مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره پنجم شماره دوم. ص ۲۵۲-۲٦٤، ۱۳۹۷

مقاله پژوهشی

مقایسه عملکرد الگوریتم‌های داده‌کاوی در پیش‌بینی بیماری عروق کرونر

هاله آیت‌اللهی[۱]، لیلا غلامحسینی[۲،۳]*، مسعود صالحی[٤]

•دریافت مقاله: ۹۷/۲/۲ •پذیرش مقاله: ۹۷/۴/۲۵

مقدمه: بیماری‌های قلبی- عروقی نخستین علت مرگ در جهان هستند و براساس برآورد سازمان بهداشت جهانی، مرگ ناشی از بیماری‌های قلبی تا سال۲۰۳۰ به ۲۳ میلیون مورد افزایش خواهد یافت. از این رو، به نظر می‌رسد استفاده از الگوریتم‌های داده‌کاوی برای پیش‌بینی بیماری عروق کرونر قلب بسیار کاربردی باشد. هدف از پژوهش حاضر مقایسه عملکرد الگوریتم‌های شبکه عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) در پیش‌بینی بیماری عروق کرونر قلب بود.

روش: پژوهش حاضر از نوع توصیفی-تحلیلی و نمونه پژوهش شامل تمام بیماران بستری مبتلا به بیماری عروق کرونر قلب در سه بیمارستان تابعه دانشگاه علوم پزشکی آجا بین سال‌های ۱۳۹۵ تا ۱۳۹۶ بود. درمجموع، ۱۳۲۴ رکورد با ۲۶ ویژگی مؤثر در این بیماری استخراج و پس از نرمال‌سازی و پاک‌سازی داده‌ها، در نرم‌افزار SPSS نسخه ۲۳ و Excel نسخه ۲۰۱۳ وارد شدند. برای قالب‌بندی داده‌ها نیز از نرم‌افزار داده‌کاوی R3.3.2 استفاده گردید.

نتایج: الگوریتم ماشین بردار پشتیبان با میانگین درصد خطای مطلق پایین‌تر (۱۱۲/۰۳)، آماره هاسمر-لمشو بالاتر (۱۶/۷۱)، حساسیت (۹۲/۲۳) و ویژگی (۷۴/۴۲) نسبت به مدل شبکه عصبی دقیق‌تر بود. همچنین، مساحت زیر منحنی راک در الگوریتم SVM بیشتر از ANN بود و می‌توان نتیجه گرفت که این مدل دارای دقت بیشتری است.

نتیجه‌گیری: در این مطالعه، الگوریتم SVM نسبت به مدل شبکه عصبی دقت و عملکرد بهتری در پیش‌بینی بیماری عروق کرونر قلب نشان داد و دارای حساسیت و صحت بالاتری بود. با این حال پیشنهاد می‌گردد که نتایج مطالعه حاضر با یافته‌های حاصل از به‌کارگیری سایر الگوریتم‌های داده‌کاوی در پژوهش‌های آتی مورد مقایسه قرار گیرد.

کلید واژه‌ها: بیماری عروق کرونر، الگوریتم‌های داده‌کاوی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان

•**ارجاع:** آیت‌اللهی هاله، غلامحسینی لیلا، صالحی مسعود. مقایسه عملکرد الگوریتم‌های داده‌کاوی در پیش‌بینی بیماری عروق کرونر. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۷؛ ۵(۲): ۲۶۴-۲۵۲

۱. دکترای انفورماتیک پزشکی، دانشیار گروه مدیریت اطلاعات سلامت، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران
۲. دانشجوی دکترای تخصصی مدیریت اطلاعات سلامت، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران
۳. مربی، گروه فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی آجا، تهران، ایران
۴. دانشیار آمار زیستی، گروه آمارحیاتی، دانشکده بهداشت عمومی، دانشگاه علوم پزشکی ایران، تهران، ایران
* **نویسنده مسئول**: تهران، خیابان ولیعصر، بالاتر از ونک، خیابان شهید رشید یاسمی، رو به روی بیمارستان خاتم الانبیاء، دانشکده مدیریت و اطلاع‌رسانی پزشکی
• **شماره تماس:** ۰۲۱۸۸۷۹۴۳۰۲ • **Email:** gholamhosseini.l@tak.iums.ac.ir

[Downloaded from jhbmi.com on 2024-10-12]

مقدمه

در طی چند دهه اخیر، شیوع بیماری‌های قلبی- عروقی در جهان رو به افزایش بوده است، به طوری‌که اکنون آن را به‌عنوان شایع‌ترین علت اصلی مرگ و ناتوانی افراد بیان کرده‌اند [۱]. این بیماری به‌معنای هر وضعیت یا بیماری است که قلب و عروق خونی آن را درگیر می‌نماید [۲]. دستگاه گردش خون را تحت تأثیر قرار داده و بیماری‌های عروقی مغز، کلیه و بیماری‌های شریانی را نیز شامل می‌شود [۳]. به‌طورکلی می‌توان گفت که طیف بالینی این بیماری‌ها از کم‌خونی بدون علامت تا آنژین صدری پایدار مزمن، آنژین ناپایدار، انفارکتوس حاد میوکارد، بیماری ایسکمیک عضله قلب و مرگ ناگهانی متغیر است [۴] و گاهی با وضعیت‌هایی مانند پرفشاری خون، بیماری عروق کرونری قلب کرونری (Coronary Artery Disease) CAD ، نارسایی مزمن قلب و بیماری عروق قلبی نیز همراه است [۵]. در سال‌های اخیر، علاوه بر عوامل خطرساز اصلی، عوامل دیگری مانند عفونت، التهاب و بیماری‌های مزمن به عنوان سایر عوامل خطر در بیماری‌های قلبی مطرح شده‌اند [۶].

در ابتدای قرن بیستم،۱۰ درصد از کل مرگ‌ها به علت بیماری‌های قلبی بوده‌اند. در پایان همین قرن نیز، موارد مرگ ناشی از این بیماری‌ها به ۲۵ درصد افزایش یافته و پیش‌بینی می‌شود که با توجه به روند رو به رشد کنونی تا سال ۲۰۲۵ میلادی بیش از ۳۵ تا ۶۰ درصد از مرگ‌های جهان به‌دلیل بیماری‌های قلبی- عروقی روی دهند [۷]. طبق گزارش سازمان بهداشت جهانی در سال ۲۰۱۷، بیش از نیمی (۵۴ درصد) از موارد مرگ در سراسر جهان به ۱۰ علت اصلی بوده و بیماری‌های قلبی که منجر به ۱۵ میلیون مرگ در سال ۲۰۱۵ شده، به عنوان بزرگ‌ترین گروه از بیماری‌های مرگ‌آفرین به‌حساب می‌آیند [۸]. بیماری‌های قلبی هر ساله میلیون‌ها نفر را به کام مرگ می کشند، و اگر تا سال ۲۰۲۰ اقدامات پیشگیرانه‌ای در این زمینه انجام نگیرد، این رقم به ۲۴/۸ میلیون نفر افزایش خواهد یافت [۹].

براساس گزارش وزارت بهداشت کشور ایران، ۳۹/۹ درصد از علل مرگ ایرانیان به علت بیماری‌های قلبی است و عوامل خطرساز این بیماری‌ها به شدت درحال گسترش هستند و در این میان بیماری عروق کرونری قلب شایع‌ترین نوع بیماری قلبی است [۱۰]. این بیماری چندعلّتی است و یک سری از عوامل خطرساز آن مانند افزایش کلسترول، پرفشاری خون، دیابت و مصرف سیگار بیشتر مورد تأکید قرار گرفته‌اند [۱۱].

