



## ارائه مدلی جهت پیش بینی بیماری دیابت با استفاده از شبکه عصبی

میثم جهانی<sup>۱</sup>/ جلال رضایی نور<sup>۲</sup>/ مهدی مهدوی<sup>۳</sup>/ اسماعیل هداوندی<sup>۴</sup>

چکیده

مقدمه: الگوریتم های فرا ابتکاری و ترکیبی از توانمندی بالایی در مدل سازی مسائل پزشکی برخوردارند. در این مطالعه از شبکه عصبی به منظور پیش بینی ابتلا به دیابت در میان افراد مستعد دیابت استفاده گردید. روش کار: پژوهش حاضر از نوع کاربردی و جامعه ی هدف آن متشکل از ۵۴۵ فرد بیمار و سالم از مرکز دیابت دانشگاه علوم پزشکی همدان جمع آوری گردید جهت پیش بینی بیماری دیابت استفاده شده است. در این مطالعه از الگوریتم ممیتیک که تلفیقی است از الگوریتم ژنتیک و الگوریتم جستجوی محلی است، به منظور به روز رسانی وزن های شبکه عصبی و توسعه دقت شبکه عصبی استفاده شده است.

یافته ها: بررسی اولیه نشان داد که دقت شبکه عصبی، ۸۸ درصد، می باشد. بعد از بروز رسانی وزن ها با الگوریتم ممیتیک دقت آن به ۹۳/۲ درصد افزایش یافت. برای مدل پیشنهادی به ترتیب حساسیت، ویژگی، ارزش اخباری مثبت، ارزش اخباری منفی، مساحت زیر منحنی ۹۶/۲، ۹۲/۴، ۹۳/۸، ۹۵/۳، ۰/۹۵۸ برای مدل الگوریتم ژنتیک، ۹۸، ۸۴/۸، ۸۸/۶، ۸۸/۲، ۰/۹۵۲ و برای مدل رگرسیون لجستیک، ۹۵/۶، ۸۴/۵، ۹۴/۷، ۸۷/۰، ۰/۹۱۶ به دست آمد.

نتیجه گیری: بر اساس یافته های این پژوهش، مدل های شبکه های عصبی در مقایسه با مدل رگرسیون از میزان خطای کمتری در تشخیص بیماری بر اساس متغیرهای فردی و سبک زندگی برخوردارند. یافته های این مطالعه می تواند به برنامه ریزان و ارائه کنندگان خدمات سلامت در برنامه های غربالگری و تشخیص به موقع بیماری دیابت کمک می نماید.

کلید واژه ها: دیابت، تکنیک پشتیبان تصمیم گیری، شبکه عصبی، الگوریتم ژنتیک، الگوریتم ممیتیک

• وصول مقاله: ۹۵/۰۳/۲۹ • اصلاح نهایی: ۹۵/۰۶/۱۳ • پذیرش نهایی: ۹۵/۰۸/۲۹

۱. کارشناس ارشد مهندسی فناوری اطلاعات، دانشکده فنی و مهندسی دانشگاه قم، قم، ایران

۲. دانشیار گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران؛ نویسنده مسئول (j.rezaee@qom.ac.ir)

۳. دکتری مدیریت خدمات بهداشتی و درمانی، موسسه مدیریت و سیاستگذاری سلامت، دانشگاه ایراسموس روتردام، روتردام، هلند

۴. استادیار گروه مهندسی صنایع، دانشکده مهندسی صنایع و کامپیوتر، دانشگاه صنعتی بیرجند، بیرجند، ایران



ضعف نظام‌های مراقبتی در شناسایی بیماری در مراحل اولیه بیماری است. تاخیر در تشخیص و پیش‌بینی دیابت و در نتیجه عدم کنترل کافی قند خون احتمال عوارض قلبی و عروقی، نارسایی کلیوی، عوارض چشمی، و درگیری اندام‌های تحتانی را به میزان قابل توجهی افزایش می‌دهد [۷-۹].

اهمیت تشخیص به موقع بیماری‌ها با استفاده از مدل‌های پیش‌بینی از دهه‌های گذشته توسط ارائه‌کنندگان خدماتی درمانی و محققین درک شده است. جهت مدیریت بهتر سلامتی و کاهش مصرف خدمات سلامت نیاز است که میزان خطر ابتلا به بیماری یا پیشرفت بیماری در میان کلیه افراد یک جمعیت مشخص، شناسایی شود. با تعیین و خوشه‌بندی میزان مخاطرات، خدمات مناسب برای هر یک از خوشه‌ها طراحی می‌شود. مدل‌های پیش‌گویی بر اساس فاکتورهای جمعیت‌شناختی، وضعیت سلامتی و الگوی مصرف خدمات سلامت به تعیین میزان ریسک افراد و احتمال استفاده از انواع خدمات می‌پردازند. این ابزارها در بهبود پیامدهای بیماران و همچنین کاهش هزینه‌های مدیریت بیماری‌ها در شبکه‌های سلامت استفاده شده‌اند [۱۰-۱۲].

در مطالعه حاضر برای پیش‌بینی بیماری دیابت از تکنیک‌های الگوریتم‌های فراابتکاری استفاده شده است. الگوریتم‌های فراابتکاری به طور فزاینده‌ای جهت توسعه مدل‌های تشخیص بیماری‌ها استفاده شده است. این الگوریتم‌ها ترکیبات بهینه را در یک زمان معقول به دست می‌آورد. همچنین الگوریتم‌های فراابتکاری جهت یافتن راه‌حلی برای مسائل با مشکلات محاسباتی ارائه شده است. الگوریتم‌های فراابتکاری از قبیل شبکه عصبی، الگوریتم ژنتیک، الگوریتم ممتیک از توانمندی بالایی در مدل‌سازی مسائل پیچیده برخوردارند. این الگوریتم‌ها جهت تعیین میزان خطر ابتلا به بیماری‌ها، پیش‌بینی ابتلا و تشخیص بیماری در مراحل پیش‌بالینی مورد استفاده قرار گرفته‌اند [۱۳، ۱۴].

