مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره ششم شماره اول. ص۲۳-۱۲، ۱۳۹۸

پیش‌بینی دیابت با بهینه‌سازی الگوریتم نزدیک‌ترین همسایه توسط الگوریتم ژنتیک

محمد مومنی[۱]*، علی محمد لطیف[۲]، مهدی آقا صرام[۳]، کاظم حاج میرزاده[۴]، ثریا غراوی[۵]، سید محمد نقیب القرا[۶]

• دریافت مقاله: ۱۳۹۷/۲/۲۶ • پذیرش مقاله: ۱۳۹۷/۷/۷

مقدمه: دیابت یا بیماری قند یک اختلال متابولیک سوخت‌وسازی در بدن است که توانایی تولید انسولین در بدن از بین می‌رود و انسولین تولیدی نمی‌تواند عملکرد طبیعی خود را انجام دهد. وجود علائم و ویژگی‌های مختلف این بیماری، تشخیص را برای پزشکان دشوار می‌کند. داده‌کاوی امکان تحلیل داده‌های بالینی بیماران برای تصمیم‌گیری‌های پزشکی را فراهم می‌کند. هدف این پژوهش، ارائه یک مدل برای افزایش دقت پیش‌بینی دیابت است.

روش: در این مطالعه، پرونده پزشکی ۱۱۵۱ بیمار مبتلا به دیابت با تعداد ۱۹ ویژگی مورد بررسی قرار گرفت. اطلاعات بیماران از پایگاه داده استاندارد UCI جمع‌آوری شد. هر یک از بیماران حداقل به مدت یک سال تحت پیگیری بودند. به‌منظور ارائه مدل پیش‌بینی دیابت از الگوریتم ژنتیک و نزدیک‌ترین همسایه استفاده شد.

نتایج: نتایج نشان داد که دقت پیش‌بینی مدل پیشنهادی برابر با ۰/۷۶ بود. همچنین برای روش‌های نایو بیز، شبکه عصبی پرسپترون چند لایه و ماشین بردار پشتیبان دقت پیش‌بینی به ترتیب برابر با ۰/۶۲، ۰/۶۵ و ۰/۷۵ به دست آمد.

نتیجه‌گیری: در پیش‌بینی دیابت، مدل پیشنهادی نسبت به سایر مدل‌های موردمقایسه، دارای حداقل میزان خطا و بیش‌ترین دقت و صحت است. روش نایو بیز، حداکثر میزان خطا و کم‌ترین دقت را دارا می‌باشد.

کلید واژه‌ها: پیش‌بینی دیابت، الگوریتم ژنتیک، الگوریتم نزدیک‌ترین همسایه، داده‌کاوی

•ارجاع: مومنی محمد، لطیف علی محمد، آقا صرام مهدی، حاج میرزاده کاظم، غراوی ثریا، القرا سید محمد نقیب. پیش‌بینی دیابت با بهینه‌سازی الگوریتم نزدیک‌ترین همسایه توسط الگوریتم ژنتیک. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۸؛ ۶(۱): ۲۳-۱۲.

۱. دانشجوی دکتری، گروه کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه یزد، یزد، ایران
۲. دانشیار، گروه کامپیوتر، دانشکده مهندسی برق و کامپیوتر، هوش مصنوعی، دانشگاه یزد، یزد، ایران
۳. دکتری علوم کامپیوتر، دانشیار، گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه یزد، یزد، ایران
۴. متخصص بیماری‌های داخلی و دانشیار، دانشکده پزشکی، دانشگاه آزاد اسلامی، واحد یزد، یزد، ایران
۵. کارشناسی ارشد مهندسی کامپیوتر نرم افزار، مربی، گروه کامپیوتر، دانشکده مهندسی کامپیوتر و برق، نرم‌افزار، مجتمع آموزش عالی اسفراین، خراسان شمالی، اسفراین، ایران
۶. پزشک عمومی، دانشکده پزشکی، دانشگاه آزاد اسلامی، واحد یزد، یزد، ایران

*** نویسنده مسئول:** یزد، دانشگاه یزد، دانشکده مهندسی برق و کامپیوتر، گروه کامپیوتر

• شماره تماس: ۰۹۱۵۹۷۳۶۰۰۳

• **Email:** mohamad.momeny@stu.yazd.ac.ir

مقدمه

دیابت یا بیماری قند یک اختلال متابولیک در بدن است. در این بیماری توانایی تولید انسولین در بدن از بین می‌رود و یا بدن در برابر انسولین مقاوم شده؛ بنابراین انسولین تولیدی نمی‌تواند عملکرد طبیعی خود را انجام دهد. دیابت دو نوع اصلی دارد. در دیابت نوع یک تخریب سلول‌های بتا در پانکراس منجر به نقص تولید انسولین می‌شود و در نوع دو، مقاومت پیش‌رونده بدن به انسولین وجود دارد که در نهایت ممکن است به تخریب سلول‌های بتای پانکراس و نقص کامل تولید انسولین منجر شود. در دیابت نوع دو مشخص است که عوامل ژنتیکی، چاقی و کم‌تحرکی نقش مهمی در ابتلای فرد دارند [۱]. بیماری دیابت یکی از عوامل مرگ‌ومیر در بیشتر کشورها است [۲].

تشابه علائم بالینی و آزمایشگاهی دیابت، احتمال بروز خطا در تشخیص را افزایش می‌دهد. تشخیص و پیش‌بینی انواع بیماری‌ها با استفاده از تکنیک‌های داده‌کاوی امکان‌پذیر است. داده‌کاوی در پزشکی به فرآیند استخراج اطلاعات معتبر از پیش ناشناخته، قابل‌فهم و قابل‌اعتماد از پایگاه داده‌های پزشکی و استفاده ازآن‌جهت پیش‌بینی، تشخیص و کمک به درمان بیماری گفته می‌شود. کشف الگوهای مفید بین بیماری و علائم بالینی و آزمایشگاهی بیمار از کاربردهای داده‌کاوی در پزشکی است. منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیرمجموعه از داده‌های بیمار و تشخیص بیماری را بیان می‌کند [۳]. در سال‌های اخیر با پیشرفت‌هایی که در تشخیص زودرس این بیماری به وجود آمده، درمان آن نیز با موفقیت بیشتری همراه شده است. اگر توده‌های پستان در اندازه کوچک کشف شوند به خوبی قابل درمان هستند. داده‌کاوی از روش‌های جدید برای تشخیص زودرس دیابت است [۴].