با توجه به نتایج یک مطالعه اپیدمیولوژیک که به بررسی علل مرگ پرداخته، از ۳۲۱۵۷۰ مورد فوت در سال ۱۳۸۸، تعداد ۸۲۳۰۷ مورد ناشی از بیماری عروق کرونری قلب بوده و با میزان ۲۵/۶ درصد اولین عامل مرگ در ایران به شمار می‌آید [۱۲]. براساس یافته‌های مطالعات پیشین، شیوع بیماری عروق کرونر در مردان بیش از زنان است و علائم بیماری در زنان به میزان متوسط حدود ۱۰ سال دیرتر از مردان ظهور می یابد [۱۳]؛ بنابراین با توجه به رشد چشمگیر بیماری‌های قلبی که هزینه بالایی بر جامعه تحمیل می‌کند، جوامع پزشکی درصدد یافتن راهی برای پیش‌بینی دقیق و زودهنگام هستند تا با انجام اقدامات پیشگیرانه از بروز این گونه بیماری‌ها پیشگیری نموده و هزینه‌های درمان را کاهش دهند [۱۴]. گفتنی است که حوزه بهداشت و درمان سرشار از اطلاعات است، ولی متأسفانه داده‌های لازم برای تصمیم‌گیری مؤثر و کشف الگوهای پنهان از آن استخراج نمی‌شود. با استخراج داده‌های مفید و کشف دانش از میان حجم انبوه داده‌های پزشکی می‌توان به شناسایی عوامل بروز، رشد یا انتشار بیماری پرداخت و اطلاعات ارزشمندی را برای شناسایی روش‌های تشخیص، پیش‌بینی و درمان بیماری‌ها در اختیار پزشکان قرار داد [۱۵]؛ بنابراین بسیاری از مراکز بهداشتی- درمانی که به دنبال راه‌حل‌های عملی برای کشف دانش از پایگاه داده پزشکی هستند، فرآیند داده‌کاوی را با هدف اتخاذ تصمیم‌های مؤثر و کارآ به‌کار می‌گیرند [۱۶]. به کارگیری فرایند داده‌کاوی برای تشخیص و پیش‌بینی بیماری‌های قلبی، یکی از زمینه‌های مورد علاقه محققان است زیرا، الگوریتم‌های مفید و کاربردی داده‌کاوی می‌توانند به سرعت ویژگی‌های هر بیماری را تفسیر نموده و به‌عنوان سیستم‌های پشتیبان تصمیم به پزشکان یاری رسانند [۱۷].

دانش نوین داده‌کاوی به‌عنوان یکی از ۱۰ دانش درحال توسعه، دهه آینده را با تحولات عظیم تکنولوژیک مواجه کرده و با برخورداری از زمینه‌های تخصصی دارای کاربردهای بسیار وسیعی در حوزه علوم پزشکی است. درحال حاضر، استفاده از روش‌های سنتی تشخیص بیماری‌های قلبی- عروقی از کارایی لازم برخوردار نبوده و بنابراین، پیش‌بینی دقیق و سریع این بیماری به منظور کاهش هزینه‌ها و پیشگیری از عوارض جبران‌ناپذیر آن ضروری به نظر می‌رسد. با توجه به مطالعات گوناگونی که در حوزه داده‌کاوی بیماری‌های قلبی- عروقی انجام شده می‌توان دریافت که تا به‌حال از الگوریتم‌های متفاوتی مانند خوشه‌بندی، طبقه‌بندی، رگرسیون و قوانین محدود انجمنی، انواع

[Downloaded from jhbmi.com on 2024-10-12]

درخت تصمیم، شبکه بیزی، شبکه عصبی مصنوعی (پرسپترون چندلایه با الگوریتم پس‌انتشار خطا) (شکل ۱) یا (الگوریتم شیب توأم مقیاس شده) و ماشین بردار پشتیبان (شکل ۲) استفاده شده است، ولی مقایسه‌ای برای تعیین دقت پیش‌بینی این الگوریتم‌ها صورت نگرفته است [۱۸-۳۳]؛ بنابراین هدف از این مطالعه مقایسه دقت پیش‌بینی بیماری عروق کرونری قلب با استفاده از الگوریتم‌های شبکه عصبی مصنوعی (Artificial Neural Network) ANN و ماشین بردار پشتیبان (Support Vector Machines) SVM بوده و وجه تمایز آن در به‌کارگیری این الگوریتم‌ها برای پیش‌بینی بیماران عروق کرونری در بیمارستان‌های نظامی بود.

روش

پژوهش حاضر از نوع کاربردی بود که به روش توصیفی- تحلیلی و با استفاده از فرآیند داده‌کاوی انجام شد. محیط پژوهش ۳ بیمارستان منتخب تابعه دانشگاه علوم پزشکی آجا بودند. در این پژوهش از میان انواع بیماری‌های قلبی تنها بیماران مبتلا به عروق کرونری قلب بین سال‌های ۱۳۹۵ تا ۱۳۹۶ انتخاب و وارد مطالعه شدند؛ بنابراین نمونه‌گیری به‌صورت سرشماری داده‌های موجود و حجم نهایی نمونه ۱۳۲۴ مورد و شامل تمامی بیماران بستری شده در بخش قلب و عروق بود. ابزار گردآوری داده‌ها شامل یک چک‌لیست بود که براساس منابع موجود و راهنماهای بیماری قلبی- عروقی و مقالات معتبر تهیه و طراحی گردید. این چک‌لیست شامل کلیه ویژگی‌ها و فیلدهای موردنیاز برای پیش‌بینی بیماری عروق کرونری قلب بود. در مرحله بعد با مراجعه حضوری به محیط‌های پژوهشی داده‌های موردنیاز از پرونده بیماران استخراج و در چک لیست مربوطه وارد شد. به‌منظور پاک‌سازی داده‌ها، در ابتدا فیلدهای خالی پرونده با بررسی پرونده پزشکی بیماران عروق کرونر قلبی مقداردهی گردید. رکوردهایی که همچنان در این مرحله دارای مقادیر از دست‌رفته بودند، از مطالعه حذف شدند. پس از آن، برخی از رکوردهایی که دارای داده‌های پرت بودند، به روش محاسبه میانگین و استنتاج از داده‌های همنوع بهینه شدند و رکوردهایی که به هیچ وجه تشابهی با سایر داده‌ها نداشتند، حذف شدند. سپس، داده‌ها در نرم‌افزار اکسل ۲۰۱۳ وارد شده و مورد تأیید متخصصان قرار گرفت. در نهایت، پس از آماده‌سازی، پردازش و پاک‌سازی داده‌ها، تعداد ۲۰۶۵ رکورد اولیه بیماران به ۱۳۲۴ رکورد کاهش یافت.

به‌منظور انجام آزمون‌های آماری از نرم افزار SPSS نسخه ۲۳ استفاده شد و برای قالب‌بندی داده‌ها نیز، نرم‌افزار داده‌کاوی R3.3.2 مورد استفاده قرار گرفت. در ادامه، مجموعه داده به دو مجموعه آموزشی و آزمایشی تقسیم گردید. برای این کار، از روش استاندارد تصادفی تقسیم‌بندی داده‌های تحت مطالعه استفاده شد؛ بنابراین ۷۰ درصد از رکوردها برای آموزش یا ساخت مدل و ۳۰ درصد نیز برای تست یا آزمون مدل‌ها استفاده شدند و پس از مدل‌سازی، ارزیابی نتایج انجام گرفت. در نهایت، براساس فرآیند داده‌کاوی و با دانش کشف شده از اجرای تکنیک‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان، دقت پیش‌بینی بیماری عروق کرونری قلب در بیمارستان‌های مورد مطالعه تعیین و مقایسه شد.

