DOI: 10.48305/jims.v44.i851.0137

مجله دانشکده پزشکی اصفهان — سال چهل و چهارم/شماره‌ی ۸۵۱/هفته‌ی سوم فروردین ۱۴۰۵

تاریخ دریافت: ۱۴۰۳/۱۲/۱۹ — تاریخ پذیرش: ۱۴۰۴/۱۲/۲۶ — تاریخ چاپ: ۱۴۰۵/۰۱/۲۳

پیش‌بینی دیابت با استفاده از الگوریتم Adaboost در مردان فعال

حمید آقاعلی‌نژاد[۱]، لیلا فصیحی[۲]، محمدحسین قهرمانی[۳]، موژان آقاعلی‌نژاد[۴]

مقاله پژوهشی

چکیده

مقدمه: دیابت، بر بیماری‌های قلبی- عروقی، از دست دادن بینایی و بیماری‌های کلیوی تأثیر می‌گذارد. تکنیک‌های داده‌کاوی به پیش‌بینی زودهنگام و درمان دقیق بیماری پزشکان را کمک می‌کند. با توجه به نقش فعالیت بدنی در پیشگیری از دیابت، این مطالعه به بررسی تأثیر آن در کنار سایر عوامل خطر پرداخت. بنابراین، هدف اصلی این مطالعه، پیش‌بینی دیابت با استفاده از الگوریتم Adaboost در مردان فعال بود.

روش‌ها: در این مطالعه‌ی توسعه‌ای- کاربردی، پرونده‌ی ۵۰۰ بیمار مرد در محدوده‌ی سنی ۲۰-۸۰ سال و با سابقه‌ی فعالیت بدنی منظم (حداقل ۳ جلسه ۹۰ دقیقه‌ای در هفته) که در ۱۰ سال گذشته در بیمارستان‌های میلاد و آیت الله کاشانی تهران مراجعه کرده بودند به عنوان نمونه انتخاب شدند. تعداد۲۰ متغیر آنتروپومتری، ژنتیکی، سابقه‌ی خانوادگی، عوامل محیطی و فیزیولوژیکی به عنوان ویژگی‌های ورودی الگوریتم انتخاب شدند. برای ارزیابی عملکرد الگوریتم پیشنهادی از معیارهای، حساسیت، صحت و دقت استفاده شد. از نرم‌افزار MATLAB (۲۰۲۴) برای تجزیه و تحلیل داده‌ها استفاده شد.

یافته‌ها: نتایج نشان داد که الگوریتم Adaboost با دقت ۷۳/۱ درصد و صحت ۷۵/۳ درصد توانست بیماری دیابت را در مردان فعال پیش‌بینی کند.

نتیجه‌گیری: این مدل می‌تواند به عنوان ابزار کمکی در غربالگری اولیه دیابت در مراکز درمانی مورد استفاده قرار گیرد.

واژگان کلیدی: دیابت؛ Adaboost؛ مردان؛ فعال؛ پیش‌بینی

ارجاع: آقاعلی‌نژاد حمید، فصیحی لیلا، قهرمانی محمدحسین، آقاعلی‌نژاد موژان. **پیش‌بینی دیابت با استفاده از الگوریتم Adaboost در مردان فعال.** مجله دانشکده پزشکی اصفهان ۱۴۰۵؛ ۴۴ (۸۵۱): ۱۳۷- ۱۴۲.

مقدمه

مداخلات سبک زندگی و/ یا داروها معمولاً برای درمان دیابت نوع ۲ تجویز می‌شوند (۱). در طول هر نوع فعالیت بدنی، جذب گلوکز به ماهیچه‌های اسکلتی فعال از طریق مسیرهای مستقل از انسولین افزایش می‌یابد (۲). سطح گلوکز خون با افزایش تولید گلوکز کبدی ناشی از تنظیم کننده گلوکز و بسیج اسیدهای چرب آزاد حفظ می‌شود (۳)، که ممکن است با مقاومت به انسولین یا دیابت مختل شود (٤). بهبود حساسیت سیستمیک و احتمالاً کبدی به انسولین به دنبال هر فعالیت بدنی می‌تواند از ۲ تا ۷۲ ساعت طول بکشد (۵, ٦). علاوه بر این، فعالیت بدنی معمولی عملکرد سلول‌های بتا (٦)، حساسیت به انسولین (۷)، عملکرد عروقی (۸) و میکروبیوتای روده (۹) را افزایش می‌دهد، که همگی ممکن است منجر به مدیریت دیلبت و همچنین کاهش خطر بیماری شوند.

