مقاله پژوهشی
مجله دانشگاه علوم پزشکی رفسنجان
دوره ۱۸، اسفند ۱۳۹۸، ۱۲۸۶-۱۲۷۰

استفاده از ماشین بردار پشتیبان در تشخیص محدودیت عملکردی بیماران دیابتی شمال‌غرب ایران در سال ۱۳۹۶: یک مطالعه توصیفی

لیلی فرجی گاوگانی[۱]، پروین سربخش[۲]، محمد اصغری جعفرآبادی[۳]، مرتضی شمشیرگران[۴]

دریافت مقاله: ۹۷/۱۰/۱۲ ارسال مقاله به نویسنده جهت اصلاح: ۹۷/۱۲/۱۲ دریافت اصلاحیه از نویسنده: ۹۸/۹/۱۷ پذیرش مقاله: ۹۸/۹/۲۷

چکیده

زمینه و هدف: ماشین بردار پشتیبان (Support vector machine; SVM) به‌عنوان یک روش آماری قوی و کارآمد در تشخیص و پیش‌بینی پیامدهای بالینی بر اساس ترکیباتی از متغیرهای پیش‌بین کاربرد دارد. هدف این پژوهش، استفاده از SVM برای تشخیص محدودیت عملکردی بیماران دیابتی و بررسی میزان صحت این تشخیص می‌باشد.

مواد و روش‌ها: این پژوهش توصیفی بر روی ۳۷۸ بیمار دیابتی مراجعه کننده به مراکز دیابتی اردبیل و تبریز در سال ۹۴-۱۳۹۳ انجام شد. جهت طبقه‌بندی بیماران دیابتی از لحاظ وضعیت محدودیت عملکردی بر مبنای متغیرهای دموگرافیک و بالینی از SVM با تابع هسته (Radial basis function; RBF) و روش اعتبارسنجی آموزش و آزمون استفاده شد. ارزیابی بر اساس شاخص‌های تشخیصی شامل حساسیت، ویژگی، صحت و سطح زیر منحنی (Receiver operating characteristic; ROC) انجام شد.

یافته‌ها: نتایج حاصل از مدل SVM نشان داد که صحت طبقه‌بندی، حساسیت و ویژگی مدل SVM در افتراق و تشخیص صحیح وجود محدودیت عملکردی در بیماران دیابتی به ترتیب برابر ۹۹٪، ۱۰۰٪ و ۹۷٪ بود. سطح زیر منحنی ROC برای قدرت تشخیصی این مدل ۰/۹۸ بود.

نتیجه‌گیری: در این مطالعه SVM برای طبقه‌بندی وضعیت محدودیت عملکردی بیماران دیابتی استفاده شد که نتایج نشان‌گر صحت و دقت مناسب مدل بود. با توجه به اهمیت طبقه‌بندی صحیح پیامدهای پزشکی بر اساس ترکیباتی از متغیرهای پیش‌بین، استفاده از روش‌هایی مانند SVM که قادر به یافتن چنین ترکیبات بهینه‌ای هستند، می‌تواند مفید باشد.

واژه‌های کلیدی: داده کاوی، ماشین بردار پشتیبان، محدودیت عملکردی، طبقه‌بندی، تابع هسته

۱- کارشناس ارشد آمار زیستی، مرکز پزشکی مبتنی بر شواهد، دانشگاه علوم پزشکی تبریز، تبریز، ایران
۲- (نویسنده مسئول) استادیار گروه آموزشی آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران
تلفن: ۳۳۳۵۷۵۸۱-۰۴۱، دورنگار: ۳۳۳۴۲۲۱۹-۰۴۱، پست الکترونیکی: p.sarbakhsh@gmail.com
۳- دانشیار گروه آموزشی آمار و اپیدمیولوژی ، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران
۴- دانشیار گروه آموزشی آمار و اپیدمیولوژی ، دانشکده بهداشت، گروه آمار و اپیدمیولوژی دانشگاه علوم پزشکی تبریز، تبریز، ایران

مقدمه

طی سال‌های اخیر، علاقه‌مندی به ارزیابی و ارتقاء کیفیت زندگی بیماران مبتلا به بیماری‌های مزمن به‌صورت یک هدف درآمده است [۱]. کیفیت زندگی در بیماران مزمن از اهمیت خاصی برخوردار است به‌طوریکه سلامت جسمی، روانی و اجتماعی بیمار را تحت تأثیر قرار داده و بسیاری از بیماران بدون امید به بهبود، زندگی می‌کنند [۲]. بیماری دیابت نوع دو یک بیماری مزمن می باشد و برای متخصصان مراقبت‌های بهداشتی هدف کلیدی در درمان بیماران دیابتی افزایش کیفیت زندگی بیماران است [۳].

در مطالعه‌ای نشان داده شده است که نمره کیفیت زندگی افراد دیابتی در مقایسه با افراد سالم در گروه سنی مشابه پایین‌تر است [۴]. هم‌چنین مطالعات نشان داده‌اند که در بیماران دیابتی نوع دو حضور مشکلات بالینی تأثیر نامطلوب بر شاخص کیفیت زندگی دارد [۵].

یکی از پیامدهای مهم بیماری دیابت محدودیت در عملکرد روزانه بیماران است. محدودیت در عملکرد فیزیکی به محدودیت در انجام انواع فعالیت‌های مختلف روزانه مانند استحمام کردن تا فعالیت‌های سودمند روزانه مانند خرید اطلاق می‌شود [۶]. مطالعات متعددی تأثیر بیماری دیابت بر ناتوانی در عملکرد روزانه افراد را نشان داده‌اند [۹-۷]. با کاهش عملکرد زندگی روزمره، کیفیت زندگی کاهش می‌یابد، بنابراین با کاهش محدودیت عملکردی می‌توان به بهبود کیفیت زندگی و ارتقاء کیفیت مراقبت در این بیماران کمک شایانی کرد [۱۰].

یکی از روش‌های پیشرفته و قوی آماری در زمینه تشخیص و طبقه‌بندی صحیح پیامدها بر اساس عوامل خطر مرتبط که اخیراً کاربرد فراوانی در علوم پزشکی پیدا کرده است، روش ماشین بردار پشتیبان (Support vector machine; SVM) است [۱۱]. با توجه به نتایج قابل قبول این روش در کاربردهای پزشکی [۱۴-۱۲] در این مطالعه بر آن شدیم تا از این روش برای طبقه‌بندی وضعیت محدودیت عملکردی بیماران دیابتی استفاده کنیم.

لذا هدف از این پژوهش استفاده از SVM برای تشخیص محدودیت عملکردی بر اساس عوامل خطر مرتبط شامل سن، جنسیت، طول مدت ابتلاء، نمایه توده بدنی و شاخص‌های کیفیت مراقبت شامل کلسترول، تری‌گلیسیرید، قند خون ناشتا، HDL (high-density lipoprotein)، فشارخون سیستولیک و دیاستولیک و هموگلوبین A1C و تعیین میزان دقت و کارآیی این تشخیص و طبقه‌بندی است.

