مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره سوم شماره چهارم. ص ۲۵۸-۲۵۱، ۱۳۹۵

مقاله پژوهشی

یک سیستم خبره تشخیص بیماری قلبی مبتنی بر ترکیب شواهد در داده‌کاوی

حمیدرضا طهماسبی[۱]، مهرداد جلالی[۲]*، حسن شاکری[۲]

• دریافت مقاله: ۹۵/۱۰/۲۲ • پذیرش مقاله: ۹۵/۱۲/۲۶

مقدمه: بیماری عروق کرونری، شایع‌ترین نوع بیماری قلبی و عامل اصلی مرگ‌ومیر در کشورهای صنعتی می‌باشد. این پژوهش با هدف طراحی یک سیستم خبره با دقت بالا برای تشخیص بیماری عروق کرونری قلب انجام شد.

روش: این مطالعه از نوع کاربردی بوده و از ۱۴ ویژگی مربوط به ۳۰۳ نفر که تحت آنژیوگرافی کرونری قرار گرفتند، استفاده شده است. برای تشخیص دقیق‌تر بیماری عروق کرونری، نتایج سه روش کلاسه‌بندی شبکه‌های عصبی، بیزین ساده و نزدیکترین k همسایه با استفاده از تئوری ترکیب شواهد دمستر-شافر ترکیب شده است. از نرم‌افزار داده‌کاوی Weka نسخه ۷٫۳ و همچنین زبان برنامه‌نویسی C# در محیط Net Framework. برای پیاده‌سازی روش استفاده گردید. برای ارزیابی کارایی، روش 10-Fold cross validation به کار برده شد.

نتایج: نتایج نشان داد که میانگین دقت (Accuracy)، حساسیت (Sensitivity) و ویژگی (Specificity) در روش پیشنهادی به ترتیب ۹۰/۱ درصد، ۸۹/۰۹ درصد و ۹۱/۳ درصد می‌باشد که این مقادیر در مقایسه با هر یک از کلاسه‌بندهای شرکت‌کننده در ترکیب بیشتر بود و همچنین نسبت به تحقیقات مشابه، دقت بهتری در تشخیص افراد دارای بیماری عروق کرونری داشت.

نتیجه‌گیری: تحلیل نتایج نشان می‌دهند که در جامعه آماری مورد مطالعه، روش پیشنهادی عملکرد بهتری در تشخیص بیماری عروق کرونری دارد و می‌تواند به عنوان یک سیستم خبره، توسط متخصصین بالینی درگیر با بیماری قلبی، با هدف کمک به تصمیم‌گیری‌های بالینی و کاهش خطاها، بهبود زمان انتظار در تشخیص بیماری و کاهش آزمایش‌های غیرضروری پزشکی استفاده گردد.

کلید واژه‌ها: بیماری عروق کرونری، سیستم خبره، تشخیص پزشکی، تئوری دمستر-شافر، کلاسه‌بندی

•**ارجاع:** طهماسبی حمیدرضا، جلالی مهرداد، شاکری حسن. یک سیستم خبره تشخیص بیماری قلبی مبتنی بر ترکیب شواهد در داده‌کاوی. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۵؛ ۳(۴): ۲۵۸-۲۵۱.

۱. دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد نیشابور، نیشابور، ایران.
۲. استادیار، دکترای مهندسی کامپیوتر، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد مشهد، مشهد، ایران.
* **نویسنده مسئول:** مشهد، دانشگاه آزاد اسلامی مشهد، گروه مهندسی کامپیوتر
• **شماره تماس:** ۰۹۱۵۱۰۴۶۱۱۷
• **Email:** ithrt20@gmail.com

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

مقدمه

دسته مهمی از مسائل در پزشکی مربوط به تشخیص بیماری‌ها می‌باشد که بر اساس آزمایش‌های مختلف بر روی بیمار انجام می‌گیرد. هنگامی که تعداد پارامترها در تشخیص بیماری زیاد شود ممکن است تشخیص بیماری حتی برای یک متخصص خبره پزشکی نیز به سختی امکان‌پذیر باشد [۱]. در سال‌های اخیر استفاده از سیستم‌های خبره و تصمیم‌یار در پزشکی، با هدف کمک به تصمیم‌گیری‌های بالینی و کاهش خطاها، بهبود زمان انتظار در تشخیص بیماری و کاهش آزمایش‌های غیرضروری پزشکی مورد توجه بوده است [۲].