دیابت نوع ۲ که زمانی دیابت غیر وابسته به انسولین یا دیابت بزرگسالان نامیده می‌شد، شایع‌ترین شکل دیابت است که ۹۰ تا ۹۵ درصد از ۱۳ میلیون مرد مبتلا به دیلبت را تحت تأثیر قرار می‌دهد (۱). نرخ دیابت در ایران به طور چشم‌گیری افزایش یافته است (۲). خطر ابتلا به دیابت نوع ۲ معمولاً با افزایش سن افزایش می‌یابد (۳). انجمن دیابت آمریکا توصیه می‌کند که افرادی که سایر عوامل خطر برای این بیماری را ندارند باید آزمایش را بعد از ٤۵ سالگی شروع کنند (٤). برخلاف افراد مبتلا به دیابت نوع ۱، افراد مبتلا به دیابت نوع ۲ انسولین تولید می‌کنند، اما کافی نیست (۵). یا بدن آن‌ها

۱- استاد فیزیولوژی ورزش، گروه تربیت بدنی و علوم ورزشی، دانشکده‌ی علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران.

۲- دکتری فیزیولوژی ورزشی، گروه تربیت بدنی و علوم ورزشی، دانشکده‌ی علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران.

۳- کارشناسی ارشد فیزیولوژی ورزشی، گروه تربیت بدنی و علوم ورزشی، دانشکده‌ی علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران.

۴- دانشجو کارشناسی ارشد فیزیولوژی ورزشی، گروه تربیت بدنی و علوم ورزشی، دانشکده‌ی تربیت بدنی و علوم ورزشی، دانشگاه تهران، تهران، ایران.

نویسنده‌ی مسؤول: حمید آقاعلی‌نژاد: استاد فیزیولوژی ورزش، گروه تربیت بدنی و علوم ورزشی، دانشکده‌ی علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران

Email: halinejad@modares.ac.ir

انسولین را تشخیص نمی‌دهد و آن‌طور که باید از آن استفاده نمی‌کند (٦). این مقاومت به انسولین نامیده می‌شود (٧).

داده کاوی استخراج الگوهای ناشناخته یا پنهان شده قبلی از یک پایگاه داده یا انبار داده عظیم است (١٠). سازمان‌های متعددی از داده کاوی برای تجزیه و تحلیل مجموعه داده‌های عظیم، برای بهبود فرآیند تصمیم گیری و به دست آوردن نتایج بلند مدت بهتر استفاده می‌کنند (١١). روش‌های داده کاوی و یادگیری ماشین باعث شده است متخصصان این حوزه به دنبال اجرای این عملیات بر روی داده‌های پزشکی و فرآیندهای موجود در حوزه‌ی سلامت باشند. مسائل متفاوتی که قبلاً توسط یک متخصص پزشکی انجام می‌شد، در حال حاضر می‌تواند توسط ماشین‌های هوشمند به نحوی بهتر همراه با کارایی بالاتر انجام پذیرد.

به منظور تشخیص و پیشگیری از دیابت یک مدل پیش‌بینی توسط Fiarni و همکاران برای پیش‌بینی وقوع سه عارضه اصلی دیابت در اندونزی، توسعه داده شد. از هفت عامل خطر برای پیش‌بینی این عوارض سن، جنس، شاخص توده‌ی بدنی، سابقه خانوادگی دیابت، فشار خون، مدت زمان ابتلا به دیابت و سطح گلوکز خون استفاده شد. آن‌ها از دو الگوریتم خوشه‌بندی k-means و روش‌های طبقه‌بندی درخت بیز ساده برای بررسی این مجموعه داده استفاده کردند (١٢).