مواد و روش‌ها

داده‌های این مطالعه که در سال ۱۳۹۶ انجام شد، برگرفته از دو مطالعه توصیفی انجام شده در شمال‌غرب ایران می‌باشد [۱۶-۱۵]. افرادی که تشخیص قطعی دیابت داشته و به مراکز ارجاعی (بیمارستان امام خمینی (ره) استان اردبیل در فاصله بهمن ماه ۱۳۹۳ لغایت اردیبهشت ۱۳۹۴ و بیمارستان‌های امام رضا (ع) و سینای تبریز در فاصله زمانی آذر ماه ۱۳۹۴ لغایت اسفند ۱۳۹۴) مراجعت داشته و واجد شرایط بودند با روش نمونه‌گیری در دسترس وارد مطالعه مذکور شدند. مراکز

فوق به‌عنوان کلینیک‌های ارجاعی بیماران دیابت در این منطقه شناخته می‌شوند.

معیارهای ورود به مطالعه شامل: تشخیص دیابت نوع ۲، سن بالاتر از ۲۵ سال، داشتن پرونده مراقبت در مراکز یاد شده حداقل برای یک‌سال، سکونت در مناطق شهری، نداشتن بیماری‌های خاص و ناتوان کننده (مانند هموفیلی، تالاسمی و غیره) بودند. معیارهای خروج از مطالعه نیز شامل: عدم تمایل به حضور در مطالعه و داشتن سایر انواع دیابت بودند. جزئیات بیش‌تر در مورد این مطالعات در چندین مطالعه ارائه شده‌اند [۱۵-۱۷]. مطالعات فوق با کد اخلاقی TBZMED.REC.1392.207 و TBZMED.REC.1394.5 در کمیته اخلاق دانشگاه علوم پزشکی تبریز به تصویب رسیده‌اند.

در مطالعه حاضر با ادغام داده‌های این دو مطالعه، حجم نمونه به ۶۹۶ نفر رسید، ولی پس از حذف نمونه‌های دارای مقادیر گمشده، در نهایت ۳۷۸ نفر با داده‌های کامل برای متغیرهای مد نظر جهت فرآیند مدل سازی باقی ماندند.

اطلاعات مورد نیاز در خصوص پارامترهای مرتبط با کیفیت مراقبت در طول یک‌سال گذشته شامل نتایج آزمایش: هموگلوبین A1C، فشارخون، پروفایل چربی (کلسترول، تری گلیسیرید، HDL)، نتیجه آخرین آزمایش قند خون ناشتا و طول مدت ابتلاء به بیماری و سایر اطلاعات مورد نیاز با استفاده از چک لیست و از طریق پرونده‌های مراقبت جمع آوری شده‌اند. هم‌چنین پرسش‌نامه استاندارد اختلال عملکردی توسط پرسش‌گر آموزش دیده در محل کلینیک دیابت تکمیل شده است.

برای ارزیابی کیفیت زندگی به‌منظور بررسی وضعیت سلامتی بیماران و تعیین نتایج مراقبت‌های بهداشتی از پرسش‌نامه ۲۶ سؤالی WHOQOL-BREF استفاده گردیده است که دو سؤال اول جهت بررسی کیفیت کلی زندگی و میزان درک خود فرد از کیفیت زندگی طراحی شده است و ۲۴ سؤال بعدی وضعیت سلامتی و کیفیت زندگی را در چهار بعد: سلامتی جسمانی با هفت سوال، سلامتی روانی با شش سوال، روابط اجتماعی با سه سؤال و محیط فیزیکی با هشت سوال مورد سنجش قرار گرفته است استاندارد سازی نسخه فارسی این پرسش‌نامه توسط دکتر نجات و همکاران در سال ۱۳۸۴ انجام شده و روایی و پایایی آن مورد تأیید قرار گرفته است [۱۷-۱۹].

نمره‌دهی پرسش‌نامه به روش لیکرت انجام گرفته است هر سؤال شامل پنج پاسخ (اصلا"، کم، متوسط، زیاد، خیلی‌زیاد) می‌باشد، به هر کدام از پاسخ‌ها به‌ترتیب از یک تا پنج امتیاز تعلق می‌گیرد و نمره‌گذاری سؤالات ۳، ۴، ۲۶ به صورت معکوس انجام می‌گیرد. نمره بالاتر در هر کدام از ابعاد نشان دهنده کیفیت زندگی بهتر است. پس از محاسبه نمره خام در هر کدام از ابعاد، می‌توان نمرات را بر اساس دستورالعمل سازمان بهداشت جهانی به مقیاس ۱۰۰-۰ یا ۲۰-۴ تبدیل کرد که در این مطالعه از بازه ۱۰۰-۰ استفاده شده است [۲۲-۲۱، ۱۸، ۶].

برای تعیین ظرفیت عملکردی از پرسش‌نامه ارزیابی کیفیت زندگی ۳۶ سؤالی Short Form Health (SF36) Survey استفاده شده است [۱۹-۱۷]. این پرسش‌نامه که توسط پرسش‌گر آموزش دیده در محل کلینیک دیابت تکمیل شده، شامل ۱۰ سؤال در مورد محدودیت عملکردی در زمینه های مختلف بود و از شرکت کنندگان خواسته شده بود که برحسب انجام فعالیت‌های فیزیکی متوسط و شدید مانند بلند کردن، خرید کردن، از پله بالا رفتن، قدم زدن، خم شدن، زانو زدن، حمام کردن و لباس پوشیدن، یکی از سه گزینه: ۱- زیاد مشکل دارم، ۲- کمی مشکل دارم و ۳- اصلاً مشکل ندارم، انتخاب کنند در برخی سؤالات کدگذاری مجدد جهت همسو شدن سؤالات صورت گرفته و امتیازها از ۰ تا ۱۰۰ متغیر می-باشد، صفر بدترین و ۱۰۰ بهترین وضعیت در مقیاس مورد نظر را نشان می‌دهد [۲۰]. سؤالات این پرسش‌نامه در جدول ۱ ارائه شده است.

جدول ۱- سؤالات پرسش‌نامه استاندارد اختلال عملکردی

۱- در فعالیت‌های شدید مثل دویدن، بلندکردن اجسام سنگین، شرکت کردن در ورزش‌های سنگین
۲- در فعالیت‌های متوسط مثل جابه‌جا کردن میز، کشیدن جاروبرقی و...
۳- در حمل‌کردن خرید روزانه
۴- در بالا رفتن از چند طبقه
۵- در بالا رفتن از یک طبقه
۶- خم‌شدن، دولا شدن، زانوزدن
۷- پیاده‌روی بیش از یک کیلومتر
۸- پیاده‌روی به فاصله چند صد متر
۹- پیاده‌روی به فاصله یک صد متر
۱۰- در استحمام یا پوشیدن لباس‌های خود

با کد بندی مجدد داده‌ها و اختصاص مقدار صفر برای افراد بدون محدودیت عملکردی و یک برای افراد با محدودیت کم و ۲ برای افراد با محدودیت‌های عملکردی زیاد، نمره کل برای هر فرد با استاندارد کردن نمره و تبدیل آن به درصد محاسبه شده است. تقسیم‌بندی طبق رفرنس، به این صورت بود که کسب نمره کل بین ۹۰ تا ۱۰۰، ویژگی عدم محدودیت عملکردی و مابین ۰ تا ۸۹ وجود محدودیت عملکردی در بیمار دیابتی قلمداد شده است [۲۲ -۲۱، ۶].