از دیدگاه داده‌کاوی، پیش‌بینی در تشخیص بیماری، جزء مسائل کلاسه‌بندی داده‌ها محسوب می‌شود. مطالعات انجام شده نشان می‌دهند که این تکنیک‌ها در تشخیص بیماری‌هایی از قبیل سرطان، دیابت و بیماری قلبی بسیار مفید بوده‌اند [۳]. برای حل یک مسئله کلاسه‌بندی، روش‌ها و الگوریتم‌های مختلفی وجود دارد. این روش‌ها در معماری، الگوریتم یادگیری یا نحوه آموزش و نمایش ویژگی‌ها با یکدیگر متفاوت‌اند [۴]. در داده‌کاوی پزشکی و به خصوص در مسائل مربوط به تشخیص بیماری‌ها استفاده از الگوریتم‌های کلاسه‌بندی با بالاترین درجه اعتماد و دقت، همیشه مورد توجه بوده است [۴].

بیماری قلبی علت اصلی مرگ‌ومیر در جهان در ده سال گذشته بوده [۶،۵] و بیماری عروق کرونری قلب (Coronary Artery Disease) CAD شایع‌ترین نوع بیماری قلبی می‌باشد که عامل اصلی مرگ‌ومیر در کشورهای صنعتی بوده و به سرعت در کشورهای در حال توسعه رو به انتشار است [۷]. به همین دلیل در مقایسه با سایر بیماری‌های تهدید کننده زندگی، این بیماری مورد توجه بسیاری از پژوهشگران حوزه پزشکی بوده است [۸]. بیماری‌های عروق کرونر قلب ناشی از تنگ شدن یا اسپاسم عروق کرونر که رگ‌های اصلی خون‌رسانی به عضله قلبی می‌باشند و در نتیجه کاهش خون‌رسانی به قلب می‌باشد [۹]. تست ورزش، اکوکاردیوگرام، اسکن هسته‌ای قلب و آنژیوگرافی از جمله راه‌های تشخیص این بیماری است [۱۰]. این قبیل آزمایش‌ها معمولاً در معرض خطا بوده [۸] و برای بیمار، رنج‌آور، وقت‌گیر و هزینه‌بر می‌باشند [۱۱]. به همین دلیل در سال‌های اخیر، توسعه و ارائه سیستم‌های خبره در تشخیص بیماری قلبی و به ویژه بیماری عروق کرونری قلب با دقت بالا برای غلبه بر این مشکلات از اهمیت خاصی برخوردار بوده است [۱،۳].

اگر چه بعضی از روش‌های کلاسه‌بندی برای تشخیص بیماری نسبت به بقیه بهتر عمل می‌کنند، ولی هیچ کلاسه‌بندی نمی‌تواند همیشه و در هر وضعیتی بر دیگری برتری داشته و یا داده‌ها را بدون هیچ‌گونه خطایی طبقه‌بندی کند [۴]. یک روش غلبه بر محدودیت‌های یک کلاسه‌بند و بهبود دقت سیستم‌های تشخیص استفاده از ترکیب کلاسه‌بندها می‌باشد [۳،۱۱]. در یک سیستم کلاسه‌بندی ترکیبی، اطلاعات خروجی کلاسه‌بندهای مختلف، با هم ترکیب شده و تصمیم نهایی کلاسه‌بندی بر اساس اطلاعات ترکیب شده، گرفته می‌شود. در این مقاله یک سیستم خبره برای تشخیص بیماری عروق کرونری قلب پیشنهاد می‌گردد که به منظور دستیابی به دقت بیشتر، نتایج سه روش کلاسه‌بندی متداول استفاده شده در تشخیص بیماری شامل شبکه‌های عصبی، بیزین ساده و نزدیک‌ترین k همسایه [۱۲] را با استفاده از تئوری ترکیب شواهد دمستر-شافر (Dempster-Shafer) ترکیب می‌کند. این تئوری تعمیم‌یافته احتمال بیزین می‌باشد که خصوصیات بازیابی صریح عدم قطعیت و قاعده ترکیب شواهد آن [۱۳] موجب استفاده از آن در این روش شده است.

Rani مجموعه داده بیماری عروق کرونری قلب را با استفاده از شبکه‌های عصبی مورد تحلیل قرار داد. در پژوهش انجام شده با به کارگیری دو مدل شبکه عصبی تک لایه و شبکه عصبی چندلایه بر روی مجموعه داده، به ترتیب میانگین دقت حدود ۸۷ درصد و ۸۳ درصد حاصل شده است [۱۴].

یک سیستم هوشمند برای تشخیص بیماری عروق کرونری قلب با استفاده از شبکه‌های احتمالی توسط Khosravanian و Ayat [۱۰] ارائه و بر روی ۱۵۲ نفر از افراد تحت آنژیوگرافی مورد آزمایش قرار گرفت. این سیستم ۶/۶ درصد خطا در تشخیص داشته است و در مقایسه با شبکه‌های عصبی پس انتشار سرعت زیادتر و تعمیم‌پذیری خوبی دارد [۱۰].