نتایج

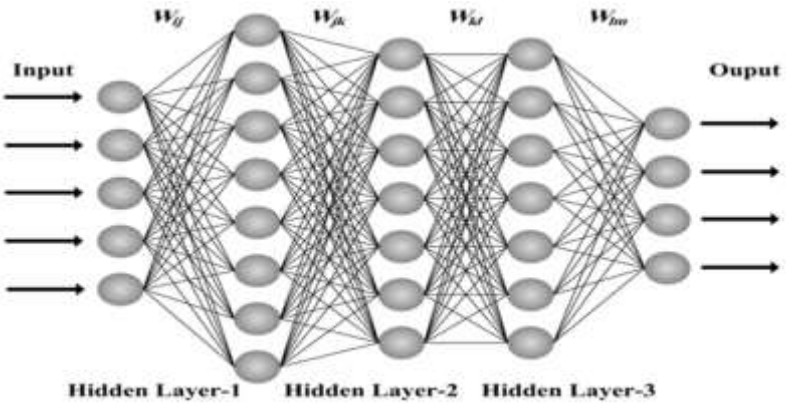
در ابتدا، براساس متغیرهای پژوهش در بیمارستان‌های مورد مطالعه به تفکیک مورد تحلیل قرار گرفت. براساس نتایج به دست آمده اکثر بیماران مرد (۶۲/۷٪، n=۸۲۹) و میانگین سنی آنان بین ۶۲-۵۴ سال بود و سایر بیماران را زنان (۳۷/۳٪، n=۴۹۴) تشکیل می‌دادند که میانگین سنی آنان بین ۶۴-۶۱ سال بود. براساس مقایسه میانگین وزنی بیماران و غیربیماران در بیمارستان‌های مورد مطالعه، در دو بیمارستان A و بیمارستان B، بین بیماری CAD (Coronary Artery Disease) و میانگین وزن اختلاف معنی‌دار (P<۰/۰۰۱) مشاهده شد. توزیع فراوانی بیماری عروق کرونری قلب برحسب وضعیت شغلی در بیمارستان‌های مورد مطالعه نشان‌دهنده آن بود که در هر ۳ بیمارستان بین داشتن یا نداشتن بیماری، اختلاف معنی‌دار وجود داشت. توزیع فراوانی بیماری عروق کرونری قلب برحسب محل سکونت حاکی از آن بود که اکثریت بیماران (۹۸٪، n=۱۰۸۲) را ساکنین شهرها تشکیل می‌دادند. توزیع فراوانی بیماری عروق کرونری قلب بر حسب وجود سابقه خانوادگی بیماری نیز نشان‌دهنده این بود که در تمامی بیمارستان‌ها اختلاف معنی‌داری (P<۰/۰۰۱) بین داشتن یا نداشتن سابقه خانوادگی بیماری (۷۹/۳٪، n=۱۰۴۹) وجود داشت.

یافته‌ها نشان داد که (۷۷/۳٪،n=۱۰۲۴) از افراد غیرسیگاری بودند و در همه بیمارستان‌ها اختلاف معنی‌داری (P<۰/۰۰۱) بین سیگاری بودن و بیماری CAD وجود داشت و همچنین، (۷۱/۴٪،n=۹۴۴) از افراد فاقد هرگونه بیماری همراه بودند.

هدف اصلی این پژوهش، تعیین دقت پیش‌بینی بیماری عروق کرونری قلب با استفاده از الگوریتم شبکه عصبی مصنوعی و

[Downloaded from jhbmi.com on 2024-10-12]

مقایسه آن با ماشین بردار پشتیبان بود. در ابتدا، تعداد ۲۶ متغیر استخراج شده از پایگاه داده بیماران قلبی بیمارستان‌های مورد مطالعه به عنوان پارامترهای لایه ورودی تعیین و با اجرای الگوریتم‌ها در برازش مدل شبکه عصبی مصنوعی سه لایه، وزن هر یک از آن‌ها محاسبه شد. سپس، با توجه به وزن محاسبه شده متغیرهای جنس، شغل، سابقه خانوادگی، مصرف سیگار، بیماری همراه، میانگین ضربان قلب، وضعیت موج تی.اس.تی، سابقه فشارخون بالا، درد قفسه سینه، کلسترول، تری‌گلیسیرید، قندخون و کراتینین به عنوان متغیرهای تأثیرگذار بر بیماری عروق کرونری قلب تعیین شدند.



شکل ۱: نمایی از پرسپترون چندلایه در شبکه عصبی مصنوعی

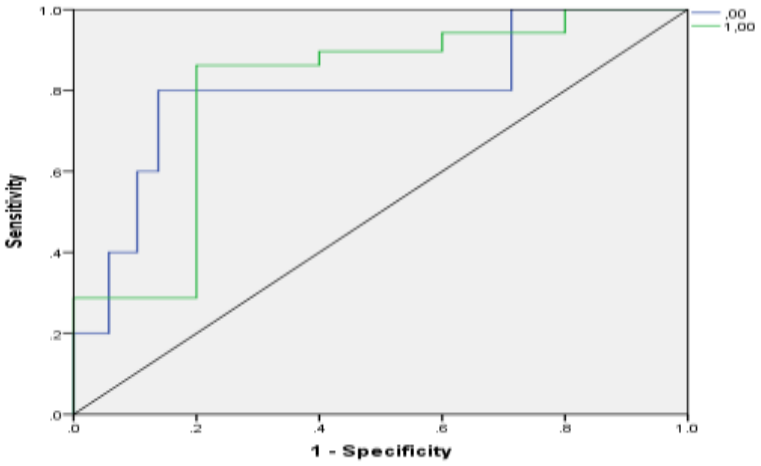
در مطالعه حاضر، به‌منظور اجرای مدل شبکه عصبی مصنوعی ۷۰ درصد از داده‌ها آموزشی و ۳۰ درصد آزمایشی در نظر گرفته شدند. در نتیجه، مشخص شد که با توجه به پایین‌تر بودن میانگین مربعات خطا و خطای نسبی در مجموعه آزمایشی، برازش داده‌ها با استفاده از مدل شبکه عصبی مصنوعی پرسپترون چندلایه مناسب و دقت آن برابر با ۰/۷۹۸ بود.

جدول ۱: شاخص‌های دقت مدل در دو مرحله آموزشی و آزمایشی در الگوریتم شبکه عصبی مصنوعی

	نمونه‌گیری	میانگین مربعات خطا	خطای نسبی	دقت
آموزشی	۷۰ درصد	۵/۳۹	۰/۰۰۲	۰/۷۹۸
آزمایشی	۳۰ درصد	۳/۸۴	۰/۰۰۲	

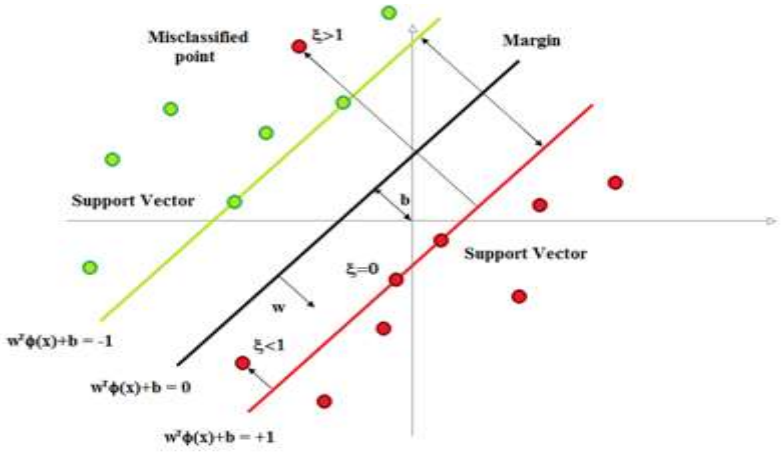
نمودار ۱ نشان‌دهنده منحنی مشخصه عملکرد (Receiver Operating Characteristic curve) ROC برای آزمون افراد سالم و بیمار است. هر عدد در این نمودار نسبت نرخ پیش‌بینی درست به نرخ پیش‌بینی غلط مدل را نشان می‌دهد. دقت مدل بستگی به این دارد که چقدر مدل آزمون توانسته است به‌درستی افراد را به سالم و بیمار (نرخ پیش‌بینی درست) تفکیک نماید. این دقت از محاسبه با استفاده از ناحیه سطح زیر نمودار ROC محاسبه می‌شود، هرچه مقدار این مساحت به ۱ نزدیک‌تر باشد، نشان‌دهنده دقت بالاتر مدل و هرچه به ۰/۵ نزدیک‌تر باشد، نشان‌دهنده دقت کمتر مدل است. همچنین، هرچه مقدار این نسبت به گوشه سمت چپ نزدیک‌تر باشد، نشان‌دهنده بیشتر بودن مساحت زیر منحنی است؛ بنابراین با توجه به نمودار ۱ در تشخیص بیماری براساس مدل شبکه عصبی مصنوعی مشخص شد که این مدل در پیش‌بینی صحیح افراد بیمار از دقت بالایی برخوردار است.