مدل‌های مبتنی بر الگوریتم یادگیری ماشین (برنامه نویسی برای استخراج یک مدل کلی از روی داده‌ها) دو نوع هستند: مدل‌های مبتنی بر الگوریتم یادگیری با ناظر (Supervised learning) و مدل‌های مبتنی بر الگوریتم یادگیری بی‌ناظر (Unsupervised learning). در مدل‌های با ناظر، برای داده‌ها گروهی از قبل برای عضویت مشاهدات وجود دارد و در مدل‌های بی‌ناظر، برای داده‌ها از قبل گروهی برای عضویت مشاهدات وجود ندارد و داده‌ها به خوشه‌هایی همگن تقسیم می‌شوند. شماری از این روش‌ها بر مبنای تابع هسته عمل می‌کنند. روش SVM نیز جزء روش‌های داده کاوی با الگوریتم با ناظر می‌باشد و کلاس‌بندی را بر اساس تابع هسته با هدف طبقه‌بندی دو یا چند گروهی و رگرسیون انجام می دهد. این روش برای اولین بار توسط Vaping در سال ۱۹۶۳ برای تفکیک و دسته‌بندی داده‌هایی که جداپذیر خطی بودند ساخته شد و در سال ۱۹۹۵ توسط Vaping و Cortes برای حالت غیر خطی تعمیم داده شد [۱۱]. امروزه ماشین‌های

بردار پشتیبان، به متداول‌ترین تکنیک‌های پیش‌بینی و تشخیصی در داده کاوی تبدیل شده‌اند [۱۲].

فرآیند ساخت مدل شامل دو مرحله آموزش و آزمون می باشد که در مرحله آموزش، مدل الگو و روابط موجود بین پیامد و پیش‌بین‌ها را یاد می‌گیرد و در انتهای فاز آموزش، قابلیت تعمیم‌دهی مدل برازش داده شده توسط داده‌های آزمون مورد ارزیابی قرار می‌گیرد. SVM پیش‌بینی‌های خود را با استفاده از ترکیبات خطی و غیر خطی بر روی مجموعه‌ای از داده‌های آموزش به‌نام بردارهای پشتیبان انجام می‌دهد. با توجه به مزایایی که این روش نسبت به روش‌های کلاسیک آماری دارد و نیاز به پیش‌فرض خاصی در مورد داده‌ها نیست، شناسایی الگوی ارتباط از روی داده‌های گروه آموزش صورت می‌گیرد و از روی الگوی تشخیص داده شده توسط مدل، داده های جدید ارزیابی می‌شوند. SVM نسبت به شبکه‌های عصبی تئوری قوی‌تری داشته و نسبت به بیش برازش مقاوم‌تر است. عدم اهمیت به ساختار داده‌ها و کلاس‌ها، قابلیت استفاده برای داده‌ها با ابعاد بالاتر و استفاده از اصول ریسک ساختاری، به منظور کاهش خطای مدل از مزایای دیگر این روش است [۱۲].

یک فرض بسیار قوی در SVM این است که داده‌ها به صورت جداپذیر خطی هستند. در این‌صورت خط جداکننده بین کلاس‌ها با استفاده از الگوریتم‌های مختلفی مانند پرسپترون (Perceptron) رسم می‌شود. در این روش مرز خطی بین دو کلاس به گونه‌ای محاسبه می‌شود که: ۱- تمام نمونه‌های کلاس ۱+ در یک‌طرف مرز و تمام نمونه‌های کلاس ۱- در طرف دیگر مرز واقع شوند و ۲- مرز تصمیم‌گیری به گونه‌ای می‌باشد که فاصله نزدیک‌ترین نمونه‌های آموزشی هر دو کلاس از یک‌دیگر در راستای عمود بر مرز تصمیم‌گیری تا جایی که ممکن است حداکثر شود [۲۴].

دو خط مرزی موازی خط جداکننده رسم می‌کنیم و با فاصله‌های مساوی از خط جداکننده آن‌قدر از هم دور می‌کنیم تا به داده‌ها برخورد کند. به داده‌هایی که به این خط برخورد کنند، بردارهای پشتیبان گفته می‌شود [۱۲]. حل معادله پیدا کردن خط بهینه به روش QP (Quadratic programming) که روش شناخته شده‌ای در حل مسائل محدودیت‌دار هستند، صورت می‌گیرد. در حالت خطی، SVM با در نظر گرفتن مجموعه آموزشی (X_k, Y_k)، $K=1,2,\ldots\ldots,m$، $X_k \in R_n$,$Y_k \in \{-1,+1\}$,ابر صفحه بهینه با حداکثر حاشیه را از طریق حل مسأله بهینه‌سازی زیر پیدا می‌کند:

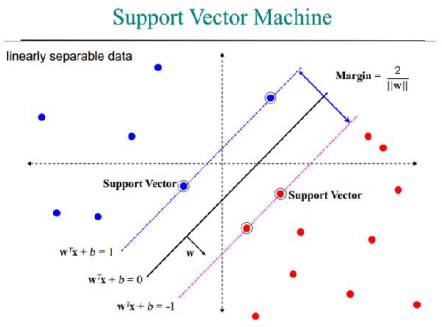
$$\min 1/2 \|w\|^2 : y_k((w.x_k)+b) \geq 1 \qquad K=1,2,\ldots\ldots,m$$

$w.x+b=0$ نشان‌دهنده صفحه جدا کننده مرزهای دو ناحیه است. فضای بین این دو کلاس حاشیه QP می‌باشد که هدف به حداکثر رساندن این فاصله می‌باشد. زمانی که این فاصله بیش‌تر باشد داده‌ها به‌خوبی از هم تفکیک شده و کیفیت خروجی و دقت و صحت طبقه‌بندی افزایش پیدا خواهد کرد (شکل ۱). یک مرز تصمیم‌گیری خطی را در حالت کلی می‌توان به‌صورت زیر نوشت:

$$w.x+b=0$$

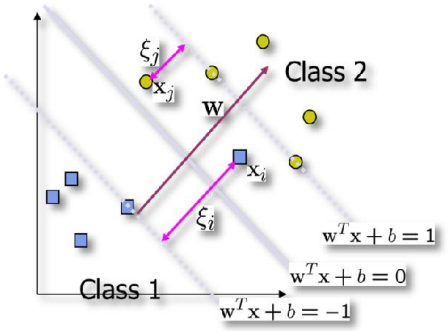
x یک نقطه بر روی مرز تصمیم‌گیری و w یک بردار n بعدی عمود بر مرز تصمیم‌گیری است. b فاصله مبدأ تا مرز

تصمیم‌گیری و w. x بیان‌گر ضرب داخلی دو بردار w و x است [۱۲].