Su و همکاران [۵] از شبکه‌های عصبی مصنوعی، درخت تصمیم، رگرسیون لجستیک و قواعد وابستگی با استفاده از عکس‌های سه بعدی بدن، بیماری دیابت را در مراجعه‌کنندگان پیش‌بینی کردند. در این تحقیق، عکس‌های سه‌بعدی و دوبعدی از تمامی اعضای بدن افراد مختلف (بیماران دیابتی و افراد سالم) ثبت گردید. سپس از این عکس‌ها ویژگی‌هایی مانند: سطح شکم، محیط پاها، حجم دست‌ها و... استخراج شد. با به‌کارگیری این متغیرها در تکنیک‌های داده‌کاوی بر پایه عکس‌های سه‌بعدی و دوبعدی بدن با دقت ۸۹٪ ابتلا به بیماری دیابت نوع دوم پیش‌بینی شد. استخراج قواعدی از قبیل «اگر حجم شکم = x و حجم پا =y و مساحت سینه =z باشد آنگاه با احتمال ۹۰٪ بیمار دیابت نوع دوم دارد» از نوآوری‌های این پژوهش بود.

Santhanam و Padmavathi [۶] برای کاهش بعد از الگوریتم ژنتیک و خوشه‌بند K-Means جهت کلاس‌بندی دیابت روش SVM (Support Vector Machine) را به کار بردند. برای ایجاد مدل از پایگاه داده استاندارد UCI(Universal Chess Interface) استفاده کردند. دقت ۹۶/۷۱٪ برای تشخیص دیابت نتیجه این مطالعه بود. در این پژوهش از الگوریتم K-Means برای حذف داده‌های نویزی، الگوریتم ژنتیک برای انتخاب ویژگی و در نهایت از الگوریتم ماشین بردار پشتیبان برای طبقه‌بندی دیابت استفاده نمودند.

Chakraborty و همکارش [۷] از بهینه‌سازی با استفاده از الگوریتم ژنتیک برای داده‌های میکروآری جهت تشخیص دیابت استفاده کردند. بهبود دقت تشخیص و پیش‌بینی دیابت با استفاده از بهینه‌سازی چندهدفه از نوآوری این پژوهش بود. استفاده از پایگاه داده استاندارد (United States Military Standard) MLL برای تشخیص دیابت از نقاط قوت این پژوهش بود. Maheta و Dabhi [۸] از برنامه‌نویسی بهینه‌سازی الگوریتم ژنتیک برای کلاس‌بندی داده‌های نامتقارن، برای پیش‌بینی از الگوریتم ماشین بردار پشتیبان و شبکه عصبی مصنوعی استفاده کردند.

دقت تشخیص و پیش‌بینی روش‌های داده‌کاوی را می‌توان با استفاده از الگوریتم ژنتیک افزایش داد. الگوریتم ژنتیک یکی از زیر گروه‌های محاسبات فرا ابتکاری است که از قوانین تکامل بیولوژیک طبیعی تبعیت می‌کند. الگوریتم ژنتیک با استفاده از قانون بقای برترین‌ها، در یک زیرمجموعه از پاسخ‌های مسئله، به دنبال به دست آوردن پاسخ‌های بهتر است. یک زیر مجموعه از پاسخ‌های ممکن برای مسئله، جمعیت اولیه را تشکیل می‌دهد. متناسب با ارزش هر یک از پاسخ‌ها، فرآیند انتخاب از جمعیت اولیه و تولیدمثل برای ایجاد نسل جدید انجام می‌شود. در هر نسل با ترکیب و تولیدمثل پاسخ‌های انتخاب شده، به کمک عملگرهایی که از ژنتیک طبیعی پیروی می‌کنند، تقریب‌های بهتری از جواب نهایی به دست می‌آید. این فرآیند باعث می‌شود که نسل‌های جدید با شرایط مسئله سازگارتر باشند [۹]، در این پژوهش سعی شد روش

نزدیک‌ترین همسایه برای افزایش دقت تشخیص و پیش‌بینی بیماری دیابت با الگوریتم ژنتیک را بهبود ببخشد.

هدف این مطالعه این است که با استفاده از مجموعه داده‌های مربوط به پرونده‌های بیماران در پایگاه داده UCI [۱۰]، مدلی برای افزایش دقت تشخیص و پیش‌بینی روش‌های نزدیک‌ترین همسایه توسط الگوریتم ژنتیک معرفی گردد.

روش

الف- گردآوری مجموعه داده‌ها

مجموعه داده‌های مربوط به دیابت، از پایگاه داده استاندارد UCI مربوط به سال ۲۰۱۴ گرفته شد. که شامل ۱۱۵۱ نمونه بود. نمونه‌هایی که فاقد اطلاعات کامل بودند از طریق روش بیش‌ترین فراوانی، تخمین زده شدند. برای هر بیمار تعداد ۱۹ ویژگی ثبت شده است. جدول ۱ نشان دهنده ویژگی‌های پایگاه داده مورد استفاده است.

قالب مناسب داده‌ها به عنوان ورودی داده‌کاوی در نتایج و خروجی تأثیرگذار است. اگر مقادیر ویژگی‌های مجموعه داده در دامنه متفاوتی قرار داشته باشند، احتمال بروز خطا در یافته‌ها افزایش می‌یابد. به قرار دادن داده‌های یک جامعه آماری در دامنه مشابه، نرمال‌سازی گفته می‌شود. در مدل پیشنهادی نحوه نرمال‌سازی به روش Max/Min و در بازه [۱-۰] است .

ب- الگوریتم ژنتیک

ایده اصلی الگوریتم ژنتیک از نظریه تکاملی داروین گرفته شده است. نظریه داروین به این شرح است که آن دسته از صفات طبیعی که با قوانین طبیعی سازگاری بیش‌تری دارند، شانس بقای بیش‌تری دارند. شایان ذکر است که نظریه تکاملی داروین هیچ اثبات تحلیلی و قطعی ندارد؛ اما از نظر تجربی و آماری تأیید شده است [۱۱].

الگوریتم ژنتیک در سال ۱۹۶۲ میلادی توسط جان هلند ارائه شد. این الگوریتم در گروه الگوریتم‌های بهینه‌سازی تصادفی قرار دارد و برای بهینه‌سازی مسائل پیچیده با فضای جست‌وجوی ناشناخته مناسب است [۱۲].

افراد جدید یک جامعه از طریق زادوولد تولید می‌شوند. شانس بقای یک فرد در نسل جدید به ترکیب خاص کروموزومی وابسته است. در مراحل زادوولد ممکن است جهش‌هایی در خصوصیات یک فرد نسل جدید رخ دهد که در نتیجه موجودی با خصوصیات عالی و سازگاری بالا تولید شود. در روند زادوولد به گونه‌های برتر در هر نسل اجازه تولیدمثل داده می‌شود و گونه‌های نامطلوب به‌تدریج از بین خواهند رفت و افراد نسل‌های جدید با گذشت زمان تکامل می‌یابند. الگوریتم ژنتیک در زیر به صورت خلاصه بیان شده است.

