤٣٩
دسته دوم	٦٦٥	٨٧	٠/٥٠٠٠	٠/٨٩٤٧	٠/٧١٤٣	٠/٧٥٨٦
دسته سوم	٦٧٧	٧٥	٠/٥٠٠٠	٠/٩١٤٩	٠/٧٧٧٨	٠/٧٦٠٠
دسته چهارم	٦٦٩	٨٣	٠/٥٢١٧	٠/٩١٦٧	٠/٧٠٥٩	٠/٨٠٧٢
دسته پنجم	٦٧٤	٧٨	٠/٥٧١٤	٠/٧٩٠٧	٠/٦٨٩٧	٠/٦٩٢٣
دسته ششم	٦٦٢	٩٠	٠/٦١٧٦	٠/٧٨٥٧	٠/٦٣٦٤	٠/٧٢٢٢
دسته هفتم	٦٧٥	٧٧	٠/٥٠٠٠	٠/٩٠٥٧	٠/٧٠٥٩	٠/٧٧٩٢
دسته هشتم	٦٦٢	٩٠	٠/٥٩٤٦	٠/٨٨٦٨	٠/٧٨٥٧	٠/٧٦٦٧
دسته نهم	٦٦٧	٨٥	٠/٧٨٧٩	٠/٨٠٧٧	٠/٧٢٢٢	٠/٨٠٠٠
دسته دهم	٦٥٧	٩٥	٠/٤٨٤٨	٠/٨٧١٠	٠/٦٦٦٧	٠/٧٣٦٨
دسته یازدهم	٦٦٦	٨٦	٠/٤٦٤٣	٠/٩١٣٨	٠/٧٢٢٢	٠/٧٦٧٤
دسته دوازدهم	٦٦٨	٨٤	٠/٦٠٦١	٠/٨٢٣٥	٠/٦٨٩٧	٠/٧٣٨١
دسته سیزدهم	٦٥٨	٩٤	٠/٤٥٩٥	٠/٨٧٧٢	٠/٧٠٨٣	٠/٧١٢٨
دسته چهاردهم	٦٦٤	٨٨	٠/٦٩٢٣	٠/٩١٩٤	٠/٧٨٢٦	٠/٨٥٢٣
دسته پانزدهم	٦٦٢	٩٠	٠/٥٥٥٦	٠/٨٥١٩	٠/٧١٤٣	٠/٧٣٣٣
دسته شانزدهم	٦٦٦	٨٦	٠/٢٥٨١	٠/٩٦٣٦	٠/٨٠٠٠	٠/٧٠٩٣
دسته هفدهم	٦٧٤	٧٨	٠/٤٨١٥	٠/٩٨٠٤	٠/٩٢٨٦	٠/٨٠٧٧
دسته هجدهم	٦٦٥	٨٧	٠/٦٩٤٤	٠/٩٢١٦	٠/٨٦٢١	٠/٨٢٧٦
دسته نوزدهم	٦٦٧	٨٥	٠/٥٧١٤	٠/٩٢٩٨	٠/٨٠٠٠	٠/٨١١٨
دسته بیستم	٦٧١	٨١	٠/٦٠٦١	٠/٨٩٥٨	٠/٨٠٠٠	٠/٧٧٧٨

بحث

سیستم هوشمند طراحی‌شده در این مطالعه که با ترکیب روش رگرسیون و شبکه عصبی پرسپترون چندلایه طراحی‌شده است موجب افزایش دقت دسته‌بندی نسبت به سایر روش‌های پیاده‌سازی شده بر این مجموعه داده شده است. در واقع تقریب تابع با استفاده از شبکه عصبی پرسپترون چندلایه موجب بهبود در شاخص‌های عملکردی سیستم پیشنهادی شده است.

تاکنون روش‌های مختلف هوش مصنوعی در تشخیص بیماری دیابت طراحی و ارزیابی‌شده است. در مقایسه‌ی نتایج این مطالعه با سایر روش‌های پیاده‌سازی در این موضوع به مواردی اشاره می‌شود که بر مجموعه داده PID پیاده‌سازی شده است. Fathi Ganji و همکاران از ترکیب الگوریتم بهینه‌سازی کلونی مورچگان و سیستم فازی برای تشخیص بیماری دیابت استفاده کردند و به‌دقت ٨٤/٢٤ رسیدند [٤]. Khosravanian و همکاران با استفاده از شبکه عصبی مصنوعی به تشخیص این بیماری پرداختند. آنها یک شبکه‌ی عصبی پس انتشار با الگوریتم گرادیان توأم مقیاس شده طراحی کردند و دقت به‌دست‌آمده را ٨٤/٣٠ گزارش کردند [٢٤]. Fayssal و همکاران ترکیب الگوریتم کلونی زنبور عسل و سیستم فازی را در تشخیص این بیماری به‌کار گرفتند و به دقت ٨٤/٢١ رسیدند [٢٥]. در پژوهشی دیگر از Khosravanian و همکاران یک سیستم استنتاج عصبی- فازی[1] با الگوریتم ژنتیک و الگوریتم بهینه‌سازی ازدحام ذرات در تشخیص بیماری دیابت