شکل ۱- داده‌های فضای ورودی در حالت جداپذیر خطی برای داده های آموزش

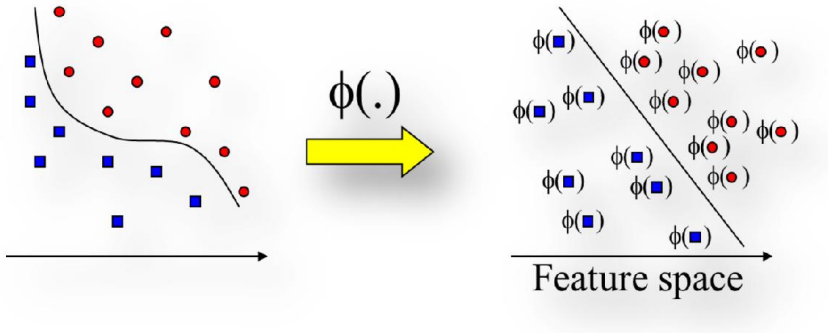
در صورتی که داده‌ها به صورت جداناپذیر خطی باشد (شکل ۲)، یعنی داده‌هایی داشته باشیم که توسط تابع $(w.x_k)+b$ به اشتباه دسته‌بندی شده باشند، متغیرهایی کمکی (Slack) به‌کار برده می‌شود که برابر مقدار لازم برای انتقال داده‌ها به مرز جداکننده ایده آل می‌باشد. بعد از انتقال، داده‌ها به صورت خطی جداپذیر خواهند بود [۲۵].



شکل ۲- داده‌های فضای ورودی در حالت جداناپذیر خطی برای داده‌های آموزش

حالت دیگر زمانی اتفاق می‌افتد که داده‌ها در فضای ورودی به صورت غیرخطی باشند، یعنی کلاس داده‌ها هم‌پوشانی داشته باشند (شکل ۳). در این صورت با استفاده از توابع غیرخطی تحت عنوان فضای ویژگی، داده‌ها به فضایی با ابعاد بیش‌تر انتقال داده می‌شود، تا در این فضا به صورت خطی از هم جدا شوند و در فضای جدید با جایگزینی x_k با $\varphi(x_k)$ مرز بهینه را می‌توان محاسبه نمود [۲۶].

فضای ویژگی دارای ابعاد بزرگ می‌باشد، به همین خاطر دارای محاسبات زیاد و پرهزینه خواهد بود. برای غلبه بر این مشکل از تابع هسته (Kernel) استفاده می‌شود که مبتنی بر ضرب داخلی می‌باشند [۲۷-۲۸].



شکل ۳- داده‌های فضای ورودی در حالت غیرخطی

هسته‌های متفاوتی وجود دارد که انتخاب هسته مناسب بستگی به ماهیت داده‌ها دارد. در حالت کلی دو نوع هسته داریم: هسته خطی و هسته غیر خطی که شامل سیگموید، چندجمله‌ای و تابع شعاعی پایه یا گوسین (Radial basis function; RBF) می‌باشد که متداول‌ترین آن‌ها RBF می‌باشد. انواع توابع هسته همراه با الگوریتم‌های مربوطه در جدول ۲ نشان داده شده است.

جدول ۲- انواع توابع هسته خطی و غیر خطی مورد استفاده در فضای ویژگی برای جداسازی داده‌های غیرخطی

الگوریتم	تابع هسته
$K(x_i,x_j)=\langle x_i,x_j\rangle$	**Linear**
$K(x_i,x_j)=(r+\gamma\langle x_i,x_j\rangle)^d$	**Polynomial**
$K(x_i,x_j)=\exp\left(-\gamma\|x_j-x_i\|^2\right)$	**Radial Basis Function (RBF)**
$K(x_i,x_j)=\tanh(\gamma X_i^T X_j+r)$	**Sigmoid**

x_i,x_j مجموعه‌ای از داده های آموزش، γ(گاما) یک پارامتر تعریف شده توسط کاربر به‌عنوان عرض کرنل، d درجه چند جمله‌ای، r مقدار اریب یا بایاس، T ماتریس واحد

در حالت کلی دو نوع SVM وجود دارد که به نوع پیامد بستگی دارد. اگر پیامد مد نظر پیوسته باشند از نوع رگرسیون و اگر گسسته باشند از نوع رده‌بندی استفاده می‌شود. یک نکته حائز اهمیت در بهبود قدرت یادگیری و قابلیت تعمیم الگوریتم SVM، در نتیجه افزایش دقت دسته‌بندی، بهینه‌سازی پارامترهای مدل است. تنظیم پارامترهای مدل تأثیر زیادی در دقت دسته‌بندی دارد که مهم‌ترین این پارامترها، پارامتر جریمه (Cost) برای کنترل سادگی تابع هسته و پارامترهای تابع هسته مانند گاما γ برای هسته گوسی می‌باشد که هر چقدر مقدار پارامتر جریمه بزرگ‌تر شود، خطای γ آموزش کم‌تر خواهد شد و از طرفی قدرت تعمیم‌دهی مدل نیز کم‌تر می‌شود. برای تعیین مقایر بهینه پارامترهای جریمه (C) و γ از روش اعتبار سنجی (Cross-validation) استفاده می‌شود [۲۹].

روش SVM برای بررسی نحوه ارتباط بین متغیرهای جنسیت، سن، طول مدت ابتلاء، نمایه توده بدنی، HDL، کلسترول، تری‌گلیسیرید، قند خون ناشتا، فشارخون سیستولیک و دیاستولیک و هموگلوبین A1C به‌عنوان عوامل خطر مرتبط و محدودیت عملکردی به‌عنوان پیامد به‌کار گرفته شد. با توجه به این که بیماران دیابتی در یکی از دو کلاس دارای محدودیت عملکردی و عدم محدودیت عملکردی قرار دارند، یعنی متغیر پاسخ، از نوع کیفی دوحالته است بنابراین نوع مدل به‌کار گرفته شده برای SVM از نوع طبقه‌بندی می‌باشد. نوع هسته به‌کار گرفته شده نیز از نوع RBF می‌باشد. پارامترهایی که برای بهینه‌سازی مدل استفاده می‌شود پارامتر C با مقدار یک و پارامتر γ برای تابع هسته غیر خطی استفاده شده با مقدار ۱ می‌باشد [۳۱-۳۰].

مهم‌ترین معیار برای ارزیابی کارایی یک الگوریتم دسته-بندی کننده، میزان دقت و دسته‌بندی آن الگوریتم است. میزان دقت نشان می‌دهد که تا چه اندازه داده‌ها به درستی در طبقه مربوط به خود قرار گرفته اند و معیارهای ارزیابی حساسیت و ویژگی که ویژگی میزان طبقه‌بندی صحیح هر کدام از افراد دارای محدودیت عملکردی و عدم محدودیت عملکردی در طبقه مربوط به خود می‌باشد.