ونگ و همکاران در یک مطالعه موردی روی ۸۶۴۰ بیمار برای شناسایی بیماران دیابتی نوع دو و آزمایش کردن شبکه مصنوعی و لجستیک چند متغیره داشته و تحلیل‌های خود را در نمودار سطح زیر منحنی مقایسه کردند و به این نتیجه رسیدند دقت شبکه مصنوعی بالاتر است [۱۵].

## مقدمه

بیماری دیابت یکی از مشکلات عمده نظام سلامت در نقاط مختلف جهان است. به طور متوسط ۱۰ درصد مردم جهان به این بیماری مبتلا هستند [۱] نزدیک به ۲۵/۸ میلیون آمریکایی در سال ۲۰۱۱ میلادی به بیمار دیابت مبتلا بوده‌اند و ۷۹ میلیون نفر دیگر نیز در خطر ابتلا به این بیماری هستند [۲]. در ایران نیز نزدیک به ۹/۳ درصد نفر مبتلا به بیماری دیابت هستند. شیوع بیماری دیابت در مناطق شهری ایران با سرعت نگران‌کننده‌ای در حال افزایش است [۳].

علاوه بر هزینه‌های خانوار، هزینه‌های اجتماعی بیماری دیابت، شامل هزینه‌های درمان بیماران و از دست رفتن بهره‌وری نیروی کار، هزینه قابل توجهی بر روی بخش عمومی ایجاد می‌نماید. به طوری که هزینه‌های درمان هر بیمار دیابتی در طول عمر آنان به طول میانگین در آمریکا در سال ۲۰۱۲ نزدیک ۸۵۲۰۰ دلار تخمین زده شده است که ۵۳ درصد آن مربوط به عوارض دیابت است [۴].

دیابت یک بیماری همه‌گیر می‌باشد که در اثر کاهش یا نبود انسولین در بدن اتفاق می‌افتد و بر دو نوع یک و دو تقسیم می‌شود. در دیابت نوع یک یا دیابت وابسته به هورمون انسولین (هورمون تنظیم‌کننده قند خون)، بدن بیمار قادر به تولید انسولین نیست. دیابت نوع یک در هر سنی اتفاق می‌افتد اما در کودکان بسیار شایع‌تر است [۵]. در دیابت نوع دوم، انسولین در مراحل اولیه به میزان کافی ترشح می‌شود، ولی به نوعی، مقاومت در برابر عملکرد طبیعی این هورمون در سطح سلول‌های بدن وجود دارد که نتیجه آن بالا رفتن سطح قند خون می‌باشد. این نوع دیابت به نوع سبک زندگی و الگوهای غذایی افراد وابسته است. در صورت تشخیص در مراحل اولیه بیماری، امکان موفقیت در درمان تا حدود ۶۰ درصد وجود دارد. دیابت نوع دوم بسیار شایع‌تر از نوع یک بوده به گونه‌ای که نزدیک به ۹۵ درصد از ابتلا به بیماری دیابت را شامل می‌شود [۱].

تقریباً نیمی از مبتلایان به بیماری دیابت از بیماری خود اطلاع ندارند [۶]. از دلایل عمده عدم تشخیص به موقع بیماری دیابت

تموراس مقایسه مطالعاتی به روی مدل‌های تشخیص دیابت داشتند و با طراحی مدلی از شبکه عصبی نشان دادند در بین شبکه عصبی و مدل‌هایی همچون رگرسیون مدل شبکه عصبی عملکرد بهتری دارد [۱۱].

در این مطالعه از شبکه عصبی به منظور پیش‌بینی ابتلا به دیابت در میان افراد مستعد دیابت استفاده شده است. از الگوریتم ممیتیک به منظور برورسانی اوزان الگوریتم شبکه عصبی استفاده می‌گردد.

### روش کار

پژوهش حاضر از نوع کاربردی و جامعه‌ی هدف آن متشکل



شکل ۱: گام‌های توسعه و مقایسه مدل‌ها

پیش‌بینی با الگوریتم ژنتیک و رگرسیون چند متغیره با ماتریس آشفستگی و نمودار مساحت زیر منحنی محاسبه گردید. پس مقادیر ویژگی‌ها از مقادیر  $\pm 1$  به مقدار اصلی برگردانده شد. در ادامه به توضیحات هر یک از این مراحل پرداخته می‌شود. جهت پیش‌بینی بیماری دیابت از متغیرهای فردی و جمعیت‌شناسی استفاده شد. بر اساس مقایسه مقادیر این متغیرها در بین افراد بیمار و سالم، پیش‌بینی ابتلا به بیماری صورت گرفت. این متغیرها عبارتند از سن، میزان قند خون ناشتا، میزان فشار خون، شاخص توده بندی و میزان تغییرات وزنی فرد. سپس مقادیر این متغیرها با ۵۴۵ فرد بیمار و سالم از مرکز دیابت دانشگاه علوم پزشکی همدان جمع‌آوری گردید. از میان این افراد، تعداد ۳۴۵ نفر افراد دیابتی و ۱۷۰ نفر نیز افراد سالم هستند. جزئیات ویژگی‌های بالینی هر دو گروه بیماران و افراد سالم در جدول یک ارائه گردیده است.

