

پیش‌بینی صحیح دوز انسولین در بیماران دیابتی نوع دو با استفاده از ترکیب سیستم‌های هوشمند مصنوعی و الگوریتم‌های داده‌کاوی

محمد فیوضی^{۱*}، جواد حدادنیا^۲، نسرین ملانیا^۳، محمد محمد زاده^۱

چکیده

مقدمه: دیابت بیماری است که علاوه بر پیشگیری، نیاز به مراقبت‌های فراوانی از قبیل پیش‌بینی صحیح میزان نوسانات سطح قند خون دارد. از مهم‌ترین عوارض این بیماری می‌توان به بیهوشی، کما و حتی مرگ اشاره کرد. امروزه در این بیماران، تعیین صحیح دوز انسولین براساس تجربه یا دانش پزشکان با استفاده از تعامل با بیماران مشخص می‌شود، هر چند که وجود خطاهای انسانی اجتناب‌ناپذیر است. در این مطالعه سعی شده است تا پیش‌بینی صحیحی از نیاز ۴۸ ساعت آینده انسولین برای بیماران به دست آید.

روش‌ها: ۱۲۴ بیمار دیابتی نوع دو و ۱۸۸ فرد مشکوک به بیماری نوع دو براساس ۱۲ ویژگی با انتخاب تصادفی از بین افراد مراجعه کننده به مرکز تحقیقات دیابت دانشگاه علوم پزشکی سبزوار در فاصله سال‌های ۱۳۸۵ تا ۱۳۹۰ مورد مطالعه و بررسی قرار گرفتند. سیستم پیشنهادی نویسندگان دارای چندین زیر سیستم از قبیل الگوریتم‌های تکاملی ($BPSO^1$) به منظور انتخاب بهترین و موثرترین ویژگی‌ها، الگوریتم‌های داده‌کاوی (SVM^2) به منظور تشخیص و دسته‌بندی ویژگی‌های موثر از غیر موثر و سیستم‌های انطباقی فازی عصبی ($ANFIS^3$) به منظور تخمین، یادگیری و تطبیق در جهت پیش‌بینی صحیح، مورد استفاده قرار گرفته است. تمام شبیه‌سازی‌ها توسط نرم افزار MATLAB انجام شده است.

یافته‌ها: سیستم پیشنهادی براساس بهترین ویژگی‌ها در بانک داده در قالب ترکیب و تعامل موفق شد به دقت بالایی با کمترین خطا دست یابد. این سیستم در مقایسه با سایر روش‌های معمول از سرعت قابل توجه، عملکرد مناسب و دقت بالایی برخوردار است. سیستم پیشنهادی در قالب ترکیب و تعامل با زیر سیستم‌ها موفق شد به نرخ ویژگی ۸۴/۱ درصد، ۹۱ درصد در حساسیت و به دقت ۹۲/۹ درصد، دست یابد.

نتیجه‌گیری: در این تحقیق با توجه به ضرورت تعیین صحیح و به موقع میزان انسولین برای بیماران، روشی نوین مبتنی بر ترکیب سیستم‌های هوشمند جهت پیش‌بینی صحیح دوز انسولین برای بیماران ارائه شده است. با به کارگیری سیستم پیشنهادی علاوه بر مشخص شدن نیاز صحیح انسولین برای بیماران از مشکلاتی همچون بستری و مراقب‌های طولانی و طاقت فرسای بیماران برای مشخص شدن نیاز انسولین آن‌ها پیشگیری شده است.

واژگان کلیدی: دیابت، بازشناسی الگو، پیش‌بینی، دوز انسولین، فرآیندهای پردازش و هوشمند مصنوعی

۱- مرکز تحقیقات سالمند شناسی، دانشگاه علوم پزشکی سبزوار، سبزوار، ایران

۲- مرکز تحقیقات فناوری های نوین پزشکی، دانشکده مهندسی برق، دانشگاه حکیم سبزواری، سبزوار، ایران

۳- بخش بیوشیمی، گروه زیست شناسی، دانشکده علوم پایه، دانشگاه حکیم سبزواری، سبزوار، ایران

* نشانی: تهران، نارمک، دانشگاه علم و صنعت، دانشکده مهندسی برق، بخش مهندسی کنترل، کدپستی: ۱۶۸۴۶۱۳۱۱۴، تلفن: ۰۹۱۸۸۱۶۰۹۳۰، پست

الکترونیک: mohammad.fiuzy@yahoo.com

مقدمه

بیماری دیابت یکی از شایع‌ترین بیماری‌های شناخته شده در دنیا می‌باشد. براساس گزارش‌ها، حدود ۲۲۰ میلیون نفر در جهان از این بیماری رنج می‌برند [۳-۱]. دیابت به ۳ نوع تقسیم می‌شود؛ نوع یک یا دیابت وابسته به انسولین که بیشتر در سن پایین و کودکان دیده می‌شود، نوع دو یا دیابت غیر وابسته به انسولین که در ۹۰ تا ۹۵ درصد بیماران دیابتی مشاهده می‌شود [۱]، و نوع سوم که بیشتر در زنان باردار دیده می‌شود و به دیابت بارداری معروف است [۴]. به منظور تشخیص بیماری دیابت علاوه بر تحقیقات متعدد، مطالعاتی براساس ویژگی‌های ۸ گانه سازمان جهانی بهداشت یعنی سن، سابقه بیماری دیابت، نمایه توده بدنی (kg/m^2)، انسولین سرم دو ساعته ($\mu\text{U/ml}$)، ضخامت پوست ماهیچه سه سر بازویی (mm)، فشار خون (mm gh)، غلظت گلوکز پلاسمای خون در دو ساعت، دفعات بارداری صورت گرفته است که با توجه به نتایج، استفاده از تکنیک‌های داده‌کاوی و هوش مصنوعی توصیه شده است [۶، ۵]. مشکلی که در رابطه با این بیماری و مخصوصاً در نوع یک آن وجود دارد، افزایش و یا کاهش بیش از حد سطح قند خون است که می‌تواند به بیهوشی، اغماء و حتی مرگ بیمار منجر شود [۷]. بنابراین استفاده از روشی مناسب به منظور پیش‌بینی و تعیین میزان صحیح دوز انسولین و در نهایت، پیش‌گیری از این عوارض گام مهمی در جهت کنترل بهتر این بیماری محسوب می‌شود. در همین راستا تحقیقاتی انجام گرفته است، از آن جمله می‌توان به مدل‌های ریاضی و مینیمال [۸-۱۰] براساس ویژگی‌های حساسیت بدن بیمار در پاسخ به گلوکز تزریقی، دوز انسولین مصرفی و سطح فعلی قند خون اشاره کرد که برای تعیین حساسیت بدن بیمار در پاسخ به گلوکز تزریقی و نیز پیش‌بینی کوتاه مدت سطح قند و انسولین است. در سال ۲۰۱۰ در [۱۱] با استفاده از شبکه‌های عصبی پیش‌بینی سطح گلوکز خون برای بیماران براساس ویژگی‌هایی در هر بیمار شامل وزن، زمان اولیه تزریق، زمان انتهایی تزریق و کربوهیدرات مصرفی، انجام شده است. در مطالعه‌ای با استفاده از ویژگی‌های ۱۳ گانه بانک اطلاعاتی NASA [۱۲] و در مطالعه‌ای دیگر توسط روش

برنامه‌ریزی LSSVM^۴ پیش‌بینی^۵ دوز انسولین صورت گرفته است [۱۳]. در یک مطالعه دیگر، یک سیستم حلقه بسته عملی آزمایشگاهی به صورت آنلاین برای کنترل گلوکز خون براساس جنسیت، سن، مدت مصرف، هموگلوبین HbA1c، دوز انسولین و کربوهیدرات مصرفی و ویژگی‌های آماری شامل میانگین و انحراف از معیار در ویژگی‌های بیان شده، تهیه شده است [۱۴]. Pappada و همکاران در سال ۲۰۰۸ تنها برای ۱۸۰-۵۰ دقیقه توانستند براساس ویژگی‌های نحوه زندگی، وضعیت روحی، دوز انسولین، وعده غذایی و کربوهیدرات مصرفی، پیش‌بینی سطح انسولین برای بیماران را انجام دهند [۱۵]. Stavroula و همکاران با استفاده از ۴ پارامتر شامل عملکرد کوتاه، متوسط، سریع، بلند مدت و کربوهیدرات مصرفی و غلظت قند خون که تابعی از زمان بعد از تزریق انسولین هستند، توسط شبکه‌های عصبی میزان دوز انسولین را برای بیماران مشخص کردند [۱۶]. Wahab و همکاران با استفاده از کنترل کننده تطبیقی بر اساس ۲ معادله دیفرانسل مرتبه اول برای انسولین و تنظیم گلوکز در ۲ زیر سیستم، تعیین دوز انسولین را انجام داده‌اند [۱۷]. در نهایت فرآیند نهایی به نتیجه رسیدن این تحقیق تحت عنوان روش پیشنهادی به این صورت است که، ابتدا داده‌ها نرمالیزه می‌شوند، سپس توسط الگوریتم ترکیبی فروکاهی (SVM - BPSO) تعداد داده‌ها (متغیر داده‌ها) کاهش یافته، سپس بانک داده براساس داده‌های جدید، وارد سیستم فازی - عصبی تطبیقی به منظور پیش‌بینی صحیح دوز انسولین می‌شود. در نهایت داده‌ها تحت عمل عکس فرآیند نرمالیزه قرار می‌گیرند تا مقدار واقعی مشخص شود.

مساله انتخاب ویژگی و یا بازشناسی الگو

انتخاب ویژگی و بازشناسی الگو در واقع برگزیدن ویژگی‌هایی است که حداکثر توان را در پیشگویی خروجی دارا باشند [۱۸]. تعریف اینکه زیر مجموعه بهینه چه می‌تواند باشد، به مسئله‌هایی که قصد داریم حل کنیم، وابسته است [۱۹]. اهمیت دادن به تمام ویژگی‌ها بر این اساس همیشه مناسب نیست. همان‌گونه که در برخی مطالعات به اثبات رسیده است [۲۰]، با اینکه انتظار می‌رود

رفته است. طی دهه گذشته محققان روی الگوریتم‌های جستجوی تکاملی مثل الگوریتم ژنتیک [۲۳]، ACO⁶ [۲۴] و ترکیب ACO/PSO⁷ [۲۵] تمرکز کرده‌اند.

روش‌ها

جمعیت مورد مطالعه، شامل ۳۱۲ فرد مبتلا به دیابت نوع دو و مشکوک به بیماری دیابت نوع دو، با نمونه‌گیری تصادفی سیستماتیک در فاصله مهرماه ۱۳۹۰ تا انتهای اردیبهشت ۱۳۹۱ با حمایت دانشگاه علوم پزشکی سبزوار در محل مرکز تحقیقات دیابت شهرستان سبزوار با کمک افرادی که به این مرکز طی فاصله سال‌های اسفندماه ۱۳۸۵ تا اسفندماه ۱۳۹۰ مراجعه کرده بودند انجام شد (تا انتهای تحقیق تنها از ۳۱۲ فرد بر اساس رابطه کوکرون، با توجه به تعداد کل جامعه نهایی آماری، امکان نمونه‌گیری میسر شد). پرسشنامه‌هایی طراحی شد. سپس توسط بیماران با همکاری پزشکان تکمیل شد، جدول ۱ نمونه‌ایی از متغیرهای ثبت شده بیماران است. نحوه مقدار دهی متغیرها بر این اساس بود که در چند متغیر مقدار واقعی آن‌ها قرار داده شده‌اند، مثلاً در ویژگی‌های شماره ۴، ۵ یا ۶ و در چند مورد دیگر با همکاری پزشکان و اظهارات بیماران از ۰ تا ۵ نمره دهی شده‌اند، مانند ویژگی‌های شماره ۳ و ۸.