[Downloaded from jhbmi.com on 2024-10-12]



نمودار ۱: منحنی مشخصه عملکرد در تشخیص بیماری براساس مدل شبکه عصبی مصنوعی

در مرحله بعدی از این مطالعه، دقت پیش‌بینی بیماری عروق کرونری قلب با استفاده از الگوریتم‌های ماشین بردار پشتیبان محاسبه شد. سپس، شاخص‌های دقت مدل در دو مرحله آموزشی و آزمایشی در الگوریتم ماشین بردار پشتیبان مقایسه شدند.



شکل ۲: نمایی از ماشین بردار پشتیبان

جدول ۲: شاخص‌های دقت مدل مراحل آموزشی و آزمایشی در الگوریتم ماشین بردار پشتیبان

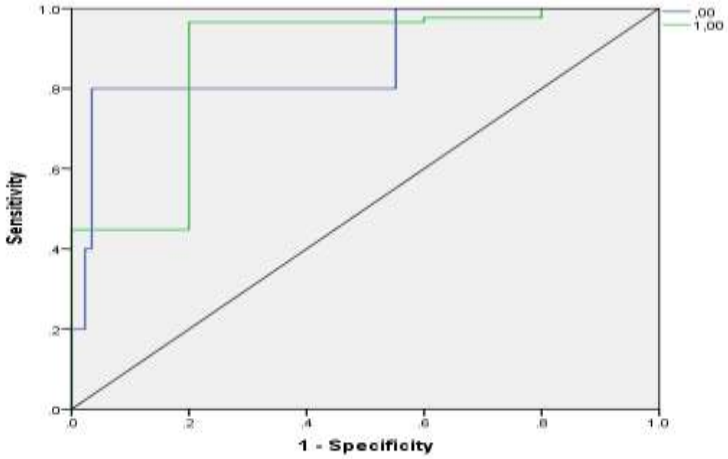
	نمونه‌گیری	مقدار F*	ضریب کاپا**	دقت***
آموزشی	۷۰ درصد	۰/۷۶۱	۰/۷۰۶	۰/۸۷۱
آزمایشی	۳۰ درصد	۰/۶۹۶	۰/۶۳۶	

*F-measure =2 * ((Precision * Recall) / (Precision + Recall))
** Cohen's Kappa coefficients = (Accuracy - expAccuracy) / (1- expAccuracy)
***Precision=TP/TP+FP

در این مرحله از شاخص‌های مقدار F و ضریب کاپا برای تعیین دقت مدل استفاده شد. با توجه به مقدار F، مدل SVM از قدرت و حساسیت متوسط رو به بالایی برای تشخیص بیماران عروق کرونری قلب برخوردار بود. همچنین، با توجه به مقادیر شاخص‌های مدل SVM و مقادیر کاپا در دوره آموزشی و آزمایشی، مدل SVM در زمینه طبقه‌بندی افراد به دو گروه سالم و بیمار، نسبت به حالت تصادفی از دقت بالاتری برخوردار بود. سپس، با مقایسه شاخص صحت مشخص شد که مدل

[Downloaded from jhbmi.com on 2024-10-12]

SVM نسبت به مدل شبکه عصبی از صحت بالاتری برخوردار بوده و طبقه‌بندی بهتری را ارائه کرده است.
از سوی دیگر، با توجه به نمودار ۲ می‌توان دریافت که مساحت سطح زیر منحنی راک در الگوریتم SVM بیشتر از مدل شبکه عصبی است و در نتیجه مدل SVM در تشخیص درست افراد بیمار نسبت به مدل شبکه عصبی بهتر عمل نموده و از دقت بالاتری برخوردار است.



نمودار ۲: منحنی مشخصه عملکرد در تشخیص بیماری براساس مدل ماشین بردار پشتیبان

به منظور مقایسه دو الگوریتم شبکه عصبی مصنوعی و ماشین بردار پشتیبان از شاخص‌های میانگین درصد خطای مطلق MAPE(Mean Absolute Percentage Error)، آزمون نیکویی برازش هاسمر-لمشو، (Hosmer-lemeshow test) حساسیت و ویژگی استفاده شد (جدول ۳).

جدول۳: شاخص‌های مقایسه برای بررسی پیش‌بینی مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان

شاخص	ANN	SVM
میانگین درصد خطای مطلق	۱۲۵/۱۷	۱۱۲/۰۳
هاسمر- لمشو	۱۲/۴	۱۶/۷۱
حساسیت	۸۸/۰۱	۹۲/۳۲
ویژگی	۷۳/۶۴	۷۴/۴۲

باتوجه به یافته‌های پژوهش، میزان پایین میانگین درصد خطای مطلق نشان‌دهنده این بود که الگوریتم SVM با خطای کمتری به برازش داده‌ها می‌پردازد. همچنین، مقدار بیشتر آزمون نیکویی برازش هاسمر- لمشو نشان از بهتر بودن الگوریتم SVM دارد. از سویی، الگوریتم SVM با میزان حساسیت (۹۲/۲۳) و ویژگی (۷۴/۴۲) از قدرت و حساسیت بالاتری نسبت به شبکه عصبی مصنوعی برای تشخیص بیماری عروق کرونری قلب برخوردار بود.

بحث و نتیجه‌گیری

بنابر یافته‌های این مطالعه، مهم‌ترین عوامل مؤثر بر ایجاد بیماری CAD جنس، شغل، سابقه خانوادگی، مصرف سیگار، بیماری همراه، میانگین ضربان قلب، وضعیت موج تی.اس.تی، سابقه فشارخون بالا، درد قفسه‌سینه، کلسترول، تری‌گلیسیرید، قندخون و کراتینین بودند. در مطالعات پیشین نیز، عوامل متعددی برای بروز و پیشرفت بیماری‌های قلبی معرفی شده است. این عوامل در ۶ گروه کلی و عبارت بودند از: عوامل زیست- محیطی، عادات زندگی، عوامل خطرساز، بیماری‌های زمینه‌ای، عوامل روانی- شخصیتی و عوامل اجتماعی [۳۴]. از سایر عوامل خطرساز متداول در ایجاد بیماری عروق کرونری قلب نیز

[Downloaded from jhbmi.com on 2024-10-12]

می‌توان به افزایش فشارخون، سبک زندگی [۳۵]، میزان بالای کلسترول خون [۳۶]، دیابت [۳۷]، چاقی [۳۸] و استعمال سیگار [۳۹] اشاره کرد که به عنوان برخی از پیش‌زمینه‌های این بیماری بوده و با درمان مناسب تأثیر سوء این متغیرها کاهش یافته است. در پژوهش حاضر، توزیع بیماری عروق کرونری قلب برحسب جنس نشان داد که میزان بروز بیماری عروق کرونری قلب در مردان بیشتر از زنان بوده و با روند افزایش سن و وزن فرد، احتمال میزان ابتلا به بیماری افزایش پیدا کرده است. نتایج پژوهش حسینیان و همکاران با این مطالعه هم‌راستا و بیانگر این مطلب بود که بین سن بالای ۶۰ سال، جنس مرد و مصرف سیگار رابطه آماری معنی‌داری با بیماری عروق کرونری قلب وجود دارد [۴۰]. در پژوهش Masethe و همکارش نیز، سیستمی برای پیش‌بینی حمله قلبی با صحت ۹۹ درصد ارائه گردید که متغیرهای آن شامل جنس، سن، نوع درد قفسه‌سینه، افزایش ضربان قلب، کلسترول، مصرف سیگار، قندخون، فشارخون، تغییرات ضربان قلب، رژیم غذایی و مصرف الکل بودند [۴۱].

در بررسی ارتباط میان بیماری CAD و وضعیت شغلی افراد مشخص شد که با توجه به آمار بالاتر بیماران عادی، احتمال بروز بیماری عروق کرونری قلب در این افراد بیشتر از کارمندان و بازنشستگان بوده است. از سوی دیگر، یافته‌های حاصل از متاآنالیز در مطالعه Kivimäki و همکاران نشان داد که خطر بروز بیماری قلبی درحدود ۴۰ درصد بوده و افزایش فشار شغلی منجر به دو برابر شدن احتمال بروز بیماری شده است؛ بنابراین نوع شغل به‌عنوان عامل خطر بیماری‌های قلبی مورد بررسی قرار گرفت [۴۲].