همان‌طور که شکل یک نشان می‌دهد در ابتدا داده‌های مورد نیاز جمع‌آوری گردید. سپس با استفاده از روش‌های پیش‌پردازش داده مقدار کمینه و بیشینه ویژگی‌های ورودی شبکه عصبی به بازه  $\pm 1$  تبدیل شد. در مرحله بعد پارامترهای مورد نیاز شبکه عصبی مانند نرخ ممتم، توابع انتقال، تابع خطا از طریق روش آزمون و خطا و تجربه مقالات دیگر به دست آید. سپس به منظور بهتر شدن وزن‌های شبکه عصبی از الگوریتم ممیتیک استفاده شد. از آنجا که الگوریتم ممیتیک تلفیقی از الگوریتم ژنتیک، الگوریتم جستجوی محلی و شبکه عصبی (جهت ارزیابی دقت نتیجه اوزان) می‌باشد، در این راستا در ابتدا پارامترهای بهینه مورد نیاز برای شبکه عصبی مانند نرخ ممتم، توابع انتقال، تابع خطا از طریق روش آزمون و خطا و مطالعه به دست آمد. در مرحله بعد، پارامترهای بهینه الگوریتم ژنتیک محاسبه شده و جهت بالا بردن دقت در پیش‌بینی در الگوریتم ممیتیک به کار برده شد. در نهایت نیز نتایج مدل

جدول ۱: ویژگی افراد مبتلا به دیابت و سالم

انحراف معیار	گروه سالم			گروه مبتلا به دیابت			سن
	حد اکثر	حداقل	میانگین	انحراف معیار	حد اکثر	حداقل	
۲۹/۳	۷۰	۱۶	۵۵	۳۲/۶۷	۷۰	۱۳	۳۶
۲۳/۴	۱۱۰	۶۷	۹۵	۹۰/۲۴	۳۰۰	۱۲۰	۱۵۶
۲۶/۸	۸۵	۴۰	۷۷	۳۰/۲	۹۷	۴۰	۷۵
۲۳/۷۵	۴۰	۱۶/۳	۲۲/۹	۹/۳	۵۵/۵	۱۹	۲۶/۸
۶/۹	۹	-۱	۳.۰	۴/۸	۳/۱	-۹	۲/۰

گام‌های پیش پردازش داده - از روش کمینه-بیشینه استفاده شد. این روش در فرمول شماره یک ارائه شده است.

$$X_{new} = \left( 2 * \frac{(X_{max} - X_{min})}{(X_{min} - X_{max})} \right) - 1 \quad (1)$$

جهت توسعه مدل، در گام نخست مقادیر هر یک از این متغیرها در جدول دو جهت طبقه بندی و دقت در تشخیص، نرمال سازی شد. جهت نرمال سازی - به عنوان یکی از

مدل‌های مختلفی از توابع انتقال و خطا مورد آزمایش قرار گرفتند که بهترین نتایج برای تابع انتقال لایه اول، تابع انتقال لایه دوم، تابع خطا، تعداد نرون ورودی، تعداد نرون میانی، به ترتیب سیگموید (Sigmoid)، تانژانت (Tangent)، میانگین توان دوم خطا، پنج، هفت است. نتایج حاصل از شبکه عصبی با نرون‌های مختلف در جدول دو ثبت گردید. همانطور در جدول دو ارائه شده است شبکه عصبی با هفت نرون و نرخ یادگیری ۰/۵۷ و نرخ منتم ۰/۷۶ دارای بهترین دقت در تست آمد.

به ترتیب کمترین و بیشترین مقدار در هر متغیر هستند. در این روش مقادیر اندازه گیری شده به محدوده ی جدیدی تبدیل شدند. این محدوده در این مقاله در محدوده  $\pm 1$  تعریف شد. بعد از نرمال سازی داده ها، مقادیر هر یک از ویژگی‌ها وارد شبکه عصبی شدند. الگوریتم‌های شبکه عصبی داده‌های خروجی از مرحله نرمال سازی را از طریق نرون‌های لایه اول دریافت کرد و آموزش قرار داد. و بعد از آموزش، داده ها در دو کلاس یک به عنوان بیماران دیابتی و منفی یک به عنوان افراد سالم تقسیم شدند. در این مطالعه به منظور آزمایش، و به دست آوردن دقت مناسب،

جدول ۲: تعیین و مقایسه پارامترهای شبکه عصبی

تعداد نرون لایه مخفی	نرخ یادگیری	نرخ منتم	دقت آموزش	دقت تست
۳	۰/۶۳	۰/۷۳	۹۳	۷۹
۷	۰/۵۷	۰/۷۶	۸۹	۸۸
۱۶	۰/۶۰	۰/۷۷	۸۹	۸۷
۳۶	۰/۶۲	۰/۷۲	۹۱	۸۷

بالا برد [۱۷]. در مطالعه حاضر بعد از آموزش شبکه عصبی برای به دست آوردن دقت بالاتر نتیجه، جهت بروز رسانی اوزان از الگوریتم

یکی از نکاتی که باید در شبکه عصبی در نظر گرفت وزن ها و پارامتر های قابل تنظیم شبکه عصبی است و در صورت انتخاب درست وزن ها می توان دقت شبکه عصبی را به میزان قابل توجهی

ممتیک استفاده شده است. سپس الگوریتم مذکور با مدل‌های توسعه یافته مقایسه شد.