۱- مجموعه‌های تصادفی از کاندیداهای جواب به عنوان جمعیت اولیه تولید می‌شوند، تولید و در هر نسل با کاندیداهای جدیدی جایگزین می‌شوند.

۲- در هر تکرار الگوریتم، جمعیت توسط تابع برازندگی ارزیابی می‌شود. سپس تعدادی از بهترین کاندیداها برای نسل بعد گزینش می‌شوند و جمعیت جدید را تشکیل می‌دهند.

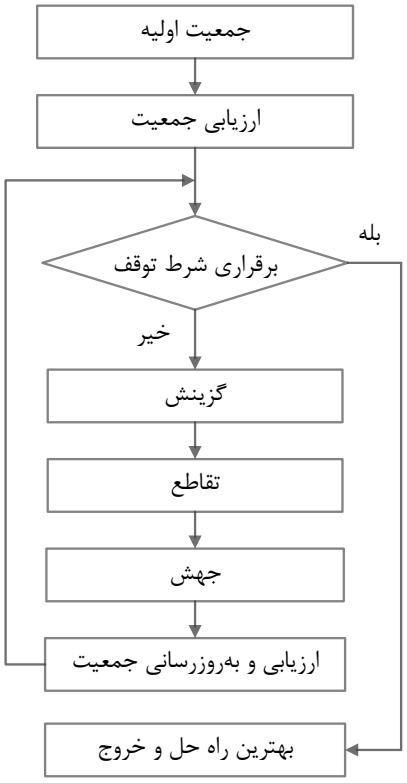
۳- تعدادی از این جمعیت با استفاده از اپراتورهای ژنتیکی نظیر تقاطع و جهش برای تولید فرزندان جدید استفاده می‌شوند.

۴- مراحل فوق تا رسیدن به یک پاسخ مناسب ادامه می‌یابد.

مراحل مطرح شده برای اجرای الگوریتم ژنتیک در قالب یک روندنما در شکل ۱ نشان داده شد.

جدول ۱: بازه مقادیر ویژگی‌های بالینی بیماران

توضیح ویژگی‌های بالینی بیماران	بازه مقادیر ویژگی‌های بالینی بیماران
نتیجه ارزیابی کیفیت	۰٪
نتیجه پیش‌نمایش (دارای علائم DR /بدون علائم DR)	۱٪
تعداد MA های کشف شده در سطح اطمینان α= ۰،۵ و .... ۱،۰	۷٪-۲٪
تعداد نقاط exudate در سطح اطمینان α= ۰،۵ و .... ۱،۰	۱۶٪-۸٪
فاصله اقلیدسی مرکز ماکولا از مرکز دیسک نوری	۱۷٪
نتایج AM/FM براساس طبقه‌بندی (دارای علائم DR /بدون علائم DR)	۱۸٪



شکل ۱: روندنمای الگوریتم ژنتیک

ج- مدل پیشنهادی

هر یک از ویژگی‌ها و یافته‌ها در تشخیص و پیش‌بینی دیابت از اهمیت خاصی برخوردار هستند. به‌بیان‌دیگر همه ویژگی‌ها دارای ارزش یکسان نیستند. به‌عنوان‌مثال در تشخیص بیماری دو ویژگی BMI (Body Mass Index) و عوامل ژنتیکی دارای اهمیت متفاوتی هستند. جدول ۲ توضیحات مربوط به هرکدام از ویژگی‌ها را نمایش می‌دهد. این‌که هر یک از ویژگی‌ها دارای چه ارزشی است و چقدر در تشخیص بیماری نقش دارد، مسئله مهمی است. در این مطالعه روشی ارائه شد که ارزش و نقش هر یک از ویژگی‌ها به‌طور دقیق مشخص و بیماری تشخیص داده شد. داده‌های موجود در پرونده بیماران با استفاده از نرم‌افزار Mathworks Matlab R2019a توصیف، شبیه‌سازی و تحلیل شد. در ادامه مدل پیشنهادی با روش‌های درخت تصمیم‌گیری، نایو بیز و نزدیک‌ترین همسایه مورد مقایسه قرار گرفت.

جدول ۲: توضیحات مربوط به ویژگی‌ها در تشخیص و پیش‌بینی دیابت

اطلاعات مشخص شده در مورد ویژگی‌ها و یافته‌ها در تشخیص و پیش‌بینی دیابت
۰) نتیجه باینری ارزیابی کیفیت که در آن ۱ نشان دهنده کیفیت بد و ۰ نشان دهنده کیفیت مناسب آن است.
۱) نتیجه دوتایی پیش از غربالگری، که در آن ۱ نشان دهنده اختلال شبکیه شدید و ۰ فقدان آن است.
۲-۷) نتایج تشخیص MA، هر مقدار ویژگی برای تعدادی از MAها در سطح اطمینان آلفا، ثابت است . . . ، ۱ ، ۰٫۵ به ترتیب
۸-۱۵) شامل اطلاعات مشابه ۲-۷) برای exudates است. به عنوان مجموعه‌ای از نقاط به جای تعداد پیکسل‌هایی که ضایعات را تولید می‌کنند، نشان داده شده است، این ویژگی‌ها با تقسیم تعداد ضایعات با قطر ROI برای جبران اندازه‌های مختلف تصویر، عادی می‌شوند.
۱۶) فاصله اقلیدسی مرکز ماکولا و مرکز دیسک نوری برای ارائه اطلاعات مهم در مورد وضعیت بیمار، این ویژگی نیز با قطر ROI عادی شده است.
۱۷) قطر دیسک نوری.
۱۸) نتیجه باینری AM
۱۹) برچسب کلاس (برچسب ذخیره‌سازی برای کلاس‌ها و Messidor ۱و۲ و۳)
۱ حاوی نشانه‌های DR می‌باشد.
۰ بدون نشانه‌های DR می‌باشد.

در روش پیشنهادی برای مشخص کردن ارزش و نقش هر یک از ویژگی‌ها در تشخیص بیماری از الگوریتم ژنتیک استفاده شد. برای هر یک از ویژگی‌ها، یک ژن تعریف شد. برای مقداردهی ژن‌ها، به‌طور تصادفی برای هر ژن یک عدد در بازه [۱-۰] اختصاص داده شد که نشان دهنده درجه اهمیت ویژگی متناظر با آن ژن است. هر اندازه مقدار ژن بزرگ‌تر باشد، نشان‌دهنده ارزش و اهمیت بیش‌تر ویژگی متناظر است.