طبق مطالعه حاضر، در بررسی ارتباط بین بیماری CAD با محل سکونت مشخص شد که احتمال بروز این بیماری در شهرنشین‌ها بیشتر از روستاییان بوده است. در همین رابطه، دلبندی با بررسی ارتباط آلاینده‌های هوای شهر تهران و مرگ ناشی از بیماری‌های قلبی در افراد بالای ۶۰ سال طی یک دوره مطالعاتی بین سال‌های ۲۰۱۳-۲۰۰۸ بیان داشت که بین این دو رابطه معناداری وجود دارد [۴۳]. Zhang و همکاران نیز در بررسی ارتباط پراکنش فضایی ذرات معلق و میزان مرگ بیماران قلبی در شهر پکن بیان داشتند که افزایش ذرات معلق باعث افزایش مرگ‌های ناشی از بیماری‌های قلبی در شهرنشینان گردید [۴۴].

در مطالعه مشابهی، Vanos و همکاران خطر مرگ ناشی از آلودگی هوا در۱۰ شهر کانادا را ارزیابی کردند. برهمین اساس، آنان دریافتند که رابطه معنی‌داری بین مرگ بیماران قلبی-تنفسی با شهرنشینی و آلاینده‌های هوای شهری وجود دارد [۴۵]. ولیکن، یافته‌های مطالعه آصف‌زاده و همکاران نشان داد که میزان مرگ‌های ناشی از بیماری‌های قلبی در مناطق روستایی با افزایش سن بالا رفته و احتمال مرگ در روستاها نسبت به شهرها دو برابر بوده است [۴۶].

در بررسی توزیع فراوانی بیماری عروق کرونر قلب برحسب سابقه خانوادگی مشخص شد که اختلاف معناداری بین سابقه خانوادگی و بیماری CAD وجود دارد. در مطالعه حاضر، سابقه خانوادگی به‌معنای وجود بیماری عروق کرونر در افراد درجه یک خانواده (پدر، مادر، خواهر، برادر) بود که زمان بروز این بیماری در مردان پیش از ۵۵ سالگی و در زنان پیش از ۶۵ سالگی در نظر گرفته شد [۴۷]. براساس مطالعه بیدل و همکاران، عوامل خطر بیماری‌های قلبی که با انسداد عروق کرونر رابطه داشتند عبارت بودند از: عوامل خطر قابل‌تعدیل مانند سیگار، HDL، کلسترول، فشارخون سیستولیک و دیاستولیک، و عوامل خطر غیرقابل تعدیل مانند سن، جنس و سابقه بیماری قلبی در بستگان درجه یک [۴۸].

براساس یافته‌های مطالعه حاضر، تنها ۲۲/۷ درصد از نمونه‌ها سیگاری و ۷۷/۳ درصد نیز غیرسیگاری بودند. با انجام آزمون کای‌اسکوئر در بیمارستان‌های مورد مطالعه مشاهده شد که بین بیماری CAD و سیگاری بودن اختلاف معنی‌داری ($P<0/001$) وجود دارد، ولی با توجه به درصد بسیار پایین افراد سیگاری می‌توان این گونه نتیجه گرفت که بین سیگاری بودن و احتمال بروز این بیماری رابطه مستقیمی وجود دارد. امروزه مطالعات نشان داده‌اند که استعمال سیگار در کشورهای آسیایی نسبت به سایر کشورهای جهان روند رو به افزایشی داشته است [۴۹]. صفدری و همکاران نیز، متغیرهای فشارخون، چربی بالا و سیگار را به عنوان مؤثرترین عوامل ابتلا به این بیماری معرفی کرده‌اند [۵۰]. مصرف سیگار یک عامل خطر قوی و مستقل برای بیماری‌های قلبی است و تمام بیماران قلبی باید از استعمال سیگار اجتناب کنند [۵۱]. درمان‌گران تأکید کرده‌اند که با کاهش مصرف سیگار خطر ابتلا به CAD می‌تواند در آینده به‌صورت چشمگیری کاهش یابد؛ بنابراین وضعیت مصرف سیگار بیماران قلبی باید به‌طور سیستماتیک ارزیابی گردد [۳۸]. براساس نتایج مطالعه بخشنده و همکاران، مصرف سیگار نقش ویژه‌ای در ایجاد بیماری‌های قلبی داشته و به‌طورکلی، در گروه افراد سیگاری جنس مرد دارای میانگین بیشتری بوده و سن ابتلا به بیماری در آن‌ها پایین‌تر بود [۵۲].

[Downloaded from jhbmi.com on 2024-10-12]

نتایج مطالعه حاضر نشان داد که ۲۸/۶ درصد از افراد مورد مطالعه دارای یک یا چند بیماریِ همراه بودند. برهمین اساس در مطالعات پیشین، بیمارانی که به‌طور همزمان به دو بیماری احتقانی قلب و بیماری مزمن انسدادی ریه مبتلا بودند، در مقایسه با بیمارانی که فقط مبتلا به یکی از این بیماری‌ها بودند، شدیدترین عوارض را به‌همراه داشتند [۵۳]. از سوی دیگر، براساس یافته‌های پژوهش رضا بیگی داورانی و همکاران بیماری دیابت، استعمال دخانیات و سابقه بیماری قلبی کمترین سهم را در بروز بیماری‌های قلبی داشت. بیشترین سهم عوامل خطرساز در بروز بیماری نیز به ترتیب عبارت بودند از: استرس، عدم فعالیت فیزیکی، اضافه وزن یا چاقی [۵۴]. درخصوص ارتباط بین میانگین ضربان قلب و بروز بیماری CAD در بیمارستان‌های مورد مطالعه مشخص شد که بین این دو متغیر اختلاف معنی‌داری وجود دارد و درنتیجه با افزایش میانگین ضربان قلب احتمال بروز این بیماری نیز افزایش می‌یافت.

بررسی توزیع فراوانی بیماری عروق کرونری قلب برحسب موج تی.اس.تی در بیمارستان‌های مورد مطالعه نشان داد که تنها ۴۵ درصد از افراد دارای موج غیرنرمال بودند و به‌طورکلی، وضعیت موج تی.اس.تی ارتباط مستقیمی با بروز این بیماری نداشت. در همین راستا، یافته‌های مطالعه‌ای با عنوان «پیش‌بینی دقیق بیماری عروق کرونری با استفاده از الگوریتم‌های بیوانفورماتیک» نشان داد که ویژگی‌های نتایج اسکن تالیوم، نوع درد قفسه‌سینه، شیب خط ST و تعداد رگ‌های اصلی درگیر دارای بالاترین وزن بوده و از مهمترین متغیرها در تشخیص افراد سالم و بیمار بوده‌اند [۵۵]. براساس یافته‌های پژوهش دیگری که برای تشخیص بیماری‌های دریچه‌ای قلب با استفاده از الگوریتم‌های شبکه عصبی مصنوعی و تحلیل مؤلفه‌های اصلی بیماری انجام شد، کاهش ابعاد سیگنال‌های قلبی اثر مثبتی بر طبقه‌بندی صداهای قلبی داشت [۵۶].

نتایج مطالعه حاضر، در مقایسه میانگین میزان تری‌گلیسیرید و کراتینین خون نشان داد که بین بروز این بیماری و میانگین میزان تری‌گلیسرید و کراتینین خون ارتباط و همبستگی مثبتی بوده و با افزایش این متغیرها میزان ابتلا به بیماری CAD نیز افزایش می‌یافت.