Abdollahi و Nouri-Moghaddam از یک روش آموزش گروهی مبتنی بر الگوریتم ژنتیک برای شناسایی و پیش‌بینی مؤثر پیامدهای دیابت استفاده کردند. یکی از با ارزشترین دارایی‌های سازمان‌های سلامت در عصر اطلاعات، داده‌ها هستند و جمع آوری، ذخیره و تحلیل ا ین داده‌ها می‌تواند یکی از عوامل موفقیت این سازمان‌ها محسوب شود (١٣). به وسیله‌ی داده‌کاوی امکان پی‌بردن به روابط، روندها و الگوهای مخفی بین داده‌ها و در نهایت دستیابی به دانش نوین میسر خواهد شد. بنابراین، هدف از مطالعه‌ی حاضر، ارائه‌ی مدلی برای پیش‌بینی بیماری دیلبت با استفاده از الگوریتم Adaboost در مردان فعال بود.

روش‌ها

جامعه‌ی مورد پژوهش، کلیه‌ی بیماران مرد مراجعه‌کننده به بیمارستان‌های میلاد و آیت الله کاشانی تهران بود. تعداد پرونده‌های اولیه ١٢٢٣ پرونده بود. بعد از بررسی پرونده‌ها و ثبت مشخصات آزمایشگاهی و تکمیل پرسشنامه فعالیت بدنی (از طریق وسایل ارتباط جمعی از قبیل واتس آپ، تلگرام و ...)، نهایتاً از بین آن‌ها تعداد ٥٠٠ پرونده مربوط به مردان فعال در دامنه‌ی سنی ٢٠ تا ٨٠ سال که در ١٠ سال گذشته به این بیمارستان‌ها مراجعه کرده بودند به عنوان نمونه انتخاب شدند. لندازه‌ی نمونه با استفاده از نرم‌افزار G*Power با لندازه اثر ٠/١٥، توان آزمون ٠/٨٩ و مقدار خطای ٠/٠٥ تعین شد.

معیارهای ورود به مطالعه شامل: جنسیت مرد، سن بین ٢٠ تا ٨٠ سال، دارای سوابق پزشکی و آزمایشات بالینی در بیمارستان و در دسترس از طریق تلفن یا اینترنت بودند. معیارهای خروج از مطالعه شامل: خستگی غیرعادی، کم خونی، اختلال عملکرد فیزیکی ناشی از بیماری و مبتلا به بیماری‌های کلیه، کبد، پاراتیروئید و تیروئید بودند. در پژوهش حاضر، منظور از فعال، مردانی بودند که به طور منظم در یک سال منتهی به پژوهش، هفته‌ای سه جلسه و هر جلسه به مدت ٩٠ دقیقه فعالیت بدنی (پیاده‌روی سریع و حرکات کششی) داشتند (١٤).

با بررسی پرونده‌های پزشکی ابتدا ٥٧ متغیر و سپس، با استفاده از نظر دو پزشک متخصص در زمینه‌ی بیماری دیابت و بررسی نتایج مقالات علمی، نهایتاً ٢٠ متغیر آنتروپومتری، ژنتیکی، سابقه‌ی خانوادگی، عوامل محیطی و فیزیولوژیکی شامل: هماتوکریت، کلسترول کل، میزان فیلتراسیون گلومرولی کلیه، فشارخون سیستولی، فشارخون دیاستولی، آسپارتات آمینو ترنسفراز، کراتین سرم، گاما گلوتامیل ترانسفراز، آلانین آمینوترانسفراز، شاخص توده بدنی، درصد چربی بدن، HbA1c، کلسترول لیپوپروتئین با چگالی بالا، کلسترول لیپوپروتئین با چگالی پایین، کبد چرب، سابقه‌ی خانوادگی، اضافه وزن، مصرف سیگار، درمان کاهش چربی و عدم یا وجود بیماری قلبی، به عنوان ویژگی‌های ورودی الگوریتم انتخاب شدند (١٥).