Vila-Francés و همکاران [۲] یک سیستم خبره برای تشخیص بیماری قلبی ارائه کرده‌اند که از شبکه‌های بیزین استفاده کردند. دقت سیستم این پیشنهادی در تشخیص صحیح افراد غیر بیمار حدود ۹۱ درصد بوده است.

Rajkumar و Reena از الگوریتم‌های درخت تصمیمی، بیزین ساده و نزدیک‌ترین k همسایه برای تشخیص بیماری قلبی استفاده کرده‌اند که با ارزیابی آن‌ها بر روی مجموعه داده با ۳۰۰۰ نمونه، الگوریتم بیزین ساده با دقتی برابر ۵۳/۳۳ درصد از دو روش دیگر بهتر بوده است [۱۵].

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

Shouman و همکاران [۶] مدلی برای افزایش دقت روش کلاسه‌بندی درخت تصمیم در تشخیص بیماران مبتلا به بیماری عروق کرونری ارائه کرده‌اند که با اعمال انواع درختان تصمیم بر روی مجموعه داده و استفاده از تکنیک رأی اکثریت به دقت ۸۴/۱ درصد دست یافتند و نشان دادند که در مقایسه با روش‌های Bagging و درخت تصمیم J48 دقت بهتری داشته است.

Negahbani و همکاران [۱۱] یک سیستم ترکیبی تشخیص خودکار CAD با استفاده از روش خوشه‌بندی C-Means فازی پیشنهاد کرده‌اند که دقتی برابر ۸۷ درصد در تشخیص درست دارد. این روش از یک روش انتخاب آماری ویژگی‌ها به‌منظور کاهش تعداد ویژگی‌ها و انتخاب مهم‌ترین آن‌ها استفاده می‌کند. با توجه به این که نوع ویژگی‌ها از انواع مختلف عددی و اسمی می‌باشند برای محاسبه فاصله بین نمونه‌ها از معیار فاصله مینکوسکی تعمیم یافته استفاده می‌کند [۱۱].

Alizadehsani و همکاران [۱۶] الگوریتم‌های بیزین ساده، ماشین بردار پشتیبان، شبکه‌های عصبی و Bagging را روی مجموعه داده با ۳۰۳ نمونه جهت تشخیص CAD به کار گرفتند. برای بهبود دقت، ابتدا مجموعه داده توسط یک روش ایجاد ویژگی، غنی شده و سپس بهره اطلاعاتی و درجه اطمینان برای تعیین مؤثرترین ویژگی‌ها روی مجموعه داده به کار برده شد. نتایج روش مذکور نشان داده است که روش ماشین بردار پشتیبان در بین الگوریتم‌های بررسی شده بهترین دقت را داشته است.

یک سیستم تصمیم‌یار بالینی برای تشخیص بیماری قلبی توسط Bashir و همکاران [۳] پیشنهاد شده است که ابتدا پنج روش کلاسه‌بندی بیزین ساده، درخت تصمیم مبتنی بر شاخص جینی، درخت تصمیم مبتنی بر بهره اطلاعاتی، یادگیرنده مبتنی بر حافظه و ماشین بردار پشتیبان بر روی مجموعه داده‌ها اعمال شده و نتیجه نهایی تشخیص با به کارگیری رأی اکثریت روی نتایج این پنج کلاسه‌بند به دست می‌آید. به کارگیری این سیستم بر روی چندین مجموعه داده مختلف مربوط به بیماری قلبی، در بهترین حالت دقت ۹۰/۰۳ درصد را به همراه داشته است که از دقت بالاتری نسبت به روش‌های جدید مقایسه شده در مطالعه مذکور برخوردار می‌باشد.

هدف این مطالعه، طراحی یک سیستم خبره مبتنی بر ترکیب شواهد، برای تشخیص بیماری عروق کرونری قلب است که علاوه کمک به تصمیم‌گیری‌های بالینی و کاهش خطاها و همچنین صرفه‌جویی در زمان و هزینه‌های آزمون تشخیصی، از اعتماد و دقت قابل قبولی برخوردار باشد.