یکی از عوامل مهم در تشخیص CAD، ارزیابی درد قفسه‌سینه است. درهمین راستا، یافته‌های مطالعه نشان دادند که بین درد قفسه‌سینه و وجود بیماری ارتباط معناداری وجود داشت [۳۸]. از سویی، در مطالعه‌ای که با هدف تعیین ارتباط بین شدت درد قفسه‌سینه با شاخص‌های فیزیولوژیک در بیماران مبتلا به بیماری شریان کرونر انجام شد، مشخص شد که شدت درد قفسه‌سینه در بیماران برابر با ۲/۱۴±۶/۵۱ است. براساس این نتایج، ارتباط معناداری بین این دو متغیر وجود نداشته و می‌توان این‌گونه استنباط کرد که درد قفسه‌سینه دارای ماهیت ذهنی بوده و نمی‌توان از این شاخص به‌منظور پیش‌بینی بیماری استفاده کرد [۵۷].

مروری بر متون نشان داد که تحقیقات متعددی برای پیش‌بینی بیماری CAD با استفاده از فرآیند داده‌کاوی انجام شده است. به‌طور مثال، در تحقیقی از سه تکنیک رگرسیون لجستیک، درخت تصمیم‌گیری، طبقه‌بندی و شبکه‌های عصبی به همراه داده‌های هشت عامل خطرساز درخصوص ۱۲۴۵بیمار استفاده شد و در نهایت، مدل شبکه عصبی پرسپترون چندلایه با دقت ۷۸/۸ درصد بهترین مدل معرفی گردید [۵۸]. در پژوهش Sajja نیز، از الگوریتم‌های بیز ساده، درخت تصمیم و پرسپترون چندلایه شبکه عصبی بر روی مجموعه داده‌ها استفاده شد. براساس یافته‌ها، دقت الگوریتم پرسپترون چندلایه شبکه عصبی برابر ۹۱/۷۵ درصد و نشان‌دهنده بهترین عملکرد بود [۵۹]. در مطالعه حاضر نیز، بیماری CAD به‌عنوان متغیر هدف انتخاب شد و تعداد ۲۶ متغیر از متغیرهای پایگاه داده بیماران قلبی به‌عنوان متغیرهای لایه ورودی استفاده گردید؛ سپس، پارامترهای لایه ورودی در برازش مدل شبکه عصبی با اجرای الگوریتم شبکه عصبی مصنوعی تعیین شدند. بر همین اساس، وزن متغیرهای لایه ورودی در برازش مدل شبکه عصبی سه لایه برآورد شده و نتایج آن ارائه شد. برای اجرای مدل شبکه عصبی۷۰ درصد از داده‌ها آموزشی و ۳۰ درصد نیز آزمایشی درنظر گرفته شدند. نتایج مطالعه بیانگر این بود که استفاده از شبکه عصبی برای برازش این داده‌ها مناسب بوده و میزان دقت کلی این مدل ۰/۷۹۸ محاسبه گردید. از سوی دیگر، نتایج مطالعه حاضر نشان داد که الگوریتم SVM با مقدار میانگین پایین‌تر درصد خطای مطلق و خطای کمتر به برازش داده‌ها می‌پردازد. مقدار بیشتر آماره هاسمر- لمشو نیز نشان از عملکرد بهتر الگوریتم SVM بر روی داده‌های موجود داشت و همچنین، در تشخیص بیماری CAD پیش‌بینی دقیق و بهتری ارائه می‌کرد. همچنین، الگوریتم SVM با دقت و حساسیت بیشتری نسبت به مدل ANN به پیش‌بینی بیماران عروق کرونری قلب پرداخته است.

Colak و همکاران نیز از مدل شبکه عصبی مصنوعی برای پیش‌بینی بیماری عروق کرونر قلب استفاده کردند. آن‌ها از روش‌های تشخیصی غیرتهاجمی برای پیش‌بینی بیماری استفاده

[Downloaded from jhbmi.com on 2024-10-12]

کردند که این روش به‌عنوان یک پیش‌آگهی در تصمیمات بالینی کاربرد داشت [۶۰].

براساس مطالعات پیشین، الگوریتم‌های SVM با دقت و صحت بالاتری به پیش‌بینی بیماری و تفکیک بیماران و افراد سالم پرداخته‌اند [۶۱-۶۳]. به‌عنوان مثال، در سال ۲۰۱۶ مطالعه‌ای به‌منظور پیش‌بینی دقیق بیماری عروق کرونری با استفاده از الگوریتم‌های بیوانفورماتیک انجام شد. نتایج نشان دادند که از بین این ویژگی‌ها، متغیر اسکن تالیوم به عنوان مهمترین ویژگی در تشخیص بیماری‌های قلبی مطرح شد و طراحی مدل‌های پیش‌بینی ماشینی مانند الگوریتم SVM با دقت ۱۰۰ درصد می‌توانست بین افراد بیمار و سالم تمایز قائل شود [۵۵]؛ با وجود مطالعات بسیاری که همگی در راستای تائید عملکرد و دقت بیشتر SVM هستند، ولیکن معدود مطالعاتی نیز وجود دارند که کارایی این الگوریتم را تائید نکرده و روش‌هایی مانند بهینه‌سازی ذرات دودویی و الگوریتم ژنتیک را برای عنوان یک مدل انتخابی برای تعیین بیماری عروق کرونر پیشنهاد داده‌اند [۶۴]. محدودیت اصلی این پژوهش، عدم دسترسی به بانک اطلاعاتی بیماران قلبی بود که برای رفع این محدودیت، با همکاری متخصصان این حوزه چک‌لیستی شامل ویژگی‌ها و فیلدهای مؤثر در پیش‌بینی بیمارهای عروق کرونری قلب طراحی گردید و سپس، پژوهشگران با مراجعه حضوری به محیط‌های پژوهشی، داده‌های لازم را از پرونده پزشکی بیماران استخراج نمودند.

+امروزه، به دلیل گسترش روزافزون فناوری‌های نوین اطلاعاتی که با تولید حجم انبوهی از داده‌ها همراه است، فرآیند کشف دانش، شناسایی روابط یا کشف الگوهای پنهان در بین داده‌ها با مشکلات جدی روبه‌رو شده است. فرآیند پیش‌بینی بیماری‌ها در تشخیص پزشکی نیز می‌تواند به عنوان یک فرآیند مهم در تصمیم‌گیری تفسیر شود که در طی آن پزشکان با توجه به یک یا چندین متغیر گوناگون از یک مجموعه داده بالینی، بیماری را پیش‌بینی می‌کنند. این فرآیند را می‌توان به روش منطقی، هدفمند، دقیق و با استفاده از روش‌های یادگیری ماشینی و به‌کارگیری الگوریتم‌های داده‌کاوی تسهیل و تسریع کرد. دانش داده‌کاوی با هدف فراهم آوردن روش‌هایی برای پردازش هوشمند حجم بالای داده‌ها و استخراج روال‌ها و قوانین حاکم بر آن مطرح شده است؛ بنابراین هرجا که مسئله تحلیل اطلاعات مطرح است، فرآیند داده‌کاوی در گستره‌ای وسیع کاربرد خواهد داشت. درحال حاضر، به دلیل افزایش چشمگیر بیماری‌های قلبی و هزینه‌های بالایی که بر جامعه وارد می‌کنند، جامعه پزشکی به دنبال برنامه‌هایی برای بررسی دقیق، پیشگیری، شناسایی زود هنگام و درمان مؤثر این بیماری‌ها می‌باشد. از این رو، پیشنهاد می‌شود با استفاده از داده‌کاوی و کشف دانش در مراکز قلب و عروق، دانش ارزشمندی ایجاد شود تا رفتار آتی بیماران قلبی قابل پیش‌بینی باشد؛ بنابراین این مطالعه با هدف پیش‌بینی بیماری عروق کرونر قلب به مقایسه دقت پیش‌بینی دو الگوریتم کاربردی داده‌کاوی با نام‌های SVM و ANN پرداخت. برهمین اساس، با به‌کارگیری الگوریتم‌های کاربردی داده‌کاوی و با توجه به متغیرهای پیشگویی‌کننده بیماری‌های قلبی می‌توان با دقت و حساسیت بیشتری به تشخیص و ترسیم فرآیند درمانی بیماران قلبی پرداخته و درنتیجه از عواقب جبران‌ناپذیر بیماری‌های قلبی و به‌ویژه بیماری عروق کرونر قلب پیشگیری نمود.