به منظور اعتبار سنجی نتایج حاصل از SVM، روش اعتبار سنجی آموزش و آزمون به کار برده شده است. به این صورت که داده‌ها به طور تصادفی به دو قسمت ۷۰ درصد (داده آموزش) و ۳۰ درصد (داده آزمون) تقسیم شدند. پس از برازش مدل بر روی داده‌های آموزش، برای اعتبار سنجی نتیجه به دست آمده و بررسی قابلیت تعمیم‌دهی مدل، بر روی داده‌های آزمون با استفاده از سطح زیر منحنی (ROC; Receiver operating characteristic) مورد ارزیابی قرار گرفت. با توجه به این که در این روش داده‌های گروه آزمون کاملاً از روند مدل‌سازی کنار گذاشته می‌شوند و در صورت داشتن حجم نمونه پایین، این کار باعث کاهش توان مطالعه می‌شود. یک روش دیگر برای سنجش اعتبار مدل، استفاده از روش اعتبار سنجی k باره است. در این روش کل داده‌ها به k دسته تقریبا هم اندازه تقسیم می‌شوند. K-1 دسته برای آموزش مدل به کار می‌رود و مدل بر روی k امین دسته آزمون می‌شود. این‌کار k بار تکرار می‌شود، به‌طوری که هر بار یکی از دسته‌ها به عنوان آزمون و بقیه به‌عنوان آموزش استفاده شود. میانگین شاخص‌های اعتبار سنجی حاصل از k دسته آزمون به‌عنوان شاخص نهایی گزارش می‌شود [۳۲].

در مطالعه حاضر، برای اعتبار سنجی مدل و ارزیابی قابلیت تعمیم‌دهی نتایج به داده‌های جدید، با استفاده از روش اعتبار سنجی ۱۰ باره، داده‌ها به ۱۰ قسمت تقریباً مساوی تقسیم شدند و پس از برازش مدل بر روی داده‌های آموزش، مدل بر روی داده‌های آزمون، مورد ارزیابی قرار گرفت و در نهایت، متوسط شاخص‌های مربوط به ارزیابی دقت طبقه‌بندی کننده حاصل از ۱۰ گروه آزمون گزارش شد. تحلیل‌ها با نرم افزار R نسخه ۳/۳/۱ و بسته نرم افزاری auc و e1071 پس از تعیین شاخص‌های مرکزی و پراکندگی مورد تجزیه و تحلیل قرار گرفت. پس از بررسی نرمال بودن داده‌ها با استفاده از آماره‌های توصیفی، برای مقایسه متغیرهای کمی نرمال و غیرنرمال در دو گروه به ترتیب از آزمون‌های t مستقل و Mann-Whitney استفاده گردید. سطح معنی‌داری در آزمون‌ها ۰/۰۵ در نظر گرفته شده است.

نتایج

جامعه مورد پژوهش، بیماران دیابتی نوع دو مراجعه کننده به مرکز دیابت بیمارستان امام خمینی (ره) در اردبیل یا در مراکز دیابت بیمارستان‌های امام رضا (ع) و سینای تبریز که در فاصله بهمن ماه ۱۳۹۳ لغایت اسفند ۱۳۹۴ دارای پرونده مراقبت بوده و ساکن آن شهر می‌باشند با سن ۲۹ تا ۷۰ سال و میانگین و انحراف معیار سنی ۸/۷۹ ± ۵۶/۱۶ سال و میانگین و انحراف معیار طول مدت ابتلاء آن‌ها ۶/۱۱ ± ۸/۹۱ سال بود. از بین ۳۷۸ بیمار دیابتی، ۲۵۵ نفر (۶۷/۴۶٪) دارای

محدودیت عملکردی بودند. از ۲۴۹ زن تحت مطالعه نیز ۱۸۰ نفر (۷۲/۳٪) دارای محدودیت عملکردی بودند در حالی که در مردان ۷۵ نفر (۵۸/۱٪) دارای محدودیت عملکردی بودند. بین زنان و مردان تفاوت معنی‌داری از لحاظ احتمال ابتلاء به محدودیت عملکردی مشاهده شد (P =۰/۰۰۵). اطلاعات پایه ای این افراد شامل متغیرهای دموگرافیک و بیوشیمیایی در جدول ۳ ارائه شده است. با توجه به جدول ارائه شده، از لحاظ آماری بین متغیرهای سن (P <۰/۰۰۱)، نمایه توده بدنی (P <۰/۰۰۱) و طول مدت ابتلاء (P=۰/۰۱۴) در دو گروه بیماران دارای محدودیت عملکردی و بدون محدودیت عملکردی اختلاف معنی‌داری مشاهده شده است.

جدول ۳- مشخصات دموگرافیک و بیوشیمیایی پایه‌ای بیماران دیابتی دارای پرونده مراقبت در بیمارستان‌های امام رضا(ع) و سینای تبریز و بیمارستان امام خمینی اردبیل در فاصله بهمن ماه ۱۳۹۳ لغایت اسفند ۱۳۹۴ به تفکیک وضعیت محدودیت عملکردی بیماران

متغیر	بدون محدودیت عملکردي (n=۱۲۳)	دارای محدودیت عملکردي (n=۲۵۵)	مقدار P
سن (سال)	۵۱/۰۲ ± ۸/۶۷	۷/۷۲ ± ۵۸/۶۵	<۰/۰۰۱
نمایه توده بدنی (کیلوگرم بر متر مربع)	۳/۸۶ ± ۲۷/۷۶	۴/۸۷ ± ۲۹/۷۴	<۰/۰۰۱
هموگلوبین A1C (%)	۱/۸۵ ± ۷/۸۵	۱/۹۴ ± ۸/۰۳	۰/۳۹۰
قند خون ناشتا (میلی‌گرم بر دسی‌لیتر)	۱۵۰ (۱۱۶-۱۹۴)	۱۵۰ (۱۱۸-۱۸۵)	۰/۹۵۸*
فشارخون سیستولیک (میلی‌متر جیوه)	۱۷/۰۱ ± ۱۲۷/۱۹	۱۷/۳۸ ± ۱۲۵/۵۲	۰/۳۸۰
فشارخون دیاستولیک (میلی‌متر جیوه)	۱۰/۵۹ ± ۷۵/۱۶	۹/۵۳ ± ۷۵/۱۳	۰/۹۸۱
کلسترول (میلی‌گرم بر دسی‌لیتر)	۴۰/۸۷ ± ۱۷۴/۵۵	۴۳/۳۶ ± ۱۷۴/۳۸	۰/۹۷۱
لیپو پروتئین با چگالی بالا (میلی‌گرم بر دسی لیتر)	۴۵ (۳۷-۵۱)	۴۸ (۴۰-۵۶)	۰/۱۰۰*
تری‌گلیسیرید (میلی‌گرم بر دسی‌لیتر)	۱۵۰ (۱۰۶-۱۹۹)	۱۶۴ (۱۱۲-۲۱۸)	۰/۱۸۵
طول مدت ابتلاء (سال)	۵/۶۰± ۷/۸۰	۶/۲۸ ± ۹/۴۵	۰/۰۱۴

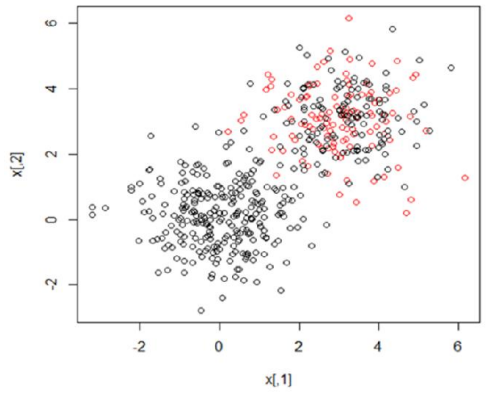
داده‌های جدول به صورت انحراف استاندارد ± میانگین و یا (چارک سوم-چارک اول) میانه گزارش شده است. از آزمون t مستقل و* آزمون Mann-Whitney و ۰/۰۵ به‌عنوان اختلاف معنی‌داری استفاده گردیده است.