رابطه کوکرون:

$$n = \frac{\frac{z^2 pq}{d^2}}{1 + \frac{1}{N} \left(\frac{z^2 pq}{d^2} - 1 \right)}$$

با داشتن اطلاعات بیشتر نتایج بهتر شود، اما در عمل مشاهده می‌کنیم که این مساله باعث افت دقت می‌گردد [۲۱]. بنابراین برای رسیدن به یک بازنمایی مناسب، انتخاب بهترین بردار ویژگی یا ساده‌سازی بردار ویژگی امری ضروری است. یک سیستم شناسایی الگوی متداول شامل ۴ بخش است، استخراج ویژگی، انتخاب ویژگی، طراحی و آموزش طبقه‌بندی کننده و سرانجام آزمایش. در این تحقیق از الگوریتم ژنتیک دودویی^۴ (GA³) برای کاوش و انتخاب ویژگی‌های موثر، الگوریتم ماشین بردار پشتیبان (SVM⁵) برای طبقه‌بندی نمونه‌ها و همین‌طور از شبکه‌های عصبی مصنوعی برای طراحی طبقه‌بندی کننده و آزمایش نهایی استفاده می‌گردد. الگوریتم‌های انتخاب ویژگی بسته به روند ارزیابی آن‌ها به دو دسته عمده تقسیم می‌شوند. اگر انتخاب ویژگی مستقل از هرگونه الگوریتم یادگیری انجام شود (یعنی به صورت یک پیش‌پردازنده کاملاً مجزا)، آن را روش فیلتر یا حلقه باز گویند. در این مورد ویژگی‌های نامطلوب قبل از استنتاج دور ریخته می‌شوند، اما، اگر روند ارزیابی با یک الگوریتم طبقه‌بندی در ارتباط باشد، روش انتخاب ویژگی را Wrapper یا حلقه بسته می‌نامند. این روش، جستجو در فضای زیر مجموعه‌ها را بر اساس تخمین دقت ناشی از انتخاب یک زیر مجموعه خاص تحت شرایط الگوریتم طبقه‌بندی مورد استفاده (به عنوان معیاری از بهینه‌گی آن زیر مجموعه) انجام می‌دهد [۲۲]. الگوریتم‌های دسته دوم معمولاً نتایج بهتری به دست می‌دهند. مهم‌ترین بخش هر روش انتخاب ویژگی حلقه بسته، الگوریتم جستجویی است، که در آن به کار

جدول ۱- پارامترها و ویژگی‌های مورد استفاده در این تحقیق

متغیرها
۱- سطح غلظت قند خون در ابتدای بازه زمانی (mg/d lit)
۲- دوز انسولین ۱۲ ساعته تزریقی به بدن (unit)
۳- سطح فعالیت بدنی (unit)
۴- سن
۵- BMI
۶- HDL
۷- LDL
۸- Total Cholesterol (mg/dl)
۹- Fast Blood Sugar (mg/dl)
۱۰- غلظت انسولین ۶ ساعته تزریقی به بدن (unit)
۱۱- سطح استرس (unit)
۱۲- کربوهیدرات مصرفی (unit)

$$y_i = \frac{x_i - \mu}{\sigma} \quad \text{رابطه ۱:}$$

فرآیند اجرای مراحل در روش پیشنهادی

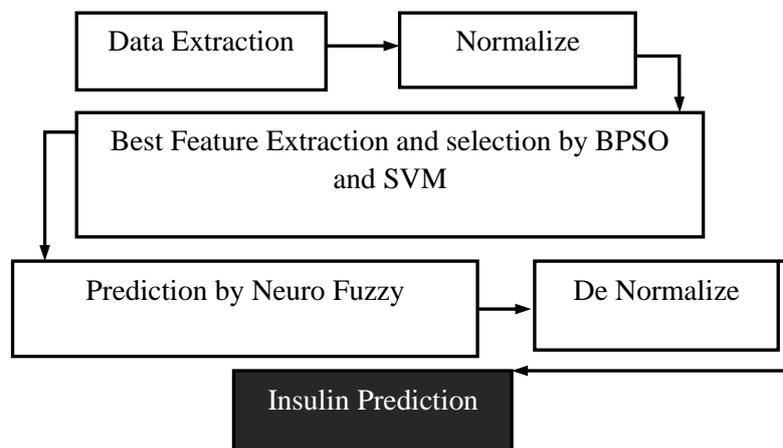
ابتدا بانک اطلاعاتی بر اساس داده‌ها تشکیل می‌شود، سپس توسط رابطه ۱ نرمالیزه می‌شود. اثبات رابطه:

$$z = \frac{x_i - \mu}{\sigma} \rightarrow \text{var}(z) = \text{var}\left(\frac{x_i - \mu}{\sigma}\right) \Rightarrow \text{by refer to } \text{var}(ax + b) = a^2 \text{var}(x) + b \text{ then } \frac{\text{var}(x) - \text{var}(\mu)}{\sigma^2} = \frac{\text{var}(x) - 0}{\sigma^2} = \frac{\sigma^2 - 0}{\sigma^2} = 1$$

$$E(z) = \frac{1}{\sigma}(E(x) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$

صحت نتایج) بهترین ویژگی‌ها انتخاب می‌شوند. مجدداً داده‌ها دنرمالایز می‌شوند؛ سپس بر اساس ویژگی‌های انتخاب شده، میزان صحیح دوز انسولین در بیماران به کمک شبکه فازی عصبی تعیین می‌شود. فلوجارت ۱ بیان کننده روش کار می‌باشد.

به راحتی مشخص است که هر مقداری وارد این رابطه شود به بازه ۰ و ۱ تبدیل خواهد شد. در این رابطه x_i بیانگر مقدار داده‌ها؛ μ نمایشگر میانگین داده‌ها و σ انحراف معیار داده‌ها می‌باشند. در ادامه توسط الگوریتم کاوشی (BPSO) به همراه الگوریتم طبقه‌بندی کننده (SVM) به منظور طبقه‌بندی در جهت برآورد دقت و



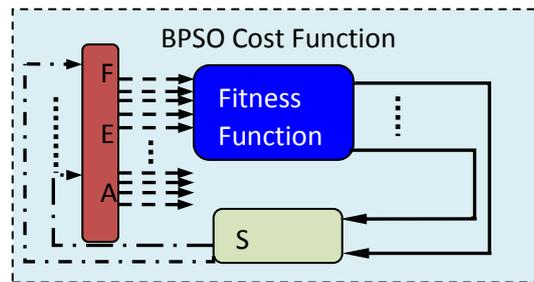
فلوجارت ۱- فلوجارت پیشنهادی به منظور پیش بینی دوز انسولین

جمعیت، مکان هر پرنده با دو مقدار بهترین (p_best , g_best) به روز می‌شود. در هر تکرار با توجه به روش الگوریتم مبتنی بر تکرار قاعده سعی و خطا تا حاصل شدن نتیجه صحیح، الگوریتم بعد از یافتن سرعت و موقعیت جدید پرنده را براساس روابط ۲ و ۳ به روز رسانی می‌کند. روند روش پیشنهادی تابع بهینه کننده به منظور اکتشاف و پیدا کردن بهترین متغیرها، در الگوریتم BPSO در فلوجارت ۲ می‌باشد.

۱-۱- الگوریتم بهینه‌سازی حرکت دسته جمعی ذرات (BPSO) الگوریتم حرکت دسته جمعی ذرات از جمله الگوریتم‌های جستجوی موازی مبتنی بر جمعیت است که توسط Kennedy و Eberhart در سال ۱۹۹۵ مطرح شد [۲۷، ۲۸] که با یک گروه از جواب‌های تصادفی (پرنده‌گان یا ذرات) شروع به کار می‌کند. هر ذره (پرنده) به صورت چند بعدی با دو بردار $X_i[t]$ و $V_i[t]$ که به ترتیب معرف موقعیت و سرعت فعلی در لحظه (t) از ذره i هستند مشخص می‌شود. در هر مرحله از حرکت

$$V_i[t+1] = w \times v_i[t] + C_1 r_1 (X^{i,pbest}[t] + X_i[t]) + C_2 r_2 (X^{gbest}[t] - X_i[t]) \quad \text{رابطه ۲:}$$

$$X_i[t+1] = X_i[t] + V_i[t+1] \quad \text{رابطه ۳:}$$



فلوچارت ۲- تابع هزینه الگوریتم حرکت دسته جمعی ذرات

محدود می‌شود، که عضوی از دامنه $[V_{min}, V_{max}]$ می‌باشد. رابطه ۳ نیز بردار موقعیت فعلی پرنده را با توجه به سرعت جدید آن بروز رسانی می‌کند. تغییر معادله بروز رسانی الگوریتم باینری در ۲ گام انجام می‌شود، الف: استفاده از رابطه ۲ برای تغییر سرعت و ب: اعمال تابع سیگموئید در رابطه ۴ برای محدود کردن سرعت بین $[0, 1]$. رابطه ۴ بازنویسی شد.

$$X_{ij}[t+1] = \begin{cases} 1 & \text{if } rand() < S(V_{ij}[t+1]) \\ 0 & \text{Other Wise} \end{cases}$$

$$S(v_{ij}[t+1]) = \frac{1}{1 + e^{-v_{ij}[t+1]}}$$

یبتی برای هر ذره، الگوریتم BPSO شروع می‌شود مانند بردار ۱.

در رابطه ۲، W ضریب اینرسی است (با در نظر گرفتن مقادیر معرفی شده در مطالعه Feng و همکاران [۲۹])؛ که در بازه $[0, 1.25]$ قرار دارد. $C1$ و $C2$ میزان تجربه شخصی و گروهی است که به‌طور معمول برای هر دوی آن‌ها از مقدار $1/5$ و یا $1/2$ استفاده شده است [۳۰، ۳۱]. $r1$ و $r2$ نیز اعداد تصادفی در بازه $[0, 1]$ می‌باشند. برای جلوگیری از واگرایی الگوریتم، سرعت پرنده به یک بازه رابطه ۴:

در اینجا به‌منظور بهینه‌سازی فضای جواب‌های مساله از مدل الگوریتم باینری حرکت دسته جمعی ذرات (BPSO) استفاده می‌کنیم. با در نظر گرفتن یک رشته دودویی ۱۲

1	0	0	1	1	0	1	0	1	1	1	0
---	---	---	---	---	---	---	---	---	---	---	---

بردار ۱:

با توجه به نتایج مطلوب [۳۲] تابع برازندگی در رابطه ۵ را پیشنهاد می‌کنیم.

که در بردار ۱ "0" نمایانگر حذف ویژگی و "1" نمایانگر انتخاب آن ویژگی می‌باشد. البته در اینجا دقت شناسایی مهم‌تر از کوچک بودن زیر مجموعه منتخب است، بنابراین رابطه ۵:

$$Fitness = \alpha \cdot Accuracy + \beta \cdot \frac{|n| - |s|}{|n|}$$

می‌کنیم. در این تحقیق دقت شناسایی مهمتر است، پس α بسیار بزرگتر از β خواهد بود. Accuracy یا میزان صحت، از مقایسه بین بردار بیماران شناسایی شده توسط سیستم و بیماران واقعی به‌دست می‌آید. پارامترهای BPSO مطابق جدول ۲ تنظیم شده‌اند. در رابطه ۵ صحت انجام کار توسط رابطه ۶ معرفی می‌شود.

در این رابطه $|n|$ کل ویژگی‌ها و $|s|$ ویژگی‌های منتخب است. جمله اول ضریبی از دقت شناسایی و جمله دوم ضریبی از نرخ کاهش ویژگی‌هاست. α و β ضرایب ثابت پیشنهادی در رابطه ۵ هستند. به‌خاطر ارجحیت دادن به دقت شناسایی نسبت انتخاب ویژگی مجموع ضرایب α و β را ثابت و برابر ۱۰۰ پیشنهاد

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Fn + Tn + Fp} \quad \text{رابطه ۶:}$$

تعداد افراد سالم است که مدل تصمیم به اشتباه آن‌ها را مبتلا به دیابت تشخیص داده است. Fn یا منفی نادرست نیز بیانگر تعداد افراد بیمار است که مدل تصمیم به اشتباه آن‌ها را سالم تشخیص داده است.

در رابطه ۶، Tp یا مثبت درست، بیانگر تعداد افرادی است که مبتلا به دیابت هستند و مدل تصمیم به درستی آن را تشخیص داده است، Tn یا منفی درست، تعداد افراد سالم است که مدل تصمیم به درستی آن را سالم تشخیص داده است. Fp یا مثبت نادرست نیز بیانگر

جدول ۲- پارامترها و مشخصات الگوریتم بهینه سازی (کاوشی) BPSO

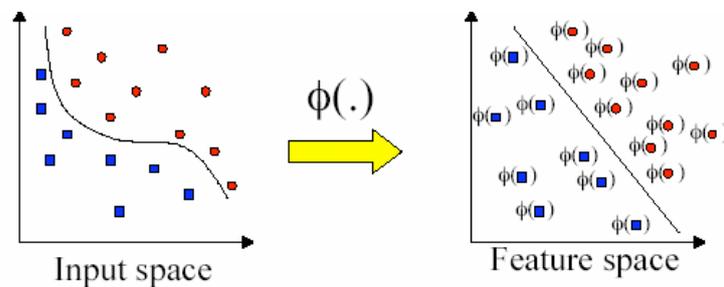
پارامترها	مشخصات
الگوریتم بهینه سازی	الگوریتم BPSO
جمعیت اولیه	۲۰ ذره
Maxiter	۱۰۰ تلاش
C1 یا ضریب تجربه شخصی	۱ [۱۹، ۲۵]
C2 یا ضریب تجربه گروهی	۲۰ [۱۹، ۲۵]
W یا ضریب اینرسی	بین ۰ تا ۱ به صورت نزولی [۲۵]
Vmax	$[۱۹، ۲۵] + 4V$
Vmin	$[۱۹، ۲۵] - 4V$

BPSO

با چند زیر مجموعه تصادفی شروع به کار می‌کند و الزام دارد تعدادی زیر مجموعه را که حاوی مفیدترین اطلاعات هستند، در نهایت معرفی کند. بردارهای ویژگی معرفی شده توسط BPSO برای آموزش طبقه‌بندی کننده SVM به کار برده می‌شوند.

الگوریتم طبقه بندی SVM

SVM یک الگوریتم یادگیری با سرپرستی است که می‌تواند در کاربردهایی نظیر جداسازی و طبقه‌بندی داده مورد استفاده قرار گیرد [۳]. این الگوریتم می‌تواند داده‌های آموزش را به دو دسته تقسیم کند، اصولاً داده‌ها توسط یک خط قابل جداسازی هستند، و در غیر این صورت الگوریتم با نگاشت ویژگی‌ها به وسیله تابع ϕ به ابعاد بالاتر سعی می‌کند توسط یک فرا صفحه با دقت بالاتری داده‌ها را تفکیک نماید (شکل ۱).