تشکر و قدردانی

این مقاله برگرفته از یافته‌های یک طرح پژوهشی مصوب در معاونت تحقیقات و فناوری دانشگاه علوم پزشکی ارتش با شماره ثبت ۵۹۴۲۷۳ بود که با همکاری اعضای هیئت علمی دانشگاه علوم پزشکی ایران اجرا شد. گردآوری داده‌ها و تهیه بانک اطلاعاتی با همکاری صمیمانه مسئولین محترم واحد فناوری اطلاعات سلامت بیمارستان‌های مورد مطالعه انجام پذیرفت که بدین‌وسیله از ایشان کمال امتنان و سپاسگزاری به‌عمل می‌آید.

تضاد منافع

نویسندگان تضاد منافعی گزارش نکرده‌اند.

References

1. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al. Executive summary: heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation 2012;125(1):188-97.

2. Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, Budaj A, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. Eur Heart J 2013;34(38):2949-3003.

3. Fuster V, Kelly BB. Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. Washington (DC): National Academies Press (US); 2010.

[Downloaded from jhbmi.com on 2024-10-12]

4. Firouzian A, Bigvand A, Akbarimlak S, Ghafarzadeh Motlagh A. Study of the Prevalence of Coronary Artery Disease risk factors in people over 15 years of age in Tehran 2000 [dissertation]. Tehran: Tehran University of Medical Sciences Tehran; 2000. Persian.
5. Rajkumar A, Reena GS. Diagnosis of heart disease using datamining algorithm. Global Journal of Computer Science and Technology 2010;10(10):38-43.
6. Go AS, Mozafarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Heart Disease and Stroke Statistics—2013 Update. A Report From the American Heart Association 2013. Circulation 2012;127(1):e6–e245.
7. Kasper DL, Fauci AS, Hauser SL, Longo DL, Jameson JL, Loscalzo J. Harrison's Principles of Internal Medicine. 18th ed. NewYork: McGraw-Hill Education; 2011.
8. World Health Organization (WHO). The top 10 causes of death; 2017 [cited 2018 Feb 16]. Available from: http://www.who.int/mediacentre/factsheets/fs310/en/.
9. McPhee JS, French DP, Jackson D, Nazroo J, Pendleton N, Degens H. Physical activity in older age: perspectives for healthy ageing and frailty. Biogerontology 2016;17(3):567-80
10. Bonow RO, Mann DL, Zipes DP, Libby P. Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine. Philadelphia, PA: Saunders; 2014.
11. U.S. National Library of Medicine (NLM). Coronary Artery Disease (CAD); 2017 [cited 2018 Mar 5]. Available from: https://medlineplus.gov/ coronary arterydisease.html.
12. Amani F, Kazemnejad A, Habibi R, Hajizadeh E. Pattern of mortality trend in Iran during 1970-2009. J Gorgan Univ Med Sci 2011; 12(4) :85-90. Persian
13. Navid H. What is coronary artery disease? Iranian society of heart disease specialists; 2014. Available http://www.iranheart.ir.
14. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthc Inform Res 2013;19(2):121-9.
15. Heidarian M, Slami Joibari M. Datamining in Blood stem cells transplant records for acute & chronic Graft Versus Host Disease (GVHD) diagnosis [dissertation]. Tabriz: Azarbaijan Shahid Madani University; 2015. Persian
16. Sudhakar K, Manimekalai DM. Study of Heart Disease Prediction using Data Mining. Journal of Advanced Research in Computer Science and Software Engineering 2014;4(1): 1157-60.
17. Hamidi H, Daraei A. Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases. International Journal of Engineering-Transactions A: Basics; 2016;29(7): 921-30. Available from: https://www. ije.ir/Vol29/No7/A/6-2298.pdf.
18. Sivagowry S, Durairaj M, Persia A. An empirical study on applying data mining techniques for the analysis and prediction of heart disease. International Conference on Information Communication and Embedded Systems (ICICES); 2013 Feb 21-22; Chennai, India: IEEE; 2013. p. 265-70.
19. Taneja A. Heart disease prediction system using data mining techniques. Oriental J of computer sci and tech 2013; 6(4):457-66. Available from: https://www. computer scijournal.org/.
20. Bhatla N, Jyoti K. An analysis of heart disease prediction using different data mining techniques. Int J Eng 2012; 1(8):1-4.
21. Dangare CS, Apte SS. Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications 2012; 47(10):44-8.
22. Mohammadpour R A, Esmaeili M H, Ghaemian A, Esmaeili J. Application of artificial neural network for assessing Coronary Artery Disease. J Mazandaran Univ Med Sci 2012; 21(86):9-17. Persian
23. Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications 2011;17(8):43-8.
24. Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, et al. Development of traditional Chinese Medicine clinical data warehouse for medical knowledge discovery and decision support. Artif Intell Med 2010;48(2-3):139-52.
25. Guner LA, Karabacak NI, Akdemir OU, Karagoz PS, Kocaman SA, Cengel A, et al. An open-source framework of neural networks for diagnosis of coronary artery disease from myocardial perfusion SPECT. J Nucl Cardiol 2010;17(3):405-13.
26. Mazaheri S, Ashoori M, Bechari Z. A model to predict heart disease treatment using data mining. Payavard Salamat 2017;11(3):287-96. Persian
27. Kazemi M, Mehdizadeh H, Shiri A. Heart disease forecast using neural network data mining technique. Journal of Ilam University of Medical Sciences 2017;25(1):20-32. Persian
28. Sabbagh Gol H. Detection of coronary artery disease using C4.5 decision tree. Journal of Health and Biomedical Informatics 2017;3(4):287-99. Persian
29. Rafeh R, Arbabi M. Data Mining techniques to diagnose diabetes using blood lipids. Journal of Ilam University of Medical Sciences 2015;23(4):239-47. Persian
30. Khosravanian A, Ayat SS. Presenting an intelligent system for diagnosis of coronary heart disease by using Probabilistic Neural Network. Health Inf Manage 2015; 12(1):3-13. Persian
31. Mahmoudi I, Askari Moghadam R, Moazzam MH, Sadeghian S. Prediction model for coronary artery disease using neural networks and feature selection based on classification and regression tree. J Shahrekord Univ Med Sci 2013; 15(5) :47-56. Persian
32. Adeli M, Biglarian A, Bakhshi E, Adeli OA. Application of artificial neural network model in predicting the mixed response of atherosclerosis disease. Razi Journal of Medical Sciences 2013; 20(113):20-8. Persian

[Downloaded from jhbmi.com on 2024-10-12]