این پژوهش در دو مرحله انجام گرفت، مرحله اول شامل جمع‌آوری داده‌های مربوط به بیماران از طریق بررسی پرونده‌ی بیمارستانی بود. مرحله‌ی دوم آموزش الگوریتم AdaBoost بر روی داده‌های جمع‌آوری شده بود. الگوریتم AdaBoost، با به کارگیری متغیرهای ورودی و تعیین متغیر هدف ایجاد شد. برای استفاده بهینه از داده‌ها باید آن‌ها را به شکلی تغییر داد که برای الگوریتم مناسب باشد (١٦). برای پرسش‌هایی که پاسخ بله و خیر داشتند از عدد صفر و یک استفاده شد. عدد یک به معنای پاسخ بله و عدد صفر به معنای پاسخ خیر در نظر گرفته شد. مجموعه داده‌ها در قالب اکسل به نرم‌افزار (MATLAB) نسخه‌ی ٢٠٢٤ انتقال داده شد و مورد تجزیه و تحلیل قرار گرفت.

الگوریتم AdaBoost

الگوریتم AdaBoost، یک روش یادگیری جمعی است و معروف‌ترین الگوریتم از خانواده الگوریتم‌های Boosting است که توسط Freund و Schapire ارائه شده است (١٧). در الگوریتم‌های یادگیری جمعی، یک نمونه توسط چندین کلاسه بند مختلف کلاسه‌بندی می‌شود و نتایج کلاسه‌بندی‌ها به شکل هوشمندانه‌ای با یکدیگر ترکیب شده و نتیجه نهایی برای آن نمونه خاص تعیین می‌گردد. اغلب، کارآیی الگوریتم یادگیری جمعی در مقایسه با تک تک کلاسه بندهای شرکت‌کننده در ساختار آن بالاتری کسب می‌نمایند. در الگوریتم یادگیری جمعی، هر

کلاسه بند، با یک زیر مجموعه تصادفی و منتخب از کل نمونه‌ها، آموزش داده می‌شود. با شکل گرفتن چندین کلاسه بند متفاوت، کلاسه بند نهایی که نتیجه نگاه جمعی است دارای کارآیی بالاتری خواهد بود (۱۸).

معیار ارزیابی

در این پژوهش جهت ارزیابی روش پیشنهادی از دو معیار دقت، صحت بهره گرفته شده است. روابط مرتبط با این معیارها در جدول ۱ آمده است.

جدول ۱. روش ارزیابی‌ها

کلاس پیش‌بینی‌شده	
منفی	مثبت
منفی کاذب(FN)	مثبت واقعی(TP)
منفی واقعی(TN)	مثبت کاذب(FP)
دقت	صحت
$\frac{TP+TN}{(TP+FP+TN+FN)}$	$\frac{TP}{(TP+FP)}$

یافته‌ها

در مجموع ۵۰۰ پرونده حاوی ۲۰ ویژگی آنتروپومتری، ژنتیکی، سابقه‌ی خانوادگی، عوامل محیطی و فیزیولوژیکی مرتبط با دیابت در این مطالعه استفاده شد. داده‌ها با استفاده از نرم‌افزار (Matlab R2024, The Mathworks Inc., USA) توصیف، و تجزیه تحلیل شده است. جدول ۲ و ۳ متغیرهای کمی و کیفی پرونده‌های آزمودنی‌ها را نشان می‌دهد.

جدول ۲. متغیرهای کمی آزمودنی‌ها

شماره	شاخص‌ها	میانگین ± انحراف استاندارد
۱	هماتوکریت	۰/۳۸ ± ۰/۰۶۵
۲	کلسترول کل (مول بر لیتر)	۴/۵۶ ± ۱/۴۰۱
۳	میزان فیلتراسیون گلومرولی کلیه (میلی‌لیتر بر دقیقه)	۷۸/۹۸ ± ۲۹/۲۹
۴	فشارخون سیستولی (میلی‌متر جیوه)	۱۴۳/۹ ± ۱۸/۵
۵	فشار خون دیاستولی (میلی‌متر جیوه)	۸۵/۳ ± ۱۰/۴
۶	آسپارتات آمینو ترنسفراز (واحد در لیتر)	۲۵/۲۱ ±۳۶/۹۱
۷	کراتین سرم (میکرومول بر لیتر)	۸۶/۹۳ ± ۸۰/۰۹
۸	گاما گلوتامیل ترانسفراز (واحد در لیتر)	۴۸/۴۹ ± ۱۰۸/۰۸
۹	آلانین آمینوترانسفراز (واحد در لیتر)	۲۶/۸۲ ± ۳۲/۳۶
۱۰	PH ادرار	۵/۷۴ ± ۰/۷۲
۱۱	تعداد لکوسیت‌ها در ادرار (واحد در لیتر)	۶۱/۴۴ ± ۶۴/۵۵
۱۲	HbA1c	۴۰/۱ ± ۴/۲
۱۳	کلسترول لیپوپروتئین با چگالی بالا (میلی‌مول بر لیتر)	۱/۱۴ ± ۰/۳۶
۱۴	کلسترول لیپوپروتئین با چگالی پایین (میلی‌مول بر لیتر)	۲/۷۱ ± ۱/۰۱