به ازای هر یک از کروموزوم‌های موجود در جمعیت، مراحل زیر تکرار می‌شود:

۱-مجموعه داده‌ها فراخوانی می‌شود.

۲- یک رکورد از مجموعه داده انتخاب می‌شود.

۳- مقادیر ویژگی‌های رکورد نرمال می‌شود.

۴- مقدار ژن مربوط به هر ویژگی از کروموزوم استخراج می‌شود.

۵- به‌طور متناظر هر یک از ویژگی‌ها در ژن مخصوص به خود ضرب شده تا مقدار ویژگی‌ها وزن‌دار شوند.

۶- مراحل یک تا پنج برای همه رکوردها تکرار می‌شود.

۷- بر اساس مقادیر وزن‌دار ویژگی‌ها، عمل کلاسه‌بندی انجام می‌شود.

تابع برازندگی نشان دهنده شایستگی یا توانایی هر کروموزوم است. با فرض متغیرهای تعریف شده در جدول ۱، تابع برازندگی در شکل ۲ نمایش داده شد. عمل کلاسه‌بندی با مقادیر ویژگی‌های وزن‌دار، که بر اساس هریک از کروموزوم‌ها محاسبه می‌گردد، به روش نزدیک‌ترین همسایه [۱۳] انجام شد. هر کروموزوم که کلاسه‌بندی را با خطای پیش‌بینی کم‌تری انجام دهد، شایسته‌تر خواهد بود. تخمین خطا بر مبنای نمونه‌گیری به روش ۱۰-Fold Cross Validation [۱۴] انجام شد.

ALGORITHM

1) Read the training data from a file
2) Read the testing data from a file
3) Normalize the attribute values in the range of 0 to 1.
4) Let $x_1, x_2 \ldots x_m$ denote the m instances from data set $\{f_1, f_2, \ldots, f_n\}$, n = Number of features
5) Let Ch denote the current chromosome from population $\{g_1, g_2, \ldots, g_r\}$, r = Number of genes
6) Assign weight Ch to each instance x_i in the training set
7) Train the weights on the whole training data set

For every training instance
- Calculate the weighted value as
- $Ch_j * x_{ij}$, where j is the attribute
- Find the K nearest neighbors based on the Euclidean distance
- Calculate the class value

End for

8) For each testing instance in the testing data set
- Find the K nearest neighbors in the training data set based on the Euclidean distance
- Predict the class value by finding the maximum class represented in the K nearest neighbors

End for

9) Calculate the error rate as

Error Rate =
*1- (# of correctly classified examples / All) * 100*

10) Fitness Function=Minimize (*Error Rate*)

شکل ۲: تابع برازندگی

عملگر انتخاب، تعدادی کروموزوم را برای تولیدمثل از جمعیت انتخاب می‌کند. در این مدل از روش انتخاب نخبه‌گرا استفاده شده است. در این روش بهترین کروموزوم‌ها در هر نسل برای تولیدمثل انتخاب می‌شوند. انتخاب نخبگان به میزان قابل‌ملاحظه‌ای کارایی الگوریتم ژنتیک را افزایش می‌دهد [۱۵]. در عمل تقاطع به صورت تصادفی بخش‌هایی از

کروموزوم‌ها با یکدیگر ترکیب شده و فرزند جدید متولد می‌شود. این موضوع باعث ارثبری فرزندان از خصوصیات والدین خود می‌شود. در شکل ۳ نحوه تولید فرزند جدید نمایش داده شد. عملگر جهش پس از اتمام عمل تقاطع بر روی فرزندان جدید اعمال می‌شود. این عملگر یک ژن از یک کروموزوم را به‌طور تصادفی انتخاب نموده و سپس محتوای آن ژن را تغییر می‌دهد. شکل ۴ نشان‌دهنده نحوه جهش می‌باشد.

۱- بردار r در ابعاد (۳۴×۱) به‌طور تصادفی در بازه [۱-۰] مقداردهی می‌گردد.
۲- به صورت تصادفی دو کروموزوم Ch_1 و Ch_2 از جمعیت انتخاب می‌شوند.
۳- فرزند جدید اول = $r * Ch_1 + (1-r) * Ch_2$
۴- فرزند جدید دوم = $(1-r) * Ch_1 + r * Ch_2$

شکل ۳: نحوه تولید فرزند جدید

پس از اتمام عمل جهش، کروموزوم‌های تولید شده به عنوان نسل جدید در نظر گرفته می‌شوند. پس از گذشت چندین نسل، مقادیر کروموزوم‌ها همگرا شده و پاسخ نهایی به‌دست می‌آید.

۱- بردار r در ابعاد یک کروموزوم (۳۴×۱) تعریف می‌شود.
۲- بردار r با توزیع نرمال در بازه [۱,۱-] مقداردهی می‌گردد.
۳- به‌صورت تصادفی یک کروموزوم از جمعیت انتخاب می‌شوند.
۴- کروموزوم جهش‌یافته از جمع کروموزوم انتخاب شده با r ایجاد می‌شود.

شکل ٤: نحوه اعمال عملگر جهش

مدل پیشنهادی با سه روش ماشین بردار پشتیبان [۱۶]، شبکه عصبی پرسپترون چند لایه [۱۷] و نایو بیز [۱۸] مقایسه شده است. ارتباط بین کلاس‌های واقعی و کلاس‌های پیش‌بینی شده با استفاده از ماتریس Confusion قابل محاسبه است. در شکل ۵، پارامترهای موردنیاز ماتریس Confusion ذکر شد.

TP: تعداد رکوردهایی که به‌درستی، مثبت تشخیص داده می‌شوند.
TN: تعداد رکوردهایی که به‌درستی، منفی تشخیص داده می‌شوند.
FP: تعداد رکوردهایی که به غلط، مثبت تشخیص داده می‌شوند.
FN: تعداد رکوردهایی که به غلط، منفی تشخیص داده می‌شوند.