بر اساس میانگین اعتبار سنجی حاصل از ۱۰ دسته آزمون، میزان دقت ۰/۹۹، حساسیت ۱۰۰ درصد و ویژگی ۹۷ درصد به دست آمد که نشان می‌دهد مقدار خطایی که در هنگام طبقه بندی داده‌ها رخ می‌دهد ۱ درصد می‌باشد، یعنی با احتمال حدود ۹۹ درصد افراد به‌درستی در کلاس مربوطه قرار گرفته‌اند.

یکی دیگر از روش‌های مناسب برای ارزیابی نتایج به‌دست آمده از یک طبقه‌بندی کننده، استفاده از منحنی ROC می-باشد [۳۳]. هر چه میزان انحراف از خط مبنا بیش‌تر باشد، کارآیی آن طبقه‌بندی کننده در شناسایی طبقات بهتر است. علاوه بر بررسی روند نمودار طبقه مورد نظر، هر چقدر سطح زیر منحنی این نمودار بیش‌تر باشد، قابلیت اطمینان این روش بیش‌تر خواهد بود [۳۴] که سطح زیر این منحنی برای داده‌های به‌کار گرفته شده برای تعیین ارزش طبقه‌بندی

محدودیت عملکردی بر مبنای عوامل خطر مورد بررسی با استفاده از روش SVM برابر ۹۸ درصد به‌دست آمده است. مقدار سطح زیر منحنی ROC با استفاده از روش رگرسیون لجستیک بر روی داده‌های مورد مطالعه با همان عوامل خطر مرتبط نیز محاسبه شده که برابر ۷۹ درصد به‌دست آمده و نشان دهنده کارآیی بهتر روش SVM در طبقه‌بندی داده‌ها در مقایسه با مدل لجستیک مرسوم می‌باشد.

بر اساس مدل SVM و با توجه به شکل ۴، می‌توان نتیجه گرفت که داده‌ها در فضای ورودی به صورت خطی از هم جداپذیر نمی‌باشند. بنابراین از توابع کرنل (Kernel functions) مناسب برای جداسازی بیماران دیابتی از لحاظ اختلال عملکردی استفاده می‌شود:



شکل ۴ -فضای ورودی مربوط به دو کلاس از بیماران دیابتی (دارای اختلال عملکردی و عدم اختلال عملکردی) دارای پرونده مراقبت در بیمارستان امام رضا (ع) و سینای تبریز در فاصله زمانی بهمن ماه ۱۳۹۳ لغایت اسفند ۱۳۹۴

جدول ۴ نشان می‌دهد که روش SVM با تابع کرنل غیر خطی RBF از ارزش و کارآیی بیش‌تری برای طبقه‌بندی بیماران دیابتی از لحاظ اختلال عملکردی نسبت به تابع کرنل خطی (Linear) برخوردار می‌باشد و بیش‌ترین دقت و حساسیت را در طبقه‌بندی محدودیت عملکردی در بیماران دیابتی دارا می‌باشد و این بیان‌گر یک الگوی غیر خطی بین داده‌های این مطالعه می‌باشد.

جدول ۴ - نتایج حاصل از برازش مدل SVM بر روی داده‌های گروه آزمون بیماران دیابتی دارای پرونده مراقبت در بیمارستان‌های امام رضا(ع) و سینای تبریز و بیمارستان امام خمینی اردبیل در فاصله زمانی بهمن ماه ۱۳۹۳ لغایت اسفند ۱۳۹۴

کرنل	حساسیت	ویژگی	میزان دقت	AUC
Linear	%۹۳	%۴۲	%۷۸	۰/۶۷۵
Radial	%۱۰۰	%۹۷	%۹۹	۰/۹۸

بحث

در پژوهش حاضر میزان خطا و هم‌چنین میزان درستی طبقه بندی افراد دیابتی از لحاظ محدودیت عملکردی، بر مبنای ویژگی‌هایی مانند جنسیت، سن، طول مدت ابتلاء و هم‌چنین ویژگی‌های کیفیت مراقبت شامل: فشارخون، قند خون، نمایه توده بدنی، HDL، کلسترول، هموگلوبین A1C و

تری‌گلیسرید مورد بررسی قرار گرفته و نتایج بیان‌گر عملکرد خوب SVM در طبقه‌بندی صحیح محدودیت عملکردی بیماران دیابتی بود. وجود ارتباط معنی‌دار و قدرت تشخیص و طبقه‌بندی این شاخص‌ها برای محدودیت عملکردی قابل انتظار است زیرا مطالعات انجام گرفته در مورد عوامل خطر مرتبط با محدودیت عملکردی در بیماران دیابتی نشان می-دهد که کنترل ضعیف دیابت (بر مبنای شاخص هموگلوبین A1C) با محدودیت عملکردی در ارتباط است و افراد با کنترل ضعیف‌تر و شاخص هموگلوبین A1C بالاتر در مقایسه با افراد با کنترل بهتر بیماری، بیش‌تر دچار محدودیت در عملکرد روزانه و ناتوانی شده‌اند [۷].

مطالعات متعددی کارآیی خوب این روش قدرت‌مند را در افتراق صحیح افراد نشان داده‌اند. در مطالعه‌ای که برای بررسی قابلیت روش‌های یادگیری ماشین در تحقیقات مربوط به دیابت انجام شده است، از روش SVM و رگرسیون برای تشخیص و پیش‌بینی بیماری دیابت استفاده شده که نتایج نشان‌دهنده بهتر بودن کارآیی الگوریتم SVM نسبت به بقیه روش‌ها (مانند لوجستیک) در تشخیص و پیش‌بینی این بیماری می‌باشد [۱۳].

نتایج پژوهش دیگری که از SVM برای تفکیک دیابتی‌ها از سایر افراد استفاده کرده است، نشان‌دهنده قدرت طبقه‌بندی خوب این روش است و سطح زیر منحنی ROC برای صحت این طبقه‌بندی برابر ۸۲/۵ درصد به‌دست آمده است. سطح زیر منحنی ROC برای افتراق افراد پره دیابت و دیابتی از افراد سالم با استفاده از SVM برابر ۷۳/۲ درصد بود [۱۴]. نویسندگان این مقاله سایتی (http://www.hugenavigator.net/DiseaseClassification-Portal/startPageDiabetes.do) طراحی کرده‌اند که در این سایت، پژوهش‌گران می‌توانند از نتایج مدل SVM حاصل از متغیرهایی مانند فعالیت فیزیکی، سن، سابقه خانوادگی، نژاد و قومیت، وزن، قد، دور کمر، شاخص توده بدنی (BMI) و فشار خون بالا برای تشخیص دیابت و پره دیابت استفاده کنند.