روش

این پژوهش، یک مطالعه‌ی کاربردی- توسعه‌ای است که بیماری عروق کرونری قلب را پیش‌بینی و تشخیص می‌دهد. جامعه آماری، مجموعه داده بیماری عروق کرونری قلب بیمارستان Cleveland موجود در مخزن داده دانشگاه ایروین کالیفرنیا [۱۷] می‌باشد که در اغلب پژوهش‌ها و مطالعات مرتبط، از این مجموعه داده استفاده شده است [۶،۲۰،۱۹،۱۸]. این مجموعه داده شامل ۳۰۳ نمونه با ۷۶ ویژگی می‌باشد که در بسیاری از آزمایش‌ها و مطالعات منتشر شده از ۱۴ ویژگی آن‌ها استفاده شده است [۱۹،۱۷] که یکی از این ویژگی‌ها وضعیت نهایی فرد مراجعه کننده می‌باشد که می‌تواند نرمال و یا بیمار CAD باشد. جدول ۱ این ویژگی‌ها را نشان می‌دهد. این داده‌ها مربوط به ۳۰۳ مراجعه کننده بدون سابقه سکته قلبی و کاردیومیوپاتی است که در بیمارستان Cleveland آمریکا تحت آنژیوگرافی کرونری قرار گرفتند و ویژگی‌های دموگرافی و بالینی آن‌ها ثبت شده است. هنگامی که حداقل یکی از عروق کرونری بیش از ۵۰ درصد تنگ شود، توسط آنژیوگرافی یک CAD تشخیص داده می‌شود [۱۱].

در روش پیشنهادی، نتایج حاصل از سه کلاسه‌بند متداول مورد استفاده در تشخیص بیماری شامل شبکه‌های عصبی، بیزین ساده و نزدیکترین k همسایه با هم ترکیب می‌شوند تا تصمیم نهایی کلاسه‌بندی براساس اطلاعات ترکیب شده اتخاذ گردد و یک سیستم خبره دقیق‌تر در تشخیص بیماری عروق کرونری قلب حاصل گردد. خروجی هر کلاسه‌بند یکی از دو کلاس نرمال و بیماری CAD می‌باشد. با توجه به ویژگی‌های بازیابی صریح عدم قطعیت و قاعده ترکیب شواهد در تئوری ترکیب شواهد دمپستر-شافر، از این تئوری برای ترکیب نتایج خروجی کلاسه‌بندها استفاده شد. به طوری که نتایج خروجی کلاسه‌بندهای مستقل را به عنوان شواهد در نظر گرفته و با روی هم‌گذاری آن‌ها، یک تابع باور به دست می‌آورد که متناظر با خروجی‌های کلاسه‌بندها می‌باشد.

در تئوری شواهد دمستر-شافر، باور (Belief) مقداری است که برای بیان قطعیت یک گزاره یا رویداد به کار می رود. یک باور معمولاً براساس یک تابع به نام تخصیص احتمال پایه BPA(Basic Probability Assignment) محاسبه می‌شود. قاعده ترکیب شواهد دمستر، دو بدنه شواهد مستقل

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

تعریف شده در یک چارچوب مشاهدات را با هم ترکیب کرده و به یک بدنه شواهد تبدیل می‌کند. در شکل ۱ دو بدنه شواهد مشاهده می‌شود که A_i (i=1,…,l) و B_j (j=1,…,m) هر کدام معرف یک گزاره و $m(A_i)$ و $m(B_j)$ به ترتیب میزان باور یا BPA برای گزاره‌های Ai و Bj می‌باشند.

$$\{A_1, A_2, ..., A_l\}, \{m_1(A_1), m_1(A_2), ..., m_1(A_l)\}$$
$$\{B_1, B_2, ..., B_m\}, \{m_2(B_1), m_2(B_2), ..., m_2(B_m)\}$$

شکل ۱: دو بدنه شواهد A و B

برای بدنه شواهد جدید حاصل از ترکیب این دو بدنه شواهد، مقدار BPA برابر است با:

$$m(C) = (m_1 \oplus m_2)(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j)} \quad (۱)$$

در روش ترکیبی پیشنهادی با سه کلاسه بند و M کلاس، خروجی کلاسه‌بندها به عنوان شواهد در نظر گرفته شده و کلاس‌ها به عنوان چارچوب مشاهدات محسوب می‌شوند. در صورت وجود عدم قطعیت برای تعیین کلاس نمونه X، این نمونه به هیچ کلاسی تعلق نمی‌یابد. با استفاده از رابطه ۱، تابع باور کامل برای هر کلاس C_q به صورت زیر به دست می‌آید:

$$Bel(C_q) = m(C_q) \qquad q = 1, ..., M+1 \quad (۲)$$

در رابطه فوق، به ازای q=M+1، میزان باور کامل به این که نمونه X، متعلق به هیچ کلاسی نیست به دست می‌آید. برای تعیین برچسب کلاس نمونه X ابتدا بزرگ‌ترین مقدار باور پیدا می‌شود:

$$C = Max\{Bel(C_1), ..., Bel(C_M), Bel(C_{M+1})\} \quad (۳)$$

سپس کلاس با بزرگ‌ترین مقدار باور، به عنوان کلاس نمونه X تعیین می‌گردد. اگر بزرگ‌ترین مقدار باور برابر $Bel(C_{m+1})$ باشد، به این معنی است که هیچ قطعیتی درباره برچسب کلاس X وجود ندارد و کلاسی برای آن تعیین نمی‌شود. با توجه به این که برای پیش‌بینی وضعیت عروق کرونری قلب شامل دو کلاس مبتلا به بیماری و نرمال می‌باشد؛ لذا مقدار M برابر ۲ می‌باشد.