شکل ۱- نگاشت داده‌ها از فضای ورودی به فضای ویژگی [۳]

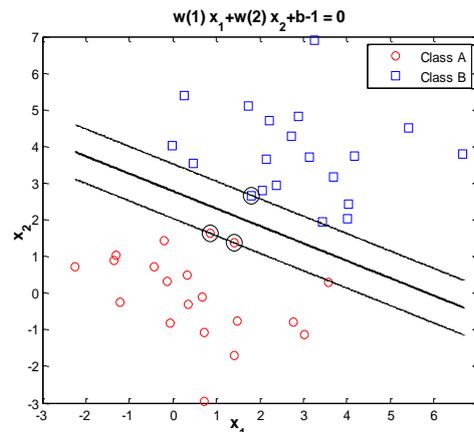
که در این روابط x متغیر ورودی، w بردار نرمال خط جدا کننده و b عرض از مبدا خط جدا کننده است. در هر تلاش طبقه‌بندی نمونه‌های خروجی از الگوریتم BPSO، توسط الگوریتم ماشین پشتیبان SVM انجام می‌شود تا در نهایت کمترین و بهترین ویژگی‌هایی که منجر به ایجاد بهترین کلاس‌بندی در داده‌ها می‌شود، تعیین شوند. به‌منظور کاهش محاسبات در کلاس‌بندی، از تابع خطی استفاده شده است. در اثر اعمال این الگوریتم در ویژگی سوم، شکل ۲ به‌دست می‌آید.

رابطه ۷ رابطه فرا صفحه طبقه‌بندی کننده داده‌ها است. روابط ۸ و ۹ روابط فرا صفحه‌های موازی براساس شرایط حداکثر حاشیه می‌باشند. در صورت ترسیم این توابع ساده فاصله بین دو صفحه حاشیه برابر $\frac{2}{\|w\|}$ می‌باشد [۳].

$$\text{رابطه ۷: } w \times x - b = 0$$

$$\text{رابطه ۸: } w \times x - b = -1$$

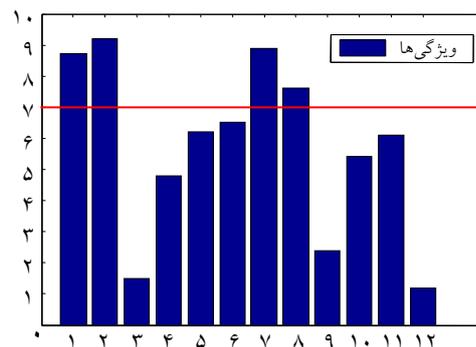
$$\text{رابطه ۹: } w \times x - b = 1$$



شکل ۲- طبقه‌بندی داده‌ها به ۲ گروه در ویژگی سوم

پشتیبان (SVM)، ویژگی‌های شماره ۸، ۷، ۲، ۱، (ویژگی‌های بهینه و مفید) که در ۱۰۰ مرتبه اجرای سیستم پیشنهادی به‌دست آمدند در شکل ۳ مشخص شدند.

پس از اجرای مراحل مختلف الگوریتم پیشنهادی به‌منظور مشخص شدن بهترین و موثرترین ویژگی‌ها، یعنی اعمال الگوریتم‌های بهینه‌سازی (کاوشی) و الگوریتم بردار ماشین

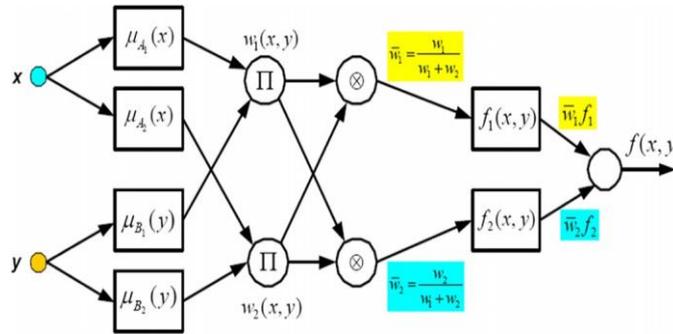


شکل ۳- مشخص شدن بهترین و مفیدترین متغیرها از سایر متغیرها در فرآیند کاوش به‌منظور مشخص شدن بهترین متغیرها

می‌شود، سپس باید به مرحله بعد، یعنی شبکه فازی - عصبی به‌منظور پیش‌بینی صحیح میزان دوز انسولین براساس الگوی

بعد از اینکه متغیرهای مهم مشخص شدند، بانک اطلاعاتی با ابعاد ۳۱۲ سطر (افراد مورد مطالعه) و ۴ ستون (ویژگی) مرتب

استفاده در شکل ۴ دیده می‌شود که دایره‌ها بیان کننده گره‌های ثابت و مربع‌ها بیان کننده گره‌های تطبیقی هستند. در لایه اول، تمامی گره‌ها تطبیقی هستند و خروجی این لایه، درجه عضویت فازی ورودی‌هاست که با روابط ۱۰ و ۱۱ به دست می‌آید [۳۳].



شکل ۴- سیستم انطباقی فازی عصبی

فازی سوگنو با دو ورودی x و y یک خروجی f داشته باشد، برای یک مدل فازی سوگنو مرتبه اول، مجموعه قوانین مشترک با دو قانون If-then به صورت زیر است. در رابطه ۱۰ و ۱۱، X داده‌های ورودی (متغیر یا ویژگی‌های بهینه) برای هر قانون و Y داده خروجی (میزان صحیح دوز انسولین) برای هر قانون می‌باشد.

در این سیستم‌ها با استفاده از مجموعه داده‌های ورودی (۴) متغیر بهینه مشخص شده برای هر نمونه) - خروجی (دوز صحیح انسولین برای هر نمونه) در مدل پیش‌بینی کننده دوز انسولین به عنوان داده‌های آموزشی، یک سیستم استنباط فازی از نوع سوگنو ایجاد می‌شود که پارامترهای تابع عضویت آن با استفاده از الگوریتم پس انتشار و روش کوچک‌ترین مربعات تنظیم می‌گردد [۳۴]. اگر یک سیستم

$$\text{رابطه ۱۰: Rule 1: If } X \text{ is } A_1 \text{ and } Y \text{ is } B_1 \text{ then } f_1 = \mu_1 \times x + q_1 y + r_1$$

$$\text{رابطه ۱۱: Rule 2: If } X \text{ is } A_2 \text{ and } Y \text{ is } B_2 \text{ then } f_2 = \mu_2 \times x + q_2 y + r_2$$

ساختار الگوریتم ANFIS استفاده شده در این تحقیق از Tahami و همکاران اقتباس شده است [۳۴] (جدول ۳).

جدول ۳- پارامترهای ساختار الگوریتم ANFIS

روش تولید سیستم فازی	الگوریتم میانگین فازی
الگوریتم آموزش	پس انتشار خطا
تعداد خوشه‌های الگوریتم میانگین فازی	۱۰
مشخصات آموزش در سیستم پیشنهادی	
۱۰۰ مرتبه	حد نهایی تکرار
۰/۹	میزان نرخ کاهشی
۱/۱	میزان نرخ افزایشی
۰	خطای هدف
۰/۰۱	میزان نرخ انتسابی اولیه

نتایج ارزیابی و آزمایش‌های نهایی

ویژگی‌های اصلی، سرنوشت‌ساز و تاثیر گذار در تعیین میزان دوز انسولین در بیماران تعیین شدند، سپس شبیه‌سازی نهایی توسط شبکه عصبی فازی-مصنوعی (ANFIS) مطابق شکل ۵ برای ۴۸ ساعت آینده بیمار به صورت هر ۶ ساعت تهیه شده است. البته مقادیر پس از اینکه توسط شبکه فازی عصبی پیش‌بینی شدند، توسط رابطه ۱۲ دنرمالیزه می‌شوند. پس از دنرمالیزه شدن، نتایج شبیه‌سازی در شکل ۵ برای بیمار تصادفی شماره ۱۵ مشخص شده است. در شکل ۵ مقدار واقعی انسولین تجویز شده برای فرد در طول ۴۸ ساعت آینده توسط بارهای فیروزه‌ای مشخص شده است، این در حالی است که مقدار پیش‌بینی شده توسط سیستم پیشنهادی توسط بارهای آبی پررنگ نشان داده شده است. در ابتدای هر بار با فرض ۰/۵ درصد، خط خطا (error bar) با نقاطی نمایش داده شده که اگر این نقاط را در بارهای فیروزه‌ای به یکدیگر متصل نماییم، خط ممتد قرمز و اگر نقاط خط خطا را در بارهای آبی پررنگ به یکدیگر متصل نماییم، خط مشکی غیر ممتد مشخص می‌شود. مقیاس مقادیر در این نمودار در حالت یکسانی (per unit) شده که در نهایت توسط رابطه ۱۲ به حالت واقعی (دنرمالیزه) بازگشت داده می‌شود.

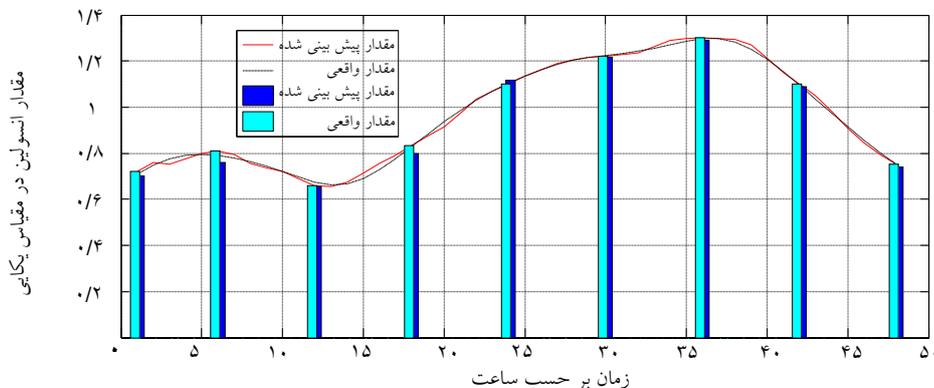
رابطه ۱۲:

$$x_i = y_i \times \sigma^2_i + \mu$$

با توجه به سیستم پیشنهادی و جدول ۳، روش تولید سیستم فازی توسط روش فازی میانگین فازی (FCM) با تعداد ۱۰ کلاستر (خوشه)، استفاده شده است. به طوری که الگوریتم آموزش شبکه انطباقی فازی - عصبی، از نوع پس انتشار خطا، با ۱۰۰ تکرار، نرخ کاهش ۰/۹، نرخ افزایشی ۱/۱ با مقدار ۰/۰۱ در هر تکرار و خطای نهایی در حد صفر تعیین شده است.

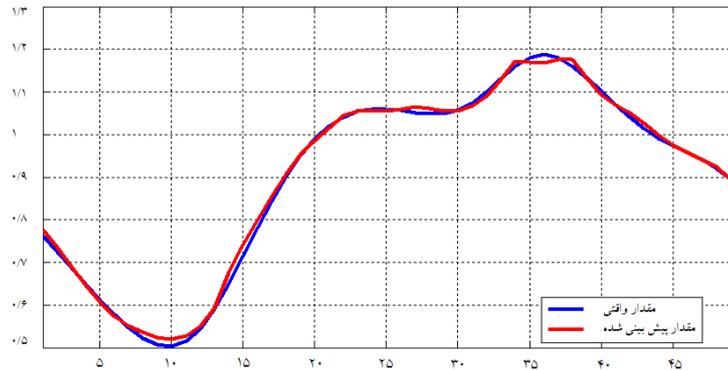
بررسی صحت آزمایش و ارزیابی نهایی

فرآیند انتخاب ویژگی که منجر به پیش‌بینی مقدار صحیح دوز انسولین می‌شود به این صورت است که توسط الگوریتم باینری حرکت دسته جمعی ذرات براساس تابع هزینه تعریف شده (کاوش به منظور بررسی ویژگی‌ها) و به کمک الگوریتم طبقه‌بندی ماشین بردار پشتیبان (SVM) موثرترین ویژگی‌ها مشخص می‌شوند، سپس شبکه فازی عصبی و بانک داده مبتنی بر ویژگی‌های موثر، پیش‌بینی صحیحی از میزان دوز انسولین برای بیماران تهیه می‌کند. برای اثبات قابلیت‌های سیستم پیشنهادی، اعتبار سنجی عرضی ۱۰ دسته‌ای (توسط K-fold) بر روی پایگاه داده ۳۱۲ نمونه‌ای موجود پیاده شد که به ۱۰ دسته تقسیم شدند. ۸ دسته ۳۱ داده‌ای و ۲ دسته ۳۲ داده‌ای. در این اعتبار سنجی، ۹ قسمت از داده‌ها برای آموزش و ۱ قسمت برای آزمایش به کار می‌روند، آموزش و آزمایش ۱۰ بار به صورت متوالی چرخشی انجام می‌شود به طوری که هر بار یک قسمت برای آزمایش کنار گذاشته می‌شود.



شکل ۵- نتایج پیش‌بینی دوز انسولین و میزان واقعی برای بیمار تصادفی شماره ۱۵

البته برای نمایش بهتر میزان تهیه شده و یا معرفی شده، از شکل ۶ استفاده می‌نماییم، که یک نمای ۳ بعدی از شبیه‌سازی و تعیین دوز دارو برای ۴۸ ساعت آینده بیمار می‌باشد.