در مطالعه دیگری در ایران از SVM برای پیش‌بینی دیابت بر اساس سن و جنسیت و چربی‌های کلسترول، LDL (Low-density lipoprotein)، HDL و تری‌گلیسرید استفاده گردید، نتایج بیان‌گر کارآیی قابل قبول این روش در تشخیص دیابت است و سطح زیر منحنی ROC برابر ۸۸/۷۷ درصد به‌دست آمده است [۳۵]. در مطالعه دیگری نشان داده شد که روش SVM در پیش‌بینی افراد دیابتی نسبت به رگرسیون لجستیک مناسب‌تر می‌باشد [۱۴].

ولی مطالعه مشابهی که از SVM برای بررسی عوامل مرتبط با محدودیت عملکردی استفاده کرده باشد و صحت طبقه‌بندی این مدل را در افتراق و تشخیص صحیح افراد از لحاظ وضعیت محدودیت عملکردی تعیین کرده باشد، در جستجوی محققین یافت نشد.

وجود داده‌های گمشده نسبتاً زیاد در دو مطالعه انجام شده در شمال‌غرب ایران [۱۶-۱۵] از محدودیت‌های مطالعه حاضر می‌باشد که امکان استفاده از همه داده‌ها را در تحلیل فراهم نکرد. البته با توجه به این که الگوی گمشدگی را می‌توان کاملاً تصادفی در نظر گرفت و نیز با توجه به محاسبات مربوط به

حجم نمونه لازم برای رگرسیون لجستیک، تعداد داده‌ها جهت انجام تحلیل‌ها کافی به‌نظر می‌رسند و نتایج نیز نشان‌دهنده قابل قبول بودن دقت برآوردها است، لذا داده‌های گمشده از تحلیل حذف شدند و جانهی انجام نشد. در صورتی که توان مدل SVM حاصل را در اختیار پژوهش‌گران قرار داد (مثل ایجاد یک سایت برای مدل مربوطه)، به سادگی می‌توان با دادن اطلاعات مورد نیاز مدل شامل مقادیر مربوط به متغیرهای پیش‌بین، وضعیت پیامد را با دقت مشخصی پیش‌بینی یا طبقه‌بندی کرد.

نتیجه‌گیری

می‌توان اظهار نمود زمانی که هدف صرفاً طبقه‌بندی (Classification) یا پیش‌ینی یک متغیر پاسخ کمی (Regression) باشد بدون نیاز به دانستن جزئیات مدل از جمله نحوه و شکل ارتباط بین متغیرها، مدل SVM کارایی بالایی دارد و در صورت داشتن تنظیمات مناسب برای پارامترهای این مدل، قدرت پیش‌بینی و طبقه‌بندی خوبی برای داده‌های جدید خواهد داشت.

به‌عنوان توصیه بهداشتی، می‌توان گفت که در بیماران دیابتی، کنترل و مراقبت دیابت می‌تواند باعث کاهش محدودیت در عملکرد روزانه و ناتوانی بیماران دیابتی گردد.

تشکر و قدردانی

نگارندگان بر خود لازم می‌دانند مراتب تشکر صمیمانه خود را از معاونت پژوهشی دانشگاه علوم پزشکی تبریز به خاطر قبول هزینه‌ها و از بیماران به خاطر شرکت در پژوهش و از مرکز پزشکی مبتنی بر شواهد به خاطر مشاوره در نگارش مقاله به‌عمل آورند.

References

[1] Yaghoubi A, Tabrizi J-S, Mirinazhad M-M, Azami S, Naghavi-Behzad M, Ghojazadeh M. Quality of life in cardiovascular patients in Iran and factors affecting it: A systematic review. *(JCVTR)* 2012; 4(4): 95.

[2] Shahrbabaki PM,Nouhi E, Kazemi M, Ahmadi F. Family Support as a Reliable Resource for Coping in Patients with Heart Failure. *Acta Medica Mediterranea* 2016;32:873-8. [Farsi]

[3] Khatib ST, Hemadneh MK, Hasan SA, Khazneh E, Sa'ed HZ. Quality of life in hemodialysis diabetic patients: a multicenter cross-sectional study from Palestine. *BMC nephrology* 2018;19(1):49

[4] Solli O, Stavem K, Kristiansen IS. Health-related quality of life in diabetes: The associations of complications with EQ-5D scores. *Health and quality of life outcomes* 2010; 8(1): 1

[5] Ose D, Wensing M, Szecsenyi J, Joos S, Hermann K, Miksch A. Impact of primary care–based disease management on the health-related quality of life in patients with type 2 diabetes and comorbidity. *Diabetes care* 2009; 32(9): 1594-6

[6] Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical functioning items in the medical outcomes study. *Med Care* 2007; 45(5 Suppl 1): S32-8

[7] Sarbakhsh P, Jafarabadi MA, Shamshirgaran SM, Jahangiry L. Identifying Factors Associated with Functional Limitation Among Diabetic Patients in Northwest of Iran: Application of the Generalized Additive Model. *International journal of endocrinology and metabolism* 2018;16(2)

[8] Lu FP, Lin KP, Kuo HK. Diabetes and the risk of multi-system aging phenotypes: a systematic review and meta-analysis. *PLoS One* 2009; 4(1): e4144

[9] Mamaghanian A, Shamshirgaran SM, Aiminisani N, Aliasgarzadeh A. Clinico-epidemiological factors of health related quality of life among people with type 2 diabetes. *(WJD)* 2017;8(8):407

[10] Zhu Y, Fish AF, Li F, Liu L, Lou Q. Psychosocial factors not metabolic control impact the quality of life among patients with type 2 diabetes in China. *Acta diabetologica* 2016; 53(4): 535-41

[11] Fatahi M. Suport Vector Machines: A survey [dissertation]. [razi]: razi university: 2015. 48p

[12] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *(CGP)* 2018; 15(1): 41-5

[13] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research. *(CSBJ)* 2017

[14] Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC medical informatics and decision making* 2010; 10(1): 16

[15] Ataei J, Shamshirgaran SM, Iranparvar Alandari M, Safaeian AR. Evaluation of diabetes quality of care based on a care scoring system among people referring to diabetes clinic in Ardabil, 2014. *Journal of Ardabil University of Medical Sciences* 2015; 15(2): 207-19. [Farsi]

[16] Shamshirgaran SM, Mamaghanian A, Aliasgarzadeh A, Aiminisani N, Iranparvar-Alamdari M, J A. Age differences in diabetes-related complications and glycemic control. Under press. *BMC Endocrine Disorders* Under press 2017; 17(1): 25