جدول ۱: مشخصات ویژگی‌های استفاده شده از مجموعه داده Cleveland

ویژگی	توضیحات	انواع مقادیر (برای ویژگی‌های دسته‌ای)
Age	سن به سال	
Gender	جنسیت	زن/مرد
Cp	نوع درد قفسه سینه	آنژین معمولی/آنژین غیرمعمولی/درد غیر آنژینی/بدون درد
Trestbps	فشارخون در حالت استراحت برحسب میلی‌متر جیوه	
Chol	کلسترول سرم خون (mg/dl)	
FBS	قند خون ناشتا بیشتر از ۱۲۰mg/dl	بله/خیر
Restecg	نتایج نوار قلب در حال استراحت	نرمال/موج ST-T غیرمعمول (موج T وارونه باشد و یا ST بالاتر از ۰/۰۵ باشد)/کوچک شدن بطن چپ
Thalach	حداکثر ضربان قلب (bpm)	
Exang	آنژین القاشده توسط فعالیت ورزشی	بله/خیر
Oldpeak	افت ST القاشده از ورزش در مقایسه با استراحت	
Slope	شیب قله موج ST	شیب بالا/صاف/شیب پایین
Ca	تعداد عروق بزرگ مشخص‌شده توسط فلوروسکوپی	۰-۳
Thal	سینتیگرافی تالیم-۲۰۱	نرمال/نقص ثابت/نقص برگشت‌پذیر
Num	تشخیص بیماری کرونری قلبی (آنژیوگرافی وضعیت بیمار)	نرمال: تنگ شدن کمتر از ۵۰ درصد/CAD: تنگ شدن بیشتر از ۵۰ درصد

در سیستم خبره پیشنهادی، ابتدا هر یک از روش‌های کلاسه‌بندی متداول شبکه‌های عصبی پرسپترون چندلایه، نزدیک‌ترین k همسایه و بیزین ساده به صورت مجزا جهت پیش‌بینی وضعیت عروق کرونری قلب روی مجموعه داده‌ها اعمال می‌گردند. بدین منظور از ماژول‌های موجود در نرم‌افزار داده‌کاوی Weka نسخه ۷٫۳ با همان پارامترهای پیش فرض استفاده شد.

روش نزدیک‌ترین k همسایه به ازای تعداد همسایه‌های مختلف بین ۳ تا ۱۳ بر روی مجموعه داده‌ها مورد بررسی قرار گرفت که بیشترین دقت به ازای k=۹ حاصل شد؛ بنابراین

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

مقدار k برابر ۹ انتخاب گردید. خروجی این کلاسه‌بند با استفاده از ماتریس تداخل به فرم عددی قابل پذیرش توسط تئوری دمستر-شافر تبدیل شد. در مرحله بعد خروجی دو کلاسه‌بند نزدیک‌ترین k همسایه و شبکه عصبی به عنوان شواهد، با استفاده از تئوری ترکیب دمستر-شافر ترکیب شده و نتایج حاصل با خروجی کلاسه‌بند بیزین ساده نیز با استفاده از تئوری مذکور، ترکیب می‌گردد. روش ترکیبی با استفاده از زبان برنامه‌نویسی C# در محیط Net Framework پیاده‌سازی گردید.

برای ارزیابی دقیق‌تر کارایی از تکنیک k-fold cross validation استفاده شده است که همانند بسیاری از مطالعات دیگر [۳،۱۱،۱۶]، مقدار k برابر ۱۰ در نظر گرفته شد. برای بررسی کارایی مدل پیشنهادی، معیارهای دقت (Accuracy)، حساسیت (Sensitivity) و ویژگی (Specificity) مورد استفاده قرار می‌گیرد. معیار دقت، بیانگر دقت مدل در پیش‌بینی درست بیماری (وضعیت عروق کرونری قلب یا CAD)، معیار حساسیت، نسبت افراد درست بیمار تشخیص داده شده توسط مدل، به کل افراد بیمار و معیار ویژگی، نسبت افراد درست نرمال تشخیص داده شده توسط مدل، به کل افراد نرمال می‌باشد. مقادیر این معیارها از ماتریس تداخل (Confusion) (جدول ۲) به دست می‌آیند.