الگوریتم ممتیک ترکیبی از الگوریتم ژنتیک و جستجوی محلی می‌باشد. الگوریتم ژنتیک کلاسیک برای پیدا کردن سطح پاسخ، سریع عمل می‌کند اما بهینه سراسری را به کندی پیدا می‌کند. بدین منظور با اضافه کردن الگوریتم‌های جستجوی محلی می‌توان الگوریتم ژنتیک را بهبود داد [۱۸]. همچنین الگوریتم ممتیک به عنوان یک پل بین یک جستجوی سراسری و جستجوی محلی می‌باشد. این الگوریتم در مقایسه با هم‌تایان معمولی خود دارای سرعت بالا فرآیند جستجو بوده است و دارای راه‌هایی با دقت بالاتر در حل مشکلات پیچیده می‌باشد. از آنجا که الگوریتم ممتیک نیز شامل پارامترهایی است که در نتیجه آزمایش موثر است بر همین اساس بعضی از پارامترهای این الگوریتم مانند اندازه

جمعیت، نرخ ترکیب، و نرخ جهش مطابق زیر مورد تست قرار می‌گیرد:

۱. نرخ ترکیب (Cross rate): معمولاً بین ۸۰ درصد - ۹۵ درصد پیشنهاد شده، اما در برخی موارد نرخ ۶۰ درصد نیز نرخ مناسب ذکر گردیده است [۱۹].
  ۲. نرخ جهش (Mutation rate): معمولاً پایین و از ۰/۲ تا ۵/۰ درصد ذکر گردیده است.
  ۳. جمعیت اولیه (Initial population): بین ۲۰-۳۰ توصیه می‌گردد اما در برخی مطالعات نیز اندازه جمعیت مطابق با اندازه کروموزم توصیه می‌شود [۲۰، ۲۱].
- با توجه به اینکه بررسی ترکیب کلیه حالات فوق با هم مشکل است، با استفاده از روش آزمایش تاگوچی [۲۲، ۲۳] چند حالت از میان کل مجموعه حالات مطابق جدول سه بررسی و بهترین پارامترها انتخاب شد.

جدول ۳: پارامترهای پیشنهادی تاگوچی

جمعیت اولیه	نرخ ترکیب	نرخ جهش
۲۰	۰/۶۰	۰/۰۵
۲۰	۰/۸۰	۰/۱۰
۲۰	۰/۹۰	۰/۲۰
۳۰	۰/۶۰	۰/۱۰
۳۰	۰/۸۰	۰/۲۰
۳۰	۰/۹۰	۰/۰۵
۵۰	۰/۶۰	۰/۲۰
۵۰	۰/۸۰	۰/۰۵
۵۰	۰/۹۰	۰/۱۰
۳۰	۰/۹۰	۰/۱۰

یکی از مدل‌هایی که جهت مقایسه در این مطالعه استفاده گردید، رگرسیون لجستیک می‌باشد. رگرسیون لجستیک از ابزارهای آماری به شمار می‌آید که به منظور مدل‌سازی و تحلیل داده‌ها از آن استفاده می‌گردد. شکل کلی این مدل به صورت زیر است:

$$\text{Log} \left( \frac{\pi}{1-\pi} \right) = \alpha + \sum \beta x$$

که در آن  $x$  بیانگر متغیرهای پیشگو است  $\beta$  ضریب برآورد شده مدل برای متغیر مستقل مربوطه و  $\pi$  احتمال ابتلا یا

نتایج پیشنهادی روش آزمایشات تاگوچی در جدول سه ارائه شده است. بهترین نرخ‌ها بعد از آزمایش مربوط به نرخ‌های جهش ۰/۰۵ و ترکیب ۰/۹ بود که منجر به دقت آموزش ۹۰ و دقت تست ۸۸ گردید (جدول سه). بعد از به دست آوردن بهترین پارامترها از الگوریتم ژنتیک و شبکه عصبی، این پارامترها به عنوان بهترین پارامترهای الگوریتم ممتیک تعیین شد و نتایج حاصل به ترتیب با دقت آموزش ۹۲/۸۴ و دقت تست ۹۳/۲۳ به دست آمد.

حساسیت مدل در تشخیص افراد بیمار از سالم است. [۲۶].

### یافته ها

همانطور که در روش کار توضیح داده شد، به منظور ساخت مدل در ابتدا پارامترهای مناسب شبکه عصبی انتخاب گردید. سپس به منظور انتخاب پارامترهای مورد نیاز الگوریتم ممیتیک محاسبه گردید و مدل مورد آزمایش قرار گرفت و نتایج هر بررسی ثبت شد. در این مرحله مدل پیشنهادی (ممیتیک) با مدل های ژنتیک، رگرسیون چند متغیره با استفاده از داده های آزمایش مقایسه گردید، که برای بررسی تاثیر انتخاب تصادفی بودن داده های آموزشی ۲۵ درصد از داده ی ما را تشکیل می دهد و این مقدار به صورت تجربی و از مقالات مورد مطالعه انتخاب گردیده است. بدین منظور، ابتدا مدل ها توسط ماتریس آشفستگی مورد مقایسه قرار گرفت. سپس مدل ها توسط نمودار سطح زیر منحنی مورد بررسی شد. در ماتریس آشفستگی مدلی مناسب است که دارای حساسیت، ویژگی، ارزش اخباری منفی، ارزش اخباری مثبت بالاتر و درست نمایی منفی پایین داشته باشد. نتایج حاصل از ماتریس آشفستگی در جدول چهار ارائه شده است. این ماتریس مدل های الگوریتم های ژنتیک، ممیتیک و رگرسیون لجستیک را با یکدیگر مقایسه می کند.