جدول ۳. متغیرهای کیفی آزمودنی‌ها

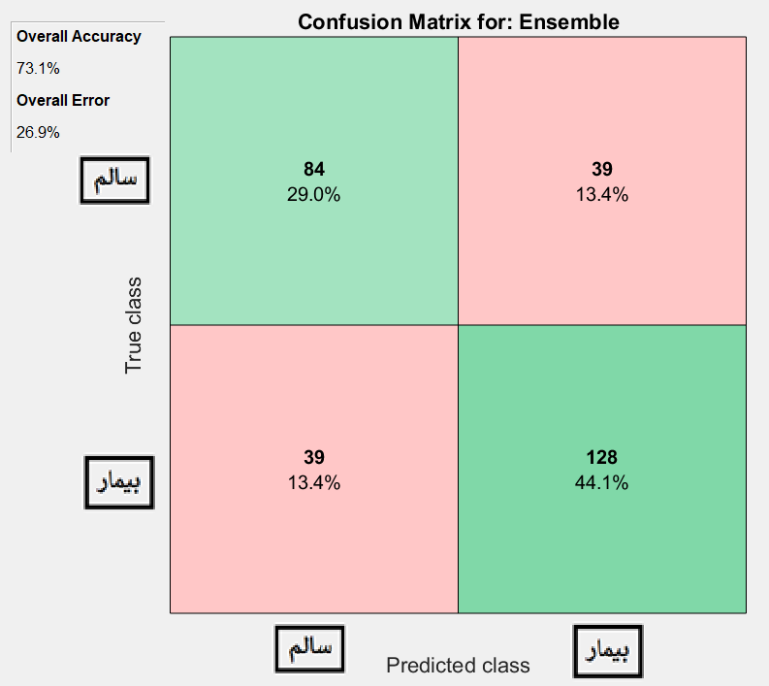
شماره	ویژگی	نوع (ناپیوسته)	درصد
۱	کبد چرب	۱ دارد	۴۲/۱۲
		۰ ندارد	۵۷/۸۸
۲	سابقه خانوادگی	۱ دارد	۸۷
		۰ ندارد	۱۳
۳	اضافه وزن	۱ دارد	۸۵
		۰ ندارد	۱۵
۴	مصرف سیگار	۱ دارد	۵۷/۰۱
		۰ ندارد	۴۲/۹۹
۵	درمان کاهش چربی (٪)	۱ دارد	۳۴/۱
		۰ ندارد	۶۶/۹
۶	عدم یا وجود بیماری قلبی	۱ دارد	۵۳
		۰ ندارد	۴۷

از الگوریتم AdaBoost، برای پیش‌بینی بیماری دیابت استفاده شد. از بین داده‌ها ۲۰ درصد برای تست و ۸۰ درصد برای آموزش الگوریتم در نظر گرفته شدند. نتایج حاصل از ماتریس درهم ریختگی این الگوریتم در شکل ۱ نشان داده شده است. نتایج نشان داد که الگوریتم AdaBoost، می‌تواند با دقت ۷۳/۱ درصد و صحت ۷۵/۳ درصد مردان فعال دارای بیماری دیابت را پیش‌بینی کند.