شکل ۵: پارامترهای موردنیاز برای ارتباط بین کلاس‌های واقعی و کلاس‌های پیش‌بینی شده

برای مقایسه مدل پیشنهادی با سایر روش‌ها از معیارهای Accuracy، Sensitivity، Specificity، Precision و F-Measure با توجه به شکل ۵ طبق روابط زیر استفاده شد:

$$Accuracy = (TP + TN) / All \quad (۱)$$

$$Sensitivity = TP / (TP + FN) \quad (۲)$$

$$Specificity = TN / (FP + TN) \quad (۳)$$

$$Precision = TP / (TP + FP) \quad (۴)$$

$$Recall = TP / (TP + FN) \quad (۵)$$

$$F\ Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (۶)$$

نتایج

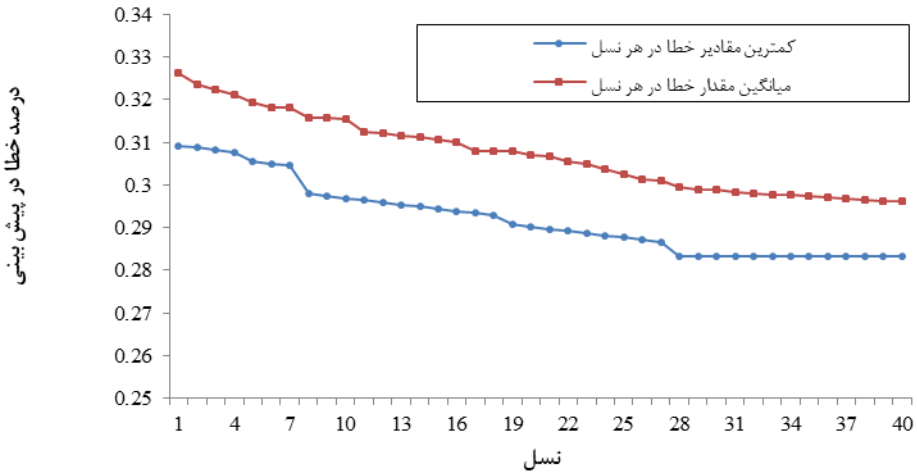
پیش‌بینی و تشخیص صحیح دیابت با استفاده از هوش مصنوعی و یادگیری ماشین، شانس درمان موفق را بالا می‌برد. در این پژوهش برای پیش‌بینی و تشخیص دیابت، از الگوریتم ژنتیک برای بهینه‌سازی نتایج الگوریتم نزدیک‌ترین همسایه استفاده شد و یک مدل جدید ارائه گردید. در روش پیشنهادی با استفاده از کاهش تعداد متغیرها و وزن‌دار کردن ویژگی‌ها با به‌کارگیری الگوریتم ژنتیک برای افزایش دقت با هدف طراحی و ارزیابی یک مدل پزشک‌یار در تعیین درجه بدخیمی دیابت انجام شد. در جدول ۳ نحوه وزن‌دار کردن ویژگی‌ها بر اساس ژن متناظر نشان داده شد.

نتایج شبیه‌سازی نشان می‌دهد که مدل پیشنهادی با دقت پیش‌بینی ۰/۷۶ از روش‌های نایو بیز، شبکه عصبی MLP (Multi Layer Perceptron) و ماشین بردار پشتیبان دقت بیش‌تری دارد. دقت بالا در تشخیص دیابت حاکی از برتری رهیافت پیشنهادی است.

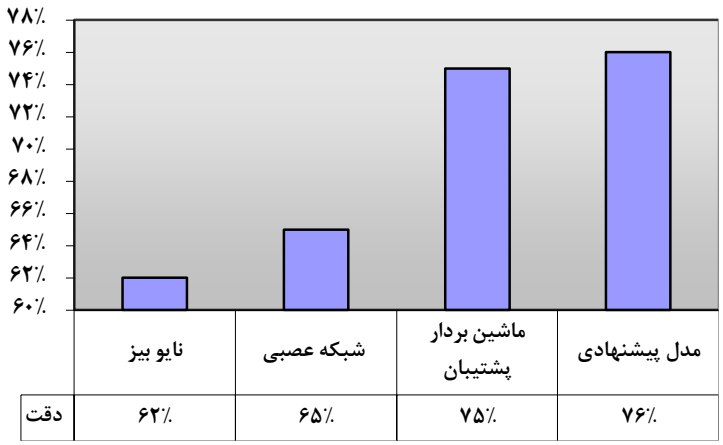
جدول ۳: وزن‌دار کردن ویژگی‌ها با استفاده از الگوریتم ژنتیک

ویژگی‌های وزن دار	مقدار اولیه	مقدار نرمال	وزن	مقدار وزن‌دار
X_3	۵۵	۰/۷۱	۰/۳	۰/۲۱
X_4	۳	۰/۳۳	۰/۲	۰/۰۷
...	...	...	...	...

شکل ۶ نمایش دهنده درصد خطا برای پیش‌بینی بیماری در الگوریتم پیشنهادی در ۴۰ نسل است. شکل ۷ نشان‌دهنده نمودار نتایج تشخیص روش‌های مختلف با معیار Accuracy است. همان‌گونه که مشاهده شد مدل پیشنهادی دقت بیش‌تری نسبت به سایر روش‌ها دارد. همچنین مقایسه نتایج پیش‌بینی بیماری با معیارهای Sensitivity، Specificity، Precision و F-Measure در جدول ۴ نمایش داده شد. نتایج مقایسه نشان دهنده برتری عملکرد مدل پیشنهادی می‌باشد. مقادیر جدول نشان دهنده عملکرد بهتر روش پیشنهادی است.



شکل ۶: درصد خطا برای پیش‌بینی بیماری در الگوریتم پیشنهادی



شکل ۷: نمودار نتایج با معیار Accuracy

جدول ٤: مقایسه نتایج با معیارهای Sensitivity، Specificity، Precision و F-Measure

معیار مقایسه	کلاس مثبت	روش نایو بیز	روش شبکه عصبی	روش ماشین بردار پشتیبان	مدل پیشنهادی
Sensitivity	دارای علائم DR	۰/۴۵۲	۰/۷۶۸	۰/۸۳۷	۰/۸۷۷
	بدون علائم DR	۰/۷۶۴	۰/۵۵۸	۰/۶۶۹	۰/۶۴۴
Specificity	دارای علائم DR	۰/۷۶۴	۰/۵۵۸	۰/۶۶۹	۰/۶۴۴
	بدون علائم DR	۰/۴۵۲	۰/۷۶۷	۰/۸۳۷	۰/۸۷۷
Precision	دارای علائم DR	۰/۶۲۹	۰/۶۰۵	۰/۶۹۱	۰/۷۰۴
	بدون علائم DR	۰/۶۱۲	۰/۷۳۰	۰/۸۲۳	۰/۸۸۴
F-Measure	دارای علائم DR	۰/۵۲۶	۰/۶۷۶	۰/۷۵۷	۰/۷۸۱
	بدون علائم DR	۰/۶۸۰	۰/۶۳۳	۰/۷۳۸	۰/۷۳۱

بحث و نتیجه‌گیری

در این مطالعه، مجموعه‌ای از الگوهای مؤثر در بروز دیابت بر اساس پایگاه داده استاندارد UCI مورد پردازش قرار گرفت. در این پژوهش تعداد قابل قبولی از بیماران با متغیرهای مناسب با حداقل داده‌های از دست رفته به کار گرفته شد. مطالعه حاضر با به‌کارگیری ویژگی‌های بیماران دیابتی با استفاده از الگوریتم ژنتیک و نزدیک‌ترین همسایه، دیابت را پیش‌بینی کرد. مدل پیشنهادی این مطالعه شامل نرمال‌سازی ویژگی‌ها، انتخاب ویژگی‌ها، وزن‌دهی به ویژگی‌ها و معرفی روشی برای پیش‌بینی دیابت با استفاده از ویژگی‌های وزن‌دار است. در این مطالعه علاوه بر شناسایی مهم‌ترین ویژگی‌ها، با استفاده از انتخاب ویژگی‌ها بر اساس الگوریتم ژنتیک به عملکردی بهتر بر حسب شاخص‌های دقت، حساسیت و ویژگی به دست آمد.