عدم ابتلا به بیماری است [۲۴]. در این مطالعه نتایج بدست آمده با مدل رگرسیون مقایسه و تحلیل گردید.

برای تحلیل میزان دقت مدل ها برای تشخیص افراد بیمار از غیر بیمار از ماتریس آشفستگی و سطح زیر منحنی استفاده شد [۲۵] در نمودار سطح زیر منحنی و ماتریس درهم ریختگی، دو پارامتر حساسیت (Sensitivity) و ویژگی (Specificity) به کار می رود. با توجه به اینکه حساسیت و ویژگی به تنهایی در پیش بینی افراد بیمار از سالم موثر نمی باشند از معیار های ارزش اخباری مثبت، ارزش اخباری منفی و میزان درست نمایی مثبت و میزان درست نمایی منفی استفاده می شود. در فرمول درست نمایی مثبت از نسبت مثبت های واقعی در بیماران بر نسبت مثبت های کاذب در افراد غیربیمار به دست می آید. بنا براین هر چه این مقدار از یک بیشتر باشد نشان از بهتر بودن مدل در افراد بیمار دارد. فرمول میزان درست نمایی منفی از نسبت منفی های کاذب در بیماران بر نسبت منفی های واقعی در افراد غیربیمار به دست می آید بنابراین، هر چه این مقدار به صفر نزدیک باشد بهتر است [۲۶,۲۷].

مساحت زیر منحنی، روشی است جهت بررسی مدل و تست آزمایشگاهی و نشان می دهد مدل طراحی شده چقدر توانسته در تشخیص بیماری موفق باشد. بدین منظور نمودار سطح زیر منحنی از نموداری با دو بعد استفاده می کند که محور X نمودار نشان دهنده میزان و محور عمودی نشان دهنده میزان

جدول ۴: مقایسه مدل ها با ماتریس آشفستگی بر حسب درصد

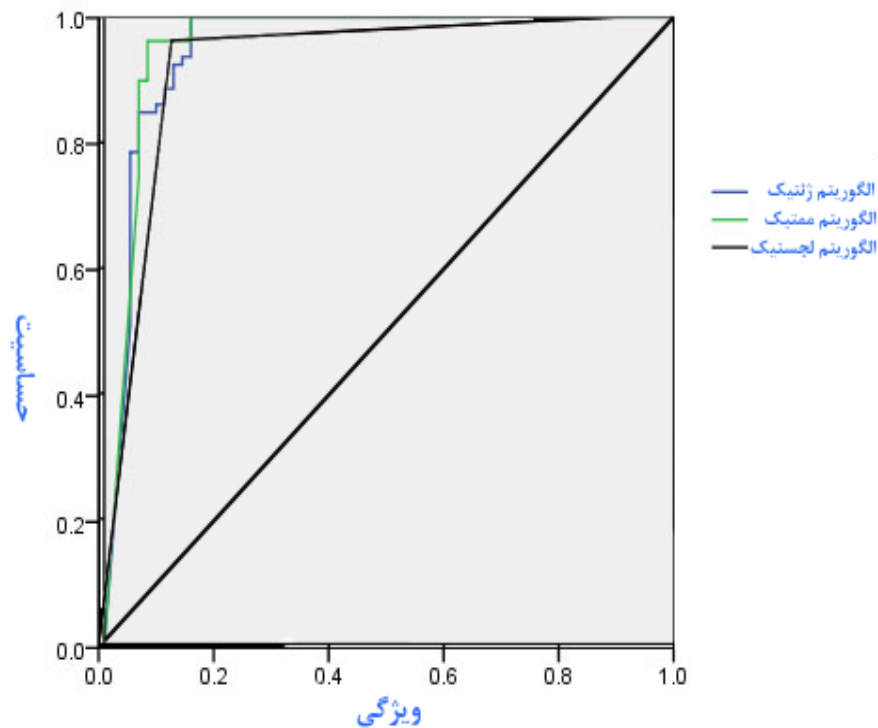
ژنتیک	ممیتیک	رگرسیون لجستیک	حساسیت
۹۸/۰۰	۹۶/۲۰	۹۵/۶۰	۸۴/۵۰
۸۴/۸۰	۹۲/۴۰	۸۴/۵۰	۹۴/۷۰
۸۸/۶۰	۹۳/۸۰	۹۴/۷۰	۸۷/۰
۹۸/۲۰	۹۵/۳۰	۸۷/۰	۶/۲۰
۶/۵۰	۱۲/۷۰	۶/۲۰	۰/۰۵
۰/۰۱۴	۰/۰۴۱	۰/۰۵	

همانطور که جدول چهار نشان می دهد در بین مدل های مقایسه شده مدل های حاصل از الگوریتم ممیتیک و الگوریتم ژنتیک دارای حساسیت بالایی می باشند. اما میزان ویژگی مدل حاصل

از الگوریتم ژنتیک در مقایسه با سایر مدل ها پایین است. همچنین در بین مدل ها، مدل پیشنهادی (الگوریتم ممیتیک) مدلی است که دارای میزان درست نمایی مثبت، ارزش

سمت بالا و چپ (۰ و ۱) نزدیک تر باشد مناسب تر است و مدل پیش بینی به حالت ایده آل خود نزدیک تر است. این نقطه انطباق کامل مدل پیش بینی شده را با مدل واقعی دارد. نقطه ی مقابل آن با مختصات (۱ و ۰) نشان دهنده عکس مدل واقعی است. در نقطه (۱ و ۰) نتیجه ی سوالاتی که منجر به این حالت شده را می توان عکس کرد.