جدول ۲: ماتریس تداخل

		نتیجه پیش‌بینی شده		
		CAD	Normal	uncertainty
نتیجه واقعی	CAD (بیمار)	TP(True Positive)	FN(False Negative)	UP
	Normal (نرمال)	FP(False Positive)	TN(True Negative)	UN

تحلیل این ماتریس در تشخیص بیماری منجر به چهار حالت مثبت حقیقی (True Positive)، منفی حقیقی (False Negative)، مثبت کاذب (False Positive) و منفی کاذب (True Negative) می‌شود. با توجه به این که در روش پیشنهادی در صورت وجود عدم قطعیت، نمونه ورودی به هیچ کلاسی تعلق نمی‌یابد و نه بیمار و نه نرمال تشخیص داده می‌شود، دو حالت تعداد نمونه‌هایی که بیمار بوده‌اند و به کلاسی تعلق نگرفته‌اند (UP) و تعداد نمونه‌هایی که نرمال بوده‌اند و به کلاسی تعلق نگرفته‌اند (UN) نیز در ماتریس تداخل منظور می‌گردد. با توجه به جدول ۲، معیارهای ارزیابی را می‌توان به صورت زیر بیان کرد:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad (۴)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad (۵)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN+UP+UN} \qquad (۶)$$

نتایج

جدول ۳ ماتریس تداخل حاصل از اعمال هر یک از سه روش کلاسه‌بندی شرکت کننده در ترکیب و همچنین روش ترکیبی پیشنهادی را برای تشخیص و پیش بینی وضعیت عروق کرونری قلب بر روی مجموعه داده CAD نشان می‌دهد. در این جدول، MLP بیانگر روش شبکه عصبی پرسپترون چندلایه، KNN بیانگر روش نزدیک‌ترین k همسایه و NB روش بیزین ساده می‌باشد. روش پیشنهادی نیز با CM (Combined Method) نشان داده شده است.

برای سه کلاسه‌بند MLP، KNN و NB که از عدم قطعیت استفاده نمی‌کنند، ستون عدم قطعیت (uncertainty) در این جدول برابر صفر شده است. در روش ترکیبی پیشنهادی نیز تمام نمونه‌ها به یکی از دو کلاس بیمار و نرمال دسته‌بندی شده و عدم قطعیت صفر شده است.

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

جدول ۳: ماتریس تداخل برای هر یک از روش‌ها و روش ترکیبی

روش		نتیجه پیش‌بینی شده		
		CAD	Normal	uncertainty
MLP	CAD	۱۰۹	۲۹	۰
	Normal	۲۹	۱۳۶	۰
KNN	CAD	۱۱۳	۲۵	۰
	Normal	۲۶	۱۳۹	۰
NB	CAD	۱۱۰	۲۸	۰
	Normal	۲۲	۱۴۳	۰
CM	CAD	۱۲۶	۱۲	۰
	Normal	۱۸	۱۴۷	۰

جدول ۴، معیارهای مربوط به کارایی شامل دقت، حساسیت و ویژگی به دست آمده برای هر یک از روش‌ها را با توجه به ماتریس‌های تداخل جدول ۳ نشان می‌دهد.

جدول ۴: کارایی روش ترکیبی و هر یک از روش‌های شرکت کننده در ترکیب (درصد)

روش	حساسیت	ویژگی	دقت
MLP	۷۸/۹۹	۸۲/۴۲	۸۰/۸۶
KNN	۸۱/۸۸	۸۴/۲۴	۸۳/۱۷
NB	۷۹/۷۱	۸۶/۶۷	۸۳/۵
CM	۹۱/۳	۸۹/۰۹	۹۰/۱

جدول ۵ مقایسه کارایی روش پیشنهادی را با دو تا از جدیدترین روش‌ها در تشخیص بیماری عروق کرونری قلب ارائه شده توسط Negahbani و همکاران [۱۱] و همچنین Bashir و همکاران [۳] نشان می‌دهد.

جدول ۵: مقایسه کارایی روش پیشنهادی با دو روش جدید دیگر (درصد)

روش	حساسیت	ویژگی	دقت
CM (روش پیشنهادی)	۹۱/۳	۸۹/۰۹	۹۰/۱
Negahbani و همکاران [۱۱]	۸۵/۰۷	۸۸/۷۶	۸۷/۱۳
Bashir و همکاران [۳]	۹۰/۹	۸۵/۱	۸۶/۸۲