این که هر یک از ویژگی‌ها دارای چه ارزشی است و چقدر در تشخیص بیماری نقش دارد، مسئله مهمی است، در این مطالعه روشی ارائه شد که ارزش و نقش هر یک از ویژگی‌ها به‌طور دقیق مشخص و بیماری دیابت پیش‌بینی و تشخیص داده شد، در روش پیشنهادی برای مشخص کردن ارزش و نقش هر یک از ویژگی‌ها در تشخیص بیماری از الگوریتم ژنتیک استفاده شد.

در مطالعه‌ای پیش‌بینی بیماری دیابت نوع دوم با ماشین بردار پشتیبان انجام شد [۱۹] در این پژوهش ۷۶۸ بیمار با هشت متغیر از قبیل فشارخون افراد و میزان انسولین تزریقی برای تشخیص دیابت استفاده شد. تعداد نمونه‌های قابل‌قبول برای داده‌کاوی از مزایای این پژوهش بود؛ اما تعداد کم متغیرها برای مدل‌سازی از معایب این روش است. در مطالعه حاضر تعداد قابل قبولی از بیماران با تعداد متغیرهای مناسب با حداقل داده‌های از دست رفته به کار گرفته شد که مزیت مدل

پیشنهادی این مطالعه نسبت به مدل پیش‌بینی بیماری دیابت نوع دوم با ماشین بردار پشتیبان است.

Barakat و همکاران [۲۰] با بهبود و بهینه‌سازی روش ماشین بردار پشتیبان در تشخیص بیماری دیابت نوع دو رسیدند. مجموعه داده مورد استفاده شامل ۴۶۸۲ نفر مراجعه کننده با ویژگی‌های جنسیت، BMI، فشارخون، کلسترول و قندخون بود. تعداد قابل‌ملاحظه نمونه‌ها در مدل‌سازی از مزایای این پژوهش بود؛ اما عیب اصلی این روش نیز تعداد کم متغیرها برای مدل‌سازی است، که مدل پیشنهادی این مطالعه با تعداد متغیرهای مناسب عملکردی بهتری برای پیش‌بینی دیابت دارد.

Worachartcheewan و همکاران [۲۱] روشی برای تشخیص بیماری با استفاده از روش ماشین بردارش پشتیبان و الگوریتم ژنتیک چند منظوره ارائه دادند. پایگاه داده UCI برای شبیه‌سازی مورد استفاده قرار گرفته است. استفاده از کمینه‌سازی ریسک ساختاری در ماشین بردار پشتیبان به جای کمینه کردن تجربی از خصوصیات ماشین بردار پشتیبان است. در روش SVM برای حل مسائل رگرسیون، ابتدا تابع رگرسیون با به‌کارگیری یک دسته تابع خطی توسط ماشین بردار پشتیبان تخمین زده می‌شود. سپس عملیات رگرسیون توسط تابعی که انحراف از مقدار واقعی در آن کمتر از تابع تلفات تعیین شده است، انجام می‌شود. در نهایت بهترین جواب با کمینه‌سازی ریسک ساختاری توسط ماشین بردار پشتیبان ارائه می‌گردد. در این مطالعه از الگوریتم ژنتیک مبتنی بر رتبه‌بندی نامغلوب (Nondominated Sorting Genetic Algorithm) برای حل مسائل بهینه‌سازی استفاده شد. ایجاد جمعیت اولیه، محاسبه توابع هدف معیارهای برازندگی، مرتب‌سازی جمعیت طبق شروط غلبه، برآورد فاصله ازدحامی، انتخاب، تقاطع، جهش، تکرار مراحل تا رسیدن به شرط خاتمه از مراحل کار این الگوریتم است. در این پژوهش سه پارامتر C ،γ و اپسیلون در نتیجه پیش‌بینی بیماری تأثیر دارند. مقدار کوچک برای پارامتر C منجر به کاهش نرخ دقت دسته‌بندی می‌شود. اگر مقدار C بزرگ باشد، دقت دسته‌بندی در مرحله آموزش افزایش می‌یابد؛ ولی کاهش دقت در مرحله آزمایش را به دنبال خواهد داشت. در ابتدا جمعیت اولیه برای پارامترهای پارامتر MP، C،γ و اپسیلون تشکیل گردید. سپس بهینه‌سازی انجام شد. مرحله بهینه‌سازی شامل فاز آموزش و فاز آزمایش است. پس از یافتن مقادیر بهینه برای پارامترها توسط الگوریتم NSGA عمل کلاس‌بندی انجام گرفت و نتایج نهایی به‌دست آمد. عیب این روش این است که تأثیر پارامتر γ در خروجی دسته‌بندی نسبت به C بیش‌تر است، مقدار بزرگ برای پارامتر γ باعث برازش بیش‌ازحد و مقدار کوچک کاهش دقت را در پی دارد. مدل پیشنهادی این مطالعه سعی کرد به جای استفاده از سه پارامتر C،γ و اپسیلون و کلاس‌بندی، از تکنیک نرمال‌سازی ویژگی‌ها، انتخاب ویژگی‌ها و وزن‌دهی به ویژگی‌ها استفاده نماید که باعث شده علاوه بر شناسایی مهم‌ترین ویژگی‌ها، با استفاده از انتخاب ویژگی‌ها بر اساس الگوریتم ژنتیک به عملکردی بهتری بر حسب شاخص‌های دقت، حساسیت و ویژگی به دست آید.

Rudziński [۲۲] از بهینه‌سازی مبتنی بر قواعد فازی جهت کلاس‌بندی استفاده کرده است و از این روش برای طبقه‌بندی انواع پایگاه‌های داده استفاده کردند. دقت بالا و میزان خطای پایین از مزایای این روش بود و پیچیدگی بالای زمانی از معایب این روش است. چون این مطالعه سعی در طراحی و ارزیابی یک مدل پزشک‌یار در تعیین درجه بدخیمی دیابت را دارد، پیچیدگی زمانی پایین از اهداف این روش می‌باشد.