[1] Adaptive Neuro-Fuzzy Inference System: ANFIS

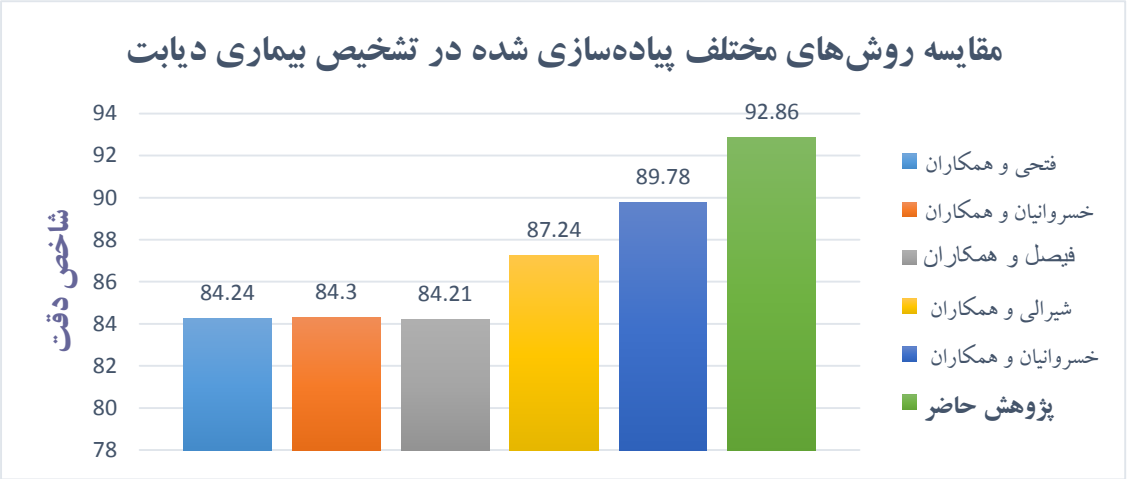
[Downloaded from journals.tums.ac.ir on 2025-01-08]

طراحی شد. دقت به‌دست‌آمده با استفاده از الگوریتم ژنتیک ۷۶/۷۶ و با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات ۸۹/۷۸ گزارش شد [۲۶].

جدول ۳ و شکل ۳ مقایسه‌ی روش‌های پیاده‌سازی شده در تشخیص بیماری دیابت را نشان می‌دهد.

جدول ۳- مقایسه روش‌های مختلف پیاده‌سازی شده در تشخیص بیماری دیابت

نویسندگان	پایگاه داده مورد استفاده	روش پیاده‌سازی	دقت	سال
فتحی و همکاران [٤]	PID	ترکیب الگوریتم بهینه‌سازی کلونی مورچگان و سیستم فازی	۸٤/۲٤	۲۰۱۱
خسروانیان و همکاران [۲٤]	PID	شبکه عصبی پس انتشار با الگوریتم گرادیان توأم مقیاس شده	۸٤/۳۰	۲۰۱۱
Fayssal و همکاران [۲۵]	PID	ترکیب الگوریتم کلونی زنبورعسل و سیستم فازی	۸٤/۲۱	۲۰۱۳
شیرالی و همکاران [۲۳]	PID	ترکیب سیستم استنتاج فازی سوگنو و الگوریتم کرم شب‌تاب	۸۷/۲٤	۲۰۱٦
خسروانیان و همکاران [۲٦]	PID	ترکیب ANFIS و PSO	۸۹/۷۸	۲۰۱٦
پژوهش حاضر	PID	روش رگرسیون مبتنی بر MLP	۹۲/۸٦	۲۰۱۸