اخباری مثبت، ارزش اخباری منفی بالا و میزان درست نمایی منفی پایین و نزدیک به صفر بهتری است پس می توان این مدل را به عنوان مدلی مناسب انتخاب کرد. نتایج حاصل از نمودار سطح زیر منحنی در شکل شش نشان داده شده است. محور افقی (X) میزان ویژگی و محور عمودی (Y) میزان حساسیت در هر مدل را نشان می دهد. هر منحنی نشان دهنده عملکرد یک مدل است. در نمودار شکل شش هر چه نقاط به



شکل ۲: نمودار سطح زیر منحنی الگوریتم ژنتیک، ممیتیک، رگرسیون لجستیک

سطح زیر منحنی ۰/۹۳۹ از کمترین دقت برخوردار است.

### بحث و نتیجه گیری

این مطالعه مدل هایی را با استفاده از الگوریتم های شبکه عصبی و ممیتیک جهت پیش بینی بیماری دیابت بر اساس مقادیر متغیرهای فردی و بالینی (میزان قند خون ناشتا، فشار خون، شاخص توده بدنی و میزان تغییرات وزن بدن) ارائه می کند. در طی این مطالعه سعی شد تا با بهبود وزن های شبکه

در سطح معنی داری  $\alpha = 0.05$  برای سه مدل مذکور، مساحت زیر منحنی های بیانگر عملکرد خوب مدل ها در تمایز بین دو گروه نرمال و بیمار است. همانطور که شکل دو نشان می دهد مناطق زیر منحنی برای مدل ممیتیک ۰/۹۵۸ است که بیانگر دقت بهتر مدل در تمایز بین دو گروه بیمار و غیر بیمار است. سپس بعد از مدل ممیتیک، مدل الگوریتم ژنتیک با مقدار به دست آمده ۰/۹۵۲ دارای مساحت زیر منحنی بیشتری نسبت به الگوریتم ممیتیک است و مدل رگرسیون لجستیک با

جهت پیش بینی ابتلا به بیماری دیابت ارائه می‌دهد. به طور سنتی از میزان قند خون ناشتا جهت تشخیص ابتلای فرد به دیابت استفاده می‌شود. مدل‌های ارائه شده در این پژوهش علاوه بر قند خون ناشتا، از فاکتورهای سبک زندگی مانند شاخص توده بدنی، میزان تغییرات وزن فرد و همچنین فشار خون فرد جهت تشخیص ابتلا استفاده می‌نماید. بنابر این نتایج می‌توان با دقت بالا با استفاده از مقادیر متغیرهای فوق به تشخیص بیماری پرداخت. در نظر گرفتن این فاکتورها در پیش بینی بیماری از اهمیت بسیار زیادی برخوردار است، با توجه به اینکه تغییرات در سبک زندگی بیمار فاکتور تعیین کننده تغییرات کلینکی مانند سطح قند خون است که در مراحل بروز بیماری قابل شناسایی است. بنابراین، این مطالعه با به کارگیری فاکتورهای سبک زندگی می‌تواند بینش بسیار ارزشمندی نسبت به پیش بینی ابتلای بیمار فراهم نماید.

انتظار می‌رود مدل ارائه شده در این مطالعه در مراحل بعدی جهت استفاده در برنامه‌های پزشکی خانواده و سایر برنامه‌های مدیریت سلامتی مانند برنامه‌های غربالگری جهت تعیین میزان خطر ابتلا و پیشرفت بیماری مورد استفاده قرار گیرد. به ویژه در کشورمان که نظام ارائه خدمت به طور فزاینده‌ای بر پیشگیری از بیماری و تشخیص بیماری در مراحل ابتدایی تاکید می‌نماید. همچنین یافته‌های این مطالعه می‌تواند در راستای تهیه نرم افزارهای مدیریت خطر و همچنین توسط ارائه کنندگان بیمه‌های درمانی جهت تعیین میزان مشارکت افراد در تامین هزینه‌های سلامتی در آینده مورد استفاده قرار گیرد. این مطالعه دارای محدودیت‌هایی نیز هست. مدل الگوریتم ممیتیک و الگوریتم‌های فرا ابتکاری در مقایسه با مدل‌های رگرسیون نیازمند زمان طولانی تری برای آموزش هستند. همچنین جهت آموزش بهتر مدل به پردازشگرهای قدرتمندی نیاز است تا اینکه ساعت‌ها اجرا شده و بتوانند مدل‌هایی با دقت بالاتر ارائه دهند. همچنین کمبود داده به منظور آموزش بیشتر مدل، داده‌های گمشده در نمونه‌ها و احتمال خطا در ثبت داده به دلیل بکارگیری روش‌های سنتی (کاغذی) از دیگر محدودیت‌های این مطالعه است [۲۹].

عصبی مدلی مناسب جهت پیش بینی بیماری دیابت ارائه گردد. در این راستا با استفاده از الگوریتم ممیتیک به بروز-رسانی وزن‌های شبکه عصبی پرداخته شد، که بعد از به دست آوردن وزن‌ها و اعمال آن به شبکه عصبی مدلی ارائه شد که دارای دقت ۹۳ درصد و با نرخ‌هایی به ترتیب با حساسیت، ویژگی، ارزش اخباری مثبت، ارزش اخباری منفی، سطح زیر منحنی، ۰/۹۵۸، ۹۵/۳، ۹۳/۸، ۹۲/۴، ۹۶/۲، به دست آمد.