33. Ghasem Ahmad L, Tolouei ashlaghi A, Poor ebrahimi A, Ebrahimi M. Using data mining techniques for prediction breast cancer recurrence. Iranian Journal of Breast Diseases 2013; 5(4): 23-34. Persian
34. Daliri N, Zakeri-Moghadam M. Aging and heart disease. Iranian Journal of Cardiovascular Nursing 2016;5(1):64-7. Persian
35. Bayturan O, Kapadia S, Nicholls SJ, Tuzcu EM, Shao M, Uno K, et al. Clinical predictors of plaque progression despite very low levels of low-density lipoprotein cholesterol. J Am Coll Cardiol 2010;55(24):2736-42.
36. Nicholls SJ, Hsu A, Wolski K, Hu B, Bayturan O, Lavoie A, et al. Intravascular ultrasound-derived measures of coronary atherosclerotic plaque burden and clinical outcome. J Am Coll Cardiol 2010;55(21):2399-407.
37. Murthy VL, Naya M, Foster CR, Gaber M, Hainer J, Klein J, et al. Association between coronary vascular dysfunction and cardiac mortality in patients with and without diabetes mellitus. Circulation 2012;126(15):1858-68.
38. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J 2012;33(13):1635-701.
39. Frey P, Waters DD, DeMicco DA, Breazna A, Samuels L, Pipe A, et al. Impact of smoking on cardiovascular events in patients with coronary disease receiving contemporary medical therapy (from the Treating to New Targets [TNT] and the Incremental Decrease in End Points Through Aggressive Lipid Lowering [IDEAL] trials). Am J Cardiol 2011;107(2):145-50.
40. Hosseiniyan AS, Maleki A, Alizadeh S, Asadi M. The relationship between the risk factors of coronary artery disease and coronary angiography findings. 6th Annual Student Congress Student Research Committee; Ardabil University of Medical Sciences and Health Services; 2014. Persian. Available from: https://www.eprints.arums.ac.ir/6444/1/cardiology.pdf.
41. Masethe HD, Masethe MA. Prediction of heart disease using classification algorithms. Proceedings of the World Congress on Engineering and Computer Science; 2014 Oct 22-24; San Fransico, USA: WCECS; 2014. p. 22-4.
42. Kivimaki M, Nyberg ST, Batty GD, Fransson EI, Heikkila K, Alfredsson L, et al. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. Lancet 2012;380(9852):1491-7.
43. Delbandi T. The impact of air pollutants in Tehran with deaths from cardiovascular disease in people over 60 years. International Conference on Urban Economy; 2016 May 18; Tehran: Scientific Society of Urban Economics of Iran; 2016. Persian
44. Zhang Q, Cheng L, Boutaba R. Cloud computing: state-of-the-art and research challenges. Journal of Internet Services and Applications 2010;1(1):7-18.
45. Vanos JK, Hebbern C, Cakmak S. Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities. Environ Pollut 2014;185:322-32.
46. Asefzadeh S, Alikhani S, Javadi HR.. Socio-economic status and mortality from cardiovascular diseases in Qazvin. J Qazvin Univ Med Sci 2013;16(4):40-6. Persian.
47. Veeranna V, Pradhan J, Niraj A, Fakhry H, Afonso L. Traditional cardiovascular risk factors and severity of angiographic coronary artery disease in the elderly. Prev Cardiol 2010;13(3):135-40.
48. Bidel Z, Hemati R, Naserifar R, Nazarzadeh M, Del Pisheh A. Association of cardiovascular risk factors and coronary arteries involvement based on angiographic findings. Journal of Ilam University of Medical Sciences 2014;22(1):147-54. Persian
49. Ansari R, Mehraeen E, Sabahi A. Need and priority assessment of Kerman's teaching hospitals to implementation of RFID technology. Journal of Health and Biomedical Informatics 2015; 2(3):141-8. Persian
50. Safdari R, Ghazi Saeedi M, Gharooni M, Nasiri M, Arji G. Comparing performance of decision tree and neural network in predicting myocardial infarction. Journal of Paramedical Science and Rehabilitation 2014; 3(2): 26-37. Persian
51. Meyers DG, Neuberger JS, He J. Cardiovascular effect of bans on smoking in public places: a systematic review and meta-analysis. J Am Coll Cardiol 2009;54(14):1249-55.
52. Bakhshandeh H, Amani F, Doostkami H. Comparison of 200 smoker and non-smoker patients with acute myocardial infarction regarding age, sex, blood pressure, diabetes, cholesterol, history of the disease and familial history admitted in Bouali hospital of Ardabil [dissertation]. Ardabil: Ardabil University of Medical Sciences; 2002. Persian.
53. Campo G, Pavasini R, Malagu M, Mascetti S, Biscaglia S, Ceconi C, et al. Chronic obstructive pulmonary disease and ischemic heart disease comorbidity: overview of mechanisms and clinical management. Cardiovasc Drugs Ther 2015;29(2):147-57.
54. Rezabeigi Davarani E, Iranpour A, Khanjani N, Mohseni M, Nazari Robati F. Cardiovascular Diseases risk factors and the relationship between knowledge level and preventive behaviors for Cardiovascular diseases among women in Kerman. Journal of Health_Based Research 2016;2(2):119-32. Persian.
55. Shafiee H, Ebrahimi M. Accurate Prediction of Coronary Artery Disease Using Bioinformatics Algorithms. Qom Univ Med Sci J 2016; 10(4) :22-35. Persian.
56. Uguz H. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. J Med Syst 2012;36(1):61-72.

[Downloaded from jhbmi.com on 2024-10-12]

57. Fakhr-Movahedi A, Ebrahimian A, Mirmohammad khani M, Ghasemi S. Relationship between chest pain severity and physiological indexes in patients with coronary artery disease. Tehran Univ Med J 2016; 74 (2):140-5. Persian

58. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Systems with Applications 2008;34(1):366-74.

59. Sajja S. Data mining of medical datasets with missing attributes from different sources [dissertation]. Ohio: Youngstown State University; 2010.

60. Colak MC, Colak C, Kocaturk H, Sagiroglu S, Barutcu I. Predicting coronary artery disease using different artificial neural network models. Anadolu Kardiyol Derg 2008;8(4):249-54.

61. Maglogiannis I, Loukis E, Zafiropoulos E, Stasis A. Support Vectors Machine-based identification of heart valve diseases using heart sounds. Comput Methods Programs Biomed 2009;95(1):47-61.

62. Ghumbre S, Patil C, Ghatol A. Heart disease diagnosis using support vector machine. International Conference on Computer Science and Information Technology (ICCSIT'2011); 2011 Dec 17- 18; Pattaya, Thailand; 2011.

63. Hanbay D. An expert system based on least square support vector machines for diagnosis of the valvular heart disease. Expert Systems with Applications 2009;36(3, Part 1):4232-8.

64. Babaoglu I, Findik O, Ülker E. A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. Expert Systems with Applications 2010; 37(4):3177-83.

[Downloaded from jhbmi.com on 2024-10-12]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2018; 5(2): 252-264

Comparing the Performance of Data Mining Algorithms in Predicting Coronary Artery Diseases

Ayatollahi Haleh[1], Gholamhosseini Leila[2, 3]*, Salehi Masoud[4]

• **Received:** 22 Apr, 2018 • **Accepted:** 12 Aug, 2018

Introduction: Cardiovascular diseases are the first leading cause of death worldwide. World health organization has estimated that the morality rate due to heart diseases will mount to 23 million cases by 2030. Hence, the use of data mining algorithms will be useful in predicting coronary artery disease. The objective of the present study was to compare the accuracy of the CAD predictions made by ANN and SVM techniques.

Methods: The present study was conducted via descriptive-analytical method. The research sample included all CAD patients hospitalized in three hospitals affiliated to AJA University of Medical Sciences from March 2016 to March 2017. Totally, 1324 records with 26 characteristics affecting the disease were extracted and after normalizing, and cleaning of the data, they were entered in SPSS statistics V23.0 & IBM Excel 2013; then, R3.3.2 data mining software was used to format data.

Results: SVM model with lower MAPE (112.03) and higher Hosmer-lemeshow statistic (16.71), sensitivity (92.23) and specificity (74.42) yielded better fitness of data and provides more accurate prediction than ANN model. On the other hand, since the area under the ROC curve in SVM algorithm was more than that in ANN, it could be concluded that this model had higher accuracy.

Conclusion: According to the results, SVM algorithm presented higher accuracy and better performance than ANN model and showed higher sensitivity and accuracy. It is suggested that in future studies, the results of the present study be compared with the findings resulted from applying other data mining algorithms.

Keywords: Coronary Artery Disease (CAD), Data mining algorithms, Artificial Neural Network (ANN), Support Vector Machine (SVM)

• **Citation:** Ayatollahi H, Gholamhosseini L, Salehi M. Comparing the Performance of Data Mining Algorithms in Predicting Coronary Artery Diseases. Journal of Health and Biomedical Informatics 2018; 5(2): 252-264.

1. Ph.D. in Medical Informaticsm, Associate Professor, Health Information Management Dept., School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.
2. Ph.D. Student in Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.
3. Lecturer, Health information Technology Dept., School of Paramedical Sciences, AJA University of Medical Sciences, Tehran, Iran.
4. Ph.D. in Biostatistics, Associate Professor, Biostatistics Dept., School of Public Health, Iran University of Medical Sciences, Tehran, Iran.

***Correspondence**: No. 6, Shahid Yasami St., Vali-e-Asr St., School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.
• **Tel:** 021-88794302 • **Email:** gholamhosseini.l@tak.iums.ac.ir

[Downloaded from jhbmi.com on 2024-10-12]