بحث و نتیجه‌گیری

در این مطالعه سیستمی خبره برای تشخیص وجود یا عدم وجود بیماری عروق کرونری قلب، پیشنهاد شد که سه نتیجه روشی که خود به تنهایی می‌توانند برای این منظور استفاده شوند به کمک تئوری ترکیب شواهد دمستر-شافر با هم ترکیب شدند. اگر چه یکی از امتیازات این تئوری، ویژگی بازیابی عدم قطعیت است به طوری که لزوماً همه نمونه‌های ورودی را به بیمار یا نرمال طبقه‌بندی نکرده و در صورت وجود عدم قطعیت، هیچ تشخیصی انجام نمی‌دهد، ولی جدول ۳ نشان می‌دهد که این ویژگی در روش ترکیبی برابر صفر شده است؛ یعنی همه نمونه‌ها طبقه‌بندی شده‌اند. این اتفاق به این دلیل می‌تواند باشد که در خروجی کلاسه‌بندهای شرکت کننده در ترکیب عدم قطعیت وجود ندارد. برای این که واقعاً عدم قطعیت منظور گردد می‌توان این کلاسه‌بندها را تغییر داد به طوری که هر یک خود، حساس به عدم قطعیت باشد. با مقایسه معیارهای حساسیت، ویژگی و دقت هر یک از روش‌های شرکت کننده در ترکیب باهم در جدول ۴ نشان می‌دهد که روش بیزین ساده (NB) در مقایسه با شبکه عصبی (MLP) از کارایی بهتری برخوردار بوده و همچنین در مقایسه با نزدیک‌ترین k همسایه (KNN)، ویژگی و دقت بالاتری دارد. از طرفی روش پیشنهادی نسبت به هر یک از کلاسه‌بندهای شرکت کننده در ترکیب، حساسیت، ویژگی و دقت بیشتری دارد و افزایش کارایی بسیار قابل توجه می‌باشد.

همان‌طور که در جدول ۵ مشاهده می‌شود، حساسیت، ویژگی و دقت روش پیشنهادی در مقایسه با روش‌های ارائه شده توسط Negahbani و همکاران [۱۱] و همچنین Bashir و همکاران [۳] بیشتر می‌باشد و روش پیشنهادی با دقت ۹۰/۱ درصد می‌تواند به عنوان یک سیستم خبره، کاربرد حائز اهمیتی برای متخصصین بالینی درگیر در بیماری قلبی و به ویژه عروق کرونری قلب داشته و ضمن کاهش خطاها، با کاهش آزمایش‌های غیرضروری منجر به استفاده بهینه از منابع موجود و کاهش زمان انتظار بیمار جهت تشخیص بیماری گردد. به‌کارگیری روش ترکیبی پیشنهادی موجب افزایش حافظه مصرفی، زمان اجرایی و پیچیدگی محاسباتی می‌گردد، ولی در کاربردهای پزشکی که دقت و اعتماد از اهمیت و اولویت بیشتری برخوردار است، به کارگیری این نوع سیستم‌ها بسیار سودمند خواهد بود. روش پیشنهادی می‌تواند به عنوان یک سیستم خبره، توسط متخصصین بالینی درگیر با بیماری قلبی، با هدف کمک به تصمیم‌گیری‌های بالینی و کاهش خطاها، بهبود زمان انتظار در تشخیص بیماری و کاهش آزمایش‌های غیرضروری پزشکی استفاده گردد.

در کارهای آینده، به کارگیری روش پیشنهادی بر روی داده‌های مربوط به بیماری عروق کرونری قلب مراکز درمانی

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

داخل کشور و بررسی کارایی و تحلیل نتایج آن می‌تواند موردتوجه قرار گیرد. به علاوه، استفاده از این روش در تشخیص سایر بیماری‌ها از قبیل سرطان و دیابت نیز مفید خواهد بود. همچنین بررسی نتایج استفاده از سایر روش‌های کلاسه‌بندی در روش ترکیبی پیشنهاد می‌گردد.

References

1. Anooj PK. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. Journal of King Saud University - Computer and Information Sciences 2012;24(1):27–40.

2. Vila-Francés J, Sanchis J, Soria-Olivas E, Serrano AJ, Martínez-Sober M, Bonanad C, et al. Expert system for predicting unstable angina based on Bayesian networks. Expert Systems with Applications 2013;40(12):5004–10.

3. Bashir S, Qamar U, Khan FH, Javed MY. MV5: A Clinical Decision Support Framework for Heart Disease Prediction Using Majority Vote Based Classifier Ensemble. Arab J Sci Eng 2014;39(11):7771–83.

4. Tahmasbi H. Evaluation of Data Mining Classification Techniques and Performances on Medical Data. 4th Iran Data Mining Conference; 2010 Oct –Nov 31-1; Tehran, Iran; 2010. Persian

5. World Health Organization (WHO). The top 10 causes of death [cited 2016 Jan 1]. Available from: http://www.who.int/mediacentre/factsheets/fs310/en/

6. Shouman M, Turner T, Stocker R. Using decision tree for diagnosing heart disease patients. Proceedings AusDM '11 Proceedings of the Ninth Australasian Data Mining Conference; 2011 Dec 1–2; Ballarat, Australia: Australian Computer Society, Inc; 2011. p. 23–30.