بحث

در دهه‌های اخیر، حجم داده‌های حوزه سلامت به طرز چشمگیری افزایش یافته است. از پرونده‌های بیماری و تاریخچه‌ی سلامت تا نتایج آزمایش‌ها و اطلاعات کلینیکی، داده‌های سلامت حاوی اطلاعات بسیار زیادی درباره وضعیت سلامت افراد و جوامع هستند. اما این حجم عظیم از داده‌ها معمولاً به صورت غیرساختاری و پراکنده ذخیره می‌شوند. در این تحقیق با استفاده از ۲۰ ویژگی مربوط به عوامل محیطی، فیزیولوژیکی، آنتروپومتری، ژنتیکی، سابقه‌ی خانوادگی و آزمایشگاهی، فرایند پیش‌بینی و تشخیص بیماری دیابت با استفاده از الگوریتم AdaBoost بررسی گردید. نتایج نشان داد که الگوریتم AdaBoost، می‌تواند با دقت ۷۳/۱ درصد و صحت ۷۵/۳ درصد مردان فعال دارای بیماری دیابت را پیش‌بینی کند.

تحت عنوان داده‌کاوی پزشکی، مطالعات گوناگون، الگوریتم‌ها و شاخص‌های مختلفی را برای پیش‌بینی بیماری دیابت بکار برده‌اند. همسو با این مطالعه، Chaves و Marques در پژوهش خود به این نتیجه رسیدند که استفاده از الگوریتم‌های داده‌کاوی برای تفسیر داده‌ها یا هدف پیش‌بینی و تشخیص بیماری دیابت، منجر به افزایش موفقیت پزشکان در شناسایی افراد مبتلابه این بیماری و تعیین جمعیت‌های در معرض خطر



شکل ۱. ماتریس درهم ریختگی الگوریتم AdaBoost

بیماری شده و به دنبال آن انجام اقدامات درمانی به موقع که می‌تواند مانع از پیشرفت بیماری و یا افزایش عوارض و مشکلات جانبی این بیماری مزمن در زندگی روزمره این گروه از بیماران شود (۱۹).

Gao و همکاران در مطالعه‌ی خود با عنوان «پردازش داده‌های دیابت نوع ۲ با الگوریتم EM و C4.5 » با ترکیب الگوریتم‌های C4.5، سیستمی برای پردازش داده‌ها برای دیابت نوع ۲ ایجاد کردند (۲۰).

Huang و همکاران مطالعه‌ای را برای شناسایی عوامل اصلی مؤثر بر کنترل دیابت با استفاده از انتخاب ویژگی در سیستم مدیریت بیمار انجام دادند (۲۱).

تعدادی از مطالعات از نظر نوع الگوریتم بکار رفته و نوع و تعداد شاخص‌های استفاده شده با مطالعه فوق همسو نبودند. در این رابطه، Han و همکاران، نرم‌افزار ماینر سریع را با استفاده از الگوریتم درخت تصمیم ID3 در پایگاه داده بیماران دیابتی به کار بردند (۲۲). Anbananthen و همکاران از شبکه عصبی مصنوعی و درخت تصمیم توسعه‌یافته با استفاده از الگوریتم C4.5 برای تشخیص افراد مبتلا به دیابت بر اساس ویژگی‌های مربوط به سن و فشار خون استفاده کردند (۲۳).

Fang، داده‌های بیماران مبتلا به دیابت را با استفاده از تکنیک‌های مختلف خوشه‌بندی کرد. ویژگی‌هایی که در این مدل‌ها مهم بودند شامل سن، سابقه‌ی خانوادگی و وزن بود، آن‌ها دقت مدل بر اساس خوشه‌بندی ۸۰ درصد گزارش کردند (۲٤).

مطالعه‌ی حاضر محدودیت‌هایی داشت و می‌تواند در تحقیقات آینده مورد توجه قرار گیرد که از آن جمله می‌توان به محدودیت جغرافیایی محل جمع‌آوری داده‌ها و تعداد رکورد اندک و در مجموعه‌ی داده اشاره نمود. یکی دیگر از مهم‌ترین محدودیت‌ها در این زمینه، نبود اطلاعات کافی مربوط به افراد سالم بود، زیرا در اکثر داده‌های ثبتی، تعداد نمونه افراد سالم بسیار کمتر از بیماران است. وجود این عدم تقارن می‌تولند منجر به عدم یادگیری کامل الگوریتم‌ها و کاهش دقت تشخیص شود. همچنین افزایش دقت در شناسایی افراد دیابتی به بزرگی پایگاه داده بستگی دارد، بنابراین در مطالعات آتی با استفاده از پایگاه‌های اطلاعاتی بزرگتر می‌توان تعداد رکوردها و دقت الگوریتم را افزایش داد.