در این مطالعه همچنین مدل ممیتیک با الگوریتم ژنتیک و مدل رگرسیون مقایسه گردید. الگوریتم ممیتیک در مقایسه با الگوریتم ژنتیک و مدل رگرسیون دارای دقت بالاتری است. نتایج این پژوهش یافته‌ای مطالعات قبلی مانند وانگ و همکاران را تایید می‌کند. آنها دو مدل شبکه مصنوعی و رگرسیون لجستیک چند متغیره را با هم مقایسه کرده‌اند. نتایج آن‌ها نشان داد که مدل شبکه عصبی از لجستیک دارای دقت بالاتری می‌باشد [۱۵].

چونجیان و همکاران، مطالعه‌ای برای شناسایی بیماران DM۲T و آزمایش کردن شبکه عصبی مصنوعی و لجستیک چند متغیره داشتند و تحلیل‌های خود را در نمودار سطح زیر منحنی مقایسه کردند و به این نتیجه رسیدند دقت شبکه عصبی مصنوعی از لجستیک چند متغیره بالاتر است [۱۵]. نتایج مطالعه حاضر نیز با نمودار سطح زیر منحنی انجام شد و نشان از دقت بالای شبکه عصبی مصنوعی دارد که با نتیجه تحقیق مذکور همخوانی دارد. در این مطالعه نیز مدل پیشنهادی با رگرسیون چند متغیره مقایسه شد که نشان داد مدل ممیتیک از رگرسیون دقت بالاتری دارد.

لنگری زاده و همکاران با استفاده از شبکه عصبی به پیش بینی تولد نوزاد نارس در مادران باردار شده پرداخته‌اند که نتایج به دست آمده از این پژوهش، نشان داد استفاده از شبکه پرسپترون چندلایه برای پیش بینی نتیجه زایمان از نظر تولد نوزاد ترم یا نوزاد نارس در مادران باردار شده از طریق فناوری‌های کمک باروری می‌تواند در پیشگیری از عوارض تولد نوزاد نارس کمک کننده باشد [۲۸].

مطالعه حاضر نیز نشان داد که استفاده از شبکه عصبی در پیش بینی دیابت کمک کننده است. این مطالعه مدل نوآورانه‌ای



## References

1. Pei E, Li J, Lu C. Effects of lipids and lipoproteins on diabetic foot in people with type 2 diabetes mellitus: a meta-analysis. *Journal of Diabetes and its Complications* 2014; 28 (4): 559–64.
2. Centers for Disease Control and Prevention (CDC). National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States. US Department of Health and Human Services, Centers for Disease Control and Prevention [serial online] 2011 Dec [cited 2014 December 5]. 201 (12): [12screens]. Available from: URL: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:National+diabetes+fact+sheet,+2011.+2011#2>.
3. Lotfi H, Saadati H. ]Prevalence of Diabetes in People aged  $\geq 30$  years: The Results of Screening Program of Yazd Province, Iran, in 2012]. *Journal of research in health sciences* 2014; 14 (1): 87–91. [Persian]
4. Zaugg S, Dogbey G. Diabetes Numeracy and Blood Glucose Control: Association With Type of Diabetes and Source of Care. *Clinical Diabetes* 2014; 32 (4): 152–157.
5. Nielsen D, Krych L. Buschard, Beyond genetics. Influence of dietary factors and gut microbiota on type 1 diabetes. *Federation of European Biochemical Societies* 2014; 588 (22): 4234–4243.
6. Zealand N, Amos A. The rising global burden of diabetes and its complications: estimates and

مطمئناً مدل‌های ارائه شده در این مطالعه را می‌توان بهبود بخشید. هر چند مدل الگوریتم ممیتیک به دست آمده دارای دقت بالایی است اما با جمع آوری داده‌های جدید می‌توان مدل را مجدد آموزش و دقت آن را افزایش داد. همچنین جهت بهبود پایایی، روش مطالعه، مدل‌های ارائه شده می‌تواند توسط سایر محققان استفاده شود که از طرفی منجر به بهبود قابلیت مدل‌ها شود و از طرفی دیگر نیاز به تدوین مدل‌های مشابه را در مطالعات جدید برطرف می‌نماید که در نهایت به صرفه جویی در هزینه و زمان منجر می‌شود.

## تشکر و قدردانی

این مقاله حاصل بخشی از پایان نامه تحت عنوان «طراحی سیستم پشتیبان تصمیم‌گیری جهت تشخیص و پیش‌بینی دیابت» در مقطع کارشناسی ارشد در سال ۱۳۹۳ کد ۲۱۷۹۶۹۴ می‌باشد که با حمایت دانشگاه قم اجرا شده است.