7. Squeri A. Coronary Artery Disease-New Insights and Novel Approa-ches. Rijeka: Intechopen; 2012.

8. Nahar J, Imam T, Tickle KS, Chen YP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications. 2013;40(4):1086–93.

9. Giri D, Acharya UR, Martis RJ, Sree SV, Lim TC, Ahamed T, et al. Automated diagnosis of Coronary Artery Disease affected patients using LDA, PCA, ICA and Discrete Wavelet Transform. Knowledge-Based Systems 2013;37:274–82.

10. 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

11. Negahbani M, Joulazadeh S, Marateb HR, Mansourian M. Coronary Artery Disease Diagnosis Using Supervised Fuzzy C-Means with Differential Search Algorithm-based Generalized Minkowski Metrics. Peertechz J Biomed Eng 2015; 1(1): 006-014.

12. Tahmasbi H, Amoozgar M, Adine H. Replacement of Missing Values and its Effect on the Classification Accuracy in Medical Data Mining. Journal of Health and Biomedical Informatics 2015; 2(1):24-32. Persian

13. Sentz K, Ferson S. Combination of evidence in Dempster-Shafer theory. California: Sandia National Laboratories; 2002.

14. Rani KU. Analysis of heart diseases dataset using neural network approach. International Journal of Data Mining & Knowledge Management Process 2011; 1(5): 1-8.

15. Rajkumar A, Reena GS. Diagnosis of heart disease using datamining algorithm. Global Journal Of Computer Science and Technology 2010;10(10):38-43.

16. Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, Boghrati R, Ghandeharioun A, et al. A data mining approach for diagnosis of coronary artery disease. Comput Methods Programs Biomed 2013;111(1):52-61.

17. UCI Machine Learning Repository. Heart Disease Data Set [cited 2016 Jul 24]. Available from: http://archive.ics.uci.edu/ml/datasets/Heart+Disease

18. Bashir S, Qamar U, Khan FH. IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework. J Biomed Inform 2016;59:185-200.

19. Muthukaruppan S, Er MJ. A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. Expert Systems with Applications 2012;39(14):11657–65.

20. Samuel OW, Asogbon GM, Sangaiah AK, Fang P, Li G. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Systems with Applications 2017; 28;68:163-72.

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

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2017; 3(4): 251-258

An Expert System for Heart Disease Diagnosis Based on Evidence Combination in Data Mining

Tahmasbi Hamid Reza[1], Jalali Mehrdad[2]*, Shakeri Hassan[2]

• **Received:** 11 Jan, 2017 • **Accepted:** 16 Mar, 2017

Introduction: Coronary Artery disease is the most common type of heart disease and one of the leading causes of death in industrialized countries. The aim of this study was to design an expert system with high accuracy for Coronary artery disease diagnosis.
Methods: In this applied study, 14 features of 303 patients underwent coronary angiography were used. Dempster-Shafer theory of evidence combination was used to combine the results of three classifying methods including Decision Tree, K-Nearest Neighbor and Neural Network, in order to design a more accurate coronary artery disease diagnostic system. The data mining tool (Weka version 3.7) and C# in Net Framework environment were used for the implementation of model. The 10-fold cross-validation was used for the efficiency assessment.
Results: According to the results, mean accuracy, sensitivity, and specificity of the proposed system were 90.1%, 89.09% and 91.3% respectively. These values were higher in comparison with each of the participated classifiers in the combination. Moreover, in comparison to the similar studies, this method showed higher accuracy for the diagnosis of coronary artery disease.
Conclusion: The results of this research indicates that in the studied population, the proposed method has better accuracy in the diagnosis of coronary heart disease. This method, as an expert system, can help clinicians in making decisions, reducing clinical errors, improving the time to get a diagnostic through reducing waiting time and reducing unnecessary medical tests.

Keywords: Coronary artery disease, Expert system, Medical diagnosis, Dempster-Shafer theory, Classification

• **Citation:** Tahmasbi HR, Jalali M, Shakeri H. An Expert System for Heart Disease Diagnosis Based on Evidence Combination in Data Mining. Journal of Health and Biomedical Informatics 2017; 3(4): 251-258.

1. Ph.D. Student, Computer Engineering Dept., Neyshabur Branch, Islamic Azad University, Neyshabur, Iran
2. Ph.D. in Computer Engineering, Assistant Professor, Computer Engineering Dept., Islamic Azad University of Mashhad, Mashhad, Iran

***Correspondence:** Mashhad Branch, Islamic Azad University, Mashhad, Iran
• **Tel:** 09151046117 • **Email:** ithrt20@gmail.com

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