شکل ۳- مقایسه روش‌های پیاده‌سازی شده در تشخیص دیابت بر مجموعه داده‌ی PID

همان‌طور که در جدول ۳ مشاهده می‌شود در این مقاله نیز از مجموعه داده‌ی PID استفاده شده است و سیستم پیشنهادی توانست با دقت مناسبی افراد را به دودسته بیمار و سالم طبقه‌بندی کند. نتایج شبیه‌سازی نشان داد شاخص دقت به‌دست‌آمده نسبت به سایر روش‌های شبیه‌سازی شده بر این مجموعه داده بهبود یافته است. دلیل این امر همان‌طور که پیش‌تر ذکر شد استفاده از توانایی شبکه عصبی مصنوعی MLP در تقریب تابع است. علاوه بر آن پیش‌پردازش داده‌ها در کم کردن خطای شبیه‌سازی و افزایش دقت پیش‌بینی مؤثر بود. اگرچه پیچیدگی‌های روش رگرسیون موجب شده است تا روش پیشنهادی در مقایسه با روش‌های مطرح‌شده در [۲٦] ازنظر سرعت اجرا ضعیف‌تر عمل کند اما رسیدن به‌دقت بیشتر در دسته‌بندی افراد و نزدیک شدن نتایج به واقعیت می‌تواند این نقص را پوشش دهد. از محدودیت‌های این مطالعه می‌توان به ۱٦ رکورد حذف شده اشاره کرد چراکه تخمین نزدن موارد حذف شده و جایگزین نشدن آنها می‌تواند در نتایج به‌دست‌آمده تأثیرگذار باشد.

در این مطالعه، سیستم هوشمندی برای دسته‌بندی افراد به دو دسته سالم و مبتلابه بیماری دیابت با استفاده از روش رگرسیون مبتنی بر شبکه عصبی پرسپترون چندلایه طراحی

[Downloaded from journals.tums.ac.ir on 2025-01-08]

شد. روش به‌کار رفته در این پژوهش تاکنون در تشخیص بیماری دیابت ارائه نشده است. نتایج حاصل از ارزیابی سیستم طراحی‌شده نشان داد سیستم پیشنهادی با دقت مناسبی افراد را به دودسته سالم و بیمار دسته‌بندی کرده است. این نتیجه از آن جهت اهمیت دارد که با پیش‌آگهی به بیماران، مراجعه و درمان به‌موقع می‌توان از عوارض ناشی از این بیماری جلوگیری کرد. در واقع، سیستم طراحی‌شده می‌تواند برای بهبود تشخیص بیماری دیابت مورد استفاده قرار گیرد.

با توجه به مطالعاتی که نشان دهنده بهبود شاخص‌های عملکردی سیستم‌های ترکیبی در تشخیص بیماری است پیشنهاد می‌شود در پژوهش‌های آتی از دیگر شبکه‌های عصبی مصنوعی به این منظور استفاده شود و نتایج حاصل با این مطالعه مقایسه شود.

سپاسگزاری

این مقاله هیچگونه تضاد منافعی برای نویسندگان ندارد و مستخرج از پژوهشی است که با حمایت مالی دانشگاه پیام نور انجام شده است.

مآخذ

1. Kar S, Das S, Kanti Ghosh P. Applications of neuro fuzzy systems: A brief review and future outline. *Appl Soft Comput* 2014; 15: 243-259.
2. International Diabetes Fedration. A Report from the International Diabetes Fedration. *Available from: http://www.idf.org*
3. Fiuzy M, Haddania J, Mollania N. Prediction and determination of correctdose of insulin in diabetic patients based on diabetes fuzzy diagnosis. *Ijdld* 2017; 16 (1):17-38.
4. Fathi Ganji M, Saniee Abadeh M. A fuzzy classification system based on ant colony optimization for diabetes disease diagnosis .*Expert* Syst 2011; 38: 14650–14659.
5. Zheng T, Xie W, Xu L, He X, Zhang Y, You M and et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int J Med Inform* 2017; 97: 120-127.
6. Perveen S, Shahbaz M, Guergachi A, Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes. *Procedia Comput Sci* 2016; 82: 115-121.
7. Kandhasamy JP, Balamuria S. Performance analysis of classifier models to predict diabetes mellitus. *Procedia Comput Sci* 2015; 47: 45-51.
8. Habibi Sh, Ahmadi M, Alizadeh S. Type 2 diabetes Mellitus screening and risk factors using decision tree: Results of data mining. Glob J Health sci 2015; 7: 5 Available from: *http://www.ccsenet.org/journal/index.php/gjhs/article/view/43726*
9. Anderson J.P, Parikh J.R, Shenfeld D.K. Reverse engineering of prediction models for progression to Type 2 diabetes. *Diabetes Technology Society* 2016; 10 (1): 6-18.
10. Nazari M, Zamani Dehkordi B, Kiomarsi Dehkordi F. [Diagnosis diabetes on the basis of information extracted from the ECG signal using artificial neural networks]. *J Shahrekord Univ Med* Sci. 2017; 19 (4): 64-77. (Persian)
11. Anderson AE, Kerr WT, Thames A, Li T, Xiao J, Cohen MS. Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: a cross-sectional, unselected, retrospective study. *J Biomed Inform* 2016; 60: 162–168.
12. A. Ramezankhani, O. Pournik, J. Shahrabi, F. Azizi, F. Hadaegh, D. Khalili. The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of type 2 diabetes. *Med Decis Making* 2016; 36 (1): 137–144.
13. Lee BJ, Ku B, Nam J, Pham DD, Kim JY. Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes. *IEEE J Biomed Health Inform* 2014; 18 (2): 555–561.
14. Mani S, Chen Y, Elasy T, Clayton W, Denny J. Type 2 diabetes risk forecasting from EMR data using machine learning. *AMIA Annu Symp Proc* 2012; 606–615.
15. Han L, Luo S, Yu J, Pan L, Chen S. Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. *IEEE J Biomed Health Inform* 2015; 19 (2): 728–734.
16. G.J. Simon, J. Schrom, M.R. Castro, P.W. Li, P.J. Caraballo. Survival association rule mining towards type 2 diabetes risk assessment. *AMIA Annu Symp Proc* 2013; 1293–1302.
17. G.J. Simon, P.J. Caraballo, T.M. Therneau, S.S. Cha, M.R. Castro, P.W. Li. Extending association rule summarization techniques to assess risk of diabetes mellitus. *IEEE Trans Knowl Data Eng* 2015; 27 (1): 130–141.
18. S. El-Sappagh, M. Elmogy, A.M. Riad. A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. *Artif Intell Med* 2015; 65 (3):179–208.
19. UCI Machine Learning Repository: Pima Indians Diabetes Data Set: *https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes*