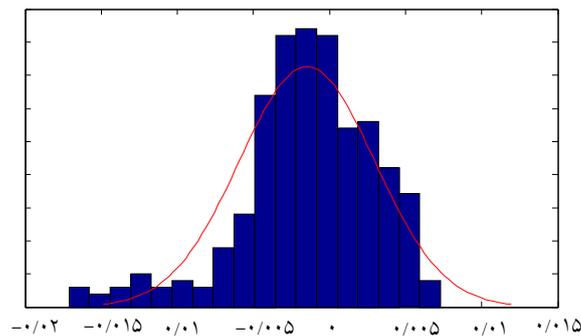
شکل ۶- پیش بینی دوز صحیح داروی انسولین برای ۴۸ ساعت آینده بیمار تصادفی شماره ۱۵

پیش‌بینی صحیح دوز انسولین بیماران، خطای مقدار صفر را در حدود ۸۰ بار تجربه کرده است، البته ناگفته نماند که این خطا به ازای هر بیمار تعریف شده است. به‌منظور بررسی و مقایسه عملکرد سیستم پیشنهادی، پارامترهایی طبق روابط ۱۴ و ۱۵ و ۱۶ تعریف می‌شوند. البته نتایج این پارامترها در جدول ۴ مشخص شده‌اند.

رابطه ۱۳:

$$\bar{e} = \frac{\sum_{i=1}^n \|y_i - \hat{y}_i\|^2}{n}$$

به‌منظور بررسی خطای عملکرد شبکه فازی عصبی، رابطه ۱۳ تعریف می‌شود که نمودار آن در شکل ۷ رسم شده است. اساس کار شبکه به‌منظور انطباق فضای ورودی به خروجی، بر طبق انطباق با تکیه بر فرآیند سعی و خطا است، خطایی است که توسط رابطه ۱۳ محاسبه می‌شود، که بیان‌کننده میانگین مربعات ناشی از اختلاف میانگین مربعات مقادیر خروجی (y_i) با مقادیر واقعی (\hat{y}_i) است. در شکل ۷ به‌طور مثال مشخص می‌شود که این شبکه برای



شکل ۷- خطای عملکرد شبکه فازی عصبی

$$\text{Speciality} = \frac{Tn}{Tn + Fp} \quad \text{رابطه ۱۴:}$$

$$\text{Sensitivity} = \frac{Tp}{Tp + Fn} \quad \text{رابطه ۱۵:}$$

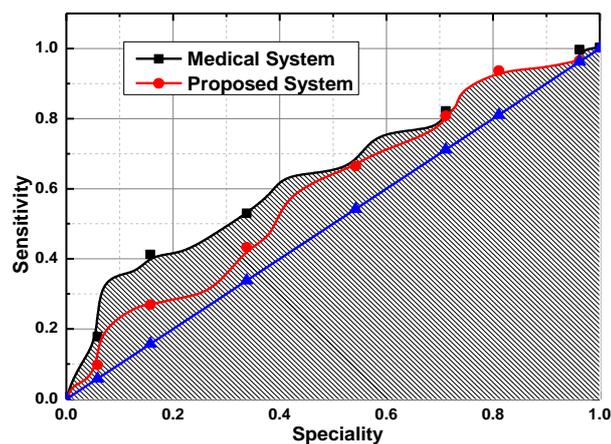
$$\text{Total Accuracy} = \frac{\text{Total True}}{\text{Total input}} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad \text{رابطه ۱۶:}$$

جدول ۴- نتایج عملکرد سیستم‌فازی عصبی در پارامترهای ۱۴، ۱۵، ۱۶

پارامتر	درصد
Sensitivity (%)	۹۱/۲
Specialty (%)	۸۴/۱
Total Accuracy (%)	۹۲/۹

می‌باشد که بیان کننده احتمال مثبت یک تست است که به بیماران مبتلا نسبت داده شده است. ویژگی در یک آزمون، بیان کننده احتمال منفی یک تست می‌باشد که به بیماران نسبت داده شده است [۳۵]. شکل ۸ نمودار ROC هر ۲ سیستم برای ۳۰ مورد بیمار تصادفی آزمایشی است که، میزان حساسیت و ویژگی آن‌ها براساس رفتار سیستم پیشنهادی و عملکرد پزشک مشخص شده است.

منحنی ROC6 مشخصه‌ای دیگری برای طبقه‌بندی صحیح و بیان عملکرد بین سیستم‌ها در شرایط کاری یکسان است. همان‌طور که در شکل ۸ می‌بینید، مشاهده حساسیت و ویژگی در هر سیستم بر حسب دیگری بیان شده است. در موارد و شرایط ایده‌آل کاری، با افزایش حساسیت، کاهش کمی در ویژگی وجود دارد، تا زمانی که به سطح بسیار بالایی از حساسیت رسیده باشیم. حساسیت، یک آزمون



شکل ۸- منحنی ROC به منظور مقایسه سیستم واقعی و پیشنهادی در داده‌های آزمایشی

پزشکان به‌طور قابل توجهی نزدیک شد (منظور بیماران مطالعه پیش رو است که برای آن‌ها دارو تجویز شده است). البته به دلیل عدم دسترسی به منابع تحقیقات قبلی که در زمینه پیش‌بینی و تعیین صحیح دوز انسولین انجام گردیده است، مقایسه قابل ملاحظه‌ای در این زمینه نمی‌توان انجام داد، اما در نهایت هدف تطبیق و یکسان بودن نتایج سیستم‌های پیشنهادی با تجویز پزشکی در هر حالت مهم است، که خوشبختانه سیستم پیشنهادی با توجه به شکل‌های (۶ و ۸) و جدول ۴ عملکرد مناسب و قابل قبولی داشته است. همچنین مقایسه نتایج به دست آمده در این تحقیق با نتایج به دست آمده از تصمیم‌گیری‌های

البته، هر چه قدر مساحت زیر نمودار سیستم پیشنهادی به سیستم واقعی نزدیک‌تر باشد نشان دهنده عملکرد نزدیک سیستم پیشنهادی به سیستم واقعی یا تصمیمات پزشکان است.

نتیجه‌گیری

در این تحقیق با توجه به ضرورت تعیین صحیح و به موقع میزان انسولین برای بیماران، روشی نوین مبتنی بر ترکیب سیستم‌های هوشمند ارائه شده است، به طوری که نتایج موجود به دست آمده به نتایج واقعی یا همان تجویزات

افزایش کارایی این شبکه‌ها در جهت پیش‌بینی صحیح باشند. البته هدف اصلی در این نوع تحقیقات به دست آوردن صحت و دقت ۱۰۰ درصدی می‌باشد، تا بتوان به سیستم‌های هوشمند مصنوعی اطمینان داشت و بعد عملی به آن‌ها بخشید. میل و رسیدن به چنین هدفی تنها در صورت دانش بالا، فهم درست و تسلط کافی از یک طرف به سیستم‌های هوشمند، و از طرفی دیگر به بیماری و داده‌کاوی امکان‌پذیر است که در آینده‌ایی نه چندان دور محقق خواهد شد.

پزشکان، نشان دهنده عملکرد قابل ملاحظه در میزان صحت پیش‌بینی سری زمانی سطح غلظت قند خون است، چرا که سیستم پیشنهادی موفق شد با در دست داشتن چند ویژگی از بیماران مبتلا به دیابت، سطح قند خون تا ۴۸ ساعت آینده بیمار را پیش‌بینی نماید. از شبکه‌های مصنوعی تطبیقی فازی عصبی به منظور انطباق، یادگیری و دسته‌بندی استفاده شد. چالش آموزش شبکه‌های عصبی مهم‌ترین دغدغه می‌باشد. ترکیب الگوریتم‌های تکاملی و سیستم‌های داده کاوی می‌تواند یک ایده مناسب برای

مآخذ

- World Health Organization, Diabetes Center, Fact Sheet N312, Retrived from: www.who.int/mediacentre/factsheets/fs312/en
- Centers for Disease Control and Prevention, National Diabetes, Fact Sheet 2011 Retrived from: www.cdc.gov/diabetes
- نعیم آبادی م، امیراحمدی چماچار ن، تهامی ا و ربانی ح. تشخیص دیابت با استفاده از SVM و MLP، چکیده نامه چهاردهمین کنفرانس دانشجویی مهندسی برق ایران؛ کرمانشاه؛ ایران؛ کرمانشاه، دانشگاه صنعتی کرمانشاه؛ ص ۲۵۱-۲۵۸.
- American Diabetes Association, Diabetes Basics. Retrived from: www.diabetes.org/diabetes-basics,
- فیوضی م، قره خانی ا، حدادنیا ج. تشخیص بیماری دیابت با استفاده از سیستم‌های فازی، درخت تصمیم‌گیری و شبکه‌های عصبی، چکیده نامه چهارمین کنفرانس ملی فناوری اطلاعات و دانش ایران؛ بابل؛ ایران، دانشگاه صنعتی کرمانشاه، ص ۱۲۵-۱۳۱.
- فیوضی م، حدادنیا ج، ملانیا ن. تشخیص بیماری دیابت براساس روش ترکیبی هوشمند. مجله علمی و ترویجی هوش مصنوعی و ابزار دقیق ایران؛ ۱۳۹۱؛ ۱ (۳۲): ۸-۱.
- قوچانی س، تهامی، س. پیش‌بینی نوسانات سطح قند خون در بیماران مبتلا به دیابت نوع ۱ با استفاده از شبکه‌های عصبی خود بازگشتی الممن. مجله علمی و پژوهشی فنی و مهندسی دانشگاه آزاد اسلامی مشهد؛ ۱۳۸۷؛ ۲ (۱): ۹۴-۸۱.
- Li J, Kuang Y, Mason CC. Modeling the glucose-insulin regulatory system and ultradian insulin secretory oscillations with two explicit time delays. *Journal of Theoretical Biology* 2006; 242:722-735.
- Leaning MS and Boroujerdi MA. A system for compartmental modelling and simulation. *Computer Methods programs biomed* 1991; 35 (2):148-157.
- Tresp V, Moody J, and DeLong WZ. Neural modelling of physiological processes. *Computational Learning Theory and Natural Learning Systems* 1994; 2(1): 250-258.
- Huang HP, Liu SW, Chien IL, Lin CH. A Dynamic Model with Structured Recurrent Neural Network to Predict Glucose-Insulin Regulation of Type 1 Diabetes Mellitus. Proceeding of 9th International Symposium on Dynamics and Control of Process Systems (DYCOPS 2010), Leuven, Belgium, pp 25-31, July 5-7, 2010.
- Lean Yu. An evolutionary programming based asymmetric weighted least squares support vector machine ensemble learning methodology for software repository mining. *Information Sciences* 2012; 191(12): 31-46.
- NASA MDP repository. Retrived from: <http://mdp.ivv.nasa.gov>,
- Fernandez de Canete J, Gonzalez-Perez S, Ramos-Diaz JC. Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes. *Computer methods and Programs in biomedicine Journal* 2008; 4(2): 128-135.
- Scott M, Pappada BS, Brent D, Rosman PM. Development of a Neural Network for Prediction of Glucose Concentration in Type 1 Diabetes Patients. *Journal of Diabetes Science and Technology* 2008; 2(5): 189-197.
- Mougiakakou SG, Prountzou A, Iliopoulou D, Nikita KS, Andriani V, Bartsocas CS. Neural Network based Glucose - Insulin Metabolism Models for Children with Type 1 Diabete. *Engineering in Medicine and Biology Society* 2006; 1: 3545-8. 28th Annual International Conference of thr IEEE.

17. Wahab A, Kong YK, and Quek C. *Model Reference Adaptive Control on Glucose for the Treatment of Diabetes Mellitus*. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems CBMS'06, pp 250-258, 2006.
18. Jensen R. *Combining rough and fuzzy sets for feature selection*. Ph.D. Thesis. School of informatics, Univ. Edinburgh, 2005.
۱۹. علیپور م، حدادنیا ج. معرفی یک روش دقیق برای تشخیص سرطان پستان. *مجله بیماری‌های سرطان پستان* ۱۳۸۹؛ دوره ۲، شماره ۲، ص ۱۵-۲۲.
20. Xiang T and Gong S. Spectral clustering with eigenvector selection. *Pattern Recognition* 2008. 41(3): 1012-1029.
21. Toosi SAZ, Sadoghi Yazdi H. *Spectrum Clustering by K-PCA*. Proceeding of 16th CSI Conference, pp 118- 124, Winter 2010.
22. Kanan H, Faez K, Hosseinzadeh M. *Face recognition system using ant colony optimization-based selected features*. Proceeding of IEEE Symp on Computational Intelligence in Security and Defense Applications pp 57-62, 2007.
23. Siedlecki W, Sklansky J. A note on genetic algorithms for large-scale feature selection. *Journal of pattern Recognition Letters* 1989; 10 (5): 335-47.
24. Liu B, Abbess H, McKay B. Classification rule discovery with ant colony optimization. *Journal of IEEE Computational Intelligence Bulletin* 2004; 3(1): 31-5.
25. Meng Y. A swarm intelligence based algorithm for proteomic pattern detection of ovarian cancer. *Computational Intelligence and Bioinformatics and Computational Biology* 2006, 24 (8): 1-7.
26. Newman DJ, Hettich S, Blake CLS, & Merz CJ. 1998 UCI. *Repository of machine learning database*, Irvine, CA: University of California, Dept. of Information and Computer Science. Retrired from: archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes
27. Pan H, Ma Q, Wei X. Research on Fault Diagnosis of Gearbox Based on Particle Swarm Optimization Algorithm. *IEEE 3rd International Conference on Mechatronics* 2006; (ICM' 3rd), pp 254-260.
28. Sheybani M, meibody M. *PSO-LA: A Novel of Optimization*. Proceeding of 12th Conference on Computer Engineering, Iran, pp 1162 – 1168, 2006.
29. Feng X, Zhang J, and Yang Z. *Adaptive Particle Swarm Optimization on Individual Level*. Proceeding of IEEE conference, China, pp. 1215- 1218, 2002.
30. Norozie Beyrami M. *Improve the convergence of the algorithm of particle swarm*. Proceeding of Conference on industrial engineering, Osco IAU, pp 205-211, 2008.
31. Kennedy J, and Eberhart R. *Particle swarm optimization*. Proceeding of Conference on Evolutionary Computation IEEE, USA, pp. 1942-1948, 1995.
32. Tahami SE, Khalil Zadeh MA. *Intelligent Diagnosis by Multi-Layer Neural Network*. Proceeding of 12th Iranian Conference of Biomedical Engineering, Tabriz, Iran, pp 89-95, 2005.
33. Kemal Polat, Salih Güne. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing* 2007; p 702–710.
34. Tahami SE, Bamashki SM, Khalil Zadeh MA. *Diabet Diagnosis by GA-NN, ANFIS*. Proceeding of 1th Join Conference of Fuzzy and Intelligent System, summer 2007, Mashhad, Iran, pp 115-121.
35. Rajeswari K, Vaithiyanathan V. Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network. *International Journal of Computer Science and Network Security* 2011; 11(4): p 56-64.