نتیجه‌گیری

در این مطالعه، تلاشی برای شناسایی و بررسی روش‌های یادگیری ماشین و داده‌کاوی برای دیابت انجام شد. این مدل می‌تواند به عنوان ابزار کمکی در غربالگری اولیه دیابت در مراکز درمانی مورد استفاده قرار گیرد.

تشکر و قدردانی

این مقاله منتج از طرح تحقیقاتی می‌باشد و این اثر تحت حمایت مادی بنیاد ملی علم ایران (INSF) برگرفته شده از طرح شماره (٤٠٢٦٧٩٣) انجام شده است. بدین‌وسیله از حمایت مادی بنیاد ملی علم ایران (INSF) تقدیر و تشکر می‌شود.

References

1. Sabag A, Chang CR, Francois ME, Keating SE, Coombes JS, Johnson NA, et al. The effect of exercise on quality of life in type 2 diabetes: a systematic review and meta-analysis. Med Sci Sports Exerc 2023; 55(8): 1353-65.
2. Ribeiro AKPdL, Carvalho JPR, Bento-Torres NVO. Physical exercise as treatment for adults with type 2 diabetes: a rapid review. Front Endocrinol (Lausanne) 2023; 14: 1233906.
3. Kanaley JA, Colberg SR, Corcoran MH, Malin SK, Rodriguez NR, Crespo CJ, et al. Exercise/physical activity in individuals with type 2 diabetes: a consensus statement from the American College of Sports Medicine. Med Sci Sports Exerc 2022; 54(2): 353-68.
4. Frankenberg NT, Mason SA, Wadley GD, Murphy RM. Skeletal muscle cell-specific differences in type 2 diabetes. Cell Mol Life Sci 2022; 79(5): 256.
5. Bajpeyi S, Tanner CJ, Slentz CA, Duscha BD, McCartney JS, Hickner RC, et al. Effect of exercise intensity and volume on persistence of insulin sensitivity during training cessation. J Appl Physiol (1985) 2009; 106(4): 1079-85.
6. Houmard JA, Tanner CJ, Slentz CA, Duscha BD, McCartney JS, Kraus WE. Effect of the volume and intensity of exercise training on insulin sensitivity. J Appl Physiol (1985) 2004; 96(1): 101-16.
7. Kirwan JP, Solomon TP, Wojta DM, Staten MA, Holloszy JO. Effects of 7 days of exercise training on insulin sensitivity and responsiveness in type 2 diabetes mellitus. Am J Physiol Endocrinol Metab 2009; 1(2): E151-6.
8. Magalhães JP, Melo X, Correia IR, Ribeiro RT, Raposo J, Dores H, et al. Effects of combined training with different intensities on vascular health in patients with type 2 diabetes: a 1-year randomized controlled trial. Cardiovasc Diabetol 2019; 18(1): 34.
9. Motiani KK, Collado MC, Eskelinen J-j, Virtanen KA, Löyttyniemi E, Salminen S, et al. Exercise training modulates gut microbiota profile and improves endotoxemia. Med Sci Sports Exerc 2019; 52(1): 94-104.
10. Fasihi L, Tartibian B, Eslami R, Fasihi H. Artificial intelligence used to diagnose osteoporosis from risk factors in clinical data and proposing sports protocols. Sci Rep 2022; 12(1): 18330.
11. Tartibian B, Fasihi L, Eslami R, Fasihi A. Comparison and prediction of hepatic encephalopathy complications in liver transplant patients using random forest algorithm in active and inactive men. Physical Treatments 2025; 15(1): 81-90.
12. Fiarni C, Sipayung EM, Maemunah S. Analysis and prediction of diabetes complication disease using data mining algorithm. Procedia computer science 2019; 161: 449-57.
13. Abdollahi J, Nouri-Moghaddam B. Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran J Comput Sci 2022; 5(3): 205-20.
14. Sargeant JA, Yates T, McCann GP, Lawson CA, Davies MJ, Gulsin GS, et al. Physical activity and structured exercise in patients with type 2 diabetes mellitus and heart failure. Practical Diabetes 2018; 35(4): 131-8.
15. Portes J, Bullón B, Gallardo I, Fernandez-Riejos P, Quiles JL, Giampieri F, et al. Prevalence of undiagnosed diabetes and prediabetes related to periodontitis and its risk factors in elderly individuals. J Dent 2023; 132: 10448.
16. Fasihi L, Tartibian B, Eslami R. Presenting a model for detecting osteoporosis in active older men using the support vector machine algorithm. Scientific Journal of Rehabilitation Medicine 2022; 11(5): 742-53.
17. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997; 55(1): 119-39.
18. Thuraka B, Pasupuleti V, Kodete CS, Chigurupati RS, Tirumanadham NKMK, Shariff V. Enhancing Diabetes Prediction using Hybrid Feature Selection and Ensemble Learning with AdaBoost. 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC); 2024: IEEE.
19. Chaves L, Marques G. Data mining techniques for early diagnosis of diabetes: a comparative study. Appl Sci 2021; 11(5): 2218.
20. Gao J, Luo SL, Jia HB, Zhang TM, Han YW. Type 2 diabetes data processing with EM and C4. 5 Algorithm. 2007 IEEE/ICME International Conference on Complex Medical Engineering, CME 2007; 2007.
21. 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.
22. Han J, Rodriguez JC, Beheshti M. Diabetes data analysis and prediction model discovery using rapidminer. 2008 2nd International Conference on Future Generation Communication and Networking; 2008: IEEE.
23. Anbananthen KSM, Sainarayanan G, Chekima A, Teo J. Artificial neural network tree approach in data mining. Malaysian Journal of Computer Science 2007; 20(1): 51-62.
24. Fang X. Are you becoming a diabetic? A data mining approach. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery; 2009: IEEE.