Richards و همکاران [۲۳] مشاهدات مربوط به مرگ زودرس بیماران دیابتی را مورد مطالعه قرار دادند. مجموعه داده‌های مربوط به ۲۱۰۰۰ بیمار موردمطالعه قرار گرفت. از نتایج این مطالعه، استفاده از روش قواعد انجمنی برای کشف قواعدی است که به‌طورمعمول توسط پزشک‌ها نادیده گرفته می‌شد؛ اما در این مطالعه سعی شد علاوه بر شناسایی مهم‌ترین ویژگی‌ها، با استفاده از انتخاب ویژگی‌ها بر اساس الگوریتم ژنتیک به عملکردی بهتر بر حسب شاخص‌های دقت، حساسیت و ویژگی دست یابد.

در پژوهشی از روش درخت تصمیم استفاده شد و نتیجه «مهم‌ترین متغیر برای کنترل قند خون، سن افراد می‌باشد» حاصل شد. این نتیجه در تصمیم‌گیری پزشکان برای تفسیر نتایج آزمایشگاهی مؤثر بود [۲۴]. Huang و همکاران [۲۵] ویژگی‌های تأثیرگذار بر روی کنترل قندخون را با استفاده از روش‌های داده‌کاوی مورد بررسی قرار دادند. در این پژوهش از درخت تصمیم و بیز ساده به منظور تحلیل داده‌کاوی استفاده کردند و مشخص شد که متغیرهای سن، مدت زمان تشخیص بیماری، نیاز به درمان انسولین، گلوکز خون و رژیم غذایی مهم‌ترین فاکتورهای اثرگذار بر روی کنترل قندخون هستند. Huang و همکاران [۲۵] برای شناسایی عوامل عمده تأثیرگذار بر کنترل دیابت، از روش‌های هوش جمعی برای انتخاب ویژگی استفاده کردند. پیچیدگی بالای زمانی از معایب

این روش بود. چون این مطالعه سعی در طراحی و ارزیابی یک مدل پزشک‌یار در تعیین درجه بدخیمی دیابت را دارد، پیچیدگی زمانی پایین از اهداف این روش می‌باشد.

Lee و همکاران [۲۶] با مطالعه بر روی ۸۸ بیمار، فاکتورهای پیش‌بینی کننده بهبودی دیابت بعد از عمل جراحی چاقی (عمل روی معده) را یافتند. نتایج تحقیق حاکی از آن بود که آلانین آمینوترانسفراز ، کلسترول و فشارخون مهم‌ترین متغیرهایی هستند که بر روی بهبود بیماران دیابتی بعد از عمل جراحی چاقی تأثیر می‌گذارند. مدل پیشنهادی مطالعه حاضر با دقت ۷۶٪ سعی کرده نه یک ویژگی مهم بلکه با شناسایی مهم‌ترین ویژگی‌ها یک مدل پزشک‌یار مطمئن در تعیین درجه بدخیمی دیابت باشد.

Han و همکاران [۲۷] با تحلیل و بررسی اطلاعات بیماران موجود در پایگاه داده یک بیمارستان و به‌کارگیری الگوریتم درخت تصمیم، وجود دیابت را پیش‌بینی کردند. مطالعه حاضر با دقت ۷۶٪ با به‌کارگیری ویژگی‌های بیماران دیابتی با استفاده الگوریتم ژنتیک و نزدیک‌ترین همسایه، دیابت را پیش‌بینی نمود.

مدل پیشنهادی با دقت ۷۶٪ دارای عملکرد بهتری نسبت به روش نایو بیز، درخت تصمیم و نزدیک‌ترین همسایه در تشخیص دیابت است. در روش پیشنهادی با استفاده از کاهش تعداد متغیرها و وزن‌دار کردن ویژگی‌ها با به‌کارگیری الگوریتم ژنتیک برای افزایش دقت با هدف طراحی و ارزیابی یک مدل پزشک‌یار در تعیین درجه بدخیمی دیابت انجام شد. مدل پزشک‌یار طراحی شده در این پژوهش در تشخیص درجه بدخیمی موفق بوده است و دسته‌بندی با دقت قابل‌قبولی انجام گردید. آزمایش‌ها و شبیه‌سازی نشان داد سیستم پزشک‌یار معرفی شده در این پژوهش بر روی مجموعه داده بیماران بومی مبتلا به دیابت بیمارستان مرتاض یزد به دقت ۹۷/۳٪ رسیده است که بالاتر از تحقیقات مشابه بر روی مجموعه داده‌های متفاوت بوده است؛ اما پیچیدگی و زمان‌بر بودن مدت اجرا از نقاط ضعف این روش بود که در تحقیقات آینده سعی در برطرف کردن این ضعف خواهد شد.

این سیستم می‌تواند به عنوان یک سیستم پزشک‌یار در مطب‌های پزشکی بسیار کاربردی باشد.

نتایج شبیه‌سازی نشان داد که مدل پیشنهادی با دقت پیش‌بینی ۰/۷۶ از روش‌های نایو بیز، شبکه عصبی پرسپترون چند لایه و ماشین بردار پشتیبان دقت بیشتری دارد. دقت بالا در تشخیص دیابت حاکی از برتری رهیافت پیشنهادی است.

تعارض منافع

بین وظایف پزشک در نحوه تشخیص وجود دیابت در بیمار مبتلا به قند و منافع شخصی پزشک تعارض به وجود می‌آید. مهم‌ترین نگرانی موجود، در اولویت قرار گرفتن تسریع پیش‌بینی وجود دیابت به جای ارائه خدمات مبتنی بر دانش تخصصی از سوی پزشکان است که می‌تواند به تنزل اعتماد بیمار به آنان منجر شود. تعارض منافع ممکن است به اشکال مختلف رخ دهد که مهم‌ترین آن صرف زمان کم برای پیش‌بینی وجود دیابت است. اگر از همه علائم و ویژگی‌ها در تشخیص دیابت استفاده شود، خطای تشخیص کاهش می‌یابد.

References

1. Kitabchi AE, Umpierrez GE, Miles JM, Fisher JN. Hyperglycemic crises in adult patients with diabetes. Diabetes Care 2009;32(7):1335-43.

2. Gardner DG, Shoback D. Greenspan's Basic & Clinical Endocrinology. 9th ed. NewYork: McGraw-Hill Medical; 2011.

3. Gastaldelli A, Gaggini M, DeFronzo RA. Role of adipose tissue insulin resistance in the natural history of type 2 diabetes: results from the san antonio metabolism study. Diabetes 2017;66(4):815-22.

4. Volkov P, Bacos K, Ofori JK, Esguerra JL, Eliasson L. Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. Diabetes 2017;66(4):1074-85.