مقدمه

بیماری دیابت یکی از شایع‌ترین بیماری‌های شناخته شده در دنیا می‌باشد. براساس گزارش‌ها، حدود ۲۲۰ میلیون نفر در جهان از این بیماری رنج می‌برند [۳-۱]. دیابت به ۳ نوع تقسیم می‌شود؛ نوع یک یا دیابت وابسته به انسولین که بیشتر در سن پایین و کودکان دیده می‌شود، نوع دو یا دیابت غیر وابسته به انسولین که در ۹۰ تا ۹۵ درصد بیماران دیابتی مشاهده می‌شود [۱]، و نوع سوم که بیشتر در زنان باردار دیده می‌شود و به دیابت بارداری معروف است [۴]. به منظور تشخیص بیماری دیابت علاوه بر تحقیقات متعدد، مطالعاتی براساس ویژگی‌های ۸ گانه سازمان جهانی بهداشت یعنی سن، سابقه بیماری دیابت، نمایه توده بدنی (kg/m^2)، انسولین سرم دو ساعته ($mu U/ml$)، ضخامت پوست ماهیچه سه سر بازویی (mm)، فشار خون ($mm gh$)، غلظت گلوکز پلاسمای خون در دو ساعت، دفعات بارداری صورت گرفته است که با توجه به نتایج، استفاده از تکنیک‌های داده‌کاوی و هوش مصنوعی توصیه شده است [۵، ۶]. مشکلی که در رابطه با این بیماری و مخصوصاً در نوع یک آن وجود دارد، افزایش و یا کاهش بیش از حد سطح قند خون است که می‌تواند به بیهوشی، اغماء و حتی مرگ بیمار منجر شود [۷]. بنابراین استفاده از روشی مناسب به منظور پیش‌بینی و تعیین میزان صحیح دوز انسولین و در نهایت، پیش‌گیری از این عوارض گام مهمی در جهت کنترل بهتر این بیماری محسوب می‌شود. در همین راستا تحقیقاتی انجام گرفته است، از آن جمله می‌توان به مدل‌های ریاضی و مینیمال [۸-۱۰] براساس ویژگی‌های حساسیت بدن بیمار در پاسخ به گلوکز تزریقی، دوز انسولین مصرفی و سطح فعلی قند خون اشاره کرد که برای تعیین حساسیت بدن بیمار در پاسخ به گلوکز تزریقی و نیز پیش‌بینی کوتاه مدت سطح قند و انسولین است. در سال ۲۰۱۰ در [۱۱] با استفاده از شبکه‌های عصبی پیش‌بینی سطح گلوکز خون برای بیماران براساس ویژگی‌هایی در هر بیمار شامل وزن، زمان اولیه تزریق، زمان انتهایی تزریق و کربوهیدرات مصرفی، انجام شده است. در مطالعه‌ای با استفاده از ویژگی‌های ۱۳ گانه بانک اطلاعاتی NASA [۱۲] و در مطالعه‌ای دیگر توسط روش

برنامه‌ریزی LSSVM^۴ پیش‌بینی^۵ دوز انسولین صورت گرفته است [۱۳]. در یک مطالعه دیگر، یک سیستم حلقه بسته عملی آزمایشگاهی به صورت آنلاین برای کنترل گلوکز خون براساس جنسیت، سن، مدت مصرف، هموگلوبین HbA1c، دوز انسولین و کربوهیدرات مصرفی و ویژگی‌های آماری شامل میانگین و انحراف از معیار در ویژگی‌های بیان شده، تهیه شده است [۱۴]. Pappada و همکاران در سال ۲۰۰۸ تنها برای ۱۸۰-۵۰ دقیقه توانستند براساس ویژگی‌های نحوه زندگی، وضعیت روحی، دوز انسولین، وعده غذایی و کربوهیدرات مصرفی، پیش‌بینی سطح انسولین برای بیماران را انجام دهند [۱۵]. Stavroula و همکاران با استفاده از ۴ پارامتر شامل عملکرد کوتاه، متوسط، سریع، بلند مدت و کربوهیدرات مصرفی و غلظت قند خون که تابعی از زمان بعد از تزریق انسولین هستند، توسط شبکه‌های عصبی میزان دوز انسولین را برای بیماران مشخص کردند [۱۶]. Wahab و همکاران با استفاده از کنترل کننده تطبیقی بر اساس ۲ معادله دیفرانسل مرتبه اول برای انسولین و تنظیم گلوکز در ۲ زیر سیستم، تعیین دوز انسولین را انجام داده‌اند [۱۷]. در نهایت فرآیند نهایی به نتیجه رسیدن این تحقیق تحت عنوان روش پیشنهادی به این صورت است که، ابتدا داده‌ها نرمالیزه می‌شوند، سپس توسط الگوریتم ترکیبی فروکاهی (SVM - BPSO) تعداد داده‌ها (متغیر داده‌ها) کاهش یافته، سپس بانک داده براساس داده‌های جدید، وارد سیستم فازی - عصبی تطبیقی به منظور پیش‌بینی صحیح دوز انسولین می‌شود. در نهایت داده‌ها تحت عمل عکس فرآیند نرمالیزه قرار می‌گیرند تا مقدار واقعی مشخص شود.

مساله انتخاب ویژگی و یا بازشناسی الگو

انتخاب ویژگی و بازشناسی الگو در واقع برگزیدن ویژگی‌هایی است که حداکثر توان را در پیشگویی خروجی دارا باشند [۱۸]. تعریف اینکه زیر مجموعه بهینه چه می‌تواند باشد، به مسئله‌هایی که قصد داریم حل کنیم، وابسته است [۱۹]. اهمیت دادن به تمام ویژگی‌ها بر این اساس همیشه مناسب نیست. همان‌گونه که در برخی مطالعات به اثبات رسیده است [۲۰]، با اینکه انتظار می‌رود

رفته است. طی دهه گذشته محققان روی الگوریتم‌های جستجوی تکاملی مثل الگوریتم ژنتیک [۲۳]، ACO⁶ [۲۴] و ترکیب ACO/PSO⁷ [۲۵] تمرکز کرده‌اند.

روش‌ها

جمعیت مورد مطالعه، شامل ۳۱۲ فرد مبتلا به دیابت نوع دو و مشکوک به بیماری دیابت نوع دو، با نمونه‌گیری تصادفی سیستماتیک در فاصله مهرماه ۱۳۹۰ تا انتهای اردیبهشت ۱۳۹۱ با حمایت دانشگاه علوم پزشکی سبزوار در محل مرکز تحقیقات دیابت شهرستان سبزوار با کمک افرادی که به این مرکز طی فاصله سال‌های اسفندماه ۱۳۸۵ تا اسفندماه ۱۳۹۰ مراجعه کرده بودند انجام شد (تا انتهای تحقیق تنها از ۳۱۲ فرد بر اساس رابطه کوکرون، با توجه به تعداد کل جامعه نهایی آماری، امکان نمونه‌گیری میسر شد). پرسشنامه‌هایی طراحی شد. سپس توسط بیماران با همکاری پزشکان تکمیل شد، جدول ۱ نمونه‌ایی از متغیرهای ثبت شده بیماران است. نحوه مقدار دهی متغیرها بر این اساس بود که در چند متغیر مقدار واقعی آن‌ها قرار داده شده‌اند، مثلاً در ویژگی‌های شماره ۴، ۵ یا ۶ و در چند مورد دیگر با همکاری پزشکان و اظهارات بیماران از ۰ تا ۵ نمره دهی شده‌اند، مانند ویژگی‌های شماره ۳ و ۸.

رابطه کوکرون:

$$n = \frac{\frac{z^2 pq}{d^2}}{1 + \frac{1}{N} \left(\frac{z^2 pq}{d^2} - 1 \right)}$$

با داشتن اطلاعات بیشتر نتایج بهتر شود، اما در عمل مشاهده می‌کنیم که این مساله باعث افت دقت می‌گردد [۲۱]. بنابراین برای رسیدن به یک بازنمایی مناسب، انتخاب بهترین بردار ویژگی یا ساده‌سازی بردار ویژگی امری ضروری است. یک سیستم شناسایی الگوی متداول شامل ۴ بخش است، استخراج ویژگی، انتخاب ویژگی، طراحی و آموزش طبقه‌بندی‌کننده و سرانجام آزمایش. در این تحقیق از الگوریتم ژنتیک دودویی^۴ (GA³) برای کاوش و انتخاب ویژگی‌های موثر، الگوریتم ماشین بردار پشتیبان (SVM⁵) برای طبقه‌بندی نمونه‌ها و همین‌طور از شبکه‌های عصبی مصنوعی برای طراحی طبقه‌بندی‌کننده و آزمایش نهایی استفاده می‌گردد. الگوریتم‌های انتخاب ویژگی بسته به روند ارزیابی آن‌ها به دو دسته عمده تقسیم می‌شوند. اگر انتخاب ویژگی مستقل از هرگونه الگوریتم یادگیری انجام شود (یعنی به صورت یک پیش‌پردازنده کاملاً مجزا)، آن را روش فیلتر یا حلقه باز گویند. در این مورد ویژگی‌های نامطلوب قبل از استنتاج دور ریخته می‌شوند، اما، اگر روند ارزیابی با یک الگوریتم طبقه‌بندی در ارتباط باشد، روش انتخاب ویژگی را Wrapper یا حلقه بسته می‌نامند. این روش، جستجو در فضای زیر مجموعه‌ها را براساس تخمین دقت ناشی از انتخاب یک زیر مجموعه خاص تحت شرایط الگوریتم طبقه‌بندی مورد استفاده (به عنوان معیاری از بهینه‌گی آن زیر مجموعه) انجام می‌دهد [۲۲]. الگوریتم‌های دسته دوم معمولاً نتایج بهتری به دست می‌دهند. مهم‌ترین بخش هر روش انتخاب ویژگی حلقه بسته، الگوریتم جستجویی است، که در آن به کار

جدول ۱- پارامترها و ویژگی‌های مورد استفاده در این تحقیق

متغیرها
۱- سطح غلظت قند خون در ابتدای بازه زمانی (mg/d lit)
۲- دوز انسولین ۱۲ ساعته تزریقی به بدن (unit)
۳- سطح فعالیت بدنی (unit)
۴- سن
۵- BMI
۶- HDL
۷- LDL
۸- Total Cholesterol (mg/dl)
۹- Fast Blood Sugar (mg/dl)
۱۰- غلظت انسولین ۶ ساعته تزریقی به بدن (unit)
۱۱- سطح استرس (unit)
۱۲- کربوهیدرات مصرفی (unit)

$$y_i = \frac{x_i - \mu}{\sigma} \quad \text{رابطه ۱:}$$

فرآیند اجرای مراحل در روش پیشنهادی

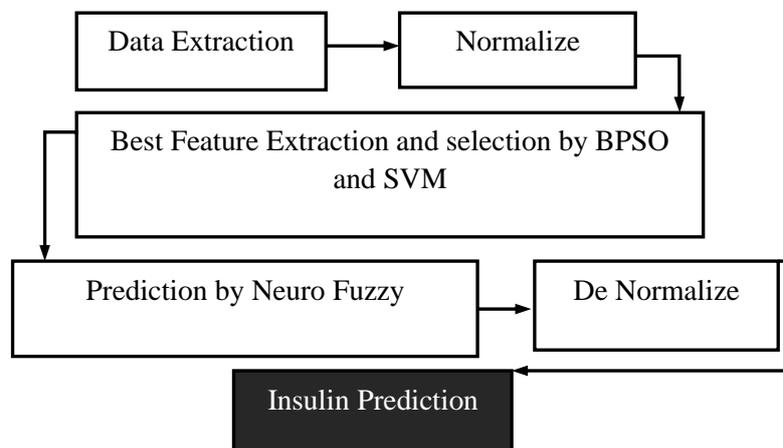
ابتدا بانک اطلاعاتی بر اساس داده‌ها تشکیل می‌شود، سپس توسط رابطه ۱ نرمالیزه می‌شود. اثبات رابطه:

$$z = \frac{x_i - \mu}{\sigma} \rightarrow \text{var}(z) = \text{var}\left(\frac{x_i - \mu}{\sigma}\right) \Rightarrow \text{by refer to } \text{var}(ax + b) = a^2 \text{var}(x) + b \text{ then } \frac{\text{var}(x) - \text{var}(\mu)}{\sigma^2} = \frac{\text{var}(x) - 0}{\sigma^2} = \frac{\sigma^2 - 0}{\sigma^2} = 1$$

$$E(z) = \frac{1}{\sigma}(E(x) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$

صحت نتایج) بهترین ویژگی‌ها انتخاب می‌شوند. مجدداً داده‌ها دنرمالایز می‌شوند؛ سپس بر اساس ویژگی‌های انتخاب شده، میزان صحیح دوز انسولین در بیماران به کمک شبکه فازی عصبی تعیین می‌شود. فلوچارت ۱ بیان کننده روش کار می‌باشد.