DOI: *10.48305/jims.v44.i851.0137*

Journal of Isfahan Medical School **Vol. 44, No. 851, 3rd Week, April 2026**

Received: 09.03.2025 Accepted: 17.03.2026 Published: 12.04.2026

Diabetes Prediction Using the Adaboost Algorithm in Active Men

Hamid Agha-Alinejad[1], Leila Fasihi[2], Mohammad Hossein Ghahremani[3], Mozhan Agha-Alinejad[4]

Original Article

Abstract

Background: Diabetes affects cardiovascular diseases, vision loss, and kidney disease. Data mining techniques help doctors predict the disease early and treat it accurately. Given the role of physical activity in preventing diabetes, this study examined its effect alongside other risk factors. Therefore, the main objective of this study was to predict diabetes in active men using the Adaboost algorithm.

Methods: In this applied-developmental study, the records of 500 male patients aged 20-80 years with a history of regular physical activity (at least 3 90-minute sessions per week) who had visited Milad and Ayatollah Kashani hospitals in Tehran in the past 10 years were selected as samples. 20 anthropometric, genetic, family history, environmental and physiological variables were selected as input features of the algorithm. The criteria of sensitivity, accuracy and precision were used to evaluate the performance of the proposed algorithm. MATLAB (2024) software was used for data analysis.

Findings: The results showed that the Adaboost algorithm was able to predict diabetes in active men with an accuracy of 73.1% and precision of 75.3%.

Conclusion: This model can be used as an auxiliary tool in the initial screening of diabetes in medical centers.

Keywords: Diabetes, Adaboost, Men, Active, Prediction

Citation: Agha-Alinejad H, Fasihi L, Ghahremani MH, Agha-Alinejad M. **Diabetes Prediction Using the Adaboost Algorithm in Active Men.** J Isfahan Med Sch 2026; 44(851): 137- 42.

1- Professor of Exercise Physiology, Department of Physical Education & Sport Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
2- PhD of Exercise Physiology, Department of Physical Education & Sport Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
3- MSc. of Exercise Physiology, Department of Physical Education & Sport Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
4- MSc Student of Exercise Physiology, Department of Physical Education & Sport Sciences, Faculty of Physical Education & Sport Sciences, Tehran University, Tehran, Iran
Corresponding Author: Hamid Agha-Alinejad, Professor of Exercise Physiology, Department of Physical Education & Sport Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran; Email: halinejad@modares.ac.ir