5. Su CT, Yang CH, Hsu KH, Chiu WK. Data mining for the diagnosis of type II diabetes from three-dimensional body surface anthropometrical scanning data. Computers & Mathematics with Applications 2006;51(6):1075-92.

6. Santhanam T, Padmavathi MS. Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis. Procedia Computer Science 2015;47:76-83.

7. Chakraborty G, Chakraborty B. Multi-objective Optimization Using Pareto GA for Gene-Selection from Microarray Data for Disease Classification. IEEE International Conference on Systems, Man, and Cybernetics; 2013 Oct 13-16; Manchester, UK: IEEE; 2013. p. 2629-34.

8. Maheta HK, Dabhi VK. Classification of imbalanced data sets using Multi Objective Genetic International

Conference on Computer Communication and Informatics (ICCCI); 2015 Jan 8-10; Coimbatore, India: IEEE; 2015. p. 1-6.
9. Holland JH. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT press; 1992.
10. Antal B, Hajdu A. An ensemble-based system for automatic screening of diabetic retinopathy. Knowledge-Based Systems 2014;60:20-7.
11. Han J, Kamber, M, Pei J. Data Mining: Concepts and Techniques. 3th ed. USA: Morgan Kaufmann; 2011.
12. Goldberg DE, Holland JH. Genetic Algorithms and Machine Learning. Machine Learning 1988;3(2):95-9.
13. Alpaydin E. Voting over multiple condensed nearest neighbors. Artificial Intelligence Review 1997;11(1):115-32.
14. Kohavi R. A Study of Cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th international joint conference on Artificial intelligence; 1995 Aug 20-25; Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc; p.1137–43.
15. De Jong KA. Analysis of the behavior of a class of genetic adaptive systems [dissertation]. Michigan: University of Michigan Ann Arbor, MI; 1975.
16. Cortes C, Vapnik V. Support-vector networks. Machine Learning1995;20(3):273-97.
17. Rosenblatt F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books; 1962.
18. Rennie JD, Shih L, Teevan J, Karger DR. Tackling the poor assumptions of naive bayes text classifiers. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning; 2003 Aug 21 – 24; Washington, DC, USA: AAAI Press; 2003. p. 616-23.
19. Purnami SW, Embong A, Zain JM, Rahayu SP. A new smooth support vector machine and its applications in diabetes disease diagnosis. Journal of Computer Science 2009;5(12):1003-8.
20. Barakat N, Bradley AP, Barakat MH. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Transactions on Information Technology in Biomedicine 2010; 14(4): 1114-20.
21. Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Pidetcha P, Prachayasittikul V. Identification of metabolic syndrome using decision tree analysis. Diabetes Res Clin Pract 2010;90(1):e15-8.
22. Rudziński F. A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers. Applied Soft Computing 2016;38:118-33.
23. Richards G, Rayward-Smith VJ, Sonksen PH, Carey S, Weng C. Data mining for indicators of early mortality in a database of clinical records. Artif Intell Med 2001;22(3):215-31.
24. Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. Artif Intell Med 2002;26(1-2):37-54.
25. Huang Y, McCullagh P, Black N, Harper R. Feature selection and classification model construction on type 2 diabetic patients' data. Artif Intell Med 2007;41(3):251-62.
26. Lee WJ, Chong K, Chen JC, Ser KH, Lee YC, Tsou JJ, et al. Predictors of diabetes remission after bariatric surgery in Asia. Asian J Surg 2012;35(2):67-73.
27. Han J, Rodriguez JC, Beheshti M. Diabetes data analysis and prediction model discovery using rapidminer. Second International Conference on Future Generation Communication and Networking; 2008 Dec 13-15; Hainan Island, China: IEEE; 2008. p. 96-9.

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2019; 6(1): 12-23

Diabetes Prediction by Optimizing the Nearest Neighbor Algorithm Using Genetic Algorithm

Momeny Mohammad[1]*, Latif Ali Mohammad[2], Agha Sarram Mehdi [3], Hajmirzazade Kazem[4], Gharravi Sorayya[5], NaghiboAlghara Seyed Mahammad[6]

• **Received:** 16 May, 2018 • **Accepted:** 29 Sep, 2018

Introduction: Diabetes or diabetes mellitus is a metabolic disorder in body when the body does not produce insulin, and produced insulin cannot function normally. The presence of various signs and symptoms of this disease makes it difficult for doctors to diagnose. Data mining allows analysis of patients' clinical data for medical decision making. The aim of this study was to provide a model for increasing the accuracy of diabetes prediction.

Method: In this study, the medical records of 1151 patients with diabetes were studied, with 19 features. Patients' information were collected from the UCI standard database. Each patient has been followed for at least one year. Genetic Algorithm (GA) and the nearest neighbor algorithm were used to provide diabetes prediction model.

Results: It was revealed that the prediction accuracy of the proposed model equals 0.76. Also, for the methods of Naïve Bayes, Multi-layer perceptron (MLP) neural network, and support vector machine (SVM), the prediction accuracy was 0.62, 0.65, and 0.75, respectively.

Conclusion: In predicting diabetes, the proposed model has the lowest error rate and the highest accuracy compared to the other models. Naïve Bayes method has the highest error rate and the lowest accuracy.

Keywords: Prediction of diabetes, Genetic algorithm, Nearest neighbor algorithm, Data mining

• **Citation:** Momeny M, Latif AM, Sarram R, Kazem Hajmirzazade K, Gharravi S, NaghiboAlghara SM. Diabetes Prediction by Optimizing the Nearest Neighbor Algorithm Using Genetic Algorithm. Journal of Health and Biomedical Informatics 2019; 6(1): 12-23. [In Persian]

1. Ph.D. Student, Electrical and Computer Engineering Dept., School of Electrical and Computer Engineering, Yazd University, Yazd, Iran
2. Assistant Professor, Electrical and Computer Engineering Dept., Yazd University, Yazd, Iran
3. Associate Professor of Computer Sciences, Electrical and Computer Engineering Dept., Yazd University, Yazd, Iran
4. Assistant Professor of Diseases and Diseases, Faculty of Medicine, Islamic Azad University, Yazd. Iran.
5. M.Sc. in Computer engineering (Software), Lecturer, Electrical and Computer Engineering Department, Computer Dept., Integrated Higher Education of Esfarayen, North Khorasan, Esfarayen, Iran
6. General Medicine, Faculty of Medicine, Islamic Azad University, Yazd, Iran

*Correspondence: Faculty of Electrical and Computer Engineering, Computer Department, Yazd University, Yazd, Iran.
• **Tel:** 09159736003 • **Email:** mohamad.momeny@stu.yazd.ac.ir