به راحتی مشخص است که هر مقداری وارد این رابطه شود به بازه ۰ و ۱ تبدیل خواهد شد. در این رابطه x_i بیانگر مقدار داده‌ها؛ μ نمایشگر میانگین داده‌ها و σ انحراف معیار داده‌ها می‌باشند. در ادامه توسط الگوریتم کاوشی (BPSO) به همراه الگوریتم طبقه‌بندی کننده (SVM) به منظور طبقه‌بندی در جهت برآورد دقت و



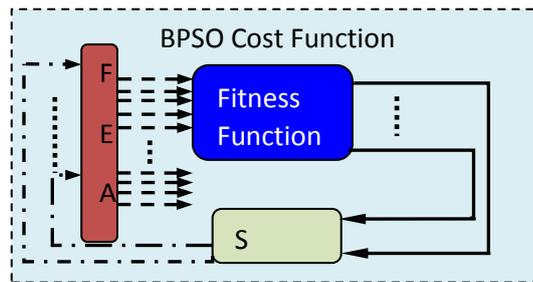
فلوچارت ۱- فلوچارت پیشنهادی به منظور پیش بینی دوز انسولین

جمعیت، مکان هر پرنده با دو مقدار بهترین (p_best , g_best) به روز می‌شود. در هر تکرار با توجه به روش الگوریتم مبتنی بر تکرار قاعده سعی و خطا تا حاصل شدن نتیجه صحیح، الگوریتم بعد از یافتن سرعت و موقعیت جدید پرنده را براساس روابط ۲ و ۳ به روز رسانی می‌کند. روند روش پیشنهادی تابع بهینه کننده به منظور اکتشاف و پیدا کردن بهترین متغیرها، در الگوریتم BPSO در فلوچارت ۲ می‌باشد.

۱-۱- الگوریتم بهینه‌سازی حرکت دسته جمعی ذرات (BPSO) الگوریتم حرکت دسته جمعی ذرات از جمله الگوریتم‌های جستجوی موازی مبتنی بر جمعیت است که توسط Kennedy و Eberhart در سال ۱۹۹۵ مطرح شد [۲۷، ۲۸] که با یک گروه از جواب‌های تصادفی (پرنده‌گان یا ذرات) شروع به کار می‌کند. هر ذره (پرنده) به صورت چند بعدی با دو بردار $X_i[t]$ و $V_i[t]$ که به ترتیب معرف موقعیت و سرعت فعلی در لحظه (t) از ذره i هستند مشخص می‌شود. در هر مرحله از حرکت

$$V_i[t+1] = w \times v_i[t] + C_1 r_1 (X^{i, pbest}[t] + X_i[t]) + C_2 r_2 (X^{gbest}[t] - X_i[t]) \quad \text{رابطه ۲:}$$

$$X_i[t+1] = X_i[t] + V_i[t+1] \quad \text{رابطه ۳:}$$



فلوچارت ۲- تابع هزینه الگوریتم حرکت دسته جمعی ذرات

محدود می‌شود، که عضوی از دامنه $[V_{min}, V_{max}]$ می‌باشد. رابطه ۳ نیز بردار موقعیت فعلی پرنده را با توجه به سرعت جدید آن بروز رسانی می‌کند. تغییر معادله بروز رسانی الگوریتم باینری در ۲ گام انجام می‌شود، الف: استفاده از رابطه ۲ برای تغییر سرعت و ب: اعمال تابع سیگموئید در رابطه ۴ برای محدود کردن سرعت بین $[0, 1]$. رابطه ۴ بازنویسی شد.

$$X_{ij}[t+1] = \begin{cases} 1 & \text{if } rand() < S(V_{ij}[t+1]) \\ 0 & \text{Other Wise} \end{cases}$$

$$S(v_{ij}[t+1]) = \frac{1}{1 + e^{-v_{ij}[t+1]}}$$

یبتی برای هر ذره، الگوریتم BPSO شروع می‌شود مانند بردار ۱.

1	0	0	1	1	0	1	0	1	1	1	0
---	---	---	---	---	---	---	---	---	---	---	---

بردار ۱:

با توجه به نتایج مطلوب [۳۲] تابع برازندگی در رابطه ۵ را پیشنهاد می‌کنیم.

$$Fitness = \alpha \cdot Accuracy + \beta \cdot \frac{|n| - |s|}{|n|}$$

می‌کنیم. در این تحقیق دقت شناسایی مهمتر است، پس α بسیار بزرگتر از β خواهد بود. Accuracy یا میزان صحت، از مقایسه بین بردار بیماران شناسایی شده توسط سیستم و بیماران واقعی به دست می‌آید. پارامترهای BPSO مطابق جدول ۲ تنظیم شده‌اند. در رابطه ۵ صحت انجام کار توسط رابطه ۶ معرفی می‌شود.

در رابطه ۲، W ضریب اینرسی است (با در نظر گرفتن مقادیر معرفی شده در مطالعه Feng و همکاران [۲۹])؛ که در بازه $[0, 1.25]$ قرار دارد. $C1$ و $C2$ میزان تجربه شخصی و گروهی است که به طور معمول برای هر دوی آن‌ها از مقدار $1/5$ و یا $1/2$ استفاده شده است [۳۰، ۳۱]. $r1$ و $r2$ نیز اعداد تصادفی در بازه $[0, 1]$ می‌باشند. برای جلوگیری از واگرایی الگوریتم، سرعت پرنده به یک بازه رابطه ۴:

در اینجا به منظور بهینه‌سازی فضای جواب‌های مساله از مدل الگوریتم باینری حرکت دسته جمعی ذرات (BPSO) استفاده می‌کنیم. با در نظر گرفتن یک رشته دودویی ۱۲

که در بردار ۱ "0" نمایانگر حذف ویژگی و "1" نمایانگر انتخاب آن ویژگی می‌باشد. البته در اینجا دقت شناسایی مهم‌تر از کوچک بودن زیر مجموعه منتخب است، بنابراین رابطه ۵:

در این رابطه $|n|$ کل ویژگی‌ها و $|s|$ ویژگی‌های منتخب است. جمله اول ضریبی از دقت شناسایی و جمله دوم ضریبی از نرخ کاهش ویژگی‌هاست. α و β ضرایب ثابت پیشنهادی در رابطه ۵ هستند. به خاطر ارجحیت دادن به دقت شناسایی نسبت انتخاب ویژگی مجموع ضرایب α و β را ثابت و برابر ۱۰۰ پیشنهاد

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Fn + Tn + Fp} \quad \text{رابطه ۶:}$$

تعداد افراد سالم است که مدل تصمیم به اشتباه آن‌ها را مبتلا به دیابت تشخیص داده است. F_n یا منفی نادرست نیز بیانگر تعداد افراد بیمار است که مدل تصمیم به اشتباه آن‌ها را سالم تشخیص داده است.

در رابطه ۶، Tp یا مثبت درست، بیانگر تعداد افرادی است که مبتلا به دیابت هستند و مدل تصمیم به درستی آن را تشخیص داده است، Tn یا منفی درست، تعداد افراد سالم است که مدل تصمیم به درستی آن را سالم تشخیص داده است. Fp یا مثبت نادرست نیز بیانگر

جدول ۲- پارامترها و مشخصات الگوریتم بهینه سازی (کاوشی) BPSO

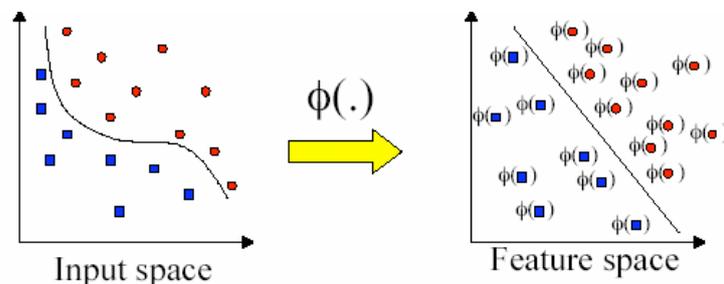
پارامترها	مشخصات
الگوریتم بهینه سازی	الگوریتم BPSO
جمعیت اولیه	۲۰ ذره
Maxiter	۱۰۰ تلاش
C1 یا ضریب تجربه شخصی	۱ [۱۹، ۲۵]
C2 یا ضریب تجربه گروهی	۲۰ [۱۹، ۲۵]
W یا ضریب اینرسی	بین ۰ تا ۱ به صورت نزولی [۲۵]
Vmax	$[۱۹، ۲۵] + 4V$
Vmin	$[۱۹، ۲۵] - 4V$

BPSO

با چند زیر مجموعه تصادفی شروع به کار می‌کند و الزام دارد تعدادی زیر مجموعه را که حاوی مفیدترین اطلاعات هستند، در نهایت معرفی کند. بردارهای ویژگی معرفی شده توسط BPSO برای آموزش طبقه‌بندی کننده SVM به کار برده می‌شوند.

الگوریتم طبقه بندی SVM

SVM یک الگوریتم یادگیری با سرپرستی است که می‌تواند در کاربردهایی نظیر جداسازی و طبقه‌بندی داده مورد استفاده قرار گیرد [۳]. این الگوریتم می‌تواند داده‌های آموزش را به دو دسته تقسیم کند، اصولاً داده‌ها توسط یک خط قابل جداسازی هستند، و در غیر این صورت الگوریتم با نگاشت ویژگی‌ها به وسیله تابع ϕ به ابعاد بالاتر سعی می‌کند توسط یک فرا صفحه با دقت بالاتری داده‌ها را تفکیک نماید (شکل ۱).



شکل ۱- نگاشت داده‌ها از فضای ورودی به فضای ویژگی [۳]

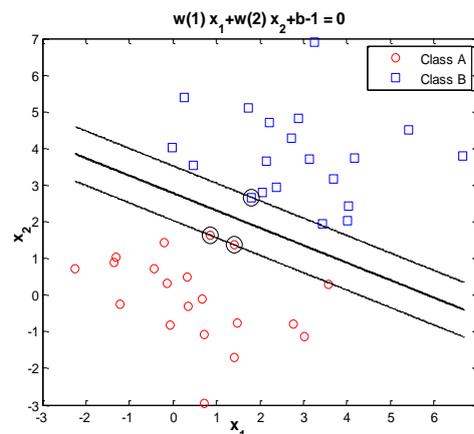
که در این روابط x متغیر ورودی، w بردار نرمال خط جدا کننده و b عرض از مبدا خط جدا کننده است. در هر تلاش طبقه‌بندی نمونه‌های خروجی از الگوریتم BPSO، توسط الگوریتم ماشین پشتیبان SVM انجام می‌شود تا در نهایت کمترین و بهترین ویژگی‌هایی که منجر به ایجاد بهترین کلاس‌بندی در داده‌ها می‌شود، تعیین شوند. به‌منظور کاهش محاسبات در کلاس‌بندی، از تابع خطی استفاده شده است. در اثر اعمال این الگوریتم در ویژگی سوم، شکل ۲ به‌دست می‌آید.

رابطه ۷ رابطه فرا صفحه طبقه‌بندی کننده داده‌ها است. روابط ۸ و ۹ روابط فرا صفحه‌های موازی براساس شرایط حداکثر حاشیه می‌باشند. در صورت ترسیم این توابع ساده فاصله بین دو صفحه حاشیه برابر $\frac{2}{\|w\|}$ می‌باشد [۳].

$$\text{رابطه ۷: } w \times x - b = 0$$

$$\text{رابطه ۸: } w \times x - b = -1$$

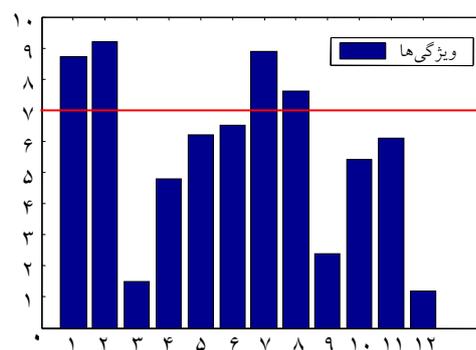
$$\text{رابطه ۹: } w \times x - b = 1$$



شکل ۲- طبقه‌بندی داده‌ها به ۲ گروه در ویژگی سوم

پشتیبان (SVM)، ویژگی‌های شماره ۸، ۷، ۲، ۱، (ویژگی‌های بهینه و مفید) که در ۱۰۰ مرتبه اجرای سیستم پیشنهادی به‌دست آمدند در شکل ۳ مشخص شدند.