[17] Shamshirgaran SM, Ataei J, Malek A, Iranparvar-Alamdari M, Aminisani N. Quality of sleep and its determinants among people with type 2 diabetes mellitus in Northwest of Iran. *(WJD)* 2017; 8(7): 358

[18] Skevington S.M, Lotfy M, O'Connell K.A. The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A Report from the WHOQOL Group. *Quality of Life Research* 2004; 13(2): 299-310

[19] Banihashemi M, Hafezi M, Nasiri-Toosi M, Jafarian A, Abbasi MR, Arbabi M, et al. Psychosocial Status of Liver Transplant Candidates in Iran and Its Correlation with Health-Related Quality of Life and Depression and Anxiety. *Journal of transplantation* 2015: 329615

[20] Montazeri A, Goshtasbi Azita, Vahdaninia M.A.Sthe Short Form Health Survey (Sf-36): Translation And Validation Study Of The Iranian Version. *PAYESH* 2006; 5(1):49-56

[21] Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical functioning items in the medical outcomes study. *Medical care* 2007; 45(5): S32-S8

[22] Gubhaju L, Banks E, MacNiven R, McNamara BJ, Joshy G, Bauman A, et al. Physical

Functional Limitations among Aboriginal and Non-Aboriginal Older Adults: Associations with Socio-Demographic Factors and Health. *PLOS ONE* 2015; 10(9): e0139364

[23] Shamshirgaran SM, Jorm L, Lujic S, Bambrick H. Health related factors among people with type 2 diabetes by country of birth: result from the 45 and Up Study. *Primary care diabetes* 2019; 13(1): 71-81

[24] Mokhtari MH, Najafi A. Comparison of Suport Vector Machine and Artificial Neural Networks classification methods in Land Extracting from Landsat TM Satellite Image. *(JWSS)* 2015; 19(72): 35-45

[25] Ahadi B, Majd HA, Khodakarim S, Rahimi F, Kariman N, Khalili M, et al. Using support vector machines in predicting and classifying factors affecting preterm delivery. *(JPS)* 2016; 7(3): 37-42

[26] Son Y-J, Kim H-G, Kim E-H, Choi S, Lee S-K. Application of support vector machine for prediction of medication adherence in heart failure patients. *(HIR)* 2010; 16(4): 253-9

[27] Huang C-L, Chen M-C ,Wang C-J. Credit scoring with a data mining approach based on support vector machines. *Expert systems with applications* 2007; 33(4): 847-56

[28] Bellotti T, Crook J. Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications* 2009; 36(2): 3302-8

[29] Zhou Y, Zhang R, Wang S, Wang F. Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy. *Sensors* 2018; 18(7)

[30] Sarbakhsh P, Gavgani LF, Jafarabadi MA, Shamshirgaran SM. Detection of Functional Limitation in Diabetic Patients Based on the Optimal Combination of Care Indicators Using Ramp AUC and Comparing its Performance With the Existing Methods. *(CJMB)* 2018; 5(2): 149-54

[31] Fong Y, Yin S, Huang Y. Combining biomarkers linearly and nonlinearly for classification using

case–control studies. *Biostatistics* 2016; 17(3): 499-522

[32] Moyo S, Doan TN, Yun JA, Tshuma N. Application of machine learning models in predicting length of stay among healthcare workers in underserved communities in South Africa. *Human resources for health* 2018; 16(1): 68

[33] Huang Y. Evaluating and comparing biomarkers with respect to the area under the receiver operating characteristics curve in two-phase the area under the ROC curve. *Statistics in Medicine* 2016; 35(21): 3792-809

[34] Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *(CJIM)* 2013; 4(2): 627

[35] Rafeh R, Arbabi M. Data Mining Techniques to Diagnose Diabetes Using Blood Lipids. *(SJIMU)* 2015; 23(4): 239-47. [Farsi]

Application of Support Vector Machine for Detection of Functional Limitations in the Diabetic Patients of the Northwest of IRAN in 2017: A Descriptive Study

L. Faraji Gavgani[۱], P. Sarbakhsh[۲], M. Asghari Jafarabadi[۳], M. Shamshirgaran[۴]

Received: 02/01/2019 Sent for Revision: 03/03/2019 Received Revised Manuscript: 08/12/2019 Accepted: 18/12/2019

Background and Objectives: Support vector machine (SVM) is a robust and effective statistical method for the diagnosis and prediction of clinical outcomes based on combinations of predictor variables. The aim of this study was to use SVM to detect the functional limitations in the diabetic patients and evaluate the accuracy of this diagnosis.

Materials and Methods: This descriptive study was conducted on 378 diabetic patients referred to the diabetic centers of Ardabil and Tabriz in 2014-2015. To classify the diabetic patients in terms of functional limitation, based on the demographic and clinical variables, SVM was used with RBF (radial basis function) kernel and the training and test validation method. Evaluation was performed based on diagnostic indices including sensitivity, specificity, accuracy and area under the ROC (receiver operating characteristic) curve.

Results: The results of SVM method showed that the classification accuracy, sensitivity, specificity of the SVM method in differentiating and correct diagnosis of functional limitations in the diabetic patients were 99%, 100% and 97%, respectively. The area under the ROC curve as the detection performance analysis of this model was 0.98.

Conclusion: In this study, SVM was used to classify the functional limitation status of the diabetic patients, and the results showed that the model had an acceptable performance. Considering the importance of classifying the medical outcomes correctly based on the combinations of predictor variables, the use of the methods such as SVM that are able to find optimal combinations could be helpful.

Key words: Data mining, SVM, Functional limitation, Classification, Kernel function.

Funding: This research was funded by Evidence-Based Medicine Center, Tabriz University of Medical Sciences.

Conflict of interest: None declared.

Ethical approval: The Ethics Committee of Tabriz University of Medical Sciences approved the study. (TBZMED.REC.1395.794).

How to cite this article: Faraji Gavgani L, Sarbakhsh P, Asghari Jafarabadi M, Shamshirgaran M. Application of Support Vector Machine for Detection of Functional Limitations in the Diabetic Patients of the Northwest of IRAN in 2015: A Descriptive Study. *J Rafsanjan Univ Med Sci* 2020; 18 (12): 1270-1286. [Farsi]

1- MSc in Biostatistics, Evidence-Based Medicine Center, Tabriz University of Medical Sciences, Tabriz, Iran, ORCID: 0000-0002-5701-0782

2- Assistant Prof., Dept. of Statistics and Epidemiology, Health Faculty, Tabriz University of Medical Sciences, Tabriz, Iran, ORCID: 0000-0002-4213-5152

(Corresponding Author) Tel: (041) 33357581, Fax: (041) 33342219, E-mail: p.sarbakhsh@gmail.com

3- Associate Prof., Dept. of Statistics and Epidemiology, Health Faculty, Tabriz University of Medical Sciences, Tabriz, Iran, ORCID: 0000-0003-3284-9749

4- Associate Prof., Dept. of Statistics and Epidemiology, Tabriz University of Medical Sciences, Tabriz, Iran, ORCID: 0000-0003-1719-2822