- projections to the year 2010. *Diabetic medicine* 1997; 14 (s5) S7–S85.
7. Manzella D, Grella R, Abbatecola, G, Paolisso, Repaglinide Administration Improves Brachial Reactivity in Type 2 Diabetic Patients. *Diabetes Care* 2005; 28 (2): 366–371.
8. Zhuo X, Zhang P. The lifetime cost of diabetes and its implications for diabetes prevention. *Diabetes Care* 2014; 37 (9): 2557–64.
9. Chavey A, Kioon M. Maternal diabetes, programming of beta-cell disorders and intergenerational risk of type 2 diabetes. *Diabetes & Metabolism* 2014; 40 (5): 323–330.
10. Channouf N, L'Ecuyer P. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Management Science* 2007; 10 (1): 25–45.
11. Temurtas H, Yumusak N. A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with Applications* 2009; 36 (4): 8610–8615.
12. Parsaeian M, Mohammad K. Original Article Comparison of Logistic Regression and Artificial Neural Network in Low Back Pain Prediction : Second National Health Survey. *Iranian Journal of Public Health* 2012; 41 (6): 86–92.
13. Vukicevic A, Jovicic G. Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced bladder cancer outcome. *Expert Systems with Applications* 2014; 41 (18): 8092–8100.
14. Patel V, Shortliffe E. The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine* 2009; 46 (1): 5–17.
15. Wang C, Li L. Evaluating the risk of type 2 diabetes mellitus using artificial neural network: An effective classification approach. *Diabetes Research and Clinical Practice* 2013; 100 (1): 111–118
16. Palar P, Tsuchiya T. A comparative study of local search within a surrogate-assisted multi-objective memetic algorithm framework for expensive problems. *Applied Soft Computing Journal* 2016; 43 (2): 1–19.
17. Nguyen D, Widrow B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Neural Networks Journal* 1990; 13 (7): C21.
18. Knowles D, Corne D. A memetic algorithm for multiobjective optimization, In: William E. Hart N Krasnogor editors. *Proceedings of the 2000 Congress on Evolutionary Computation; 2000 July 16-19; California, USA*. New York: Institute of Electrical and Electronics Engineers;2000. 325-332
19. KHODAVAISI H. [Comparing the Exchange Rates Predicted by STAR Non-linear Models and Alternative Models]. *Economic Modelling* 2013; 7 (3): 85–105. [Persian]
20. Leardi R, Boggia R. Genetic algorithms as a strategy for feature selection. *Journal of Chemometrics* 1992; 6 (5): 267–281.
21. Leardi R. Application of a Genetic Algorithm To Feature Selection Under Full Validation

- Conditions and To Outlier Detection. Journal of Chemometrics 1994; 8 (7): 65–79.
22. Tchognia N, Hartiti B. Application of Taguchi approach to optimize the sole gel process of the quaternary Cu<sub>2</sub>ZnSnS<sub>4</sub> with good optical properties Herv e. Optical Materials 2016; 57 (7): 85–92.
23. Yang Y. Design optimization of cutting parameters for tuning operation based on taguchi mehod . Journal of Material Processing Technology 1997; 84(1): 123–129.
24. Upadhyaya S, Farahmand K. Comparison of NN and LR classifiers in the context of screening native American elders with diabetes. Expert Systems with Applications 2013; 40 (15): 5830–5838.
25. Sokolova M, Japkowicz N. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. Advances in Artificial Intelligence 2006; 4307(3): 1015–1021.
26. Fawcett T. An introduction to ROC analysis. Pattern recognition letters 2006; 27(8): 861–874.
27. Kim J, Lee J. Data-Mining-Based Coronary Heart Disease Risk Prediction Model Using Fuzzy Logic and Decision Tree. Healthcare Informatics Research 2015; 21 (3): 167–174.
28. Langarizadeh M, Ghazi Saeedi M. [Predicting Premature Birth in Pregnant Women via Assisted Reproductive Technologies using Neural Network]. Journal of Health Administration 2016; 18 (62): 42–51. [Persian]
29. Rezaeenour J, Yari-Eili M, Roozbahani Z, Ebrahimi M. Prediction of Protein Thermostability by an Efficient Neural Network Approach. Journal of Health Management and Informatics 2016; 3(4):102-110.

# Proposing a Model for Predicting Diabetes Patients by Neural Network

Jahani M<sup>1</sup>/ Rezaeenour J<sup>2</sup>/ Mahdavi M<sup>3</sup>/ Hadavandi E<sup>4</sup>

## Abstract

**Introduction:** Meta-heuristic and combined algorithms have a great capability in modelling medical decision making. This study used neural networks in order to predict Type 2 Diabetes (T2D) among high risk individuals.

**Methods:** This study was an applied research. Data from 545 individuals (diabetic and non-diabetic), in Diabetes Clinic of Hamedan University of Medical Sciences, were used to develop predictive diabetes models. Memetic algorithms which are a combination of genetic algorithm (GA), local search algorithm, and neural networks were applied to update weights and improve predictive accuracy of neural network models. In the first step, optimum parameters for neural networks such as momentum rate, transfer functions, and error functions were examined through trial and error and other studies.

**Results:** The preliminary analysis showed that the accuracy of neural networks was 88 percent. The use of memetic algorithm improved its accuracy to 93.2 percent. Among models, regression model had the least accuracy. For the memetic algorithm model the amount of sensitivity, specificity, positive predictive value, negative predictive value, and area under the curve were 96.2, 95.3, 93.8, 92.4, and 0.958, respectively. These parameters for GA were 98.0, 84.8, 88.6, 98.2, and 0.952 and for the logistic regression model were 95.6, 84.5, 94.7, 87.0, and 0.916, respectively.

**Conclusions:** Models developed by neural networks have a higher predictive accuracy than the regression model. The results of this study can contribute to risk management and planning of health services by providing healthcare decision makers with more accurate predictive models based on clinical and life style characteristics of individuals.

**Key words:** Diabetes, Decision Support Techniques, Neural network, Genetic Algorithms, Memetic algorithm

• Received: 18/Jun/2016 • Modified: 3/Sep/2016 • Accepted: 19/Nov/2017

1. MSc of Information Technology Engineering, Faculty of Technology and Engineering, University of Qom, Qom, Iran
2. Associate Professor of Department of Industrial Engineering, Faculty of Technology and Engineering, University of Qom, Qom, Iran; Corresponding Author (j. rezaee@qom. ac. ir)
3. PhD of Health Services Management and Organizations, Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, Netherlands
4. Assistant professor of Department of Industrial Engineering, Faculty of Computer and Industrial engineering, Birjand University, Birjand, Iran