[Downloaded from journals.tums.ac.ir on 2025-01-08]

20. Ghazanfari M, Alizadeh S, Teimourpour B. Data Mining & Knowledge Discovery. *Tehran: iust*; 2011. [In Persian]
21. Tang RL Z. An improved neural network model and its applications. Information in Computer Science, 8 200; 10: 1881-1888.
22. Raghu S, Sriraam N. Optimal configuration of multilayer perceptron neural network classifier for recognition of intracranial epileptic seizures. *Expert syst* 2017; 89: 205-221.
23. Shirali M, Madmoli Y, Roohafza J, Karimi H, Baboli Bahmaei A, Ertebati S. [Improvement diagnosis of diabetes using a combination of sugeno fuzzy inference systems and firefly algorithms]. *Ijdld* 2017; 15(3):172-176. [In Persian]
24. Khosravanian A, Ayat S. Design a Back Propagation Neural Network for Diabetes Diagnosis. *Proceedings of the 4th Communication & Telecommunication Applied Science & Technology 2014*. Shiraz. Iran.
25. Fayssal B. and Chikh MA. Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm. *Comput Methods Programs Biomed* 2013; 112: 92-103.
26. Khosravanian A, Rahmanimanesh M. Designing a Neuro-Fuzzy Inference System Using Genetic Algorithm and Particle Swarm Optimization Algorithms in diagnosis of diabetes. *Proceedings of the 4th International Conference on Applied Research in Computer Engineering and signal processing*. 2016; Tehran-Iran.

[Downloaded from journals.tums.ac.ir on 2025-01-08]

A PHYSICIAN ASSISTANT INTELLIGENCE SYSTEM BASED ON ARTIFICIAL NEURAL NETWORK FOR DIABETES DIAGNOSIS

Asieh Khosravanian[1], Saeed Ayat[2*]

1. *Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran*
2. *Department of Computer Engineering and Information Technology, Payame Noor University, Iran*

ABSTRACT

Backgrounds: Early detection of diabetes is critical to avoid complications and damage caused by this disease. The purpose of this paper is designing an intelligent system for Diabetes prediction (healthy or patient) by using regression method based on Multilayer Perceptron Neural Network.
Methods: In this descriptive-analytic study, an intelligent system is designed to classification diabetes patients. The system is simulated by MATLAB software 2015 (8.5.0.197613). In this study, used PID dataset in UCI Machine Learning Repository. The dataset is contained 768 records from Indian women and 8 diagnostic factors for Diabetes.
Results: The data were then divided randomly in 20 groups for training and testing, after preprocessing. 90% of the data is used for training phase and 10% for the test phase. The results obtained based on sensitivity, specificity, accuracy and precision were 0.4815, 0.9804, 0.8077 and 0.9286, respectively.
Conclusion: The obtained results, showed superiority of designed intelligent system to classify individuals (healthy and patient) in comparison with other methods implemented on this dataset. Using MLP-Regression has increased the accuracy of the proposed system.

Keywords: Diabetes, Regression, Multilayer Perceptron Neural Network, PID dataset

* Taleghani Blvd, Payame Noor University, Najaf Abad, Isfahan, Tel: +983142727123, Fax: +983142727123, E-mail: dr.ayat@pnu.ac.ir

[Downloaded from journals.tums.ac.ir on 2025-01-08]