پس از اجرای مراحل مختلف الگوریتم پیشنهادی به‌منظور مشخص شدن بهترین و موثرترین ویژگی‌ها، یعنی اعمال الگوریتم‌های بهینه‌سازی (کاوشی) و الگوریتم بردار ماشین

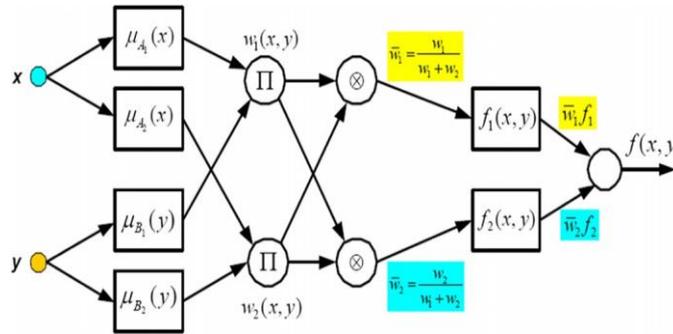


شکل ۳- مشخص شدن بهترین و مفیدترین متغیرها از سایر متغیرها در فرآیند کاوش به‌منظور مشخص شدن بهترین متغیرها

می‌شود، سپس باید به مرحله بعد، یعنی شبکه فازی - عصبی به‌منظور پیش‌بینی صحیح میزان دوز انسولین براساس الگوی

بعد از اینکه متغیرهای مهم مشخص شدند، بانک اطلاعاتی با ابعاد ۳۱۲ سطر (افراد مورد مطالعه) و ۴ ستون (ویژگی) مرتب

استفاده در شکل ۴ دیده می‌شود که دایره‌ها بیان کننده گره‌های ثابت و مربع‌ها بیان کننده گره‌های تطبیقی هستند. در لایه اول، تمامی گره‌ها تطبیقی هستند و خروجی این لایه، درجه عضویت فازی ورودی‌هاست که با روابط ۱۰ و ۱۱ به دست می‌آید [۳۳].



شکل ۴- سیستم انطباقی فازی عصبی

فازی سوگنو با دو ورودی x و y یک خروجی f داشته باشد، برای یک مدل فازی سوگنو مرتبه اول، مجموعه قوانین مشترک با دو قانون If-then به صورت زیر است. در رابطه ۱۰ و ۱۱، X داده‌های ورودی (متغیر یا ویژگی‌های بهینه) برای هر قانون و Y داده خروجی (میزان صحیح دوز انسولین) برای هر قانون می‌باشد.

در این سیستم‌ها با استفاده از مجموعه داده‌های ورودی (۴) متغیر بهینه مشخص شده برای هر نمونه- خروجی (دوز صحیح انسولین برای هر نمونه) در مدل پیش‌بینی کننده دوز انسولین به عنوان داده‌های آموزشی، یک سیستم استنباط فازی از نوع سوگنو ایجاد می‌شود که پارامترهای تابع عضویت آن با استفاده از الگوریتم پس انتشار و روش کوچک‌ترین مربعات تنظیم می‌گردد [۳۴]. اگر یک سیستم

$$\text{رابطه ۱۰: Rule 1: If } X \text{ is } A_1 \text{ and } Y \text{ is } B_1 \text{ then } f_1 = \mu_1 \times x + q_1 y + r_1$$

$$\text{رابطه ۱۱: Rule 2: If } X \text{ is } A_2 \text{ and } Y \text{ is } B_2 \text{ then } f_2 = \mu_2 \times x + q_2 y + r_2$$

ساختار الگوریتم ANFIS استفاده شده در این تحقیق از Tahami و همکاران اقتباس شده است [۳۴] (جدول ۳).

جدول ۳- پارامترهای ساختار الگوریتم ANFIS

روش تولید سیستم فازی	الگوریتم میانگین فازی
الگوریتم آموزش	پس انتشار خطا
تعداد خوشه‌های الگوریتم میانگین فازی	۱۰
مشخصات آموزش در سیستم پیشنهادی	
۱۰۰ مرتبه	حد نهایی تکرار
۰/۹	میزان نرخ کاهشی
۱/۱	میزان نرخ افزایشی
۰	خطای هدف
۰/۰۱	میزان نرخ انتسابی اولیه

نتایج ارزیابی و آزمایش‌های نهایی

ویژگی‌های اصلی، سرنوشت‌ساز و تاثیر گذار در تعیین میزان دوز انسولین در بیماران تعیین شدند، سپس شبیه‌سازی نهایی توسط شبکه عصبی فازی-مصنوعی (ANFIS) مطابق شکل ۵ برای ۴۸ ساعت آینده بیمار به صورت هر ۶ ساعت تهیه شده است. البته مقادیر پس از اینکه توسط شبکه فازی عصبی پیش‌بینی شدند، توسط رابطه ۱۲ دنرمالیزه می‌شوند. پس از دنرمالیزه شدن، نتایج شبیه‌سازی در شکل ۵ برای بیمار تصادفی شماره ۱۵ مشخص شده است. در شکل ۵ مقدار واقعی انسولین تجویز شده برای فرد در طول ۴۸ ساعت آینده توسط بارهای فیروزه‌ای مشخص شده است، این در حالی است که مقدار پیش‌بینی شده توسط سیستم پیشنهادی توسط بارهای آبی پررنگ نشان داده شده است. در ابتدای هر بار با فرض ۰/۵ درصد، خط خطا (error bar) با نقاطی نمایش داده شده که اگر این نقاط را در بارهای فیروزه‌ای به یکدیگر متصل نماییم، خط ممتد قرمز و اگر نقاط خط خطا را در بارهای آبی پررنگ به یکدیگر متصل نماییم، خط مشکی غیر ممتد مشخص می‌شود. مقیاس مقادیر در این نمودار در حالت یکسانی (per unit) شده که در نهایت توسط رابطه ۱۲ به حالت واقعی (دنرمالیزه) بازگشت داده می‌شود.

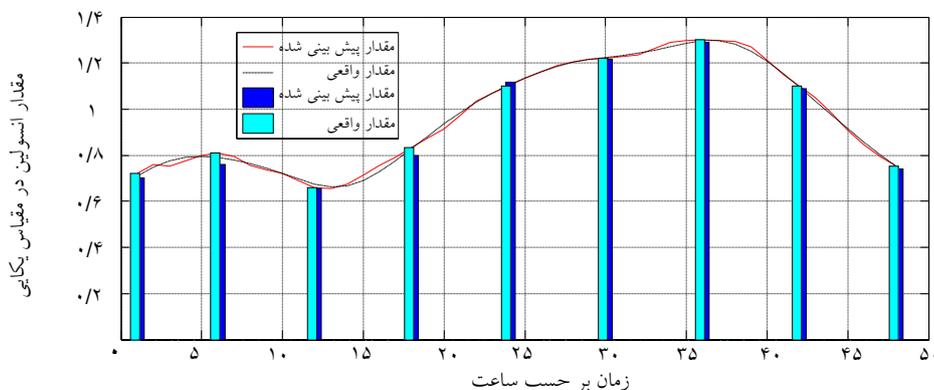
رابطه ۱۲:

$$x_i = y_i \times \sigma^2_i + \mu$$

با توجه به سیستم پیشنهادی و جدول ۳، روش تولید سیستم فازی توسط روش فازی میانگین فازی (FCM) با تعداد ۱۰ کلاستر (خوشه)، استفاده شده است. به طوری که الگوریتم آموزش شبکه انطباقی فازی - عصبی، از نوع پس انتشار خطا، با ۱۰۰ تکرار، نرخ کاهش ۰/۹، نرخ افزایشی ۱/۱ با مقدار ۰/۰۱ در هر تکرار و خطای نهایی در حد صفر تعیین شده است.

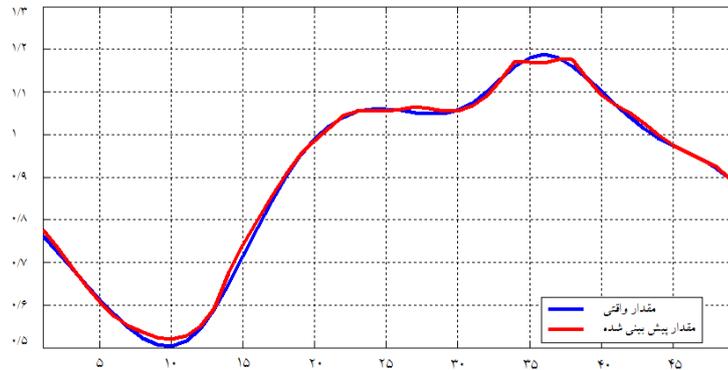
بررسی صحت آزمایش و ارزیابی نهایی

فرآیند انتخاب ویژگی که منجر به پیش‌بینی مقدار صحیح دوز انسولین می‌شود به این صورت است که توسط الگوریتم باینری حرکت دسته جمعی ذرات براساس تابع هزینه تعریف شده (کاوش به منظور بررسی ویژگی‌ها) و به کمک الگوریتم طبقه‌بندی ماشین بردار پشتیبان (SVM) موثرترین ویژگی‌ها مشخص می‌شوند، سپس شبکه فازی عصبی و بانک داده مبتنی بر ویژگی‌های موثر، پیش‌بینی صحیحی از میزان دوز انسولین برای بیماران تهیه می‌کند. برای اثبات قابلیت‌های سیستم پیشنهادی، اعتبار سنجی عرضی ۱۰ دسته‌ای (توسط K-fold) بر روی پایگاه داده ۳۱۲ نمونه‌ای موجود پیاده شد که به ۱۰ دسته تقسیم شدند. ۸ دسته ۳۱ داده‌ای و ۲ دسته ۳۲ داده‌ای. در این اعتبار سنجی، ۹ قسمت از داده‌ها برای آموزش و ۱ قسمت برای آزمایش به کار می‌روند، آموزش و آزمایش ۱۰ بار به صورت متوالی چرخشی انجام می‌شود به طوری که هر بار یک قسمت برای آزمایش کنار گذاشته می‌شود.



شکل ۵- نتایج پیش‌بینی دوز انسولین و میزان واقعی برای بیمار تصادفی شماره ۱۵

البته برای نمایش بهتر میزان تهیه شده و یا معرفی شده، از شکل ۶ استفاده می‌نماییم، که یک نمای ۳ بعدی از شبیه‌سازی و تعیین دوز دارو برای ۴۸ ساعت آینده بیمار می‌باشد.



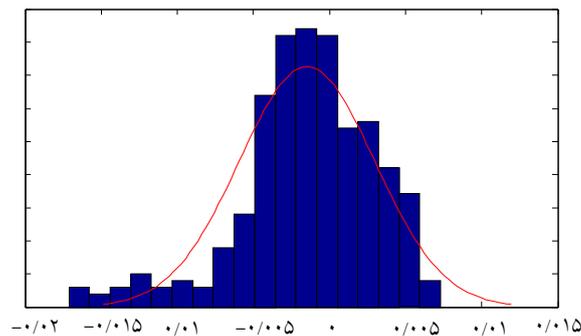
شکل ۶- پیش بینی دوز صحیح داروی انسولین برای ۴۸ ساعت آینده بیمار تصادفی شماره ۱۵

پیش‌بینی صحیح دوز انسولین بیماران، خطای مقدار صفر را در حدود ۸۰ بار تجربه کرده است، البته ناگفته نماند که این خطا به ازای هر بیمار تعریف شده است. به‌منظور بررسی و مقایسه عملکرد سیستم پیشنهادی، پارامترهایی طبق روابط ۱۴ و ۱۵ و ۱۶ تعریف می‌شوند. البته نتایج این پارامترها در جدول ۴ مشخص شده‌اند.

رابطه ۱۳:

$$\bar{e} = \frac{\sum_{i=1}^n \|y_i - \hat{y}_i\|^2}{n}$$

به‌منظور بررسی خطای عملکرد شبکه فازی عصبی، رابطه ۱۳ تعریف می‌شود که نمودار آن در شکل ۷ رسم شده است. اساس کار شبکه به‌منظور انطباق فضای ورودی به خروجی، بر طبق انطباق با تکیه بر فرآیند سعی و خطا است، خطایی است که توسط رابطه ۱۳ محاسبه می‌شود، که بیان‌کننده میانگین مربعات ناشی از اختلاف میانگین مربعات مقادیر خروجی (y_i) با مقادیر واقعی (\hat{y}_i) است. در شکل ۷ به‌طور مثال مشخص می‌شود که این شبکه برای



شکل ۷- خطای عملکرد شبکه فازی عصبی

$$\text{Speciality} = \frac{Tn}{Tn + Fp} \quad \text{رابطه ۱۴:}$$

$$\text{Sensitivity} = \frac{Tp}{Tp + Fn} \quad \text{رابطه ۱۵:}$$

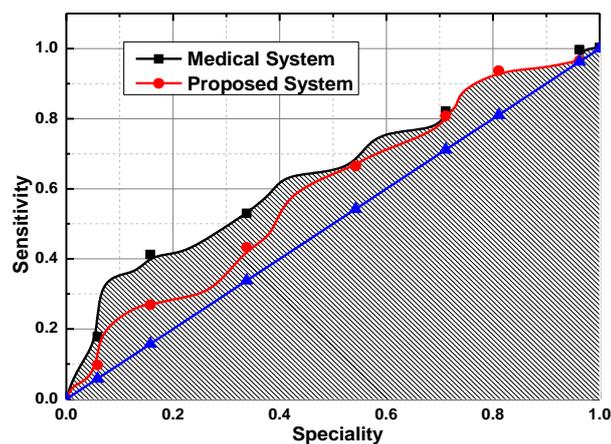
$$\text{Total Accuracy} = \frac{\text{Total True}}{\text{Total input}} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad \text{رابطه ۱۶:}$$

جدول ۴- نتایج عملکرد سیستم‌فازی عصبی در پارامترهای ۱۴، ۱۵، ۱۶

پارامتر	درصد
Sensitivity (%)	۹۱/۲
Specialty (%)	۸۴/۱
Total Accuracy (%)	۹۲/۹

می‌باشد که بیان کننده احتمال مثبت یک تست است که به بیماران مبتلا نسبت داده شده است. ویژگی در یک آزمون، بیان کننده احتمال منفی یک تست می‌باشد که به بیماران نسبت داده شده است [۳۵]. شکل ۸ نمودار ROC هر ۲ سیستم برای ۳۰ مورد بیمار تصادفی آزمایشی است که، میزان حساسیت و ویژگی آن‌ها براساس رفتار سیستم پیشنهادی و عملکرد پزشک مشخص شده است.

منحنی ROC6 مشخصه‌ای دیگری برای طبقه‌بندی صحیح و بیان عملکرد بین سیستم‌ها در شرایط کاری یکسان است. همان‌طور که در شکل ۸ می‌بینید، مشاهده حساسیت و ویژگی در هر سیستم بر حسب دیگری بیان شده است. در موارد و شرایط ایده‌آل کاری، با افزایش حساسیت، کاهش کمی در ویژگی وجود دارد، تا زمانی که به سطح بسیار بالایی از حساسیت رسیده باشیم. حساسیت، یک آزمون



شکل ۸- منحنی ROC به منظور مقایسه سیستم واقعی و پیشنهادی در داده‌های آزمایشی

پزشکان به‌طور قابل توجهی نزدیک شد (منظور بیماران مطالعه پیش رو است که برای آن‌ها دارو تجویز شده است). البته به دلیل عدم دسترسی به منابع تحقیقات قبلی که در زمینه پیش‌بینی و تعیین صحیح دوز انسولین انجام گردیده است، مقایسه قابل ملاحظه‌ای در این زمینه نمی‌توان انجام داد، اما در نهایت هدف تطبیق و یکسان بودن نتایج سیستم‌های پیشنهادی با تجویز پزشکی در هر حالت مهم است، که خوشبختانه سیستم پیشنهادی با توجه به شکل‌های (۶ و ۸) و جدول ۴ عملکرد مناسب و قابل قبولی داشته است. همچنین مقایسه نتایج به‌دست آمده در این تحقیق با نتایج به‌دست آمده از تصمیم‌گیری‌های

البته، هر چه قدر مساحت زیر نمودار سیستم پیشنهادی به سیستم واقعی نزدیک‌تر باشد نشان دهنده عملکرد نزدیک سیستم پیشنهادی به سیستم واقعی یا تصمیمات پزشکان است.

نتیجه‌گیری

در این تحقیق با توجه به ضرورت تعیین صحیح و به‌موقع میزان انسولین برای بیماران، روشی نوین مبتنی بر ترکیب سیستم‌های هوشمند ارائه شده است، به‌طوری‌که نتایج موجود به‌دست آمده به نتایج واقعی یا همان تجویزات

افزایش کارایی این شبکه‌ها در جهت پیش‌بینی صحیح باشند. البته هدف اصلی در این نوع تحقیقات به دست آوردن صحت و دقت ۱۰۰ درصدی می‌باشد، تا بتوان به سیستم‌های هوشمند مصنوعی اطمینان داشت و بعد عملی به آن‌ها بخشید. میل و رسیدن به چنین هدفی تنها در صورت دانش بالا، فهم درست و تسلط کافی از یک طرف به سیستم‌های هوشمند، و از طرفی دیگر به بیماری و داده‌کاوی امکان‌پذیر است که در آینده‌ایی نه چندان دور محقق خواهد شد.

پزشکان، نشان دهنده عملکرد قابل ملاحظه در میزان صحت پیش‌بینی سری زمانی سطح غلظت قند خون است، چرا که سیستم پیشنهادی موفق شد با در دست داشتن چند ویژگی از بیماران مبتلا به دیابت، سطح قند خون تا ۴۸ ساعت آینده بیمار را پیش‌بینی نماید. از شبکه‌های مصنوعی تطبیقی فازی عصبی به منظور انطباق، یادگیری و دسته‌بندی استفاده شد. چالش آموزش شبکه‌های عصبی مهم‌ترین دغدغه می‌باشد. ترکیب الگوریتم‌های تکاملی و سیستم‌های داده کاوی می‌تواند یک ایده مناسب برای

مآخذ

- World Health Organization, Diabetes Center, Fact Sheet N312, Retrived from: www.who.int/mediacentre/factsheets/fs312/en
- Centers for Disease Control and Prevention, National Diabetes, Fact Sheet 2011 Retrived from: www.cdc.gov/diabetes
- نعیم آبادی م، امیراحمدی چماچار ن، تهامی ا و ربانی ح. تشخیص دیابت با استفاده از SVM و MLP، چکیده نامه چهاردهمین کنفرانس دانشجویی مهندسی برق ایران؛ کرمانشاه؛ ایران؛ کرمانشاه، دانشگاه صنعتی کرمانشاه؛ ص ۲۵۱-۲۵۸.
- American Diabetes Association, Diabetes Basics. Retrived from: www.diabetes.org/diabetes-basics,
- فیوضی م، قره خانی ا، حدادنیا ج. تشخیص بیماری دیابت با استفاده از سیستم‌های فازی، درخت تصمیم‌گیری و شبکه‌های عصبی، چکیده نامه چهارمین کنفرانس ملی فناوری اطلاعات و دانش ایران؛ بابل؛ ایران، دانشگاه صنعتی کرمانشاه، ص ۱۲۵-۱۳۱.
- فیوضی م، حدادنیا ج، ملانیا ن. تشخیص بیماری دیابت براساس روش ترکیبی هوشمند. مجله علمی و ترویجی هوش مصنوعی و ابزار دقیق ایران؛ ۱۳۹۱؛ ۱ (۳۲): ۸-۱.
- قوچانی س، تهامی، س. پیش‌بینی نوسانات سطح قند خون در بیماران مبتلا به دیابت نوع ۱ با استفاده از شبکه‌های عصبی خود بازگشتی الممن. مجله علمی و پژوهشی فنی و مهندسی دانشگاه آزاد اسلامی مشهد؛ ۱۳۸۷؛ ۲ (۱): ۹۴-۸۱.
- Li J, Kuang Y, Mason CC. Modeling the glucose-insulin regulatory system and ultradian insulin secretory oscillations with two explicit time delays. *Journal of Theoretical Biology* 2006; 242:722-735.
- Leaning MS and Boroujerdi MA. A system for compartmental modelling and simulation. *Computer Methods programs biomed* 1991; 35 (2):148-157.
- Tresp V, Moody J, and DeLong WZ. Neural modelling of physiological processes. *Computational Learning Theory and Natural Learning Systems* 1994; 2(1): 250-258.
- Huang HP, Liu SW, Chien IL, Lin CH. A Dynamic Model with Structured Recurrent Neural Network to Predict Glucose-Insulin Regulation of Type 1 Diabetes Mellitus. Proceeding of 9th International Symposium on Dynamics and Control of Process Systems (DYCOPS 2010), Leuven, Belgium, pp 25-31, July 5-7, 2010.
- Lean Yu. An evolutionary programming based asymmetric weighted least squares support vector machine ensemble learning methodology for software repository mining. *Information Sciences* 2012; 191(12): 31-46.
- NASA MDP repository. Retrived from: <http://mdp.ivv.nasa.gov>,
- Fernandez de Canete J, Gonzalez-Perez S, Ramos-Diaz JC. Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes. *Computer methods and Programs in biomedicine Journal* 2008; 4(2): 128-135.
- Scott M, Pappada BS, Brent D, Rosman PM. Development of a Neural Network for Prediction of Glucose Concentration in Type 1 Diabetes Patients. *Journal of Diabetes Science and Technology* 2008; 2(5): 189-197.
- Mougiakakou SG, Prountzou A, Iliopoulou D, Nikita KS, Andriani V, Bartsocas CS. Neural Network based Glucose - Insulin Metabolism Models for Children with Type 1 Diabete. *Engineering in Medicine and Biology Society* 2006; 1: 3545-8. 28th Annual International Conference of thr IEEE.

17. Wahab A, Kong YK, and Quek C. *Model Reference Adaptive Control on Glucose for the Treatment of Diabetes Mellitus*. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems CBMS'06, pp 250-258, 2006.
18. Jensen R. *Combining rough and fuzzy sets for feature selection*. Ph.D. Thesis. School of informatics, Univ. Edinburgh, 2005.
۱۹. علیپور م، حدادنیا ج. معرفی یک روش دقیق برای تشخیص سرطان پستان. *مجله بیماری‌های سرطان پستان* ۱۳۸۹؛ دوره ۲، شماره ۲، ص ۱۵-۲۲.
20. Xiang T and Gong S. Spectral clustering with eigenvector selection. *Pattern Recognition* 2008. 41(3): 1012-1029.
21. Toosi SAZ, Sadoghi Yazdi H. *Spectrum Clustering by K-PCA*. Proceeding of 16th CSI Conference, pp 118- 124, Winter 2010.
22. Kanan H, Faez K, Hosseinzadeh M. *Face recognition system using ant colony optimization-based selected features*. Proceeding of IEEE Symp on Computational Intelligence in Security and Defense Applications pp 57-62, 2007.
23. Siedlecki W, Sklansky J. A note on genetic algorithms for large-scale feature selection. *Journal of pattern Recognition Letters* 1989; 10 (5): 335-47.
24. Liu B, Abbess H, McKay B. Classification rule discovery with ant colony optimization. *Journal of IEEE Computational Intelligence Bulletin* 2004; 3(1): 31-5.
25. Meng Y. A swarm intelligence based algorithm for proteomic pattern detection of ovarian cancer. *Computational Intelligence and Bioinformatics and Computational Biology* 2006, 24 (8): 1-7.
26. Newman DJ, Hettich S, Blake CLS, & Merz CJ. 1998 UCI. *Repository of machine learning database*, Irvine, CA: University of California, Dept. of Information and Computer Science. Retrived from: archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes
27. Pan H, Ma Q, Wei X. Research on Fault Diagnosis of Gearbox Based on Particle Swarm Optimization Algorithm. *IEEE 3rd International Conference on Mechatronics* 2006; (ICM' 3rd), pp 254-260.
28. Sheybani M, meibody M. *PSO-LA: A Novel of Optimization*. Proceeding of 12th Conference on Computer Engineering, Iran, pp 1162 – 1168, 2006.
29. Feng X, Zhang J, and Yang Z. *Adaptive Particle Swarm Optimization on Individual Level*. Proceeding of IEEE conference, China, pp. 1215- 1218, 2002.
30. Norozie Beyrami M. *Improve the convergence of the algorithm of particle swarm*. Proceeding of Conference on industrial engineering, Osco IAU, pp 205-211, 2008.
31. Kennedy J, and Eberhart R. *Particle swarm optimization*. Proceeding of Conference on Evolutionary Computation IEEE, USA, pp. 1942-1948, 1995.
32. Tahami SE, Khalil Zadeh MA. *Intelligent Diagnosis by Multi-Layer Neural Network*. Proceeding of 12th Iranian Conference of Biomedical Engineering, Tabriz, Iran, pp 89-95, 2005.
33. Kemal Polat, Salih Güne. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing* 2007; p 702–710.
34. Tahami SE, Bamashki SM, Khalil Zadeh MA. *Diabet Diagnosis by GA-NN, ANFIS*. Proceeding of 1th Join Conference of Fuzzy and Intelligent System, summer 2007, Mashhad, Iran, pp 115-121.
35. Rajeswari K, Vaithiyanathan V. Fuzzy based modeling for diabetic diagnostic decision support using Artificial Neural Network. *International Journal of Computer Science and Network Security* 2011; 11(4): p 56-64.

THE PREDICTION OF CORRECT DOSE OF INSULIN IN MELLITUS PATIENT BY USING ARTIFICIAL INTELLIGENT SYSTEM AND COMBINED DATA MINING ALGORITHM

Mohammad Fiuzy^{1*}, Javad Haddadnia², Nasrin Mollania³, Mohammad Mohammad zedeh¹

1. Healthy ageing Research Center, University of Medical Sciences, Sabzevar, Iran

2. Research center for Advanced medical Technologies, Electrical Engineering Faculty, Hakim Sabzevari University, Sabzevar, Iran

3. Biochemistry Section, Department of Biology, Faculty of Basical Science, Hakim Sabzevari University, Sabzevar, Iran

ABSTRACT

Background: Diabetes is such diseases that need high quality beside prevention such as correctly predict fluctuations in blood glucose levels. The main complications of the disease can be anesthesia, coma and even death. Today, in these patients, the correct dose of insulin determined based on experience or doctors knowledge, and interact between the patients and physician, although there is an inevitable human errors.

Methods: In this study based on applied method, 124 patients and 188 healthy subjects based on 12 features by Random Selection, Who had been referred to Research Center for Diabetic in Sabzevar university of Medical Science since 2006 to 2011 were studied. The proposed system has several subsystems, such as evolutionary algorithms (BPS 1) to select the most effective features, Data Mining Algorithms (SVM 2) to detect and classify the features from the non-effective features. Adaptive Neuro fuzzy systems (ANFIS 3) to estimate learn and adaptation in order to correctly predict have been used.

Results: In this study, we try to use artificial intelligence systems to determine the correct dose of insulin for diabetics. The proposed system combines the best attributes in the database in the form the interaction was able to achieve high accuracy with the lowest error. The proposed system based on best features in the database in the interaction form was able to achieve high accuracy with the lowest error. The proposed system in the form of composition and interaction with the subsystem was able to achieve carefully 84.1% in specificity, 91% in sensitivity and 92.9% in accuracy.

Conclusion: In this research, due to the importance of correct and timely determination of insulin for diabetics, a new method based on the combination of intelligent systems is presented. Thus, the results obtained in previous articles and studies provide significantly improved.

Keywords: Artificial Intelligent, Insulin, Pattern Recognition, Prediction

* Department of Control, School of Electrical Engineering, Iran University of Science and Technology, P.O BOX : 1684613114, E-mail: mohammad.fiuzy@yahoo.com